Emerging heterogeneous integrated photonic platforms on silicon

Abstract: Silicon photonics has been established as a mature and promising technology for optoelectronic integrated circuits, mostly based on the silicon-on-insulator (SOI) waveguide platform. However, not all optical functionalities can be satisfactorily achieved merely based on silicon, in general, and on the SOI platform, in particular. Long-known shortcomings of silicon-based integrated photonics are optical absorption (in the telecommunication wavelengths) and feasibility of electrically-injected lasers (at least at room temperature). More recently, high two-photon and free-carrier absorptions required at high optical intensities for third-order optical nonlinear effects, inherent lack of second-order optical nonlinearity, low extinction ratio of modulators based on the free-carrier plasma effect, and the loss of the buried oxide layer of the SOI waveguides at mid-infrared wavelengths have been recognized as other shortcomings. Accordingly, several novel waveguide platforms have been developing to address these shortcomings of the SOI platform. Most of these emerging platforms are based on heterogeneous integration of other material systems on silicon substrates, and in some cases silicon is integrated on other substrates. Germanium and its binary alloys with silicon, III–V compound semiconductors, silicon nitride, tantalum pentoxide and other high-index dielectric or glass materials, as well as lithium niobate are some of the materials heterogeneously integrated on silicon substrates. The materials are typically integrated by a variety of epitaxial growth, bonding, ion implantation and slicing, etch back, spin-on-glass or other techniques. These wide range of efforts are reviewed here holistically to stress that there is no pure silicon or even group IV photonics per se. Rather, the future of the field of integrated photonics appears to be one of heterogenization, where a variety of different materials and waveguide platforms will be used for different purposes with the common feature of integrating them on a single substrate, most notably silicon.

Keywords: integrated photonics; integrated optics; silicon photonics; mid-infrared photonics; optoelectronics; electrooptics; photodetectors; semiconductor lasers; optical modulators; nonlinear optics; silicon; germanium; gallium arsenide; indium phosphide; tantalum pentoxide; silicon nitride; silicon dioxide; lithium niobate.

DOI 10.1515/nanoph-2014-0024
Received September 20, 2014; accepted December 16, 2014

1 Introduction

The word “photonics” was coined over thirty 30 ago in analogy to “electronics”. Since then, several related terms, such as microphotonics, integrated photonics and optoelectronic integrated circuits (OEICs), have been adopted and commonly used in analogy to microelectronics, integrated electronics, etc. Several similarities between the two fields – including using more or less the same family of materials (most notably semiconductors), the possibility of integrating several micrometer- to submicrometer-size devices on a single chip, fabrication methods based on well-established techniques conducted in cleanroom facilities – were and still are implicit in this analogy. However, as discussed below, the microelectronic chips have been increasingly dominated by the complementary metal-oxide-semiconductor (CMOS) technology, while a ubiquitous platform for microphotonics has been elusive. Instead, different photonic materials and technologies have been developed for different purposes and if there has been any path to integration, it has been one of heterogeneously merging these different platforms. This article reviews how the field of integrated photonics has been drifting into such heterogeneous integrated solutions, especially on silicon substrates.
In layman’s terms, the story of microelectronic platforms, to a large extent, has been one of unification, while that of microphotonics has been one of balkanization. It is true that heterogeneous technologies, such as BiCMOS – a hybrid platform based on silicon-germanium (SiGe) heterojunction bipolar transistors (HBTs) and CMOS field-effect transistors – as well as III–V compound-semiconductor-based HBTs and high-electron-mobility-transistors (HEMTs) have their own important applications and market share, but the microelectronics industry, particularly at very-large scale of integration, has been dominated by Si CMOS. The reasons for this converging force of CMOS are perhaps beyond the scope of this work but briefly include the electronic design advantages of the architecture in terms of high input impedance, negligible static power consumption, self-isolation, as well as ease of layout, very high yield and scalability of the process, relative inexpensiveness of high-quality Si wafers, and last but not least availability of a native oxide for the material. It should be noted that using strained layers of SiGe in drain and source regions of p-channel MOS transistors has been pursued as one means to enhance carrier mobility, but for the most part CMOS is a homogeneous platform on Si. The younger field of microphotonics has meanwhile followed a different path in terms of the number of platforms developed for various needs.

Starting with the invention of gallium-arsenide (GaAs) “homojunction” lasers in 1962 [1], the structures used for semiconductor lasers (perhaps the most important integrated photonic device next to photodetectors) evolved into double-heterostructures [2, 3] for room-temperature operation in 1977 [4], and later to quantum-confined structures for higher performance at room temperature in 1978 [5]. However, not a single direct-bandgap material system is able to cover the wide optical spectrum range. For instance, GaAs and indium-phosphide (InP) and their ternary and quaternary alloys are used for different windows of telecommunication (telecom) infrared wavelengths of 750–1700 nm. Nitride semiconductors [indium gallium aluminum nitride (InGaN) alloys] span visible to ultraviolet wavelengths. In parallel, crystalline and amorphous dielectrics, such as quartz and silica, have been traditionally pursued in integrated optics for waveguides and other passive devices, while high-performance electrooptic (EO) modulators have been mostly fabricated on lithium niobate (LiNbO3).

This balkanization of integrated photonics suggests that unfortunately no “magical” material or material system has been developed to date, whose unique collective properties can satisfy the somewhat orthogonal needs of the modern integrated photonics, including the choice of wavelength, efficient light emission and detection, large EO coefficient, strong optical nonlinear properties, ease of fabrication, integrability, scalability, reliability, inexpensiveness, etc.

2 Can silicon homogenize integrated photonics?

In the last quarter of century or so, silicon has been promoted and aggressively pursued as the elusive unifying material for integrated photonics. The field started by a series of seminal papers by R. A. Soref and coworkers [6–10]. The argued advantages of silicon over III–V OEICs have been the possibility of demonstrating much more compact optical waveguides with lower optical propagation loss, better processing yield, low cost and the potential compatibility with CMOS electronics. Several review papers [11–15] and books [16–20], have been written on silicon photonics, including some works by this author. For completeness, the basics of this exciting technology are briefly reviewed in the next paragraphs.

Figure 1 shows the idea of a silicon-based “superchip” [10]. Clearly, incorporating laser diodes, HBTs, HEMTs and BiCMOS devices on other materials are implicitly envisioned, i.e., the need for heterogeneous integration was acknowledged from the early days of silicon photonics. This vision is to a large extent in agreement with the state of the art, as seen throughout this paper. Figure 2 eloquently shows the importance and advantages of the versatile silicon-on-insulator (SOI) waveguide platform [11], which has been progressing rapidly since 1991 [21, 22]. The SOI platform has effectively become the benchmark platform for silicon photonics and several devices and photonic integrated circuits have been implemented on it since the mid-1990s [23–27]. The standard SOI waveguide technology is schematically represented later in the left structure of Figure 4. A thin film (220 nm to 3 μm) of silicon sits on top of a silicon dioxide (SiO2) insulator layer, the so-called buried oxide or BOX, that serves as a lower-index bottom cladding layer for vertical optical confinement. The BOX layer is thick enough to prevent mode leakage into the silicon substrate. Techniques such as Separation by Implantation of OXygen (SIMOX) process [28], or Bond and Etch-back SOI (BESOI) [29] can be employed to produce the wafers, but hydrogen implantation followed by thermal wafer splitting (the SmartCut™ process to produce Unibond® wafers [30]) has perhaps been the most successful approach to produce SOI wafers. An interesting review of these SOI wafer production methods can be found elsewhere [31].
A silicon-based solution for photonics would certainly be very rewarding and revolutionary. Hence, numerous researchers around the globe have been intensely working on advancing the field on various fronts with the ultimate hope of achieving a reliable technology that can be utilized as a universal platform for various optical functionalities and ultimately be integrated with microelectronic circuits monolithically on a single silicon chip. The endeavor has certainly been successful to the extent that a handful of companies have been marketing a variety of OEIC products for some years, most notably optical transceivers for data communications (datacom), Ethernet and other short-haul applications.

However, it was clear from the beginning that there is no “pure silicon” photonics, i.e., other materials are inevitably required. First, silicon does not absorb light at wavelengths above the material’s bandgap (~1.1 μm). Instead, Ge or SiGe photodetectors have been employed, as reviewed in Section 3. Accordingly, it has been argued that group-IV photonics is a more accurate name than silicon photonics.

Second, the bandgap of bulk crystalline Si is indirect, i.e., the light emission process is phonon-mediated and the associated spontaneous recombination is very long, leading to a very low internal quantum efficiency [13, 32]. Consequently, achieving lasing in bulk Si has not been possible. Several approaches have been alternatively pursued. Unfortunately, for a variety of reasons, room-temperature electrically-injected silicon-based lasers remain elusive to date [13]. A pragmatic solution has been the hybridization of silicon photonic chips with III–V lasers, as reviewed in Section 4.

In reality and with the lack of any other practical solution, the silicon-photonic companies typically incorporate III–V lasers and germanium photodetectors, one way or the other, and silicon is mostly used for passive devices and modulators. For instance, Figure 3 shows a schematic of how one of Luxtera Corporation’s early optical transceiver chips incorporates such non-silicon devices [33]. As depicted in Figure 4, it can be argued that “groups III-to-V” or “groups III-IV-V” photonics are semantically more accurate than silicon or group IV photonics for the established commercial technology. It is noted that the schematics of Figure 4 are very crude and generic and many variations exist, as seen throughout this paper. For instance, two cases for III–V monolithic
integration is shown with respective pros and cons described in Section 4.

Third, despite a lot of effort for using crystalline silicon as an active region because of its high third-order optical nonlinearities, the omnipresent optical loss induced by two-photon- and free-carrier absorptions (TPA and FCA) has been very problematic [13, 34, 35]. Using thin films of other nonlinear materials “hybridized” on silicon substrates has been chipping away another core prospect of the SOI platform. These heterogeneous platforms for third-order nonlinear applications are reviewed in Section 5.

Fourth, although silicon optical modulators operating based on the free-carrier plasma effect have come a long way and their performance has been steadily improving, the inherent tradeoff between modulation bandwidth and depth remains a challenge and the devices cannot compete with telecom EO modulators in this regard. Hybridization of thin films of other materials known for their strong EO effect, e.g., LiNbO$_3$ on silicon substrates could be a possible route forward. Such materials also possess strong second-order optical nonlinearity and may enable compact chips for entangled photon generation and other quantum-optic-related applications. Efforts on such materials are reviewed in Section 6.

Fifth, silicon is an ideal material for nonlinear integrated optics at wavelengths above \( \sim 2.2 \mu m \). That is, the wavelength range in which TPA vanishes and the material becomes transparent even at high optical intensities, until multiphonon absorption kicks in and the loss increases again above \( \sim 6.5 \mu m \). The 2.2–6.5 \( \mu m \) window covers the longer portion of the traditionally-defined short-wave length infrared (SWIR of 1.4–3.0 \( \mu m \)) range as well as the shorter portion of the mid-wavelength infrared (MWIR of 3.0–8.0 \( \mu m \)). Using silicon for integrable devices in the atmospheric window of 3–5 \( \mu m \), and sometimes the wider 2.2–6.5 \( \mu m \) range, has been loosely defined as mid-infrared (or mid-IR) silicon photonics. However, the BOX layer of the standard SOI technology is lossy across most of this range and hence other heterogeneous platforms such as silicon-on-sapphire, silicon-on-nitride, and air-clad waveguides have been developed, as reviewed in Section 7.

All in all, with the lack of any integrated optical material system for a whole range of applications, researchers have been reluctantly accepting that, despite their challenges and disadvantages, heterogeneous platforms appear to be the practical way forward. Still, silicon remains one of the most attractive substrate materials due to its low-cost, availability of large wafers, ease of handling and thermal

Figure 3 Images of optoelectronic chips produced by Luxtera Corporation and how (A) III–V lasers and (B) Ge photodetectros are integrated with Si waveguides [33].

Figure 4 Simple schematic of the SOI waveguide platform, and how germanium and III–V compound semiconductor layers ought to typically be integrated in order to achieve hybrid photodetectors and lasers, respectively, suggesting that in reality there exists no pure silicon or even group IV photonics, rather “groups III-to-V photonics” or “groups III-IV-V photonics” are semantically more accurate.
cycling, and well-established processing techniques, as well as possibility of integrating with CMOS electronics. The philosophy is to use silicon as much as possible but hybridize it with other materials when better performance can justify the challenges of hybridization.

3 Germanium photodetectors on silicon

As mentioned in the Introduction, the least controversial shortcoming of the benchmark SOI platform is its incapability of absorbing photons with $\lambda > 1.1$ $\mu$m. For the O- to U-bands of telecom wavelengths (1.26–1.67 $\mu$m), Germanium’s cutoff wavelength of $\sim 1.8$ $\mu$m appeals as a perfect choice. Ge-on-Si photodetectors have been in the making for three decades [36–51].

The challenge has been the material’s large lattice mismatch with silicon ($\sim 4\%$). Thick Ge layers would lead to surface roughness, threading dislocations and increased dark photocurrent. Also, Ge does not have a stable oxide and passivating it to avoid surface recombination – which further increases the dark current – is another challenge. Dislocations can be controlled by graded SiGe buffered layers or in thin strained layers, both of which can reduce the dark current. Waveguide photodiodes with thin strained Ge layers on Si are more appealing for integrated chips. The small waveguide dimensions also result in shorter carrier transit time and increased detection bandwidth. The tradeoff is the length of the devices, that can be tens of micrometers long in order to increase the responsivity. One of the lowest reported dark current densities is $\sim 40$ mA/cm$^2$, achieved in a 1.3-$\mu$m-wide and 4-$\mu$m-long waveguide at $-1$ V of bias [51]. The corresponding responsivity and 3-dB bandwidth are 0.8 A/W and 45 GHz, respectively. Evidently and despite all the efforts, relatively high dark current remains a disadvantage for Ge-on-Si waveguide photodetectors as compared with normal-incidence counterparts and III–V detectors.

To highlight the fabrication challenges of heterogeneous integration of Ge on Si substrates, the processing steps of one of the earlier high-speed photodetectors (40 Gb/s) by Intel Corporation is presented here [44]. Figure 5 shows schematics of the device, how it is integrated with a SOI waveguide, and its lateral cross-section. The following processing steps were conducted in the selected region where the photodetector was intended to be fabricated: (a) implanting the substrate with boron to reduce parasitic capacitance; (b) oxide deposition and patterning for defining the Ge region; (c) selective epitaxial growth of 1.3-$\mu$m-thick Ge; (d) chemical-mechanical polishing (CMP) of $\sim 0.5$ $\mu$m of the grown Ge; (e) thermal annealing to reduce the threading dislocation density; (f) $n$-doping of the Ge layer; (g) selective excess $p^{++}$-doping of silicon for improved ohmic contacts; (h) dopant activation by rapid thermal annealing; (i) metallization and patterning to form the diode contacts.

Figure 5 (A) Side-view, (B) cross-section and (C) micrograph of one of the early high-speed Ge waveguide photodiodes developed at Intel Corporation [44].
The extra steps for Ge hybridization are the selective epitaxial growth, CMP, and thermal annealing. Such extra, sometimes complicated, fabrication steps are typically the price tag for any heterogeneous technology, more examples of which are discussed in this paper. Meanwhile, one advantage of Ge, over other materials to be discussed, is that, despite its large lattice mismatch, thin films of it can be epitaxially grown on Si with little complications. Not all other materials can be as easily grown on Si, and hence other heterogeneous integration techniques, such as wafer- and flip-chip-bonding, ought to be devised as seen in the following sections.

4 III–V Lasers on silicon

Attaining lasing is perhaps the holy grail of silicon photonics. Unlike compound semiconductors, bulk crystalline Si is an indirect bandgap material and light emission is mediated by phonon emission or absorption. This second-order emission process is very weak (with an internal quantum efficiency of \(\sim 10^{-6}\)) and the spontaneous recombination lifetime is in the millisecond range [13]. In principle, silicon can provide high material gain at high carrier injection levels for lasing. However at high injection levels, FCA and Auger recombination prevent achieving any net optical gain [13, 32].

Since the early 1990s, several approaches have been pursued to increase the quantum efficiency of luminescence in bulk Si. They include, but not limited to, Erbium (Er)-doped Si, Si-rich oxide or Si nanocrystals (NCs), Er-doped Si NCs, deposition of boron dopant with SiO2 nanoparticles mix on Si, surface texturing, dislocation loops, and stimulated Raman scattering. However, for a range of specific reasons beyond the scope of this paper and reviewed elsewhere by B. Jalali and this author in 2006 [13], unfortunately none of these approaches have been able to offer room-temperature electrically-injected lasers. Since then, an interesting development has been room-temperature lasing in strained, \(n\)-type-doped Ge on Si [52]. Strain makes Ge an “almost” direct bandgap material and \(n\)-doping populates the engineered direct conduction band. Evidence of room-temperature electrically-injected lasers at a high doping level of \(4\times10^{19}\) cm\(^{-3}\) and tensile strain of about 0.2% in the 1520–1700 nm range was reported in 2012 [52]. However, a high optical loss of 90 cm\(^{-1}\), induced by high doping levels right in the middle of the active region of the waveguides, results in high threshold currents and the approach does not appear to be ready for prime-time applications yet. Another recent photonic material development has been germanium tin (GeSn), a group IV binary compound semiconductor alloy with direct band-gap below 0.8 eV at certain compositions. No GeSn laser has been demonstrated yet, but electroluminescence devices spanning \(\sim 1530–1900\) nm wavelengths have been reported [53].

The disappointing reality of the lack of a viable approach for achieving Si-based lasers has not affected the attractiveness of silicon photonics. Instead, hybrid and heterogeneous approaches have been aggressively pursued. The most straightforward hybrid approach that comes to mind is to flip-chip bond or solder-based assembly a completely grown and processed functional III–V laser onto the silicon photonic circuit and couple the laser output into silicon waveguides via some coupling mechanism. These approaches are indeed as old as silicon photonics itself [54–57]. Variations of this approach are fairly established and adopted by the silicon photonic industry, e.g., Luxtera Corporation [33, 58], NEC Corporation [59], Toshiba Corporation [60], and Fujitsu Laboratories [61]. Perhaps such more straightforward per-device flip-chip bonding approaches suffice the present low-volume, low-throughput needs of the industry but for future high-volume demanding markets, heterogeneous monolithic integration of III–V wafers and Si wafers, or already processed circuits, is more desirable.

Epitaxial growth of III–V heterostructures on Si substrates has been pursued for over 25 years [62]. Similar to Ge-on-Si epitaxial growth, however, the challenge is the large lattice mismatch (4.1%) between GaAs and Si, as well as 250% difference of their thermal expansion coefficients. Direct growth of GaAs-based heterostructures on Si creates very high densities of surface misfit dislocations \(\sim 10^{10}\) cm\(^{-2}\), which transform into threading dislocations propagating into the active region of lasers during the growth. Consequently, the lasers have very short lifetime caused by catastrophic failure. It turns out that the lattice mismatch of Ge and GaAs is very small (0.077%) and their thermal expansion coefficients match over large temperature ranges. Hence, the same graded SiGe epitaxial growth technology discussed for SiGe photodetectors can be utilized as buffer layers in order to achieve low-dislocation GaAs-based epilayers on Si. A challenge has been the growth of a polar (GaAs) over a nonpolar (Si) surface, a phenomenon long known as antiphase disorder [63]. Certain substrate preparation and initial growth precautions can significantly alleviate the issue [64–67]. AlGaAs/GaAs quantum well lasers on relaxed SiGe buffers layers on Si have been reported [68]. Surface threading dislocation densities were as low as \(2\times10^{6}\) cm\(^{-2}\) allowing continuous-wave, room-temperature lasing at a wavelength of
858 nm. Yet, relatively low differential quantum efficiency of 0.24 and high threshold current density of 577 A/cm² impaired the devices.

Similarly, other approaches for growing GaAs-based laser on Si, including strained-layer superlattices (SLS) [69, 70] and more recently quantum dot dislocation filters (see Figure 6A) [71, 72], typically suffer from high threshold current and short lifetimes. At any rate, GaAs-based lasers, even their longer-wavelength quantum-dot dot types, become increasingly less efficient light emitters at wavelengths above silicon’s bandgap, and InP-based lasers are more suitable for integrating with Si and for telecom wavelengths. In addition to the thermal expansion coefficient difference and polar/nonpolar antiphase disorder issues, the problem with direct InP growth on Si is the even larger (8.1%) lattice mismatch between the materials, prohibiting using graded SiGe buffers. Even employing InGaAs SLS buffer layers is not as effective as in GaAs. Initial works resulted in high dislocations densities in the order of 10⁸–10⁹ cm⁻² [76]. Progress was later made for this method [77–79], yet the dislocation densities remain in the 10⁷ cm⁻² range, which is an order of magnitude higher than lattice-matched growth. A somewhat radical solution for direct growth of III–V lasers emitting in the 1550-nm telecom wavelength is using another material system, e.g., gallium antimonide (GaSb)-based lasers. Strained GaInSb/AlGaAsSb quantum-well lasers have been grown on Si, but at room temperature the devices could operate only under pulsed bias and with a high threshold current density of 5 kA.cm⁻² [80].

The tough circumstances may leave no option other than relying on some bonding techniques. As discussed, a very practical, yet low-yield and high-cost, approach is flip-chip bonding of completely processed individual lasers on silicon photonic chips. A more monolithic approach for large-scale integration is to first conduct wafer (full or partial) or die bonding of III–V epilayers onto silicon photonic chips, followed by processing the III–V devices. An advantage is that aligning III–V devices with Si waveguides during lithography is typically easier this way than during chip bonding or assembly.

Indirect bonding methods, including adhesive (e.g., based on polymers) and eutectic bonding (based on metals such as In–Au and Pb) may be convenient ways to effectively “glue” the two wafers. However, bonding by polymer materials, such as the benzocyclobutene (BCB) family [81], may be fine for lab demonstration but is not reliable for practical applications due to weak formed

---

**Figure 6** Schematics of four heterogeneous approaches to integrate III–V laser on Si substrates: (A) monolithic epitaxial growth of InGaAs/GaAs quantum-dot regions buffered by 10 layers of inactive dots to filter propagation of threading dislocations into the active region [72]; (B) evanescently-coupled InP lasers bonded on SOI waveguides [73]; (C) tape-waveguide vertical coupling form III–V gain regions into SOI [74]; and (d) bonding III–V gain regions into trenches made in silicon for planar coupling between the two [75].
bonds. Also, metallic layers of eutectic bonding cause optical loss.

A more reliable approach is direct wafer (or fusion) bonding of InP and Si layers. The idea can be traced back to Lord Rayleigh [82], who bonded optically-polished glass plates at room temperature. In modern semiconductor processing, two atomically smooth (<1 nm roughness) and clean surfaces are brought to physical contact at low temperatures (to form atomic bonds with no worry of the lattice-mismatch hurdle) and subsequently the bonds are strengthened by thermal annealing. The research in this area was motivated by bonding two Si wafers in the 1980s for electronic applications, pioneered by researchers at IBM [29] and Toshiba [83]. Interestingly, Ref. [29] reports one of the very first SOI wafers achieved by the aforementioned bonding and etch-back technique.

Taking advantage of hydrophilic surfaces (e.g., SiO₂) or assisting the process by oxygen plasma can enhance the initial bonding, as follows. One of the earliest III–V lasers bonded to Si wafers is by C. Seassal et al. [84], with Ecole Centrale de Lyon and LETI-CEA in France, where they managed to demonstrate optically-pumped InP microdisk lasers bonded to Si wafers. Two SiO₂ layer were formed by plasma-enhanced chemical vapor deposition (PECVD) on both InP and Si wafers, and then polished, cleaned, bonded and annealed at 200°C for 60 minutes to increase the bonding strength. The InP substrate was subsequently removed by selective wet etching using an In₀.₅₃Ga₀.₄₇As sacrificial (etch-stop) layer, exposing the heterolayers for microdisk laser processing, i.e., 6-μm diameter disks were patterned by standard dry etching techniques. An electrically-injected InAsP/InGaAsP microdisk laser was later demonstrated [85]. One issue with this approach is the quality of the PECVD oxide and the strength of the created bonds and their stability at elevated temperatures [86]. Low heat dissipation of the underlying dielectric SiO₂ layer [see case (a) in Figure 4], leading to degraded laser performance, is another disadvantage of using hydrophilic SiO₂ bonding layers.

One challenge for wafer-bonding approaches is how to couple light into Si waveguides. Based on the concept of vertical coupling via evanescent coupling (originally proposed and demonstrated in the context of III–V microdisk lasers stacked on another III–V bus waveguide [87]), the mentioned French research groups managed to couple light from optically-pumped microdisk lasers into silicon strip waveguides by decreasing the total SiO₂-SiO₂ hydrophilic surfaces from 800 nm down to 300 nm [88].

Pioneered by a group at Uppsala University, Sweden, plasma-assisted direct bonding of semiconductors has been shown to be an interesting technique to avoid SiO₂, hydrophilic layers and their complications [86, 89–91]. Based on this bonding technique, researchers at the University of California, Santa Barbara (UCSB), demonstrated optically-pumped [92] and later the first electrically-injected [93] evanescently-coupled edge-emitting Fabry-Perot AlGaInAs/InP lasers coupled into Si waveguides (see Figure 6B). A hybrid optical mode, mostly residing in the SOI waveguide, has partial overlap with the laser separate confinement structure, allowing coupling between the two modes. Since then and in collaboration with Intel Corporation, the UCSB group has impressively matured this platform and managed to demonstrate more complicated photonic circuitry, i.e., demonstrate and integrate lasers [including Fabry-Perot, microdisk, microring and distributed Bragg reflector (DBR) types], semiconductor optical amplifiers, electroabsorption and phase modulators, as well as photodetectors. Detailed reviews of these works can be found elsewhere [73, 94]. Also, other researchers have shown very similar lasers with InGaAsP, rather than AlGaInAs, multi-quantum well active regions [95].

Meanwhile, arguing that the evanescently-coupled lasers should better have more confinement in the InP-based gain region than the Si waveguides, research groups at Ghent University and IMEC, Belgium, in collaboration with the mentioned French groups, demonstrated somewhat different electrically-pumped evanescently-coupled lasers [96]. Another approach pursued by the same groups is using III–V gain regions and distributed reflectors on SOI waveguides linked by tapered adiabatic mode converters for realizing tunable lasers (see Figure 6C) [74, 97]. Finally, Skorpios Technologies has envisioned and developed a planar platform in which unprocessed III–V epilayers are metal-bonded into grooves made in silicon substrate, such that the processed laser is eventually butt-coupled into SOI waveguides [75] (see case (b) in Figure 4 and Figure 6D). This approach avoids the inherent shortcoming of vertically-coupled lasers in terms of guided optical mode distribution between Si and InP waveguides and offers efficient direct optical coupling between the two planar (horizontally-aligned) regions. Other advantages include better heat dissipation due to the omission of the BOX underneath the laser, allowing laser operation up to 80°C, as well as hermetically sealing the lasers in SiO₂ encapsulation, eliminating the need for their packaging.

To summarize this Section, it is apparent that any existing and perhaps upcoming laser technology on Si has its own pros and cons. Thus, it is hard to judge which of so many reviewed competing approaches will prevail. But it is likely that some heterogeneous integration methods, based on III–V lasers pre- or post-process bonded to Si, will be the most relevant solutions in the foreseeable future.
5 Third-order nonlinear photonics on silicon

Silicon has a centrosymmetric lattice structure, thus it inherently lacks second-order optical nonlinearity. It has high third-order optical nonlinearity (both Kerr and Raman types), though. Specifically, silicon waveguides have a much high nonlinear parameter \( \gamma = (\omega \times n_2)/(c \times A_w) \) compared to traditional integrated-optics platforms, e.g., silica fibers or planar waveguides. Here, \( n_2 \) is the nonlinear refractive index, \( \lambda \) and \( \omega \) are the wavelength and angular frequency of light, respectively, and \( c \) is the speed of light in vacuum. From the early 2000s, pioneered by B. Jalali’s group at UCLA, both Raman-based [98–103] and Kerr-based [104, 105] nonlinear effects have been explored in silicon photonics. Numerous studies have been published in the area of nonlinear silicon photonics by several groups, reviews of which can be found elsewhere [13, 106]. An omnipresent problem in the telecommunication wavelengths, however, is the TPA at required high optical intensities and the even higher FCA induced by it [13, 34, 35]. Certain approaches can reduce the carrier lifetime, alleviate the problem and modestly improve the nonlinear device performances. Active carrier sweep-out in reverse-biased waveguide diodes has been proposed [35] and demonstrated [107]. This author and coworkers demonstrated energy harvesting based on two-photon photovoltaic effect to achieve energy-efficient nonlinear devices [108–112]. Nonetheless, in all of these sweep-out-based solutions, the carrier lifetime remains relatively high and FCA cannot be significantly subdued due to free-carrier screening of the junction electric field [113]. Difficulties of coupling light into higher order modes of specially-designed waveguides notwithstanding, cladding pumping of nonlinear silicon photonic devices is another potential solution [114]. Also, introducing midgap states through high-energy irradiation of ions, e.g., helium, is another means of modestly reducing the carrier lifetime, although it comes at the expense of higher linear propagation loss. Therefore, only small values of net gain in Raman amplifiers has been observed [115].

Several researchers have totally moved to other nonlinear materials such as hydrogenated amorphous silicon (a-Si:H) [116–118], silicon nitride (SiN) [119–121], Hydex\textsuperscript{TM} [122, 123], chalcogenide glass [124–126], and tantalum pentoxide (Ta\textsubscript{2}O\textsubscript{5}) [127–132] waveguides on silicon substrates to totally circumvent the nonlinear loss problem of SOI waveguides. The argument is that the improved nonlinear figure of merit, \( \text{FOM} = n_2/(\beta_{\text{TPA}} \times \lambda) \) in these materials – with negligible TPA coefficient, \( \beta_{\text{TPA}} \) – is more important than the previously argued high \( \gamma \) of Si. The common idea for all of these heterogeneous platforms is to use deposition, spinning or other methods of thin film formation to create waveguides on a dielectric bottom cladding layer (usually SiO\textsubscript{2}) on Si substrates and exploit the materials’ nonlinear optical properties for functions such as supercontinuum generation, four-wave mixing, self-phase modulation, cross-phase modulation, third-harmonic generation, stimulated Raman scattering, etc. To give an example of such materials, a novel platform based on Ta\textsubscript{2}O\textsubscript{5} developed at the author’s lab at CREOL is reviewed in the following [131, 132].

Table 1 summarizes \( n_2 \) and \( \beta_{\text{TPA}} \) in all the materials discussed in the previous paragraph in comparison to SiO\textsubscript{2}, as a benchmark. The literature of a-Si: H is inconsistent and seemingly is very dependent on the deposition condition, but very high values of \( n_2 \) have been claimed [116, 118]. However, a tradeoff between \( n_2 \) and \( \beta_{\text{TPA}} \) is observable and overall the considerable value of \( \beta_{\text{TPA}} \) (which can be much higher than crystalline Si [116]) limits the FOM (e.g., \( \sim 5 \) in Ref. [118]). The other materials of interest in Table 1 have negligible \( \beta_{\text{TPA}} \) and hence measurably or potentially possess very high FOM. Among them, Ta\textsubscript{2}O\textsubscript{5} has an \( n_2 \) three times higher than SiN and comparable to chalcogenide glasses. Meanwhile, its refractive index of \( \sim 2.1 \) is higher than Hydex’s (\( \sim 1.7 \)) and its damage threshold is much higher than of that chalcogenide glasses. The reduction of the waveguide dimensions in the low-loss Ta\textsubscript{2}O\textsubscript{5} sub-micron waveguide fabrication technique described below reduces \( A_w \). Consequently, the \( \gamma \) of Ta\textsubscript{2}O\textsubscript{5} waveguides can be much higher than the competing materials of Table 1.

Sputtering and standard lithography and reactive ion etching processes have been previously used to form Ta\textsubscript{2}O\textsubscript{5} waveguides [128, 130]. However, sputtered waveguides

<table>
<thead>
<tr>
<th>Material</th>
<th>( n_2 \times 10^{-20} \text{ m}^2/\text{W} )</th>
<th>( \beta_{\text{TPA}} \text{ (cm/GW)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO\textsubscript{2}</td>
<td>3.2</td>
<td>Negligible</td>
</tr>
<tr>
<td>Hydex</td>
<td>11</td>
<td>Negligible</td>
</tr>
<tr>
<td>SiN</td>
<td>26</td>
<td>Negligible</td>
</tr>
<tr>
<td>Ta\textsubscript{2}O\textsubscript{5}</td>
<td>72.3</td>
<td>Negligible</td>
</tr>
<tr>
<td>Chalcogenide</td>
<td>(-100)</td>
<td>(6.2 \times 10^{-4})</td>
</tr>
<tr>
<td>Crystalline Si</td>
<td>250</td>
<td>0.7</td>
</tr>
<tr>
<td>a-Si:H</td>
<td>2100–4200</td>
<td>0.25–4.1</td>
</tr>
</tbody>
</table>
suffer from high propagation loss, e.g., as high as 8.5 dB/cm in 18-μm-wide waveguides [130]. Also, compact (sub-micron) \(\text{Ta}_2\text{O}_5\) waveguide are not easily feasible based on such conventional schemes. Based on our proposed scheme, called selective oxidation of the refractory metal (SORM) Tantalum (Ta), we have recently demonstrated the first high-index contrast \(\text{Ta}_2\text{O}_5\) submicron waveguides and microring resonators on Si substrates (see Figure 7) [131, 132].

The improved fabrication process of Ref. [132] starts with sputtering Ta, rather than \(\text{Ta}_2\text{O}_5\), and making stripes out of the refractory metal by electron-beam lithography and chlorine-based selective etching. The etching recipe is highly selective on Ta over the photoresist, allowing smooth sidewalls and hence low propagation loss. The Ta stripes are consequently oxidized to form \(\text{Ta}_2\text{O}_5\) channel waveguides. Ridge waveguides can be easily achieved too by two iterations of Ta sputtering, the first of which is fully oxidized to form a \(\text{Ta}_2\text{O}_5\) slab region. Waveguides with propagation loss of 4.9 dB/cm, microring resonators with loaded quality factor of \(6 \times 10^4\) and grating couplers for convenient in- and out-coupling are demonstrated.

Figure 7 summarizes some of the demonstrated waveguide properties [132]. These demonstrations suggest that \(\text{Ta}_2\text{O}_5\)-on-Si is potentially a viable platform for high-performance nonlinear integrated photonic applications.

### 6 Second-order nonlinear photonics on silicon

There have been some reports on breaking the lattice symmetry of bulk Si by applying stress on Si waveguides to take advantage of the induced second-order optical nonlinearity as well as linear EO (or Pockels) effect [134–136].

![Diagram](image.png)

**Figure 7** (A) Micrograph of \(\text{Ta}_2\text{O}_5\) waveguides on Si; (B) measured and simulated transmission spectrum of grating couplers on the platform; (C) measured spectral response of the microring resonators fabricated on the platform in the 1550 nm range; (D) zoomed version of one of the resonances in (C), also showing fitted data to it [132].
In this respect, this author and coworkers have been able to induce electronically-tunable birefringence in SOI waveguides by blanketing them with a piezoelectric transducer made of lead zirconium titanate (PZT) [137, 138]. However, the amount of induced birefringence is in the range of $3 \times 10^4$, which is large enough for electrically tuning the phase-mismatch of transverse-electric (TE) and transverse-magnetic (TM) modes in parametric nonlinear optical effects, but not enough to take advantage of the effect for second-order nonlinear effect itself. It is true that the induced strain can be higher using stressed cladding layers, such as SiN, and in smaller waveguide cross-sections [134–136], but overall the extent to which second-order optical nonlinearity and/or Pockels effect can be induced by such methods is questionable considering that single-element Si is a nonpolar crystal.

As far as optical modulation function in silicon photonics is concerned, the free-carrier plasma effect was proposed in the early works by Soref et al. [8–10] as a viable alternative to the Pockels effect. In essence, carrier injection, depletion or accumulation in $p$-$n$ junction diodes or MOS capacitors, typically incorporated into Mach-Zehnder interferometers (MZI) or microrings, can be used to modulate the refractive index or optical loss. Numerous works on such modulators have been reported, a review of which can be found elsewhere [139]. Although Si optical modulators with adequately high bit rate (~50 Gb/s) and low half-voltage length-product $V_{NL}L$ (~2.8 V.cm) have been reported in MZIs, the devices only offer modest (3–5.5 dB) of modulation depth or extinction ratio [140, 141]. Such low modulation depth may be enough for near-term very short-haul links but this certainly remains a shortcoming for the long-term requirements of the industry, where at least 7 dB of modulation depth at high speeds is demanded [139]. Indeed, the modulation depth times the bit rate may be a better figure of merit, which is ~150 to ~280 dB.Gbs$^i$ in the mentioned state of the art works [140, 141]. In clear contrast, the benchmark LiNbO$_3$ modulators boast 20 dB or better modulation depth (the industry standard for digital communications) [142] and 40 Gb/s or higher bit rates are commonplace in commercial long-haul telecom systems, giving a figure of merit of at least 800 dB.Gbs$^i$. Up to 100 GHz modulation bandwidth has been measured in LiNbO$_3$, with careful RF design of traveling-wave electrodes as early as 1998 [143], hence even higher modulation depth times bit rates are feasible in principle, the complications of detection and electronic circuitry for modulation and detection notwithstanding.

It again appears that heterogeneous integration of Si with another material with strong second-order optical nonlinearity, as well as linear EO or electro-absorption (EA) effect, may be the way forward. Nitride semiconductors have high second-order nonlinearities. GaN has been bonded to Si by PECVD SiO$_2$ with measured second-order nonlinear susceptibility, $\chi^{(2)}$, of ~16 pm/V [144]. However, as mentioned in Section 4, the quality of PECVD oxide is low for bonding. Also, the uniformity of the thin GaN films obtained by CMP, after removing the (111) Si substrate, on which they were originally grown, is an issue. Sputtering of AlN on Si wafers coated with SiO$_2$ has also been reported [145]. However, the crystalline quality of sputtered AlN is expectedly low, leading to a weak $\chi^{(2)}$ of ~4 pm/V. Nonlinear polymer materials have shown strong EO effect and, their long-term reliability notwithstanding, have potential for high-performance modulators in Si waveguides coated with them [146].

Telecom wavelengths are far from silicon’s bandgap. Hence, EA modulators based on the Franz-Keldysh effect (FKE) are not possible on Si. But both the quantum-confined Stark effect (QCSE) and FKE in SiGe can be exploited at longer wavelengths. Si$_{0.6}$Ge$_{0.4}$-Si quantum-well waveguide EA modulators based on QCSE operating at 1.15 μm wavelength were demonstrated at the University of Michigan in 1998 [147]. Later, another team employed Ge-Si$_{1-x}$Ge$_x$ quantum wells in surface-illuminated modulators to red-shift the operating wavelength to ~1.45 μm [148]. More recently, GeSi, i.e., Ge epitaxial layer comprising <1% of silicon, has been used to exhibit strong FKE around 1.55 μm and demonstrate EA modulators with 10 dB extinction ratio [149]. Mellanox Technologies has been recently maturing this technology to integrate it into their OEICs [150]. Another approach, pursued at UCSB, is to take advantage of the quantum-well intermixing technique in order to shift the bandgap of InP-based heterostructures and achieve EA modulators on the same elipayers used for lasers in a hybrid OEIC on Si substrates [73].

An alternative approach pursued at the author’s group at CREOL is heterogeneous integration of LiNbO$_3$ on Si [151]. With high nonlinear optical coefficients (e.g., $d_{31}=25.2$ pm/V), LiNbO$_3$ is an ideal material for nonlinear integrated photonics. Also, single-crystalline LiNbO$_3$ has large EO coefficients ($r_{31}=31$ pm/V and $r_{33}=8$ pm/V) and a wide transparency range (0.4–5 μm wavelength) [152]. Indeed, standard LiNbO$_3$ waveguides are widely regarded as the best vehicle for EO modulation in the photonic industry [142] with impressively high modulation bandwidths (over 100 GHz [143]). As mentioned before, LiNbO$_3$ modulators undoubtedly offer higher performance (in terms of modulation bandwidth and modulation depth) compared to Si-based optical modulators.

LiNbO$_3$ waveguides are traditionally formed by diffusion of titanium, the process of annealed proton exchange
or implantation of dopants (e.g., oxygen ions) in bulk wafers [142]. Any of these processes can only slightly alter the refractive index of the material, i.e., the index contrast of the obtained stripe waveguides is rather small (0.1–0.2) and hence the guided modes are weakly confined. In EO modulators, this shortcoming leads to large device cross-sections (widths of several microns) and consequently large $V_{\text{fL}}$ of 10–20 V cm depending on modulation frequency and characteristic impedance [143, 153]. Indeed, for current very high-speed applications, the frequency response of LiNbO$_3$ modulators is limited by the long lengths of the electrodes, thus using travelling-wave electrodes becomes essential [142, 143]. Large cross-section waveguides also imply the need for high-power sources to achieve high optical density for the onset of optically-induced $\chi^{(2)}$ nonlinearity. Furthermore, for weakly-confined waveguides the bending loss becomes significantly high for small radii. Consequently, the possibility of integrating several of the devices on the same chip for advanced vector modulation formats, e.g., quadrature amplitude modulation (QAM), becomes difficult and cost-prohibitive when the number of MZIs exceeds a few. Monolithic integration of only four LiNbO$_3$ MZIs for 16-level QAM modulation with a large module size of 90×13.5×7 mm$^3$ is the state of the art [154].

Achieving thin films of LiNbO$_3$ on a material like SiO$_2$ with much lower index and a method for submicron ridge or channel formation (instead of the traditional stripe type) can potentially alleviate the discussed problems and yield high-contrast submicron waveguides. However, fabrication of high-quality thin films single-crystalline form of LiNbO$_3$ has been proven to be very challenging. Chemical vapor deposition [155], sol-gel processing [156], RF sputtering [157–159], and pulsed-laser deposition [160, 161] have been pursued with little success. That is, the EO coefficients are much smaller for films obtained by any of these methods typically have low optical quality, i.e., the EO coefficients are much smaller (up to more than 10 times) than the mentioned bulk values.

Alternatively, crystal ion slicing has been successfully employed to obtain high-quality thin films of crystalline LiNbO$_3$ [162–165]. A combination of deep helium ion implantation, selective etching and epoxy-resin bonding has been applied to transfer 5-to-10-μm-thick LiNbO$_3$ layers to silicon substrates [162]. LiNbO$_3$-on-LiNbO$_3$ thin films and modulators, buffered by a SiO$_2$ layer, attained by ion implantation and wafer bonding have been also reported [163, 164]. Interestingly, the EO properties of the films was not degraded during the implantation process. However, the argon ion beam milling process employed to achieve ridge waveguides causes high propagation loss. In another report, the adhesive polymer BCB, was used instead for wafer bonding and demonstrate electrooptically tunable LiNbO$_3$-on-LiNbO$_3$ microring resonators [165]. Polymers, like epoxy-resin [162] or BCB [165], may be simple and straightforward solutions for lab demonstrations, but, as mentioned before, they are generally not reliable for practical applications due to the weak nature of the formed bonds.

The author’s group has been able to address all the above challenges and for the first time demonstrate reliable LiNbO$_3$-on-Si thin films as well as low-loss waveguides, microring resonators and record-low $V_{\text{fL}}$ modulators on them [151]. The highly repeatable and reliable process used for attaining thin films (400–500-nm thick) of LiNbO$_3$ on Si begins with ion implanting commercially available LiNbO$_3$ wafers to weaken bonds for future slicing. A standard silicon substrate is coated with a thermally-grown SiO$_2$ layer to be later employed as waveguide cladding layer. The use of thermally-grown SiO$_2$, rather than PECVD oxide, is critical in forming strong bonds. The two wafers are then bonded at room temperature, after which they are heated to 200°C to not only slice the LiNbO$_3$ wafer at the bonds weakened by ion implantation, but also to strengthen the formed bond at the LiNbO$_3$-SiO$_2$ interface. CMP is the last step applied on the bonded LiNbO$_3$ thin films to attain smooth surfaces for waveguide processing.

Another historic challenge for LiNbO$_3$ integrated photonics has been the difficulty of etching the material to produce waveguides with smooth sidewalls. Ion beam milling [164, 165], plasma etching [166, 167], and wet etching [168–171] are the main techniques pursued. Wet etching in particular has produced fairly low-loss waveguides [171]. However, slow etch rate, undercutting, asymmetric profiles and large feature sizes are common shortcomings of wet etching recipes and submicron waveguides cannot be achieved.

One solution is to guide light in SOI waveguides and bond a thin film of LiNbO$_3$ on top of the SOI chip. Both free-standing [172] and BCB-bonded films have been reported [173]. The shortcomings of this method is not only the instability of the free-standing structures and/or the unreliability of polymers but also the fact that only the evanescent tail of the SOI waveguide mode reaches out to the LiNbO$_3$ top cladding, leading to small (e.g., 11% [173]) overlap of the EO active region with the optical mode.

A novel approach proposed at CREOL is rib-loading LiNbO$_3$ slab waveguides with an index-matching material, such that most of the optical energy still resides in the slab region [151]. Coincidently, the index of Ta$_2$O$_5$ matches that of LiNbO$_3$ (~2.1). Hence, the same SORM technology described for attaining Ta$_2$O$_5$ waveguides on SiO$_2$-coated
Si [131, 132] can be applied on thin films of LiNbO3 bonded to Si and readily achieve waveguides. Figure 8A and B show a surface-scanning micrograph (SEM) of a fabricated device as well as the device schematic and calculated optical mode [151]. Strong overlap of the optical field with the LiNbO3 is evident in the simulated mode in Figure 8C. Figure 8D shows the modulation response of the devices at low frequencies. \( V_\pi L \) as low as 4 V-cm is obtained in push-pull configured MZI modulators. The reported \( V_\pi L \) is three times lower than the best commercial devices. Calculations show \( V_\pi L \) as low as 1.2 V-cm is possible. Preliminary high-speed characterization, using \( S_{21} \) parameter of a network analyzer, suggests operation up to 10 GHz. High-speed measurements based on travelling-waveguide electrodes for velocity matching of RF and optical waves are underway. A clear advantage of the results in Figure 8D is that the modulation depth is expectedly very high (~20 dB) and is a clear advantage over free-carrier modulated Si modulators. With fitting a sinusoidal function to the output of the MZIs, highly linear transmission (<3% error in the fitting) was extracted, confirming the devices’ suitability for analog communications in addition to the more common digital data- and telecommunications. Finally, Figure 9 summarizes the improvement achieved by the presented approach in terms of waveguide effective area and minimum radii for negligible bending loss over conventional diffused waveguides and where it stands with respect to other integrated photonic platforms.

### 7 Mid-infrared integrated photonic platforms

Optics in the mid-wave infrared (MWIR or mid-IR) wavelengths, particularly the atmospheric transmission window of 3–5 μm, has been researched on for over 60 years [174]. Historically, mid-IR optics has been motivated by applications such as light detection and ranging (LIDAR) and infrared countermeasures. Over the past decade, mid-IR photonics has found a host of civilian applications, e.g., Fourier transform infrared (FTIR) spectroscopy, remote biological and chemical sensing for environmental monitoring and free-space optical communications [175–177]. More recently, clinical applications, like tissue photoablation, have also emerged [178–180]. These applications would have not been possible without successes in high-performance uncooled commercial mid-IR optical sources [181, 182] and photodetectors [183, 184]. However, most current systems are based on assemblage of discrete components, i.e., true integration is lacking. Device miniaturization and monolithic integration would hence be a great

---

**Figure 8** (A) Micrograph, (B) schematic, (C) optical mode simulation and (D) measured modulation response of the first demonstrated LiNbO3-on-Si modulators [151].
advancement in the field of mid-IR photonics, a goal that can be achieved by silicon photonics.

Mid-IR silicon photonics has potential applications in the host of applications mentioned above, as well as lab-on-a-chip biochemical sensing for trace-gas detection and bio-agent sensing for environmental and homeland security monitoring, as well as industrial process control. It can also have several nonlinear-optical applications [185, 186]. The first demonstration of mid-IR Raman amplifiers [187, 188] suggested that nonlinear silicon photonic devices can be achieved. Indeed, TPA vanishes in the mid-IR regime (above ~2.2 μm to be more precise) and three-photon absorption are negligible at very high intensities [189]. Thus, the discussed omnipresent problem of nonlinear silicon devices in the near-IR, i.e., TPA-induced FCA, is inherently nonexistent in the mid-IR. Analytic and numerical works conducted at CREOL suggest that high-performance 3–5 μm nonlinear silicon photonic devices can be attained provided that low-loss waveguides and sources with high beam-quality are available [133, 190].

However, the standard SOI platform is generally not ideal for mid-IR photonics because the bottom cladding material, SiO$_2$, is very lossy over 2.5–2.9 μm and above 3.5 μm wavelengths [191]. Replacing SiO$_2$ with sapphire (transparent up to 4.3 μm), SiN (transparent up to 6.6 μm) or air (suspended silicon membrane, transparent up to 6.9 μm) as the bottom waveguide cladding layer, as well as Ge waveguides directly on Si (Ge-on-Si transparent up to 6.9 μm), were first proposed by Soref et al. [191]. All of these waveguide platforms have been demonstrated by other researchers, as reviewed below.

Available silicon-on-sapphire (SOS) wafers, originally developed for microelectronic applications, led to the demonstration of first Si-based mid-IR waveguides [192–194]. Low-loss waveguides and grating couplers on SOS were later demonstrated [195]. The author’s group has also fabricated SOS waveguides (see Figure 10A). However, we have not been able to achieve low propagation loss (~12 dB/cm at 1.55 μm). The reason for this is not clear, but private communications with some other researchers, who have been using available SOS wafers in the market, suggest that quality variation of the Si thin layer grown epitaxially on sapphire substrates can cause this. Indeed, there has been some report on inherent twinning defects in SOS [196]. At any rate and as mentioned, sapphire is transparent only up to 4.3 μm and silicon-on-nitride (SON) (see Figure 10B), air-clad all-silicon waveguides (see Figure 10C) and Ge-on-Si offer wider transparency ranges.

Ge-on-Si were first reported by Chang et al. [197]. Reduced-pressure chemical vapor deposition was employed to grow a 2-μm thick monocrystalline Ge relaxed layer on a Si and ~3-μm-wide stripes were fabricated by photolithography. These waveguides are inherently large due to the not very high index contrast between Ge and Si (~4 vs. 3.4) but for the same reason have relatively low propagation loss of ~3.6 dB/cm at 5.8 μm wavelength. More recently, Ge-on-Si thermo-optic phase shifters have been reported [198].

The author’s group has demonstrated SON and all-silicon waveguides for the first time [199, 200]. We have also proposed and demonstrated the silicon on lithium niobate (SiLN) waveguide and modulator platform (see Figure 10D) [201, 202]. These three platforms are reviewed in detail in the following.

In principle, the discussed ion implantation and wafer slicing steps used for LiNbO$_3$-on-Si wafers can be employed to achieve SON wafers. But with the availability of SOI wafers, perhaps there is no need to repeat those steps in order to attain Si thin films. Rather, back-side etching of the Si substrate and the BOX layer can expose the thin Si layer. Accordingly, Figure 11 shows the processing steps for achieving SON dies [199]. Briefly, SON handling dies are coated with low-stress PECVD SiN. Both CMP [203, 204] and spin-on-glass (SOG) [205, 206] techniques can in principle be used to achieve direct bonding. SOG was preferred due to its lower cost and ease of fabrication. Hydrophilic SiO$_2$ layers were deposited on both dies before bonding for improved adhesion. Initial room-temperature bonding at atmospheric pressure was followed by 60 min of annealing at 450°C. Then, the silicon handling wafer and the BOX layer are removed, resulting in a Si thin film residing on a SiN cladding layer on a Si substrate. Standard
techniques were then used to fabricate ridge waveguides (see Figure 10B). At 3.39 μm wavelength, the propagation loss of the characterized waveguides is 5.1 dB/cm for the TM mode.

The next Si-based platform discussed in this review paper is perhaps not heterogeneous per se but indeed an all-silicon optical platform (ASOP) [200], i.e., air-clad suspended membrane waveguides. Nonetheless, since the processing steps for this truly homogeneous platform are based on bonding techniques discussed throughout this paper and it can be applied to enjoy the largest transparency range of ~2.5 octaves offered by silicon (1.2–6.9 μm), it is discussed here for completeness.

Suspended membrane waveguides for mid-IR were first demonstrated by local removal of BOX in SOI waveguides via forming an array of periodical holes in the top silicon layer adjacent to the rib waveguides [207]. However, due to the via forming and isotropic wet etching steps, the attained membranes become too wide to be mechanically and thermally stable, i.e., they are prone to vibrations, bowing and low thermal dissipation. The processing steps for the approach developed at CREOL, shown in Figure 12, address these shortcomings. An inverted SOI wafer is fusion-bonded to a bulk silicon substrate pre-patterned with trenches. Similar to the discussed SON processing steps, the SOI substrate and BOX are subsequently removed and a thin Si film on the buried trench is exposed. The air-filled trench acts as the bottom

Figure 10  Four different mid-infrared waveguide technologies pursued at CREOL, their schematics and their measured propagation losses: (A) silicon-on-sapphire (SOS) (unpublished); (B) silicon-on-nitride (SON) [199]; (C) all-silicon optical platform (ASOP) [200]; (D) silicon-on-lithium-niobate (SiLN) [201, 202].

Figure 11  Fabrication steps of SON dies [199].

Figure 12  Fabrication steps of ASOP dies [200].
In an unpublished second generation of the devices, the width has been reduced to 12 μm. The depth of the trench can be freely controlled and the formed fusion-sealed internal channel can be potentially used for microfluidic or gas-sensing applications.

The last mid-IR platform discussed in this review paper is based on bonding a Si thin film on a LiNbO₃ substrate. The motivation for developing this silicon-on-lithium-niobate or SiLN platform at CREOL is primarily to achieve EO modulator in the mid-IR range. Microring resonators and grating couplers characterized at 1.55 μm [201] and EO MZI modulators at 3.4 μm [202] have been demonstrated. As shown in Figure 13A, room-temperature plasma-assisted bonding was employed to directly bond an SOI die with a LiNbO₃ die. Then, the dies are annealed at low temperature. Wet etching of the SOI backside, exposes a 220 nm-thick crystalline silicon layer on LiNbO₃. The relatively large index contrast of Si versus LiNbO₃ (∼3.4 vs. ∼2.1) provides an opportunity to achieve compact oxide-less channel (0.6-μm wide in Figure 10D) Si waveguide directly on an EO-active substrate. Microring resonators were demonstrated at near-IR wavelengths. Their propagation loss is as low as 3.0 dB/cm for the TM mode at 1.55 μm (see Figure 13B) with a ring radius of 130 μm [201]. According to mode simulations and at 3.4 μm wavelength, ∼24% of the optical mode resides in LiNbO₃. Hence, relatively strong EO modulators can be envisaged, as we have very recently demonstrated (see Figure 13C) [202]. This is the first demonstration of a mid-IR EO modulator. The devices attain a modulation depth of 8 dB, a $V_p L$ of 26 V cm at low frequencies, a propagation loss of 2.5 dB/cm, and an insertion loss of ∼3 dB, all characterized at a wavelength of 3.4 μm.

8 Concluding remarks

Herbert Spencer, the famous nineteenth century English polymath once wrote: “From the remotest past which Science can fathom, up to the novelties of yesterday, that in which Progress essentially consists, is the transformation of the homogeneous to the heterogeneous” [208]. Such a transformation appears to be the destiny of progress in the science and technology of electronic-photonic integrated circuits. It may be unwise to make predictions for any field of science and engineering, but based on the review of the field in this paper, the following remarks are cautiously made. It appears that in the foreseeable future, heterogeneous technologies, that combine group

waveguide cladding when waveguide ridges are formed by standard lithographic techniques. An SEM of the device is shown in Figure 10C. The waveguides have a TM-mode low propagation loss of 2.8 dB/cm at 3.39 μm. This is the first demonstration of a homogeneous all-silicon optical platform (ASOP). The first generation devices have a 16-μm-wide trenches and the membranes are mechanically very stable [200].
IV and III–V semiconductors, and perhaps LiNbO₃, and other materials such as nitride semiconductors and high-index dielectrics, will most likely dominate integrated optoelectronics. Electronics will most likely be based on Si CMOS and lasers will most likely be on III–Vs. Prevailing materials and device platforms for other functionalities are much more debatable and predicting them is more difficult, hence less assertive adjectives are chosen in the following. Photodetectors will likely be on Ge, but perhaps on III–Vs. Modulators could be based either on free-carrier-plasma effect in Si, EO or EA effect in III–Vs or EA in GeSi, or EO in LiNbO₃. Passive devices (waveguides, filters, etc.) will likely be on Si due to its inexpensiveness and low loss.

Acknowledgments: The author would like to thank the past and present members of his research group, J. Chiles, S. Khan, J. Ma, P. Rabiei and A. Rao, for their contributions to some of the works reviewed here. The author would also thank J. Chiles and A. Rao for critically reading the manuscript. The works are being supported by the United States’ Office of Naval Research (ONR) Young Investigator Program (YIP) under the Grant Number 11296285 and the National Science Foundation CAREER Program under Award Number ECCS-1150672.

References

[31] Celler GK, Cristoloveanu S. Frontiers of silicon-on-insulator.


