Heterogeneous Nonlinear Integrated Photonics

Sasan Fathpour, Senior Member, IEEE

(Invited Papers)

Abstract—Integrated photonics has been pursued on a large variety of materials and platforms over more than half a century. In recent decades, silicon has emerged as the preferred optical and substrate material. However, since many optical functionalities cannot be, inherently or conveniently, implemented on silicon-based devices, heterogeneous integration of other materials has been an inevitable constituent of silicon photonics from its inception. This paper reviews recent progress in the heterogeneous integration of dielectrics and compound semiconductors as the core optical material for ultracompact nonlinear integrated photonics, preferably on silicon substrates. Occasionally, important devices on native substrates of the core nonlinear material are also reviewed. The focus is on ultracompact photonic devices and circuits formed on submicron films of lithium niobate (LiNbO3), and the aluminium-gallium-arsenide (AlGaAs) and aluminium-gallium-nitride (AlGaN) families for second-order nonlinear processes, such as second-harmonic generation, and techniques to address phase-matching in them. Integration of second- and third-order nonlinear devices on the same chip appears to be a future trend, thus the related preliminary results are reviewed. A broader introduction to heterogeneously integrated photonics, shortcomings of conventional silicon photonics for application in nonlinear optics, as well as an overview of modeling second-order nonlinear integrated-waveguide devices are also provided for completeness.

Index Terms—Integrated photonics, photonic integrated circuits, silicon photonics, heterogeneous integration, nonlinear optics, nonlinear optical devices, second-harmonic generation, optical waveguides, lithium niobate, thin film, gallium arsenide, gallium nitride, aluminium nitride.

I. INTRODUCTION

INTEGRATED photonics owes its birth to two groundbreaking technologies pioneered in the 1960s: optoelectronics, on III-V compound semiconductors [1], [2], and integrated optics, primarily on dielectric waveguides [3]. Meanwhile, silicon (Si) photonics has been a subject of extensive research since the mid-80s, as a potential alternative to these parent technologies [4]. “Silicon-based optoelectronics” [5] was primarily motivated by its potential advantages over optoelectronic integrated circuits (OEICs); the same technology nowadays referred to as III-V photonic integrated circuits (PICs). Realized on silicon-on-insulator (SOI) wafers [6], [7], silicon photonics provides high-contrast, low-loss and bendable waveguides superior to III-V PICs, as well as the bulky dielectric planar counterparts of traditional integrated optics for passive and nonlinear applications, with the additional possibility of electronically manipulating silicon’s optical properties.

Other argued advantages of silicon photonics include higher processing yield, lower production cost, mechanical robustness of Si substrates and ease of handling thermal cycles on them, large-scale integration and compatibility with foundry processing. Integration with silicon-based electronics is another historically argued advantage [5]. However, this is currently much less critical, as the trend is to have photonics and electronics on separate chips, since it is not economical to fabricate relatively bulky photonic circuits with an expensive manufacturing technology node necessary for the electronics (with much smaller feature sizes).

Nonetheless, it was clear from inception that there is no ‘pure’ silicon photonics, as other materials are inevitably required for a variety of functionalities that cannot be easily achieved on silicon [5]. Some clear shortcomings of silicon-only photonics are lack of efficient light generation, in the indirect-bandgap material, and negligible photodetection, below its bandgap energy. Addressing these shortcomings was critical for the development of silicon photonics for its primary targeted application of optical transceivers.

Integrating thin layers of germanium (Ge) in waveguide photodetector configuration has been a long practice [8], hence group-IV photonics is another common term for the technology. Meanwhile, despite tremendous amounts of effort [9], dependable silicon-based lasers remain elusive to date. Instead, various methods have been successfully developed for hybrid integration of III-V lasers and are in use in commercial transceiver chips [10]. Accordingly, ‘groups III-to-V photonics’ may semantically be a more accurate descriptive name for the current state of silicon photonics.

Regardless of semantic subtleties, silicon photonics is nowadays a fairly mature technology, and is adopted by several companies for datacom applications. In addition, the versatile technology is accessible to fabless researchers through a few global foundries. It is not the intent of this paper to review silicon photonics, nor heterogeneous integration of Ge and III-V materials for optical transceiver applications, found in several other publications [9]–[12]. It is stressed, though, that heterogeneous integration of Ge photodetectors and III-V lasers with SOI waveguides complicates the processing steps, increases the cost and thereby diminishes some of the argued advantages of silicon photonics. It should be also reluctantly admitted that the same general argument applies to the heterogeneous solutions presented here for nonlinear-optic applications.
This paper focuses on recent heterogeneous solutions on silicon substrates for nonlinear integrated photonics and applications envisioned around them. In occasional cases and since the intended nonlinear-optic functionality is essentially unaffected, heterogeneous devices on native substrate of the host nonlinear material are also discussed. In addition, due to limited space, the emphasis is on second-order nonlinearity. But for some future directions and applications, discussed in Section VI.C, it is desirable to integrate both second- and third-order nonlinear devices on the same chip. Furthermore, nonlinear absorption is a common problem in both nonlinearities, if not addressed properly. Hence, a brief description of non-resonant electronic (or elastic) third-order nonlinearity and its unfortunate hurdles in conventional silicon photonics, and heterogeneous solutions to circumvent them, are also presented.

II. SHORTCOMINGS OF SILICON FOR NONLINEAR INTEGRATED PHOTONICS

Traditionally, second-order nonlinear integrated optics has been pursued in waveguides fabricated out of materials with noncentrosymmetric crystals. As discussed later, these devices have their own downsides, such as low optical confinement and bulkiness. On the other hand, crystalline silicon possesses a centrosymmetric symmetry lattice structure and hence lacks the linear electrooptic (EO) or Pockels effect and the related second-order optical nonlinearity. This apparent dilemma can be straightforwardly solved by heterogeneously integrating thin layers of noncentrosymmetric materials on a low-refractive index cladding layer, e.g., silicon dioxide (SiO₂), on Si substrates, in order to simultaneously achieve high second-order nonlinearity and high confinement/compactness, while enjoying partial or full compatibility with silicon foundry processing. Example materials are ferroelectrics, such as lithium niobate (LiNbO₃), and compound semiconductors such as the aluminium-gallium-arsenide (AlGaAs) and aluminium-gallium-nitride (AlGaN) families, as presented in detail throughout this paper.

In the 2000s, there were tremendous efforts in exploiting the strong third-order optical nonlinearities of SOI waveguides [9], [10], [12], [13]. The unfortunate omnipresent problem here is two-photon absorption (TPA), occurring below ~2.2 μm wavelength, and its induced free-carrier absorption (FCA) that severely hinder performance of the devices. In addition, the accompanying free-carrier dispersion, related to FCA through the nonlinear Kramers-Kronig relations, can further complicate the design and performance of nonlinear devices.

Techniques such as carrier sweep-out, doping, and intentional impurity implantation can be used to reduce the carrier lifetime and alleviate the nonlinear absorption issue [9]. Nonetheless, for third-order nonlinear integrated photonics, researchers have for the most part migrated to heterogeneous integration of other materials with negligible TPA. Making planar waveguides out of many of these materials can be traced back to decades ago, e.g., tantalum pentoxide (Ta₂O₅) [14]. The modern approaches, however, typically focus on compact 2-D waveguides and reduced loss, as well as more advanced devices such as microresonators and grating waveguides (e.g., see [15] on Ta₂O₅).

Additionally, dispersion engineering for fairly sophisticated integrated chips and advanced nonlinear functionalities, such as supercontinuum generation (SCG), as well as potential compatibility with silicon photonics for large circuit integration, are among current considerations. Topical reviews on third-order nonlinear integrated photonics have been published in recent years [16]–[18]. Some very recent advancements include works on silicon nitride (SiN) [19], [20], chalcogenide glass (ChG) [21], AlGaAs [22], GaN [23], and titanium dioxide (TiO₂) [24].

TPA vanishes at above 2.2 μm pump wavelengths and although the third-order nonlinearities somewhat weaken too, silicon is overall expected to be an ideal material for nonlinear integrated photonics at these wavelengths up to about 6.9 μm, where multiphonon absorption occurs. The 2.2–6.9 μm nonlinear transparency range of silicon loosely overlaps the mid-infrared (mid-IR) wavelengths defined as 3–5 μm and sometimes 3–8 μm with a host of potential applications [9], [10], [12]. The problem here is that the buried oxide (BOX), i.e., the SiO₂ insulating layer of SOI, is lossy at most of this range. Alternatively, several schemes have been devised and implemented for eliminating the BOX layer. Such beyond-SOI silicon photonics for mid-IR applications are reviewed elsewhere [10], [25], [26]. Similar BOX-free platforms may be essential to be developed for the second-order nonlinear materials mentioned above, if mid-IR operation is intended. More discussions on this subtopic is presented in Section VI.B.

III. REVIEW OF BASIC THEORY

Armstrong et al. [27] developed the fundamental theory of nonlinear light-matter interaction in 1962. Nowadays, excellent textbooks present detailed and more advanced formulations in bulk [28]–[30], optical fibers [31] and integrated optics [32]. A brief introduction is provided in the following.

Generally, nonlinearities of light-matter interaction can phenomenologically be described by defining second, third, and higher-order optical susceptibilities, \( \chi^{(2)} \), \( \chi^{(3)} \), etc., respectively, in addition to the ordinary linear susceptibility, \( \chi^{(1)} \). These susceptibilities relate the polarization field, \( \mathcal{P}(r, t) \), induced by the electric field of an optical field, \( \mathcal{E}(r, t) \), by the tensor relation

\[
\mathcal{P} = \varepsilon_0 \left( \chi^{(1)} \mathcal{E} + \chi^{(2)} : \mathcal{E} \mathcal{E} + \chi^{(3)} : \mathcal{E} \mathcal{E} \mathcal{E} + \ldots \right),
\]

where \( \varepsilon_0 \) is the vacuum permittivity. Ignore the higher-order nonlinearities, and suppose input harmonic waves of the form \( \mathcal{E}(r, t) = \Re \left\{ \sum_q \mathcal{E}(r; \omega_q) e^{i\omega_q t} \right\} \), at angular frequencies \( \omega_q \). Then, the output spectrum, \( \omega_q \), related to \( \mathcal{P}(r, t) = \Re \left\{ \sum_q \mathcal{P}(r; \omega_q) (r) e^{i\omega_q t} \right\} \), has a richer spectrum and gives rise to several second- and third-order processes, as conceptually shown in Fig. 1 and tabulated in Table I, for \( \chi^{(2)} \). Accordingly, the Cartesian \( ijk \) amplitude component
of the nonlinear polarization at \( \omega_q \) are

\[
\tilde{P}_i(r; \omega_q) = e_0 \sum_j \chi^{(1)}_{ij} (r; \omega_q) \tilde{E}_j(r; \omega_q) + \frac{1}{2} e_0 \sum_{j' k} \chi^{(2)}_{ijjk} (r; \omega_q) \tilde{E}_j(r; \omega_q) \tilde{E}_k(r; \omega_{q'}) + \frac{1}{4} e_0 \sum_{j' k' l} \chi^{(3)}_{ijkl} (r; \omega_q) \tilde{E}_j(r; \omega_q) \tilde{E}_k(r; \omega_{q'}) \times \tilde{E}_l(r; \omega_{q''}).
\]

Energy conservation is implied in (2) with the parenthesized \((v v')\) and \((v' v'\)) notations in the summations. For example, \( \omega_q = \omega_v + \omega_{v'} \) for the \( \chi^{(2)} \) summation, if sum-frequency generation (SFG) at \( \omega_q \) is intended. Phase matching conditions should also be satisfied for efficient interactions, e.g., \( k_q \approx k_v + k_{v'} \) for SFG. For a few processes, the phase matching condition is automatically satisfied (see Table I), but tackling the phase mismatch in most of the other second-order nonlinear processes is an omnipresent challenge. Table I provides more detail on the frequency relation among the interacting waves for \( \chi^{(2)} \) processes, as well as their input, \( A_i^{\text{in}} \), and output, \( A_i^{\text{out}} \), amplitude conditions, corresponding to \( i = 1, 2 \) and 3 in Fig. 1.

Several spatial-lattice symmetry relations reduce the number of nonvanishing and independent \( \chi^{(2)}_{ijk} \) and \( \chi^{(3)}_{ijkl} \) tensor elements. Particularly, \( \chi^{(2)} \) completely vanishes for centrosymmetric crystals. For noncentrosymmetric crystals, it is more common to use the nonlinear coefficients \( d_{ijk} = \chi^{(2)}_{ijk} / 2 \), which is often contracted to a \( 3 \times 6 \) matrix, denoted by \( d_{ij} \), with a few independent elements, depending on the crystallographic point group [29]. In practice, for fixed propagation direction and polarization fields, we can usually define an effective nonlinear coefficient, \( d_{\text{eff}} \), and reduce the tensor relation to a convenient scalar form, as exemplified in the Appendix.

In principle, the 81-element \( \chi^{(3)} \) susceptibility tensor virtually exists in any material. In practice, though, most of the employed materials are isotropic and only \( \chi^{(3)}_{1111}, \chi^{(3)}_{1122}, \chi^{(3)}_{1212}, \chi^{(3)}_{1221} \) are independent [29]. Also, \( \chi^{(3)}_{1111} = \chi^{(3)}_{1122} + \chi^{(3)}_{1212} + \chi^{(3)}_{1221} \) for isotropic media. In addition, these three elements usually have roughly the same magnitude and hence it is common to define a nonlinear refractive index as

\[
n_2 = 3 \text{Re} \{ \chi^{(3)}_{1111} / 4 \epsilon_0 n_0^2 c \},
\]

such that the total refractive index \( n = n_0 + n_2 I \), where \( n_0 \) is the medium’s linear index, \( c \) is the speed of light in vacuum and \( I \) is the optical intensity. It is finally noted that resonant inelastic nonlinearities based on electron-phonon interactions, i.e., Raman and Brillouin scattering, can also be modeled by their own \( \chi^{(3)} \) susceptibility tensors, but their review is beyond the scope of this work.

Modeling and analysis of \( \chi^{(2)} \) and \( \chi^{(3)} \) elastic processes in bulk materials, based on coupled-amplitude differential equations governing \( A_i \), can be traced back to early days of nonlinear optics [27] and is elaborated on in the aforementioned textbooks [28]–[30]. Also, partially thanks to close resemblance to the well-established theory in optical fibers [31], the modeling literature on third-order nonlinear integrated photonics is also fairly rich. However, in the case of \( \chi^{(2)} \) nonlinearities, variations of the bulk theory in integrated devices are very limited in texts [32], [33] and sparse in the broader literature (e.g., see [34]–[37]). Thereby, and for completeness, coupled-mode equations for \( \chi^{(2)} \) processes and their important features in terms of efficiency and phase matching in modern waveguides are revisited and presented in the Appendix.

IV. CONVENTIONAL SECOND-ORDER NONLINEAR INTEGRATED OPTICS

Franken et al. reported the first nonlinear-optic experiment, namely second-harmonic generation (SHG) in crystalline quartz, in 1961 [38]. Within a decade or so from this historic observation, several techniques were devised for increasing the efficiency and overcoming the issue of phase mismatch among the interacting waves in several materials.
Birefringence phase matching, i.e., exploiting anisotropic material dispersion, is perhaps the most common and well-known method in bulk nonlinear optics [39]. Other early examples are periodic total internal reflection in unguided thick films for spatial modulation of nonlinearity [27], [40], periodic index perturbation using material modulation in bulk structures [41], and reliance on the acoustooptic effect [42].

Advantages of integrated-optic schemes, devised in order to address the phase matching condition, were also recognized soon. Dispersion of guided modes was initially commonplace. For instance, slab waveguides were used to demonstrate Cherenkov radiation in zinc sulfide (ZnS) [43] and modal phase-matching (MPM) in GaAs [44]. Also, periodic perturbation of effective index was proposed [45] and later demonstrated in aluminium oxide (Al₂O₃) on quartz [46], and in GaAs heterostructures [47]. The first nonlinear experiment in 3-D waveguides demonstrated Cherenkov radiation from titanium-diffused LiNbO₃ [48]. Stegeman and Seaton [49] have authored a great review of these early works in nonlinear integrated optics. Below, the focus is on recent techniques that are more relevant to the modern needs.

Cubic compound semiconductors typically have very high optical nonlinearity. In the GaAs-based family, AlGaAs is preferred as it offers a wider transparency range, particularly to avoid severe TPA at below the half-bandgap [50], as discussed in more detail later in Section VI.B. Despite their high nonlinearity, intrinsic birefringence is nonexistent in these isotropic cubic semiconductors. Instead, ‘form’ birefringence can be induced in waveguides with AlGaAs/Al₂O₃ multilayers [51], [52], but the waveguides typically suffer from high propagation loss at the second-harmonic wavelength and hence limited efficiency. MPM, between various quasi-transverse electric (quasi-TE) or quasi-transverse magnetic (quasi-TM) modes, can be used in 3-D semiconductor waveguides, as well [53], [54]. Here, the conversion efficiency is generally limited due to small spatial overlap between the interacting modes.

Quasi-phase matching (QPM) in single-mode waveguides is generally advantageous over MPM. A general and theoretical description of QPM is presented in the Appendix. In essence, the objective is to provide an extra phase term by an engineered periodic effect in order to counteract the dispersion-induced phase mismatch among the interacting waves.

For instance, in semiconductors with the zincblende lattice structure, different orientations have opposite nonlinear coefficient signs. Accordingly, mechanical stacking of series of GaAs crystalline plates was explored to achieve QPM for SHG experiments, in two reports in 1976 [55], [56]. Later, techniques for epitaxial growth of AlGaAs on a periodically patterned template were developed and SHG and difference-frequency generation (DFG) experiments were conducted on waveguides based on them [57]–[60]. However, high propagation loss is a typical shortcoming of conventional III-V waveguides and the issue is aggravated by corrugations the patterning templates. Consequently, the SHG conversion efficiency in this approach is limited (e.g., to 23 %/W in one of the best reports [60]). Perhaps, this is why such orientation-patterning technique on semiconductors has been pursued more often in bulk devices. Two review papers on nonlinear frequency conversion in semiconductors can be found elsewhere [61], [62].

As elaborated more in Section VI.B, another general problem in AlGaAs waveguides is TPA in the relatively low-bandgap materials and the associated FCA that limits their nonlinear transparency range. In other words, dielectric and wide-gap semiconductors are advantageous in this regard.

Discovering periodic poling has led to attractive techniques for QPM in ferroelectric nonlinear materials. Periodic poling is induced by applying a modulating thermal, electrical or other physical force, in order to permanently flip the polarity of alternating atomic domains and hence periodically modulate the sign of \( \chi^{(2)} \). The method was first demonstrated in LiNbO₃ bulk crystals by displacing the rotation axis during the Czochralski growth from the symmetry axis of a temperature gradient [63]. Modulating the bias current during crystal growth is another method exploited for SHG [64].

Miyazawa discovered that diffusing Ti dopants into the +c face of LiNbO₃ wafers causes ferroelectric domain reversal, i.e., results in a wafer-surface spontaneous polarization opposite to that of the substrate [65]. Lithographic patterning and liftoff methods were subsequently employed to generate a periodic Ti layer on LiNbO₃ and generate green [66] and blue [67] light via SHG. In an alternative method, modulating out-diffusion of LiO₂ from LiNbO₃ by a periodic SiO₂ mask was exploited to generate blue light [68]. These in- and out-diffusion methods, however, offer only shallow inverted domains. Room-temperature direct electron-beam lithography has been also explored for domain reversal in small areas [69].

Applying an external electric field via metallic contacts at room temperature to induce permanently periodically-poled ferroelectric materials in large areas was first reported in 1993 [70] and is considered a breakthrough in second-order nonlinear integrated optics. More than 20 mW of blue-color second-harmonic light was generated with an efficiency of 600 %/W.cm² from <200 mW of input fundamental power at 852-nm wavelength. Such periodically-poled lithium niobate (PPLN) devices have been the dominant commercial technology for \( \chi^{(2)} \)-based nonlinear optics since then.

Periodic poling is in principle applicable to other ferroelectric materials, including lithium tantalate (LiTaO₃), potassium titanyl phosphate (KTiOPO₄ or KTP), and potassium niobate (KNbO₃) waveguides [32]. The technique can also be applied to non-cubic semiconductors, e.g., those with wurtzite lattice structure. An example is GaN grown on hexagonal lattice-type substrates like sapphire. Accordingly, patterning GaN/AlN stripes on sapphire allows epitaxial growth of alternating polarity of GaN layers, i.e., periodic poling of the nonlinear coefficient. SHG experiments on planar waveguides on these films has been reported with a normalized efficiency of 18 %/W [71].

In more recent years, other techniques have been devised for phase matching, as well. An example is taking advantage of the rich dispersion properties of whispering-gallery modes in optical microresonators [72]. For instance, in one scheme called cyclic phase matching [73], the notion that the effective
index of quasi-TE mode remains constant around the cavity is exploited. Accordingly, the mode’s azimuthal number can be phase-matched to the periodic index of a quasi-TM mode at four points along the circumference of a microresonator lying in the XCrystallographic plane. The progress in this field is recently reviewed elsewhere [74], [75]. But the approach typically relies on suspended microtoroid or microdisk structures, which do not easily lend themselves to integration-friendly technologies. For example, it is not trivial to couple these devices with traditional waveguides, and the reported experiments on them heavily rely on prism or tapered optical-fiber coupling schemes.

V. HETEROGENEOUS SECOND-ORDER NONLINEAR INTEGRATED PHOTONICS

As mentioned in Section II, silicon has a centro-symmetric crystal and inherently lacks $\chi^{(2)}$ nonlinearity. Nonetheless, some second-order nonlinearity is observed in stressed Si waveguides. The origin of this weak nonlinearity is debatable, but likely originates from the stressed heteropolar SiN cladding [76]. Another Si-based method is electric field-induced second harmonic generation (EFISHG), in which a DC electric field interacts with $\chi^{(3)}$ in order to effectively induce second-order nonlinearity [77]. A. Rao and this author have reviewed these approaches elsewhere [78], but since they are technically not based on heterogeneous integration of $\chi^{(2)}$ materials, they are not further discussed here. In the following subsections, recent heterogeneous works are categorized based on the employed material systems.

A. Lithium Niobate on Silicon

LiNbO$_3$ waveguides are most commonly formed by titanium (Ti) in-diffused [79] and proton exchange [80]. The in-diffusion method suffers from photorefractive damage in the visible range. The proton exchange method not only has a higher optical damage threshold, but also possesses the extra advantage of providing higher waveguide contrast and step index profile. Meanwhile, improved resistance to photorefractive damage can be attained in magnesium-doped lithium niobate crystals (MgO:LiNbO$_3$) [81].

The in-diffused and proton exchange methods, however, only slightly alter the refractive index of the material, hence the guided modes are weakly confined. An example cross-section for Ti-diffused waveguides is shown in Fig. 2(a). The large cross-section of waveguides implies the need for high-power sources to onset $\chi^{(2)}$ nonlinearities. Furthermore, weakly-confined waveguides exhibit bending loss that becomes significantly large for small radii. Consequently, incorporating several of the devices on the same chip is not feasible for integrated photonic applications.

Achieving thin films (200 to 600 nm) of LiNbO$_3$ on an insulator cladding with a much lower index (e.g., SiO$_2$), along with a technique to achieve low-loss ridge or channel waveguides alleviates the above problems and can yield high-contrast submicron LiNbO$_3$ waveguides. Schematically, such envisioned waveguide structures are shown in Fig. 2(b). The schematics are not in scale, and in reality the final waveguide cross-section of waveguides is reduced from $>30 \, \mu m^2$ in conventional waveguides in Fig. 2(a) down to $<1 \, \mu m^2$ in the ultracompact devices in Fig. 2(b).

Also shown in Fig. 2, is the geometrical differences between the crystalline axes, XZY, of the two platforms if periodic poling is intended. In order to exploit the highest nonlinear coefficient in LiNbO$_3$, $d_{33}$, it is common in conventional waveguides to use Z-cut wafers and vertical (x-direction) poling fields. In the thin-film platform, however, it is more convenient to place the poling electrodes laterally, so the once-applied quasi-DC field will be along the y direction of the waveguide, i.e., parallel to the Z crystalline axis to take advantage of $d_{33}$ for efficient nonlinear optics. This can be achieved in Y-cut, as well as X-cut, films.

Fabrication of such single-crystalline thin films of LiNbO$_3$ has been proven to be very challenging. Chemical vapor deposition [82], sol-gel processing [83], RF sputtering [84]–[86], and pulsed-laser deposition [87], [88] have been pursued with little success in terms of optical quality. Adhesive polymer (e.g., benzocyclobutene or BCB) bonding and thermal slicing has been employed for wafer bonding and demonstrate thin-film of LiNbO$_3$ on LiNbO$_3$ substrates [89]. Polymer-bonding is simple and straightforward for lab demonstrations. However, polymers are in general not attractive for practical applications due to the low quality and reliability of the formed bonds.

LiNbO$_3$-on-Si wafers were first demonstrated in 2013 [90] and later commercialized [91]. The process resembles the SmartCut™ method of SOI, to some extent, and employs ion implantation, wafer bonding and thermal slicing with careful temperature consideration to circumvent the large thermal expansion coefficient between LiNbO$_3$ and Si. The end product is a 200-to-600-nm-thick film of LiNbO$_3$ bonded to a Si wafer with a SiO$_2$ buffer in between. More details on the processing can be found in [90].

For lateral confinement of waveguides, two options are available, namely, direct etching the thin films and rib-loading them with an index-matching material that is easier to be etched. For rib-loading, the index of aforementioned low-loss Ta$_2$O$_5$ [15], [92], appropriate composition of ChG [93], as well as SiN match that of LiNbO$_3$ (∼2.1). Accordingly, we have applied Ta$_2$O$_5$ [90], ChG [94] and SiN [95]–[99]
for rib-loading LiNbO\(_3\) thin films, in different generations of our EO and nonlinear devices. More recently, we have made chips with direct etching of LiNbO\(_3\) [100].

Based on these ultracompact waveguides, we demonstrated the first LiNbO\(_3\)-on-Si Mach-Zehnder (MZ) [90] and microring [94] modulators. Subsequently, 50-GHz MZ modulators were achieved, with a half-wave-voltage product, \(V_{\pi} \cdot L\), of 3.1 V·cm, which is several times lower than commercial modulators [95]. Although the employed EO or Pockels effect is in principle a second-order phenomenon (see Table I), the devices are not discussed further here. It is also noteworthy that there are other approaches based on bonding LiNbO\(_3\) on top of SOI waveguides, at the cost of lower optical confinement. We have recently reviewed the emerging field of ultracompact lithium niobate photonics for EO modulators [101], [102], but these other platforms have not been used for nonlinear-optic applications yet.

More recently, we have employed the discussed ultracompact LiNbO\(_3\) waveguide platform for nonlinear- and quantum-optic experiments. Our first report was on pulsed SHG on ultracompact periodically-poled thin films [96]. The schematic of the device shown in Fig. 3(a), which closely follows the envisioned one in Fig. 2(b). SHG at \(\sim 784\) nm is evident in the spectrum given in Fig. 3(b). This report may be considered the first demonstration of PPLN on Si substrates.

As detailed in the Appendix, another method to attain QPM is mode-shape modulation (MSM). A schematic of such a grating waveguide is shown in Fig. 3(c) [97]. In essence, the periodic perturbation of the guided modes acts as an extra phase term at our disposal, in order to achieve QPM. We have demonstrated such poling-free QPM for SHG under pulsed pumping, as reshowed in Fig. 3(d).

Furthermore, with improved processing and higher optical confinement in LiNbO\(_3\), PPLN waveguides lend themselves to continuous-wave (CW) SHG experiments [100]. A CW diode laser with \(\sim 1\) mW power is fiber-coupled in and out of the chip. A micrograph of the chip is shown in Fig. 3(e). After accounting for coupling loss at the chip facets, we obtain an on-chip nonlinear conversion efficiency of 1,230 %/W·cm\(^2\) in 0.6-mm-long devices [100]. Shown in Fig. 3(f) are more recent (unpublished) measured SHG response versus pump wavelength exhibiting even higher conversion efficiency of \(\sim 4,600\) %/W·cm\(^2\) near the phase-matching peak in 0.3-mm-long devices.

As explained in the Appendix and evidenced from more examples given below on other platforms, interpreting such apparently large length-normalized efficiencies in short devices needs careful deliberation of the propagation loss, as the quadratic dependence of output power versus length is strictly applicable to the low-loss regime. Also, the %/W and/or the absolute (unitless) power conversion efficiencies may be more meaningful figures for most applications. These latter efficiencies depend on the propagation loss too [78], [96], but are typically maximized at fairly long devices. Nonetheless, an undeniable advantage of the modern ultracompact devices, over bulkier older-generation counterparts, is the fact that the powers required to onset the desired nonlinearity is much smaller (e.g., 1 mW in the above demonstration [97]), as the nonlinear interaction of the mixing waves scales with the inverse of effective areas of the guided modes [78], [96].

The very same devices used for pulsed SHG experiments [96] were used for SPDC experiments in order to generate correlated biphotons [88], [99]. Indeed, the PPLN waveguide in Fig. 3(a) is quasi-phase-matched for a type-0
Fig. 4. SPDC in ultracompact LiNbO$_3$-on-Si integrated photonics: (a) Counts per second measured at detectors for signal and idler photons; (b) Measured photon pair coincidence rate at detectors; (c) Photon pair generation rate on-chip [98].

interaction, i.e., when all interacting waves are TE-polarized. A Ti:sapphire laser centered at 792 nm, with 81.8 MHz repetition rate, and <0.5 ps pulse-width, is coupled into the integrated photon pair source. The generated telecom band photon pairs are detected by tungsten-silicide superconducting nanowire single photon detectors. Signal, idler, and coincidence rates are detected for different powers coupled into the waveguide, as shown in Fig. 4. Initial results indicate a pair generation rate of $\sim 1$ MHz/mW. We also spectrally resolved the second-order coherence of biphoton pairs using the time-of-flight fiber spectrometry. Spectrally-resolved coincidence-to-accidental ratios (CAR) of over 300 are measured on the same chips [99].

B. Lithium Niobate on Insulator

Resembling the discussed LiNbO$_3$-on-Si platform, a similar thin-film wafer technology has been developed on LiNbO$_3$ substrates. Dubbed lithium niobate-on-insulator (LNOI) and commercialized by a vendor [103], research teams have been relying on these wafers for nonlinear integrated-optic devices (in addition to EO modulators), as reviewed below.

It should be emphasized that thin-film technologies based on bonding methods of native substrates is typically easier to achieve, as they avoid fabrication complications, such as thermal expansion coefficient mismatch for dissimilar materials. However, heterogeneous integration on silicon substrates is obviously preferred for several reasons mentioned in the Introduction (cost, scalability, potential compatibility with silicon photonics, ease of handling thermal cycles), as well as issues related to pyroelectric charges on bulk LiNbO$_3$ substrates.

At any rate and based on the accessible LNOI wafers, MPM between a 1385-nm pump at the fundamental quasi-TE$_{11}$ mode and the second-harmonic at the quasi-TE$_{12}$ mode on waveguides formed by proton exchange has attained 48 %/W.cm$^2$ SHG efficiency [104]. Also, Chang et al. [105] have reported on QPM in thin-film PPLN on LNOI waveguides, rib-loaded with SiN. SHG with CW pumping at 1530 nm and efficiency of 160 %/W.cm$^2$ was achieved in 4.8-mm-long thin-film PPLN waveguides. Wang et al. [106] demonstrated modal phase-matching between the fundamental and the second even-order modes of waveguides, defined by direct dry etching of LiNbO$_3$, that is between a 1550-nm pump at the quasi-TE$_{11}$ mode and the second-harmonic at the quasi-TE$_{21}$ mode, leading to a normalized SHG conversion efficiency of 41 %/W.cm$^2$. The same paper reported on SHG in grating waveguides, i.e., achieving QPM in periodic structures resembling the discussed MSM techniques. Here, both the pump and the second harmonic reside in their fundamental modes, resulting in 6.8 %/W.cm$^2$ efficiency. More recently, 4-mm-long PPLN devices on LNOI demonstrated 2,600 %/W.cm$^2$ efficiency [107], which is not only lower than the aforementioned $\sim 4,600$ %/W.cm$^2$ efficiency in LiNbO$_3$-on-Si (Fig. 3(f)), but also requires much higher pump power of 220 mW fed from an Erbium-doped fiber amplifier.

In the context of resonator-based works, cascaded stimulated Raman scattering, SHG and SFG have been also observed in LNOI microrings [108]. SHG (with a weak efficiency of $3.6 \times 10^{-6}$/mW) and SPDC (with a CAR of 43.1) was measured on X-cut LNOI microdisk resonators with a radius of 45 $\mu$m and a film thickness of 300 nm, via the aforementioned cyclic phase matching technique between quasi-TE and TM whispering-gallery modes [109]. An earlier report on a very similar LNOI platform (radius of 51 $\mu$m and film thickness of 700 nm) showed a higher SHG efficiency of $1.1 \times 10^{-3}$/mW [110].

The apparent low efficiency of phase matching via whispering gallery modes could be improved in the future. However, as mentioned before, coupling into microdisks suspended on silica pedestals is challenging and in general the scheme remains unattractive as a true integrated photonic platform, unless novel fabrication and processing solutions are developed to conveniently merge the devices with conventional waveguide-based photonic circuits.

C. Semiconductors on Silicon and Insulator

Selected early works on semiconductor-based nonlinear integrated optics using birefringent, modal and QPM techniques were briefed in Section IV. Here, some of the more
Fig. 5. SHG in ultracompact AlN-on-insulator integrated photonics: (a)-(b) Modal dispersion properties for attaining phase matching between the fundamental pump and higher-order signal [114]; (c)-(d) Schematic of microring chips and their resonance properties at the pump and second-harmonic [115].

recent works -- based on heterogeneous integration of compound semiconductors on a lower-index bottom clad layer, typically an insulator material like SiO₂ -- and their application in nonlinear integrated photonics are reviewed. The materials include the wide-bandgap semiconductor nitrides (AlN and GaN) and GaAs.

Nonlinear properties of AlN have been known for decades [111] and SHG experiments have been reported on AlN films sputtered on sapphire or silicon substrates [112], [113]. Pernice et al. [114] reported on the first deposition of c-axis-oriented (polycrystalline wurtzite lattice) AlN thin films on oxidized Si substrates and fabricated waveguides and microring devices for SHG on them. As reshown in Fig. 5(a)-(b), MPM was achieved between a 1550-nm pump at the fundamental mode and the second-harmonic wave at a higher-order even mode. The estimated value of $d_{\text{eff}}$ is in the order of 2.3 pm/V, which is fairly small and can be attributed to the polycrystalline nature of the sputtered thin film. In a follow-up paper [115], the same research team utilized microrings, with 30 μm radius, and measured loaded quality-factor, $Q$, of $\sim 2.3 \times 10^5$ and $\sim 1.6 \times 10^5$ at the pump and signal wavelengths, respectively (see Fig. 5(c)-(d)). Despite an even lower estimated value of $d_{\text{eff}} \sim 0.5$ pm/V, a high SHG normalized efficiency of 2,500 %/W at a low pump power of 27 mW, corresponding to an absolute conversion efficiency of 12% is realized.

As discussed before, periodic poling is possible in wurtzite GaN and has been investigated for SHG [71]. More recently, thin films of GaN on (111) Si substrates were transfer-bonded on another oxidized Si wafer and subsequently underwent substrate removal and chemical-mechanical polishing to attain GaN-on-insulator wafers [116]. MPM between a 1550-nm pump at the fundamental mode and the second-harmonic signal at the quasi-TE₃₁ mode was used in microring resonators with a radius of 40 μm for SHG characterization. The conversion efficiency of this device is $\sim 0.02$ %/W at $\sim 110$ mW of pump power.

Some of the early works on the widely-studied (Al)GaAs waveguides were discussed in Section IV. In recent years, thin-film AlGaAs-on-insulator on InP substrates have been demonstrated by wafer bonding and substrate removal techniques for optical comb generation via third-order, $\chi^{(3)}$, nonlinearity [22]. Very recently, Chang et al. [117] have applied similar bonding techniques to achieve a 150-nm-thick [001]-oriented GaAs die on an oxidized Si substrate with a thin (5 nm) SiN adhesion layer in between. Subsequently, 1.5-μm-wide channel waveguides were formed for SHG experiments with propagation loss of 1 to 2 dB/cm. MPM was achieved between a 2-μm-pump wavelength at the fundamental quasi-TE mode and second harmonic at the fundamental quasi-TM mode. As reshown in Fig. 6, normalized conversion efficiency of 13,000 %/W.cm² was measured, corresponding to $\sim 255$ %/W and $\sim 0.25\%$ in 1.4-mm-long waveguide and $\sim 1$ mW on-chip power.

Such low-power microwatt-range on-chip output powers might suffice the needs of some modern applications, e.g., octave-spanning optical comb generation, but would require efficient in- and out-coupling schemes. Also, the choice of pumping at 2-μm wavelength may be desirable for the self-referencing technique for comb stabilization [118], [119] in the telecom wavelength range. However, GaAs's bandgap prohibits operating at SHG signal wavelengths below $\sim 850$ nm for other applications. Also, for any nonlinear process, a high-power pump at below $\sim 1700$ nm would suffer from TPA and FCA, similar to the case of silicon discussed in Section II. Indeed, as a direct bandgap semiconductor, GaAs's TPA coefficient is $\sim 30$ times higher than silicon's [120]. More discussion on this issue is presented later in Section VI.B.
VI. ROAD AHEAD

A. Conventional Frequency Conversion

As evident from the presented literature, SHG has been the primary second-order process studied so far on heterogeneous integrated photonic platforms (of course, not neglecting the fairly advanced EO modulators, as well as some discussed preliminary SPDC experiments). This is sensible for proof-of-concept demonstrations on platforms in their infancy, since SHG is perhaps the easiest nonlinear process to realize in terms of measurement and/or device design and fabrication sophistication.

Other important optical-frequency conversion processes, listed in Table I, have yet to be demonstrated on heterogeneous platforms. Among them, DFG is a reachable near-term goal mentioned in Section II, the BOX layer, common in all these platforms, is lossy in most of this range. Also unfortunate is that SiO₂, and other dielectrics like SiN, generally become increasingly absorbent at wavelengths in the visible and shorter wavelengths, due to Rayleigh scattering and electronic resonances.

We have previously demonstrated the first Si-on-SiN waveguide platform, transparent up to 6.5 μm [121]. In principle, similar platforms can be developed on other high-index semiconductors, e.g., GaAs. However, SiN-based bottom cladding layer cannot be justified for waveguides on LiNbO₃ and large-bandgap semiconductors with comparable refractive indices. All-air-clad platforms are potential solutions for these materials, as we have previously implemented in the case of Si [122]–[124]. Fabrication challenges at smaller dimensions (roughly proportional to wavelength) notwithstanding, similar air-clad waveguide structures can be explored for visible-to-UV nonlinear integrated photonics.

Meanwhile, TPA and the FCA induced by it would arguably make any platform on low-bandgap semiconductors too lossy for a pump wavelength below its half-bandgap. As mentioned before, that is a long pump wavelength of 1700 nm in GaAs with a large TPA coefficient of 15 cm/GW. A mole fraction of 18% (Al₀.₁₈Ga₀.₈₂As) would blue-shift the half-gap wavelength to ~1.5 μm with a small TPA coefficient of 0.05 cm/GW [50]. AlGaAs-on-insulator dies and waveguides with closely similar composition (Al₀.₁₇Ga₀.₈₃As) have been developed in the past, albeit on InP substrates and for χ⁽³⁾-based applications [22]. Expectedly, similar thin films of AlGaAs can be developed on Si substrates and waveguides can be designed for χ⁽²⁾ processes, as well.

Finally, it is worth mentioning that there are several other nonlinear materials, e.g., the aforementioned LiTaO₃, KTP, and KNbO₃, as well as potassium dihydrogen phosphate (KH₂PO₄ or KDP), the β-phase barium borate (β-BaB₂O₄ or BBO) and barium titanate (BaTiO₃), some of which possess comparable or even higher nonlinear coefficients than those discussed in this paper [29], [32]. For a variety of
reasons, they have never matured into reliable integrated-optic platforms, but there is always a chance to revisit them, and particularly investigate their prospects for thin-film-on-silicon heterogeneous integration. For instance, 50-nm-thick BaTiO3 thin films grown on SOI substrates and rib-loaded with SiO2 (for lateral optical confinement) has been recently demonstrated [125]. The waveguides attain a propagation loss of 6 dB/cm, which is reasonably low to consider them for $\chi^{(2)}$ nonlinear experiments.

C. Nonlinear Integrated Photonic Circuits

Beyond the basic milestones of Section VI.A, that mimic standard experiments pursued in bulk optics for decades, demonstrating more sophisticated nonlinear photonic circuitry is expected to become trendy in the coming years. Such chips would include several passive and active linear, EO, and/or $\chi^{(2)}$- and $\chi^{(3)}$-based devices on the same chip that may require heterogeneous integration of multiple materials. Two applications are selected and briefly explained in the following.

1) Quantum Photonic Integrated Circuits: The emerging field of quantum information has found many applications in areas such as quantum cryptography and secure communication, quantum sensing and imaging, and quantum computing [126], [127]. Many of these applications have relied on single photons or entangled photon pairs together with linear optical processing in bulk optics, but quantum photonic integrated circuits are beginning to offer sophisticated and high-performance implementations [128].

Generation of entangled photon pairs by SPDC is one of the most efficient ways to attain sources for such quantum-optic applications. Albeit on conventional bulky LiNbO3 waveguides, a fairly sophisticated on-chip demonstration is shown in Fig. 8 [129]. The example circuit includes phase modulators (Segment I) for deterministic separation of degenerate photon pair generated at 1560-nm wavelength in the two PPLN waveguides via SPDC (Segment II), and followed by a quantum-interference photonic-based device (Segment III).

There are a few preliminary reports on SPDC experiments in ultracompact LiNbO3 integrated photonics [98], [99], [110]. With improved SPDC efficiencies, such high-contrast thin-film waveguide platforms allow denser and more complicated circuits feasible for more advanced quantum photonic demonstrations, such as hyperentanglement [130].

2) Cascaded $\chi^{(2)}$- and $\chi^{(3)}$ Integration: The vast majority of demonstrations in the long history of research in nonlinear integrated optics have been standalone devices for a single and specific functionality, as we saw several examples of on SHG throughout this paper. Emerging applications, however, require not only passive devices such as directional couplers, filters, delay lines, etc., but also integration of devices for different $\chi^{(2)}$ and $\chi^{(3)}$ processes on the same chip. A prime example is frequency-stabilized optical comb generation, where in conjunction with some linear-optic devices, octave-spanning SCG is generated on a $\chi^{(3)}$ material, while an SHG device is used for stabilizing the comb teeth by $f$-to-$2f$ carrier-envelope offset (CEO) locking [118], [119]. As mentioned in Section III, any material has $\chi^{(3)}$ nonlinearity, so a noncentrosymmetric $\chi^{(2)}$ material can, in principle, be used for both purposes. (Al)GaAs and (Al)GaN have both strong second- and third-order nonlinearity. Indeed, the same AIN waveguide has been appropriately exploited to simultaneously generate SCG, third-harmonic generation, SHG, and DFG [131]. In addition, the CEO frequency was directly detected on the same chip for optical-frequency comb self-referencing.

Alternatively, different materials can be selected for $\chi^{(2)}$ and $\chi^{(3)}$ functionalities. So far, only a couple of preliminary reports have appeared on addressing the fabrication challenges of integrating different $\chi^{(2)}$ and $\chi^{(3)}$ materials on the same chip. One example is heterogeneous integration of LiNbO3 and SiN waveguides on the same Si wafer, albeit with no nonlinear experiments realized on either device [132]. To this end, this author and coworkers have demonstrated integration of LiNbO3 and ChG waveguides on the previously-discussed LiNbO3-on-Si platform with some basic $\chi^{(3)}$ experiments on the ChG segment [133]. Coupling between the $\chi^{(3)}$ and $\chi^{(2)}$ waveguides at different elevations is achieved through low-loss mode-converting tapers mediated by a SiN layer between ChG and LiNbO3 (Fig. 9(a)). In addition, the ChG waveguides were designed for $\chi^{(3)}$-based four-wave mixing, which was successfully characterized on the cascaded waveguides (Fig. 9(b)). Furthermore, we have recently demonstrated octave-spanning SCG ($\sim$1 to $\sim$2 $\mu$m) on similar standalone ChG waveguides pumped at 1550 nm with ultralow-power and ultrashort (26 pJ and 240 fs) pulses (Fig. 9(c)) [21]. Accordingly and as schematically shown in Fig. 9(d), a complete (to-be-demonstrated) photonic circuit should include SiN passive filters, electrodes for poling LiNbO3, or other means of efficient SHG, e.g., MSM, in order to enforce $f$-to-$2f$ referencing by converting the 2-$\mu$m tail of the octave-spanning SCG to 1-$\mu$m and electronically beating it off-chip with the 1-$\mu$m tail for frequency-stabilized comb generation. Achieving tunable phase delays on the same the $\chi^{(2)}$-$\chi^{(3)}$ chips, as well as integrating III-V-based pump sources and photodetectors are more challenging milestones to consider in the future.
and waves at $\varphi_1$.

A. Theory of Second-Order Nonlinear Integrated Photonics

Suppose a multimode waveguide with modes indices $n$ and waves at $\omega_n$ angular frequencies propagating along the $z$ direction (see Fig. 2). Considering photon conservation of energy in the slowly-varying envelope approximation and for CW waves, we can expand the frequency-domain electric and magnetic fields, $\tilde{E}_q(r)$ and $\tilde{H}_q(r)$, as

$$\tilde{E}_q(r) = \sum_n \sum_v \left[ A_{vn}^{(n)}(z) e^{(n)}_v(x, y) e^{-j\beta_v^{(n)}z} \right],$$

$$\tilde{H}_q(r) = \sum_n \sum_v \left[ j A_{vn}^{(n)}(z) h^{(n)}_v(x, y) e^{-j\beta_v^{(n)}z} \right],$$

where $\beta_v^{(n)} = n \omega_n c / \epsilon_0$ are the modal propagation constants. Transverse components, $i$, of the eigenmodes $e_i^{(n)}$ and $h_i^{(n)}$ are orthonormal, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( e_i^{(m)\ast} \times h_i^{(n)} + e_i^{(n)\ast} \times h_i^{(m)} \right) \cdot u_\zeta \, dxdy = \delta_{mn}. \tag{A.3}$$

Then, using the perturbation theory it can be shown that

$$dA_q^{(n)}(z)/dz = -j \omega_q \sum_{mm'} \kappa_{X(2)}^{(mm'n)} A_{qj}^{(om)} A_{qk}^{(vm')} e^{-j(\beta_v^{(n)} + \beta_v^{(m')} - j\beta_v^{(n)})z}; \tag{A.4}$$

where $\kappa_{X(2)}^{(mm'n)}$ is the nonlinear coupling efficiency and $\Delta\beta_{SFG}^{(mm'n)}$ is the phase-mismatch factor related to the group-velocity dispersion (GVD). For example, the coupled-mode equations for SFG are

$$\frac{dA_{1q}^{(n)}}{dz} = -j \omega_3 \sum_{mm'} \kappa_{SFG}^{(mm'n)} A_{1j}^{(om)} A_{2k}^{(vm')} e^{-j\Delta\beta_{SFG}^{(mm'n)}z}, \tag{A.7}$$

$$\frac{dA_{1q}^{(m)}}{dz} = -j \omega_1 \sum_{mm'} \kappa_{SFG}^{(mm'n)} A_{2j}^{(m'n)} A_{3k}^{(nn')} e^{+\Delta\beta_{SFG}^{(mm'n)}z}, \tag{A.8}$$

$$\frac{dA_{2q}^{(m)}}{dz} = -j \omega_2 \sum_{mm'} \kappa_{SFG}^{(mm'n)} A_{1j}^{(m'n)} A_{3k}^{(nn')} e^{+\Delta\beta_{SFG}^{(mm'n)}z}. \tag{A.9}$$

It is reminded that the crystal axis, XYZ, may not necessarily match the $xyz$ coordinate of the waveguide. For instance, to take advantage of the highest $d_{31}$ coefficient, $d_{33}$ in trigonal group 3m crystals like LiNbO$_3$, the $Z$ axis is aligned in either the $x$ or $y$ direction of the waveguide with $z$-propagating direction (see Fig. 2). Also, the dispersion of $d_{31}$ is ignored, i.e., $d_{31}^{(1)} \approx d_{31}^{(2)} \approx d_{eff}^{(3)} \equiv d_{eff}$. Then, for quasi-TE and quasi-TM modes, i.e., $E_i^{(om)} = \tilde{E}_X^{(om)} = \tilde{E}_i^{(om)} \approx 0$. 

Fig. 9. (a) Design and simulation of ultracompact LiNbO$_3$-on-Si waveguide integrated with ChG waveguides via vertical waveguide couplers mediated by a SiN layer in between [133]; (b) Measured FWM results on the left waveguide in (a) [133]; (c) Standalone ChG-on-Si waveguides with measured octave-spanning SCG [21]; (d) Conceptual fully-integrated ChG-LiNbO$_3$ circuit for frequency-stabilized optical comb generation application.
we attain
\[
\kappa_{\text{SHG}}^{(m'n')} = 2\varepsilon_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d_{\text{eff}}(r) \tilde{e}_1^{(m)}(x, y) \tilde{e}_2^{(n')}(x, y) \times \tilde{e}_3^{(m'n')} (x, y) \, dx \, dy. \tag{A.10}
\]

Another important example is SHG. For simplicity, only single-mode waveguides are briefed, in order to avoid the complications of accounting for both the degenerate \(E_i^{(m)} E_i^{(m)}\) and nondegenerate \(E_i^{(m)} E_j^{(n'n')}\) terms in multimode waveguides. If the corresponding propagation loss terms, \(a_q\), for each mode and frequency, and defining two effective \(d\)-coefficients, \(d_{\text{eff}}\), are also included, the coupled-wave equations become
\[
\frac{dA_2}{dz} = -2j\omega \kappa_{\text{SHG}}^{(2\omega)} A_1^2 e^{-j\Delta\beta_{\text{SHG}}} - \frac{a_2 A_2^2}{2}, \tag{A.11}
\]
\[
\frac{dA_1}{dz} = -j\omega \kappa_{\text{SHG}}^{(\omega)} A_1^2 A_2 e^{+j\Delta\beta_{\text{SHG}}} - \frac{a_1 A_1^2}{2}, \tag{A.12}
\]
\[
\kappa_{\text{SHG}}^{(2\omega)} = \varepsilon_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d_{\text{eff}}^{(2\omega)} e_1^2 e_2^2 \, dx \, dy, \tag{A.13}
\]
\[
\kappa_{\text{SHG}}^{(\omega)} = 2\varepsilon_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d_{\text{eff}}^{(1)} e_2 e_1 e_1 \, dx \, dy, \tag{A.14}
\]
\[
\Delta\beta_{\text{SHG}} = 2\beta_1 - \beta_2. \tag{A.15}
\]

Again, in the 3m crystals example, \(d_{\text{eff}}^{(2\omega)} \approx d_{\text{eff}}^{(1)} = d_3\). The \(dA_2/dz\) equation lends itself to analytical solutions in the undepleted pump approximation, i.e., assuming \(A_1\) is not a function of \(z\). Also, assuming that \(d_{\text{eff}}\) is not a function of \(z\), we can define the power conversion efficiency in term of the device length, \(L\), and input and output powers \(P_q = A_q A_q^*/4\), as
\[
\eta_{\text{SHG}} = \frac{P_3}{P_1(0)} = 16\varepsilon_0^2 (\kappa_{\text{SHG}}^{(2\omega)})^2 P_1(0) L^2 \sin^2 \left(\frac{\Delta\beta_{\text{SHG}} L}{2}\right). \tag{A.16}
\]

The loss terms, \(a_q\), can also be incorporated in the analytical solution by the transformation \(A_q = A_q e^{-a_q z/2}\). However, it is also noted that with different field normalizations, e.g., such that \(P_q = A_q A_q^*/4\), and by defining \(\kappa_{\text{SHG}}^{(2\omega)}\) half of what we did, the factor of 16 in the above equation can be absorbed in \((\kappa_{\text{SHG}}^{(2\omega)})^2\) [32], [36]. At any rate and for phase-matched devices, \(\Delta\beta_{\text{SHG}} = 0\), and \(\eta_{\text{SHG}} \propto P_1(0) L^2\), thus it is common to normalize to \(P_1(0)\), as well as to \(L^2\), and express the efficiency in units of \%/W and \%/W.cm\(^2\), respectively, as power- and length-independent figures of merit.

It may, however, sometimes be misleading to use these latter efficiencies, since for \(P_1(0) \ll 1\) W, absurdly high \%/W numbers might be calculated, while the actual \(\eta_{\text{SHG}}\) could be dismal. Or, for \(L \ll 1\) cm, the \%/W.cm\(^2\) value keeps increasing while the more relevant \%/W for a reasonable length might be very small. It is also meaningless to apply the \%/W.cm\(^2\) efficiency to very lossy or pulsed configurations, where the \(L^2\) dependency is generally not valid.

A major challenge for \(\chi^{(2)}\) effects is devising techniques to eliminate or alleviate efficiency suppression due to the phase-mismatch, \(\Delta\beta\). Birefringence phase matching has been extensively studied in bulk optics (see Section IV), but it is not generally applicable to guided waves, since waveguide dispersion dominates over material dispersion. The aforementioned MPM, such that \(\Delta\beta \approx 0\), is sometimes possible, for example by having \(\omega\) in the fundamental and the second harmonic \(2\omega\) in a higher-order mode, at the expense of lower mode overlap integral. However, this feature may not be always at our disposal due to the constraints of the GVD design. Also, it is not always convenient or desirable to operate in higher-order modes.

If \(d_{\text{eff}}(r)\) and/or \(\tilde{e}(r)\) are properly and periodically modulated in the \(z\)-direction, the suppressing impact of phase mismatch can be compensated to a large extent, even in single-mode waveguides. Such methods are generally called QPM, as several experimental examples are given in this paper.

As discussed in Section IV, periodically poling \(d_{\text{eff}}(z)\) in ferroelectric materials is one of the most well-established and efficient techniques to achieve QPM. Alternatively, the previously discussed MSM method, or grating-assisted QPM, can be exploited by periodically perturbing \(\tilde{e}(q)\), hence the nonlinear coupling efficiencies. This method avoids the complications of poling and is applicable to non-ferroelectric materials, such as cubic compound semiconductors, but it is necessarily less efficient than periodically poling, as proven below. In essence, periodic poling is a \(\chi^{(2)}\)-modulation perturbation for QPM, while MSM is a perturbation mediated by the linear or \(\chi^{(1)}\) susceptibility.

Mathematically, both of these QPM methods can be modeled by assuming periodic nonlinear \(\kappa(z)\) (dropping the sub- and superscripts). With \(\tilde{k}_p\) Fourier components, \(p\), for a periodicity \(K = 2\pi/\Lambda\) we can write
\[
\kappa(z) \equiv \tilde{k}_0 + \sum_{p \neq 0} \tilde{k}_p e^{-jpKz}, \quad \tilde{k}_p = \frac{1}{\Lambda} \int_0^\Lambda \kappa(z) e^{jpKz} dz. \tag{A.17}
\]

The DC or \(p = 0\), Fourier component, \(\tilde{k}_0\), is zero in the case of periodic poling with \(d_{\text{eff}}(z)\) flipping from \(+d_{\text{eff}}\) to \(-d_{\text{eff}}\) domains in one period, but is nonzero for MSM. \(\tilde{k}_0\) is the original (unmodulated) nonlinear coupling coefficient in the MSM scheme.

QPM occurs at only one integer \(p\), that is when
\[
\Delta\beta_p = \Delta\beta_{\text{SHG}} + \frac{2\pi p}{\Lambda} \approx 0. \tag{A.18}
\]

Hence, similar to Bragg grating waveguides, we have to choose the order of grating, \(p\), and the corresponding \(\Lambda\) for QPM and then can neglect all the other orders. Accordingly,
\[
A_2(z) = -2j\omega A_1^2(0) \times \left(\int_{0}^{z} \tilde{k}_0 e^{-j\Delta\beta_{\text{SHG}} z'} dz' + \int_{0}^{z} \tilde{k}_p e^{-j\Delta\beta_p z'} dz'\right). \tag{A.19}
\]
and
\[
P_2(z) = \frac{1}{P_1(z)} \frac{3}{2} \left[ z^2 \kappa_0^{2} \sin^2 \left( \frac{\Delta \beta_{\text{SHG}} z}{2} \right) + \kappa_p^2 \sin^2 \left( \frac{\Delta \beta_p z}{2} \right) \right]
+ \frac{2 \kappa_0 \kappa_p \sin \left( \frac{\Delta \beta_{\text{SHG}} z}{2} \right) \sin \left( \frac{\Delta \beta_p z}{2} \right)}{2} \cdot \cos \left( \frac{\Delta \beta_p - \Delta \beta_{\text{SHG}}}{2} z \right) \right].
\]

(A.20)

Clearly, the \( \kappa_0 \)-dependent terms of MSM suppress the efficiency. In the case of periodic poling with \( \kappa_0 = 0 \),
\[
P_2(z) = \frac{1}{\sqrt{P_1(z)}} \left( \frac{3}{2} \right) \frac{1}{2} \kappa_p^2 \sin^2 \left( \frac{\Delta \beta_p z}{2} \right),
\]

(A.21)

which is identical to what we had before in (A.16), except that \( \kappa_{\text{SHG}}^{(2\omega)} \) is replaced with \( \kappa_p \) and \( \Delta \beta_{\text{SHG}} \) replaced with \( \Delta \beta_p \). The output-input relation is still quadratic versus length, but since the harmonics \( \kappa_p < \kappa_{\text{SHG}} \), QPM devices (\( \Delta \beta_p \approx 0 \)) cannot be as efficient as ideally phase-matched devices (\( \Delta \beta_{\text{SHG}} \approx 0 \)).

However, the aforementioned mode overlap integral reduction by operating the harmonic in a higher-order mode of a multimode waveguide in the MPM technique is generally even less efficient and also unattractive due to difficulties of dealing with high-order modes.

It is finally emphasized that the above model is strictly applicable to CW mixing waves and long pulses that can be approximated as quasi-CW. For ultrashort pulses, the theory is more complicated and involves accounting for effects such as pulse spectra, the nonlinear-device optical frequency response, as well as temporal walk-off among the interacting waves in long devices. Such theories have been developed for bulk nonlinear-optic configurations that, in addition, suffer from spatial walk-off [134], [135]. Modifying these existing theories and applying them to the present case of nonlinear integrated photonics is beyond the limited space of this paper.

ACKNOWLEDGMENT

The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the U.S. Government.

REFERENCES


Sasan Fathpour (S’01–M’04–SM’13) received the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2005. He then joined the Electrical Engineering Department, UCLA, as a Post-Doctoral Fellow. He joined the Faculty of CREOL, The College of Optics and Photonics, University of Central Florida (UCF), in 2008 and was promoted to Associate and Full Professor in 2014 and 2018, respectively. He holds a joint appointment with the Department of Electrical and Computer Engineering, UCF.

His current research interests include heterogeneous integrated photonics, nonlinear integrated optics, silicon photonics, and nonconventional optical waveguide platforms operating in the mid-wave- and near-infrared and visible wavelength ranges. He is the Co-Founder of Partow Technologies LLC, a spinoff company from his CREOL team’s research that focuses on commercializing nanophotonic thin-film technologies for telecom and sensing applications.

His Ph.D. dissertation, under the supervision of Prof. P. Bhattacharya, was on gallium-arsenide-based semiconductor lasers and spin-polarized light-emitting diodes, where he, for instance, demonstrated the highest-bandwidth quantum dot lasers to date. His post-doctoral research, at Prof. B. Jalali’s group, was primarily on energy harvesting in silicon photonics and received him the UCLA Chancellor’s Award for Post-Doctoral Research.

He is a fellow of OSA and a Senior Member of SPIE. He was a recipient of the NSF CAREER Award in 2012 and the ONR Young Investigator Award in 2013. He has also received several research and teaching incentive awards at UCF, including the Reach for the Stars Award in 2015.

He is an author or co-author of over 160 journal and conference papers, book chapters, and patents. He is the Co-Editor of the book *Silicon Photonics for Telecommunications and Biomedicine*, published by CRC Press in 2012. He has been a Technical Program Committee Member of several conferences, including CLEO from 2016 to 2018 and Frontiers in Optics (FiO 2016). He has been a Guest Editor of the SPIE’s Journal of Nanophotonics from 2014 to 2015, the Symposium Co-Chair of the 2015 Materials Research Society Fall Meeting, and the Chair from 2014 to 2016, and the Vice Chair from 2012 to 2014 of Short Courses at the Conference on Lasers and Electrooptics, CLEO.