This paper presents the recent progress on integrated second-order nonlinear waveguides on silicon substrates for second-harmonic generation. In particular, demonstrations of thin-film lithium niobate, III–V compound semiconductor and dielectric waveguides integrated on silicon substrates are reviewed. For completeness, the fundamentals of the nonlinear optical processes involved are briefly introduced. Methods demonstrated for phase matching, e.g., periodic poling and mode-shape modulation, in the compact integrated devices are discussed. Finally, an outlook for how integrated photonics may benefit from the progress in this field is provided.

1. Introduction

Optical three-wave mixing\cite{1-2} has been pursued for the generation of coherent light from the ultraviolet to the infrared since 1961.\cite{1-22} The phenomenon of three-wave mixing occurs in transparent non-centrosymmetric materials that exhibit a sufficiently strong nonlinear response to intense coherent radiation. Waves of often broadly separated frequencies, commonly referred to as the pump, signal, and idler, with angular frequencies $\omega_p$, $\omega_s$, and $\omega_i$, are coupled to each other, leading to frequency conversion. Second-harmonic generation (SHG), sum- and difference-frequency generation (SFG and DFG), spontaneous parametric down conversion (SPDC), optical parametric oscillation (OPO), and optical parametric amplification (OPA) are some typical three-wave mixing processes. These were originally demonstrated based on bulk crystals in bench top configurations,\cite{1-10} and subsequently in integrated waveguides.\cite{17-22} Optical phase matching is required in these processes to compensate for the dispersion of the interacting waves to ensure efficient energy transfer.\cite{23} Thus, phase matching is one of the key factors that ought to be engineered in waveguide implementations of three-wave mixing. A significant advantage of using waveguides is the increase in the nonlinear overlap between the interacting waves, leading to an increase in the nonlinear efficiency.\cite{24} In particular, SHG, where two pump photons of frequency $\omega$ are converted to a single signal photon at frequency $2\omega$, has received significant attention, due its many applications, e.g., frequency stabilization and imaging microscopy. This review paper will largely use SHG to elucidate the progress made in second-order nonlinear integrated photonics on silicon.

2. Nonlinear Coupled-Mode Equations

Before delving into the existing integrated photonic solutions for second-order nonlinear optics on silicon, we present a theoretical formulation of SHG, which is a special case of three-wave mixing with $\omega_p = \omega_s = \omega_i/2$. Clearly, the following formulation can be readily extended to other three-wave mixing processes. Dropping the subscript notations for the involved angular frequencies, the nonlinear polarization response at the idler frequency $2\omega$, $P_{2\omega}$, is given by\cite{1,2,3}

$$P_{2\omega} = 2dE_0E_\omega.$$ \hspace{1cm} (1)

where $E_\omega$ is the pump electric field at frequency $\omega$, and $d$ is the nonlinear tensor, equal to half the second-order nonlinear susceptibility tensor, $\chi^{(2)}$. We utilize the local normal-mode expansion (LNME),\cite{25} which is based on using the local eigenmodes of a spatially varying waveguide. The normalized field amplitudes of the second harmonic (SH) signal $a_{2\omega}(z)$, and the pump $a_\omega(z)$, at frequencies $2\omega$ and $\omega$, respectively, satisfy the following coupled-mode equations in a periodic waveguide\cite{25}

$$\frac{d}{dz}a_{2\omega}(z) = -i(a_\omega(z))^2 e^{i\Delta k z}f(z) - \frac{a_{2\omega}}{2} a_{2\omega}(z),$$ \hspace{1cm} (2a)

$$\frac{d}{dz}a_\omega(z) = -ia_{2\omega}(z)a_{2\omega*}(z)e^{-i\Delta k z}f^*(z) - \frac{a_\omega}{2} a_\omega(z),$$ \hspace{1cm} (2b)

where $a_{2\omega}$ and $a_\omega$ are the waveguide propagation losses at frequencies $2\omega$ and $\omega$, respectively, and $\Delta k_0$ is the phase mismatch between the signal and the pump waves, averaged over one period of propagation length. $f(z)$ is a locally varying nonlinear coupling of the propagation equations that captures all the effects of the modulation of the waveguide, and has the form\cite{25}
f(z) = \sqrt{2\omega} \left( \frac{\mu_0}{\varepsilon_0} \right)^{1/2} \exp \left[ -i \int_0^z \left\{ 2\beta_{\omega}(\zeta) - \beta_{2\omega}(\zeta) + \Delta\beta_0 \right\} d\zeta \right] \Gamma(z),
\Gamma(z) = \int d(x, y, z) \left\{ \left( \int d(x, y, z) \right)^2 \right\}^{1/2},
(3)

where \( \beta_{2\omega}(z) \) and \( \beta_{\omega}(z) \) are the local propagation constants, \( n_{\text{eff}}^2(z) \) and \( n_{\text{eff}}^0(z) \) are the local effective indices, and \( E^0(x, y, z) \) and \( E^2(x, y, z) \) are the local transverse mode profiles of the signal and pump waves, respectively. As mentioned, \( d(x, y, z) \) is the second-order nonlinear \( d \) tensor that mediates the nonlinear process.

The origin of \( \Delta\beta_0 \) lies in the dispersion of the effective index of the waveguide, which has contributions from the material dispersions of the core and cladding materials, and the dispersion due to the optical confinement offered by the waveguide.[26,27] This leads to different phase velocities for the pump and SH waves. Efficient SHG requires a mitigation of \( \Delta\beta_0 \) by inducing an appropriate \( z \)-dependent variation in \( \beta(z) \). Phase matching, i.e., the compensation of the different phase velocities, is crucial for achieving high-efficiency nonlinear frequency conversion. The term quasi-phase matching (QPM) is used when this compensation is periodic. The accepted measure of the efficacy of the phase matching is the normalized conversion efficiency, which is a figure of merit of the waveguide itself, independent of the input power and the mode of operation, pulsed or continuous wave (CW). It is calculated, in units of \( 1/(W \cdot \text{cm}^2) \), as
\[
\eta_0 = \left( \frac{1}{\Lambda} \int_0^\Lambda f(z) \exp\left[ iz(-2\pi q/\Lambda + \Delta\beta_0)\right] dz \right)^2,
(4)
\]
where \( \Lambda \) is the period along the propagation length, and \( q \) is an integer chosen to minimize \( \Delta\beta_0(2\pi q/\Lambda) \). This amounts to calculating the Fourier series expansion of \( f(z) \) and choosing the coefficient that best cancels out \( \Delta\beta_0 \) in the phase. As an illustration, the evolution of the pump and the SH signals is simulated in Figure 1. Two different values of \( \eta_0 \), 50%//(W \cdot \text{cm}^2) (solid traces) and 1000%//(W \cdot \text{cm}^2) (dashed traces), are used as examples. These roughly correspond to a traditional (Section 3) and a novel integrated (Section 4.1) nonlinear waveguide for SHG. The input CW pump power is 10 mW for both cases, which is representative of a typical integrated photonic laser source. A strong increase in frequency conversion with an increase \( \eta_0 \) is clearly seen.

Several materials have been used for three-wave mixing in bulk crystals using free-space Gaussian beams. Table 1 lists some key materials, along with their largest nonlinear coefficient, normalized to the \( d_{33} \) coefficient of lithium niobate (LN), viz., 30 pm/V. LN is chosen as the benchmark, since it is perhaps the most heavily used material in this field. Included in

Table 1 are some III–V compound semiconductors. Nonlinear waveguides have been able to utilize only a few of these materials, due to limitations in material processing and phase matching.

The approaches used for phase matching in integrated photonic waveguides can be largely classified into two categories. The first well-established category is QPM, and is outlined in Sections 3 and 4. The second category relies on minimizing \( \Delta\beta_0 \) (to zero) through modal dispersion, i.e., engineering the dispersion of the waveguide by choosing suitable materials for the core and cladding and an appropriate waveguide geometry. Some examples that utilize this approach are outlined in Section 5. The following sections review some particular waveguide implementations of 3D optics, with an aim of emphasizing the prominent pursued approaches, rather than forming an exhaustive chronological list.
3. Periodically Poled Lithium Niobate (PPLN)

The most popular approach to phase matching has been QPM. Periodically poled lithium niobate (PPLN) waveguides, based on QPM, represent the most successful implementation of $\chi^{(2)}$ waveguides.\[17–22\] LN offers high nonlinear coefficients and a broad transmission window.\[1\] QPM is realized through the periodic reversal of the crystal domain of LN, which is also a ferroelectric material. Called periodic poling, the process induces a reversal in the sign of certain coefficients of the $d$ tensor, which are then chosen to mediate the nonlinear frequency mixing. PPLN has led to numerous successful demonstrations of three-wave mixing. High-power conversion in SHG has been demonstrated as well (Figure 2).\[28,29\] Periodic poling has also been demonstrated in other ferroelectrics and polymers.\[17,18\]

However, conventional PPLN waveguides are not readily compatible with modern integrated photonics. These waveguides are formed in LN single crystal wafers by the diffusion of metals, such as titanium,\[30\] or protons.\[31\] They offer low refractive index contrast, with weak optical confinement, large waveguide cross-sections (\(\approx 10\,\mu m^2\)), and large bending radii. In contrast, modern integrated photonics is largely based around oxidized silicon substrates, with tightly confined dry etched waveguides of submicron cross-section. While dry plasma etching and wet etching have been pursued as alternative approaches to defining large LN waveguides,\[32,33\] these methods have not yet been successfully demonstrated for small compact waveguides. The incompatibility of conventional PPLN waveguides with silicon substrates has driven the pursuit of thin film PPLN waveguides on silicon, as follows.

4. Thin Film LN on Silicon

The ideal solution to realizing efficient $\chi^{(2)}$ nonlinearities on silicon lies in forming hybrid compact LN waveguides on silicon. Such an approach aims at combining the benefits of conventional PPLN, i.e., the material properties of LN, with the benefits of modern integrated photonics, i.e., compact waveguides on robust and inexpensive silicon substrates. This has been realized by our group using thin-film LN on silicon.\[34\] Bulk single crystal LN wafers are bonded at room temperature onto oxidized silicon substrates after ion implantation, and are thermally sliced to form thin films of LN, as illustrated in Figure 3. Tightly confined waveguides are formed by rib-loading these thin films with refractive-index-matched materials. Rib loading entirely circumvents the requirement of etching the LN to induce lateral confinement, since LN is a very hard material to dry etch, especially for low-loss submicron cross-section waveguides. Previously, we have developed materials such as tantalum pentoxide, and chalcogenide glass ribs in order to achieve low-loss waveguides,\[35–37\] and used the hybrid waveguides to demonstrate optical modulators.\[34,38,39\] More recently, we have used silicon nitride (SiN), which is well established for waveguiding applications, due to its broad transmission window and ease of processing, to demonstrate optical modulators,\[40\] PPLN\[41\] (Section 4.1), and mode-shape modulation, a specific variant of grating-assisted QPM, in thin film LN on silicon,\[42\] as described in more detail in Section 4.2.

Table 1. Common $\chi^{(2)}$ materials and their largest nonlinear coefficient normalized to $d_{33}$ of lithium niobate.\[1\]

<table>
<thead>
<tr>
<th>Material</th>
<th>$d_{ij}/d_{33,\text{LN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiNbO$_3$</td>
<td>1</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>0.014</td>
</tr>
<tr>
<td>KD$_2$PO$_4$</td>
<td>0.014</td>
</tr>
<tr>
<td>$\beta$-BaB$_2$O$_4$</td>
<td>0.073</td>
</tr>
<tr>
<td>LiTaO$_3$</td>
<td>0.867</td>
</tr>
<tr>
<td>AlN</td>
<td>0.033</td>
</tr>
<tr>
<td>GaN</td>
<td>0.260</td>
</tr>
<tr>
<td>GaAs</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Figure 1. The coupled-mode equations presented in Eqs. (2a) and (2b) are simulated for $\eta_0$ of $50\%/(W\,cm^2)$ (solid traces) and $1000\%/(W\,cm^2)$ (dashed traces). The critical impact of $\eta_0$ for low CW input power applications (10 mW in this example) is visible in the pump depletion achieved in the higher $\eta_0$ trace.

Figure 2. a) Cross-section of a PPLN wafer after etching in hydrofluoric acid to enhance visibility of the poled regions. The vertical stripes correspond to alternating poled domain orientations.\[28\] b) SHG from 1064 to 532 nm (green) using a PPLN waveguide, with 42% single-pass internal power conversion. Reproduced with permission\[28\]. Copyright 1997, The Optical Society.
Compact PPLN waveguides are formed on silicon using the thin film bonding and SiN bonding described above. The Y-cut LN slab is 400 nm thick, bonded onto a 2000-nm-thick SiO₂ lower cladding layer on a silicon substrate, rib loaded by a SiN rib that is 2000 nm wide and 400 nm tall (Figure 4a).

The poling period is determined to be ≈5 μm through eigenmode simulations for TE-polarized SHG, with a TE pump at 1580 nm (Figure 4b and c). Including propagation loss, the normalized conversion efficiency is estimated to be 1400 W⁻¹ cm⁻², with a mode overlap integral of 2.3 μm². The increase in optical confinement yields an increase in nonlinear power conversion, as indicated in Figure 5. The submicron PPLN waveguides offer a consistent increase in frequency conversion (around an order of magnitude) relative to conventional PPLN for low input powers. Depending on the power level, conventional PPLN would require propagation lengths greater than four times that of thin film PPLN for a given power conversion. Additionally, at high input powers thin film PPLN offers high efficiency power conversion in less than 1 cm of length, even at higher propagation losses. The fabricated waveguides are pumped using a pulsed source and the SHG signal is measured on an optical spectrum analyzer as shown in Figure 6a–c, phase matching around a fundamental wavelength of 1580 nm. The quadratic nature of the nonlinear SHG process is confirmed (see Figure 6d).

4.2. Mode-Shape Modulation in Thin Film LN on Silicon

As an alternative to the intensive fabrication requirements of obtaining high-fidelity periodic poling, we have recently also demonstrated mode-shape modulation, a variant of grating-assisted QPM in thin-film LN waveguides,[42] shown in Figure 7a and b. The waveguide is comprised of a 600-nm-thick LN slab, which is rib-loaded by a grating waveguide of SiN. The periodicity of the grating is fixed by the extent of the lateral sinusoidal variation of the grating, as determined through simulations using an eigenmode solver. The pump and harmonic eigenmodes are shown in Figure 7c at a grating width of 1095 nm. The periodic spatial variation of the eigenmodes induced by the lateral grating induces a periodic spatial variation in f(z). In particular, the nonlinear growth of the SH field is driven by the first-order term in the Fourier series expansion of f(z), which cancels out the exp(iΔβC0) phase terms in Eqs. (2a) and (2b), while the constant term in the series expansion results in fast oscillations of the SH field, shown in Figure 7d. Measurements on the fabricated waveguide are shown in Figure 7c and f.

While the normalized conversion efficiency of conventional PPLN is somewhat higher (≈40% W⁻¹ cm⁻²), this approach circumvents the challenges of achieving high-fidelity periodic poling in thin film LN, and is broadly applicable to materials which cannot be poled. The physical dimensions of the waveguide grating strongly influence the SHG process. The thickness of the LN thin film influences the grating induced waveguide propagation loss through the amount of scattering encountered by the waveguide modes. The Fourier series coefficients of f(z) also depend on this. Additionally, the extent of the grating also affects the nonlinear conversion. A stronger grating increases the nonlinear coupling. However, a stronger grating also increases the propagation loss. Therefore, a delicate balance is required between the propagation loss and f(z).

Figure 3. a–c) LN thin-film wafer bonding scheme on silicon, including ion implantation, wafer bonding, and thermal crystal slicing steps. d) A picture of a 3” LN thin film wafer bonded to a 4” oxidized silicon wafer. Panel (d) Reproduced with permission[34]. Copyright 2013, The Optical Society.

Figure 4. a) A schematic of a thin film PPLN on Si waveguide. b and c) Eigenmode simulations in COMSOL™ at the pump (1580 nm) and the SH (790 nm). Reproduced with permission[41]. Copyright 2016, The Optical Society.
5. Other Approaches to $\chi^{(2)}$ Waveguides on Silicon

A few other approaches which have been explored toward realizing second order nonlinear waveguides on silicon are reviewed in this section.

5.1. $\chi^{(2)}$ in III–V Compound Semiconductors and Dielectrics on Silicon

As noted in Table 1, GaAs has a very high second-order nonlinear coefficient. However, implementing efficient phase matching in GaAs waveguides has been elusive. For instance, poling techniques are not applicable since the material is not ferroelectric. Other techniques, such as orientation patterning, grating-assisted quasi-phase matching, and modal phase matching have been pursued instead with varying degrees of success.[43–46] However, none of these techniques has been investigated on silicon substrates for second-harmonic generation to date. It is noted that recently, thin-film AlGaAs-on-insulator nanowaveguides on indium phosphide substrates have been demonstrated for third-order nonlinear optics.[47] It is expected that such waveguides will be realized on silicon substrates in the future, and second-order nonlinear effects can be realized on them by adapting guidelines already established for typical AlGaAs nanowaveguides.[48]

 Nonetheless, aluminum nitride (AlN) is a III–V compound semiconductor recently pursued for SHG on silicon.[49–51] Prior to the pursuit of waveguides, the $\chi^{(2)}$ response of sputtered AlN thin films was confirmed via optical reflection measurements for SHG.[52] However, these thin films are typically polycrystalline. Therefore, the effective $\chi^{(2)}$, as confirmed by SHG measurements, is often limited to a few pm/V. This naturally limits the conversion efficiency. Nevertheless, when used in a doubly resonant high quality factor microring configuration, in one particular case, AlN has yielded impressive SHG conversion efficiency (Figure 8a), as high as 2300%/W.[51]

Another III–V compound semiconductor that has been investigated for SHG on silicon is gallium nitride (GaN).[53,54] Crystalline GaN thin films have been transferred to oxidized silicon via bonding, backside removal, and chemical–mechanical polishing, and patterned to form waveguides and microrings (Figure 8b).[53] Aided by its crystalline nature, the GaN thin films exhibit a higher $\chi^{(2)}$ than the polycrystalline AlN thin films. However, the quality factor of the GaN microrings is lower than the AlN counterparts, resulting in an overall lower conversion efficiency, as seen in Figure 8c.[53]

In contrast to III–V compound semiconductors, an interesting dielectric in which SHG has been demonstrated is silicon nitride (SiN).[55–57] SiN has been very successfully used for demonstrating $\chi^{(3)}$ nonlinear optical effects, such as supercontinuum generation in straight waveguides and frequency comb generation in microring resonators.[58–61] A $\chi^{(2)}$ response can be elicited in SiN through applying strain, due to its heteropolar bonds. Silicon dioxide (SiO2) claddings have been used to break the symmetry of SiN (Figure 8d and e),[55] and electric field-induced second harmonic generation (EFISHG) has been demonstrated, as well.[56] However, the performance of SiN waveguides for SHG has been limited by the low

---

*Figure 5.* Numerical simulations for CW SHG power using PPLN waveguides – (a and b). For low input powers (100 μW), thin film PPLN consistently presents stronger SHG than conventional PPLN. c and d) Thin film PPLN offers ~50% power conversion at higher CW input powers (1 W) in less than 1 cm of propagation, regardless of propagation loss, while conventional PPLN solutions require longer lengths. Reproduced with permission.[41]. Copyright 2016, The Optical Society.

*Figure 6.* a) Measured frequency-doubled signal from thin film PPLN around 788 nm. b) Output spectra, displaced by 60 dB, for increasing input pump powers (bottom to top). c) SHG traces with input average powers of 1.67, 0.87, and 0.31 mW. d) Quadratic dependence of the 788 nm signal power on the input pump power with a slope of 1.91 on a log scale. Reproduced with permission[41]. Copyright 2016, The Optical Society.
nonlinearity itself. It is important to note that several recent reports suggest that the $\chi^{(2)}$ response of SiN is influenced not only by strain and interface effects, but also defects and the overall material composition.\[^{62-66}\]

Each of the above three platforms have relied on modal phase matching, where the phase velocities of the pump and SH modes are engineered to be equal. In essence, the dispersion due to the optical confinement of the interacting waveguide modes nullifies the dispersion of the core and cladding materials. Higher order waveguide modes are used for higher frequency modes due to the normal dispersion of the constituent materials.\[^{55,56,67-69}\] However, the use of higher-order modes compromises the nonlinear mode overlap and thus the nonlinear conversion efficiency.

5.2. $\chi^{(2)}$ in Silicon Waveguides

Silicon waveguides using the silicon-on-insulator (SOI) platform are well established as building blocks for large-scale integrated devices.
silicon limits light generation to around 1.1 μm. The limitations of silicon also become relevant. The bandgap of silicon and the transmission and absorption properties are below 2.2 μm, with a strong two-photon absorption around 0.5 μm, indicating limited transparency and high absorption. This is expected to help analyze the maximum achievable conversion efficiency in silicon waveguides. However, the performance of silicon waveguides is still debatable, and the resolution of this uncertainty is expected to help analyze the maximum achievable response, which may be extracted from silicon waveguides using this approach.

An alternative approach to SHG in Si uses EFISHG, where a d.c. electric field at zero frequency interacts with the strong nonlinear tensor of Si to enable SHG. EFISHG has demonstrated high effective nonlinear coefficients in Si, at the cost of high propagation losses. These waveguides typically have SiN cladding layers that break the crystal symmetry in the Si waveguides. It has been accordingly argued that the majority of the nonlinear response originates from the SiN cladding layer and not the Si core. At this moment, it is fair to state that the physical origin of the nonlinear response in silicon waveguides is still debatable, and requires further investigation. The resolution of this uncertainty is expected to help analyze the maximum achievable response, which may be extracted from silicon waveguides using this approach.

In addition to the discussed complications of the aforementioned approaches on Si, the transmission and absorption limitations of silicon also become relevant. The bandgap of silicon limits light generation to around 1.1 μm at the lower wavelength edge. Additionally, strong two-photon and free-carrier absorption effects below 2.2 μm limit the spectral windows at which Si could be optically pulse-pumped for frequency generation. On the other hand, the heterogeneous approaches of Sections 4.1, 4.2, and 5.1 benefit from the material properties and prior comprehensive characterization of the employed nonlinear material. This comes at the cost and complications of the heterogeneous integration process. At present, the overall nonlinear device performance offered by heterogeneously integrated materials trumps the performance achieved so far in silicon waveguides. However, the performance of the silicon waveguides may still potentially be improved, perhaps even to the point of competing with the performance of heterogeneously integrated solutions.

6. Outlook

Several schemes for second-order nonlinear optics have been established in waveguides on silicon. Each of these schemes presents trade-offs between performance and different aspects of fabrication. Realizing the full potential and benefit of these solutions would involve the integration of different optical functionalities on a common substrate, preferably silicon. This can be achieved through the heterogeneous integration of different materials on the same silicon chip, where each material system is already established for a particular set of functions. This is the overarching approach commonly employed for integrating active laser sources on silicon photonics chips. Numerous other waveguide devices have been realized using this approach as well. While the dense integration of multiple functions on a photonic chip has been repeatedly demonstrated, the pursuit of integrated systems which utilize second-order nonlinear processes is still nascent. One can easily envision frequency referencing, photon pair generation, and tunable optical parametric oscillators as some second-order functionalities which could be integrated with existing on-chip laser sources. Difference frequency generation could be used in conjunction with on-chip mode-locked lasers to generate mid-infrared light for spectroscopic applications.

Regardless of the details of the particular physical configuration employed, all of these are exciting developments realized in the last several years. The variety of these implementations can only assist in the pursuit of new densely integrated complex systems that can be achieved through the heterogeneous integration of different materials on a chip, which would not have been possible without the introduction of the second-order nonlinear process onto silicon.

Acknowledgments

The authors thank DARPA DODOS program and the Office of Naval Research (ONR) Young Investigator Program (YIP). 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 Department of Defense or the U.S. Government.

Conflict of Interest

The authors declare no conflict of interest.


