

# High-Speed Silicon Photonics Modulators

*This paper reviews optical modulators in silicon, which has become a very rich topic, including Mach–Zehnder modulators and ring modulators. The tradeoffs in design to achieve high speed with minimal drive voltage and insertion loss are studied.*

JEREMY WITZENS<sup>1b</sup>, Senior Member IEEE

**ABSTRACT** | The realization of gigahertz bandwidth modulators out of silicon-based technology in the early 2000s marked a cornerstone of silicon photonics development. While modulation speeds have since progressed well above 50 GHz and satisfy the bandwidth requirements of current and emerging modulation formats, concurrently obtaining low drive voltages and low insertion losses remains a very active area of research. While modulators generally come in two categories, direct absorption and those relying on embedded phase shifters, the focus of this paper lies on the latter capable of supporting both complex-valued modulation and optically broadband operation. The paper provides an overview of the current state of the art, as well as of currently explored improvement paths. First, common phase shifter configurations, aspects related to electrical driving, and associated power consumption are reviewed. Slow-wave, resonant, and plasmonic enhancements are further discussed. The reader is familiarized with the optimization of these devices and provided with a sense of the limitations of current technology and the potential of novel hybrid material integration.

**KEYWORDS** | Electro–optic modulators; integrated photonics; optical communications; resonantly enhanced modulators; silicon photonics; slow-wave modulators

## I. INTRODUCTION

Since the first demonstration of a silicon photonics (SiP) modulator with gigahertz (GHz) modulation frequencies in 2004 [1], substantial efforts have been made to improve modulation efficiency, bandwidth, and insertion losses. Notably, electro–optic (E/O) cutoff frequencies in excess of 50 GHz have been reached with the more common

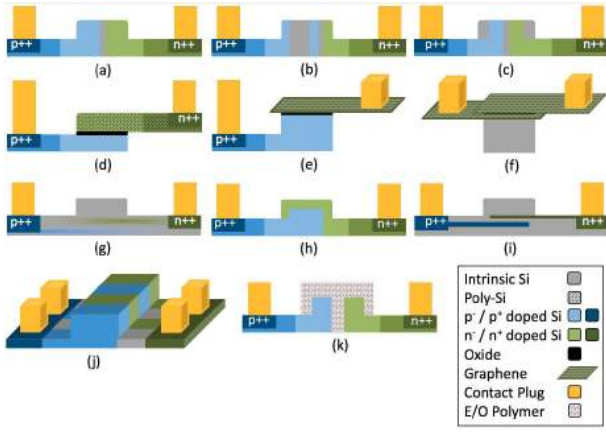
carrier depletion modulators [2], and even better performance with speeds in excess of 100 GHz [3] as well as phase shifting efficiencies  $V_{\pi}l$  better than  $0.5 \text{ V} \cdot \text{mm}$  [4] have been demonstrated with semiconductor organic hybrid (SOH) polymer integration. Nonetheless, modulators remain a key area of investigation in SiP since they still have a few weak points, improvement of which would go a long way in further increasing the attractiveness of the technology: Relatively high interface coupling losses associated to hybrid light source integration [5], [6] are compounded by the free-carrier absorption losses [7] occurring in carrier depletion/accumulation modulators, straining optical power budgets as short reach communications standards move to signaling rates of 50 Gbaud, four-level pulse amplitude modulation (PAM-4) and beyond [8]. Transceiver developers have thus to weigh the high yield and facilitated electronic integration afforded by mature complementary metal–oxide–semiconductor (CMOS) processes, the high stability and material quality of silicon (Si), and relatively easy to handle capacitive or resistive loads associated to SiP modulators against the reduced insertion losses afforded by monolithic integration of externally modulated lasers (EMLs) [9]. As SiP moves to being applied to high-performance, long-haul telecommunications [10], SiP modulators also have to be benchmarked against the characteristics of much larger but also typically lower loss lithium niobate modulators [11]. Quickly increasing application-driven requirements as well as recent progress in SiP modulator architectures and hybrid material phase shifters sustain a fast moving research field.

In the following, we will review the state of the art of SiP high-speed phase shifter technologies, aspects related to electronic driver integration and power consumption, as well as enhanced modulator architectures such as slow-wave, resonance, and plasmonic enhanced phase shifters. The primary focus lies with near-infrared (NIR) 1310- and

Manuscript received April 15, 2018; revised September 29, 2018; accepted October 16, 2018. Date of current version November 20, 2018.

The author is with the Institute of Integrated Photonics, RWTH Aachen University, Aachen 52074, Germany (e-mail: jwitzens@iph.rwth-aachen.de).

Digital Object Identifier 10.1109/JPROC.2018.2877636



**Fig. 1.** Diagrams showing phase shifter configurations discussed in the main text: (a) baseline PIN diode; (b) PIPIN diode; (c) PIN phase shifter with counter doping in the waveguide corners; (d) silicon-insulator-silicon capacitor with top polysilicon; (e) silicon-insulator-graphene capacitor; (f) graphene-insulator-graphene capacitor; (g) vertical junction phase shifter fabricated based on implantation with vertical dopant profile determined by implantation energies; (h) all-around PIN junction fabricated with angled implants; (i) vertical junction fabricated with epitaxy; (j) interleaved PIN junctions; and (k) slot waveguide filled with a poled polymer.

1550-nm O- and C-band modulators relying on embedded phase shifters, which enable both amplitude and phase modulation and remain operational over, or are adaptable to, a very wide range of wavelengths [12]. Direct absorption modulators relying, e.g., on the Franz-Keldysh effect [13] or the quantum confined Stark effect (QCSE) in (silicon-)germanium [14], as well as extension of SiP modulators to other wavelength ranges such as the mid-infrared (MIR) [15]–[17] or to the visible silicon nitride (SiN) platform [18], [19] are not the main focus of this paper.

## II. PHASE SHIFTER TECHNOLOGIES

### A. All-Silicon Phase Shifters and Their Performance Metrics

The most commonly used actuation mechanisms in high-speed SiP phase shifters are carrier depletion in reverse biased PIN diodes or carrier accumulation in silicon-insulator-silicon capacitors (SISCAPs); see Fig. 1(a) and (d). In both, modification of the free-carrier concentrations inside the waveguide results in a modification of the material's refractive index via the free-carrier plasma effect [7]. In the PIN phase shifter, the space charge region is enlarged as the reverse bias applied across the diode is increased, thus sweeping out carriers from the waveguide. In the SISCAP phase shifter, typically operated in accumulation mode, carriers of opposing polarity accumulate on either side and in the immediate vicinity of the insulator region, increasing the number of free carriers inside the waveguide as the bias voltage is increased.

Important metrics for these phase shifters are thus the amount of carriers that can be moved in and out of the waveguide by a given drive voltage change, i.e., the capacitance per unit length (also called the linear capacitance  $C_l$ ), as well as the overlap  $\Gamma$  of the accumulation or depletion regions with the optical mode. Unless the mode profile can be shrunk, a higher linear capacitance is thus typically required to lower  $V_\pi l$ , i.e., the voltage  $V_\pi$  required to obtain a  $\pi$  phase shift given a unit phase shifter length. The product of  $V_\pi$  with the phase shifter length  $l$  is a characteristic that remains approximately unchanged as the length of the phase shifter is rescaled, with deviations from this exchange relation associated to phase shifter nonlinearities.  $V_\pi l$  is thus an essential performance metric and one way to quantify phase shifter efficiency of particular relevance when drive voltages are limited by the driver's integrated circuit (IC) technology.

However, the power dissipated to actuate the device at high speed also needs to be considered. At a fixed drive voltage, a higher total capacitance  $C$  (assumed here to be voltage independent, as a simplification) also means a higher power consumption, as the minimum energy per bit to actuate the phase shifter is [20]

$$E_{\text{bit}} = \frac{1}{4} C V_{DD}^2 \quad (1)$$

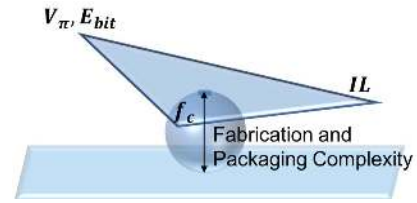
wherein  $V_{DD}$  is the supply voltage of the driver output stage applying a voltage across the phase shifter. Note that (1) only corresponds to the power consumption associated to charges actually switched across the phase shifter and does not include the power dissipated in other device elements, such as a resistive termination in a traveling wave (TW) device. Moreover, it does not include internal power consumption inside the driver in earlier signal processing stages. Thus, it is only to be considered as the fundamental limit; a more holistic discussion of modulator power consumption is the object of the next section. However, since other sources of power dissipation, both inside the modulator and in the later driver stages, tend to scale together with (1), it is indicative of power consumption scaling. Generally, increasing the phase shifter capacitance tends to reduce the power consumption, as it also reduces the required drive voltage and thus allows lowering the supply voltage:  $E_{\text{bit}}$ , as predicted by (1), scales with the square of  $V_{DD}$ , but only increases linearly with  $C$ .

Unfortunately, the capacitance cannot be indefinitely scaled up in PIN-based devices without other metrics incurring severe penalties: Increasing the capacitance requires reducing the width of the space charge region (over which the electric field is dropped) and thus increasing the dopant concentrations. Since this also results in increased waveguide losses mediated by free-carrier absorption, one eventually runs into a wall: Due to the sublinear dependence of the diode capacitance on the dopant concentrations and the supra-linear dependence of the waveguide losses (as predicted by the Drude model due to reduced

carrier mobilities with increasing concentrations [7]), the waveguide losses rise faster than the phase shifter efficiency unless other changes are applied to the phase shifter configuration [21]. This brings us to a tradeoff between phase shifter efficiency and the second cornerstone of phase shifter performance, the insertion losses (in decibels)  $IL = \alpha_{dB} \cdot l$ , wherein  $\alpha_{dB}$  are the optical losses per phase shifter length typically expressed in decibels per centimeter (the subscript differentiates  $\alpha_{dB} = 10\alpha/\ln(10)$  from the optical power decay rate  $\alpha$  expressed as an inverse  $1/e$  decay length). The tradeoff between  $V_\pi$  and insertion loss (IL) is typically given by the product  $V_\pi l \cdot \alpha_{dB}$ . With a unit of V·dB, it corresponds to the required drive voltage for a phase shifter sized for obtaining 1 dB of optical losses.

The SISCAP configuration operated in accumulation regime, in which high carrier concentrations of both polarities can accumulate on either side of the isolator irrespectively of the doping concentrations, offers a way to very significantly increase the phase shifter capacitance without excessively increasing free-carrier-induced losses other than those directly associated to the dynamically switched charges actively contributing to phase modulation. For this reason, some of the most efficient all-Si phase shifters are SISCAP-based devices [22]–[25]. However, here too, a minimum amount of doping is required once the third performance metric is added to the equation (Fig. 2): Modulation bandwidth as characterized by the E/O cutoff frequency  $f_c$ , the frequency at which the E/O  $S_{21}$  of the device drops by 3 dBe. Since the actuation speed of phase shifters is limited by their RC time constant, sufficient doping is required to maintain a low enough series resistance. The very high efficiency of SISCAP devices stems from their high capacitance when operated in the accumulation regime. However, this same capacitance limits their RC limited bandwidth (efficiency and speed can actually be traded off against each other in SISCAPs by choosing the bias point). As a drawback, SISCAP phase shifters typically rely on polysilicon for the top Si layer. Deposited on top of oxide, it cannot be straightforwardly epitaxially grown and results in both higher waveguide losses and reduced conductivity. Resulting requirements for higher doping further burden the insertion losses. While polysilicon can, in principle, be recrystallized, with seeding in a farther region in which the polysilicon is put in direct contact with the underlying single-crystal Si after local removal of the interposed oxide [26], this constitutes a further deviation from standard CMOS processing flows. It has, however, been successfully implemented to yield improved devices with a demonstrated  $V_\pi l \cdot \alpha_{dB}$  below 7 V·dB [27], an outstanding performance for an all-Si phase shifter.

The PIN depletion-type phase shifter remains widely popular however, due to its ease of fabrication as well as its straightforward compatibility with high-speed operation in the several tens of gigahertz. A number of techniques have been devised to improve its characteristics, in particular its  $V_\pi l \cdot \alpha_{dB}$ , by a judicious shaping of the dopant regions



**Fig. 2.** Designing an application-specific modulator is typically a balancing act between three metrics traded off against each other within the limitations of a given fabrication flow: 1) phase shifting efficiency ( $V_\pi$ ) or energy per bit ( $E_{bit}$ ); 2) insertion losses (IL); and 3) E/O cutoff frequency ( $f_c$ ).

inside the waveguide. Dopants are placed where they are the most needed, i.e., in the regions that get dynamically depleted in which they actively contribute to index modulation, and restricted elsewhere to the minimum required to ensure a sufficiently low series resistance for the targeted RC limited bandwidth.

One such approach is the PIPIN structure [28], in which an additional P-doped region is placed in the middle of a comparatively wide intrinsic region, leaving most of the waveguide core otherwise intrinsic [Fig. 1(b)]. This P-type center region is ideally dimensioned so as to be fully depleted at the highest utilized reverse voltage. Another approach, counter doping of the waveguide corners, removes free carriers where they contribute neither to refractive index modulation nor to electrical connectivity [29], [30] [Fig. 1(c)].

Other approaches have focused on increasing the volume in which the free-carrier concentration is being modulated by increasing the junction area. One such approach consists in implementing vertical rather than horizontal P(I)N junctions. Such can be realized by overlaying implantation regions with suitable implantation conditions to result in a targeted vertical profile [31] [Fig. 1(g)]. Since the rib waveguides used in SiP phase shifters are typically wider than high, this results in an increased junction size and a higher capacitance. This technique was further refined with angled implants to realize an all-around junction wrapping around the waveguide [32], [33] [Fig. 1(h)]. Vertical junctions proved particularly useful in tiny disk resonators [34], in which the light fills a comparatively wide portion of the device cross section as a consequence of the whispering gallery mode reaching toward the inner regions of the disk. Thus, horizontal junctions would only overlap with an even further reduced portion of the mode area. Since fabrication of vertical junctions by implantation is facilitated by flat surfaces, vertical junction phase shifters have been developed based on implantation of unetched Si films followed by deposition of an amorphous Si stripe overlaying the junction to create a strip loaded waveguide [35].

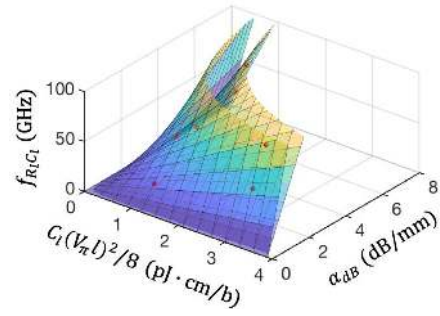
While implantation-based techniques do allow some control over dopant distribution [29], [30], fine placement of dopants in the vertical direction remains difficult.

In order to increase control over dopants, it has been proposed to implement phase shifters with help of successive epitaxial steps, with either *in situ* doping during growth [36] or alternating implantation and epitaxial overgrowth steps [21] [Fig. 1(i)]. While technological complexity for the fabrication of such devices is also considerably increased, particularly if they are to be combined with low loss interconnect waveguides, unprecedented level of control over dopants may yield high-performance Si-only phase shifters with high cutoff frequencies in the 30–50-GHz range.

Interleaved PIN junctions [Fig. 1(j)] present a further way to increase the junction capacitance per unit waveguide length that is relatively straightforward to fabricate [37]–[39], while maintaining moderate dopant concentrations. Thus  $V_\pi l \alpha_{dB}$  can scale favorably [37]. However, preventing parasitic capacitive loading arising from extension of the crosswise junctions outside the waveguide, where they do not contribute to phase shifting but result in increased power consumption and reduced bandwidth, presents technological and design challenges.

Finally, it should be mentioned that performance is not the only driving force behind novel phase shifter development. In particular, compatibility with commercial CMOS processes is an important characteristic for cost-effective deployment of the technology [40]. In recent years, first SiP modulators have been demonstrated in zero-change digital electronics CMOS processes [41], [42], a step toward monolithic integration of SiP interconnects into predesigned processor cores [43].

Instead of the metrics  $V_\pi l$  (in V·cm) and  $V_\pi l \alpha_{dB}$  (in dB·V) discussed above, which are particularly useful in a context in which the drive voltage is limited by the chosen IC technology, one can also use the corresponding metrics  $C_l(V_\pi l)^2/8$  (in cm·pJ/b) and  $C_l(V_\pi l)^2 \alpha_{dB}/8$  (in dB·pJ/b). These relate to  $E_{bit}$ , the minimum energy per bit as defined above [see (1)] rather than drive voltage, again for a phase shifter with either a unit length or insertion loss. These were derived for a Mach–Zehnder modulator (MZM) operated for non-return-to-zero (NRZ) ON–OFF keying (OOK) in dual-drive (DD) push–pull configuration [21] but are also applicable to other configurations (modulo a proportionality constant). In either case, these metrics are incomplete if one does not also take the RC constant limited cutoff frequency into account, since diverse methods used to improve the phase shifter efficiency also impact the bandwidth. In the following,  $f_{R_l C_l}$  denotes the upper limit of the phase shifter cutoff frequency as limited by its linear resistance  $R_l$  and capacitance  $C_l$  (its intrinsic RC-limited cutoff frequency). Strikingly, in a survey of Si-only phase shifters done in 2015 taking  $\mathcal{M} = f_{R_l C_l} \cdot (C_l \cdot (V_\pi l)^2 \alpha_{dB}/8)^{-1}$  as a figure of merit (FOM)—with a unit of GHz/(dB·pJ/b) balancing out bandwidth, insertion losses, and power consumption (Fig. 2)—it was found that  $\mathcal{M}$  was between 1.5 and 2.7 GHz/(dB·pJ/b) for most well-designed phase shifters (Fig. 3). More



**Fig. 3.** Graphical representation of the experimentally recorded characteristics of the all-silicon phase shifters surveyed in [21]. The two surfaces correspond to  $\mathcal{M} = 1.3$  and  $\mathcal{M} = 2.35 \text{ GHz}/(\text{dB} \cdot \text{pJ/b})$ . Somewhat surprisingly, given the wide range of phase shifter topologies and individual phase shifter characteristics, the points remain well clustered between the two surfaces.

recent, improved Si-only phase shifters do not reach significantly beyond this limit: While the device reported in [27] has an outstanding  $V_\pi l \alpha_{dB}$  and its estimated power consumption metric  $C_l(V_\pi l)^2 \alpha_{dB}/8 = 4.3 \text{ dB} \cdot \text{pJ/b}$  also beats that of all the previously surveyed devices, its cutoff frequency remains modest as 25-Gb/s eye diagrams already show some vertical eye opening reduction. While the analog cutoff frequency is not reported, one infers that  $\mathcal{M}$  stays on the order of  $3 \text{ GHz}/(\text{dB} \cdot \text{pJ/b})$ . Given the wide variety of Si-only-based devices, this tight range is remarkable and raises the question how to achieve the higher performance required by emerging standards.

Irrespectively of design details, there is a fundamental lower limit to the insertion losses associated to achieving a given phase shift, since at least the carriers required to induce the necessary refractive index change have to be present in one of the phase shifter states. The minimum losses (in decibels) associated to reaching a  $\pi$  phase shift are given by  $IL_\pi = (10/\ln(10)) \cdot 2\pi \Delta k / \Delta n$ , with  $\Delta n$  and  $\Delta k$  the free-carrier-induced changes of the real and imaginary parts of the refractive index. Based on the free-carrier refraction and absorption data reported at a 1550-nm wavelength in [7], this minimum loss can be calculated to range from 0.75 to 3 dB as average electron and hole concentrations in the modulated region are ramped up from  $10^{17} \text{ cm}^{-3}$  to  $10^{19} \text{ cm}^{-3}$ , wherein the dependency on carrier concentration stems from the supra-linear increase of the free-carrier-induced losses. Since higher concentrations also improve the efficiency of the phase shifters, simply reducing the dopant concentrations is not always an option to reduce  $IL_\pi$ . Note that these fixed losses are in apparent contradiction with the scaling law associated to the metric  $V_\pi l \alpha_{dB}$ , which implies losses can be traded off against drive voltage: The latter is only valid when the dynamic losses associated to modulation of the space charge region (PIN phase shifter) or carrier accumulation (SISCAP) are low compared to the fixed losses induced



from the overall waveguide doping needed for electrical connectivity (which is typically the case in high-speed phase shifters).

## B. Heterogeneous Material Integration

To improve phase shifters beyond the limits described above, one may consider introducing other materials, even though this requires further deviations from standard CMOS processes, creating further challenges for cost-effective manufacturing. One approach is to use other semiconductor materials for part of the phase shifter with mobilities and free-carrier effective masses more favorable to efficient low-loss modulation. From the Drude model, free-carrier-induced refraction and absorption can be modeled as [7]

$$\Delta n = -\frac{q^2 \lambda_0^2}{8\pi^2 c_0^3 \epsilon_0 n} \left( \frac{\Delta N}{m_{ce}^*} + \frac{\Delta P}{m_{ch}^*} \right) \quad (2)$$

$$\Delta \alpha = \frac{q^3 \lambda_0^2}{4\pi^2 c_0^3 \epsilon_0 n} \left( \frac{\Delta N}{m_{ce}^* \mu_e} + \frac{\Delta P}{m_{ch}^* \mu_h} \right) \quad (3)$$

with  $q$  the elementary charge,  $c_0$  the speed of light in vacuum,  $\lambda_0$  the free-space wavelength,  $\epsilon_0$  the free-space permittivity,  $n$  the refractive index,  $\Delta N$  and  $\Delta P$  the electron and hole concentrations,  $m_{ce}^*$  and  $m_{ch}^*$  the electron and hole conductivity effective masses, and  $\mu_e$  and  $\mu_h$  the electron and hole mobilities. While empirical corrective factors are typically applied to these formulas [44], they do exemplify the dependencies on free-carrier effective masses and mobilities, particularly in the NIR in which other mechanisms such as intervalence band absorption (IVBA) do not play a predominant role. It can thus be seen that low effective masses boost both  $\Delta n$  and  $\Delta \alpha$ , not changing the fundamental limit to  $lL_\pi$  or improving  $V_\pi l \alpha_{dB}$  assuming the carrier scattering time  $\tau$  remains unchanged ( $\mu = q\tau/m_c^*$ ), but potentially allowing shortening a more efficient device. On the other hand, further increase of the mobility straightforwardly allows improving insertion losses.

Within the CMOS processing compatible group IV material system, SiGe has, for example, been utilized as an alternative material for the P-side of the junction [45], [46]. Compressive strain applied to heteroepitaxially grown SiGe sandwiched between Si layers, as a consequence of the lattice mismatch, results in reduction of the hole effective mass, in turn enhancing the free-carrier-induced electrorefraction and electroabsorption. Unfortunately, here  $\Delta \alpha$  is similarly enhanced to  $\Delta n$ , so that this effect is not suitable to reduce  $V_\pi l \alpha_{dB}$ . On the other hand, the phase shifting efficiency could be almost doubled, which naturally leads to a reduction of power consumption since the phase shifter length can be halved and the capacitance thus significantly reduced, all other things remaining equal.

On the other hand, in order to improve  $V_\pi l \alpha_{dB}$  by means of improved semiconductor material properties, one has to improve mobility beyond what can be accomplished by

simply modifying the effective mass [assuming (2) and (3) to be exact, a simplification, this would boil down to a requirement for an improved carrier scattering time]. Since Si-only phase shifters are burdened particularly by the N-side of the junction, as the relevant material FOM in view of obtaining a low  $V_\pi l \alpha_{dB}$ ,  $\Delta n / \Delta k$ , is about 3.5x worse for electrons than for holes at a typical  $10^{18} \text{ cm}^{-3}$  carrier concentration [47], it appears intuitive to replace the N-side of the device with another suitable material such as InGaAsP (with a suitable composition to yield a bandgap above C-band photon energies), for which  $\Delta n / \Delta k$  is substantially improved for electrons. In [47] and [48], this approach was followed in a SISCAP-type structure, replacing the top, N-doped polysilicon part by N-doped InGaAsP heterogeneously integrated with direct wafer bonding. Much higher performance phase shifters were fabricated with a 3–5x smaller  $V_\pi l$  and an 8x lower  $V_\pi l \alpha_{dB}$  compared to polysilicon-based SISCAP devices [22], [23]. While demonstrated bandwidths are so far limited to 2 GHz, with excessive series resistances arising from processing issues identified as a limitation in [47] and [48], in every other aspect these devices provide a sizeable improvement relative to Si-only phase shifters, with comparable metrics [49] even to the SOH modulators [3], [4] discussed below.

Two-dimensional semiconductors have further gained substantial attention in recent years as materials for SiP modulators [50], [51]. After an initial focus on graphene-based direct absorption modulators [52], [53], graphene-based phase shifters have also been demonstrated [54]. Both rely on the Pauli-blocked Burstein–Moss shift: Shifting the Fermi level of graphene 0.4 eV by means of a voltage applied across a silicon–insulator–graphene stack [Fig. 1(e)] also shifts its absorption edge, from 0 eV (ideal graphene does not have a bandgap) to the 0.8 eV required to make it transparent at 1550 nm. As the Fermi level rises, the conduction band fills, consequently also increasing the minimum energy of photons that can be absorbed, since electrons have to transition from the valence band to free states in the conduction band. The absorption edge is at twice the Fermi-energy shift, since, as a consequence of electron impulse conservation, the energy of the valence electrons that can absorb light is also reduced by a corresponding amount. If the absorption edge is moved beyond 1550 nm, graphene becomes transparent in the C-band, but further shifts of the absorption edge still induce a refractive index change as predicted by the Kramers–Kronig relations. This is uniquely enabled by the reduced dimensionality of 2-D materials: Similarly filling the conduction band of a conventional bulk semiconductor would require a very large number of carriers associated to a large capacitance. Moreover, screening effects in a conventional SISCAP structure limit the carrier accumulation region to the material in the immediate vicinity of the isolator. Finally, the high mobility of graphene has the potential to maintain free-carrier losses at very low levels. On the other hand, since this phase shifter relies on the absorption edge being close to the modulated wavelength,

the optical passband of the device is restricted similarly to QCSE-based devices.

The phase shifter reported in [54] remains burdened by high optical losses ( $\sim 236$  dB/cm) and a relatively low cutoff frequency (5 GHz). However, promising improvement paths have been identified such as sandwiching the graphene film between hexagonal boron–nitride (hBN) layers to actually obtain the high mobilities typically associated to graphene or replacing the silicon–insulator–graphene stack by a graphene–insulator–graphene bilayer overlaid over an undoped Si waveguide [Fig. 1(f)], thus fully removing losses associated to doping of the silicon. Thus, these results might be further improved upon to yield truly competitive phase shifters. Extension of these configurations to other 2-D semiconductors such as transition metal dichalcogenides (TMDs), which have bandgaps in the visible, might also enable E/O switching functionality at visible wavelengths in SiN photonic integrated circuit (PIC) platforms [55]. WS<sub>2</sub> monolayers, for example, have a photoluminescence energy of  $\sim 1.95$  eV [56] corresponding to an emission wavelength of 636 nm, just slightly blue shifted from 640 nm, a wavelength commonly used for exciting fluorophores used in stimulated emission depletion (STED) microscopy. WSe<sub>2</sub> might also enable the implementation of modulators for  $\sim 850$ -nm wavelengths compatible with vertical cavity surface emitting laser (VCSEL) technology [57].

Another means to improve the efficiency of SiP modulators has been to combine Si with spun-on organic materials with a high second order nonlinearity. Silicon is a centrosymmetric material that does not possess, by itself, a second-order nonlinearity directly converting an applied radio-frequency (RF) field into a refractive index change (Pockels effect). While surface effects as well as the application of inhomogeneous strain can be used to break the centrosymmetry of Si [58]–[60], it has been later shown that free-carrier effects can play an important role in the measured apparent nonlinearities [61]–[63], so that the applicability of this effect to yield efficient high-speed modulators remains in question. Nonlinear polymers on the other hand can yield very high and ultrafast nonlinearities, with the in-device Pockels coefficient  $r_{33}$  as high as 390 pm/V exceeding more than 10x the nonlinearity of lithium niobate [64]. The high magnitude of the Pockels effect is further enhanced by infiltrating slot waveguides [65] [Fig. 1(k)] in the region of highest optical field with the polymer, maximizing the overlap between the region in which the refractive index is modulated and the optical mode. The two halves of the waveguide are further doped to enable electrical connectivity. They are electrically isolated from each other by the narrow slot across which the applied voltage is dropped, thus forming a capacitor and generating an RF electric field whose strength is inversely proportional to the slot width. After spinning-on the nonlinear polymer, individual chromophores have to be aligned in what is called the poling process. Since early results suffering from highly

reduced in-device poling due to surface anchoring effects and moderate bandwidths of a few gigahertz [66], these devices have evolved to being among the most efficient and highest speed realized in a Si platform. However, while important progress has been made in the reliability of polymer-based modulators [67], few reliability studies are yet available specifically in regards to the SOH configuration. More recently, substantial efforts have also been made toward the integration of ferroelectric perovskites such as LiNbO<sub>3</sub> [68], [69], BaTiO<sub>3</sub> [70] or lead zirconate titanate (PZT) [19] into SiP platforms as an alternate way to add a Pockels effect to the material system.

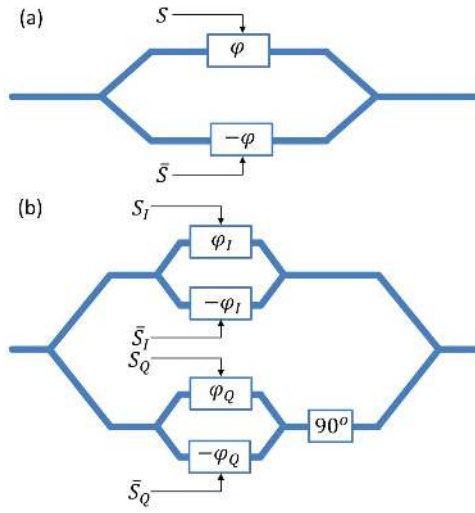
To a large extent, the high attractiveness of slot waveguide modulators stems from the high optical field enhancement inside the slot, arising as a consequence of its reduced refractive index compared to that of the rest of the Si waveguide core (see Section IV-A for a more detailed analysis). One may envision applying the same type of optical field enhancement to a semiconductor-only-based phase shifter by sandwiching a lower refractive index semiconductor between two Si layers [with the field enhancement applied to the transverse magnetic (TM) polarization if heteroepitaxy results in a vertical stack]. Silicon carbide (SiC) is, for example, such a semiconductor, that in its zinc blende ( $\beta$ -)polytype has a refractive index of  $\sim 2.55$  that is considerably smaller than that of Si ( $\sim 3.48$ ) as a consequence of its much increased optical bandgap. While it is notoriously difficult to grow with a high quality on Si, progress continues to be made in this direction [71].

### III. MACH-ZEHNDER MODULATORS, DRIVING SCHEMES, AND RELATED POWER CONSUMPTION

Here we review modulator architectures with straight phase shifters, corresponding driving schemes, and associated power consumption. Advanced architectures that are not in a rectilinear configuration or rely on advanced optical design techniques such as slow-wave propagation are covered in Section IV.

#### A. Mach–Zehnder Modulator Architectures

Rectilinear (nonresonant) modulators typically consist in either the baseline MZM [Fig. 4(a)], or nested MZMs allowing independent modulation of the in-phase (I) and quadrature (Q) components of the optical signal [Fig. 4(b)]. While the baseline MZM is suitable for amplitude-shift keying (ASK) such as OOK or PAM-4, as well as two-level phase-shift keying (PSK), the nested modulators are used for modulation of both the real and complex parts of the electrical field, as required by, e.g., quadrature phase-shift keying (QPSK), quadrature amplitude modulation (QAM), or coherent orthogonal frequency-division multiplexing (OFDM). Complex-valued modulation can also be used for single sideband (SSB) modulation in direct detection systems featuring improved



**Fig. 4.** (a) MZM operated in push-pull configuration with two complementary dual-drive (DD) signals. Opposite phases are applied to the two phase shifters. (b) Nested MZMs allowing independent modulation of the in-phase (I) and in-quadrature (Q) electrical field components. Each of the MZMs is operated in DD configuration.

robustness against dispersion, such as used, e.g., in SSB discrete multitone (DMT).

Assuming phase shifters not to induce dynamic losses, the transfer function of the baseline MZM is given by

$$E_{\text{Out}} = E_{\text{In}} \cdot e^{i \frac{\varphi_2 + \varphi_1}{2}} \cdot \cos\left(\frac{\varphi_2 - \varphi_1}{2}\right) \quad (4)$$

$$P_{\text{Out}} = P_{\text{In}} \cdot \frac{1 + \cos(\varphi_2 - \varphi_1)}{2} \quad (5)$$

wherein  $E_{\text{In}}$  and  $E_{\text{Out}}$  are the E-field amplitudes and  $P_{\text{In}}$  and  $P_{\text{Out}}$  are the optical power levels at the input and output of the modulator.  $\varphi_1$  and  $\varphi_2$  are the phase shifts induced in its lower and upper branches. In order to minimize chirp and to reduce the voltage range that needs to be sourced by the drivers, MZMs are typically operated in push-pull configuration with a DD signal resulting in  $\varphi_2 = -\varphi_1 = \varphi$ , i.e., the two phase shifters are actuated with opposite signals according to the data stream. This also reduces the power consumption, as driving two phase shifters with approximately half the signal swing results in halving the dissipated power.

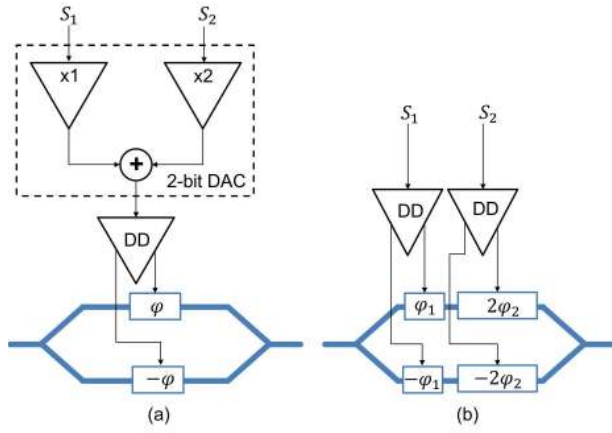
In ASK, the modulator is typically biased around an average phase difference of  $\langle \varphi_2(t) - \varphi_1(t) \rangle = 2\langle \varphi(t) \rangle = \pi/2 \bmod \pi$ , the quadrature condition. This maximizes the slope of the power transfer function (5) and thus the small signal transfer function of the modulator. Maximal optical signal strength is obtained with a dynamic swing of  $\pm\pi/2$  around the bias point, so that each phase shifter is only required to generate up to  $\pi/2$  phase shift. For two-level PSK, the modulator is typically biased around a phase  $2\langle \varphi(t) \rangle = \pi \bmod 2\pi$  (zero-transmission point). The maximal signal strength is then obtained with a dynamic swing of  $\pm\pi$ , so that a full  $V_\pi$  needs to be ideally applied to both arms.

Even though the MZM is a nominally balanced structure, optically broadband and passively biased around a predetermined average phase, mismatch between the arms resulting from fabrication tolerances typically leads to a phase error that needs to be compensated post-fabrication with an active control system: The average phase error is monitored, e.g., by replacing the output coupler by a directional waveguide coupler (DWC) and comparing the average power in the two output waveguides during modulation with a direct current (DC) balanced signal (quadrature is then reached when the average power levels at both outputs are equal [72]). A phase shift is applied accordingly until the phase error is nulled.

Since the required phase correction is also sensitive to temperature gradients applied across the MZM, which may occur if the MZM is copackaged with a laser or modulator driver dissipating a large amount of power in its close vicinity, as well as to absolute temperature if the interferometer is sufficiently imbalanced due to fabrication tolerances, the phase correction is typically dynamically adjusted with dedicated, efficient, but slow phase shifters. These are typically either implemented in the form of thermal phase shifters exploiting the thermo-optic coefficient of Si [73] or in the form of forward biased PIN junctions [74], whose large diffusion capacitance makes them very efficient, but also slow. Thermal phase shifters result in significant additional power consumption unless substantial efforts are made to improve the thermal isolation of the device from the rest of the chip [75], [76]. Forward biased PIN junctions, on the other hand, incur the unavoidable losses associated to free-carrier absorption, although they can be designed with the entire waveguide core contained inside the intrinsic region, so that free-carrier-induced losses can be maintained close to the minimum  $\text{IL}_\pi$ . A number of studies have been made to passively bias MZMs with a remanent correction [77]–[81], so that a dynamic control system and the associated electronics can be dispensed with. The practicability of such schemes remains, however, to be proven in industrial practice, contingent on MZMs being sufficiently well balanced to be insensitive to overall environmental temperature swings, on temperature gradients inside packaged devices to be sufficiently well controlled to not induce differential phase shifts, and on the long term stability of the correction.

The nested IQ modulator combines two baseline MZMs embedded within an overarching PIC. After splitting the light at the first splitter, both the  $I$ - and  $Q$ -MZMs generate independent, real-valued fields. A static,  $90^\circ$  phase shift at the output of the  $Q$ -MZM rotates its output signal to generate the imaginary field component. An output coupler sums the  $I$ - and  $Q$ -signals into an overall complex-valued signal [82]

$$E_{\text{Out}} = \frac{E_{\text{In}}}{2} \cdot [\cos(\varphi_I) + i \cdot \cos(\varphi_Q)] \quad (6)$$



**Fig. 5.** PAM-4 DD-MZM architectures with (a) a combination of the two binary signals in the electrical domain with a 2-b DAC; and (b) a combination of the two signals in the optical domain with properly sized phase shifters.

wherein each of the two MZMs is operated as an individual two-level PSK modulator. Generalization of (4)–(6) to include dynamic losses induced by phase shifters is important to take into account all link impairments (reduced extinction, distortion, chirp) in high-performance links [83].

Due to the nonlinearities of the modulator transfer functions (4)–(6), signals typically need to be pre-conditioned and fed through a high-speed digital-to-analog converter (DAC) in case of multilevel signaling, resulting in additional high-speed processing in the electrical domain [Fig. 5(a)]. The DAC conversion can, however, also be handled in the E/O domain by segmenting the phase shifters [84], [85]. This is illustrated in Fig. 5(b) for the example of PAM-4 modulation: The two multiplexed signals corresponding to the least significant bit (LSB) and most significant bit (MSB) of the PAM-4 signal are fed to individual phase shifter segments, each driven with the same voltage swing but sized according to the corresponding bit number. Segmented phase shifters have also been adapted to higher modulation formats such as QAM-16 [85], wherein the phase shifters of both the  $I$ - and  $Q$ -MZMs are segmented in order for each to receive two individual electrical signals. The odd nonlinearities introduced, respectively, by the relevant equations (5) and (6) can be addressed by slightly resizing the amplifiers in Fig. 5(a) or the phase shifter lengths in Fig. 5(b).

A further concern in the optical design of MZMs resides in obtaining optically broadband operation: Since the free-carrier dispersion effect is very broadband and MZMs are by nature balanced devices, this is in principle possible. Practical difficulties reside, e.g., in designing sufficiently broadband splitters and combiners, with advanced techniques relying, e.g., on subwavelength structures [86] or on adiabatic power splitters [87]. In the case of TW

devices, phase matching (see below) also needs to be maintained over the entire wavelength range of operation, which in turn bounds the allowable group velocity dispersion of the utilized waveguides.

## B. Driving of Lumped Element Modulators

We now continue with a review of how to drive modulators, the consequences on electrical power consumption, as well as the resulting tradeoffs between phase shift efficiency ( $V_\pi$ ), insertion losses (IL), bandwidth ( $f_c$ ), and overall power consumption and packaging complexity (Fig. 2).

The simplest phase shifter configuration is the lumped element (LE). This refers to a phase shifter that is sufficiently small compared to the RF wavelength of the electrical signal inside the chip to be considered a point load from the electrical perspective. The voltage delivered to an LE phase shifter is then approximated as being identical along its entire length. A typical criterion might be for its length to be below  $1/10^{\text{th}}$  of the free-space RF wavelength  $\lambda_{\text{RF}}$  divided by the RF effective index  $n_{\text{RF}}$ . However, the transition away from the LE approximation is gradual, with an increasing penalty applied to the modulator's cutoff frequency, so that a clear cutoff criterion is hard to define.

In this configuration, the cutoff frequency of the phase shifter is subjected to two limitations: 1) the speed at which the phase shifter can be actuated; and 2) the transit time of light through the device. The phase shifter actuation speed is typically limited primarily by its capacitance and series impedance, both resistive and inductive. The latter results from the intrinsic resistance of the phase shifter itself as well as parasitics from the connecting signal traces, including wire or bump bonds, as well as the output impedance of the driver or the characteristic impedance of the RF environment. The highest possible cutoff frequency of the phase shifter, its intrinsic limitation in absence of external peaking compensation or signal preemphasis, is simply determined by its linear capacitance  $C_l$  (pF/cm) and its linear resistance  $R_l$  ( $\Omega \cdot \text{cm}$ ). Since the total capacitance and resistance of the phase shifter are, respectively, given by  $C_l \cdot l$  and  $R_l/l$ , the intrinsic cutoff frequency of the device  $f_{R_l C_l}$  is equal to  $1/2\pi R_l C_l$  irrespectively of the length of the device. The latter does matter, however, once the extrinsic impedance  $Z_{dr}$  of the driving circuitry is taken into account. If  $Z_{dr}$  is a pure resistance, the cutoff frequency is simply given by  $1/2\pi(R_l C_l + l C_l Z_{dr})$ , from which the dependency on the phase shifter length is immediately visible. In order to realize the full benefit of LE driving, it is highly desirable to place the driver in the immediate vicinity of the phase shifter to free the electronic design from the constraint of maintaining a  $50 \Omega$  environment, and to design the output impedance of the driver to be as low as possible while maintaining a reasonable power consumption [88] (reducing the output impedance typically requires a larger transistor sizing in the output stage).



A series inductance  $L$  on the other hand, as, e.g., induced by wire bonds, is slightly more complex to analyze, as it can be both beneficial or detrimental depending on its value and on the other system parameters: Together with the phase shifter capacitance  $C = lC_i$ , it forms an LC tank with a resonance frequency  $2\pi f_{LC} = 1/\sqrt{LC}$ . If the resonance frequency is close to or slightly above the 3-dB cutoff frequency given by the capacitance and the resistive elements, it can extend the RC limited bandwidth of the phase shifter by means of a peaking response. On the other hand, if  $f_{LC}$  is substantially below  $f_{RC}$ , it will significantly distort the electrical signal delivered to the phase shifter, limiting the useful device bandwidth. A simple rule of thumb can thus be derived for the maximum allowable inductance given a targeted bandwidth and phase shifter capacitance. Since the total capacitance of a practical, nonenhanced rectilinear SiP phase shifter optimized for CMOS compatible drive voltages is on the order of 1 pF [21] and a typical pair of wire bonds results in an inductance of  $\sim 400$  pH (assuming a 300- $\mu$ m wire bond length and a 25- $\mu$ m wire diameter), one calculates a typical  $f_{LC}$  of  $\sim 8$  GHz. In other words, wire bonding is becoming increasingly challenging for nonenhanced straight phase shifters as data rates increase. Thus, more complex integration and assembly schemes, such as impedance matched coplanar wire bonds or flip-chipping with low inductance microbumps, become necessary at high speeds.

On the other hand, the transit time of free carriers through the semiconductor is typically much less of a limiting factor, since the regions in which the carrier concentrations are being actively modulated are typically limited to a few tens of nanometers. Given Si's saturation velocity, transit time does not play a role for modulators with cutoff frequencies below a few hundreds of gigahertz, i.e., for about any carrier accumulation/depletion-based device that has been implemented to date.

The second bandwidth limitation of LE modulators is rather associated to the time  $\tau_g$  taken by photons to transit through the length of the device [89]: We assume the same electrical signal to be applied throughout the entire length of the phase shifter and to be simply given by  $V = V_0 e^{i\omega_{RF}t} = V_0 e^{i\omega_{RF}(t_0 + z \cdot n_g / c_0)}$ , with  $V_0$  the amplitude of the electrical signal applied to the phase shifter,  $\omega_{RF}$  its angular frequency,  $t_0$  the time at which light enters the phase shifter,  $z$  the coordinate along the phase shifter length, and  $n_g$  the group index of light inside the phase shifter. The frequency-dependent phase shift is then

$$\frac{\Delta\varphi(\omega_{RF})}{\Delta\varphi(0)} = \frac{1}{l} \int_0^l e^{i\omega_{RF}\left(\frac{z \cdot n_g}{c_0}\right)} dz = \frac{1}{\tau_g} \int_0^{\tau_g} e^{i\omega_{RF}t} dt \quad (7)$$

wherein the second integral is obtained by a change of variable  $t = z \cdot n_g / c_0$ .  $\Delta\varphi(\omega_{RF})$  is complex valued, as the E/O S<sub>21</sub> of the modulator can also have a phase delay. This

then results in

$$\left| \frac{\Delta\varphi(\omega_{RF})}{\Delta\varphi(0)} \right|^2 = \sin^2\left(\frac{\omega_{RF}\tau_g}{2}\right) \quad (8)$$

which results in a photon transit time limited cutoff frequency  $f_\tau \simeq 0.44/\tau_g$ . In a simple, straight phase shifter without pronounced slow-wave effects, the photon transit time limitation does not add a significant constraint beyond the assumption of LE operation. The expression for  $f_\tau$  can be reexpressed as  $f_\tau \simeq 0.44 \cdot (\lambda_{RF}/l n_{RF}) \cdot (n_{RF}/n_g) \cdot f_{RF}$ . The assumption of LE operation already implies that  $\lambda_{RF}/l n_{RF} \gg 1$ . Moreover, in the absence of pronounced slow light effects, one may assume that  $n_{RF}/n_g$  is on the order of 1 or larger. (Refractive indices are typically larger at RF than at optical frequencies and the RF mode can be further slowed down by capacitive loading; see Section III-C. However, the waveguide mode is also better confined in a high index core.) Thus,  $f_\tau$  can be seen to be typically larger than the maximum frequency at which LE operation is maintained. This does not hold true, however, when the light is slowed down, e.g., by meandered, periodic, or resonant structures (Section IV). In that case, it may well be that  $n_g \gg n_{RF}$  and the bandwidth extracted from (8) may become a dominant limitation even for structures that are LE from an electrical perspective.

In order to discuss the power consumption of SiP modulators, we introduce the device-specific RF power consumption. Here, RF refers to the fact that this is the power consumption associated to the delivery of the high-speed signal to the phase shifters, excluding bias point stabilization by the control circuitry. Since the latter can also constitute an important part of the total power, it is important to complete the electrical power budget accordingly (particularly for the resonantly enhanced devices described in Section IV-D). Another important aspect of this metric is that it only captures the power consumption associated to currents flowing from the supply voltage  $V_{DD}$  into the device and from the device back to ground. Power consumption internal to the driver is not accounted for. However, the latter also depends on the modulator characteristics, as the output stages of the driver have to be sized according to the load: For example, driving a higher capacitance LE requires a higher current sourcing capability and thus a larger sizing of the output stage, also increasing internal driver power consumption. Similarly, reducing the driver output impedance to match a higher power TW device with a lower characteristic impedance also requires increasing the sizing of the driver. Generally, the driver sizing, and thus also its internal power consumption, can be assumed to scale with the delivered power. However, the scaling factor also depends on the driver architecture, which is constrained by the nature of the load: An output driver stage for an LE load can be implemented, for example, as a simple CMOS NOT

gate [90], while a TW device typically needs to be driven by an impedance matched output buffer [91], [92]. Finally, in today's data center architectures, in which transceivers are connected to a switch chip via lossy printed circuit board (PCB) transmission lines, drivers need to incorporate retiming and signal reshaping, so that, in view of the fixed power overhead, reducing modulator power intake below a certain threshold becomes of limited practical relevance unless it is accompanied by far reaching architectural changes requiring a tighter level of integration between transmitter and electrical switch chip [93].

As already described above, the average RF power consumption of an LE phase shifter is given by (1). This then provides the rationale for the FOMs  $C_l(V_{\pi}l)^2/8$  (in cm·pJ/b) and  $C_l(V_{\pi}l)^2\alpha_{\text{RF}}/8$  (in dB·pJ/b) trading off modulator power consumption with, respectively, phase shifter length and phase-shifter-induced insertion losses: In an MZM operated in push-pull, DD configuration, the maximum dynamic phase shift that needs to be applied by either phase shifter is  $\pi/2$ , corresponding to a voltage swing  $V_{\pi}l/2l$ . Since the capacitance of the two phase shifters taken together is  $2lC_l$ , the total RF power consumption is  $C_l(V_{\pi}l)^2/8l$ , with longer phase shifters resulting in lower power consumption due to the reduction in drive voltage overcompensating the increased capacitance.

For a phase shifter technology with a given  $V_{\pi}l\alpha_{\text{RF}}$  (V·dB) and  $C_l(V_{\pi}l)^2\alpha_{\text{RF}}/8$  (dB·pJ/b), the tradeoffs in the sizing of the phase shifters can now be discussed: Increasing  $l$  results in increased insertion losses, but a reduced drive voltage requirement for achieving full extinction as well as a reduced power consumption. The drive voltage is primarily constrained by the voltage swing that can be delivered by the chosen IC technology. The tradeoff between insertion losses and modulator power consumption is, however, more subtle.

Not only does the internal power consumption of the driver have to be taken into account, but also the assumption of full extinction can be significantly off in a real scenario. For one, given an available drive voltage, one can find a phase shifter length for which the optical signal contrast at the output of the modulator, the optical modulation amplitude (OMA), is maximized. For a modulator biased at the quadrature point and operated for any form of ASK, assuming both phase shifters to have a length  $l$ , the outer OMA is given by

$$\begin{aligned} \text{OMA [dB}_m] - P_{\text{in [dB}_m]} \\ = 10\log_{10} \left( \sin \left( \frac{V_{dr} \cdot \pi \cdot l}{V_{\pi}l} \right) \right) - \alpha_{\text{dB}}l. \end{aligned} \quad (9)$$

Excessively short modulators are penalized primarily by insufficient phase shifts, while excessively long modulators are penalized by high absorption losses. Thus, given the significant absorption losses typically occurring in SiP modulators, the optimum phase shifter length can be significantly below the length required to obtain a full  $\pi/2$  phase shift.

Moreover, maximizing the OMA is not necessarily the optimum as, at a fixed drive voltage, the modulator RF power consumption scales with the length of the phase shifter, while the OMA scales sublinearly due to the non-linearity of the sine function. An optimization trading off OMA with overall power consumption also needs to take the laser into account, since a reduction in OMA has to be compensated by a higher laser power in an optimized link. Due to the low wall plug efficiency of NIR semiconductor lasers, on the order of 15% for uncooled lasers, required increases in laser power come at a steep price. Moreover, less efficient utilization of the maximum available laser power from a single laser diode, typically below 100 mW for a high power distributed feedback laser (DFB) at module operation temperature, also leads to a reduction in channel count in parallel single mode (PSM) solutions and has thus important consequences on channel count scalability and module manufacturing cost. Thus, modulator dimensioning typically has to be cooptimized with driver as well as overall system design.

Last, in an LE design, the maximum length of the phase shifters is, of course, limited by the constraint of maintaining LE operation. Due to the limited efficiency of most SiP phase shifters, this is typically problematic unless the more efficient SISCAP or SOH slot waveguide phase shifters or enhancement techniques such as slow light or resonant enhancement are used. However, even with the less efficient phase shifter types, the scaling rules described above remain relevant as distributed drivers [94]–[96] can be utilized, in which case the device-specific power consumption remains the same. Moreover, even for well-designed TW devices, the overall modulator RF power consumption, including the power dissipated in the termination, can be assumed to be a multiple of the power derived above [21], as described in the following.

## C. Traveling Wave Modulators

In a TW modulator, the electrical signal is transported by a transmission line (TL) in which it propagates forward at the same speed than packets of light inside the waveguide. The phase velocity of the transmission line, as given by the RF effective index  $n_{\text{RF}}$ , has to be matched to the waveguide group index  $n_g$ . In order to show this, (7) can be generalized into

$$\frac{\Delta\varphi(\omega_{\text{RF}})}{\Delta\varphi(0)} = \frac{1}{V_0l} \int_0^l |V(z)| e^{i\omega_{\text{RF}}\left(\frac{z \cdot n_g}{c_0}\right) + i\varphi_{\text{RF}}(z)} dz \quad (10)$$

where, as before,  $z$  is the coordinate along the length of the waveguide and is curvilinear in the most general case as the waveguide can be wrapped (see meandered modulators in Section IV-B).  $|V(z)|$  and  $\varphi_{\text{RF}}(z)$  are the magnitude and phase of the RF field dropped across the phase shifter at coordinate  $z$  at a fixed time  $t_0$ . In the most general case, they can be extracted from a full 3-D RF device simulation. Assuming the phase shifter to be straight and to be connected to a transmission line with

wave number  $\beta_{\text{RF}} = \omega_{\text{RF}} n_{\text{RF}} / c_0$  and linear losses  $\alpha_{\text{RF}}$  ( $\alpha_{\text{RF}}/2$  in Neper per length describing the inverse characteristic decay length of the RF voltage), (10) simplifies into

$$\frac{\Delta\varphi(\omega_{\text{RF}})}{\Delta\varphi(0)} = \frac{1}{l} \int_0^l e^{i \frac{\omega_{\text{RF}}}{c_0} (n_g - n_{\text{RF}}) z - \frac{\alpha_{\text{RF}}}{2} z} dz. \quad (11)$$

In addition to the intrinsic RC limited bandwidth of the phase shifter, the bandwidth of such TW modulators is limited by both deviations from phase matching, i.e.,  $n_g \neq n_{\text{RF}}$ , as well as by the finite transmission line losses  $\alpha_{\text{RF}}$ . The bandwidth of a phase mismatch limited modulator (assuming no TL losses) is simply given by  $f_\tau \simeq 0.44/|\tau_g - \tau_{\text{RF}}|$ , where  $\tau_g$  again denotes the transit time of the light as given by its group index and  $\tau_{\text{RF}}$  is the transit time of the RF signal, as given by its phase velocity/RF effective index (the previously given formula for LE modulators is simply the limit in which  $\tau_{\text{RF}} \rightarrow 0$ ). The bandwidth of a TL-loss limited modulator (assuming perfect phase matching) is given by the frequency  $f_\alpha$  at which the RF loss coefficient reaches the value  $\alpha_{\text{RF}} = 1.48/l$  (at which the TL losses reach -6.4 dB). Hence, an accurate model of the TL line losses has first to be determined in order to model the modulator cutoff frequency. Importantly, while in RF engineering TL losses are usually assumed to scale with the square root of the RF frequency (as a consequence of the metal resistance growing with frequency due to a reduced skin depth), in SiP modulators the RF TL losses grow much faster, scaling rather as the square of the frequency once losses arising from loading by the phase shifters become dominant [97]–[99]. As described below, this simple trend can be used to trade off  $V_\pi$  and cutoff frequency in TW-based devices by resizing the length of the device.

As a first approximation, the SiP phase shifter can be added as an additional parallel RC load to the classic telegraph line model [97]. This is not entirely self-evident, since it assumes currents inside the Si to flow predominantly to and from the waveguide perpendicularly to the main waveguide axis, as opposed to also flowing along the direction of the waveguide. This turns out to be true for typical phase shifter configurations as a consequence of the conductivity of the doped Si serving to transport the signals to and from the waveguides remaining much below that of the metal lines [97]. In addition, care has to be taken to pick silicon-on-insulator (SOI) wafers whose Si substrate has ultrahigh purity/high resistance, as induced eddy currents can otherwise be sources of RF losses [97], [98].

Since currents flow laterally to and from the Si waveguide, the transverse electromagnetic (TEM) field symmetry, which implies only longitudinal currents, is broken and the phase velocity of the RF wave can be slowed beyond what would be predicted from the refractive indices of the surrounding materials. This is reflected in the additional capacitive load of the phase shifter being added

to the equivalent circuit model without the usual decrease in linear TL inductance  $L_{\text{TL}}$ :

In a simple coplanar TL design, bringing the two metal lines closer together increases the linear TL capacitance  $C_{\text{TL}}$  but also decreases the linear TL inductance  $L_{\text{TL}}$ , leaving the phase velocity unchanged. The product  $L_{\text{TL}}C_{\text{TL}}$  is more or less fixed by the phase velocity of the unloaded transmission line, since

$$\omega_{\text{RF}} \sqrt{L_{\text{TL}}C_{\text{TL}}} = \frac{\omega_{\text{RF}} n_{\text{RF}0}}{c_0} \quad (12)$$

with  $n_{\text{RF}0}$  the RF effective index of the unloaded transmission line. Since the unloaded TL supports a quasi-TEM mode,  $n_{\text{RF}0}$  is essentially given by the field overlap with the surrounding materials (not strictly TEM due to their inhomogeneity, but nonetheless a useful approximation).

Here, on the other hand, the linear phase shifter capacitance  $C_l$  is added to the capacitive load independently of  $L_{\text{TL}}$ . At low RF frequencies below the RC cutoff of the phase shifter, at which its capacitance is not screened from the TL, the linear waveguide capacitance  $C_l$  can be simply added to the “regular” TL capacitance  $C_{\text{TL}}$  yielding the RF wave number

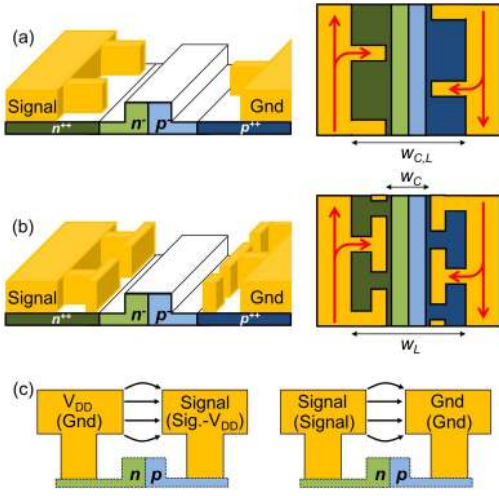
$$\text{Re}(\beta_{\text{RF}}) \simeq \omega_{\text{RF}} \sqrt{L_{\text{TL}}(C_{\text{TL}} + C_l)}. \quad (13)$$

Similarly, the characteristic impedance is modified as

$$Z_{\text{RF}} \simeq \sqrt{\frac{L_{\text{TL}}}{C_{\text{TL}} + C_l}}. \quad (14)$$

$L_{\text{TL}}$  and  $C_{\text{TL}}$  can be adjusted by varying the spacing between the metal lines. Unfortunately, for a fixed  $C_l$ , i.e., for a fixed phase shifter design, this only allows setting one of  $\text{Re}(\beta_{\text{RF}})$  (phase matching) or  $Z_{\text{RF}}$  (impedance matching): Increasing the distance between the metal lines reduces  $C_{\text{TL}}$  and increases  $L_{\text{TL}}$ . As a consequence, both  $\text{Re}(\beta_{\text{RF}})$  and  $Z_{\text{RF}}$  increase. However, small deviations from phase matching can be compensated by implementing small periodic detours in the waveguide (if  $n_g < n_{\text{RF}}$ ) [99] or adding additional distributed capacitive loads to the transmission line in slow-wave TLs (if  $n_g > n_{\text{RF}}$ ) as described below [100], [101]. With these additional degrees of freedom, both characteristic impedance and phase matching can be precisely dialed in.

One difficulty lies in the fact that separating the metal lines from each other also increases the linear series resistance of the phase shifters, reducing their RC limited cutoff frequency and increasing TL line losses. This can be remedied by extending metal finger electrodes from the metal lines to the SiP phase shifter [Fig. 6(a)]. If these finger electrodes are sufficiently far apart to prevent direct capacitive coupling between them, longitudinal currents in this region remain suppressed, so that the RF properties of the TL remain dictated by the initial electrode spacing [99].



**Fig. 6.** (a) and (b) Segmented electrode configurations.

(a) Electrode extensions reduce the phase shifter's series resistance without significantly impacting phase velocity and characteristic impedance of the transmission line (TL). The linear inductance and linear capacitance of the TL (in addition to the phase shifter capacitance) remain primarily determined by the distance  $w_{C,L}$  between the main coplanar electrodes. (b) T-shaped extensions are shaped to significantly increase the linear capacitance of the TL, further slowing down the RF wave. The linear inductance remains determined by the distance  $w_L$ , while the capacitance between the opposing top bars of the T-shaped extensions plays a significant role (the linear capacitance becomes determined by  $w_C$  in the limit where they entirely fill the length of the TL). (c) Driving scheme in which shielding between the two MZM arms is not necessary, as the two MZM arms taken together effectively behave as a GSG coplanar TL from an RF perspective. Two driving schemes are indicated by labels; the top one is more amenable to driver integration, and the lower one (in parenthesis) is more adequate when signals are brought in with RF cables referenced to a common ground.

One may go even one step further and replace the simple fingers by T-shaped structures [Fig. 6(b)] [100], [101]. This then adds an additional linear capacitance to the equivalent TL model and further slows down the RF wave in a slow-wave TL. While complex, multi-metal-layer CMOS back-end stacks provide many more degrees of freedom to tailor TL properties, the techniques outlined above require only one low RF-loss metal layer, which can be particularly beneficial from a cost saving perspective in dedicated photonic processes.

Finally, RF signal coupling between adjacent transmission lines transporting complementary signals to the two phase shifters can severely degrade the high-frequency performance of modulators. A simple, but clearly unsatisfying approach from a system integration perspective consists in allowing for significant spacing between the modulator arms. Full shielding with a multilayer metal stack pushes complexity into the required CMOS back-end-of-line (BEOL) [30]. Alternative solutions have been found, e.g., by using wire bonds to short multiple ground lines with each other (compensating for the lack of additional metal planes) [102] or by implementing driving schemes in which a single, clean RF supermode

of the coupled TLs is excited with a well-defined, phase matched velocity [99] [Fig. 6(c)]. This is arguably the simplest solution since it does not require any additional technological complexity. It has been shown to enable close packed modulator arrays with minimal RF cross talk [103].

RF signal losses inside the TW phase shifters can now be straightforwardly analyzed. In addition to the usual TL losses associated to the resistance of the metal lines, which scale as  $\text{dB/cm}/\sqrt{\text{Hz}}$  and may remain dominant at low frequencies, at higher frequencies RF losses are dominated by ohmic losses associated to the series (linear) resistance  $R_l$  of the phase shifters and scale as  $\text{dB/cm}/\text{Hz}^2$ , growing much faster with frequency. They can be expressed as [97]

$$\alpha_{\text{RF}} = \omega_{\text{RF}}^2 C_l^2 R_l Z_{\text{RF}}. \quad (15)$$

As expected, the linear loss coefficient grows with increasing phase shifter resistance  $R_l$ . It also grows with the phase shifter capacitance  $C_l$  since an increased capacitance leads to a larger phase shifter current intake, and thus to faster power dissipation. Finally, it grows with the TL's characteristic impedance  $Z_{\text{RF}}$ . The power entering the transmission line is given by  $V_0^2/Z_{\text{RF}}$ , wherein  $V_0$  is the root mean square (RMS) voltage supplied by the driver at the beginning of the TL, and evolves as  $P(z) = |V(z)|^2/Z_{\text{RF}}$ . The larger this power, the longer it takes to dissipate with linear power losses  $dP/dz = -\omega_{\text{RF}}^2 C_l^2 R_l |V(z)|^2$ . The dependency on  $Z_{\text{RF}}$  can be leveraged, since it allows a direct tradeoff between power consumption and cutoff frequency, provided the latter remains below the intrinsic RC limited phase shifter cutoff frequency and remains limited by transmission line losses: By redesigning the TL to feature a lower characteristic impedance, power consumption is increased, but  $\alpha_{\text{RF}}$  is reduced and in turn the cutoff frequency increased (at equal phase shifter length). This is one reason why high-speed modulators are sometimes designed with a reduced  $25 \Omega$  impedance.

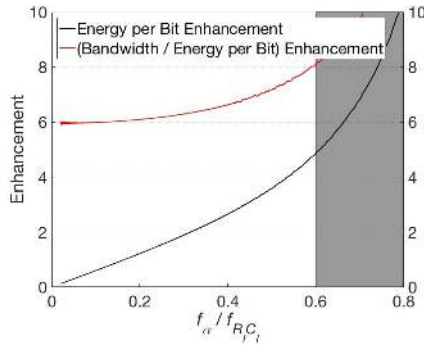
A similar tradeoff can be reached by resizing the length of the phase shifters. As described above, the TL loss limited cutoff frequency is given by  $l\alpha_{\text{RF}} = 1.48$ , resulting in

$$f_{\alpha} = \frac{\sqrt{1.48}}{2\pi C_l \sqrt{R_l Z_{\text{RF}}}}. \quad (16)$$

Since the required drive voltage scales as  $V_0 \propto 1/l$ , improving the bandwidth of the modulator by shortening it also results in an increase of the required drive voltage (assuming extinction to be maintained) and thus of power consumption as well. In comparison, reducing the characteristic impedance does not result in increased drive voltage requirements and results in a more moderate power consumption increase.

This scaling law can also be verified with empirical data: In [104], for example, two TW modulators were reported





**Fig. 7.** Enhancement of energy per bit ( $E_{\text{bit}}$ ) dissipated in the phase shifters of a segmented modulator with distributed driver compared to the energy per bit dissipated in both the phase shifter and termination of a TW modulator (black curve), as a function of the ratio of the transmission line loss limited cutoff frequency ( $f_\alpha$ ) to the intrinsic RC cutoff frequency limitation of the phase shifter ( $f_{RL}C_L$ ). While the power dissipation enhancement associated to a distributed driver configuration depends strongly on  $f_\alpha/f_{RL}C_L$ , once the overall device cutoff frequency  $f_c$  is factored in, the enhancement of the ratio  $f_c/E_{\text{bit}}$  appears to be in a range 6–8. The grayed out area corresponds to a design space in which the TW modulator's cutoff frequency is mostly limited by  $f_{RL}C_L$ , in which case it would appear rational to either increase the characteristic impedance of the modulator to reduce its power consumption at minimal cost in terms of bandwidth, to increase the phase shifter length, or to redesign the phase shifter. For a complete discussion, see [21].

with the same phase shifter design (53-GHz RC limited cutoff frequency) but different phase shifter lengths of 1.25 and 2.5 mm. These resulted in overall modulator cutoff frequencies of 23 and 17 GHz, respectively, both significantly below the RC limited cutoff and limited by the TL losses. The cutoff frequency of the shorter modulator is almost  $\sqrt{2}$  times larger than the longer one as a result of halving of the modulator length (with a slight RC cutoff effect for the shorter modulator).

The general argument is often made that TW devices have a much higher power consumption than LE devices due to the necessity of adding a termination at the end of the TL, which constantly consumes power as opposed to only when the device is switched. This is actually a rather coarse simplification, as the power consumption of TW and LE devices are linked to each other: If a TW device cutoff frequency is limited by TL RF losses—which is typically the case in high-speed devices in which this limits the phase shifter length—the RF power is actually also largely consumed inside the phase shifter at maximum operating speeds as opposed to being primarily consumed in the termination. A detailed analysis shows that TW devices have a power consumption typically 6–8 times larger than an LE device with comparable bandwidth and phase shifter technology [21] (Fig. 7) assuming that both are well-optimized devices. Thus, the phase shifter FOM  $C_L(V_\pi l)^2/8$  remains relevant, even though a scaling factor has to be considered.

## D. Distributed Drivers

Finally, a third modulator driving scheme consists in the combination of segmented phase shifters with distributed drivers [94]–[96]. This driving scheme combines the advantages of LE and TW driving, at the cost of higher circuit, packaging, and integration complexity. In this driver configuration, the signal from the transmission line is first amplified by distributed amplifier elements prior to being applied to phase shifter segments. Each segment is driven as an LE, with the associated RF power consumption. On the other hand, the phase shifter segments are driven with the appropriate delays relative to each other to match the transit time of the light. Thus, the cumulative length of the phase shifter segments taken together is not limited by phase matching considerations. The TL can be considered to transport information, including control over appropriate delays. Signal power, on the other hand, is provided by the hooked up amplifier elements in an ideal manner.

Of course, such a driving scheme is more complex than a simple LE driver, also leading to some overhead in internal driver power consumption. Moreover, it requires tight connectivity between the distributed driver and the modulator. This can be elegantly handled if electronics and photonics are both monolithically integrated in the same chip [94], [96]. In case of hybrid integration of a dedicated electronic chip with a pure optical chip, flip-chip attachment of the driver onto a larger SiP chip by means of microbump arrays provides both a high degree of connectivity and reduced inductive parasitics [105].

## IV. FOLDED, SLOW-WAVE, AND RESONANT ENHANCEMENT

After reviewing phase shifter technologies in Section II, as well as rectilinear phase shifters and modulator driving schemes in Section III, various architectures that have been developed to increase the efficiency of modulators in the optical domain are surveyed here. These can be categorized as: 1) folded or meandered structures; 2) slow-wave modulators; and 3) resonantly enhanced modulators. They are mostly applied to LE modulators, even though some concepts such as slow light are also, in principle, applicable to TW-type driving. While the enhancement factors reported in many works often focus on the power reduction of an LE modulator after application of the enhancement technique, one should mention that an even more important enhancement can consist in shrinking the device to enable LE driving in the first place, since the large power overhead associated to TW devices or the integration complexity of distributed drivers can thus be avoided. Even though these enhancement schemes may seem quite different at a first glance, they are analyzed in a common framework and shown to share common tradeoffs.

We start this section with a discussion of the simple overlap integral used to assess the phase shifting efficiency of baseline, straight phase shifters. The discussion

continues with an overview of folded, slow-wave, and resonant devices. The section closes with some general considerations on the RF power reduction/optical bandwidth tradeoff seen in these devices.

### A. Overlap Integral Exemplified With a Slot Waveguide Phase Shifter

In a regular optical waveguide, i.e., with a simple translation invariance along  $z$ , the change in effective index  $\Delta n_{\text{eff}}$  is related to the change in refractive index across the waveguide  $\Delta n(x, y)$  by an overlap integral that describes how the field intensity overlays the region of refractive index change. In a high index contrast material system such as SiP, one has to rigorously derive the overlap integral from the complete, fully vectorial version of Maxwell's equations, resulting in

$$\Delta n_{\text{eff}} = \frac{\iint n |E|^2 \Delta n \cdot dS}{Z_0 \iint \text{Re}(\vec{E} \times \vec{H}^*) \cdot d\vec{S}} \quad (17)$$

with  $Z_0$  the impedance of free space and  $\vec{E}$  and  $\vec{H}$  the electric and magnetic (H) fields.  $dS$  is a surface element and the double integral is taken over the entire waveguide cross section. As intuitively expected, (17) predicts that the effective index change induced by a refractive index perturbation  $\Delta n$  of bounded extent is maximized if the perturbation is located at the position of highest E-field intensity. Since a refractive index change corresponds to a change of the electric dipole density, it stands to reason that the E-field is the relevant quantity in the numerator. The denominator is simply a normalization of the mode to unit power. If a constant  $\Delta n$  is applied to a bounded area  $S$ , (17) simplifies as

$$\Delta n_{\text{eff}} = \frac{\iint_S n |E|^2 \cdot dS}{Z_0 \iint \text{Re}(\vec{E} \times \vec{H}^*) \cdot d\vec{S}} \Delta n = \Gamma \Delta n \quad (18)$$

where  $\Gamma$  is referred to as the overlap of the index perturbation with the optical mode. While in a low index contrast system such as a glass fiber  $Z_0 \text{Re}(\vec{E} \times \vec{H}^*) \cdot \vec{e}_z \simeq n |E|^2$ , with  $\vec{e}_z$  the unit vector along the  $z$ -direction, and the overlap integral thus is a number between 0 and 1 as one would expect based on the naming, in a high index contrast system such as SiP,  $\Gamma$  can also take a value above 1. As we will see, this is related to the slow light effect: The ratio between  $n |E|^2$  (related to the optical energy density) and  $\text{Re}(\vec{E} \times \vec{H}^*) \cdot \vec{e}_z$  (the energy flux along the main axis of the waveguide) depends on the speed of light in the waveguide and is enhanced as the light is slowed down.

The efficiency of a phase shifter is thus maximized by moving the optically active region, in which the refractive index is modulated, to the region of highest field intensity. In a slot waveguide, this is done by placing the slot in the center of the waveguide. An additional enhancement arises

from the slot also boosting the E-field as a consequence of the continuity of the displacement (D-)field in the direction perpendicular to the interface: For transverse electric (TE) polarized light primarily polarized along the in-plane  $x$ -axis, the  $D_x$  field component is the same right before and right after the interface defining the slot boundaries. Since the slot is typically filled with a polymer with refractive index  $n_{\text{slot}}$  substantially lower than the refractive index of Si  $n_{\text{si}}$ , the E-field inside the slot is boosted by  $(n_{\text{si}}/n_{\text{slot}})^2$  and  $\Gamma$  by  $(n_{\text{si}}/n_{\text{slot}})^3$  provided the slot is small enough to not otherwise modify the E-field [97].

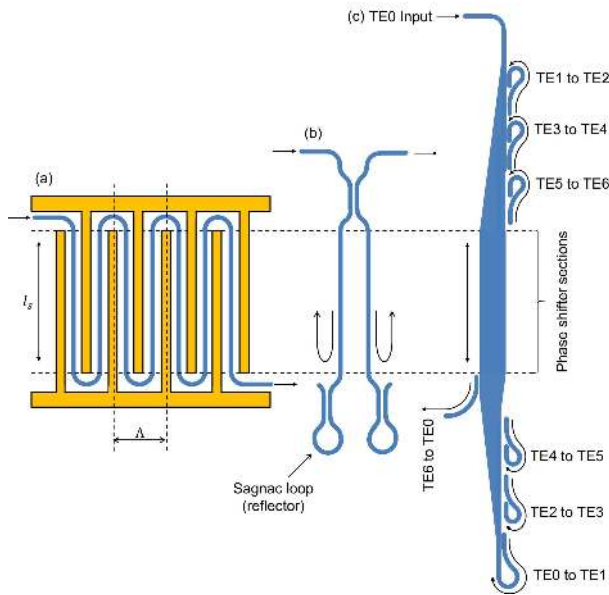
### B. Folded Modulators

The first general category of enhancement we will look at is referred to here by the somewhat general term “folded modulator” and refers to a modulator in which either 1) the phase shifter section is folded, e.g., meandered [21], in order to reduce its dimensions; or in which 2) the light is sent back multiple times through the same phase shifter section [106], [107] without building up a resonance. These essentially boil down to the same from a more abstract perspective. Also, these enhancement schemes come with no or only relatively modest (2x) RF power enhancement if compared to the unfolded structure driven by a distributed driver. However, these techniques may shrink the phase shifter dimensions to enable LE driving in the first place and consequently contribute to a substantial power or integration complexity reduction.

The simplest idea to reduce the size of a phase shifter might be to fold it in tight bends, i.e., to meander it [Fig. 8(a)]. In such a configuration, a long optical path length can be folded into a short electrical path. The E/O cutoff frequency will at some point be limited by the transit time of the light ( $\tau_g$ ) as described in Section III, (8) in the LE approximation and (10) in the more general case. Moreover, past a critical lumped capacitance, driving of the structure becomes problematic, particularly if wire bonds with their high parasitic inductance are involved.

One may rather fold the light back into the same waveguide, e.g., by using a Sagnac loop [106] [Fig. 8(b)]. The light can then travel twice through the same waveguide, doubling the phase shift efficiency. The capacitance is effectively halved compared to an unfolded modulator, resulting in a power enhancement factor of 2, while other metrics such as the effective  $V_\pi l_{\alpha_{\text{dB}}}$  remain the same since the light travels twice through the phase shifter. This architecture precludes, of course, the use of a TW configuration, since phase matching cannot be obtained for both propagation directions jointly. If the phase shifter is single mode and the polarization of the light maintained, a reflection can only be implemented at one end without inducing a resonance (coupling back into the same mode, with the same polarization and direction of propagation, closes a resonance).

In a recent publication [107], it was proposed to repeatedly fold the light back into the same multimode



**Fig. 8.** Folded phase shifter architectures. (a) Meandered phase shifter and (b) and (c) phase shifters in which the light is cycled multiple times without building up a resonance. In (b), an MZM is configured for the light to transit twice through each phase shifter: A Sagnac loop reflects light back at the end. In (c), the light is sent seven times through the same phase shifter. The waveguide mode is iterated at each reflection to prevent building up a resonance (adapted from [107]). Active waveguide volumes in (a) and (c) are similar.

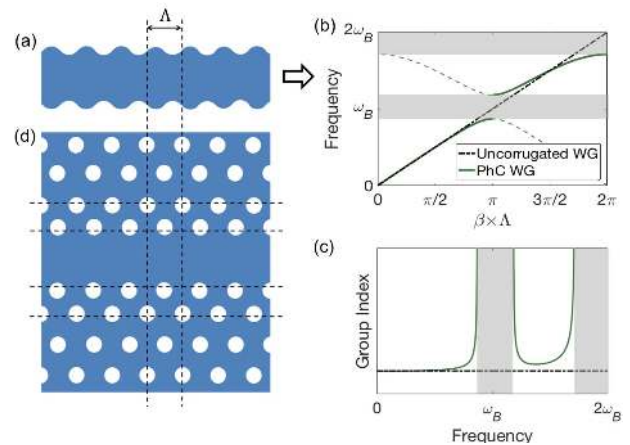
phase shifter, modifying the waveguide mode at each reflection in order to prevent the buildup of a resonance [Fig. 8(c)]. However, since the required waveguide width also scales with the number of supported modes, the actual enhancement remains modest: If the light is folded back  $N$  times through a multimode phase shifter whose width is also widened by a factor  $\sim N$  to support that number of modes, the enhancement is voided. Either the mode overlap will be reduced by a factor  $N$ , thus reverting to a single pass phase shifter efficiency, or the volume of the region in which a refractive index change occurs has to be increased by a factor  $N$ , so that the power consumption also grows accordingly in case of high-speed modulation. An enhancement of 2 could be obtained by also rotating the polarization of the light by  $90^\circ$  at each reflection, since light could then travel twice through the phase shifter with a given mode index but an orthogonal polarization; however, there would be additional difficulties associated to polarization rotation in the implementation of such a structure. Of course, there are some corrective factors to these simple scaling rules, such as changing mode confinements. However, the achievable enhancement can be seen to be clearly bounded for a high-speed phase shifter (for a thermal phase shifter, as shown in [107], the entire multimode waveguide can be efficiently heated).

Interestingly, at an abstract level, this scheme is equivalent to the meandered modulator: In the latter, the array

of optically uncoupled waveguides between bends can be considered a multimodal structure with a natural base given by the individual, uncoupled waveguide modes and the supermode index simply given by the number of the waveguide carrying the light. The array of bends on either side are then seen as reflectors that increment the supermode number by one.

### C. Slow-Wave Modulators

In a slow-wave modulator, the group velocity of light is reduced in order to increase the phase shifting efficiency. One approach to slow down light in a waveguide is to apply a simple corrugation [Fig. 9(a)]. This leads to a stopband, i.e., a forbidden optical frequency region, to open in the dispersion diagram of the waveguide [Fig. 9(b)]. Light with frequencies falling inside that stopband cannot propagate and instead gets reflected back upon entering the waveguide. This stopband opens around the angular frequency  $\omega_B$  corresponding to the Bragg condition  $2\pi/\Lambda = 2\omega_B n_{\text{eff}}/c_0$ , with  $\Lambda$  the period of the corrugation. The stopband has a frequency width that grows with the strength of the corrugation. As a side effect of this stopband opening, the group velocity of the light, given by the slope of the dispersion diagram  $v_g = \partial\omega/\partial\beta$ , with  $\omega$  the angular frequency of the light and  $\beta$  its wave number, tends to zero at the stopband edges. While light is slowed, the amount of slowing is very sensitive to the optical



**Fig. 9.** Slow-wave phase shifters: (a) corrugated waveguide; (b) corresponding dispersion diagram; and (c) frequency-dependent group velocities. The corrugation opens stopbands (shaded in gray) in frequency ranges close to the Bragg conditions for reflection of light back into the waveguide (multiples of  $\omega_B$  in absence of “baseline” waveguide dispersion). As the frequency of light propagating inside the waveguide approaches the edges of the stopbands, the group velocity tends to zero resulting in a slow light enhancement of the induced phase shift. In a photonic crystal waveguide (d), displacement of one or several rows of holes relative to their nominal positions (horizontal dashed lines) gives the degrees of freedom to flatten the group velocity in an extended slow light wavelength range, at the cost of a less confined mode compared to simple nanowire waveguides (adapted from [108]).

frequency, due to the curvature of the dispersion diagram [Fig. 9(c)]. Photonic crystal (PhC) waveguides in which light inside a central line defect interacts with several rows of periodically placed holes [Fig. 9(d)] have been designed to feature large and relatively constant slowing over extended frequency ranges [108], [109].

There are several ways to visualize how this enhancement mechanism works. An intuitive approach typically taken in the photonic crystal community is based on analyzing changes applied to the waveguide dispersion diagram under the effect of a refractive index change. This analysis starts with the converse of (18), which describes the change of the effective index  $\Delta n_{\text{eff}}$  at a fixed optical wavelength  $\lambda_0$  under the effect of a refractive index change  $\Delta n$  (the end result that we actually need to further model the modulator properties). We start by analyzing the simpler problem of how the wavelength  $\lambda_0$  changes at a fixed wave number  $\beta$ , assuming that the refractive index is changed by a fixed, small percentage  $\Delta n = \varepsilon \cdot n$  throughout the structure. The wavelength then changes by an amount  $\Delta \lambda_0$  such that  $\Delta \lambda_0 / \lambda_0 \simeq \Delta n / n = \varepsilon$ . Using  $\Delta \lambda_0 / \lambda_0 \simeq -\Delta \omega / \omega$ , we find that the dispersion diagram [Fig. 9(b)] is shifted up or down accordingly.  $\Delta \omega$  at fixed  $\beta$  is now converted into  $\Delta \beta$  at fixed  $\omega$  taking the slope  $v_g$  of the dispersion diagram into account. This results in  $\Delta \beta \equiv 2\pi \Delta n_{\text{eff}} / \lambda_0 = \omega / v_g \cdot (\Delta n / n)$ , with readily apparent slow light enhancement  $\Delta n_{\text{eff}} = n_g (\Delta n / n)$ .

Of course, this could also have been directly deduced from (17), with the benefit of also allowing for arbitrary spatial profiles  $\Delta n(x, y)$  so long as they remain sufficiently small to be in the domain of validity of perturbation theory. In order to make the group velocity explicitly apparent, we replace the Poynting vector in (17) by the product of the energy density with the group velocity. Since corrugated or PhC waveguides do not have a continuous, but a discrete translation invariance, (17) needs to be generalized to a 3-D equivalent covering the entire unit cell (UC). Averaged over an entire cell, energy equipartition in the E- and H-fields is further invoked to yield

$$\begin{aligned} \Delta n_{\text{eff}} &= \frac{n_g \iiint \Delta n \cdot n |E|^2 \cdot dV}{Z_0 c_0 \iiint n^2 \varepsilon_0 |E|^2 \cdot dV} \\ &= \frac{\iiint n_g (\Delta n / n) \cdot n^2 |E|^2 \cdot dV}{\iiint n^2 |E|^2 \cdot dV} \end{aligned} \quad (19)$$

where the volume integrals are taken over the entire UC. One might observe that it takes a very similar form to the overlap integral for low index contrast systems, with the additional  $n_g$  in the numerator and  $n$  in the denominator accounting for slow light effects in high index contrast material systems.

A couple of remarks can now be made in regards to slow-wave modulators: If no excess losses are added by the slow light waveguide itself, the  $V_{\pi} \alpha_{\text{dB}}$  of the modulator remains unchanged. Indeed, carrier-induced absorption losses (associated to the imaginary part of  $\Delta n$ )

are enhanced by exactly the same factor as the effective index change. Moreover, photon transit time limitations to the cutoff frequency also remain exactly the same: For a given phase shifter length, the photon transit time gets increased by the same multiplicative factor by which  $V_{\pi} l$  gets divided. In other words, one may shrink the phase shifter as a consequence of the slow light effect; the photon transit time and the insertion losses, however, remain unchanged for a given targeted phase shifting capability. The power consumption, on the other hand, decreases as a consequence of the smaller device length and scales directly with the slow light enhancement if applied to an LE device. The enhancement is even more pronounced if an initially TW phase shifter is shrunk to being LE driven.

Some amount of excess losses is also generally incurred due to the slow light waveguides themselves. While great progress has been made in the reduction of excess losses from slow light PhC waveguides [110], power consumption remains extremely sensitive to excess losses at the system level: Assuming a link to require on the order of 10 mW of optical power—a typical number for SiP—and further assuming a typical 15% laser wall-plug efficiency, losing only 1 dB of power to excess losses would result in an additional  $\sim 17$  mW laser power consumption per channel if the OMA cannot be recovered by means of increasing a limited drive voltage. Such overall considerations need to be taken into account in practical transceiver architectures when assessing optical enhancement techniques.

Importantly, simply taking  $n_g / n$  as the FOM for slow light enhancement can be very misleading, since it is only the correct metric if the effective area of the optical mode remains unchanged. If, when moving from a regular nanowire waveguide to a slow light PhC waveguide, the mode becomes wider,  $\Delta n$  also needs to be applied to a larger area in order for the phase shifting range not to be penalized by this. In other words, the correct metric for the slow light enhancement of high-speed phase shifters is  $n_g S_{\text{eff}0} / n_{g0} S_{\text{eff}}$ , with  $S_{\text{eff}0}$  the effective mode area of a nonenhanced reference design relying on a compact cross-section single-mode waveguide,  $n_{g0}$  its group index, and  $S_{\text{eff}}$  the effective mode area of the slow light phase shifter. Assuming the refractive index modulation to be applied in a small region where the refractive index and the E-field take the values  $n_m$  and  $E_m$ , this effective area is given by

$$S_{\text{eff}} = \frac{\iiint_{UC} n^2 |E|^2 \cdot dV}{\Lambda \cdot n_m^2 |E_m|^2}. \quad (20)$$

To develop this point, one may consider the meandered modulator introduced in Section IV-B as a segmented waveguide in which one meander forms the unit cell. The period  $\Lambda$  of the segmented waveguide is then two times the bend radius [Fig. 8(a)]. Further referring to the length of the straight sections between bends as  $l_s$ , the light is slowed by a factor  $(\pi \Lambda + l_s) / \Lambda$ . The mode volume and the phase shifter's capacitance are, however, also increased by the same factor relative to a straight waveguide, so that



the power consumption enhancement factor, assuming both straight and meandered phase shifters to be LEs, is—as expected—exactly 1. While the meandered modulator presents all the formal prerequisites for an optically broadband, arbitrary slow-waveguide configuration, it does not present any power enhancement due to its equally demultiplied mode volume. A somewhat contrived example, it exemplifies the crucial importance of taking both mode volume and slow light effects into account.

This is indeed an important aspect to take into consideration when designing slow light PhC modulators: For example, in [108] and [109], optically wideband slow light enhancement was obtained by shifting the second or third row of holes, as counted away from the waveguide core [Fig. 9(d)]. However, interaction with these outer holes is only made possible by a relatively delocalized waveguide mode. To maintain a high spatial overlap, one may implement the modulator with wide, vertical junctions or with interleaved PIN junctions [111], although the increased junction capacitance then also burdens the RF power consumption.

One may note that in order to achieve an enhancement, i.e., slow light without increasing the mode area, the optical energy densities inside the waveguide necessarily have to rise (since transported power density is the product of energy density with group velocity). This property is shared with resonantly enhanced devices, in which resonant enhancement of the modulator power consumption also scales with the enhancement of the power circulating inside the cavity. Thus, in both categories of devices, attention should also be given to nonlinear effects, including nonlinear optical losses, when high enhancement factors are combined with moderate input power levels. Moreover, opening of the stopband creates a dead zone in the optical spectrum in which the device cannot be operated, which follows similar constraints as resonantly enhanced devices, albeit much more favorably due to the small periodicity of typical slow-wave structures (see Section IV-F).

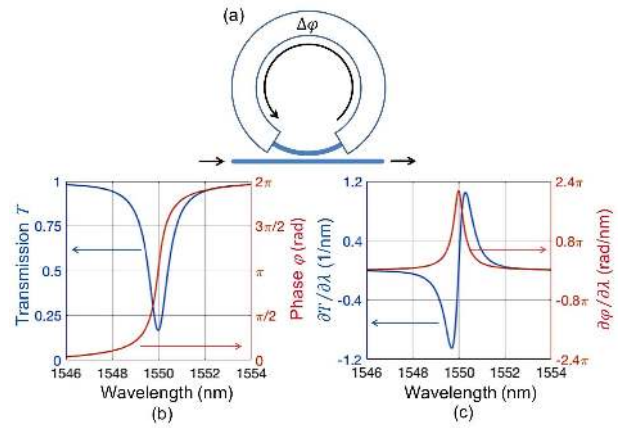
#### D. Resonant Enhancement

Another approach for the enhancement of electrically LE modulators consists in embedding the phase shifter in a resonator, e.g., a ring resonator [112] [Fig. 10(a)] or, even smaller, a 1-D PhC cavity embedded in a Si waveguide [113]. Resonant ring modulators (RRMs) can straightforwardly support high cutoff frequencies in excess of 50 GHz [114].

The basic idea behind resonant modulators is to shift the resonance frequency  $f_r$  by means of index modulation, wherein the resonance frequency shift  $\Delta f_r$  is given by

$$\frac{\Delta f_r}{f_r} = -\frac{\Delta n_{\text{eff}}}{n_g} = -\frac{\Delta n \cdot \Gamma}{n_g}. \quad (21)$$

Inclusion of the group index  $n_g$  in the denominator factors in that the effective index changes both due to the



**Fig. 10. RRMs. (a) Diagram of the device. (b) Power and phase transfer functions of an overcoupled ring resonator as a function of wavelength. (c) Slopes of the transfer functions indicative of the intensity and phase modulation efficiencies versus carrier wavelength.**

refractive index modification  $\Delta n$ , resulting in  $\Delta n_{\text{eff}}$ , the effective index change at a fixed wavelength, as well as due to the resonant wavelength change, which needs to be self-consistently taken into account. Since the optical transfer function of a resonant device is highly dependent on whether the optical carrier falls within or without the resonance, shifting the resonance in and out of the optical carrier can have a drastic effect on both the transmitted amplitude and phase. For example, the transfer function of a single bus waveguide coupled to a ring resonator features periodic extinctions at each of the resonances [115], so that the transmitted amplitude can be modulated this way (Fig. 10).

The full width at half maximum (FWHM) of a resonance is a function of the quality factor of a cavity ( $Q$ ) and is given by  $\delta f = f_r/Q$ . It gives the characteristic quantity by which the resonance needs to be shifted to obtain substantial modulation. For example, shifting the resonance frequency away from the optical carrier by half the FWHM results in an OMA 3 dB below the power entering the device (3-dB modulation penalty), assuming the ring to be critically coupled, i.e., to have a coupling strength to the waveguide resulting in full extinction on resonance [115]. The  $Q$ -factor in turn is a function of the internal optical losses inside the cavity. The higher the  $Q$ -factor, the lower the resonance wavelength shift, and consequently the  $\Delta n_{\text{eff}}$ , that are required. Thus, these devices can be very efficient.

Unfortunately, in this basic configuration, not only does the resonant enhancement constrain the carrier to coincide with the resonance frequency of the resonator—making it an optically narrowband device—but also reducing the resonator linewidth/increasing the  $Q$ -factor limits the E/O cutoff frequency of the RRM. In a cavity ring down experiment in which the sourced optical power is

suddenly turned off, the power and the E-field inside a resonator exponentially decay with  $1/e$  time constants given, respectively, by  $Q/\omega_r$  and  $2Q/\omega_r$  (factor 2 results from the E-field scaling as the square root of the optical power). These also give the characteristic time constants with which the light stored inside the device reacts to other changes, such as, for example, a refractive index change. Since functioning of the device is contingent on the field amplitude inside the device tracking the electrically induced dynamic resonance detuning,  $2Q/\omega_r$  is also the characteristic time constant with which the modulator can be operated. Since it is the storage time inside the device, it is analogous to the optical transit time  $\tau_g$  in rectilinear LE devices. A more detailed analysis of the dynamical aspects of RRM [116], [117], [114] reveals that the actual cutoff frequency depends on the exact frequency detuning of the optical carrier relative to resonance and can be approximated as  $f_r/2Q + |f_0 - f_r|$  for a critically coupled RRM operated in ASK mode, where  $f_0$  is the optical carrier frequency. The exact E/O transfer function features a peaking response analogous to the response of inductively enhanced peaking amplifiers in the electrical domain [114]. Since in ASK operation the small-signal modulation efficiency of the RRM is maximized for a detuning  $|f_0 - f_r| \simeq 0.29 \cdot f_r/Q$ , a typical bandwidth is  $\simeq 0.8 \cdot f_r/Q$ .

Thus, for high-speed devices, it is necessary to reduce the  $Q$ -factor to obtain a low enough photon storage time, thus also reducing the modulation efficiency. If dopant concentrations can be freely chosen, the best way to increase waveguide losses is to increase the P- and N-doping in the embedded phase shifter, since this has the beneficial side effect of also improving its  $V_\pi l$  as well as reducing its series resistance and thus improving its RC time constant (increasing dopant concentrations typically results in the resistance decreasing faster than the capacitance increasing), at no additional cost since the additional waveguide losses are actually needed here. For this reason, high-speed RRM design optimization tends to converge to higher dopant concentrations than for rectilinear MZM phase shifters, making it hard to implement both types of devices effectively in a single technology run without process splits or additional implants.

RRMs can be implemented in different configurations. For example, one may combine modulation with multiplexing by coupling the RRM to a second waveguide in which the modulated signal is being dropped [117]. However, this would be quite inefficient since efficient power transfer from the bus to drop waveguide requires optical RRM losses to be dominated by coupling rather than internal waveguide losses. Since this would leave very little of the “loss budget” (determined by the targeted  $Q$ -factor) for free-carrier absorption, it would strongly constrain doping concentrations and thus the performance of the phase shifter. Thus, modulation and multiplexing with two separate rings can be more efficient [118].

The modulation efficiency of a critically coupled RRM used for ASK and biased with the optical carrier at the point where the Lorentzian transfer function of the ring resonance has its highest slope, further assuming that the phase shifter covers the entire circumference of the ring, can be derived as [114]

$$\frac{1}{P_{\text{in}}} \frac{\partial P_{\text{out}}}{\partial n_{\text{eff}}} \simeq 1.3 \frac{Q}{n_g} = \frac{1.3\pi}{\alpha\lambda_0} \quad (22)$$

where the formula for the  $Q$ -factor of a critically coupled RRM,  $Q = \pi n_g / \alpha \lambda_0$ , i.e., half the intrinsic  $Q$ , was used. Since the modulation efficiency is limited by the linear phase shifter loss coefficient  $\alpha$ , one may compare it to a rectilinear modulator with phase shifters of identical cross section, sized to have a cumulative length  $l_{\text{MZM}} = 2l$  equally split over both arms

$$\frac{1}{P_{\text{in}}} \frac{\partial P_{\text{out}}}{\partial n_{\text{eff}}} = \frac{\pi l_{\text{MZM}}}{\lambda_0} e^{-\frac{\alpha}{2} l_{\text{MZM}}} \quad (23)$$

wherein (23) is obtained by a small-signal linearization of (9) around the quadrature point. This number is maximized if the cumulative length of the two MZM phase shifters is chosen to verify  $l_{\text{MZM}} = 2/\alpha$ , in which case (23) takes the value  $\simeq 0.74Q/n_g$ . In other words, the maximum OMA achievable by either type of device at a given drive voltage is almost the same. OMA, as limited by the available voltage swing, is not fundamentally improved, which is why RRM also tend to have large modulation penalties at low drive voltages notwithstanding the resonant enhancement. The optical transit time of the rectilinear modulator, with this sizing, can be derived as  $2Q/\omega_r$  so that it would also have a comparable E/O bandwidth limitation if operated as an LE ( $f_r \simeq 1.4f_r/Q$ ). As for a slow-wave modulator, the advantage of the RRM is rather reduced RF power consumption stemming from the reduced phase shifter length.

The ratio between modulation efficiencies of RRM and MZI, as given by (22) and (23), can be further expressed as

$$1.3 \frac{F}{\pi} \frac{l_{\text{RRM}}}{l_{\text{MZM}}} e^{\frac{\alpha}{2} l_{\text{MZM}}} \simeq 3.5 \frac{F}{\pi} \frac{l_{\text{RRM}}}{l_{\text{MZM}}} \quad (24)$$

where the exponential was again evaluated assuming  $l_{\text{MZM}} = 2/\alpha$  on the right-hand side of (24).  $l_{\text{RRM}}$  corresponds to the length of the phase shifter embedded inside the RRM and is assumed here to fill the entire circumference.  $F$  is the finesse of the resonator, defined as the ratio of the free spectral range (FSR), i.e., the frequency difference between adjacent resonances, and the FWHM. It takes the value  $\pi/l_{\text{RRM}}\alpha$  for the critically coupled RRM. Thus, the phase shifter length inside the RRM,  $l_{\text{RRM}}$ , is reduced by a factor  $3.5F/\pi$  compared to an MZM with comparable OMA, resulting in a reduced power consumption (the enhancement factor reduces to  $1.3F/\pi$  if compared to a short MZM not suffering from substantial waveguide losses).

At resonance, the power circulating inside a critically coupled ring is  $F/\pi$  larger than the power coming in from the bus waveguide. The electrical power reduction factor thus also corresponds to the amount by which the optical power is boosted inside the RRM.  $F$  also scales with the ratio of the FSR over the optical passband of the device and is thus a metric of how much the optical passband has been restricted. Finally, it corresponds to the amount by which light has been slowed down relative to a single round-trip pass, since the relevant time delay  $2Q/\omega_r$  can be expressed as  $(F/\pi) \cdot \tau_{\text{RT}}$ , with  $\tau_{\text{RT}}$  the time taken by the light to make one round trip around the cavity. Thus, an RRM follows similar exchange relations as a slow-wave modulator.

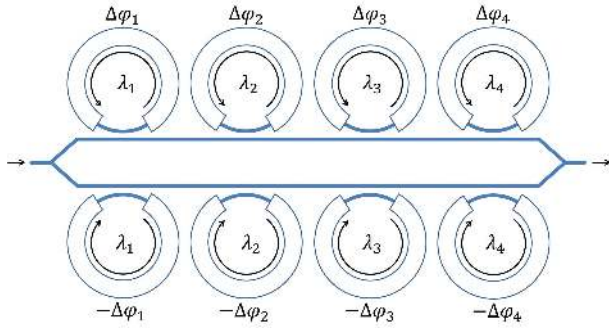
Importantly, the FSR depends on the circumference of the ring and grows as  $l_{\text{RRM}}$  is shrunk, so that the enhancement factor can be increased by further shrinking of the device [since this reduces the capacitance without changing the modulation efficiency given by (22) provided free-carrier absorption remains dominant over bending losses]. Thus, while the drive voltage related modulation penalty is limited by the required E/O bandwidth (limiting the  $Q$ -factor), the power enhancement factor is not. As we will see, it is thus possible to design devices with a relatively wide optical bandwidth that maintain a sizeable resonant enhancement (Section IV-E).

The resonant devices described so far share the common handicap of only functioning for a very narrow range of optical carrier frequencies. Thus, the ring resonance needs to be precisely tuned relative to the wavelength of the laser. While tracking the resonance of the ring and correcting it with a phase tuner is not a major problem from a data acquisition perspective, maintaining operation over extended durations in thermally unstable environments is subjected to the finite tuning range and substantial power consumption of the phase tuners: Uncooled Datacom modules are typically required to function for a case temperature range spanning at least 0 °C to 70 °C, which can be extended to -5 °C to 85 °C in certain specifications [119]. A system has to track significant temperature swings within that range without interrupting data transmission. In addition, at startup, an initial correction has to be applied in order to align the ring resonance with the optical carrier. The resonance frequency of fabricated ring resonators is subjected to variability caused by varying Si film thicknesses in SOI wafers as well as variations in fabricated waveguide widths. While important progress has been made in improving repeatability as SiP fabrication has moved to 193-nm lithography (and better) and 300-mm wafers, resonance frequency repeatability across chips and wafers remains on the order of  $\sigma = 1.8$  nm (std. dev.) [120]. Given the thermo-optic coefficient of Si ( $1.8 \times 10^{-4} \text{ K}^{-1}$  at 1550 nm and room temperature), an initial correction of 4 nm ( $\sim 2$  std. dev.) by itself would already require a thermal phase shifter to apply a temperature swing of  $\sim 60$  °C. Added to additional temperature swings during operation, the maximum tem-

perature applied by the thermal phase shifter could easily reach significantly above 100 °C, not only resulting in a high power consumption, but also posing challenges in regards to the long term reliability of BEOL CMOS stacks. Replacing thermal tuners by other relatively slow but efficient phase shifter technologies such as forward biased PIN diodes results in its own set of difficulties, such as excess carrier-induced losses spoiling the  $Q$ -factor or excess length spoiling the FSR and the finesse of the cavity. For these reasons, reduced power thermal phase trimmers [75], [76] and the reduction of resonant modulators' thermal sensitivity [121], [122] have been very active fields of research. Finally, clever leveraging of wavelength-division multiplexed (WDM) system architectures can be used to alleviate the thermal control problem: While the absolute resonant frequency of ring resonators varies significantly across dies, the relative resonance frequency of nearby rings designed to have a given frequency offset is much more reproducible, with a std. dev. as low as 0.15 nm reported in [120]. Consequently, one may assume the resonances of nearby RRM to shift together in the same direction. If the FSR of the RRM is chosen to coincide with the total width of the WDM communication band, as a ring resonance moves out of the communication band, another one of the same ring enters it. This can be leveraged by reassigning rings to channels whose optical carrier frequencies result in the smallest initial tuning, minimizing the necessary tuning range at system startup [123]. Since, in general, deep (D-)WDM systems already need some amount of global (and power hungry) temperature stabilization, this may yield a practical control system in such a configuration. The frequency selectivity of RRM might even become an advantage to address a specific channel from a number of carriers generated by an optical comb source [124].

## E. Resonator-Assisted Mach-Zehnder Modulators

Resonator-assisted (RA)-MZMs form another category of resonantly enhanced modulators (Fig. 11). Here, ring resonators are used to obtain an enhanced phase shift rather than directly extinguishing the light [125]–[127]. The phase delay introduced by an optical resonator stems from its functionality as a storage element delaying the propagation of the E-field by a time constant  $2Q/\omega_r$  at resonance in a cavity ring down experiment. In an overcoupled ring resonator, for which the coupling strength to the bus waveguide is larger than the critical coupling strength, the phase delay applied by the ring resonator to light transiting in the bus waveguide depends monotonously on the frequency of the delayed light with a slope simply given by  $\tau_g = -\partial\varphi/\partial\omega = KQ/\omega_r$  at resonance.  $K$  is dependent on the coupling strength and varies from 2 to 4 as one transitions from critical coupling to a highly overcoupled regime. As for the RRM described above, dynamically shifting the resonance frequency relative to the optical carrier frequency serves to modify this transfer function,



**Fig. 11. RA-MZM for amplitude modulation of four independent WDM channels.** In this case, the ring resonance linewidth is restricted by the requirement of preventing interchannel cross talk. Alternatively, the rings can be optimized to obtain optically wideband operation, in which case several rings are typically collectively driven on each arm to obtain sufficient phase shifts.

in this case, the magnitude of the phase delay applied to the light, with an efficiency (phase shift per actuation voltage) that also grows with the  $Q$ -factor (Fig. 10). Here too, the actual enhancement in terms of power consumption reduction is a function of the finesse. Since not only phase but also amplitude is being modified by the rings, both aspects have to be included in a complete analysis.

By including a ring on both branches of an RA-MZM and actuating them with complementary signals, push–pull operation can be obtained for amplitude modulation. Such a configuration has also been used for higher order amplitude modulation, such as PAM-4 [128]. By adding an additional  $90^\circ$  phase shift to one of the MZM arms and driving the two rings with independent signals, a very compact  $IQ$ -modulator can also be implemented such as required for SSB [129] or higher order modulation. For example, the E-field transmitted through a critically coupled resonator is zeroed on resonance and flips its sign as the resonance is shifted through the carrier frequency, mimicking the functionality of a nested MZM biased around the 0-point in a conventional  $IQ$ -modulator. Moreover, RRM's tuned to different communication channels can be cascaded in each MZM arm to achieve multichannel operation (Fig. 11).

Direct amplitude modulation with a single ring as described in Section IV-D also results in spurious phase modulation, so that some amount of chirp is unavoidable, worsening the effect of dispersion and limiting transmission distances. Using two rings in push–pull configuration in an RA-MZM significantly reduces the chirp and thus enables transmission of amplitude modulated signals over longer distances [128]. Operating two rings per channel in an RA-MZM also increases the obtained OMA compared to a single ring. Moreover, several collectively driven rings can be cascaded on each arm to increase the achieved phase shift [130]–[134]. While this does not allow improving the OMA beyond the limit constrained by optical phase shifter losses [conf. (9)] as a consequence of cascaded rings

also compounding the latter, this does allow maintaining the OMA while at the same time broadening the optical passband and maintaining resonant enhancement:

Overcoupling a single resonator reduces its phase shifting efficiency, but also its insertion losses. One can actually show that, in the absence of excess losses due to coupling junctions and waveguide bending, for a highly overcoupled ring, the induced losses and the phase shifting efficiency at resonance are both equivalent to those of a rectilinear phase shifter with a length  $2F \cdot l_{\text{RRM}}/\pi$  [133]. In other words, as for a slow-wave modulator, the phase shift efficiency (the equivalent  $V_\pi l$ ) is enhanced, but  $V_\pi \alpha_{\text{dB}}$  remains unchanged. As a consequence, the maximum OMA at a given drive voltage and at an optimized effective phase shifter length  $l_{\text{eff}} = 2F \cdot N \cdot l_{\text{RRM}}/\pi$ , with  $N$  the number of rings, is still limited by (9) and cannot be improved beyond that limit by cascading more rings. On the other hand, cascading a larger number of rings allows reducing their  $Q$ -factor while maintaining  $l_{\text{eff}}$ , thus increasing the optical bandwidth of the device, while at the same time maintaining the OMA.

The highly overcoupled resonators used here as enhanced phase shifters can be described by exactly the same theoretical framework as other forms of slow-wave modulators. The power circulating inside the cavities at resonance is simply given by  $2F/\pi$  times the power in the bus waveguide, thus, it is subjected to the same enhancement factor as the phase shift efficiency due to the slowing down of the light associated to the resonator storage time  $\tau_g = 4Q/\omega_r$ : The group index is enhanced by a factor  $(4Q/\omega_r)/(l_{\text{RRM}}n_g/c_0)$ , the ratio between actual time delay and the time delay  $\tau_{\text{RT}}$  that would have been introduced by the unfolded phase shifter, which is also equal to  $2F/\pi$ .

A main challenge associated to the implementation of such a device is to highly overcouple the rings (to reduce the  $Q$ -factor and thus obtain a large optical bandwidth) while at the same time maintaining a high finesse (to maintain a resonant enhancement) without introducing excess bending or coupling losses (that would spoil the  $V_\pi \alpha_{\text{dB}}$  and thus burden the overall power consumption associated to, e.g., the laser). In particular, to maintain a high finesse while reducing the  $Q$ -factor, the device size has to be shrunk, making it harder to achieve high coupling with low excess losses (small junction lengths) and increasing bending losses due to small radii. By loading each arm of an MZM with  $N = 5$  collectively driven rings, it has been possible to show a modulation efficiency of  $V_\pi l = 0.19 \text{ V} \cdot \text{cm}$  enhanced by a factor 6.8 compared to a rectilinear phase shifter with comparable PIN junction design, with an E/O cutoff frequency of 23.5 GHz limited by the high impedance of the  $50 \Omega$  test environment (the intrinsic E/O cutoff frequency of the phase shifter was modeled as 46 GHz). This modulation efficiency stayed within 3 dB of its maximum within a 3.8-nm optical bandwidth. After cavity design optimization, excess losses associated to bending and excess coupling losses in the five resonators were reduced to a total of 1.2 dB [132]. While some of



the RF power consumption reduction is voided by the increased insertion losses, the main benefit of the device is to enable LE driving with a regular PIN phase shifter and an acceptable modulation penalty of  $-4.8$  dB at  $2 V_{pp}$  drive signals. LE operation is thus obtained without the circuit and packaging complexity associated to distributed drivers. Compared to a TW modulator, power reduction is substantial [133] [ $E_{bit} = 144$  fJ in the definition of (1)].

Importantly, while the MZM had to be actively biased to its quadrature point, the rings themselves were not individually tuned. By increasing the number of rings, the optical passband can be further increased with the objective of accommodating cumulative variability of both ring resonances across chips and lasing wavelength, as well as environmental temperature swings [134].

## F. Resonator-Assisted Versus PhC Slow-Wave Modulators

Since the cascaded rings described above are also a form of slow-wave phase shifter, one might wonder how the tradeoff between enhancement and optical passband compares to PhC-based slow-wave modulators. In the case of rings, it has already been established that the enhancement scales with the finesse as  $2F/\pi$ . Since the finesse is the ratio of the free spectral range with the linewidth of the cavity,  $\pi/2F$  can also be seen as the integral of the optical passband over an FSR (the factor  $\pi/2$  results from the integral of a Lorentzian). A very similar scaling can be seen for PhC slow-wave modulators, as shown here with a simplifying assumption.

We assume that the main effect of the PhC is to open a stopband and to locally modify the dispersion diagram of the waveguide, but that the position of the stopbands can still be predicted from the dispersion diagram of an equivalent unperturbed waveguide [dashed-dotted line in Fig. 9(b)]. This assumption is reasonable in case the slowing is induced by a corrugation opening a stopband, but, as we will see below with the extreme example of a meandered modulator, it is by no means generally true.

The frequency range covered by a single transmission band of the PhC waveguide is given by

$$\Delta\omega = \int_0^{\frac{\pi}{\Lambda}} \frac{\partial\omega}{\partial\beta} d\beta = \int_0^{\frac{\pi}{\Lambda}} \frac{c_0}{n_g} d\beta \quad (25)$$

where the upper bound of the integral is given by the boundary of the first Brillouin zone, where the stopband opens.

If the enhanced group index takes a value of at least  $n_{gE}$  over a frequency range  $\Delta\omega_E$ , we get

$$\Delta\omega_E \leq \int_0^{\frac{\pi}{\Lambda}} \frac{c_0}{n_{gE}} d\beta = \frac{\pi}{\Lambda} \frac{c_0}{n_{gE}} \quad (26)$$

and the maximum achievable group index enhancement is limited by the tradeoff relation  $\Delta\omega_E \cdot n_{gE} \leq (\pi/\Lambda)c_0$ .

Assuming the group index  $n_{g0}$  of the unperturbed waveguide does not change with frequency (dispersionless waveguide), the difference between the center frequencies of two adjacent stopbands is given by  $\Delta\omega_B = (c_0/n_{g0}) \cdot (\pi/\Lambda)$  and

$$\frac{n_{gE}}{n_{g0}} \leq \frac{\Delta\omega_B}{\Delta\omega_E}. \quad (27)$$

$\Delta\omega_E$  plays a similar role as the resonator linewidth, in that it is the useful frequency range over which enhancement is obtained (the optical bandwidth of the modulator).  $\Delta\omega_B$  plays a similar role to the FSR of a resonator (the period over which the enhancement repeats). Typically, only the first and portions of higher PhC transmission bands can be used, due to the rest being above the light line and thus corresponding to lossy free space coupled Bloch modes, but this is not so different from ring-assisted modulators in which typically only a single resonance is being effectively used. While the FSR of a ring is determined by the optical length of its circumference,  $n_g l_{RRM}$ ,  $\Delta\omega_B$  is determined by the characteristic length of the distributed feedback inside the PhC waveguide,  $2n_{g0}\Lambda$ . Since the latter is on the same order as the wavelength, it can be immediately seen that it is significantly smaller than the circumference of even the smallest ring or disk resonators. Hence, similarly to the advantages of DFB over Fabry-Perot lasers, distributed feedback in PhCs allows a substantial improvement of the effective FSR of the device, thus also increasing the width of a single continuous frequency band of operation ( $\Delta\omega_E$ ) at a fixed enhancement. On the other hand, ring-based slow-wave phase shifters can be built entirely from high confinement single-mode waveguides, with high overlap between active phase shifter regions and the optical field, as opposed to lesser confined PhC waveguide modes in optically broadband slow-wave modulators (see Section IV-C).

The meandered modulator described in Section IV-B [Fig. 8(a)], analyzed as a slow-wave segmented waveguide, does not follow (27) and remains optically broadband irrespective of slowing of the light, i.e., irrespective of how much  $l_s$  is increased. Here, the assumption of the dispersion of the segmented waveguide being aligned with that of an underlying unperturbed waveguide of constant group index  $n_{g0}$  (i.e., the dispersion diagram reaching the edges of the first Brillouin zone at multiples of  $\omega_B$ ) does not hold: Instead of resulting from distributed reflection, slowing is entirely due to broadening of the Bloch mode (rays following a zigzag path rather than a straight line, as in higher order waveguide modes). Conversely, there is also no enhancement factor.

## V. PHASE SHIFTERS WITH SURFACE PLASMONS

Since there are clear limitations to slow-wave or resonant enhancement, other improvement paths are being explored to address the limitations of the phase shifters themselves. While a number of phase shifter configurations

have already been described in Section II, enhancement stemming from light propagating as surface plasmon polaritons (SPPs) is treated separately here, as its description requires the overlap and slow light enhancement concepts described in details in Section IV.

SPPs are optical waves that propagate at the interface between a metal and a dielectric. Below the plasma frequency  $\omega_P$  of the metal, the real part of its dielectric constant  $\varepsilon_M(\omega) \simeq 1 - \omega_P^2/\omega^2$  (as approximated by the free electron gas model) is negative and can take very large values far below the plasma resonance. As the frequency of light grows and approaches the plasma frequency, the magnitude of the metal's dielectric constant decreases and reaches a point, slightly below the plasma frequency, at which it is exactly opposite to the dielectric's dielectric constant  $\varepsilon_D$ . At this point, the dispersion diagram of the SPP diverges with the wave number  $\beta = (\omega/c_0) \cdot (\varepsilon_M \varepsilon_P / (\varepsilon_M + \varepsilon_P))^{1/2}$  taking very large values and the group velocity, the derivative  $v_g = d\omega/d\beta$ , also tending to zero. These large wave numbers are also accompanied by a very short evanescent decay length of the light in both the metal and the dielectric, with the decay rate given by  $\alpha_e = \sqrt{\beta^2 - (2\pi n_D/\lambda_0)^2}$  in the latter. Consequently, SPPs are a means to dramatically reduce the effective mode area, reducing the volume in which the refractive index needs to be changed while maintaining a high overlap, while at the same time also obtaining a substantial slow light effect.

SPPs also suffer from high losses due to the high overlap with the metal as well as the slow light enhancement of these losses. However, since the slow-wave effect also applies to the refractive index change, the phase shifter's  $V_\pi l_{\text{dB}}$  is only burdened by the unenhanced metal losses. Shortening of the phase shifters to a few tens of micrometers results in  $V_\pi$  and IL that might be in range of practical applications, while at the same time providing ultrasmall power consumption [135], [136]. The slot waveguide concept has been extended to SPPs by forming a slot in a thin metal film and infiltrating it with an E/O polymer. New device concepts also rely on the metal side

of the SPP waveguide to be implemented by highly doped transparent oxides such as indium–tin–oxide (ITO) [137]. Modifying the free-carrier concentration inside the ITO shifts the plasma resonance, a further highly efficient tuning mechanism.

## VI. CONCLUSION

While silicon photonics modulators based on PIN diode and silicon–insulator–silicon capacitor phase shifters have been maturing and are being implemented increasingly as standardized basic cells, novel materials and modulator topologies remain very active fields of research. Limitations of Si-only phase shifters, in particular their insertion losses, are triggering a high level of interest into hybrid devices with lower losses and increased modulation speeds. As industry discusses the next form factors that will follow 400G modules, renewed interest has appeared in further improvement of SiP modulators.

New materials, in particular III–V compound semiconductors [47], [48], offer a path to improving the characteristics of silicon-only phase shifters. Their introduction into CMOS process flows, however, also results in increased manufacturing costs, so that implementation in industrial practice is contingent on system architectures first moving beyond the performance limits of currently deployed technology. Resonant and slow-wave enhancements might also find their way into industrial practice in CWDM systems, in which temperature stabilization is already a prerequisite [124]. As photonic–electronic copackaging advances, increased pressure on modulator footprint and power consumption raises the interest in enhanced modulators sufficiently optically broadband to be compatible with uncooled, free running operation—an objective toward which substantial progress has been made [134].

This paper aimed at providing the reader with an overview of the current state of the art of silicon photonics modulators, as well as with a sense of fundamental tradeoffs, sizing rules, and topological choices used in their design. ■

## REFERENCES

- [1] A. Liu *et al.*, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature*, vol. 427, pp. 615–618, Feb. 2004.
- [2] D. Patel, A. Samani, V. Veerasubramanian, S. Gosh, and D. V. Plant, "Silicon photonic segmented modulator-based electro-optic DAC for 100 Gb/s PAM-4 generation," *IEEE Photon. Technol. Lett.*, vol. 27, no. 23, pp. 2433–2436, Dec. 1, 2015.
- [3] L. Alloatti *et al.*, "100 GHz silicon–organic hybrid modulator," *Light, Sci. Appl.*, vol. 3, Mar. 2014, Art. no. e173.
- [4] R. Palmer *et al.*, "High-speed, low drive-voltage silicon–organic hybrid modulator based on a binary-chromophore electro-optic material," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2726–2734, Aug. 15, 2014.
- [5] B. Snyder, B. Corbett, and P. O'Brien, "Hybrid integration of the wavelength-tunable laser with a silicon photonic integrated circuit," *J. Lightw. Technol.*, vol. 31, no. 24, pp. 3934–3942, Dec. 15, 2013.
- [6] A. Martir-Mártir *et al.*, "Hybrid silicon photonics flip-chip laser integration with vertical self-alignment," in *Proc. Conf. Lasers Electro-Opt. Pacific Rim (CLEO-PR)*, Jul./Aug. 2017, pp. 1–4.
- [7] R. A. Soref and B. R. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.*, vol. JQE-23, no. 1, pp. 123–129, Jan. 1987.
- [8] Draft Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers and Management Parameters for 200 Gb/s and 400 Gb/s Operation, IEEE Standard P802.3bs/D3.2, Jun. 2017.
- [9] O. Ozolins *et al.*, "100 GHz externally modulated laser for optical interconnects," *J. Lightw. Technol.*, vol. 35, no. 6, pp. 1174–1179, Mar. 15, 2017.
- [10] C. R. Doerr *et al.*, "Single-chip silicon photonics 100-Gb/s coherent transceiver," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2014, Art. no. Th5C.1.
- [11] A. Rao and S. Fathpour, "Compact lithium niobate electrooptic modulators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 4, Jul./Aug. 2018, Art. no. 3400114.
- [12] C. Doerr *et al.*, "O, E, S, C, and L Band silicon photonics coherent modulator/receiver," in *Proc. Opt. Fiber Commun. Conf. Exhib. (OFC)*, Mar. 2016, Art. no. Th5C.4.
- [13] Y. Luo *et al.*, "Strong electro-absorption in GeSi epitaxy on silicon-on-insulator (SOI)," *Micromachines*, vol. 3, no. 2, pp. 345–363, Apr. 2012.
- [14] J. Frigerio *et al.*, "Giant electro-optic effect in Ge/SiGe coupled quantum wells," *Sci. Rep.*, vol. 5, Oct. 2015, Art. no. 15398.
- [15] M. A. Van Camp *et al.*, "Demonstration of electrooptic modulation at 2165 nm using a silicon Mach-Zehnder interferometer," *Opt. Express*, vol. 20, pp. 28009–28016, Dec. 2012.
- [16] D. J. Thomson *et al.*, "Optical detection and modulation at 2  $\mu\text{m}$ –2.5  $\mu\text{m}$  in silicon," *Opt. Express*, vol. 22, pp. 10825–10830, Apr. 2014.
- [17] R. Ponce *et al.*, "Design of a high-speed germanium–tin absorption modulator at mid-infrared wavelengths," in *Proc. 14th Int. Conf. Group IV Photon. (GFP)*, Aug. 2017, pp. 19–20.

- [18] N. Hosseini *et al.*, "Stress-optic modulator in TriPlex platform using a piezoelectric lead zirconate titanate (PZT) thin film," *Opt. Express*, vol. 23, no. 11, pp. 14018–14026, May 2015.
- [19] K. Alexander *et al.*, "Nanophotonic Pockels modulators on a silicon nitride platform," *Nature Commun.*, vol. 9, Aug. 2018, Art. no. 3444.
- [20] D. A. B. Miller, "Energy consumption in optical modulators for interconnects," *Opt. Express*, vol. 20, no. S2, pp. A293–A308, Mar. 2012.
- [21] S. S. Azadeh *et al.*, "Low  $V_{\pi}$  silicon photonics modulators with highly linear epitaxially grown phase shifters," *Opt. Express*, vol. 23, no. 18, pp. 23526–23550, Aug. 2015.
- [22] B. Milivojevic *et al.*, "Silicon high speed modulator for advanced modulation: Device structures and exemplary modulator performance," *Proc. SPIE*, vol. 8990, Feb. 2014, Art. no. 899013.
- [23] B. Milivojevic, S. Wiese, S. Anderson, T. Brenner, M. Webster, and B. Dama, "Demonstration of optical transmission at bit rates of up to 321.4 Gb/s using compact silicon based modulator and linear BiCMOS MZM driver," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 768–774, Feb. 15, 2017.
- [24] K. Debnath *et al.*, "All-silicon carrier accumulation modulator based on a lateral metal-oxide-semiconductor capacitor," *Photon. Res.*, vol. 6, no. 5, pp. 373–379, May 2018.
- [25] C. Gunn and T. L. Koch, "Silicon Photonics," in *Optical Fiber Telecommunications V A*, I. P. Kaminow, T. Li, and A. E. Willner, Eds., 5th ed. Cambridge, MA, USA: Academic, 2008.
- [26] J. Shin *et al.*, "Epitaxial growth technology for optical interconnect based on bulk-Si platform," in *Proc. 11th IEEE Int. Conf. Group IV Photon. (GFP)*, Aug. 2014, pp. 3–4.
- [27] J. Fujikata, S. Takahashi, M. Takahashi, M. Noguchi, T. Nakamura, and Y. Arakawa, "High-performance MOS-capacitor-type Si optical modulator and surface-illumination-type Ge photodetector for optical interconnection," *Jpn. J. Appl. Phys.*, vol. 55, Mar. 2016, Art. no. 04EC01.
- [28] M. Ziebell *et al.*, "40 Gbit/s low-loss silicon optical modulator based on a p-pn diode," *Opt. Express*, vol. 20, no. 10, pp. 10591–10596, Apr. 2012.
- [29] X. Tu, T.-Y. Liow, J. Song, M. Yu, and G. Q. Lo, "Fabrication of low loss and high speed silicon optical modulator using doping compensation method," *Opt. Express*, vol. 19, no. 19, pp. 18029–18035, Sep. 2011.
- [30] X. Tu *et al.*, "Silicon optical modulator with shield coplanar waveguide electrodes," *Opt. Express*, vol. 22, no. 19, pp. 23724–23731, Sep. 2014.
- [31] M. R. Watts, W. A. Zortman, D. C. Trotter, R. W. Young, and A. L. Lentine, "Low-voltage, compact, depletion-mode, silicon Mach-Zehnder modulator," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 159–164, Jan./Feb. 2010.
- [32] F. Y. Gardes, D. J. Thomson, N. G. Emerson, and G. T. Reed, "40 Gb/s silicon photonics modulator for TE and TM polarizations," *Opt. Express*, vol. 19, no. 12, pp. 11804–11814, Jun. 2011.
- [33] Z. Yong *et al.*, "U-shaped PN junctions for efficient silicon Mach-Zehnder and microring modulators in the O-band," *Opt. Express*, vol. 25, no. 7, pp. 8425–8439, Apr. 2017.
- [34] M. R. Watts, W. A. Zortman, D. C. Trotter, R. W. Young, and A. L. Lentine, "Vertical junction silicon microdisk modulators and switches," *Opt. Express*, vol. 19, no. 22, pp. 21989–22003, Oct. 2011.
- [35] Y. Maegami *et al.*, "High-efficiency strip-loaded waveguide based silicon Mach-Zehnder modulator with vertical p-n junction phase shifter," *Opt. Express*, vol. 25, no. 25, pp. 31407–31416, Dec. 2017.
- [36] S. S. Azadeh *et al.*, "Epitaxially grown vertical junction phase shifters for improved modulation efficiency in silicon depletion-type modulators," *Proc. SPIE*, vol. 9516, May 2015, Art. no. 95160T.
- [37] H. Xu *et al.*, "High speed silicon Mach-Zehnder modulator based on interleaved PN junctions," *Opt. Express*, vol. 20, no. 14, pp. 15093–15099, Jun. 2012.
- [38] H. Yu *et al.*, "Performance tradeoff between lateral and interdigitated doping patterns for high speed carrier-depletion based silicon modulators," *Opt. Express*, vol. 20, no. 12, pp. 12926–12938, May 2012.
- [39] H. Xu, X. Li, X. Xiao, Z. Li, Y. Yu, and J. Yu, "Demonstration and characterization of high-speed silicon depletion-mode Mach-Zehnder modulators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, Jul./Aug. 2014, Art. no. 3400110.
- [40] T. Pinguet *et al.*, "Monolithically integrated high-speed CMOS photonic transceivers," in *Proc. 5th IEEE Int. Conf. Group IV Photon. (GFP)*, Sep. 2008, pp. 362–364.
- [41] J. M. Shainline *et al.*, "Depletion-mode polysilicon optical modulators in a bulk complementary metal-oxide semiconductor process," *Opt. Lett.*, vol. 38, no. 15, pp. 2729–2731, Aug. 2013.
- [42] M. H. Idjadi and F. Aflatouni, "Integrated Pound-Drever-Hall laser stabilization system in silicon," *Nature Commun.*, vol. 8, Oct. 2017, Art. no. 1209.
- [43] C. Sun *et al.*, "Single-chip microprocessor that communicates directly using light," *Nature*, vol. 528, no. 7583, pp. 534–538, Dec. 2015.
- [44] M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, "Free-carrier electrorefraction and electroabsorption modulation predictions for silicon over the 1–14- $\mu\text{m}$  infrared wavelength range," *IEEE Photon. J.*, vol. 3, no. 6, pp. 1171–1180, Dec. 2011.
- [45] Y. Kim, M. Takenaka, T. Osada, M. Hata, and S. Takagi, "Strain-induced enhancement of plasma dispersion effect and free-carrier absorption in SiGe optical modulators," *Sci. Rep.*, vol. 4, Apr. 2014, Art. no. 4683.
- [46] J. Fujikata, M. Noguchi, J. Han, S. Takahashi, M. Takenaka, and T. Nakamura, "Record-high modulation-efficiency depletion-type Si-based optical modulator with in-situ B doped strained SiGe layer on Si waveguide for 1.3  $\mu\text{m}$ ," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sep. 2016, pp. 361–363.
- [47] J.-H. Han, F. Boeuf, J. Fujikata, S. Takahashi, S. Takagi, and M. Takenaka, "Efficient low-loss InGaAsP/Si hybrid MOS optical modulator," *Nature Photon.*, vol. 11, pp. 486–491, Aug. 2017.
- [48] T. Hiraki *et al.*, "Heterogeneously integrated III–V/Si MOS capacitor Mach-Zehnder modulator," *Nature Photon.*, vol. 11, pp. 482–486, Aug. 2017.
- [49] J. Witzens, "Silicon photonics: Modulators make efficiency leap," *Nature Photon.*, vol. 11, pp. 459–462, Aug. 2017.
- [50] Z. Sun, A. Martinez, and F. Wang, "Optical modulators with 2D layered materials," *Nature Photon.*, vol. 10, no. 4, pp. 227–238, Apr. 2016.
- [51] N. Youngblood and M. Li, "Integration of 2D materials on a silicon photonics platform for optoelectronics applications," *Nanophotonics*, vol. 6, no. 6, pp. 1205–1218, 2016.
- [52] M. Liu *et al.*, "A graphene-based broadband optical modulator," *Nature*, vol. 474, no. 7349, pp. 64–67, May 2011.
- [53] Y. Hu *et al.*, "Broadband 10 Gb/s operation of graphene electro-absorption modulator on silicon," *Laser Photon. Rev.*, vol. 10, no. 2, pp. 307–316, 2016.
- [54] V. Soriano *et al.*, "Graphene-silicon phase modulators with gigahertz bandwidth," *Nature Photon.*, vol. 12, pp. 40–45, Jan. 2018.
- [55] S. Romero-Garcia *et al.*, "Photonic integrated circuits for multi-color laser engines," *Proc. SPIE*, vol. 10108, Mar. 2017, Art. no. 101080Z.
- [56] H. Zeng *et al.*, "Optical signature of symmetry variations and spin-valley coupling in atomically thin tungsten dichalcogenides," *Sci. Rep.*, vol. 3, Apr. 2013, Art. no. 1608.
- [57] E. P. Haglund *et al.*, "Silicon-integrated hybrid-cavity 850-nm VCSELs by adhesive bonding: Impact of bonding interface thickness on laser performance," *IEEE J. Sel. Topics Quantum Electron.*, vol. 23, no. 6, Nov./Dec. 2017, Art. no. 1700109.
- [58] S. V. Govorkov, V. I. Emel'yanov, N. I. Koroteev, G. I. Petrov, I. L. Shumay, and V. V. Yakovlev, "Inhomogeneous deformation of silicon surface layers probed by second-harmonic generation in reflection," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 6, no. 6, pp. 1117–1124, Jun. 1989.
- [59] R. S. Jacobsen *et al.*, "Strained silicon as a new electro-optic material," *Nature*, vol. 441, pp. 199–202, May 2006.
- [60] M. Cazzanelli *et al.*, "Second-harmonic generation in silicon waveguides strained by silicon nitride," *Nature Mater.*, vol. 11, pp. 148–154, Feb. 2012.
- [61] C. Schriever *et al.*, "Second-order optical nonlinearity in silicon waveguides: Inhomogeneous stress and interfaces," *Adv. Opt. Mater.*, vol. 3, no. 1, pp. 129–136, Jan. 2015.
- [62] S. S. Azadeh, F. Merget, M. P. Nezhad, and J. Witzens, "On the measurement of the Pockels effect in strained silicon," *Opt. Lett.*, vol. 40, no. 8, pp. 1877–1880, Apr. 2015.
- [63] M. Borghi *et al.*, "High-frequency electro-optic measurement of strained silicon racetrack resonators," *Opt. Lett.*, vol. 40, no. 22, pp. 5287–5290, Nov. 2015.
- [64] C. Kieninger *et al.* (Sep. 2017). "Ultra-high electro-optic activity demonstrated in a silicon-organic hybrid (SOH) modulator." [Online]. Available: <https://arxiv.org/abs/1709.06338>
- [65] V. R. Almeida, Q. Xu, C. A. Barrios, and M. Lipson, "Guiding and confining light in void nanostructure," *Opt. Lett.*, vol. 29, no. 11, pp. 1209–1211, Jun. 2004.
- [66] R. Ding *et al.*, "Demonstration of a low  $V_{\pi}$  L modulator with GHz bandwidth based on electro-optic polymer-clad silicon slot waveguides," *Opt. Express*, vol. 18, no. 15, pp. 15618–15623, Jul. 2010.
- [67] D. Jin *et al.*, "EO polymer modulators reliability study," *Proc. SPIE*, vol. 7599, Mar. 2010, Art. no. 75990H.
- [68] C. Wang, M. Zhang, B. Stern, M. Lipson, and M. Lončar, "Nanophotonic lithium niobate electro-optic modulators," *Opt. Express*, vol. 26, no. 2, pp. 1547–1555, 2018.
- [69] P. O. Weigel *et al.* (Mar. 2018). "Hybrid silicon photonic-lithium niobate electro-optic Mach-Zehnder modulator beyond 100 GHz bandwidth." [Online]. Available: <https://arxiv.org/abs/1803.10365>
- [70] C. Xiong *et al.*, "Active silicon integrated nanophotonics: Ferroelectric BaTiO<sub>3</sub> devices," *Nano Lett.*, vol. 14, no. 3, pp. 1419–1425, 2014.
- [71] T. Nishiguchi, M. Nakamura, K. Nishio, T. Isshiki, and S. Nishino, "Heteroepitaxial growth of (111) 3C-SiC on well-lattice-matched (110) Si substrates by chemical vapor deposition," *Appl. Phys. Lett.*, vol. 84, no. 16, pp. 3082–3084, Apr. 2004.
- [72] J. Witzens, B. Analui, S. Mirsaidi, T. Sadagopan, B. Welch, and A. Narasimha, "Integrated control system for laser and Mach-Zehnder interferometer," U.S. Patent 7916377B2, Nov. 3, 2008.
- [73] A. Masood *et al.*, "Comparison of heater architectures for thermal control of silicon photonic circuits," in *Proc. IEEE Conf. Group IV Photon. (GFP)*, Aug. 2013, pp. 83–84.
- [74] G.-R. Zhou *et al.*, "Effect of carrier lifetime on forward-biased silicon Mach-Zehnder modulators," *Opt. Express*, vol. 16, no. 8, pp. 5218–5226, Apr. 2008.
- [75] P. Dong *et al.*, "Thermally tunable silicon racetrack resonators with ultralow tuning power," *Opt. Express*, vol. 18, no. 19, pp. 20298–20304, Sep. 2010.
- [76] J. E. Cunningham *et al.*, "Highly-efficient thermally-tuned resonant optical filters," *Opt. Express*, vol. 18, no. 18, pp. 19055–19063, 2010.
- [77] Y. Shen, I. B. Divliansky, D. N. Basov, and S. Mookherjee, "Perfect set-and-forget alignment of silicon photonic resonators and interferometers," in *Proc. Opt. Fiber Commun. Conf. Expo. Nature Fiber Optic Eng. Conf.*, Mar. 2011, Art. no. PDP3.
- [78] S. Grillanda *et al.*, "Post-fabrication trimming of athermal silicon waveguides," *Opt. Lett.*, vol. 38, no. 24, pp. 5450–5453, Dec. 2013.
- [79] S. Spector, J. M. Knecht, and P. W. Juodawlkis, "Localized in situ cladding annealing for



- post-fabrication trimming of silicon photonic integrated circuits," *Opt. Express*, vol. 24, no. 6, pp. 5996–6003, Mar. 2016.
- [80] S. Romero-García, B. Shen, F. Merget, and J. Witzens, "Packaged MZIs passively balanced by means of multimode grating couplers," in *Proc. Photon. North Conf.*, May 2016, p. 1.
- [81] A. P. Knights, Z. Wang, D. Paez, and L. Dow, "Electrical trimming of the resonance of a silicon micro-ring resonator," in *Proc. Conf. Group IV Photon.*, Aug. 2017, pp. 29–30.
- [82] P. Dong, X. Liu, S. Chandrasekhar, L. L. Buhl, R. Aroca, and Y.-K. Chen, "Monolithic silicon photonic integrated circuits for compact 100+ Gb/s coherent optical receivers and transmitters," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, Jul./Aug. 2014, Art. no. 6100108.
- [83] D. J. Thomson *et al.*, "High performance Mach-Zehnder-based silicon optical modulators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 6, Nov./Dec. 2013, Art. no. 3400510.
- [84] S. Tanaka *et al.*, "Ultralow-power (1.59 mW/Gbps), 56-Gbps PAM4 operation of Si photonic transmitter integrating segmented PIN Mach-Zehnder modulator and 28-nm CMOS driver," *J. Lightw. Technol.*, vol. 36, no. 5, pp. 1275–1280, Mar. 1, 2018.
- [85] A. Shastri *et al.*, "Ultra-low-power single-polarization QAM-16 generation without DAC using a CMOS photonics based segmented modulator," *J. Lightw. Technol.*, vol. 33, no. 6, pp. 1255–1260, Mar. 15, 2015.
- [86] A. Maese-Novo *et al.*, "Wavelength independent multimode interference coupler," *Opt. Express*, vol. 21, no. 6, pp. 7033–7040, Mar. 2013.
- [87] K. Sohelmainen, M. Kapulainen, M. Harjanne, and T. Aalto, "Adiabatic and multimode interference couplers on silicon-on-insulator," *IEEE Photon. Technol. Lett.*, vol. 18, no. 21, pp. 2287–2289, Nov. 1, 2006.
- [88] A. Moscoso-Mártir *et al.*, "Co-integration of a temperature tolerant low impedance resonantly enhanced silicon photonics modulator," in *Proc. Int. IEEE Conf. Group IV Photon. (GFP)*, Aug. 2017, pp. 101–102.
- [89] J. Leuthold *et al.*, "Silicon-organic hybrid electro-optical devices," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 6, Nov. 2013, Art. no. 3401413.
- [90] M. Webster, K. Lakshmi Kumar, C. Appel, C. Muzio, B. Dama, and K. Shastri, "Low-power MOS-capacitor based silicon photonic modulators and CMOS drivers," in *Proc. Opt. Fiber Commun. Conf. Exhib. (OFC)*, Mar. 2015, Art. no. W4H.3.
- [91] B. Goll *et al.*, "A monolithically integrated silicon modulator with a 10 Gb/s 5 V<sub>pp</sub> or 5.6 V<sub>pp</sub> driver in 0.25 μm SiGe: C BiCMOS," *Frontiers Phys.*, vol. 2, Nov. 2014, Art. no. 62.
- [92] P. Rito *et al.*, "A 40 Gb/s 4 Vpp IQ modulator driver in 0.13 μm SiGe: C BiCMOS technology for 25 Ω Mach-Zehnder modulators," in *IEEE MTT-S Int. Microw. Symp. Dig. (IMS)*, May 2015, pp. 1–4.
- [93] A. Ghiassi, "Is there a need for on-chip photonic integration for large data warehouse switches," in *Proc. 9th Int. Conf. Group IV Photon. (GFP)*, Aug. 2012, pp. 27–29.
- [94] A. Narasimha *et al.*, "A 40-Gb/s QSFP optoelectronic transceiver in a 0.13 μm CMOS silicon-on-insulator technology," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2008, Art. no. OMK7.
- [95] N. Qi *et al.*, "A 25 Gb/s, 520 mW, 6.4 Vpp silicon-photonic Mach-Zehnder modulator with distributed driver in CMOS," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2015, Art. no. W1B.3.
- [96] D. Petousi *et al.*, "Monolithically integrated high-extinction-ratio MZM with a segmented driver in photonic BiCMOS," *IEEE Photon. Technol. Lett.*, vol. 28, no. 24, pp. 2866–2869, Dec. 15, 2016.
- [97] J. Witzens, T. Baehr-Jones, and M. Hochberg, "Design of transmission line driven slot waveguide Mach-Zehnder interferometers and application to analog optical links," *Opt. Express*, vol. 18, no. 16, pp. 16902–16928, Aug. 2010.
- [98] H. Yu and W. Bogaerts, "An equivalent circuit model of the traveling wave electrode for carrier-depletion-based silicon optical modulators," *J. Lightw. Technol.*, vol. 30, no. 11, pp. 1602–1609, Jun. 1, 2012.
- [99] F. Merget *et al.*, "Silicon photonics plasma-modulators with advanced transmission line design," *Opt. Express*, vol. 21, no. 17, pp. 19593–19607, Aug. 2013.
- [100] L. Chen, C. R. Doerr, P. Dong, and Y.-K. Chen, "Monolithic silicon chip with 10 modulator channels at 25 Gbps and 100-GHz spacing," *Opt. Express*, vol. 19, no. 26, pp. B946–B951, Dec. 2011.
- [101] R. Ding *et al.*, "High-speed silicon modulator with slow-wave electrodes and fully independent differential drive," *J. Lightw. Technol.*, vol. 32, no. 12, pp. 2240–2247, Jun. 15, 2014.
- [102] H. Xu, X. Li, X. Xiao, Z. Li, Y. Yu, and J. Yu, "Demonstration and characterization of high-speed silicon depletion-mode Mach-Zehnder modulators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, Jul./Aug. 2014, Art. no. 3400110.
- [103] M. R. Billah *et al.*, "8-channel 448 Gbit/s silicon photonic transmitter enabled by photonic wire bonding," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2017, Art. no. Th5D.6.
- [104] S. S. Azadeh, J. Müller, F. Merget, S. Romero-García, B. Shen, and J. Witzens, "Advances in silicon photonics segmented electrode Mach-Zehnder modulators and peaking enhanced resonant devices," *Proc. SPIE*, vol. 9288, Sep. 2014, Art. no. 928817.
- [105] L. Carroll *et al.*, "Photonic packaging: Transforming silicon photonic integrated circuits into photonic devices," *Appl. Sci.*, vol. 6, Dec. 2016, Art. no. 426.
- [106] X. Li *et al.*, "Highly efficient silicon michelson interferometer modulators," *IEEE Photon. Technol. Lett.*, vol. 25, no. 5, pp. 407–409, Mar. 1, 2013.
- [107] Y.-C. Chang, S. P. Roberts, B. Stern, and M. Lipson (Oct. 2017). "Resonance-free light recycling." [Online]. Available: <https://arxiv.org/abs/1710.02891>
- [108] J. Li, T. P. White, L. O'Faolain, A. Gomez-Iglesias, and T. F. Krauss, "Systematic design of flat band slow light in photonic crystal waveguides," *Opt. Express*, vol. 16, no. 9, pp. 6227–6232, Apr. 2008.
- [109] H. C. Nguyen, S. Hashimoto, M. Shinkawa, and T. Baba, "Compact and fast photonic crystal silicon optical modulators," *Opt. Express*, vol. 20, no. 20, pp. 22465–22474, Sep. 2012.
- [110] J. Li, L. O'Faolain, S. A. Schultz, and T. F. Krauss, "Low loss propagation in slow light photonic crystal waveguides at group indices up to 60," *Photon. Nanostruct., Fundam. Appl.*, vol. 10, no. 4, pp. 589–593, Oct. 2012.
- [111] Y. Terada, H. Ito, H. C. Nguyen, and T. Baba, "Theoretical and experimental investigation of low-voltage and low-loss 25-Gbps Si photonic crystal slow light Mach-Zehnder modulators with interleaved p/n junction," *Frontiers Phys.*, vol. 2, no. 2, Nov. 2014, Art. no. 61.
- [112] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometer-scale silicon electro-optic modulator," *Nature*, vol. 435, pp. 325–327, May 2005.
- [113] A. Shakoob, K. Nozaki, E. Kuramochi, K. Nishiguchi, A. Shinya, and M. Notomi, "Compact 1D-silicon photonic crystal electro-optic modulator operating with ultra-low switching voltage and energy," *Opt. Express*, vol. 22, no. 23, pp. 28623–28634, Nov. 2014.
- [114] J. Müller *et al.*, "Optical peaking enhancement in high-speed ring modulators," *Sci. Rep.*, vol. 4, Sep. 2014, Art. no. 6310.
- [115] A. Yariv, "Universal relations for coupling of optical power between microresonators and dielectric waveguides," *Electron. Lett.*, vol. 36, no. 4, pp. 321–322, Feb. 2000.
- [116] W. D. Sacher and J. K. S. Poon, "Dynamics of microring resonator modulators," *Opt. Express*, vol. 16, no. 20, pp. 15741–15753, 2008.
- [117] L. Zhang, Y. Li, J.-Y. Yang, M. Song, R. G. Beausoleil, and A. E. Willner, "Silicon-based microring resonator modulators for intensity modulation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 149–158, Jan./Feb. 2010.
- [118] Y. Liu *et al.*, "Silicon Mod-MUX-Ring transmitter with 4 channels at 40 Gb/s," *Opt. Express*, vol. 22, no. 13, pp. 16431–16438, Jun. 2014.
- [119] *Quad Small Form-Factor Pluggable (QSFP) Transceiver Specification*, Multi-Source Agreement (MSA), Small Form Factor (SFF), Dec. 2006.
- [120] S. K. Selvaraja, W. Bogaerts, P. Dumon, D. Van Thourhout, and R. Baets, "Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using CMOS fabrication technology," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 316–324, Jan./Feb. 2010.
- [121] V. Raghunathan, W. N. Ye, J. Hu, T. Izuhara, J. Michel, and L. Kimerling, "Athermal operation of silicon waveguides: Spectral, second order and footprint dependencies," *Opt. Express*, vol. 18, no. 17, pp. 17631–17639, 2010.
- [122] S. S. Djordjevic *et al.*, "CMOS-compatible, athermal silicon ring modulators clad with titanium dioxide," *Opt. Express*, vol. 21, no. 12, pp. 13958–13968, Jun. 2013.
- [123] Y. Zheng, P. Lisherness, M. Gao, J. Bovington, K.-T. Cheng, and S. Yang, "Power-efficient calibration and reconfiguration for on-chip optical communication," *J. Opt. Commun. Netw.*, vol. 4, no. 12, pp. 955–966, Dec. 2012.
- [124] A. Moscoso-Mártir *et al.*, "8-channel WDM silicon photonics transceiver with SOA and semiconductor mode-locked laser," *Opt. Express*, vol. 26, no. 19, pp. 25446–25459, Sep. 2018.
- [125] A. Leinse, M. B. J. Diemeer, A. Rousseau, and A. Driessen, "A novel high-speed polymeric EO Modulator based on a combination of a microring resonator and an MZI," *IEEE Photon. Technol. Lett.*, vol. 17, no. 10, pp. 2074–2076, Oct. 2005.
- [126] D. M. Gill *et al.*, "CMOS-compatible Si-ring-assisted Mach-Zehnder interferometer with internal bandwidth equalization," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 45–52, Jan./Feb. 2010.
- [127] A. M. Gutierrez *et al.*, "Ring-assisted Mach-Zehnder interferometer silicon modulator for enhanced performance," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 9–14, Jan. 1, 2012.
- [128] R. Li *et al.*, "High-speed low-chirp PAM-4 transmission based on push-pull silicon photonic microring modulators," *Opt. Express*, vol. 25, no. 12, pp. 13222–13229, Jun. 2017.
- [129] B.-M. Yu, J.-M. Lee, C. Mai, S. Lischke, L. Zimmermann, and W.-Y. Choi, "A monolithically integrated Si optical single-sideband modulator," in *Proc. IEEE Conf. Group IV Photon. (GFP)*, Aug. 2017, pp. 165–166.
- [130] S. Akiyama, T. Kurahashi, T. Baba, N. Hatori, T. Usuki, and T. Yamamoto, "A 1 V peak-to-peak driven 10-Gbps slow-light silicon Mach-Zehnder modulator using cascaded ring resonators," *Appl. Phys. Express*, vol. 3, no. 7, 2010, Art. no. 072202.
- [131] S. Akiyama, T. Kurahashi, K. Morito, T. Yamamoto, T. Usuki, and S. Nomura, "Cascaded-ring-resonator-loaded Mach-Zehnder modulator for enhanced modulation efficiency in wide optical bandwidth," *Opt. Express*, vol. 20, no. 15, pp. 16321–16338, Jul. 2012.
- [132] S. Romero-García *et al.*, "High-speed resonantly enhanced silicon photonics modulator with a large operating temperature range," *Opt. Lett.*, vol. 42, no. 1, pp. 81–84, Jan. 2017.
- [133] S. Romero-García *et al.*, "Passively biased resonantly enhanced silicon photonics modulator with high optical bandwidth," *Proc. SPIE*, vol. 10108, Feb. 2017, Art. no. 1010802.
- [134] S. Romero-García *et al.*, "Broadband, Temperature Tolerant and Passively Biased Resonantly Enhanced Mach-Zehnder Modulators," in *Proc. 13th Annu. IEEE Int. Conf. Nano/Micro Eng. Mol. Syst.*, 2018.
- [135] A. Melikyan *et al.*, "High-speed plasmonic phase modulators," *Nature Photon.*, vol. 8, no. 3, pp. 229–233, 2014.
- [136] M. Ayata *et al.*, "High-speed plasmonic modulator in a single metal layer," *Science*, vol. 358, pp. 630–632, Nov. 2017.
- [137] V. J. Sorger, N. D. Lanzillotti-Kimura, R.-M. Ma, and X. Zhang, "Ultra-compact silicon nanophotonic modulator with broadband response," *Nanophotonics*, vol. 1, no. 1, pp. 17–22, 2012.



## ABOUT THE AUTHORS

**Jeremy Witzens** (Senior Member, IEEE) was born in Schiltigheim, France, in 1978. He received the Engineering Diploma from the Ecole Polytechnique, Palaiseau, France, in 2000 and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, USA in 2001 and 2005.



From 2006 to 2009, he was successively a Sr., Staff, and Sr. Staff Engineer at Luxtera Inc. From 2009 to 2010, he was a Pr. Research Scientist at the University of Washington, Seattle, WA, USA. Since 2011, he has been a Professor at the

RWTH Aachen University, Aachen, Germany, where he leads a research group working on silicon photonics. He is the author of over 90 journal and conference papers and he holds 22 patents. His research interests include the application of silicon photonics to short and long reach communications as well as to sensing and instrumentation in the life sciences.

Prof. Witzens was the recipient of a research award from the European Research Council (ERC) as a Starting Independent Scientist in 2011 and was the General Chair of the IEEE Conference on Group IV Photonics in 2017 as well as of the Tenth Sino-German Joint Symposium on Opto- and Microelectronic Devices and Systems in 2018.