Non-volatile reconfigurable silicon photonics based on phase-change materials

Zhuoran Fang, Rui Chen, Jiajiu Zheng, Arka Majumdar

Abstract— The traditional ways of tuning a silicon photonic network are mainly based on the thermo-optic effect or the free carrier dispersion. The drawbacks of these methods are the volatile nature and the extremely small change in the complex refractive index ($\Delta n < 0.01$). In order to achieve low energy consumption and smaller footprint for applications such as photonic memories, optical computing, programmable gate array, and optical neural network, it is essential that the two optical states of the system exhibit high optical contrast and remain non-volatile. Phase change materials (PCMs) such as Ge$_2$Sb$_2$Te$_5$ provide an excellent solution, thanks to the drastic contrast in refractive index between two states which can be switched reversibly and in a non-volatile fashion. Here, we review the recent progress in the field of non-volatile reconfigurable silicon photonics based on PCMs. We start with a general introduction to the material properties of PCMs that have been exploited in integrated photonics and discuss their operating wavelengths. The various photonic switches that are built upon these PCMs are reviewed. Lastly, we review the recent applications of PCM-based photonic integrated circuits and discuss the potential future directions of this field.

Index Terms—phase change materials, silicon photonics, reconfigurable photonics

I. INTRODUCTION

TRADITIONAL means of tuning silicon photonic integrated circuits (PICs) primarily relies on thermo-optic or free carrier dispersion effects. Thermo-optic phase shifters built on silicon-on-insulator (SOI) platform have achieved device footprint <10µm [1] but are slow and power hungry. Phase shifters based on free carrier dispersion effect can significantly reduce the power consumption and increase the modulation speed but suffer in terms of large device length (>100 µm) as a result of small refractive index change (usually $\Delta n < 10^{-3}$) [2], [3]. To overcome these limitations, the silicon photonics community has started to head toward a hybrid approach where foreign materials are integrated on SOI waveguides to act as a tunable medium [4]. One of the promising candidates is organic electro-optic (EO) polymer which exhibits large Pockels effect and fast tuning speed [5]–[7]. However, these organic compounds have tendency to degrade at temperature >100°C, which renders them unsuitable to interface with the active electronics that heat up during the operation. Lithium Niobate (LN), an EO material which has long been used in free-space EO modulators, has recently gained traction thanks to the development of advanced dry etching techniques, which leads to extremely low-loss (0.25dB/cm$^1$) waveguides [8]. The complementary-metal-oxide-semiconductor (CMOS) compatibility of LN also makes it attractive for large-scale PIC [9]. Nevertheless, devices based on LN tend to have large footprint (>1mm) due to the minimal refractive index change as a result of small Pockels coefficient. Nano-opto-electro-mechanical (NEOM) devices [10] based on plasmonic-silicon hybrid waveguide offer attojoule switching and small footprint but are relatively slow in operation. Additionally, mechanical devices often suffer from low yield and reliability issues. Silicon-plasmonic modulators that harness the free-carrier dispersion effect of ITO [11] can support operation speed up to 1.1GHz and has device footprint down to 1.4µm but often suffer from high insertion loss. Finally, all the aforementioned tuning methods are volatile, necessitating a constant supply of the electric power. Although memristor switch based on the filamentation of amorphous silicon [12] can provide non-volatility, it is plagued by high insertion loss and has very slow switching speed.

Phase change materials (PCMs), exemplified by Ge$_2$Sb$_2$Te$_5$ (GST), can mitigate these fundamental limitations, thanks to the non-volatile phase transition and strong index modulation (typically $\Delta n > 1$). Thus PCMs can enable reconfigurable PICs for various applications including photonic switches [13], [14], photonic memory [15], programmable gate arrays [16], [17], optical computing [18], and optical neural network [19]. The suitability of PCMs for applications in non-volatile reconfigurable photonics comes from its large contrast in complex refractive index upon phase transition [20], long-term stability of multiple crystallographic phases [21], reversible switching between the amorphous and crystalline states by both optical and electrical means [21], cyclability of up to $10^5$ [22], sub-nanosecond switching speed [23], and low switching energy density down to 12aJ/mm$^2$ [22], [24]. Most importantly, PCMs can be readily deposited by standard physical vapour deposition techniques, such as thermal evaporation and sputtering, onto arbitrary substrates without the requirement for lattice matching. They can be patterned by either liftoff [14] or plasma etching [25], [26], making them CMOS compatible and hence can be scaled up to large systems. Finally, many of the

Zhuoran Fang, Rui Chen, Jiajiu Zheng, and Arka Majumdar are with the Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA.

Arka Majumdar is also with the Department of Physics, University of Washington, Seattle, WA 98195, USA (e-mail: arka@uw.edu).
PCMs, such as GST and Ag-In-Sb-Te (AIST), have been very well-characterized for electronic memory[27], [28] and for rewritable compact disks[29], [30]. Hence, these materials are compatible with high-volume manufacturing.

On the other hand, it should be clear that several limitations exist for PCMs-based reconfigurable photonics. First, the phase transition conditions can be very difficult to identify as the parameter windows for crystallization and amorphization are relatively small. For example, a pulse meant for amorphization could easily cause ablation, if the power is too high. Similarly, if the pulse width is too long, crystallization may occur instead of amorphization. Secondly, the widely quoted cyclability of $10^{10}$[22] is in fact extrapolated. The highest cyclability reported so far in experiment is in the range $10^{11}$ to $10^{12}$ [22], [31] and typical cyclability is around $10^9$[32]. This means that PCMs will not be suited for high-speed switching operation, for example a PCM-based switch that operates at 1GHz will fail in only 1 second!! Hence, for practical applications in silicon photonics, PCMs may be more justified to be used in on-chip reconfigurable light routers or reconfigurable all-purpose photonic processor, similar to the field programmable gate arrays (FPGAs) in the electronic domain, where fast tuning speed is not critical and only a limited number of reconfigurations is required. Additionally, multi-level operation of PCMs is inherently stochastic under electrical excitations[32]. In other words, the transmission level attained by multi-stage crystallization/amorphization is generally not deterministic because the atomic configuration after each melt-quench process will never be the same. As a result, the crystalline domains formed by the SET pulses will have different sizes, shapes, and distribution for each SET-RESET cycle. The implication of such phenomenon is that one continuous PCM patch on silicon waveguide cannot support deterministic multi-level operation. Instead, for $n$ number of transmission levels required, $n$ number of PCM patches could be used. The $N_i$ level, where $i$ is an integer ranging from 1 to $n$, will be expressed by the complete crystallization or amorphization of $i$ number of PCM patches. Lastly, most non-volatile reconfigurable photonic platforms demonstrated so far are based on the archetypical PCM GST which has non-negligible optical loss due to the strong bond to band absorption in the visible wavelengths and the near-IR. Therefore, GST becomes impractical for large-scale photonic integrated circuits (PICs)[33], [34] and multipurpose photonic processors[16], [17] where light is guided through numerous phase change photonic routers. While device engineering can circumvent some of the losses at 1550nm[35], [36], this comes at the expense of increased device length, and such approach does not work near visible wavelength, where the optical loss is very large. To circumvent this limitation, a few wide-bandgap PCMs has emerged which provides low-loss or even zero loss operation from the visible to mid IR[37]–[40].

In this paper, we review the recent progress in the reconfigurable silicon photonics based on PCMs. We will start with the material properties of PCMs and the selection criterion for PCMs to be used in reconfigurable integrated photonics. A few commonly used PCMs will be discussed and compared for their best operating wavelength range. The tuning mechanisms will also be discussed and compared depending on the practical applications. We then move onto reviewing the integrated photonic switches based on PCMs, including device geometry, footprint, and performance. The subsequent applications based on one or an array of these switches will be reviewed. Finally, we discuss some unsolved issues and present an outlook for this vibrant research field by suggesting potential future directions. We would like to point out that several review papers on similar topics have also been published and the readers are encouraged to explore further[41]–[45].

II. PHASE CHANGE MATERIALS FOR INTEGRATED PHOTONICS

Phase change materials (PCMs) are chalcogenides that can exhibit two states with different micro-structural properties, namely crystalline or amorphous, under room temperature. The two states can be switched reversibly on a nanosecond timescale using thermal pulses, which can be created either by electrical or optical means. More specifically, to turn a crystalline PCM into amorphous state, the PCM is heated above its liquidus temperature $T_L$ through Joule heating by a short (tens of nanosecond) voltage or laser pulse[46] [Fig. 1a]. The temperature gradient between the melted region and the surrounding materials results in a high cooling rate which quenches the materials into amorphous state. The crystallization process, albeit thermodynamically favorable, is kinetically hindered as atomic mobility is low. The amorphization process is generally termed RESET. To switch the amorphous state back to crystalline state, the PCM is held at a temperature above the glass transition temperature $T_G$ and below $T_L$ by a long (hundreds of nanoseconds) voltage or laser pulse. This temperature is maintained for a significantly long period of time to allow long-range atomic diffusion. The crystallization process is generally termed SET. The optical constants and electrical resistivity also change drastically upon phase change primarily due to the change in electronic bandgap, giving a stark contrast of properties between the two phases. For these reasons, PCMs are well-suited for memory applications and in the early days, they were primarily used as a storage medium in rewritable compact disk and later in random access memory. The selection criterion for functional PCMs therefore focused previously on low electrical resistivity - to cause sufficient Joule heating in a PCM memory cell[47] - and high optical absorption - to facilitate laser writing of compact disk[29]. Following such selection criterion, scientists discovered that the majority of functional PCMs are found on the Ge-Te-Sb Ternary phase diagram[48] [Fig. 1b]. The first group lies on the pseudobinary GeTe-Sb$_2$Te$_3$ tie line, such as GeSb$_{2}$Te$_4$ and GeSb$_{2}$Te$_5$, whereas the other two groups are Ge$_{1-x}$Sb$_x$(e.g. Ge$_{1}$Sb$_{8}$) and doped Sb:Te which were developed later to improve the switching speed[47], [49].
In recent years, the integration of PCMs with nanophotonic structures has introduced a new paradigm for reconfigurable integrated photonics[15], [18], [50], [51]. For these integrated photonic applications, however, the previous selection criterion for functional PCMs has become obsolete as low optical loss becomes crucial due to the proximity of the optical mode to the PCMs in nanophotonic waveguides. The PCMs designed for random access memory and rewritable disk, typified by GST, may no longer fulfill the task because they normally suffer from strong band to band absorption from the visible to near IR and free carrier absorption in the mid and long-wave IR. The selection criterion has to be redefined to accommodate the low optical absorption and high refractive index change as required for integrated photonic applications. This is quantified by the figures of merit (FOMs)

$\frac{\Delta n}{k_a}$ and $\frac{\Delta n}{k_c}$, where $\Delta n$ is the refractive index contrast between two states, $k_a$ and $k_c$ are the extinction coefficients in the amorphous and crystalline states, respectively. To this end, several emerging PCMs have been proposed that exhibit high FOMs in various technologically important wavelengths. Table I summarizes the FOMs of a few recently reported low-loss PCMs compared with GST at the typical quantum emitter (nitrogen vacancy centers in diamond) wavelength (633nm), the telecommunication O-band (1310nm), and the telecommunication C-band (1550nm).

The stark contrast in complex refractive index of PCMs upon phase transition is believed to arise from the drastic change in the bonding type[20]. In amorphous state, atoms tend to form covalent bonds with their closest neighbors and the number of bonds formed will depend on the number of valence electrons, following the 8-N rule. As the PCMs crystallize, they are forced into a fixed crystal structure that will require sharing of valence electrons between neighboring atoms, disobeying the 8-N rule. In the early days, it was believed that valence electrons in crystalline PCMs resonate across the neighboring bonds forming a hybrid wave function similar to the ones found in Benzene rings [Fig. 1c], hence termed ‘resonant bonding’[20], [52]. More recent studies[53], [54] suggested that the bonding is drastically different from the resonant bonding found in Benzene rings and graphite, but lies between metallic bonds and covalent bonds. In other words, the valence electrons are neither fully localized as in covalent bonding, nor fully delocalized as in metallic bonding [Fig. 1d], hence termed metavalent bonding. It is the increased electron delocalization in crystalline PCMs that gives rise to a strong electronic polarization which then results in a stark change in optical dielectric constants upon phase change.

PCMs-based photonic switches operate by optical phase modulation, absorption modulation, or both. GST at 1550nm, for example, typically operates by both mechanisms owing to the simultaneous large change in refractive index ($\Delta n \sim 2.7$) and extinction coefficient ($\Delta k \sim 1$) at this wavelength. A similar PCM GSST[37], synthesized by substituting the Te in GST by Se, is based on the same mechanism and has zero loss in the amorphous state at 1550nm, giving an enormous $\frac{\Delta n}{k_a}$. Although GSST provides much lower material absorption loss, the $\Delta n$ is
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Reduced by over 36% compared with GST which stipulates a larger device footprint and higher switching energy due to the longer PCM required and higher crystallization temperature of GSST. Hence, GSST may be more suited for large-scale PICs where the insertion loss of individual switch is critical, but GST will still be superior if device footprint and energy consumption are of higher priority than the insertion loss. Another PCM that operates based on similar mechanisms is GeTe which has very large change in refractive index in the visible and finds applications in tunable reflective display[55], [56]. However, it is less investigated in integrated photonics because of lower FOMs compared with GST or GSST near 1550nm[57], [58]. While absorption modulation is beneficial for cavity-based switches, where both the optical phase and loss change will give a pronounced extinction ratio[13], [51], it does not work for applications which require phase-only modulation, such as Mach-Zhender Interferometers (MZIs) or directional coupler-based switches. This can be addressed by two new classes of PCMs, SbS2 and SbSe3, which offer near zero loss in both amorphous and crystalline states[38], [39], [59]. SbSe3, in particular, has zero loss at both 1310nm and 1550nm and a large \( \Delta n \sim 1 \), making it an ideal candidate for non-volatile phase-only control in near IR silicon photonics[60]. SbS2, on the other hand, is a wide-bandgap PCM that has broad transparency window from 610nm to near IR in the amorphous state[38], [40]. This makes it the only candidate among the four that supports low-loss operation in the quantum emitter wavelength (633nm), where it exhibits a stark index contrast of 0.94. However, since SbS2 has non-negligible loss in the crystalline state near 633nm, some device engineering[35], [36] will be required to circumvent the insertion loss in the crystalline state.

To summarize, at 1550nm GST and GSST are ideal for applications where the loss in one state (i.e. crystalline state) can be exploited, such as microring switches[37], [51] and photonic memories[50], [50], [61]. On the other hand, SbSe3 will be preferred[59] for phase shifter type applications. At 1310nm, GST and GSST become too lossy even in the amorphous state, which makes SbSe3 and SbS2 the only candidates to work in this wavelength. Since SbSe3 has larger \( \Delta n \) than SbS2, SbSe3 will still be preferred for applications requiring phase-only control. Meanwhile, SbS2 will be the most ideal option for visible photonics as it exhibits the largest FOMs among the four at 633nm. Finally, it should be pointed out that not all chalcogenides are phase change materials even if it does exhibit two or more distinct states. For example, GeSe has a transparency window in amorphous state from 800nm to near IR and a \( \Delta n \sim 0.3 \) near 1550nm when annealed at 400°C for an hour[62] which makes it attractive for low-loss operation[63]. However, we found that such transition is irreversible as optical switching with laser pulses could not turn the material back to the amorphous state. Coincidentally, such one-way transition of GeSe was also observed in a separate work[64]. It was found that the material evolved into a more equilibrium amorphous state upon thermal annealing and the change in refractive index is caused by the formation of new bonds, without undergoing crystallization.

Finally, we discuss the common tuning methods and their pros and cons for PCM-based integrated photonics. Thermal annealing, such as rapid temperature annealing or hot plate annealing, in inert atmosphere for a few minutes will typically result in the crystallization of PCMs. This method has the advantage of switching very large area PCMs but is extremely slow and provides only one-way transition as the conditions for amorphization cannot be satisfied. Additionally, individual addressing of PCMs on a device is not possible which limits the scalability of the systems. Meanwhile, the use of free-space laser to switch PCMs can be traced back to the early days where PCMs were first developed for rewritable disks[21], [29]. Such mechanism relies on the fact that most PCMs has bandgap in the visible wavelength and hence will strongly absorb the blue or green laser light. The self-heating of the PCMs caused by optical absorption results in the crystallization or amorphization depending on the laser power and pulse duration. Free-space laser switching can be very fast (~GHz), for example femtosecond laser has been used to switch PCMs[65]. Although free-space laser switching allows local addressing of the PCM, very precise alignment and bulky optics to control the laser pulses prohibit creating a large PIC. In addition, laser switching is found to be less effective when transparent PCMs are used, such as SbS2, as the material has low absorption in the optical frequency[39]. Alternatively, the PCMs can be switched by light confined in a waveguide via evanescent coupling[15], [26], [61], which offers better scalability than free-space laser. For example, 16 \( \times \) 16 array of GST-based directional coupler switches has already been demonstrated with individual control of each switch[66]. However, increasing the compute density of selected topics in quantum electronics.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2021.3120713, IEEE Journal of Selected Topics in Quantum Electronics

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

further will be difficult as routing of light becomes extremely complicated. Similar to the switching by free-space laser, on-chip optical switching also highly depends on the PCM’s absorption. For this reason, large PCM patches (~µm) such as GST are normally difficult to be completely switched this way because light will attenuate along the PCM itself. By the time light reaches the rear end of PCM film, it will likely carry insufficient energy to reach the threshold for phase transition[18].

Apart from optical switching, PCMs can also be electrically switched either by connecting the PCMs into a circuit or using an external heater. When electrical current flows through PCM that forms part of the circuit, the Joule heating of PCM itself causes the phase to change. This switching method is based on the same mechanism that was previously developed for the phase change memory[31] so is termed “memory switching”. To enhance the PCM’s interaction with the waveguide mode, the PCM is inserted in a plasmonic nanogap between two tapered waveguides[67]. Metal electrodes are normally used to provide electrical access to the PCMs and are placed close to each other (<100nm) to reduce the resistance. This, however, induces a large ohmic loss due to the proximity of metals to the waveguide mode. Transparent conductors such as Indium Tin Oxide (ITO), Fluorine-doped Tin Oxide (FTO) and graphene can be used as electrodes to reduce the insertion loss. The fundamental limitation with memory switching is that only very small volume of the PCM can be switched due to an effect called filamentation[68]. The modification to optical mode from phase change is hence minimal – only 0.5% change in transmission was reported in [67] - so will unlikely be useful for any practical applications such as optical computing or photonic memories. A vertical platform[61]. Such on-chip optical switching approach has also been demonstrated on SOI platform[76] with smaller device.

III. WAVEGUIDE INTEGRATED SWITCHES BASED ON PCMs

In this section, we will review the waveguide integrated switches based on PCMs. This can be further subdivided into three categories of switches – waveguide based, MZI, and directional coupler. Waveguide switches typically rely on the change in extinction coefficient of PCM to modulate the light transmission through a waveguide. Rios et al. first demonstrated that multiple transmission levels (eight levels) can be achieved by switching a GST patch on Si3N4 waveguides near 1550nm[15]. The switching happens via optically pumping the GST on waveguide by evanescent coupling, as shown in Fig. 2a top. Wavelength-selective operation is also demonstrated at three different wavelengths using microrings of different radii. Such multi-level photonic memory framework has been significantly improved recently, achieving 34 transmission levels or 5-bit in a single memory cell on Si3N4 platform[61]. Such on-chip optical switching approach has also been demonstrated on SOI platform[76] with smaller device.
footprint and the faster switching speed. Although free-space laser heating has been slowly superseded by on-chip optical switching, a recent work has used free-space laser writing to show phase-only control of the silicon PICs with low-loss PCM $\text{Sb}_2\text{Se}_3$ on waveguides\cite{59}. An MZI is used to demonstrate the large phase shift (over $10\pi$) induced by laser annealing the $\text{Sb}_2\text{Se}_3$ with no change in the extinction ratio and insertion loss. Over 600 cycles have been demonstrated with extinction ratio of 1.5dB \cite[see Fig. 2b]{15}. Optical switching can also combine electronic readout by connecting the PCM – in this case GeTe nanowires - into the circuit which allows mixed-mode operation\cite[57]{57}, but electrical addressing is not demonstrated \cite[Fig. 2c]{71}. A recent work has shown that both mixed-mode programming and mixed-mode readout can be realized on the same device by inserting GST into a plasmonic gap \cite{67}. However, only very small volume of GST can be switched, so the actual thickness of the GST was not reported, and no mode simulation was shown to corroborate such claim.

ITO heater is also explored in a similar work where an MZI design is used to obtain a larger modulation depth (~7dB)\cite{71}. Multi-level crystallization was observed but reversible switching in optical domain was not reported. Poor cyclability of ITO, particularly amorphization, is common in both works which points to potential issue with sputtered ITO material. The surface roughness and abrupt resistivity increase of ITO at 650K could lead to regional thermal hot spots that cause nonuniform heating of the GST\cite{77} and hence only small regions of GST near the hot spots are amorphized.

To tune large areas of PCMs, external heaters must be used. Kato et al. first proposed to use transparent ITO heater on SOI waveguide to trigger the phase change of GST\cite{72}. Although more than 10 cycles of switching were observed by measuring the GST resistance, cyclability in the optical transmission was poor (one cycle), the crystallization is very slow (100ms), and the optical contrast was minimal (~0.6%), as shown in Fig. 2d. Further improvement of the optical contrast requires the complete crystallization and amorphization of large area of PCMs.

To tune large areas of PCMs, external heaters must be used. Kato et al. first proposed to use transparent ITO heater on SOI waveguide to trigger the phase change of GST\cite{72}. Although more than 10 cycles of switching were observed by measuring the GST resistance, cyclability in the optical transmission was poor (one cycle), the crystallization is very slow (100ms), and the optical contrast was minimal (~0.6%), as shown in Fig. 2d. Further improvement of the optical contrast requires the complete crystallization and amorphization of large area of PCMs.

To tune large areas of PCMs, external heaters must be used. Kato et al. first proposed to use transparent ITO heater on SOI waveguide to trigger the phase change of GST\cite{72}. Although more than 10 cycles of switching were observed by measuring the GST resistance, cyclability in the optical transmission was poor (one cycle), the crystallization is very slow (100ms), and the optical contrast was minimal (~0.6%), as shown in Fig. 2d. Further improvement of the optical contrast requires the complete crystallization and amorphization of large area of PCMs.

Figure 2. (a) Switching PCMs on $\text{Si}_3\text{N}_4$ waveguide via evanescent coupling. Multiwavelength write and erase using microring resonators of different radii (bottom). (b) Optical switching of low-loss $\text{Sb}_2\text{Se}_3$ integrated with an MZI switch. Multilevel operation and high cyclability are demonstrated with near-zero loss. (c) Mixed mode operation of GeTe nanowire on $\text{Si}_3\text{N}_4$ waveguide with both optical and electronic readouts. (d) Plasmonic nanogap enhanced photonic switch based on GST that can be addressed and readout both electrically and optically. Figure (a) is reprinted from Ref. [15]. Figure (b) is reprinted from Ref. [59]. Figure (c) is reprinted from Ref. [57]. Figure (d) is reprinted from Ref. [67].
has been proposed to function as heaters that provide high cyclability (>1000 cycles), fast switching speed (~MHz), and large on-off contrast (~18.6dB or \( \Delta T / T_0 = 300\% \)), and foundry/CMOS compatibility[14], [69], [70]. Although the size of the GST used in [70] is comparable to that in [72], the contrast is over 15 times larger. This is because the silicon heater completely switches the GST while the ITO heater only partially does so. 4-level of transmission can also be achieved with partial crystallization from low-energy pulses, as shown Fig. 3(b). The advantages of doped silicon heater arise from the fact that it has stable resistance up to current saturation, which could go up to >45mA [70], smooth surface of crystalline silicon, and uniform conductivity across the doping region. However, uniformly doping the entire waveguide will generally lead to high insertion loss, for example 0.5dB/\( \mu \)m insertion loss from doping was reported in[69]. To avoid the excessive loss from free carrier absorption, the silicon can be doped into a PIN junction[14], where the waveguiding happens in the intrinsic region [Fig. 3c]. In this way, the insertion loss due to doping can be reduced to 0.007dB/\( \mu \)m. Doped silicon heater is particularly well-suited for transparent PCMs which have wide bandgaps[38], [39] do not tend to absorb in the near IR wavelengths. On-chip optical switching will be difficult in this regard. For example, recently an electrically reconfigurable phase shifter has been demonstrated based on doped Si heater and low-loss PCMs Sb:Se[60] that implements applications from 2 × 2 switch to optical trimming. The use of dense atomic layer deposited (ALD) Al2O3 capping also improves the device cyclability to over 3000 cycles[14], [26]. Despite the great progress of doped silicon and ITO heaters in switching PCMs, they still suffer from slow switching speed (~10MHz) and high switching energy (~nJ). A recent work has proposed to use graphene as the transparent conductor for high speed (~80MHz) and energy efficient (19.2aJ/nm\(^3\) and 6.6aJ/nm\(^3\)) non-volatile photonic switch[73], as shown in Fig. 3d. The advantage of graphene as transparent heater comes from the ultra-low heat capacity and high in-plane thermal conductivity[79], [80]. PCM tuning based on graphene heater was later experimentally demonstrated[74], although not integrated with photonic structures. The switching speed is relatively slow and energy consumption is high due to the unoptimized heater design and high phase transition threshold for GSST[37]. Finally, a slot waveguide design has been proposed recently to maximize the field overlap between the waveguide mode and the GST[81]. The field confinement in the GST-filled nanogap leads to a gigantic \( \Delta n_{eff} \sim 1.2 \) and \( \Delta \alpha \sim 35 \) dB/\( \mu \)m of the hybrid waveguide when the GST phase changes.

IV. INTEGRATED CAVITY-ENHANCED SWITCHES BASED ON PCMS

Integrated photonic cavities, such as microring resonators and photonic crystal nanobeams can enhance the light-matter interaction via temporal and spatial mode confinement. Due to the increased light-matter interaction, the optical phase and amplitude changes are amplified even when the material index changes are small. PCMs are particularly effective in modulating light when integrated with these photonic cavities.
both the optical phase and absorption modulation effect of the PCMs contribute to high extinction ratio when integrated with a cavity. In addition, for loss-loss PCMs such as Sb$_2$S$_3$ and Sb$_2$Se$_3$ that have near zero loss in both states, photonic cavities or MZIs are necessary to reveal any change in optical transmission since only the real part of the PCM's refractive index changes. Rudé et al. first demonstrated the tuning of a silicon racetrack resonator by optically switching a GST patch capped on the ring using a pulsed laser[82]. Multi-stage crystallization and amorphization were shown with an extinction ratio of 12dB [Fig. 4a], significantly higher than what most waveguide switches can achieve. The broadening of the resonance linewidth when the GST is crystallized indicates the increase in loss inside the microring as the microring moves into the undercoupling regime[83]. The red shift of the spectrum is due to the increase in the real part of the GST refractive index. The transient response of the microring transmission was also investigated with a fast photodiode, revealing the thermo-optic effect of the GST [Fig. 4a bottom] under optical excitation. However, a detailed analysis of optical loss induced by the GST is still lacking. Rios et al. integrated GST on both MZIs and microring resonators on Si$_3$N$_4$ platform and extract the loss per unit width of GST near 1550nm[84], see Fig. 4c. They found the amorphous state attenuation coefficient is 0.1dB/µm and in crystalline state the loss is around 1.1dB/µm. No reversible switching was demonstrated in this work because the GST phase transition is actuated by thermal annealing and the extinction ratio of the microring is relatively small (~15dB) since the coupling conditions of the microring is not optimized. A recent work[13] has demonstrated that by sweeping the length of GST capped on the silicon microrings, the critical coupling condition can be identified and hence the maximum optical contrast can be extracted, see Fig. 4b. Both the spectral shift and loss per unit length of GST near 1550nm were calculated. Multi-level switching by laser pulses was also observed with a gigantic extinction ratio of ~33dB thanks to the critical coupling. A non-volatile microring switch has also been demonstrated with the broadband transparent PCM GSST[37] by laser switching. The large optical absorption of GSST in the crystalline state causes the complete destruction of the high Q resonance, giving rise to a large extinction ratio of over 40dB [Fig. 4d].

Due to the difficulty in alignment and diffraction-limited resolution of laser switching, the PCM community is gradually moving towards on-chip actuation of the PCMs for high density integration and scalability. Microring resonators are particularly well-suited for on-chip optical switching of PCMs thanks to the large intensity build up inside the microrings. Although microrings have been used to achieve multiwavelength operation of photonic memory, they are explored in more details in[51] where multi-level operation and cyclability of more than 1000 cycles have been demonstrated in a 1 × 2 microring switch [Fig. 5a]. However, the insertion loss is high ~5dB and the extinction ratio is only around 6dB. A low-loss (insertion loss<1dB) 1 × 2 microring switch[26] was demonstrated recently based on nano-patterned GST disks encapsulated by ALD Al$_2$O$_3$ with extinction ratio over 20dB. Such low loss and high transmission contrast are made possible...
by the conformal and dense ALD Alumina capping that prevents the deformation and reflowing of the molten GST, as shown in Fig. 5b.

Although on-chip optical switching provides better scalability than free-space laser, it is still not desirable for high-density integration due to the increased complexity in optical routing. Furthermore, on-chip optical switching cannot switch wide bandgap PCMs that have zero absorption in the near IR.[37]–[39]. To overcome this limitation, Zheng et al.[14] demonstrated a non-volatile silicon microring switch based on a PIN heater to actuate the GST patterned on the microring. High extinction ratio of over 15dB is achieved with a cyclability over 3000 cycles [Fig. 5c]. Based on a similar design, transparent PCMs Sb$_2$Se$_3$ was recently integrated with silicon microheaters to build an electrically reconfigurable phase shifter in Si microrings[60]. A microring switch based on another wide bandgap PCM Sb$_2$S$_3$[40] was also demonstrated using an ITO integrated heater with insertion loss of 0.48dB. Figure (a) is reprinted from [51]. Figure (b) is reprinted from [26]. Figure (c) is reprinted from Ref. [14]. Figure (d) is reprinted from Ref. [40].

V. APPLICATIONS OF PCM IN INTEGRATED PHOTONICS

In this section, we will review the applications enabled by the PCM-based integrated photonics, including photonic memory, neuromorphic computing, optical neural network, convolutional processing, and optical routing. Finally, we will suggest a few prospective applications that have not yet been extensively studied.

A. Integrated photonic memory.

The huge amount of data generated nowadays drives a growing demand for large-scale data storage. Current digital computer paradigm based on the Von Neumann architecture consists of memory, central processing unit, and I/O. The control unit for command instruction and arithmetic unit for logic operation work in series which causes the well-known Von Neumann bottleneck. More specifically, Von Neumann bottleneck states that the amount of data accessible to the CPU is always less than the data stored in the memory. This bottleneck has since motivated the pursuit of on-chip optical interconnect for high-performance computing. However, the on-chip optical interconnect has practical limitations due to the need for optic-electronic or electronic-optic signal transduction. This calls for a new paradigm, where data retrieval from the storage and logic operations can be performed in parallel.

PCMs can offer a solution by allowing optical computation and storage taking place at the same time. In 2012, Pernice and Bhaskaran[86] proposed an all-optical memory on Si$_3$N$_4$
platform with GST [Fig. 6a]. The proposed device consists of a
microring resonator coupled with add, drop, and control
waveguides with a GST thin film deposited on a small segment
of the ring. The state of the GST can be changed between
amorphous, partially crystalline, and fully crystalline states by
sending high intensity optical control signals in the control
waveguide. The microring is designed to be nearly critically
coupled when GST is in the amorphous state, where high
intensity can be detected at the drop port. Since GST has large
real and imaginary refractive index contrasts, both the coupling
and resonance conditions are changed once phase change
occurs. The transmission from through port can be varied
between level “0” at amorphous state and level “1” at crystalline
state with a 10 dB contrast. The design was soon experimentally
realized by Rios et al.[84], as shown in Fig. 4c. In this work, the
loss of both the amorphous and crystalline states was evaluated
with balanced MZIs and ring resonators, and the spectral
differences between the two states were clearly shown. Multi-
bit optical memory was then achieved using on-chip optical
switching[15], leveraging the freely accessible partially
crystalline states of the GST. The “Write” (amorphization) step
is initiated by a single 100 ns pulse while “Erase”
(crystallization) operation is achieved by six consecutive 100 ns
pulses with decreasing power. It is worth noting that full
crystallization with a single optical pulse is generally hard in
practice, since not all parts of the GST thin film crystallized
simultaneously. Pumping with a sequence of pulses can
overcome this challenge by gradually heating up the whole thin
film and the decreasing power of each pulse prevents
reamorphization of already crystallized parts. Wavelength
division multiplexing (WDM) was also demonstrated: GST on
three microring resonators with different resonance
wavelengths can be set or reset by different wavelengths of light
[Fig. 2a]. Leveraging both multi-level operation and WDM,
Feldmann et al. continue to demonstrate an all-optical 256-cell
photonic phase-change memory with a capacity of 512 bits[87].
In contrast to the previous technique where a pulse train of
decreasing amplitude is used for “Erase” operation, pulse-width

Figure 6. Integrated photonic memories. (a) A theoretical proposal of an all-optical memory on SiN₄ platform with GST. (b) Pulse-width modulation for the access of arbitrary memory state. (c) 5-bit photonic memory realized by the dual-step single-pulse scheme. Figure (a) is reprinted from Ref. [86]. Figure (b) is reprinted from Ref. [88]. Figure (c) is reprinted from Ref. [61].
modulation (PWM) enables access of any arbitrary intermediate state regardless of the current state by using the same pulse amplitude but different pulse durations. A PWM twelve-pulse train with pulse period 30 ns, pulse duration from 5 to 20 ns and peak power 1.4 mW was used to achieve 12 distinguishable levels. Each level can be achieved with a specific pulse duration. PWM was also used for logic operations, including OR and NAND gates. Li et al. developed a dual-step single-pulse programming method to partially crystallize GST to an arbitrary intermediate state and showed 34-level (~5-bit) optical memory. A single dual-step pulse with programmable pulse width and energy was used for both amorphization and crystallization. To access arbitrary intermediate state, a dual-step pulse was used with a 50 ns high amplitude pulse which erases the current state, followed by a low amplitude pulse to determine the portion of amorphization. It was shown that such technique can improve the operation speed drastically from tens of kHz to a few MHz.

B. PCM-based optical information processing.

In this section we will review several works that exploit the above-mentioned photonic memories to implement system level applications such as all-optical abacus, logic gates, photonic synapses, and optical neural networks.

1) Optical arithmetic, vector-matrix multiplication, and convolutional neural networks

Feldmann et al. showed a chip-scale all-optical abacus using an array of photonic memory cells. As shown in Fig. 7a, multiple levels of transmission are mapped to digits 0 (amorphous state) and 10 (fully crystalline state). Addition of one is mapped to stepwise crystallization, which is implemented with identical pulses, each with 12 pJ power and picosecond duration. After the PCM reaches its fully crystalline state, the device is reset using ten 19 pJ pulses. In analogy to the well-known calculator abacus, each crystallization step is effectively shifting one bead to the right. The authors also showed that by using two-pulse pump scheme with PCMs on the waveguide crossing and matching arrival time of the two pulses, highly accurate random access in a 3 × 3 array can be achieved.

Inspired by the in-memory computing in the electronic domain, researchers have also used the PCM-loaded waveguides for optical computation. Both scalar-scalar and vector-matrix multiplication (VMM) were demonstrated using in-memory photonic computing, as shown in Fig. 7b. Double-step pulses, as opposed to the previously used consecutive pulses with reduced amplitude, are used for “Erase” operation to improve the speed. Moreover, the pulse width of the “Write”/ “Erase” pulses is optimized to obtain a reasonable trade-off between transmission contrast and
1) Photonic neuromorphic computing

The realization of brain-inspired computing generally requires the emulation of neural dynamics, which involves the maintenance of equilibrium potential and the ‘spiking’ of the membrane potential upon multiple excitations. PCMs become ideal for mimicking such dynamics due to their non-volatile nature and the thresholding behavior. This led to electronic emulation of the synaptic events using Phase change memory [91].

Similarly, the synaptic activity can also be mimicked in the photonic domain using the PCM-loaded waveguides. Cheng et al.[92] first demonstrated an on-chip photonic synapse with discrete GST islands on a tapered Si₃N₄ waveguide. Compared to a rectangular waveguide with a single GST patch on top, the new design is shown to have more controlled absorption and less resonance effects. The photonic synapse is firstly set to fully crystalline state by annealing and defined as level “0” with transmission T₀. By sending in single optical pulses with duration of 50 ns, the relative transmission ΔT/T₀ is changed by 7% [Fig. 8a(ii)], which is set as level “3”. Sending in 50, 100 and 1000 optical pulses with the same parameters would shift the level to “2”, “1” and “0”, respectively. An exponential relation is found between ΔT/T₀ and number of pulses applied [Fig. 8a(iii)], which can be used to implement the all-optical spike-timing-dependent plasticity (STDP). The implementation of all-optical STDP was theoretically proposed as well [Fig. 8a(iv)].

By connecting multiple synaptic cells in series, Feldmann et al.[19] demonstrated an all-optical spiking neurosynaptic network with self-learning capabilities [Fig. 8b]. As shown in Fig. 8b(i), each photonic neuron consists of several weighting

Figure 8. Photonic neuromorphic computing. (a) An on-chip integrated photonic synapse. (b) All-optical spiking neurosynaptic networks with self-learning capabilities. Figure (a) is reprinted from Ref. [92]. Figure (b) is reprinted from Ref. [19].
input ports, where the weights are set by the PCM on the waveguides. Optical pulses with different wavelengths are input into each port. After passing the ring MUX, they are coupled into the bus waveguide and acted on the larger ring at the end of the bus waveguide. By setting the weights in the input, the state of PCM on the larger ring can be varied, which operates as an activation function. The result can be read out by probing the larger ring’s transmission. With this idea, both supervised and unsupervised learning are shown in a 2×2 pattern recognition task [Fig. 8b(ii)]. The unsupervised learning is accomplished by adding a feedback path from the activation output to the weighting cells. The weighting cells with both the input and feedback pulse overlap in time are enhanced, while others are depressed. The proposed spiking neural networks are scalable by stacking multiple neurons to form a layer and can be extended to multiple layers with WDM [Fig. 8b(iv)]. A larger scale single-layer spiking neural network with 4 neurons was used to accomplish a pattern recognition task with 3×5 cells [Fig. 8b(iii)].

C. Optical routing

On-chip optical routing is essential for large-scale optical interconnects[93] used in data center and photonic processors for signal processing[16], [17] and computing[33], [66]. PCMs based photonic switches can provide a solution towards a low-loss, compact, and energy efficient optical routers for optical interconnects that consume zero energy at the static state. Despite the advantages, PCMs exhibiting a large change in refractive index also often suffer from strong optical absorption, especially in the crystalline state, as shown in Table 1. This unavoidably leads to high insertion loss of the photonic switch in the crystalline state. Such insertion loss, typically around a few dBs, is manageable for a single device but becomes significant in a large network consisting of numerous switches. To circumvent this high loss, a three-waveguide geometry design has been recently proposed[35], [36]. The structure consists of a directional coupler with a third PCM-clad Si waveguide sandwiched between two Si waveguides [Fig. 9c]. The waveguide widths are designed such that phase matching conditions can be satisfied between the amorphous PCM-Si waveguide and the bare Si waveguide. Light then couples from the input waveguide to the central waveguide and then outcouples into the cross port. When the PCM is changed into the crystalline state, the stark index contrast significantly modifies the effective index of the central hybrid waveguide and breaks the phase matching condition. Light in the input port then exits from the same waveguide, i.e., the through port as if the central waveguide does not exist. Since the light does not propagate through the crystalline PCM, the insertion loss is minimized. The three-waveguide geometry thus realizes a compact 2×2 switch. Meanwhile, removing the central waveguide and placing the PCM on the adjacent waveguide will result in a 1×2 switch [Fig. 9a]. The design was experimentally verified using GST [Fig. 9b and 9d] which results in an ultracompact (~30µm), low insertion loss (~1dB), and broadband (over 30nm) directional coupler switch in the telecommunication wavelength[36]. A similar design has also been proposed based on the broadband transparent GSST[35], leveraging the near zero loss (amorphous state) and large index change of GSST at 1550nm.

Figure 9. Low-loss broadband directional coupler switches based on GST. (a) Design of the 1×2 directional coupler switch based on GST-on-SOI hybrid waveguide. (b) Experimental optical transmission of the 1×2 directional coupler switch in amorphous and crystalline state of GST. (c) Design of the 2×2 directional coupler switch based on GST-on-SOI hybrid waveguide. (d) Experimental optical transmission of the 2×2 directional coupler switch in amorphous and crystalline state of GST. Figures (a), (b), (c), and (d) are reprinted from Ref. [36].
A different approach is to employ transparent PCMs and bypass any material loss. An electrically reconfigurable 2×2 switch based on the low-loss PCM Sb$_2$Se$_3$[60] is shown in Fig. 10a. Light can be diverted into either port (bar or cross) by switching the Sb$_2$Se$_3$ on either arm of the MZI. A very compact footprint of 6µm long Sb$_2$Se$_3$ is used to achieve π/2 phase shift with a low insertion loss of 0.03dB/µm. Similar device function is also implemented by laser writing Sb$_2$Se$_3$ on Si waveguides[59], 1 × 2 switch is realized using an inverse-designed silicon MMI splitter capped with 23nm of Sb$_2$Se$_3$ [Fig. 10b]. The crystalline PCM film is selectively amorphized by laser pulses to write an index distribution onto PCM film. It is the distribution of this index perturbation that allows the routing of light to one or both ports. The beam splitter achieves a small footprint of 6µm × 33µm, large extinction ratio of 8dB, and an insertion loss of 0.5dB.

D. Prospective: Multipurpose programmable photonic processor based on PCMs

Traditional, PICs are fabricated with fixed functionality, similar to the electronic application specific integrated circuits (ASICs). Although the PICs are reliable for completing one specific task, they are not flexible when new functionality is needed, in which case a completely new chip has to be fabricated. Hence, a multipurpose programmable photonic processor, the optical analogue of electronic field programmable gate array (FPGA), is highly desired[4]. Such processor can be programmed to implement different tasks such as optical filters, signal photon routing, and optical signal processing. It is well-know that the FPGA has numerous advantages compared to ASICs such as simpler design flow, faster time to market, higher application versatility[94]. Hence, we will expect the same benefits in the photonic domain.

Zhuang et al.[95] demonstrated a photonic signal processor using a grid of thermally tunable MZIs in a two-dimensional Si$_3$N$_4$ mesh network. They implemented a radiofrequency filter with only two square unit cells, limited by the relatively large Si$_3$N$_4$ phase shifters. Pérez et al.[16] implemented a multipurpose signal processing core with hexagonal unit cells, where they demonstrated 20 different functionalities with 7 hexagonal unit cells. Zhang and Yao[17] demonstrated similar functionality using a micro-disk array. However, most programmable photonics are based on thermo-optic phase shifters. These phase shifters are inherently large because of the relatively small TO coefficients of Si and Si$_3$N$_4$ and the thermal crosstalk between the devices. They are also power hungry due to the constant current flow and volatile nature of the TO effect. Fig. 11d shows the schematic of a nonvolatile programmable photonic platform based on PCM-loaded 2 ×2 switches that can overcome these limitations. Thanks to the large refractive index change of PCMs, the size of single photonic switch can be significantly reduced, and more devices can be integrated in the same chip area. Secondly, the static power required to hold the state of the whole system is zero due to the nonvolatile nature of PCMs. Thus, the power efficiency can be improved dramatically. Finally, we would like to comment on the thermal crosstalk as switch density increases. The thermal crosstalk will unlikely affect the PCM state in adjacent switches because (1) the heating is very localized and (2) the PCMs required a high temperature threshold for switching. Heater transfer simulation[14] shows that the temperature of the silicon slab drops below the crystallization temperature of GST ~1.5 µm away from the central heating region during amorphization. The current will only flow in the doped regions, causing Joule heating, while the adjacent heaters are separated by tens of microns to leave room for metal contacts. Therefore, it is unlikely that the thermal crosstalk will cause GST phase transition in adjacent switches.

E. Prospective: Self-reconfiguring network

Another related direction could be on-chip self-reconfigurable optical network[96]. Such a network can perform any linear operation[97] or couple between any input...
and output. A large body of theoretical research has been performed on such a network. Recently, experimental demonstrations have also showed automatic descrambling of multi-mode signal[98] and arbitrary manipulation of free space beams[99]. However, a key component in this type of self-reconfiguring network is a static phase-shifter, i.e., a phase-shifter that does not consume any energy to hold the state. Such a capability is not possible right now, and most experimental efforts rely on thermo-optic shifters. Low insertion loss PCM-based switches can significantly reduce the energy consumption in these self-configuring optical networks.

VI. OUTLOOK
After reviewing the different types of integrated photonic switches based on PCMs and their applications, we would like to provide an outlook to the field and suggest some future directions. Firstly, on-chip electrical switching and the search for low-loss PCMs and heater geometries will continue to be the forefront of this research, as driven by the need for scalable PCM-based PICs. As the PCM-based photonic arrays become larger, the insertion loss of individual switch adds up and eventually the loss of the whole system will be intolerable. GST-based microring switches, for example, generally have insertion loss of $\geq 1\text{dB}$. Even with optimized device design and state-of-the-art fabrication, the insertion loss will still be limited by the intrinsic material absorption which is around 0.1dB/µm for amorphous GST and 1dB/µm for crystalline GST. Hence, low-loss PCMs will potentially bring the insertion loss of PCM-based photonic switches down below 0.1dB which is necessary for large-scale PICs. With the discovery of low-loss PCMs in both near IR[38], [39] and MIR[37], a transparent PCMs in both amorphous and crystalline states across the visible range will be an essential topic to pursue, which may open up exciting applications in integrated quantum optics and non-volatile display. On the other hand, on-chip electrical switching will eventually supersede optical switching as routing of light becomes increasingly complicated in a large network. On-chip electrical switching also has the added benefit of compatibility with both transparent and lossy PCMs. Although low-loss, energy efficient, and high endurance electrically reconfigurable photonics switches have been demonstrated using doped silicon heaters[14], [70], so far there is little success in functional photonic switches based on transparent conductors, such as transparent conducting oxides and graphene. The switches are generally plagued by high insertion loss, slow switching speed, small modulation depth, and have limited number of cyclability[71], [72], [74]. However, transparent conductors are crucial for realizing PCM-based switches on Si$_3$N$_4$ platforms which have wide applications in integrated quantum photonics, due to its wide bandgap, and waveguide-based LiDAR, due to its low Kerr nonlinearity. Therefore, an electrically tunable, PCM-based, Si$_3$N$_4$ photonic switch that has low insertion loss, fast switching speed, high modulation depth, and high cyclability will be highly desirable. Once a Si$_3$N$_4$ compatible, electrically reconfigurable photonic switch is realized, we can envision that a non-volatile general-purpose programmable integrated photonic processor can be built that

Figure 11. Multipurpose programmable photonic processors. (a) Photonic signal processor for microwave photonic applications. Metallic heaters are used to control the SiN phase shifters. (b) A microdisk resonator arrays for signal processing. The size of the whole system is reduced to 0.4×0.4 mm$^2$. Metallic heaters are used to control the Si micro-disk resonators. (c) A multipurpose programmable processor based on hexagonal grids. Schematic of the proposed photonic processor; each grid is a hexagon with side length of around with a total of 7 grides in one chip. Metallic heaters are used to control the SiN phase shifters. (d) A proposed multipurpose photonic processor based on GST-loaded tunable beam splitter. Figure (a) is reprinted from Ref. [95]. Figure (b) is reprinted from Ref. [17]. Figure (c) is reprinted from Ref. [16].
cover wide operation range from visible to MIR. The processor can be used to perform various tasks including on-chip light routing, optical signal processing, photonic quantum computing, and optical neural network.

ACKNOWLEDGMENT

The research is funded by National Science Foundation (NSF-1640986, NSF-2003509), ONR-YIP Award, DRAER Labs and Intel. Part of this work was conducted at the Washington Nanofabrication Facility / Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site at the University of Washington, which is supported in part by funds from the National Science Foundation (awards NNCI-1542101, 1337840 and 0335765), the National Institutes of Health, the Molecular Engineering & Sciences Institute, the Clean Energy Institute, the Washington Research Foundation, the M. J. Murdock Charitable Trust, Altatech, ClassOne Technology, GCE Market, Google, and SPTS.

REFERENCES


This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2021.3120713, IEEE Journal of Selected Topics in Quantum Electronics


