POLITECNICO DI TORINO
Repository ISTITUZIONALE

System-Level Simulation for Integrated Phase-Change Photonics

(Article begins on next page)

26 November 2024

System-Level Simulation for Integrated Phase-Change Photonics

Santiago García-Cuevas Carrillo, Alessio Lugnan [ID], *Graduate Student Member, IEEE*, Emanuele Gemo [ID], Peter Bienstman [ID], *Member, IEEE*, Wolfram H. P. Pernice [ID], *Member, IEEE*, Harish Bhaskaran [ID], *Senior Member, IEEE*, and C. David Wright [ID]

Abstract—Conventional computing systems are limited in performance by the well-known von Neumann bottleneck, arising from the physical separation of processor and memory units. The use of electrical signals in such systems also limits computing speeds and introduces significant energy losses. There is thus a pressing need for unconventional computing approaches, ones that can exploit the high bandwidths/speeds and low losses intrinsic to photonics. A promising platform for such a purpose is that offered by integrated phase-change photonics. Here, chalcogenide phase-change materials are incorporated into standard integrated photonics devices to deliver wide-ranging computational functionality, including non-volatile memory and fast, low-energy arithmetic and neuromorphic processing. We report the development of a compact behavioral model for integrated phase-change photonic devices, one which is fast enough to allow system level simulations to be run in a reasonable timescale with basic computing resources, while also being accurate enough to capture the key operating characteristics of real devices. Moreover, our model is readily incorporated with commercially available simulation software for photonic integrated circuits, thereby enabling the design, simulation and optimization of large-scale phase-change photonics systems. We demonstrate such capabilities by exploring the optimization and simulation of the operating characteristics of two important phase-change photonic systems recently reported, namely a spiking neural network system and a matrix-vector photonic crossbar array (photonic tensor core). Results show that use of our behavioral model can significantly facilitate the design and optimization at the system level, as well as expediting exploration of the capabilities of novel phase-change computing architectures.

Index Terms—Integrated photonics, neuromorphic computing, phase change materials, photonic tensor core.

Santiago García-Cuevas Carrillo, Emanuele Gemo, and C. David Wright are with the Department of Engineering, University of Exeter, EX4 4QF, Exeter, U.K. (e-mail: s.garcia-cuevas-carrillo@exeter.ac.uk; eg402@exeter.ac.uk; david.wright@exeter.ac.uk).

Alessio Lugnan and Peter Bienstman are with the Ghent University/imec, 9052 Gent, Belgium (e-mail: alessio.lugnan@ugent.be; peter.bienstman@ugent.be).

Wolfram H. P. Pernice is with the Institute of Physics, University of Münster, 48149 Münster, Germany (e-mail: wolfram.pernice@uni-muenster.de).

Harish Bhaskaran is with the Department of Materials, University of Oxford, OX1 2JD Oxford, U.K. (e-mail: harish.bhaskaran@materials.ox.ac.uk).

I. INTRODUCTION

THE von Neumann architecture is one of the pillars underpinning the progress of computing technologies over many decades. In essence, this architecture is based on a central processing unit (CPU) which carries out the computations and a separated memory that stores data and instructions. The physical separation of processing and memory gives rise to the well-known von Neumann bottleneck [1], which limits the performance of the overall computing system. The consequences of the von Neumann bottleneck have been mitigated by three inter-related developments in conventional digital electronic systems, namely the exponential growth in the number of transistors per chip (Moore's Law) [2] and the number of computations carried out per Joule (Koomey's Law) [3], as well as the constancy in the power density as transistors have been reduced in size, otherwise known as Dennard scaling [4]. However, these trends that have been valid for many decades have started to break down in recent years, due to effects such as thermal injection and quantum-mechanical tunneling in nanometric-sized transistors [5]–[7]. There is thus a pressing need for new, non-von Neumann computing approaches that do not separate the essential operations of processing and memory; approaches that also offer inherent capabilities for the parallel processing required for a world of 'big data' [8].

Neuromorphic computing (see e.g., [9]) and in-memory computing (see e.g., [10]) are two such approaches. Moreover, carrying out processing in the optical domain rather than electrically opens up a route to the processing of large amounts of data in parallel (using for example wavelength division multiplexing) in a most energy efficient manner [11]–[13].

One potentially promising technology for optical neuromorphic and in-memory computing is that using chalcogenide phase-change materials combined with integrated Si-photonics [14]–[20]. Phase-change materials (PCMs) are alloys that present a very pronounced contrast in their complex refractive index between their amorphous and crystalline phases. Moreover, this change is reversible, very fast (on the order of the nanoseconds) and can be reversed over many cycles [21], [22]. When PCMs are deposited as small 'cells' on top of an integrated photonic waveguide (see Fig. 1), it is possible to control the optical amplitude and/or phase of the optical signal transmitted by the waveguide, according to the crystalline fraction of the cell. This provides a means by which the state of the PCM cell can
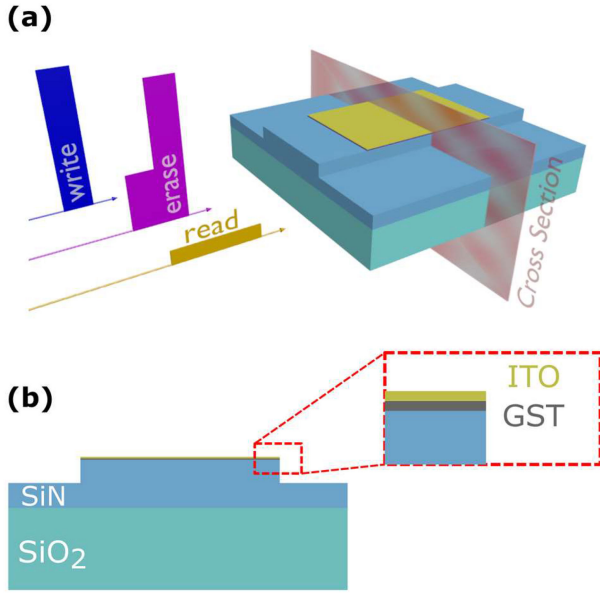
Fig. 1.   (a) Schematic representation of the basic unit cell and the different stimuli used to operate it. The basic unit cell consists of a rib (or sometimes a ridge) waveguide with a 10 nm phase-change material (here $Ge_2Sb_2Te_5$, or GST for short) covered by 10 nm of ITO. The profiles for the input pulses corresponding to the write, erase and read operations are represented on the left. The write pulse has an ideally squared shape and a relatively short duration, whereas the erase pulse is generally of longer duration and with a different profile to favor tailored control of the crystallization process. (b) Cross-section of the basic unit cell that shows the geometry and the materials used.

be optically 'read'. Further, the fraction of crystallized material in the PCM cell can also be controlled optically by sending appropriate 'write' (amorphization) and 'erase' (crystallization) pulses down the waveguide, as shown in Fig. 1.

The basic integrated phase-change photonic device shown in Fig. 1 has been shown capable of providing memory [23]–[25], arithmetic [15], [26] and neuro-synaptic [14] functionalities. PCM cells can also be combined with photonic micro-ring resonators to provide wavelength-division multiplexed memories [24], neuron [17] and synaptic [18] mimics. Indeed, progress with integrated phase-change photonic device technology has been rapid in recent years, and the first small-scale systems capable of performing pattern recognition tasks or matrix-vector multiplication operations have now been published [17], [20]. One of the key aspects in order to gain insights into the operation of such systems, to optimize their design and, importantly, to guide the successful realization of practicable large-scale phase-change photonic processing systems, is the development of compact models and their combination with commonly used (commercial) photonic system simulators. The target characteristics of a suitable compact model are simplicity (for fast computation) combined with an acceptable accuracy in predicting the response of real devices. In particular, we aim here to reduce simulation times for the programming (write/erase) of single integrated phase-change photonic devices from the several to many hours typically required by fully physical modeling (e.g., finite-element simulations of optical and thermal

characteristics combined with nucleation and growth type phase-change models) to the seconds to tens of seconds timescale required to enable, ultimately, the simulation of photonic integrated circuits (PICs) containing hundreds or thousands of such devices.

It is just such a compact model that we develop in this paper. In Section II we describe the model as applied to simulation of the read, write and erase processes in integrated phase-change photonic devices, while in Section III we integrate this model within the circuit simulator of IPKISS, a commercial photonic integrated circuit (PIC) design platform, and we simulate the performance of a prototype phase-change photonic neuromorphic processor [17]. Finally, in Section IV we apply our compact model to simulation of the performance of a phase-change photonic crossbar array capable of carrying out fast, low-energy, matrix-vector multiplication [20]. We thus concentrate on the development of the compact model and its use for simulation of integrated phase-change photonic devices in selected memory and computing applications. We do not explore integration/interfacing aspects of the active components (lasers, photodetectors etc.) required for realization of functionally-complete integrated photonic systems, and the reader is referred to many excellent reviews (e.g., [11]–[13]) for such information.

II. COMPACT BEHAVIORAL MODEL

The overall operation of the model is the result of the interaction of two sub-models, namely the readout model and the write-erase model. The readout model is concerned with extracting the value of the optical transmittance and accumulated optical phase through the integrated phase-change device for any particular amorphous/crystalline distribution in the PCM cell. The write-erase model, on the other hand, determines how the amorphous/crystalline distribution in the cell changes according to the optical write-erase optical pulses sent to the device. In this way, the resulting material phase distribution in the PCM cell determined by the write-erase model serves as input to the readout model to determine the characteristics (attenuation and phase constants) of the optical propagation mode inside the device.

A. Readout Model

The phase-change cell is considered to be divided in two distinct regions along the direction of propagation, as shown in Fig. 2(a). One of the regions has an optical propagation constant that corresponds to the phase-change material layer in its fully crystalline state. In the other region, the temperature has (as a result of a previously applied 'write' pulse) been high enough to melt the material and therefore it is amorphized. Note, however, that the amorphized region does not necessarily encompass the whole width of the PCM cell, and in such cases the value of the propagation constant is determined using effective medium theory (see supplementary materials, Section 2.4). The boundary between the two regions is located at the position, z_{int}, such that for $z < z_{int}$ the PCM is (partially) amorphous and for $z > z_{int}$ it is fully crystalline. The readout value (optical transmittance

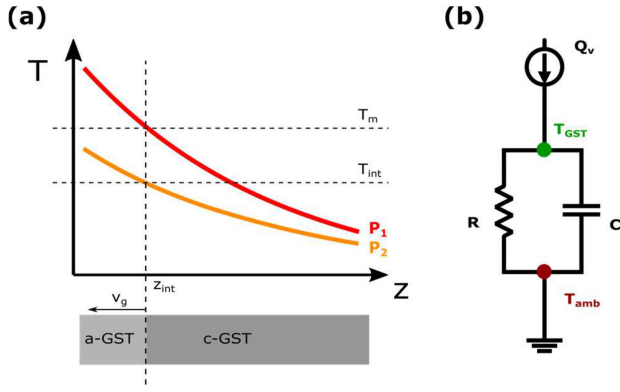**(a)**   **(b)**

Fig. 2. (a) Representation of the temperature distribution along the direction of propagation. The temperature distribution for the input power P_1 (red) has enough power to drive the temperature in a part of the cell above melting temperature, T_m, resulting in amorphization. The amorphized part of the cell is delimited by the position, z_{int}. If another pulse P_2 with lower power is sent and the temperature distribution along the cell is below melting temperature, the crystallization process is activated and the amorphous/crystalline interface will move. (b) The electrical circuit analogy used in the thermal model to describe the temperature evolution in the cell.

and accumulated optical phase) is then obtained by considering the combined contribution of these two regions. More specifically, the transmitted power P is calculated using the following formula:

$$P = P_0 \cdot \exp\left\{-2k_0\left[n_{\mathrm{Ia}}z_{\mathrm{int}} + n_{\mathrm{Ic}}\left(L - z_{\mathrm{int}}\right)\right]\right\} \quad (1)$$

where P_0 is the input power to the device, k_0 is the free space wave-vector, and n_{Ia} and n_{Ic} are the imaginary part of the complex effective refractive index for the partially amorphized and fully crystalline regions respectively.

Similarly, the accumulated phase θ is calculated using the following expression:

$$\theta = k_0 \cdot \left[n_{\mathrm{Ra}}z_{\mathrm{int}} + n_{\mathrm{Rc}}\left(L - z_{\mathrm{int}}\right)\right] \quad (2)$$

where n_{Ra} and n_{Rc} are the real part of the complex effective refractive index for the partially amorphized and fully crystalline regions respectively, and L is the total length of the PCM cell in the direction of propagation.

B. Write-Erase Model

The main task of the write-erase model is to determine the position of the amorphous/crystalline interface z_{int} after the application of write and erase excitation pulses. During amorphization, the position of this boundary is dictated by the position at which the temperature along the cell is equal to the PCM melting temperature, T_m, see Fig. 2(a). During crystallization, the temperature at the amorphous/crystalline interface, T_{int}, will be less than the melting temperature, and the position of z_{int} is determined by the (time-varying) value of T_{int} and the crystal growth velocity at T_{int}. The melting (amorphization) and crystallization cases are represented in Fig. 2(a) by the two different instantaneous exponentially decaying temperature profiles (red and yellow).

The heat transfer aspect of the write-erase model is described using a lumped circuit parameter model (see Fig. 2(b)) similar to that developed in [27] for electrical PCM devices. Here, a parallel RC tank circuit determines the current (heat flow) and potential (temperature). The ordinary differential equation which governs the evolution of temperature at position $z = 0$ (position of the input port in the PCM cell) is then described by:

$$C\frac{dT_{\mathrm{GST0}}}{dt} = \frac{T_{\mathrm{amb}} - T_{\mathrm{GST0}}}{R} + \frac{2k_0 n_{\mathrm{Ic}}P_0}{W_{\mathrm{eff}}} \quad (3)$$

where T_{GST0} is the temperature in the phase-change material at $z = 0$, R and C are the substrate thermal insulance (units $\mathrm{m^2 \cdot K \cdot W^{-1}}$) and heat capacity per unit area (units $\mathrm{J \cdot K^{-1} \cdot m^{-2}}$) respectively, and W_{eff} is the width of the PCM cell where the heat is generated (see supplementary materials Section 3 for a detailed derivation of this equation).

It is convenient to point out that we can estimate the temperature at any point along the direction of propagation in the cell from Eq. (3). We know the heat source (second term in right hand side of Eq. (3)) will change with a factor $e^{-\alpha z}$ (α being the attenuation constant of the PCM cell) in the direction of propagation z. Therefore, the temperature in the PCM cell and the rate of change of this temperature (time derivative term) are also going to vary according to the same exponentially decaying factor as the power. Thus, it is only necessary to calculate the temperature (time) evolution at $z = 0$ with equation (3) to have the approximate temperature along the complete cell. This simplifies the calculations and increases the speed of the model drastically as compared with our previously described behavioral model in [28].

The temperature calculated with Eq. (3) is introduced in a phase-change model to calculate the crystal growth velocity, and hence position, of the amorphous/crystalline boundary. The phase-change model used is the same as in [27]. The dependence of the growth velocity v_g with temperature below a threshold temperature T_{th} is described using an Arrhenius dependence:

$$v_g(T) = v_{g_\infty}\exp\left(\frac{-E_a}{k_B T}\right) \quad (4)$$

where v_{g_∞} is the growth velocity for $T \to \infty$, E_a is the activation energy, and k_B is the Boltzmann constant. On the other hand, for temperatures above T_{th} the dependence of v_g with T is given by:

$$v_g(T) = \frac{4r_{\mathrm{atom}}k_B T}{3\pi\lambda^2 R_{\mathrm{hyd}}} \cdot \frac{1}{\eta(T)} \cdot \left[1 - \exp\left(-\frac{\Delta G(T)}{k_B T}\right)\right] \quad (5)$$

where r_{atom} is the atomic radius, λ is the diffusional jump distance, R_{hyd} is the hydrodynamic radius, η is the viscosity, and ΔG is the Gibbs free energy difference between the liquid and the crystalline phase written here as:

$$\Delta G(T) = \Delta H_m \cdot \frac{T_{\mathrm{melt}} - T}{T_{\mathrm{melt}}} \cdot \frac{2T}{T_{\mathrm{melt}} + T} \quad (6)$$

In the above expression, T_{melt} is the melting temperature of the phase-change material and ΔH_m is the heat of fusion. The
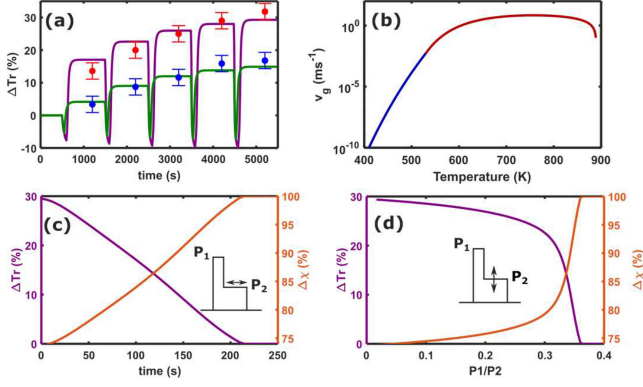
Fig. 3. (a) Comparison of experimental data (points and error bars) and the data produced by the model (solid lines). The sequence of pulses of increasing power in the experiment is 4.65, 5.24, 5.62, 5.86, 6.01 mW. This sequence of pulses produces the blue and red set of points for 50 and 100 ns duration pulses respectively in a 5 μm long cell. The green and purple lines are the data produced by the compact model when the transmittance is calculated with the same input parameters as in the experiments. (b) Growth velocity v_g plot produced by the parametric phase-change model. (c) Evolution of the change in transmittance of the phase-change cell and change in crystallinity (calculated as $(L - z_{int})/L$ where L is the length of the cell and z_{int} is the position of the amorphous and crystalline interface) for the two-step pulse scheme in [25]. The first part of the pulse has 100 ns width and $P_1 = 6.01$ mW power, and the second part of the pulse (represented by P_2 in the inset) has a fixed power of 2.4 mW. We have represented the change in transmittance and crystallization for gradually increasing duration of the second part of the pulse. (d) In this case the power and duration of the first part of the pulse is the same as in (c) but the duration is fixed to 250 ns, and the power P_2 is varied from 0.017 to 0.4 times the power of the first part of the pulse P_1 (i.e., 0.1 mW to 2.4 mW).

dependence of the viscosity with temperature is described by the well-known MYEGA model [29]:

$$\log_{10}\eta \ (T) = \log_{10}\eta_\infty + (12 - \log_{10}\eta_\infty)\frac{T_g}{T}$$

$$\times \exp\left[\left(\frac{m}{12 - \log_{10}\eta_\infty} - 1\right)\left(\frac{T_g}{T} - 1\right)\right] \tag{7}$$

where η_∞ is the viscosity for $T \to \infty$, T_g is the temperature at which the viscosity has a value of 10^{12} Pa·s, and m is the fragility given by $m = \frac{\partial \log_{10}\eta(T)}{\partial(T_g/T)}|_{T=T_g}$. For further details on the implementation of the model, see Section 7 of the supplementary information.

C. Experimental Data Fitting

The model described above was tested against, and fitted to, experimental data from [25] and [30]. More specifically, experimental results from [30] were used to test and fit the write (amorphization) and readout models, and results from [25] were used for the erase (re-crystallization) part of the model. The parameters involved in the fitting of the model to the results (as shown in Fig. 3(a)) are R, C, and W_{eff} in the thermal part, as well as the value of the imaginary part of the effective refractive index, n_{Ia}, in the partially amorphized region of the PCM cell. The values of v_{g_∞}, E_a, η_∞, T_g, and T_{th} are the parameters used for fitting in the parametric phase-change model. The fitting for these parameters is carried out guided by values reported in

the literature, and all values used are listed in the supplementary materials Section 1, along with details of the fitting process itself.

The fitting of models to experimental data is carried out in two parts concerned with processes involving a net decrease in crystallinity (amorphization) and vice-versa (crystallization). In the part concerned with amorphization, experimental data from [30] for the change in transmittance of the device as a function of pulse duration and power is used to carry out the fitting. Specifically, several pulses of different powers and two pulse durations of 50 and 100 ns were sent to a 5μm long cell, with results being shown Fig. 3(a). It is possible to see a good fit between the data from the model (solid lines) and experimental data (points with error bars).

The second part of the fitting is concerned with the erasing (re-crystallization) process; we here test our model using the two-step pulse erase approach detailed in [25] (and shown schematically in the insets of Fig. 3(c) and (d)). The first part of the pulse is relatively short and of high power; its function is primarily elevating the temperature of any pre-existing crystalline regions in the PCM cell to above melting. The second part of the pulse is less intense than the first part and longer; its function being to maintain for sufficient time the temperature in the PCM at one conducive to achieving crystallization to a desired level (to erase or program the cell). Thus, as part of the fitting process, we send an erase/program pulse with the first part having a power 6.01 mW and duration 100 ns, and the second part set to 2.4 mW (0.4 times the power of the first part of the pulse) and 200 ns duration. The objective in the fitting is to achieve complete re-crystallization for such a pulse, as has been observed experimentally over different datasets in which re-crystallization using this pulse profile has been demonstrated [25], [26], [30]. Note that the initial state of the cell (prior to the two-step crystallizing pulse) during this process is that achieved after the last of the five (100 ns) write pulses shown in Fig. 3(a).

In order to evaluate the fitting and further test the re-crystallization aspect of the model with the parameters obtained as a result of the fitting, we run two set of simulations following the experiments in [25]. In the first simulation, for a fully crystalline cell, we apply the same type of two-step pulse described above (6.01 mW for 100 ns followed by 2.4 mW), but this time the duration of the second part of the pulse is varied from 0 to 250 ns. The values of the change in transmission and the change in crystallization as a result of this operation are represented in Fig. 3(c). Note that the value of the change in transmittance at 200 ns duration for the second part of the crystallization pulse is very close to zero, indicating that the constraint previously set in the fitting process (for complete crystallization) is appropriately achieved. A second simulation was also carried out, but in this case instead of varying the duration of the second half of the pulse with a fixed power, the duration was fixed at 250 ns and the power of the second part of the two-step pulse was varied from 0.017 to 0.4 times the power in the first part of the pulse, Fig. 3(d).

From the above results, it can be seen that for the case where the power of the second part of the pulse is fixed but its duration is altered (Fig. 3(c)) we observe a near linear relationship between

the duration of the second part of the pulse and the readout value achieved. However, in the case where the second part of the programming pulse has a fixed duration but a variable power (Fig. 3(d)), there is a rather rapid change in device transmission, reflecting a rapid change in crystallized fraction, once the power of the second part of the pulse reaches around 0.3 that of the first part. These results match well with experimental trends shown in [25].

As discussed above, an important part of the model fitting process is concerned with matching as realistically as possible the crystallization dynamics seen experimentally. A necessary condition in this respect is that the growth velocity v_g should be continuous at the temperature T_{th} at which we move from the crystallization model described by Eq. (4) to that of Eq. (5) Such a condition is indeed met, as shown in Fig. 3(b) where we plot the value of v_g as calculated using Eq. (4) (blue line) and (5) (red line) and incorporating values for all the relevant parameters derived using the above fitting processes.

Generally speaking, the application of the above described behavioral model consists of the collection of a series of parameters that can be identified as a 'library' for a specific integrated phase-change photonic device. The parameters that form the library for any specific device are the effective refractive index of the cell, and the parameters for both the thermal and phase change models. These parameters are defined/determined for any device using a combination of FEM techniques and fitting to experimental data. More specifically, for the fully crystalline region of the PCM cell the effective refractive index is calculated using FEM, whereas for the real and imaginary parts of the amorphized refractive index the values are found by fitting and the effective medium approximation respectively (see supplementary materials Sections 1, 2.1 and 2.4 for further details). The parameters for both the thermal model (R, C, and W_{eff}) and the phase change model (v_{g_∞}, E_a, η_∞, T_g, and T_{th}) are found by carefully fitting to experimental data (write and erase experiments), as has been done in this section. All such parameters, along with the geometry of the cell (length and thickness of PCM layer) allow for the creation of a model library for any particular device (akin to the SPICE type libraries used for modeling electronic devices). Although in the work described above we have concentrated on what is often called the basic unit cell comprising here a straight rib type waveguide fabricated in SiN, suitable libraries can be readily developed for other device types (indeed in the following sections we apply models developed for neuronal and synaptic mimics, and phase-change photonic crossbars), different device geometries/configurations (ridge and rib waveguides, Si and SiN waveguides etc.), and different phase-change materials (e.g., GeSbTe, GeSbSeTe, GeTe, AgInSbTe, SbS etc.).

III. COMBINATION WITH INTEGRATED PHOTONICS DESIGN FRAMEWORK: NEURO-SYNAPTIC SYSTEM SIMULATION

The model described in the previous sections runs extremely quickly on even modest computing resources (the computing time is on the order of seconds for time-domain simulations
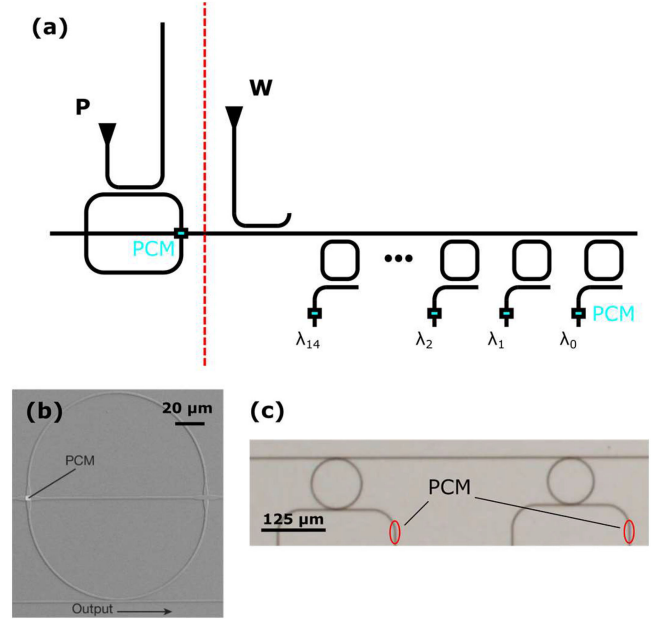

Fig. 4. (a) Schematic of the neuro-synaptic system from [17] to be simulated in this work. The red dashed line divides the system in the two parts to be studied, namely the individual synapses and collector on the right, and the activation unit on the left. The port named as W is used to program the individual synapses and the port P is used to read the state of the phase-change element in the activation unit and generate the neuron's output spike (it can also be used to write, erase or apply a bias power, see later in the main text). Cyan squares in the diagram denote the position of the phase-change material cells. (b) Scanning electron microscope image of a fabricated activation unit. (c) Optical microscope image of two synapses and their connection to the collector bus. Figs 4(b) and (c) are adapted with permission from [17].

covering tens of write-erase cycles in a single cell - see the supplementary information Section 7.3 for more detail of model execution times), and as such is suitable for the design and simulation of the performance of large-scale systems where many phase-change integrated photonic devices would be present. To do this most effectively, and to enable the ready interfacing and combination of phase-change photonic devices with standard Si-photonics components (couplers, splitters, resonators etc.), we have also integrated our compact model into the commercially available PIC design software tool IPKISS 3.3.0 (more specifically, we have used the photonic integrated circuit simulator Caphe [31] built in the IPKISS 3.3.0 framework, see supplementary information Section 4 for further details). We now use the comprehensive simulation tool that we have developed to explore the design and optimization of a neuro-synaptic integrated phase-change photonic system.

The neuro-synaptic system we investigate is that previously published in [17]. In that work, both neurons and synapses were realized using phase-change photonic devices, the former by integrating a phase-change cell with a micro-ring resonator, the latter by using straight waveguide devices similar to those discussed in previous sections. Each neuron was connected to fifteen plastic synapses, with the outputs of all synapses connected to a single bus and directed towards the input of the integration (neuron) unit, which activates when sufficient power is received, see Fig. 4(a) (we also show an example of a fabricated
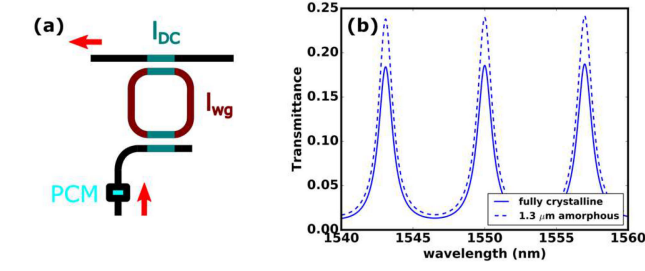
Fig. 5. Synapse simulation, (a) Part of the neuro-synaptic system that comprises the synapse. The arrows represent the input and output ports in the simulation. (b) Calculated transmittance of a synapse in its fully crystalline state (solid line) and a partially amorphous state, specifically for z_{int} at 1.3 μm (dotted line).
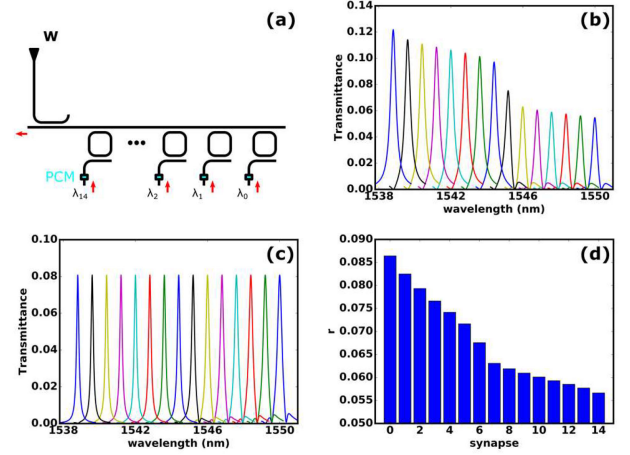

Fig. 6. Collector optimization. (a) Part of the neuro-synaptic system that comprises the collector along with the synapses. The arrows represent the input and output ports in the simulation. (b) Transmittance spectra for the inputs and output in (a) when the synapses are not optimized. (c) Response of the collector for all the synapses after optimization for equal transmittance at resonance for the same crystallinity state (equal synaptic weighting for the same crystallinity). Please, note that the 15 resonant peaks plotted in (b) and (c) correspond to the responses of the 15 ring resonators linking the synapses to the collector. (d) Ratio of the length of the couplers to the length of the waveguides for each synapse in the optimized solution.

integration unit and synapses from [17] in Fig. 4(b) and (c) respectively). In the following subsections, we simulate and optimize each of the basic components of this neuro-synaptic system (a single synapse, the collector, and the activation unit), and demonstrate in simulation its application to simple pattern recognition tasks (for details of how these basic building blocks can be connected together to form multi-neuron, multi-layer spiking neural networks, see [17]).

A. Synapse and Collector Simulation

A single synapse comprises a PCM cell deposited onto a rib waveguide and connected to an add-drop ring resonator, as shown in Fig. 5(a). The frequency response expected from this configuration is similar to that of a bare add-drop ring resonator, but with the output transmittance signal weighted by changes in the crystallinity of the PCM cell. This response is represented in Fig. 5(b) for the fully crystalline state and for a partially amorphized ($z_{int} = 1.3$ μm) synapse, which approximately corresponds to the position of the amorphous and crystalline interface after sending a pulse with 6 mW power and 100 ns duration (as shown in Fig. 3(a)).

The model for the synapse of Fig. 5(a) requires additional input parameters to those described previously for the basic unit cell type device. Those parameters are the resonant order of the ring m, the desired resonant frequency λ_R, the ratio of the length of the directional couplers to the length of the waveguides forming the ring $r = \frac{l_{DC}}{l_{wg}}$ (see Fig. 5(a)), and the gap of the directional couplers g_{DC} (see supplementary materials Section 2.3 for details on the modeling of the directional couplers). Using n_{eff} and λ_R we can obtain the total length of the ring using the expression $L_{Ring} = \frac{m \cdot \lambda_R}{n_{eff}}$. In figure Fig. 5(b) we show the response of one modeled synapse with $\lambda_R = 1550$ nm, $m = 180$, $r = 0.2$ and $g_{DC} = 250$ nm. This corresponds to a 173.7 μm long ring with 14.48 μm long directional couplers (with $n_{eff} = 1.606$ at 1550 nm wavelength). The calculation of the transmittance in Fig. 5(b) has been carried out from the input to the through port for a synapse incorporating a 3 μm long PCM cell. The spectrum is represented both for the fully crystalline state of the phase-change material and for the position of the amorphous/crystalline interface at 1.3 μm. This difference in crystallinity is reflected in the difference in transmittance

($\sim$29%) between the two spectra, and forms the basis of the synaptic weighting operation.

For proper operation of the neuro-synaptic system, the contribution to the activation unit from each individual synapse (from the input port to the port at the output of the collector as shown in Fig. 6(a)) should be identical for identical crystallinity states (identical weightings) of each synapse. In general, it is not possible to have the previously described response in a system where the different ring resonances are relatively close to each other (0.8 nm in this case). This is because the outputs of the synapses which are further away in the layout from the activation unit are affected by the resonant behavior of the synapses closer to it. This effect is shown in Fig. 6(b) where the collector transmittance of an un-optimized set of fifteen fully crystalline synapses is plotted (all synapses have the same value for r set to 0.1 in this case). Fig. 6(c), on the other hand, shows the optimized transmittance plot where the synapses now show the desired equal transmittance for equal crystallinity. The optimization is here achieved by obtaining the appropriate coupling efficiency for the directional couplers of each synapse. The optimization is carried out by searching the appropriate value of r, while leaving L_{Ring} constant (the corresponding λ_R, $m = 185$, and $g_{DC} = 250$ nm remain fixed and equal to those in Fig. 6(b)). The optimum solution (optimized value for r) is plotted in Fig. 6(d). Note that we have only represented the 185th order resonant peak in Fig. 6(b) and Fig. 6(c) for clarity (please consult the supplementary materials Section 5.1 to see the plot containing higher order peaks). Note also that the length of the rings, the spectral separation of the peaks and the working interval of the system are kept similar to those shown in Ref. [17]. Moreover, we show a comparison of the results of the models
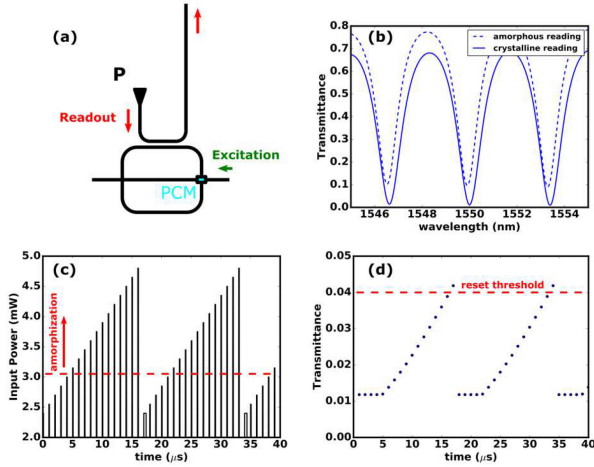
Fig. 7. (a) Representation of the activation unit, as well as the readout and excitation ports. (b) Transmittance of the activation unit for the fully crystalline state and the fully amorphized state ($z_{int} = L$) of the neuron's PCM cell. (c) Stream of gradually increasing power pulses of 100 ns duration and 1 μs period sent to the activation unit through the excitation port. The threshold for amorphization is indicated by the dashed red line. (d) Transmittance of the activation unit measured through the readout port after the interaction with the pulses in (c). The dashed line represents the threshold for re-crystallization. Note that simulations in (c) and (d) are carried out for a 1550 nm wavelength.

for the directional couplers against experimental results in the supplementary information Section 2.3.

B. Activation (Neuron) Unit Response

The sum of all the signals from all the synapses passes from the collector to the activation (or neuron) unit that comprises a micro-ring resonator with its own PCM cell (see Fig. 4(b)). If the total optical power reaching the activation unit is above a certain threshold, its PCM cell is amorphized and an output spike should be generated in (or more properly passed through) the readout waveguide P shown in Fig. 7(a). In a little more detail (for full details of the neuron operation see [17]), changes in the crystallinity of the neuron's PCM cell lead to changes in the transmission of the readout waveguide, and this change in transmission is converted to a spike output by periodically sending probe pulses down the readout waveguide. Such changes in transmission occur because the resonant frequency of the neuron's micro-ring resonator shifts towards shorter wavelengths with increasing pulse energy sent to its PCM cell. At the same time, the transmission in the neuronal ring resonator increases because of reduced absorption in the PCM cell, and thus changes the coupling between ring and waveguide. Simulations using our compact model reveal just such a behavior, as shown in Fig. 7(b), confirming that the model captures the essential operation of the activation unit. The physical lengths of the optical elements in the activation unit (waveguides, directional couplers and crossings) are chosen so that the resonance for the fully crystalline state of the neuron's PCM cell occurs at 1550 nm (see supplementary information Section 2 for further details about the optical elements).

Next, the response of the activation unit to input excitation pulses in the time domain is also examined. A stream of trapezoidal pulses with 100 ns duration and 5 ns rise/fall time is sent to the device through the excitation port (post-synaptic collector waveguide), with 1000 ns period. The power of the pulses was gradually increased from 2.4 mW in increments of 0.15 mW until the transmittance in the cell reaches a threshold value fixed at 0.04, whereafter a re-crystallization pulse of 2.4 mW and 400 ns was applied, and the process repeated, see Fig. 7(c). The stream of excitation pulses causes a transmission response in the device that resembles very much that of a rectified linear unit (ReLU), when captured through the readout port, Fig. 7(d), thus offering (as shown in experimentally in [17]) the essentials of a neuronal mimic.

C. Pattern Detection

All the obtained responses in the previous sections for the synapses and their joint operation with the collector, as well as the activation (neuron) unit are now combined to provide a neuro-synaptic system capable of carrying out the exemplar processing task of pattern detection. The first step to prepare the system to carry out a pattern recognition operation is programming the different synapses according to the pattern to be recognized (as done experimentally in [17]). To do this, some synapses are set to the fully crystalline state, while others are amorphized, resulting in a difference in transmitted optical power through them, as explained in Section III.A. This scheme is illustrated in Fig. 8(a), where the inputs to all the synapses are represented by a 5 by 3 array used to represent the pattern of interest. Fully crystalline synapses are shown in grey and amorphized synapses in white, and in this case it is possible to see that the arrangement of the crystalline (grey) synapses forms the letter 'A'. The amorphized synapses in this case result in 27.20% increment in transmittance from the fully crystalline state in each individual input. We show in the supplementary materials section 5.3 all the details of the amorphization of the synapses (which here is achieved by sending a pulse of the right frequency and amplitude through the port W in Fig. 4).

Once the system is programmed, it is tested with different input patterns. In Fig. 8(b), two different input patterns (taken from the set of possible letters A, B, C and D used for this task) are presented to the system trained (programmed) to recognize the letter A. In the first case (top of Fig. 8(b) and left in Fig. 8(c)), the input pattern presented to the synapse array (as represented by the red arrows) matches that of the programmed pattern. This configuration gives the maximum power at the output of the collector bus, since the transmittance of the programmed (amorphized) synapses is significantly larger than that of the fully crystalline synapses. The case of a non-matching pattern, representing the letter C, is illustrated in the bottom panel of Fig. 8(b) (and in the right half of Fig. 8(c)). Here, it is possible to see that the resulting power at the output of the collector will be lower than that for the fully matching pattern (the 'A'), since some of the synapses that couple the power to the collector are in the fully crystalline state and thus have a lower transmittance. Finally, the contrast (change in transmission) generated in the
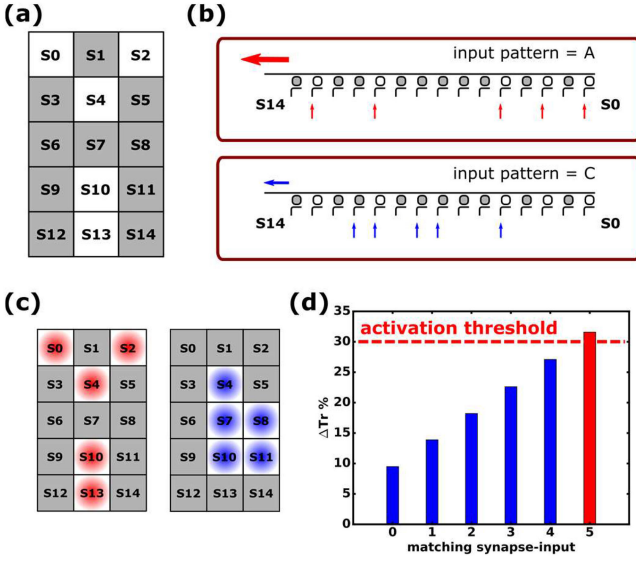
Fig. 8. (a) Array-like representation of the different synapses in the system. Amorphized (programmed) synapses are shown in white, whereas crystallized (non-programmed) synapses are in grey. The contour of the non-programmed (grey) synapses forms a letter A, and therefore this configuration of synapses is used to recognize this symbol. (b) Schematic representation of the programmed and non-programmed synapses in the layout of the system, from synapse S0 on the right to S14 on the left. The arrows at the input of the synapses represent the demultiplexed signal which forms the input pattern to the system. Two cases are represented: the case on the top (red arrows) represents an input pattern that is coincident with the amorphized synapses in the programmed pattern; bottom represents the input pattern correspondent to a letter C. The arrow at the end of the collector schematically represents the summation of the energy transmitted by the synapses, and is bigger in the case of the input pattern matching the programmed pattern (top panel) than otherwise (bottom panel), due to the increased transmittance of the programmed synapses. (c) Array representation of the two patterns for the symbols 'A' and 'C'. Note here the complementary aspect of the input pattern and their correspondence with the excitation of the different programmed synapses. Note as well that A and C pattern representations share two programmed synapses in common. (d) Change in transmittance in the readout of the system as a function of the number of matching programmed synapses in the input pattern. The pattern A in (c) corresponds to the only case that surpass the activation threshold (the input produced by the C pattern would produce a change in transmission corresponding to the excitation of two programmed synapses at the input).
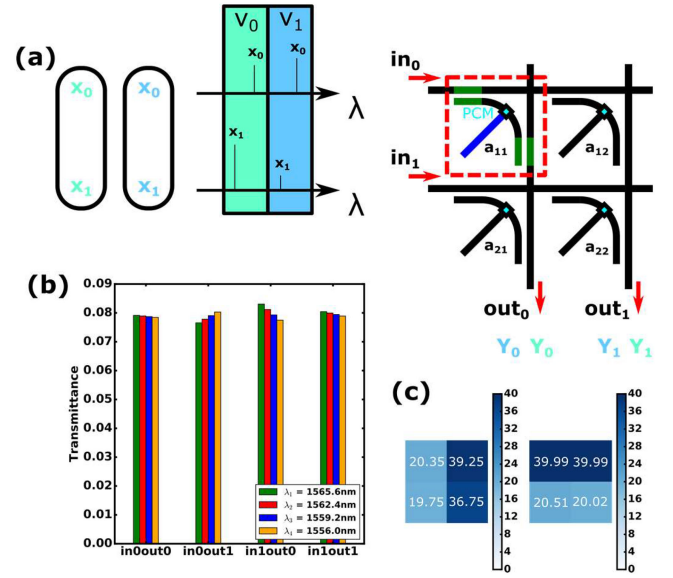

Fig. 9. (a) Schematic representation of parallel matrix-vector multiplication. In this example two vectors with two components are supplied simultaneously to the matrix-vector multiplier. The components of the vector are codified at different wavelengths in the working wavelength interval of the array. The first components of the vectors are introduced through port in_0 and the second components are introduced through port in_1. Two matrix-vector multiplications are the carried out simultaneously. Note that the four cyan squares in the represented device denote the position of the phase-change cells. (b) Optimized response of the matrix-vector multiplier. The transmittance of all the wavelengths through the different input and output ports is optimized to be as similar as possible for identical phase-states of each PCM cell (here fully crystalline). (c) Length of the directional couplers in μm for the optimal solution in (b). The plot on the left represents the length of the horizontal directional couplers and the plot on the right represent the length of the vertical directional couplers (here for an exemplar 2x2 array).

activation unit as a function of the number of inputs that match the programmed synapses is represented in Fig. 8(d). Only with 5 matching inputs, the number required to correctly recognize the pattern 'A', does the activation unit reach the required threshold for neuron 'firing'. Extended information on the pattern recognition task can be found in the supplementary information Section 5.2.

IV. COMBINATION WITH INTEGRATED PHOTONIC DESIGN FRAMEWORK: VECTOR-MATRIX MULTIPLIER

Integrated photonics combined with phase-change materials also offers the possibility of performing important arithmetic operations such as matrix-vector multiplication, a central element in much AI (artificial intelligence) based computing [26], [32], [33]. The great attraction of a photonic approach over a purely electronic approach resides in the large amounts of information that can be transmitted and processed in parallel by exploiting wavelength division multiplexing (WDM). Moreover,

by storing matrix values using non-volatile PCM cells (i.e., using a form of in-memory computing), inference processing can be carried out using extremely low energy. Indeed, we have recently demonstrated an integrated phase-change photonic matrix-vector multiplier with computing speeds in the TMAC/s (tera multiply-accumulate operations per second) range coupled with the potential for energy efficiencies in the tens to hundreds of femto-Joule per MAC [20].

The basic configuration and operation of the matrix-vector (MV) multiplier from [20] is shown in Fig. 9(a). Interconnected multiplicative units are arranged in a rectangular photonic cross-bar configuration, with inputs (vectors) applied to the rows, the matrix elements stored in phase-change cells at the crossings, and outputs taken from the bottom of the columns. The multiplications at each individual cell in the MV multiplier are of the form $a \cdot b = c$ and are performed as explained in [20], with the values of the amplitude of the optical signal and the phase-state of the PCM cell being codified in a and b respectively. Components of the vectors are also codified in different wavelengths (within the working interval of the array) and introduced through the input row ports, Fig. 9(a). In this way, each component of the vector corresponds to one row. Once inside the MV multiplier, each input vector component is multiplied by each matrix element on that row simultaneously, and the results of such multiplications are then summed in

each column to generate the correct outputs. To do this, two directional couplers divert the right amount of light to each PCM cell, and couple light back to the vertical waveguides that comprise the columns, Fig. 9(a).

For proper operation of the MV multiplier, each set of directional couplers must be carefully adjusted to compensate for the fact that the further a PCM cell is to the right in the array, the more light is absorbed along the row waveguides before the input reaches it (and the higher a cell is in a column, the more light is absorbed in the column waveguide before it reaches the output). The design of such couplers is straightforward using our compact model integrated with IPKISS. Specifically, we optimize the lengths of all the directional couplers so all input-output combinations in the vector-matrix have the same transmittance for the same level of crystallinity in the PCM cells. Results are shown in Fig. 9(b) and confirm (near) uniform contribution for each input-output combination, as required. The results of the optimization, in terms of length of the directional couplers that should be used, is shown in Fig. 9(c). (For further details of the optimization of the array, see the supplementary information Section 6.3).

The device that performs the basic operation shown in Fig. 9 can be scaled up and used to parallelize the operations that are carried out in the convolutional layers of a convolutional neural network [20]. Each individual convolution kernel can then be applied in parallel to the input data using WDM, with the only restriction of how much information it is possible to codify being the working wavelength interval and the optical dispersion of the various optical elements in the system. As an example, we show in Fig. 10 an arrangement for parallel processing (parallel convolutions) for data comprising an image of the handwritten digit '4'. The correspondence between the data in the image to be convolved and the codified input vectors is shown in Fig. 10(a). Here, a sixteen-wavelength mask is moved through the image in each iteration until it is completely convolved. The four groups of inputs that form the mask, which are represented in different colors (red, yellow, green, and cyan), are introduced simultaneously into a 4 by 4 MV multiplier (or tensor core) of the type shown in Fig. 9(a) and processed in parallel by applying different kernels programmed in the columns of the array, as shown in Fig. 10(b).

We now demonstrate the application of our behavioral model, integrated with IPKISS 3.3.0, to simulate the performance of a phase-change MV multiplier array (or photonic tensor core) of the form shown in Fig. 10 for convolutional processing. Specifically, we program a kernel for left-edge detection into one column of the array, as shown in Fig. 10(b), and then apply this to the image of the handwritten '4' from Fig. 10(a). Results are shown in Fig. 11, with Fig. 11(a) showing the expected results (i.e., mathematically evaluated result) and Fig. 11(b) the result found using our simulation tool – an excellent correspondence is clear.

In the above case, modeling of the convolutional performance of the MV multiplier was carried out for a system with no impairments (i.e., programmed levels ideal, perfect calibration via the reference array, zero noise in input signals). A significant benefit of having a system-level model is that the effects of
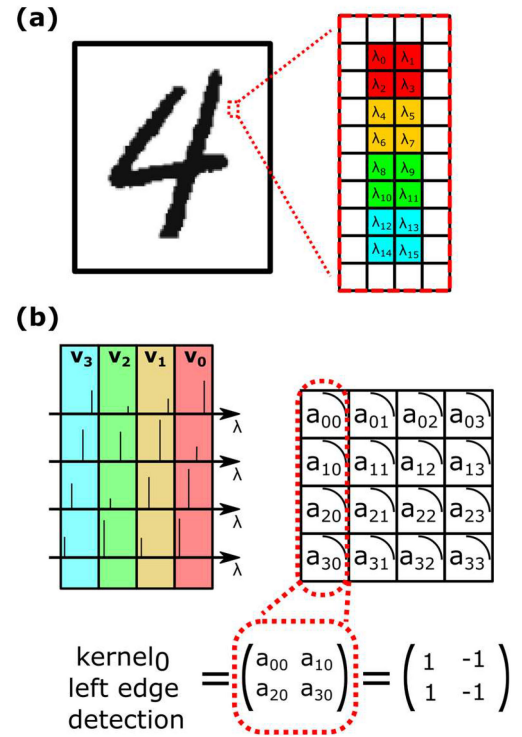

Fig. 10. (a) Image of a handwritten digit '4' taken as example to illustrate how convolution processing (MV multiplication) is carried out in the photonic phase-change crossbar. The pixels in the image are codified in the intensity of optical pulses at different wavelengths. Different colours represent groups of pixels in the image to be convolved to form a new pixel in the convolved image. (b) Clusters of pixels are flattened and arranged in four vectors that are given as inputs to the crossbar. The different convolutional kernels are programmed into (PCM cells of) the columns of the crossbar, as illustrated here for the example for a left edge detection kernel. It is worth highlighting that the four different vectors are operated with the four different programmed kernels simultaneously (for further details see Ref. [20]).
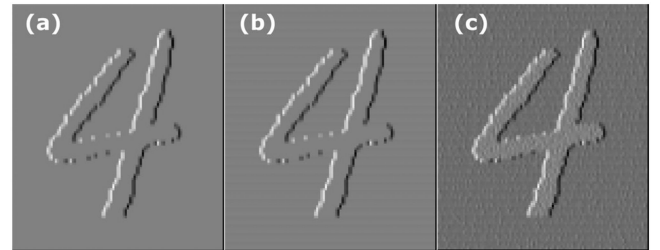

Fig. 11. (a) Result of the convolution of the image shown in Fig. 10(a) with a kernel for left edge detection. (b) Same as (a) but calculated using our behavioral model (integrated with IPKISS). (c) Same as (b), but adding effects of programming errors, temperature difference between the tensor and reference cores, and Gaussian noise in the input vectors.

such impairments can be effectively modeled too. By way of example, we here consider three impairments likely to affect real physical systems: (i) the variability in the programmed weights in both the tensor core and the reference core, (ii) a temperature difference between the tensor and reference cores, and (iii) the presence of Gaussian noise in the source used to codify the image (i.e., noisy input vectors). The need for the reference core mentioned in (i) arises because the device readout value is not absolute, but depends on the power levels chosen to

represent the input vectors. To account for this, it is necessary to have another (ideally identical) matrix, called reference core, from which reference values are obtained that allow for proper calibration of the MV outputs (for further details, please see the supplementary material Section 6).

To study the effect caused by differences in the programmed weights between the tensor core and the reference core, we use the standard deviation of the programming error found in [25], which is 0.416% (relative to the fully crystalline transmittance). This standard deviation is applied when the tensor core and the reference core are both programmed. With regards to the effect of temperature, we here introduce a temperature difference in the chip that would make the temperature of the reference core to be 30 K higher than that of the tensor core (which has all its multiplicative units at 'room' temperature, T = 293.15 K). As for noise in the input vectors, we assume a Gaussian noise with standard deviation of 15 (relative to a maximum level of 255 to represent white). The effect of these combined impairments on the convolution process carried out by the MV multiplier array results in the output image shown in Fig. 11(c). Here, the temperature variation and programming errors result in a change of the background grey level representing the white background regions of the original image, whereas the effects of noise in the input vectors appear evident over the whole extent of the figure.

V. CONCLUSION

A compact behavioral model capable of simulating, in a fast, efficient and accurate manner, the performance of integrated phase-change photonic devices has been developed. We have used this model to explore in detail the write (amorphization) and erase (re-crystallization) processes in the so-called basic unit cell device, comprising a straight rib waveguide (here in SiN) with a phase-change material (here GST) cell deposited atop. Very good agreement between model simulations and published experimental data was obtained. Moreover, our model is readily incorporated with commercially available PIC design platforms (here we use the well-known IPKISS software from Luceda Photonics), so enabling the design, simulation and optimization of large-scale phase-change photonics systems that combine phase-change devices with more standard integrated photonic components (couplers, resonators, splitters etc.). We demonstrate such capabilities by exploring the optimization and simulation of the operating characteristics of two important phase-change photonic systems recently investigated experimentally in the literature, namely a spiking neural network system and a matrix-vector photonic crossbar array (photonic tensor core). Results show that use of our behavioral model can significantly facilitate the design and optimization at the system level, as well allowing for the exploration of the capabilities of new proposed architectures in phase-change photonics computing field. Although here we have concentrated on devices fabricated in SiN and with GST as the chosen phase-change material, our approach is easily extended (by developing appropriate sets of device model parameters) to SOI type devices and alternative PCM compositions. In summary, the work reported in this paper can be regarded as a very significant step towards the creation of SPICE-like libraries and process design kits for the burgeoning field of integrated phase-change photonics.

REFERENCES

[1] J. von Neumann, *The Collected Works of John Von Neumann*. Oxford, U.K.: Pergamon Press, 1963.

[2] G. E. Moore, "Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff.," *IEEE Solid-State Circuits Soc. Newslett.*, vol. 11, no. 3, pp. 33–35, Sep. 2006.

[3] J. G. Koomey, S. Berard, M. Sanchez, and H. Wong, "Implications of historical trends in the electrical efficiency of computing," *IEEE Ann. Hist. Comput.*, vol. 33, no. 3, pp. 46–54, Jul. 2011.

[4] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct. 1974.

[5] Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. Dev.*, vol. 46, no. 2.3, pp. 213–222, Mar. 2002.

[6] T. N. Theis and P. M. Solomon, "In quest of the next switch: Prospects for greatly reduced power dissipation in a successor to the silicon field-effect transistor," *Proc. IEEE*, vol. 98, no. 12, pp. 2005–2014, Dec. 2010.

[7] T. N. Theis and H. S. Philip Wong, "The end of Moore's law: A new beginning for information technology," *Comput. Sci. Eng.*, vol. 19, no. 2, pp. 41–50, Mar. 2017, doi: 10.1109/MCSE.2017.29.

[8] T. Ben-Nun and T. Hoefler, "Demystifying parallel and distributed deep learning: An in-depth concurrency analysis," *ACM Comput. Surv.*, vol. 52, no. 4, Aug. 2019, Art. no. 65, doi: 10.1145/3320060.

[9] T. Chang, Y. Yang, and W. Lu, "Building neuromorphic circuits with memristive devices," *IEEE Circuits Syst. Mag.*, vol. 13, no. 2, pp. 56–73, Second Quarter 2013.

[10] A. Sebastian, M. L. Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nature Nanotechnol.*, vol. 15, no. 7, pp. 529–544, Jul. 2020, doi: 10.1038/s41565-020-0655-z.

[11] T. Ferreira De Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Progress in neuromorphic photonics," *Nanophotonics*, vol. 6, no. 3, pp. 577–599, Mar. 2017, doi: 10.1515/nanoph-2016-0139.

[12] B. J. Shastri, A. N. Tait, T. F. de Lima, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Principles of neuromorphic photonics," *Encycl. Complex. Syst. Sci.*, pp. 83–118, Aug. 2018, doi: 10.1007/978-1-4939-6883-1_702.

[13] B. J. Shastri *et al.*, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, pp. 102–114, Jan. 2021, doi: 10.1038/s41566-020-00754-y.

[14] Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, "On-chip photonic synapse," *Sci. Adv.*, vol. 3, no. 9, Sep. 2017, Art. no. e1700160, doi: 10.1126/sciadv.1700160.

[15] J. Feldmann *et al.*, "Calculating with light using a chip-scale all-optical abacus," *Nature Commun.*, vol. 8, no. 1, Nov. 2017, Art. no. 1256, doi: 10.1038/s41467-017-01506-3.

[16] I. Chakraborty, G. Saha, A. Sengupta, and K. Roy, "Toward fast neural computing using all-photonic phase change spiking neurons," *Sci. Rep.*, vol. 8, no. 1, Aug. 2018, Art. no. 12980, doi: 10.1038/s41598-018-31365-x.

[17] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, May 2019, doi: 10.1038/s41586-019-1157-8.

[18] I. Chakraborty, G. Saha, and K. Roy, "Photonic in-memory computing primitive for spiking neural networks using phase-change materials," *Phys. Rev. Appl.*, vol. 11, no. 1, Jan. 2019, Art. no. 014063, doi: 10.1103/PhysRevApplied.11.014063.

[19] C. Wu, H. Yu, S. Lee, R. Peng, I. Takeuchi, and M. Li, "Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network," Accessed: Jan. 2021. [Online]. Available: Apr. 2020.[Online]. Available: http://arxiv.org/abs/2004.10651

[20] J. Feldmann *et al.*, "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, no. 7840, pp. 52–58, 2021, doi: 10.1038/s41586-020-03070-1.

[21] S. Raoux and M. Wuttig, *Phase Change Materials: Science and Applications*. Berlin, Germany: Springer, 2009.

[22] S. Raoux, F. Xiong, M. Wuttig, and E. Pop, "Phase change materials and phase change memory," *MRS Bull,* vol. 39, no. 8, pp. 703–710, 2014, doi: 10.1557/mrs.2014.139.

[23] C. Rios, P. Hosseini, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "On-chip photonic memory elements employing phase-change materials," *Adv. Mater.*, vol. 26, no. 9, pp. 1372–1377, Mar. 2014, doi: 10.1002/adma.201304476.

[24] C. Ríos *et al.*, "Integrated all-photonic non-volatile multi-level memory," *Nature Photon.*, vol. 9, no. 11, pp. 725–732, Nov. 2015, doi: 10.1038/nphoton.2015.182.

[25] X. Li *et al.*, "Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell," *Optica*, vol. 6, no. 1, Jan. 2019, Art. no. 1, doi: 10.1364/OPTICA.6.000001.

[26] C. Ríos *et al.*, "In-memory computing on a photonic platform," *Sci. Adv.*, vol. 5, no. 2, Feb. 2019, Art. no. eaau5759, doi: 10.1126/sciadv.aau5759.

[27] A. Sebastian, M. L.. Gallo, and D. Krebs, "Crystal growth within a phase change memory cell," *Nature Commun.*, vol. 5, no. 1, pp. 1–9, Jul. 2014, doi: 10.1038/ncomms5314.

[28] S. G. C. Carrillo *et al.*, "Behavioral modeling of integrated phase-change photonic devices for neuromorphic computing applications," *APL Mater,* vol. 7, no. 9, Sep. 2019, Art. no. 091113, doi: 10.1063/1.5111840.

[29] A. Aladool, M. M. Aziz, and C. D. Wright, "Understanding the importance of the temperature dependence of viscosity on the crystallization dynamics in the $Ge_2Sb_2Te_5$ phase-change material," *J. Appl. Phys.*, vol. 121, no. 22, Jun. 2017, Art. no. 224504, doi: 10.1063/1.4985282.

[30] C. Rios *et al.*, "Controlled switching of phase-change materials by evanescent-field coupling in integrated photonics," *Opt. Mater. Exp.*, vol. 8, no. 9, Sep. 2018, Art. no. 2455, doi: 10.1364/OME.8.002455.

[31] M. Fiers *et al.*, "Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach," *J. Opt. Soc. Amer. B*, vol. 29, no. 5, May 2012, Art. no. 896, doi: 10.1364/josab.29.000896.

[32] M. Le Gallo *et al.*, "Mixed-precision in-memory computing," Accessed: Jun. 2018. [Online]. Available: https://arxiv.org/pdf/1701.04279.pdf

[33] G. W. Burr *et al.*, "Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element," in *Proc. IEEE Int. Electron Devices Meet.*, 2015, pp. 29.5.1–29.5.4.