CHEMICAL REVIEWS

pubs.acs.org/CR

Review

Phase-Change Memory for In-Memory Computing

Published as part of Chemical Reviews special issue "Neuromorphic Materials".

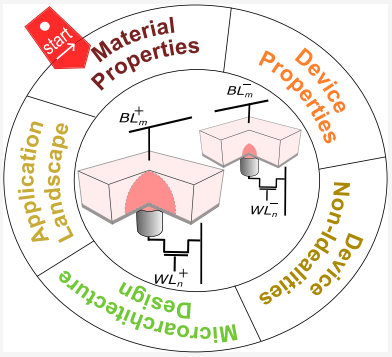
Ghazi Sarwat Syed,* Manuel Le Gallo, and Abu Sebastian

Cite This: *Chem. Rev.* 2025, 125, 5163−5194

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In-memory computing (IMC) is an emerging computational approach that addresses the processor-memory divide in modern computing systems. The core concept is to leverage the physics of memory devices and their array-level organization to perform computations directly within the memory array. Phase-change memory (PCM) is a leading memory technology being explored for IMC. In this perspective, we review the current state of phase-change materials, PCM device physics, and the design and fabrication of PCM-based IMC chips. We also provide an overview of the application landscape and offer insights into future developments.



CONTENTS

1. INTRODUCTION

In the present day, semiconductor materials are typically described in physics textbooks as having a periodically ordered arrangement of atoms, a disordered amorphous arrangement, or a mixture of both. However, this was not always the case. For a long time, textbooks mainly discussed crystals as the only viable materials for functional semiconductor devices. This belief was largely because the initial semiconductor devices, such as the transistor, were created from crystalline materials such as germanium and silicon. In the early 1960s, this belief started to be challenged. The research centered around chalcogenide glasses,[1] which consist of elements from the oxygen group of the periodic table, including sulfur, selenium, and tellurium. By introduction of thin films of tellurium alloyed with neighboring elements such as arsenic and antimony, studies began to demonstrate that these glasses could switch between resistive and conducting states, similar to the behavior of transistors. Furthermore, some materials stayed conductive until a stronger electrical current pulse was applied to revert

Received: September 2, 2024
Revised: April 24, 2025
Accepted: April 29, 2025
Published: May 22, 2025

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

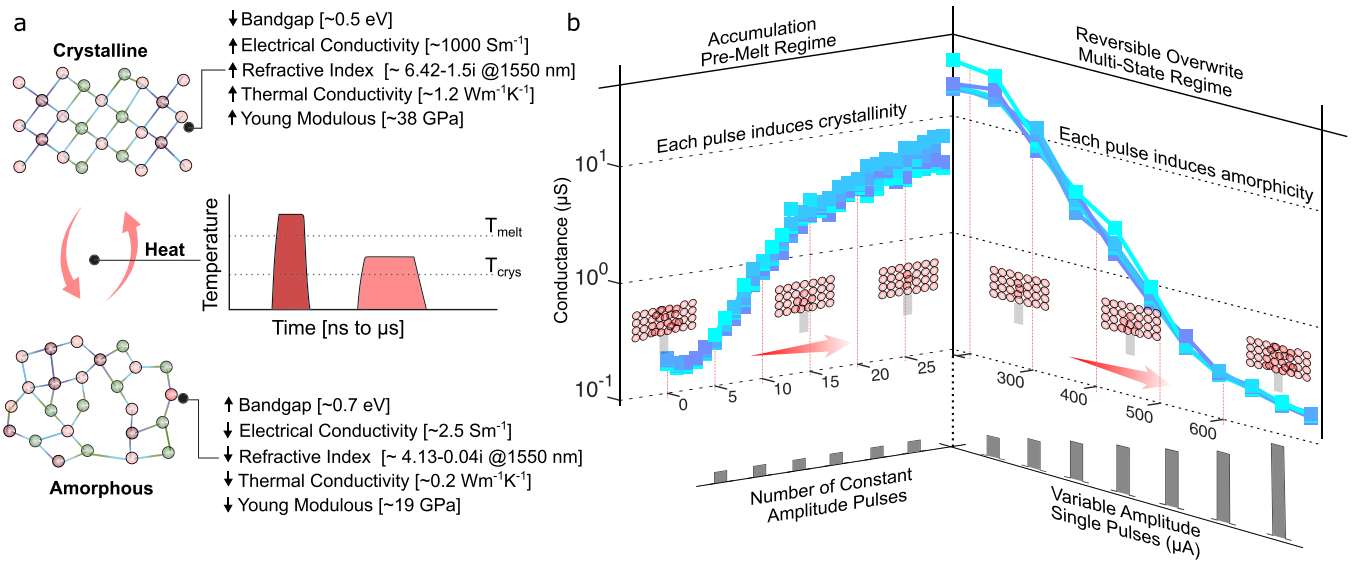


Figure 1. Phase change mechanism and operational regimes. (a) This artistic depiction illustrates heat-induced solid-state transitions between amorphous and crystalline phases within a phase change material, such as $Ge_2Sb_2Te_5$. These phases exhibit distinct electrical, thermal, and mechanical properties.[10−13] In PCM devices, nanosecond time-scale current pulses generate Joule heating. (b) Illustration of the two key physical attributes that facilitate computation using PCM devices. The right panel depicts the direct overwrite regime using melt-quench dynamics. Programming curves demonstrate achievable conductance values in response to partial RESET pulses of varying amplitudes. Increasing the amplitude of the RESET pulse results in the formation of a larger amorphous volume, largely independent of the previous phase state. The left plot illustrates the accumulative property, displaying how conductance values evolve over successive applications of a constant amplitude SET pulse. As the amorphous region shrinks due to crystallization dynamics, the device conductance gradually increases.

them to the resistive state. This became what is currently known as phase-change memory (PCM). The early discovery of PCM has made it arguably the most mature memristive technology.[2] PCM stores information by altering the atomic configurations within a nanoscale volume of material, resulting in a change in device conductance. However, unlike mainstream metal-oxide-based memristive devices, PCM employs volumetric switching instead of filamentary switching. This volumetric switching is facilitated by specific materials such as $Ge_2Sb_2Te_5$, which can be reversibly toggled between amorphous and crystalline phases with differing electrical and thermal properties. Both of these transitions are assisted by Joule heating, i.e., the transition from the crystalline to the amorphous phase relies on a melt-quench process, while the reverse transition primarily relies on crystal growth[3−5] (see Figure 1a). One of the appealing characteristics of PCM is its ability to retain stored data for long durations, typically up to 10 years at room temperature, while also enabling data to be written in just tens of nanoseconds. This characteristic continues to position PCM as a candidate for high-density nonvolatile data storage.[6−9]

Another particularly interesting emerging application, regarded as a natural transition for PCM due to its similarities with data-storage technology, is computational memory or in-memory computing (IMC).[14,15] In this paradigm, PCM devices are utilized not only for data storage but also for performing certain computational tasks. For example, by using PCM devices and two fundamental physical laws of electrical engineering, Ohm's law and Kirchhoff's current law, simple circuits can implement basic operations related to deep neural networks (DNNs) with mere femtojoules of energy. This approach eliminates the need to transfer data back and forth over power-hungry and high-latency interconnects between physically separated computing units and memory in conventional computers. Additionally, the potential for PCM devices to achieve very small cell sizes could allow large DNNs to be held entirely on-chip in a nonvolatile manner.

There are two key properties leveraged in PCM-based IMC[16] (see Figure 1b). The first property is the ability of a PCM device to store a range of conductance levels, not just two. This is important because it enhances the storage density, where one device can encode an entire synaptic weight in the case of DNNs. Programming a continuum of conductance states in devices is easily achieved by varying phase configurations (or the amount and geometry of the amorphous volume in an otherwise crystalline phase-change material) through the application of partial amorphization pulses. The second property is the accumulative property resulting from crystallization dynamics. In this scheme, the device conductance evolves according to the number of (constant amplitude) crystallization pulses that encode a computational problem with the result of the computation stored in place. This property allows for fine adjustment of the conductance levels in the devices and is essential, for example, in emulating critical dynamics required in certain computations. These capabilities, combined with a scalable structure, termed a crossbar, that can be integrated readily using already established back-end-of-the-line (BEOL) processes present another key advantage. Through massive parallelization of computations, many distinct computations on the same data, the technology allows for nontrivial reductions in the time and space complexities for certain computational tasks.

In this review, we aim to provide readers with a comprehensive account of PCM-based IMC technology. The review is divided into five parts. The first part offers an overview of the history of key advancements in materials, device concepts, and chip designs that have brought PCM to its current state. The second part delves into the material science and physics of phase-change materials, highlighting the key properties that make these materials such fascinating

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

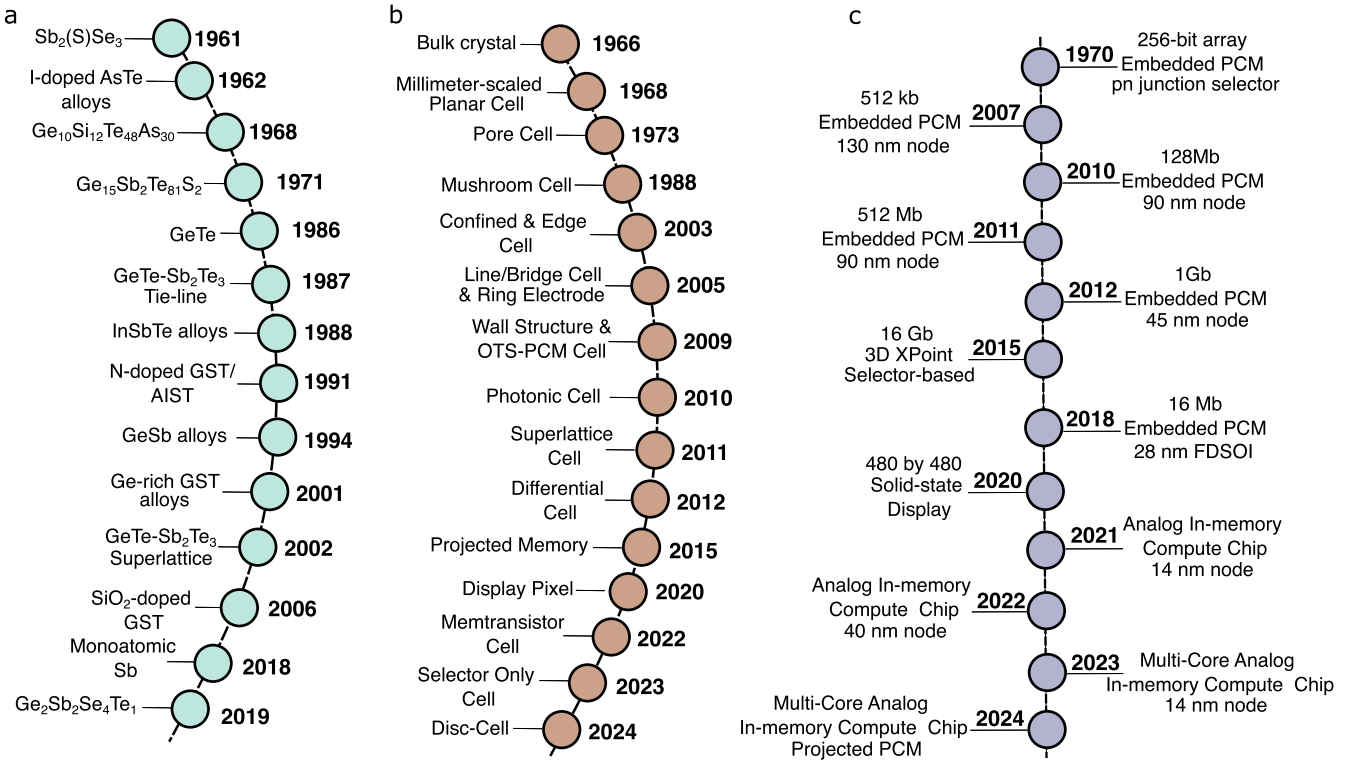


Figure 2. Historical overview of PCM. Timeline plots illustrate the evolution of PCM (a) materials, (b) devices, and (c) chips.

material systems. The third part explores the key device properties that underpin the capabilities and performance metrics of PCM devices. The fourth part compares PCM-based IMC chips with standard memory chips, highlighting the key differences (and thus the research opportunities). The fifth part surveys various applications that leverage PCM-based IMC. Finally, we present an outlook on the opportunities and challenges.

2. PART 1: EVOLUTION OF MATERIALS, DEVICES, AND CHIPS

2.1. Materials Engineering

The early observations of phase-change phenomena in solid-state electrical devices can be traced back to experiments with stibnite[17] (Sb_2Se_3) and ternary glass compositions[18] of As−Te−I (see Figure 2a). However, these materials were hindered by a limited number of switching cycles, presumably due to the sublimation and decomposition of elements due the high operational electrical currents.[19] A breakthrough occurred in 1968 with the observation of stable switching in As−Te−Ge glass, specifically within the composition of $Si_{12}Te_{48}As_{30}Ge_{10}$, which demonstrated a notable memory effect.[1] Minor modifications to this composition were found to sustain a low-resistance state after switching, maintaining this state even in the absence of an applied voltage. From the early 1970s through the mid-1980s, research focused on Te-based eutectic alloys with low melting points, such as Te−Ge and Te−Sb, which were favored due to their ability to easily transition into an amorphous state. Among these, GeTe was notable for its fast crystallization and large properties contrast,[20] yet its substantial volume expansion of 10% presented challenges for enduring multiple phase transitions. This prompted a shift toward exploring GeTe-based compositions within the GeTe−Sb_2Te_3 tie-line,[21] including variants like $Ge_{14}Sb_{29}Te_{57}$ (commonly referred to in atomic fraction ratio as $Ge_1Sb_2Te_4$), $Ge_{19}Sb_{25}Te_{56}$, $Ge_{22}Sb_{22}Te_{56}$, and $Ge_{25}Sb_{18}Te_{55}$. These compositions undergo a two-step crystallization process, resulting in a ∼4% volume change.[22] The transition involves progressing from the amorphous to the metastable cubic phase and subsequently to the stable hexagonal phase. $Ge_{22}Sb_{22}Te_{56}$, commonly referred to as $Ge_2Sb_2Te_5$ (GST),[23] offers good amorphous-phase stability with activation energies exceeding 2.2 eV, making it a preferred material for data storage applications.[21] Further exploration into another chalcogenide family, centered on the Sb_xTe_y eutectic alloy and commonly denoted as doped SbTe compounds,[24,25] involved doping with In (e.g., $In_x(Sb_7Te_3)_{1-x}$) or with Ag and In (e.g., $Ag_xIn_y(Sb_{70}Te_{30})_{1-x-y}$, also referred to as AIST). These compositions exhibited growth-dominated characteristics, contrasting with the nucleation-driven tie-line compositions, and swiftly garnered considerable attention for optical data storage. The role of individual elements also became clearer with such studies. Sb-rich compositions enabled fast crystallizing materials, while Ge-rich (Te-poor) alloys showed higher thermal stability against crystallization.

In the 2000s, as interest in electrical memory devices surged, the focus shifted toward engineering materials capable of low programming currents, higher retention (amorphous phase stability), enhanced endurance, and faster crystallization speeds. Achieving low programming currents and higher retention involved increasing the electrical resistivity of the films, which was accomplished through Ge-rich compositions or doping with elements like[26−28] N, C, O, and dielectrics such as SiO_2,[29] as well as reducing the physical thickness of the material.[30] It is worth noting that since dopant concentrations could reach nontrivial amounts (up to 15 at. %), the term doping is often replaced by alloying in the context of phase-

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

change materials. Improving endurance has involved developing materials with minimal volume changes during phase transitions, as well as compositions with reduced elemental and phase segregation propensities. Even monatomic composition, specifically using only Sb, has been explored as a phase-change material.[31] Techniques like templated crystallization from nanoparticles like Au and substructure stabilizing dopants such as Sc have demonstrated improvements in the crystallization switching speeds.[32] More recently, chalcogenide-based materials have garnered attention as promising materials for integrated optical memories and reconfigurable photonic devices. This interest arises from their unique combination of large optical contrast (with Δn ranging from approximately 0.5 to 3.5). The requirement to reduce excessive loss from free carrier absorption has spurred interest in newer compositions such as $Ge_2Sb_2Se_4Te_1$,[33] and has renewed the interest in Sb_2Se_3 and Sb_2S_3 materials.[34] Significant advancements have been made in the high-throughput screening of phase-change materials. Unlike traditional sequential synthesis, where bulk materials are synthesized, deposited, and characterized separately, these methods enable combinatorial synthesis, i.e., the deposition of compositional gradients on a single substrate, also allowing for in situ characterization of the material properties. The approach utilizes controlled codeposition from separate targets onto temperature-controlled substrates with independently operated shutters.[35] Additionally, closed-loop autonomous systems that integrate in situ real-time measurements with machine learning have been developed to accelerate discovery of newer materials.[36]

2.2. Device Engineering

Comparable to the first functional transistor, the electrical configuration of early PCM devices involved bulky electrical probes, typically featuring one fixed and one mobile probe, making lateral contact with the chalcogenide crystals (Figure 2b). Subsequently, investigations into thin-film-based phase-change films commenced. Among the initial realizations were planar and pore-type structures,[37] which exhibited reversible switching characteristics. It became apparent early on that the energy needed to switch PCM was directly proportional to the volume of phase-change material,[38] and contact areas to the electrode, prompting efforts to minimize feature size. Since the first memory chip was reported,[39] PCM devices have decreased in size by a factor of 1000, and their switching currents have also been reduced by a factor of 1000. Among the first successful device designs were mushroom-type[40] and confined-type[41] devices, which were made at the 180 and 230 nm technology node. These device designs provided vertical integration and became a starting point for two modern-day approaches in PCM device engineering: contact minimization and volume minimization, respectively. The advent of electron lithography enabled the creation of small contacts between electrodes and ultrathin phase-change films, achieving cross-sectional areas as small as 225 nm^2 in the early 2000s. This development gave rise to the formation of line or bridge cell structures.[42] However, due to their lateral span, these structures could not be integrated with high density, limiting their applicability to serving as test structures for material screening and fundamental device studies. In the pursuit of reducing programming currents even further, newer device structures were introduced. These included designs utilizing ring-type electrodes[43] and wall-type structures.[44] The former employed a bottom electrode composed of a thin film of metal arranged in an annular shape, surrounding a core made of dielectric material instead of a single solid metal. On the other hand, the latter design aimed to generate Joule heating not within the phase-change material itself but within the bottom electrode and then transfer it to the phase-change material. This approach led to the bottom electrode being commonly referred to as a 'heater'. Following these approaches, some specific modern-day device structures exhibit programming energies in the tens of femtojoules via extreme volume/contact scaling.[45−48] In efforts to improve the integration density, a thin-film two-terminal Ovonic Threshold Switch (OTS) as the selector of a PCM device was soon demonstrated.[49]

Commercial PCM products notably utilized device designs established before 2010. However, over the past decade, there has been a surge in the development of new device designs that can be tailored for emerging applications. These include functional on-chip photonic[50,51] PCM devices, as well as microheater-based structures enabling solid-state phase-change display pixels[52] and RF switches. Additionally, significant advances have been made in computational PCM unit cells, incorporating multisynaptic devices[53] and employing differential configurations.[54,55] Novel device structures, such as the projected-type, which incorporate a noninsulating layer to redirect current flow around the amorphous volume, have been developed to enhance computational precision.[56,57] Furthermore, structures like memtransistive devices, designed to execute more sophisticated computations, have also emerged.[58] Significant advancements have also been made in enabling devices with ultralow programming currents, such as superlattice structures formed by stacking chalcogenide films,[59] and the disc-type device, which confines the active switching region to a compact geometry, thereby enhancing Joule heating efficiency and thermal confinement.[60] Furthermore, selector-only devices have been introduced that use chalcogenide materials that do not rely on phase transitions induced by Joule heating for conductance modulation. These devices combine memory and selector functionalities into a single unit[61] and due to reduced thermal crosstalks, promise a scalability that surpasses the OTS-PCM cells.

2.3. Chip Scale Demonstration

The earliest lab-scale demonstrations of PCM chips occurred in 1969 with a passive crossbar[19] based on an As−Ge−I alloy (see Figure 2c). By 1970, a 256-bit array of amorphous semiconductor memory cells, connected in series with silicon p−n junction diodes to minimize sneak currents, was developed.[39] However, the volume of phase-change material required for device operation was too large, resulting in a significant power consumption during programming operations. Subsequently, in 1978, a 1024-bit PCM chip was demonstrated. Although the programming voltages and currents in this chip were lower than those in earlier attempts, they still fell short of being competitive with the memory technologies of the time. From the 1980s to the early 2000s, attempts to develop reliable PCM cells have faced significant challenges, leading to device degradation and operational instability. Consequently, interest in electrical memory cells using phase-change materials had gradually declined. However, since the 1990s, phase-change materials have found widespread use in optical memory devices and continue to serve as information storage medium in CDs, DVDs, and Blu-Ray disks.

The success of optical storage using phase-change materials sparked renewed interest in PCM in the early 2000s.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

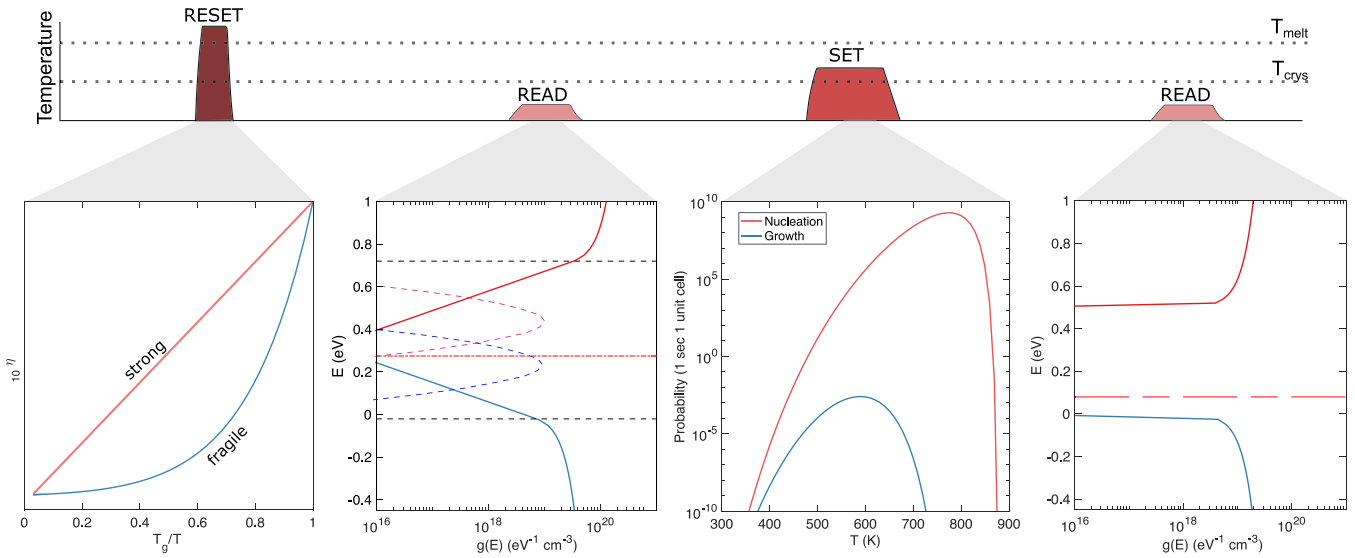


Figure 3. Conceptual illustrations of programming and read operations in PCM. (a) A RESET operation transitions the PCM device to a low conductance state by heating the phase-change material above its melting temperature, followed by rapid cooling. The high fragility in viscosity induces amorphization. (b) A low conductance value is measured during the READ operation, which can be attributed to charge transport mediated by the amorphous phase's defect states. These are shown as Gaussian-distributed profiles for acceptor and donor-type trap states in the energy band diagram ($g(E)$ represents the density of states). Note that the Fermi level is pinned in the middle of the bandgap due to the equal concentration of the donors and acceptors states. (c) A SET operation transitions the PCM device to a high-conductance state by crystallizing a previously amorphous region by heating the phase-change material above its crystallization temperature. Material and temperature-dependent nucleation and growth rates, as illustrated in the panel determine the crystallization speed. (d) The higher conductance is measured during the READ operation, which can be attributed to the disappearance of the localized traps states.

Companies such as Intel, Samsung, STMicroelectronics, Micron, and SK Hynix obtained licenses to begin manufacturing their own PCM chips of various sizes. In 2008, Numonyx, a memory company formed by Intel and STMicroelectronics and later acquired by Micron in 2010, introduced the first PCM product consisting of 128-Mbit chips in the 90 nm CMOS process.[62] Meanwhile, in 2007, Hitachi demonstrated[63] a 512-kbit memory in 130 nm CMOS technology. Micron also introduced a 45 nm 1-Gbit PCM chip in 2012, supplied to Nokia for integration into mobile phones, but it was withdrawn in 2014.[64] In 2011, IBM demonstrated 512-Mbit 2bit/cell multilevel memory cells 90 nm,[65] followed by 4bit/cell in 2013.[66] A significant advancement in PCM technology occurred with the announcement of 3D Xpoint memory by Intel and Micron.[67] This technology supposedly utilized a Ge-rich phase-change alloy as the storage component and another chalcogenide glass as the selector element. The technology was introduced in 2015 (although only commercialized in 2018) under the Intel Optane brand, offering low-latency, low-capacity nonvolatile memory options ranging from 16 to 64 GB. The production of 3D Xpoint memory, however, ceased in 2022, supposedly owing to an inadequate market need.[68] In 2023, SK Hynix demonstrated its version of a four-tier 3D Xpoint memory.[69] STMicroelectronics also commercially introduced a chip manufacturing process based on 28 nm fully depleted silicon-on-insulator (FDSOI) for Ge-rich phase-change alloy-based embedded PCM, for automotive, microcontroller applications.[70] In the realm of computational PCM, IBM demonstrated IMC cores based on 256 × 256 and 512 × 512 crossbar arrays[71,72] in 14 nm technology, designed to perform deep neural network inference. These cores utilized doped GST-based mushroom-type PCM devices. In 2022, TSMC introduced their AIMC chip in 40 nm.[73] Subsequently, IBM presented a fully integrated chip featuring 64 AIMC cores interconnected through an on-chip communication network, which also included digital activation functions and additional processing capabilities.[74] An alternative version of this chip with an analogue communication fabric was also demonstrated.[75] In 2023, further advancements were made with demonstrations[76] of AIMC cores based on mushroom-type projected PCM devices.

3. PART 2: MATERIAL PROPERTIES

Three key characteristics make chalcogenide phase-change materials functional. First, there is a significant contrast in properties between the amorphous and crystalline states. Second, the phase transitions occur within time scales that are relevant for technological applications. Third, and somewhat counterintuitively, these phases exhibit high stability at room temperature despite the rapid transitions. While many materials, including prototypical dielectrics and metallic alloys, undergo solid-state transformations, they typically lack one or more of these critical features. This section explores the essential material properties that enable the unique functionalities of phase-change materials.

3.1. Property A: Glass Formation Ability

The unusually significant nonlinear temperature-dependent viscosity $\eta(T, t) \propto \exp\left(-\frac{Q}{k_B T}\right)t\Big|_{T=T_g}$ plays a key role in the amorphization (vitrification) in phase-change materials (see the first panel in Figure 3). Here, Q and T_g are the material-dependent activation energy of viscosity and the glass-transition temperature, respectively, and k_B is the Boltzmann constant. Generally, during melt-quenching, as the temperature of the melt decreases, the viscosity η increases (in other words, the atomic diffusivity decreases), reaching a point where the structure can no longer keep up with the temperature change,

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

resulting in a glassy structure. A high viscosity in the supercooled liquid is a fundamental requirement for glass formation. Good glass formers are strong liquids that exhibit resistance to structural changes with temperature due to strong interatomic interactions (e.g., covalent bonding), leading to minimal reorganization over a wide temperature range. Viscosity, thus behaves in an almost Arrhenius fashion, increasing by several orders of magnitude over a large temperature range. Examples of good glass formers include inorganic materials such as oxides, like silica. In contrast, liquids with softer interatomic interactions, such as organic polymers (with, e.g., ionic, van der Waals bonding) have high fragility, causing viscosity to marginally change near melt-temperature, but change rapidly by several orders of magnitude over a small temperature range near T_g. Interestingly, although inorganic, phase-change materials behave similarly to organic liquids and polymers in their glass formation ability, in that $\eta(T_g)$ is very nonlinear.[77] This characteristic is typically measured in terms of fragility $\left(m = \frac{\partial \log \eta(T)}{\partial(T_g/T)}\Big|_{T=T_g}\right)$, and phase-change materials have $m \rightarrow 100$ (e.g., 90 for liquid GST), much higher than strong liquids ($m < 20$, e.g. SiO_2).

This behavior is associated with fragile-to-strong transitions, which are linked to temperature-dependent reordering in the atomic configurations within the supercooled liquid, resulting from more mobile atomic arrangements to rigid ones. Indeed, measurements have identified and quantified liquid−liquid phase transitions in chalcogenide materials.[78,79] Without these transitions, phase-change materials would not form glass at practical cooling rates. Furthermore, at elevated temperatures (around T_{crys}, which is the crystallization temperature), high fragility enhances atomic mobility in the amorphous phase, leading to rapid crystallization within nanoseconds during SET operation. For reference, the crystallization temperature (T_{crys}) for GST is 165 °C, and in other prototypical GeTe-Sb_2Te_3 compositions, T_{crys} increases with the GeTe content,[21] rising from 90 °C for Sb_2Te_3 up to 190 °C. As a comparison T_{crys} of SiO_2 is 1200 °C. Conversely, the requirement for a strong-to-fragile transition significantly restricts atomic mobility during read-out operations (below T_g), thereby slowing down crystallization. This ensures the desired retention of the amorphous phase (in GST, approximately 10^8 s at 300 K). Heuristically, the stability of the amorphous phase is expressed as $t_{failure} \propto \exp\left(-\frac{E_{crys}}{k_B T}\right)$, where E_{crys} is the activation energy of crystallization. Typically, eutectic alloys exhibit poor retention due to their low T_g, generally between 0.33 and 0.66 of the melt temperature.[80] In compounds like GST, doping with Ge and dielectrics has been shown to raise the activation energies[81] from 2.2 to 3.1 eV, roughly equivalent to a 50 °C increase in the crystallization temperature.

The rapid crystallization behavior of phase-change materials at elevated temperatures poses a significant challenge for characterizing their liquid-state properties. The supercooled liquid regime ($T_{melt}-T_{crys}$) is often inaccessible using standard techniques for extracting η data. This also makes accurate determination of T_g difficult. The limitation has led to the use of unconventional methods, such as ultrafast differential scanning calorimetry,[77] and analytical data-fitting using transmission electron microscopy measurements on crystal growth.[82] However, there are limitations to these approaches. For example, phase-change materials exhibit a breakdown of the Stokes−Einstein equation[83,84] in their viscosity-diffusivity relationship; the inverse proportionality decouples near T_{melt} and room temperature. This can impact the interpretations made from the fitted traces across a wide temperature range.

3.2. Property B: Crystallization Kinetics

As previously discussed, phase-change materials are marginal glass formers. This observation is also consistently made in the low values of the Turnbull parameter ($T_{rg} = T_g/T_m$) that establishes an inverse empirical relation to the nucleation rate. T_{rg} varies between 0.45 and 0.55 for different phase-change materials,[23,85] implying the materials have a high degree of propensity for crystallization. Indeed, crystallization can occur so swiftly that achieving a stable amorphous state necessitates cooling rates between 10^9 and 10^{11} K/s to prevent crystallization. Due to the inherently low thermal conductivity of phase-change materials, rapid cooling is achievable primarily by efficiently transferring heat to adjacent materials. The efficiency of this quenching process improves with an increase in the surface-to-volume ratio of the phase-change region and thus in nanoscaled devices.

The crystallization kinetics in PCM devices at elevated temperatures can be driven by either nucleation or growth processes (see the third panel in Figure 3), and this area has been and continues to be a subject of intensive investigation. Nucleation involves a stochastic process where a crystalline nucleus gradually attains a critical size at which it becomes stable and can grow instead of dissolving. As a consequence, nucleation is expected to occur with equal probability or homogeneously in all parts of the volume. The attainment of this critical-size nucleus requires an incubation period. The size of the critical nucleus is influenced by temperature and is determined by the difference in bulk free energy between the amorphous and crystalline phases (which reduces the critical size as it increases), the interfacial energy density between these phases (which increases the critical size as it increases), and the strain energy (which increases the critical size as it increases).

On the other hand, crystal growth (i.e., the addition of more and more monomers) occurs once the nucleus reaches this critical size and is a deterministic process. Crystal growth velocity is strongly temperature-dependent and is governed by two opposing factors: the difference in free energy between the crystalline phases and liquid, which increases growth velocity as it rises, and the viscosity of the material, which reduces growth velocity as it increases. Thus, the amorphous volume takes the expressions, $u_a = u_{a,o} - \int_{t_0}^{t_x} v_g(T)\, dt$, where $u_{a,o}$ denotes the initial amorphous thickness, $v_g = v_{g\infty} \exp\left(-\frac{E_G}{k_b T}\right)$ is the temperature-dependent growth velocity. Depending upon whether nucleation or growth exercises dominant control over the phase transformation process, phase-change materials become differentiated into nucleation-driven and growth-driven.[86] Examples of nucleation-dominated phase-change materials include $Ge_2Sb_2Te_5$, and $Ge_4Sb_2Te_2$, while Ag and In doped SbTe, GeSb, GeSnSb, and $Ge_3Sb_6Te_5$ compositions are examples of growth-dominated systems.[4]

However, on the technologically relevant scales of nanometres and nanoseconds in PCM devices, the distinction between these two conjugate crystallization models has been suggested to blur out.[87] This is because the significance of

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

nucleation diminishes, and the process of crystallization tends to be dominated by the growth of crystals; as the volume of the crystallizing material decreases, the proportion of atoms in contact with the surrounding environment, which in devices can be a crystalline phase of the phase-change material-increases. Therefore, in the compact cells of phase-change material utilized in memory devices, the likelihood of nucleation and growth from the boundary/interface between the crystal and amorphous phase is anticipated to be significantly greater than that of homogeneous-type nucleation. This behavior is further exacerbated by the fact that a substantial number of nuclei are already present due to the melt-quenched nature of the amorphous state. Within electrical devices, such nuclei grow with a peak v_g that exceeds > 1 nm ns^{-1} at elevated temperatures.[82,87]

Experimental evidence that can unambiguously distinguish the contrasting crystal growth behaviors within the same device type, such as variations due to phase configuration and length scale, has yet to be demonstrated. Additionally, because of high computational costs, it remains challenging to model nucleation and heterogeneous growth through molecular dynamics simulations over extended time scales and larger length scales. To address these challenges, newer approaches have been developed, including those based on trainable interatomic potentials using machine learning.[88,89] However, simulations covering the entire volume of the active material in real devices, while accounting for factors like preseeded nuclei, interfaces, and existing electro-thermal gradients, have yet to be demonstrated.

3.3. Property C: Electronic and Atomic Structure

In chalcogenides, strong and weak atomic bonds coexist. For example, Te atoms form 2-fold covalent bonds, creating −Te−Te−Te− coil-like chains that assemble into a hexagonal crystal structure, held together by weak van der Waals forces. These weak interchain forces arise, in part, from the presence of nonbonding lone-pair (LP) *p*-electrons on the chalcogen atoms, which contribute to the material's polarizability. Melting primarily involves breaking these weaker van der Waals interactions while preserving the stronger covalent bonds. As a result, unlike silicon—which exhibits a high melting temperature (T_{melt} = 1140 °C) due to its fully covalent network—tellurium melts at a much lower temperature (T_{melt} ∼ 450 °C). This bonding asymmetry is also manifested in the electrical properties of the melt: while many chalcogenides maintain semiconducting behavior in the molten state, typical III−V semiconductors become metallic. However, Te exhibits poor glass-forming ability because its 2-fold coordination favors rapid crystallization. To address this, Te−Sb−Ge alloys were developed, where dopants stabilize interchain cross-links through three- and 4-fold coordination. Nonetheless, binary eutectic alloys such as Te−Ge and Te−Sb often undergo phase segregation and fail to retain amorphous phases at moderate temperatures. To overcome these limitations, stoichiometric ternary Ge−Sb−Te compounds ($Ge_xSb_yTe_{1-x-y}$) were introduced, offering improved thermal stability (T_{melt} ∼ 600 °C) and greater resistance to phase separation.

The inherent structural disorder in the amorphous phase leads to significant electronic consequences. While short- and medium-range order is preserved, variations in bond lengths and bond angles introduce a distribution of electronic states that broaden the band edges. This broadening gives rise to Urbach tails, exponentially decaying states extending into the band gap, as illustrated in the second panel of Figure 3. As a result, the optical band gap in these materials is typically narrower than the mobility band gap. Compounding this effect, the amorphous phase is suggested to carry a high density of localized electronic states arising from LP *p*-electrons. These are valence-alternation pairs, comprised of overcoordinated (C^{3+}) and under-coordinated (C^{1-}) chalcogen atoms.[90] These LP-induced defect states are argued as a hallmark of chalcogenide glasses and are responsible for features which distinguish them from other covalently bonded semiconductors, such as amorphous -silicon and -III−V compounds. For example, these characteristics include the absence of a dark electron spin resonance signal, and in the middle of the band gap the Fermi level pinned near the middle of the band gap, restricting its ability to shift in response to impurity doping.[90]

In the crystalline state, technologically relevant phases of phase-change materials often exhibit high-symmetry atomic arrangements, such as cubic structures, indicative of dominant *p*-bonding. This bonding configuration is susceptible to Peierls distortions[23] and favors octahedral coordination over the sp^3 bonding typical of most III−V semiconductors. For instance, $Ge_2Sb_2Te_5$ adopts a rocksalt structure during device operation in which Te atoms occupy one face-centered cubic sublattice, while Ge and Sb atoms randomly populate the other. Within this structure, Te atoms form 2-fold coordinated chains, with Ge atoms linking these chains to Sb-containing planes. The crystalline phase also contains a significant fraction (up to 20%) of vacant Ge and Sb sites,[91] giving rise to Gaussian-shaped, acceptor-like trap states with exponentially decaying wave functions. Upon thermal treatment, vacancy ordering has been associated with increased electrical resistivity, indicative of a metal-to-insulator transition.[92] Lastly, the nature of atomic bonding in the crystalline state remains one of the most debated topics in the field. One prevalent theory proposes an entirely new bonding mechanism, suggesting that phase-change materials exhibit electron sharing and transfer characteristics distinct from the more familiar covalent, metallic, and ionic bonds as well as the weaker hydrogen and van der Waals bonds.[93,94] Conversely, an opposing theory argues that no new bonding mechanism is required. Based on electronic wavefunction analysis, the study demonstrates that phase-change materials are governed by a single, unified chemical-bonding framework consisting of conventional two-center covalent bonds and multicenter lone-pair−antibonding 'hyperbonds'.[95,96]

3.4. Property D: Structural Relaxation and Noise Characteristics

When a phase-change material is rapidly quenched from its molten state, the amorphous state thus formed is a highly stressed glass state, with atoms frozen in unstable configurations. The noncrystallinity in the phase-change material defines excess Gibbs free energy (difference between the amorphous and supercooled liquid state) in the system. As a result, the disordered amorphous structure, as a function of time and temperature relaxes toward the lower energy metastable supercooled liquid configurations.[97,98] This process is manifested markedly in the embodiment of time-dependent changes in the material properties, including viscosity, density, and the electronic bandgap. More generally, structural relaxation follows three distinct phases. The first is the onset phase wherein, for some amount of time, the properties do not change. In the second phase, relaxation is most profound, and

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

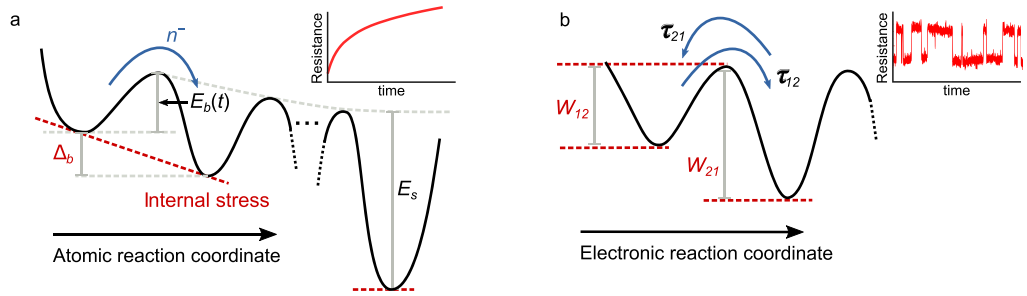


Figure 4. Structural relaxation and read noise. (a) The amorphous state is initially unstable and undergoes a series of transitions toward the more energetically favorable supercooled glass state. The activation energy for atomic reconfigurations increases as they involve the movement of atoms that have become progressively stabilized in their positions through prior relaxation steps. The inset shows an increase in device resistance over time due to structural relaxation. (b) A double-well potential model highlighting electronic transitions/switching between two energy minima, separated by a potential barrier. The inset shows fluctuations in device resistance from these transitions. It is further assumed that a distribution of double-well potentials exists to create the $1/f$ noise behavior typically observed in the devices.

the properties have been observed to change proportionally to log t. Finally, approaching the supercooled liquid, the glass reaches a saturation phase and the properties no longer continue to change. Notably, these processes are accelerated with the temperature. However, tracking structural relaxation processes through all three phases is experimentally challenging since the onset occurs in a few nanoseconds, while saturation can take significantly longer times (years) under ambient conditions. Within PCM devices, the amorphous state relaxes after RESET, and the observable metrics, such as electrical conductance and threshold-switching voltage, change due to structural relaxation. This process is commonly referred to as drift.[99]

It is commonly agreed that drift correlates with the consumption of midgap defects and bandgap widening due to local reordering,[100] via a slow evolution of the bond network toward structures with chemical order and coordination numbers similar to those of the crystalline phase. This thereby affects the read-out characteristics of PCM devices. In order to quantitatively capture such a commonly measured log t drift behavior in phase-change materials, the relaxation process most likely proceeds such that the defects with lower activation energies are removed first, followed by those with higher activation energies. In the Gibbs approach, relaxation occurs by the removal of pre-existing defect states with different activation energies. Therefore, the distribution of activation energies for the relaxation of defects serves as the parameter that tracks the state of relaxation of the material at any point in time. In an alternate approach, called collective relaxation,[101] the driving force behind this relaxation process arises from the difference between the local energy minima of adjacent states. The essential idea is that the atomic configurations that are frozen during the glass transition relax collectively (Figure 4a). As the system approaches equilibrium, this driving force diminishes, and the activation energy $E_b = E_s \times (1 - \sum(t))$ that must be overcome for the next relaxation step increases monotonically, resulting in higher energy barriers for subsequent relaxation steps. Here, $\sum$ quantifies the relaxation state of the glass ($\sum = 1$ represents an infinitely unrelaxed state, and $\sum = 0$ indicates that the system is approaching equilibrium). Notably, unlike the Gibbs approach, this model characterizes relaxation with a single activation energy that evolves over time, thus, providing a physically plausible picture of the relaxation process. More recently, using measurements on powdered samples, atomic rearrangements have been linked to the commonly observed relaxation processes in glasses—namely, β and α relaxations. The β relaxation represents a secondary, faster process that is structurally distinct from the slower α relaxation.[102]

Another characteristic of phase-change materials is that the read-out behavior is further characterized by a power spectral density that scales inversely with the frequency, leading to the so-called $1/f$ or flicker noise. The underlying mechanism that leads to flicker noise remains debated, however, double-potential-based models (see Figure 4b), typically used in the characterization of trapping-detrapping dynamics of carriers in transistor channels, have been proposed.[103,104] Here, units (carriers or atoms) can reversibly toggle with some distribution in the time constants between two energy minima (trap states) separated by a potential barrier W. The general approach to arrive at a spectrum $S(f) \propto 1/f$ is to assume that there are many Lorentzian/radio-telegraphic noise ($1/f^2$) fluctuation events, each with a relaxation time $\tau_{12/21} = \tau_0 \exp(W_{12/21}/k_bT)$, where τ_0^{-1} is the attempt frequency to surpass the barrier W. If it is then assumed that W is distributed uniformly, then this approach yields a $1/f$ spectrum. In the intermediate states, however, random-telegraphic noise, where fluctuations are more pronounced between two tractable electronic states, becomes notable. These have a strong dependency on ambient temperature and applied electric field. Read noise is also generally found to increase with the absolute resistance value. Of all the properties of phase-change materials, read-noise is the least explored. In IMC technology, as we will discuss later, the read-noise determines the ultimate limits to achievable compute precision. Consequently, understanding its physics has become more crucial than before.

3.5. Property E: Carrier Transport

In most undoped phase-change materials, current−voltage (I−V) measurements show that the crystalline state exhibits a relatively linear (ohmic) response, while the amorphous state demonstrates distinctly nonlinear behavior under applied voltage.[105,106] In the amorphous OFF state, also referred to as the subthreshold regime near room temperature, the current−voltage characteristics reveal multiple field-dependent regimes: a subexponential increase at low fields, an exponential increase at higher fields, and eventually a transition to an even steeper dependence at very high fields. These behaviors can be described using a multiple-trapping model in which charge carriers are repeatedly trapped and released between localized/bound and extended states. This framework invokes transport mechanisms such as Poole and Poole−Frenkel conduction,

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

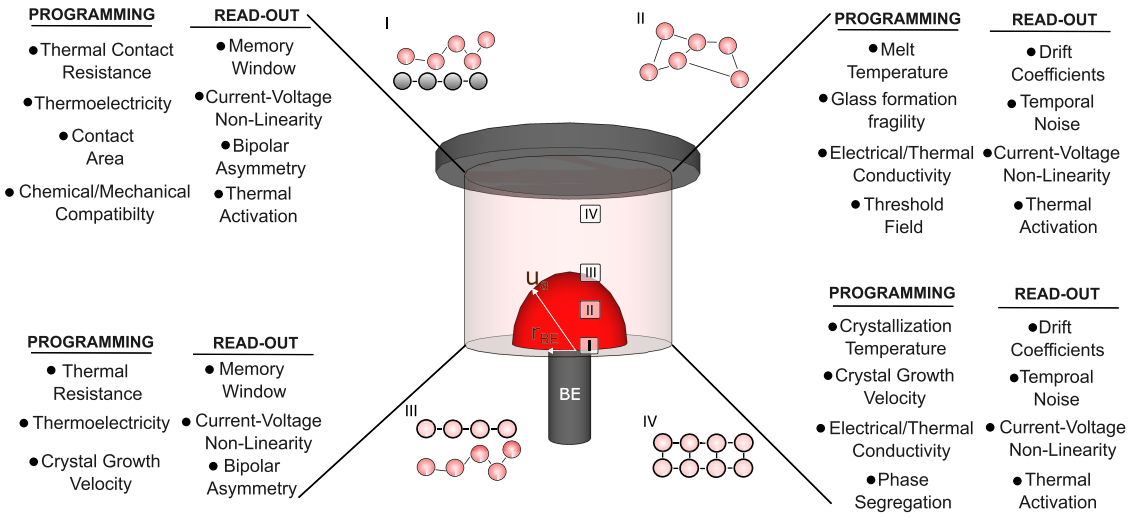


Figure 5. The various structural components associated with a prototypical PCM device. The device consists of various structural components, with conductance varying according to the geometrical scaling of an amorphous dome of size (u_a) within the bulk phase change material. Additionally, interfaces contribute electrical resistance, which also changes with the dome size. Overall, the device's behavior during programming and read-out operations is governed by various electrical, thermal, chemical, and mechanical properties of the highlighted components. When analyzing and designing computational memory devices, these parameters are carefully considered.

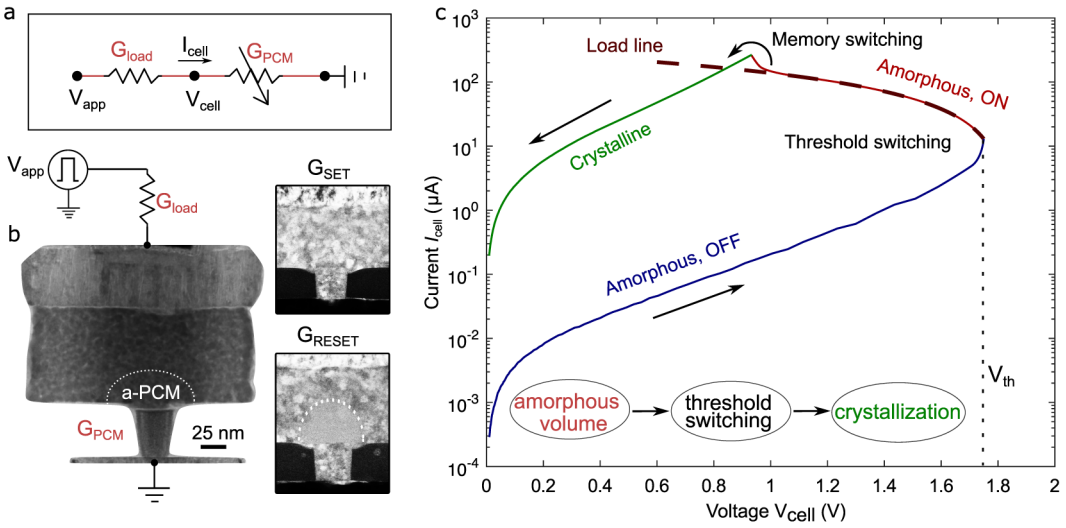


Figure 6. Threshold Switching. (a) A resistor model illustrating a PCM device in series with a load resistor (access device). The programming current and voltage dropped on the PCM device are key parameters governing the RESET and SET operations, respectively. (b) A transmission electron micrograph of a mushroom-type phase-change memory. Insets are TEM images of devices in SET and RESET phase configurations. (c) Illustration of a typical IV characteristic in a PCM device. At low applied voltages, the conductivity of the amorphous off-state is ohmic. As the voltage increases, the conductivity initially rises exponentially and then superexponentially at even higher voltages until the threshold switch occurs. Threshold switching occurs along the load line, determined by the load resistor. The surge in the current flowing through the device can crystallize the amorphous volume.

where thermal emission from ionizable defect centers, modeled as generating a Coulomb potential, plays a central role. In general, a transition from a linear to Poole dependence of current on voltage is observed at low voltages, $\ln\left(\frac{I}{I_0}\right) = \alpha \cdot V$. At higher voltages, Poole−Frenkel type behavior is observed, described by $\ln\left(\frac{I}{I_0}\right) = \beta \cdot \sqrt{V}$ Here, V is the applied voltage, I is the measured current, and the remaining terms represent other measured quantities or constants.[107,108] Modeling a two-center Coulombic potential in three dimensions[109] helps explain the observed transitions under an applied voltage (electric field (F)). This electric field, F has been understood to lower the potential barrier or activation energy (E_a), facilitating the release of carriers from bound to free states. The resulting increase in carrier density leads to enhanced electrical conductivity and the transition between different conduction regimes depends on the distance between the two centers. It is also well established that at very high electric fields and low temperatures, transport can occur via thermally assisted tunneling or direct tunneling through the potential barrier, resulting in a stronger field dependence of conductivity. At low temperatures, variable-range hopping becomes the dominant conduction mechanism, as thermal emission of carriers is

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

significantly suppressed. As in amorphous oxide semiconductors and silicon,[110] the role of tail states in subthreshold transport in prototypical phase-change materials such as GST remains an area requiring deeper understanding. Additionally, the presence of charged defect states suggests that phase-change materials exhibit weaker field-effect behavior under electrostatic gating, since the charges screen the applied electric field.[111] Lastly, note that while the above description holds for typical compositions, it may not always capture a complete picture. Materials such as AIST exhibit deviations from the behavior, for example in the Poole-Frenkel based transport,[107] underscoring the need for further composition dependent investigations of the carrier transport mechanisms.[112]

4. PART 3: DEVICE PROPERTIES

Since the atomic arrangements in PCM devices change during programming, they are structurally nonstationary devices. In other words, while the bulk and contact characteristics of standard nanodevices are fixed by fabrication, in PCM devices, these dynamically change during the device operation. Broadly, a PCM device, such as the mushroom type can be described using four components: (I) *electrode amor PCM* interface, (II) *bulk amor* PCM volume, (III) *amor crys* PCM interface, and (IV) *bulk crys* PCM volume. It has been understood that each component contributes and/or individually governs the various programming and read-out metrics (we will see in a later section that applications can leverage the PCM devices for their read-out, programming characteristics, or both). Therefore, designing a PCM device is both a material selection and an engineering problem, where many parameters are to be optimally engineered. In Figure 5, we have emphasized the key parameters related to an electrical PCM device.

4.1. Threshold Switching

Just as the nonlinearity attribute of fragility is so crucial for achieving amorphization, a nonlinear $I-V$ characteristic, is crucial for achieving crystallization. This is a critical feature of PCM technology. Upon reaching a certain voltage, referred to as the threshold switching voltage (V_{th}), the conductivity of the amorphous phase surges due to a feedback mechanism, leading to a phenomenon known as negative differential conductance,[106,107] characterized by a sudden drop in voltage (voltage snapback). For a given material, the threshold voltage scales linearly with its amorphous volume dimensions. However, the absolute values are material-dependent, specified by the band gap dependent threshold field (E_{th}), leading to $V_{th} = E_{th} \times u_a$. Binary GeSb compositions are shown to have the smallest threshold field[113] in the sub-10 V/μm, compared to GST, in the range of 50 V/μm.

If the current through the device is monitored in voltage mode, as illustrated in Figure 6, the apparent negative differential conductance often corresponds to that of the external load resistor (G_{load}) connected in series with the PCM device to cap the current, since the conductance of the PCM device becomes higher than G_{load} after threshold switching. Within an array configuration, the negative differential conductance is managed by the nonlinear selector device or transistor coupled to the PCM device. The phase immediately after threshold switching is generally termed the amorphous ON state, indicating that while the phase remains amorphous, its electrical conductivity has significantly increased. Persistent current flow through a PCM device in this state for an adequate duration leads to memory switching, meaning the device transitions to total crystallization, and the $I-V$ characteristic of the amorphous ON state converges with that of the crystalline state.

To gain insight into the significance of nonlinear field-dependent transport, consider a hypothetical situation where the amorphous state exhibits purely ohmic behavior. If the device resistance is 5 MΩ, and if a mere 5% of the power needed for melting the device (e.g., 900 μW) could result in crystallization, a voltage of $\sqrt{P \cdot R} = 15$ V would be required to induce crystallization. This would clearly make PCM not a usable technology. Another key technological relevance of the highly nonlinear $I-V$ characteristic is the near-infinite read endurance exhibited by the PCM devices. The mechanism behind threshold switching has been a topic of ongoing discussion and remains unclear. Numerous models have been suggested to elucidate this phenomenon in PCM devices; they generally fall into two categories: thermal models, which attribute the switching to an electrothermal instability within the device, and models that propose a purely electronic basis for the switching. In the absence of experimental evidence, there is increasing consensus that models based on filamentation and rupture of crystalline regions are less plausible. More recently, it has been suggested through measurements on high-bandgap chalcogenide materials that threshold switching might be linked to subtle structural rearrangements within the amorphous phase.[114] These rearrangements involve structural motifs with aligned bonds, which increase the density of states around the Fermi level and, in turn, enhance the electrical conductivity. However, these findings have not yet been experimentally validated in terms of $I-V$ characteristics in prototypical phase-change materials, and the similarities of the motifs with the crystalline lattice are still to be confirmed.

4.2. Device Failure Modes

State-of-the-art PCM devices can switch reversibly between SET and RESET states over 10^9 times. Such a number is generally achieved when their stability to electrical and thermal stresses is improved. The main mechanism that affects reversible switching between conductance states arises from both elemental and phase segregation, which occurs within the phase-change materials due to crystallization during the SET operation and melting during the RESET operation. The instability is often thermodynamic, where, for instance, Te-based amorphous eutectic compositions can separate into Te-poor and Te-rich phases within the device's active region. This is why stoichiometric compounds are preferred, as they are more stable (see Figure 7a). However, issues can still arise during melting; for example, $Ge_2Sb_2Te_5$ does not melt congruently, but phase segregates into a Te-rich liquid phase and a Ge-rich solid phase below its melting point.

These effects are intensified by the large electrical and thermal gradients within the device (see Figure 7b), leading to the electrothermal migration of elements,[115,116] which is commonly observed and modeled by

$$\frac{\partial C}{\partial t} = \nabla \cdot \left[D \left(\underbrace{\nabla C}_{\text{concentration gradient}} + \underbrace{\frac{Z^*}{kT} C \nabla V}_{\text{field gradient}} + \underbrace{\alpha C \nabla \ln T}_{\text{thermal gradient}} \right) \right]$$

In this equation, C represents the atomic concentration, D denotes the diffusivity that varies with phase and temperature,

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

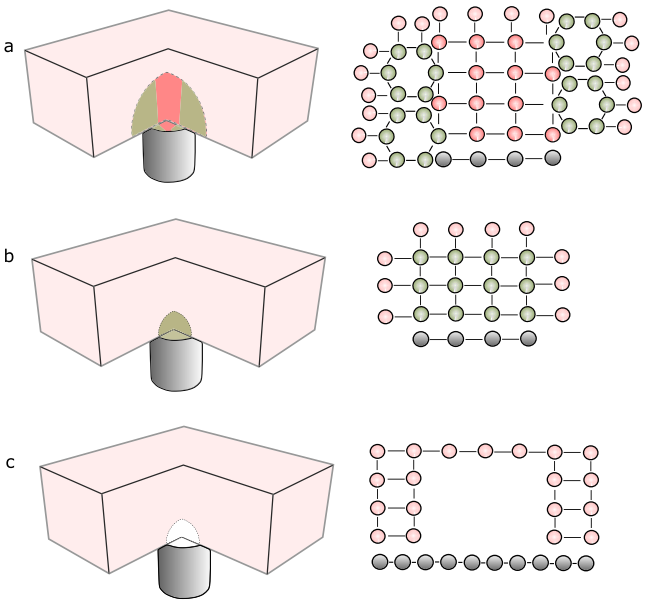


Figure 7. Modes of device failure. (a) Nonstoichiometric materials, such as eutectic compositions, can undergo phase segregation during programming operations. (b) High electrothermal fields during programming can induce ion migration, leading to elemental segregation, such as the accumulation of Sb near the electrode. (c) Cyclic volume changes in the device, due to differences in atomic density between the amorphous and crystalline states, can create localized stresses, potentially causing delamination or void formation. Mechanisms (a) and (b) often result in the device being stuck in the SET state, while mechanism (c) leads to the device being in an electrically open state.

Z^* is the effective charge, and α signifies the thermodiffusion coefficient for each species. The terms enclosed in brackets describe the atomic flux: the first term arises from Fick's first law, indicating diffusion driven by concentration gradients. The second term accounts for diffusion influenced by an electric field, while the third term captures the effect of diffusion due to temperature gradients. In $Ge_2Sb_2Te_5$, the segregation of positively charged Sb toward the bottom electrode, and negatively charged Te toward the top electrode, is a commonly noted failure-motion. This motion is attributed to the higher electronegativity (5.49 eV) of Te compared to Ge and Sb (4.6 eV and 4.85 eV). These mechanisms also explain the dependency of programming on the polarity of the voltage (i.e., the good or bad programming polarity[117]) in the electrical phase-change devices: under bad polarity, the device shows poor memory window and very limited endurance. Single elemental Sb-based PCM devices have been suggested to combat phase segregation-induced failures, however, their endurance limits are yet to be investigated.[31]

Another prominent failure mechanism in devices is related to the change in mass density with the phase transformation.[118] SET to RESET transitions can incur severe volume shrinkage (6.5−9.6%), due to the tighter atomic packing in the amorphous phase. Consequently, this causes void formation in the device, which finally limits the cyclability (see Figure 7c). In cyclic switching experiments, low programming currents, followed by device failure are commonly attributed to void formation. Material compositions with high thermal stability, and minimal-density change characteristics, such as oxygen-incorporated GeTe,[119] have been suggested to reduce such a stress-induced failure mode. These compositions, however, can be prone to phase segregation. The contact between the phase-change region and the adjacent dielectric (e.g., silicon oxide, silicon nitride) is yet another fundamental concern, since chalcogens show a strong bonding preference for each other over the dielectric for reasons relating to the minimization of strain and surface energies. This results in poor adhesion,[120] and consequently, physical peeling of the phase change film during programming, when the active region is hot. Adhesion with the chalcogenide material is typically achieved with interfacial layers including metal layers and carbon-based films.[76]

4.3. Nonidealities

A key challenge when computing with PCM devices is the temporal variations in the device conductances and the precision of programming a numeric value to a conductance state.[121,122] Consequently, this affects the accuracy of the computation.[123,124]

The temporal variations arise from intrinsic material physics. The conductance of a programmed state as a function of time can be expressed as

$$G(t, T) = \underbrace{G_0 \cdot \exp\left(-\frac{E_a(t, T)}{K_b T}\right)}_{\text{drift and temperature dependency}} + \underbrace{\eta_G(t)}_{\text{read noise}}$$

Here, G_0 is the conductance prefactor and E_a is the activation energy for charge transport. The conductance drift arises from structural relaxation in the RESET state, where the device conductance decreases with time: $G(t) = G(t_0)\left(\frac{t}{t_0}\right)^{-\nu}$ where $G(t_0)$ is the conductance measured at time t_0, and ν is the drift exponent. The drift coefficient is both material- and RESET state-dependent, typically in the range of 0.1−0.15. For example, if at t = 1 s the device conductance reads 200 μS, it will read 143 and 71 μS after 10 and 1000 s, respectively. We also note that in measurements of as-deposited amorphous films, the power law can be extended to include a virtual age t_s of the sample: $G(t) = G(t_0)\left(\frac{t+t_s}{t_0}\right)^{-\nu}$. Here, t_s represents the time or extent to which the film has drifted before measurement, during deposition and storage. Care must be taken in reporting ν values in such cases, as they can be misinterpreted when fitted with the standard power law form.

The second contribution arises from the variations in ambient temperature. Because, phase-change materials are typically low-bandgap semiconductors (0.40−0.90 eV), the variations in the ambient temperature can lead to pronounced fluctuations in the conductance. This is due to the thermally activated nature of electrical transport and is governed by the activation energy. This behavior is again state-dependent. For example, the crystalline state of GST has E_a of 0.04−0.12 eV, while the amorphous state 0.20−0.30 eV. Therefore, the amorphous phase configurations are more susceptible to temperature variations.[125]

The third contribution to conductance fluctuation is from additive $1/f$ noise. Similar to the conductance drift, $1/f$ shows a strong state dependency. The spectral density of the full RESET state can be more than 2 orders of magnitude higher than the full SET state. More generally, the magnitude of noise has a normal distribution with one standard deviation ranging between 5−10% of the mean conductance in the resting state.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

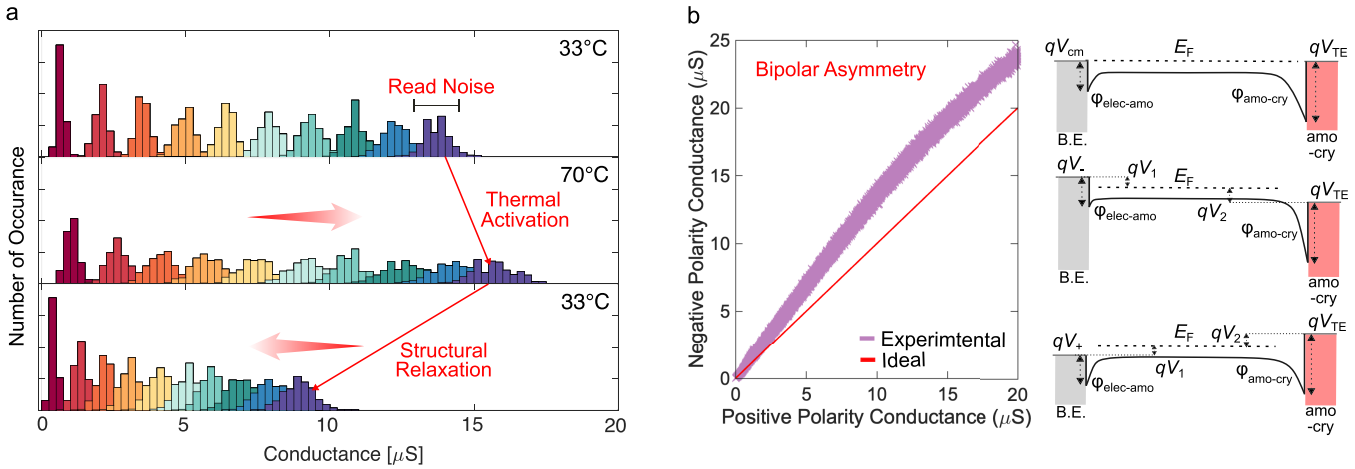


Figure 8. Nonidealities in computational PCM. (a) Device data after programming exhibits variability, depicted by wide distributions in analogue conductance states (each unique state represented in a different color) due to imprecise programming and temporal changes in conductance (top panel). This behavior is governed by additive read noise and multiplicative conductance drift. Elevated ambient temperature increases device conductivity through thermal carrier excitation and accelerates structural relaxation (middle panel). Returning to room temperature markedly reduces mean conductance values (bottom panel). (b) Conductance measured in negative voltage polarity as a function of conductance measured in positive voltage polarity. Bipolar current asymmetry in the device results in differences in the absolute values of the current based on the voltage polarity. Band diagrams of the PCM device. The device exhibits Schottky barriers for holes at both interfaces, with the amorphous−crystalline interface having a higher barrier than the electrode−amorphous interface. The bias polarity changes the band profile.

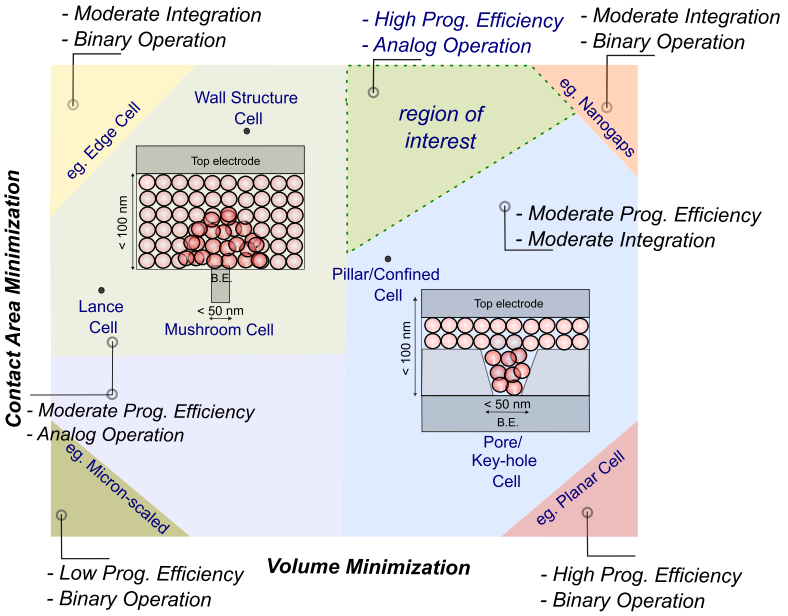


Figure 9. Device architectures. A schematic illustrating various device concepts optimized using two primary strategies: contact area reduction with the phase change material, and volumetric minimization of the phase change material.

Other sources of compute imprecision include the programming noise and the dependency of the conductance at low fields on the voltage polarity.[124] The former determines the precision with which numerical values can be mapped into the conductance values of the PCM devices. A closed-loop programming scheme is generally used in the mapping operation. Here, programming pulses are iteratively shaped to program the conductance of PCM devices within some error margins.[126] It has been shown that the $1/f$ spectral characteristic of read noise has a significant effect on the near-term temporal evolution of the conductance distribution during programming. That is, the convergence of programming to a target conductance value is fundamentally limited by read noise, which has a time-dependent distribution $(\eta_G(t))$. Thus, a noise-free programmed state is $G(t) = Z - \eta_G(t)$, where Z is the actual programmed conductance value that may or may not be within the defined tolerance margins. Due to additive read noise, the conductance values tend to be uniformly distributed around the target conductance at the time of convergence. However, over time, read-out measurements taken after convergence reveal that these states begin to diverge from their intended values. More generally, at a constant temperature, and time t_c from the last iterative programming pulse it can be shown that the ultimate precision of programming is twice the variance of $\eta_G(t_c)$.[127] All the above-described nonideal behaviors are shown in Figure 8a.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

The conductance dependence on polarity arises when $G = \underbrace{f(V, \text{sign}(V))}_{\text{current asymmetry}}$. This occurs when there are physical asymmetries in the device, such as when the charge transport barrier heights and/or the contact areas differ at the interfaces between the forcing and sensing current leads. In a GST-based mushroom-type PCM device, Schottky barriers for holes appear at both interfaces, with the barrier of the amorphous−crystalline interface being greater than that of the electrode−amorphous interface. When a positive bias is applied to the top electrode relative to the bottom electrode, the configuration resembles back-to-back diodes, with the amorphous−crystalline interface being reverse-biased and the electrode-amorphous interface being forward-biased. Conversely, this configuration is reversed when the polarity is negative. Consequently, a larger current flows when the dominant diode, namely, the amorphous−crystalline interface, is forward-biased. This scenario occurs when the applied bias polarity is negative. This dependency on voltage leads to numeric errors when performing matrix-vector multiplications (MVMs) with signed inputs (see Figure 8b).

4.4. Cell Geometry

PCM technology demands relatively high currents and power for programming because it necessitates heating of the materials to induce phase transitions. While the SET operation does require surpassing the threshold voltage, the power (and current) used in the SET pulse is typically 40−80% of that required for the RESET pulse. Therefore, the RESET pulse is the primary concern when evaluating whether the access device can provide enough current. In contrast, the SET pulse usually determines the write speed of the PCM technology. Since switching in the devices is governed by heating, two mechanisms are expected to dictate the efficiency with which heating occurs: the efficiency of Joule heat generation in the phase-change film and the resistance to the dissipation of heat, out of the phase-change film. This has led to the development of two primary strategies in device engineering: minimizing the contacts and reducing the phase-change material volume (see Figure 9). By significantly shrinking the bottom electrode or the phase-change material volume, the current density can increase, ensuring that the majority of the electric power is dissipated either at the interface between the phase-change material and the bottom electrode or within the bulk of the phase-change material. This power dissipation results in a temperature rise,[128] which is then balanced by heat being transported away from the device. Indeed, echoing Dennard's principles of transistor scaling, reducing the size of PCM cells and their contacts has resulted in performance improvement. Since the introduction of the first memory chip, the size of PCM devices has reduced by a factor of 1000, accompanied by a similar reduction in switching currents and a 10-fold decrease in switching voltages. Today, exploratory lab scale devices feature RESET currents below 10 μA at a feature size of 5 nm.[45,129] In the realm of commercial technology, where PCM devices are constructed as large arrays, the technology satisfies modest RESET currents of ∼200 μA, at a feature size of 40 nm.

Combining the simplified Fourier heat transfer equation with the Joule heating equation gives the formula $T_a = T_{amb} + \underbrace{I_{melt}^2 \times R_{ON}}_{\text{Joule heating}} \times \underbrace{R_{TH}}_{\text{heat dissipation}}$, which outlines the programming efficiency. Here, T_a represents the temperature at the 'hotspot' or actively heated region, significantly increasing from the ambient temperature T_{amb} due to power dissipation through an electrical resistance R_{ON}. This is the so-called dynamic ON resistance, which encompasses resistive contributions from the melt volume, the crystalline phase-change region, and the electrodes in series.[130] Notably, both R_{ON} and the thermal resistance R_{TH} strongly depend on the device's geometry and the phase configuration of the amorphous marks. At first approximation, I_{melt} thereby decreases proportionally from phase-change material electrode contact and phase-change material volume reductions since both these schemes increase the electrical and thermal resistances. Examples of contact area minimized devices include, mushroom-type. The contact area is more commonly defined by the critical dimension achievable with lithography. Sublithographic electrodes (defined as $\eta \times F$, where $\eta < 1$) can be defined with a thin film based edge type contacts, such as in edge-type devices,[131] as well in ring electrode type[43] mushroom-type devices. In the latter, the heater electrode consists of a thin ring of metal surrounding a center dielectric core, which reduces both the RESET current and the variability by decreasing the effective area. Volume minimized devices, include the pore type,[37] bridge or line,[46] and confined or pillar type cells.[41]

The cell designs discussed so far are known as self-heating types, primarily because the hot spot is located within the phase-change material itself. An alternative design uses a bottom electrode, referred to as a heater, to generate the heat required for the phase transition in the adjacent phase-change material.[44] In these designs, the heater primarily contributes to R_{ON}, rather than the phase-change channel. As a result, the active heating region is within the heater itself, not at the interface between the bottom electrode and the phase-change material or within the phase-change material. Examples of this architecture include lance-type and wall-structured cells, where the bottom electrode is designed to be resistive and typically longer than in mushroom-type cells. It is also worth noting that in mushroom-type devices if the phase-change material has very low resistivity or the heater has higher resistivity, the devices would not operate as a self-heating type.

Extreme contact and volume minimization, however, pose challenges for large-scale integration in BEOL. The former approach is limited by lithographic and process capabilities, while the latter approach suffers from fabrication challenges and elemental segregation. Both approaches can result in high variability and low reliability when approaching lithographically defined dimensions in the tens of nanometers, thus limiting large-scale manufacturability. Furthermore, improvements in PCM-based IMC hinge on achieving not only high device density and low programming currents but also the capability for multilevel or analogue programming.[15] This is fundamentally required for more precise mapping of numeric values into the nonvolatile conductance states of devices. Therefore, while certain device architectures may achieve programming energies in the tens of femtojoules from extreme reductions in material volume, such miniaturization can impair the analogue programming capabilities due to the minimal material available for melting. Devices featuring nanogap, edge-type, and confined structures, such as Intel Optane, exemplify this limitation and function in binary mode due to their small switchable volume.[45,131] Based on the discussion above, it could be tempting to consider that larger devices with smaller contact areas and increased volume could enable analogue programming. Yet, despite their potentially suitable volume for

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

this purpose, their larger physical size and the associated high thermal resistances could lead to slower quenching rates, challenging the implementation of precise melt-quench dynamics that is needed for analogue programming. This would be in addition to their poor integration density. In addition to optimizing programming currents and analogue programming capabilities, it is crucial to design a cell architecture with moderate threshold voltages.[113] These voltages should be substantially higher than the noise floor of circuits yet small enough to align with the operating points of individual access devices. For instance, in bridge cells, enhancing programming currents and analogue efficacy can be achieved by extending the length of the phase-change material channel. However, this extension leads to higher threshold voltages, as the required field scales proportionally with the length of the amorphous mark.

4.5. New Device Concepts

There have been two main approaches to improving PCM devices. The first approach involves material engineering, such as selectively doping elements into a host material or exploring new materials[31] like Sb. The second approach involves design engineering of the device. For example, the projected PCM.[132−134] These devices have a noninsulating projection (liner) segment (see Figure 10a), optimally engineered in

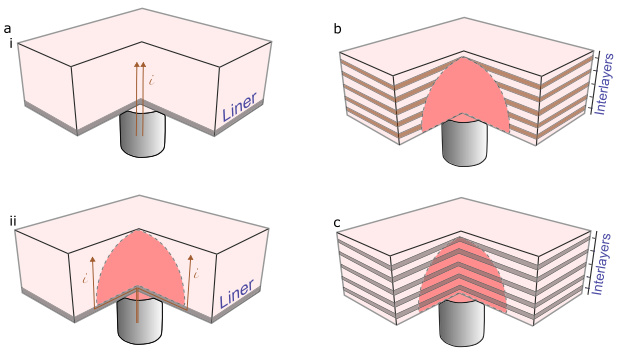


Figure 10. Schematic illustration of emerging PCM device concepts. (a) A projected-type PCM device. (i) Illustrates the phase configuration in the SET state, where the current flows from the bottom electrode into the crystalline phase-change material. (ii) Illustrates the phase configuration in the RESET state, where a substantial portion of the current bypasses the amorphous volume by flowing through the liner, thus masking the nonideal characteristics associated with the amorphous phase. (b) A superlattice-type PCM. The phase-change material consists of highly oriented thin films of two different materials that repeat periodically. In the active region, these thin films intermix during RESET, whereas during the SET operation, they are suggested to separate into distinct phases. (c) A heterostructure-type PCM. The functional layer consists of ultrathin phase-change films separated by noninsulating layers. During programming, only the phase-change layer undergoes structural changes.

electrical properties that run parallel to the phase-change material segment. During the programming process, the projection segment has minimal impact on device operation due to the highly nonlinear $I-V$ characteristics of phase-change materials. However, the projection segment decouples the read-out from the noisy electrical properties of the amorphous phase configurations of the phase-change material. Thus, in effect, the device read-out characteristics become dictated by the properties of the stable projection segment. Individual lateral (bridge cell) projected PCM devices have demonstrated in-memory scalar multiplication with an arithmetic precision equivalent to 8-bit fixed-point.[57] Similarly, individual confined-type have shown improvements in computational precision.[135] However, both of these structures pose challenges for large-scale array fabrication. Lateral devices, aside from their large footprint, are susceptible to etch damage on the sidewalls, while confined cells struggle with filling long trenches. Both issues contribute to significant device-to-device variability. Indeed for these reasons, large array demonstrations have been made with mushroom type devices: projected-mushroom type devices have been integrated onto multitile IMC chips fabricated in 14 nm CMOS technology, achieving drift and noise reduction, thus improving inference accuracies.[76] However, although mushroom type device geometry offers a large programming window and is manufacturable, the efficacy of projection itself is fundamentally lesser compared to geometries like the bridge cells, where the current flows parallel to the liner. Research is ongoing to develop manufacturable device structures that achieve also higher projection efficiencies.

Superlattice films represent another device innovation (Figure 10b). Instead of using a single thick film of phase-change material, alternating layers of two different phase-change materials, each only a few atoms thick, constitute the active volume.[59] The low cross-plane thermal and high cross-plane electrical conductivity of the films, created from van der Waals like gaps, provide a strong electro-thermal confinement effect that enhances programming efficiency.[136,137] While superlattice films were initially studied for optical disks[138] to reduce programming power, in the electrical domain, they also offer reductions in conductance drift and noise.[139] In another embodiment, the alternating layers may comprise phase-change material and noninsulating transition-metal dichalcogenide films, creating the so-called heterostructure-type device[140] (see Figure 10c). Such devices are suggested to benefit from electro-thermal confinement and improved endurance limits. The latter arises from the noninsulating layer acting as a diffusion barrier against long-range mass transport, thus preventing compositional changes in the phase-change material layers. Some examples of these periodic structures include $GeTe/Sb_2Te_3$, $TiTe_2/Sb_2Te_3$, $GeSb_2Te_4/Sb_2Te_3$, and Sb_2Te_3/GST225 stacks.

The primary challenge with projected-type devices lies in the inverse relationship between projection efficacy and memory window size since the projection layer determines the RESET conductance value.[135,141] Consequently, the most effective projected device will have a RESET conductance value very close to the SET state, which limits its functionality to binary nonvolatile states. Achieving the desired balance between reducing nonidealities and maintaining a sufficient number of distinct programmable states requires careful selection of the projection segment. Not all materials are suitable for projection layers; they must exhibit excellent compatibility with phase-change materials and BEOL processes, including chemical, thermal, and mechanical compatibility to avoid issues like interdiffusion and decohesion at the electrode and dielectric interfaces. They should also allow for integration using PCM-friendly etch chemistry, resist oxidation, and enable rapid synthesis. Currently, ultrathin carbon-based and metal-nitride films are preferred.[76] For heterostructure- and superlattice-type devices, the understanding of the fundamental mechanisms of operation remains incomplete. For instance, the reduced

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

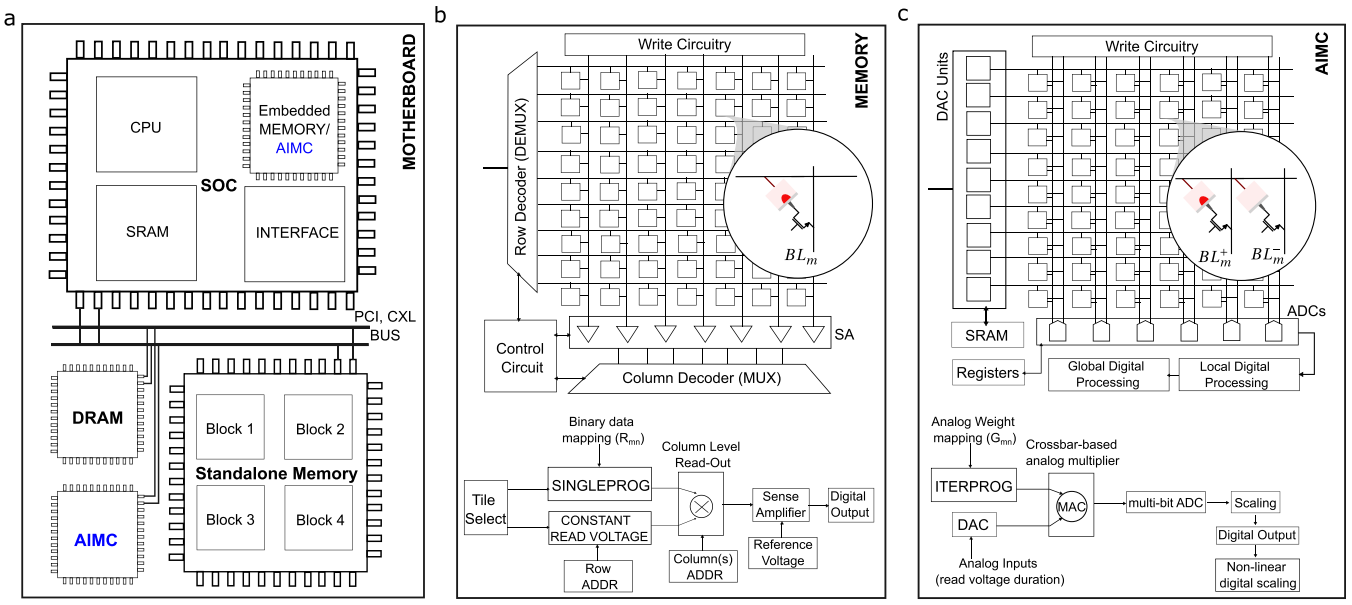


Figure 11. PCM in a computing system. (a) An artistic representation of a circuit board housing a packaged chip, such as a SoC, connected to external memory and computing via communication buses. PCM can be integrated into an SoC as embedded computational memory as well as standard memory. (b) A schematic of the key components in a PCM memory chip. The crossbar comprises one PCM device at each crosspoint that is uniquely addressed for read and write operations. Programming is typically done with open-loop single-shot pulses. (c) A schematic illustration of the key components constituting an AIMC chip. The crossbar comprises unit cells at each crosspoint, where each crosspoint may have one or more PCM devices. Multiple devices are addressed during read operations, and write operations are performed with closed-loop programming schemes.

conductance drift in heterostructure devices and the RESET conductance values can be attributed to the role of transition-metal dichalcogenide layers as projection layers. The low crystallization temperatures in superlattice films may indicate compositional changes in the amorphous volume, which can be associated with lower drift coefficients. Additionally, the poor retention from the low crystallization temperatures, and film delamination poses a challenge when integrating these devices at a wafer scale[142] and will require further investigation. Recent studies have demonstrated that superlattices based on $Ge_4Sb_6Te_7$ exhibit higher crystallization temperatures,[143] though further evidence is needed to confirm this. Additionally, superlattice compositions incorporating Sb_2Te_3, including the aforementioned material, unexpectedly show signs of lower melting temperatures. This is intriguing, as conventional phase-change materials generally have similar melting points.[144] Further investigation is required, as this phenomenon could suggest that lower programming currents result not only from thermal confinement but also from reduced melting temperatures.

5. PART 4: CHIP DESIGN AND INTEGRATION

The adoption of PCM for in-memory computations heralds a new technology expected to complement, rather than replace, existing PCM memory technology. From a system perspective, the primary difference between a memory chip and an accelerator chip lies in terms of the basic instructions supported by the chip and the application domain. However, fabrication, chip layout, packaging, and interfacing with the system are expected to remain largely similar between the two technologies. Consider a hypothetical system-on-a-chip (SoC), as illustrated in Figure 11a. The chip encompasses a CPU, local cache memory, and interface circuitry fabricated in the front-end-of-the-line (FEOL). Both memory and analogue-in-memory chips (AIMC) can be seamlessly embedded (integrated) into the chip using BEOL integration. Crucially, in this format, one enables an embedded analogue-in-memory accelerator. Alternatively, in another approach, where the SoC chip interacts with standalone components via high-speed serial computer expansion bus standards like PCI Express and CXL, standalone AIMC, and memory modules can be incorporated as devices on the main printed circuit board, namely the motherboard. In this scenario, a standalone analogue-in-memory accelerator is utilized, communicating directly with the SoC. While embedded AIMCs are constrained by chip area rather than communication latency due to their proximity to the CPU, standalone accelerators are hindered by communication delays and energy expense from data movement, but can incorporate very large compute throughput. Notably, both embedded and standalone accelerators can integrate PCM. Additionally, for neural network applications, especially those with large model parameters, a topology utilizing PCM for nonvolatile storage, as well as for AIMC can be envisioned.

5.1. Microarchitecture Differences

The microarchitecture for IMC largely resembles that of existing memory technologies.[145,146] In both cases, a memory array consists of a crossbar structure with PCM devices positioned at the cross-points of word lines and bit lines. However, there are significant differences primarily stemming from the nature of the read-out operation. In the memory applications (see Figure 11b), largely developed for binary storage, the crossbar is engineered for ultralow latencies in read and write operations. Here, a single PCM device is placed at each cross-point, and the devices are read out individually or at the word length level with a fixed read voltage. During the read operation, word lines select one or more devices in parallel along a row, and their corresponding bit lines sense their bit

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

value using a sense amplifier. During the write operation, single predefined pulses are applied with each pulse tailored for SET and RESET operations. Often a memory chip comprises several tiles, each equipped with a PCM array, a column decoder, a row decoder, and a local sense amplifier. The sense amplifier acts as a 1-bit analogue-to-digital converter, distinguishing between low and high conductance states by comparing the sensed output with predefined reference voltages. Due to the typically large read-out margins, there is no need for precise programming to a specific conductance state, thus reducing the write latency. The partitioning into tiles is necessary to maintain a small array size. This minimizes *IR* drops and capacitive effects on the word and bit lines for faster write and read-out operations. Tiles are further organized into partitions and planes, overall leading to a layout that is very similar to the banks, blocks, and page hierarchy in the standard memory microarchitecture. Since the objective is to maximize memory density (devices/bits per mm^2), more estate is prioritized for the PCM crossbars, and therefore digital peripherals such as large-footprint sense amplifiers and write drivers are shared among partitions.

For IMC, such as in the task of neural network inference, the PCM crossbars are designed to compute MVMs (Figure 11c). In performing the MVM operation $A\vec{x} = \vec{y}$, the conductance states of the PCM devices can be viewed as the elements of the matrix A of size m, n, while $\vec{x} = (x_1, x_2, ..., x_m)^T$ is the input vector, mapped to read voltage signals applied to the word-lines (rows). $\vec{y}$ is the output and represents the summed output current along the bit lines (columns). Because MVM consists of multiple scalar multiplications followed by summations—collectively known as multiply−accumulate (MAC) operations—each read operation effectively becomes a compute instruction, engaging several PCM devices in parallel. Every $A_{m,n}x_m$ multiplication operation is enabled by the Ohms law locally within each PCM device, and the additions of the partial products by the current summation law, achieved on the bit lines. It can be shown that for a crossbar, 1 [MVM]=n [MAC]=n×(m [MULT]+(m−1) [ADD])≈m×n [MULT]+m×n [ADD]=2×m×n [OP], where 1 [OP]=1 operation. Because x_m and A can take continuous analogue values, more key distinction features emerge. Every row may encode a distinct read-out signal, which can be represented either by varying amplitude or duration of the voltage signal. The latter approach uses a constant voltage within the Ohmic regime of the device, thereby negating imprecision from nonlinearity in the $I-V$ characteristics of the devices. Each cross-point will host at least two devices, each with a dedicated bit-line feeding into the analogue-to-digital (ADC) converter. This is called the differential pair scheme, enabling the encoding of both positive and negative numbers. Additionally, for each polarity, there may be more than one PCM device organized in a parallel combination so that the conductances add up. This so-called multisynapse scheme creates a wide range for mapping numeric values onto the conductance states of the devices. Furthermore, it is crucial to precisely map the numeric values onto the conductance states of the PCM devices. To that end, single-shot SET and RESET pulses are replaced with a closed-loop read-verify scheme, where each device is programmed to a precise value limited only by the read-noise, as discussed in the earlier section. In summary, within a crossbar, every cross-point hosts not a single PCM device but a unit cell that comprises multiple PCM devices, each programmed with precision. Furthermore, as the signals on the bit lines are naturally analogue, their digitization requires that the sense units are multibit ADC converters.

An in-memory computing chip, much like the standard memory chip, may comprise multiple tiles.[74] Each tile may host a PCM crossbar and several peripheral circuitries designed for self-sufficient computations to enable compute pipelining (supporting low-latency execution for a wide range of neural network architectures without introducing unnecessary redundancy). These include digital-to-analogue converter (DAC)-based programming units based on the instructions provided by the finite-state machine for programming operations, pulse-width-modulated (PWM) units for generating read-voltage pulses to the PCM array, and ADC units for digitizing the output current signals. Each tile may further host digital-processing circuitry such as fused multiply add units and registers for signal correction and performing activation operations such as ReLU. Multiple tiles may be partitioned to share additional digital circuitry units, including general-purpose microprocessors that can perform more demanding nonlinear computations such as normalization and sigmoid/tanh activations, as well as offload some of the linear operations from AIMC tiles to create a fine-grained computation pipeline for end-to-end inference. The matrix is mapped to the target values using programming techniques such as iterative programming, which involves multiple steps to achieve convergence, as we will discuss in the following section. The availability of dedicated circuits discussed so far at the tile level makes programming across the entire AIMC array in parallel possible. Specifically, the number of devices that can be programmed in parallel is determined by the number of driver units, enabling spatial parallelism along the array diagonal. Similarly, during iterative programming, all devices are programmed at each time step. The number of iterations corresponds to the time it takes for the last device(s) to converge, meaning the total programming time is determined by the time required for the entire chip.

The main challenges to overcome in PCM-based IMC chips are limited device-level precision and compute density. While this poses material challenges for PCM, it also influences the peripheral circuitry by limiting its available area while ensuring that it does not restrain precision. To improve the compute precision up to four- or five-bit fixed-point arithmetic, it is essential to minimize the temporal conductance fluctuations (such as noise, conductance drift, and temperature dependence). Promising directions for PCM include projected PCM devices and PCM devices with alternating stacks of phase-change nanolayers and confinement nanolayers. To improve the compute density, in addition to scaling both the PCM devices and the associated access transistors, high-density arrays need to be integrated at the back end of a CMOS wafer. In addition, real workloads involve a variety of different operations other than MVMs that need to be implemented in separate digital computing units. Therefore, power-hungry ADCs are needed at the crossbar outputs, which limit the energy efficiency. Moreover, a fast and flexible communication scheme together with highly efficient pipelining of the digital compute units and intermediate SRAM storage is necessary to ensure that they do not dominate the latency and power consumption.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

Table 1. Comparison between Data Storage and AIMC Technologies

technology comparison		
parameters	data storage	AIMC
read-out operation	one device is read out per column of a memory crossbar	ideally, all devices across all columns are read out in parallel, where each device contributes to the compute operation
performance metric	read/write latency	number of compute operations per Watt and mm^2
number of conductance states	typically up to ternary	a continuum number of states
conductance values	high values for SET and low values RESET	low values for SET, intermediate, and RESET
number of devices/ cross-point	one	at least two in a differential pair configuration
retention	85 °C/10 years	85 °C/day(s)
endurance	>10^9	>10^5

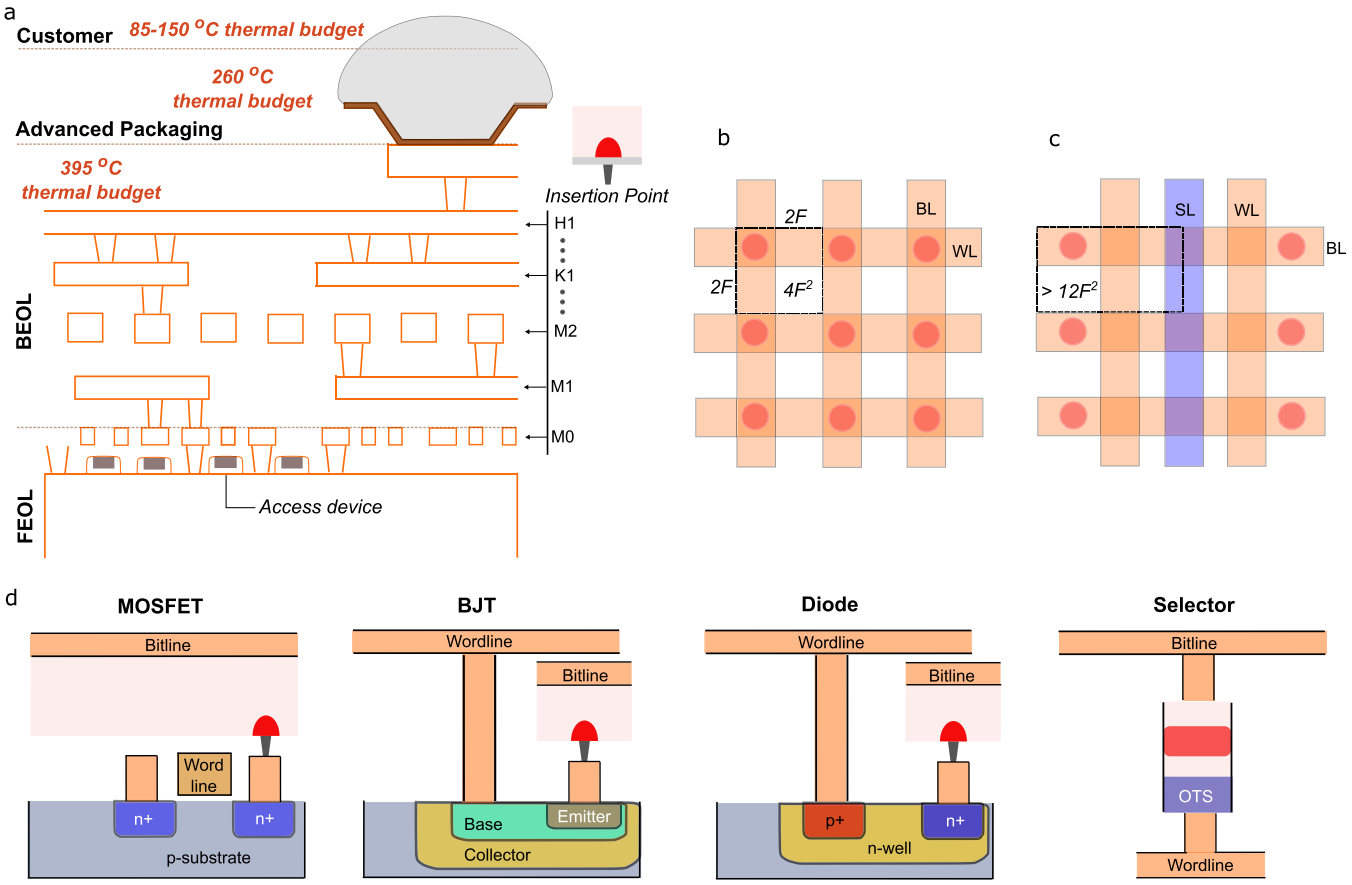


Figure 12. PCM integration. (a) An illustration of the various integration levels and their thermal constraints in chip manufacturing. In the front-end-of-the-line, transistors are integrated. In the back-end-of-the-line, PCM devices are inserted into select metal layers. Packaging of the chip onto chip carriers/modules follows this step. (b) Single selector-based PCM device. (c) Single transistor-based PCM device. (d) PCM devices integrated with various types of access devices: MOSFET, BJT, diode, and selector, such as OTS.

5.2. Differences in Device Requirements

It is important to note that, despite the similarities, data storage and analogue IMC have two different use cases, which necessitate different device specifications (see Table 1). Briefly, the primary performance metric for analogue IMC is increasing the array size so that many linear compute operations can be performed at high energy and areal efficiencies, measured in TOPSW^{-1} (tera operations per second per Watt) and TOPS mm^{-2} (tera operations per second per unit area of the chip), respectively. This involves enhancing integration density by reducing the size of selectors through lowering the programming currents and minimizing *IR* drops on the interconnects by decreasing the conductance values of the SET and RESET states. This is notably different from the data storage use case where the mandate is to read and write data quickly, thus requiring high conductance values of the SET and RESET states. Additionally, the write frequency is expected to be quite low for IMC especially in the context of DNN inference. This may significantly reduce the endurance requirement for a modest 10^5 cycles, as well as diminish the need for long-term weight retention, for example, requiring stability for only 1 day at 85 °C. Similarly, this could ease the requirement for high programming speed. However, it may be generally appreciated to be able to program the devices in sub-μs.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

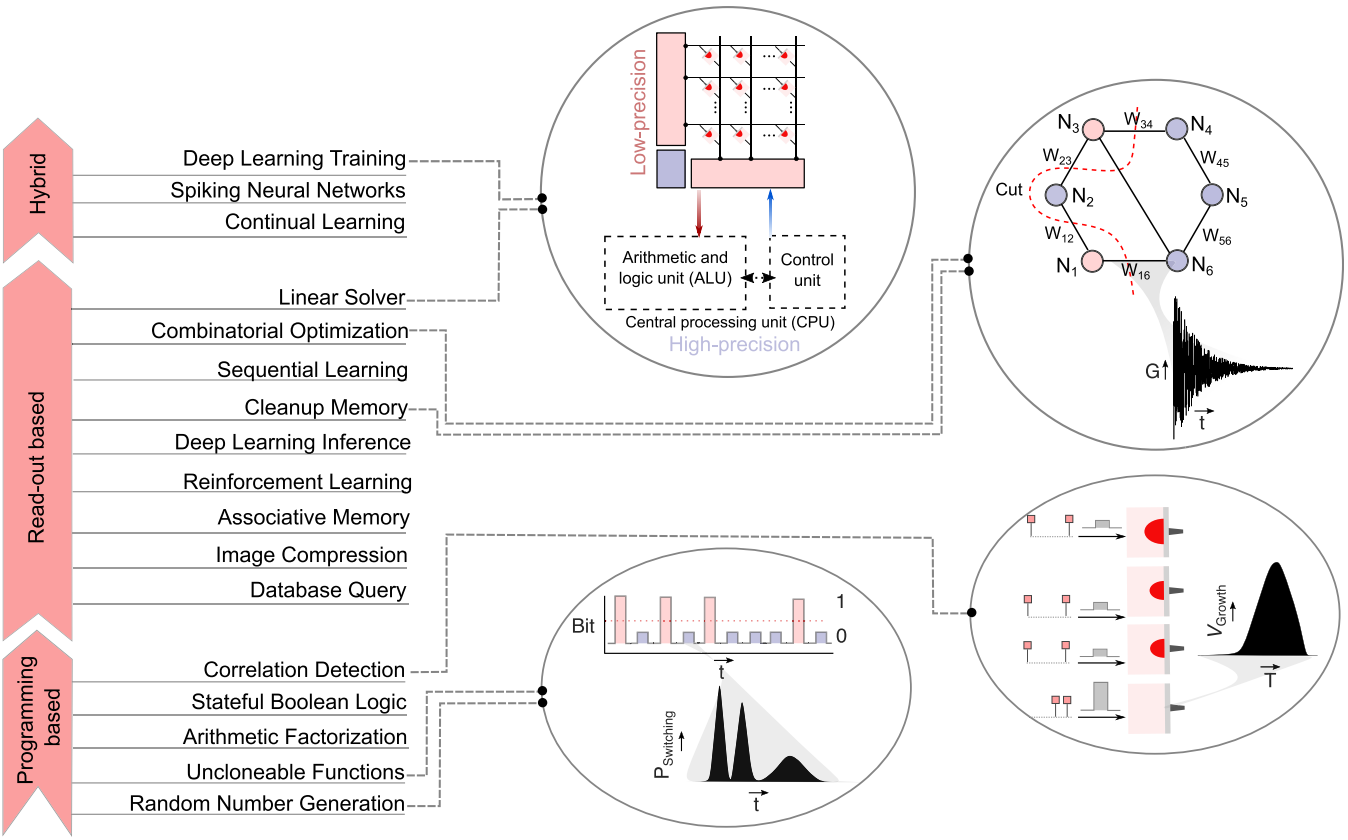


Figure 13. Application landscape. The applications are grouped into three main categories based on the mode with which the PCM device computes. These are read-based, programming-based, and hybrid.

5.3. Device Integration

It is no exaggeration to state that the effectiveness of a PCM device is intrinsically tied to its compatibility with the foundry process line (Figure 12a). This compatibility is largely determined by (a) the phase-change material's capacity to support large-scale manufacturability, ensuring chemical stability within the process-line thermal budget (typically 400 °C for BEOL), and (b) the absence of ionizable and mobile species[147] such as Ag, Al, Au, and Cu. The former, high temperature step is required for passivation of transistor gates, for example, through high-pressure deuterium annealing, to enhance the electrical stability and reliability. The latter requirement arises from the need to prevent the diffusion of conductive metals into dielectrics and the silicon channel. Consequently, materials prone to phase segregation, such as eutectic compositions, are generally deemed incompatible, as are compositions doped with mobile elements such as AIST. Furthermore, the selection criteria extend to compatibility with chip packaging. Given that a significant application of PCM is in microcontrollers, such as in boot-up memory, it becomes crucial to consider the code retention from soldering reflow (typically subject to a peak temperature of 260 °C for 2 min, according to JEDEC standard[148]). Meeting automotive specifications, which typically require two years of high-temperature data retention at 150 °C, presents a significant challenge. Materials with low crystallization temperatures, such as GST, are unable to meet these requirements. The thermal requirements also extend to the customer side, where chips must retain data for 10 years at 85 °C (although, as previously discussed, this can be relaxed to days for AIMC). This constraint limits the use of materials such as Sb_2Te_3 and monatomic Sb. Compositions like Sb_2Te_3 and Sb_2Te_1 have crystallization temperatures below[149] 100 °C, while ultrathin (approaching only 3 nm) Sb retains data for only a few seconds at room temperature, and even less so at elevated temperatures. Also note that, more generally, even for compositions with modest crystallization temperatures, within a device, the intermediate states have poorer retention than the fully RESET state. Therefore, when investigating retention, it is essential to also study programmable states that are very close to the SET state. However, certain applications that may require low retention,[74,150] such as mixed-precision linear solvers can leverage the low retention characteristics of materials with low crystallization temperatures, making them viable for use in these domains.

The crossbar array for IMC consists of metal lines/interconnects arranged orthogonally with synaptic elements located at their intersection points. Each synaptic element comprises a PCM device connected to a nonlinear access or selector device, to provide current-controlled programming and to reduce sneak path currents.[151] The choice of access device significantly impacts integration density and analogue computing capabilities. The selector should exhibit high electrical conductivity in its on-state (to sustain high programming currents) and very low electrical conductivity in its off-state (to eliminate sneak paths) and require minimal space (to enhance integration density). Broadly, selector devices fall into two categories: FEOL or BEOL integrated. Front-end integrated selectors include bipolar junction transistors (BJTs), diodes, PN junctions, and metal-oxide-semiconductor field-effect transistors (MOSFETs). BJTs and

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

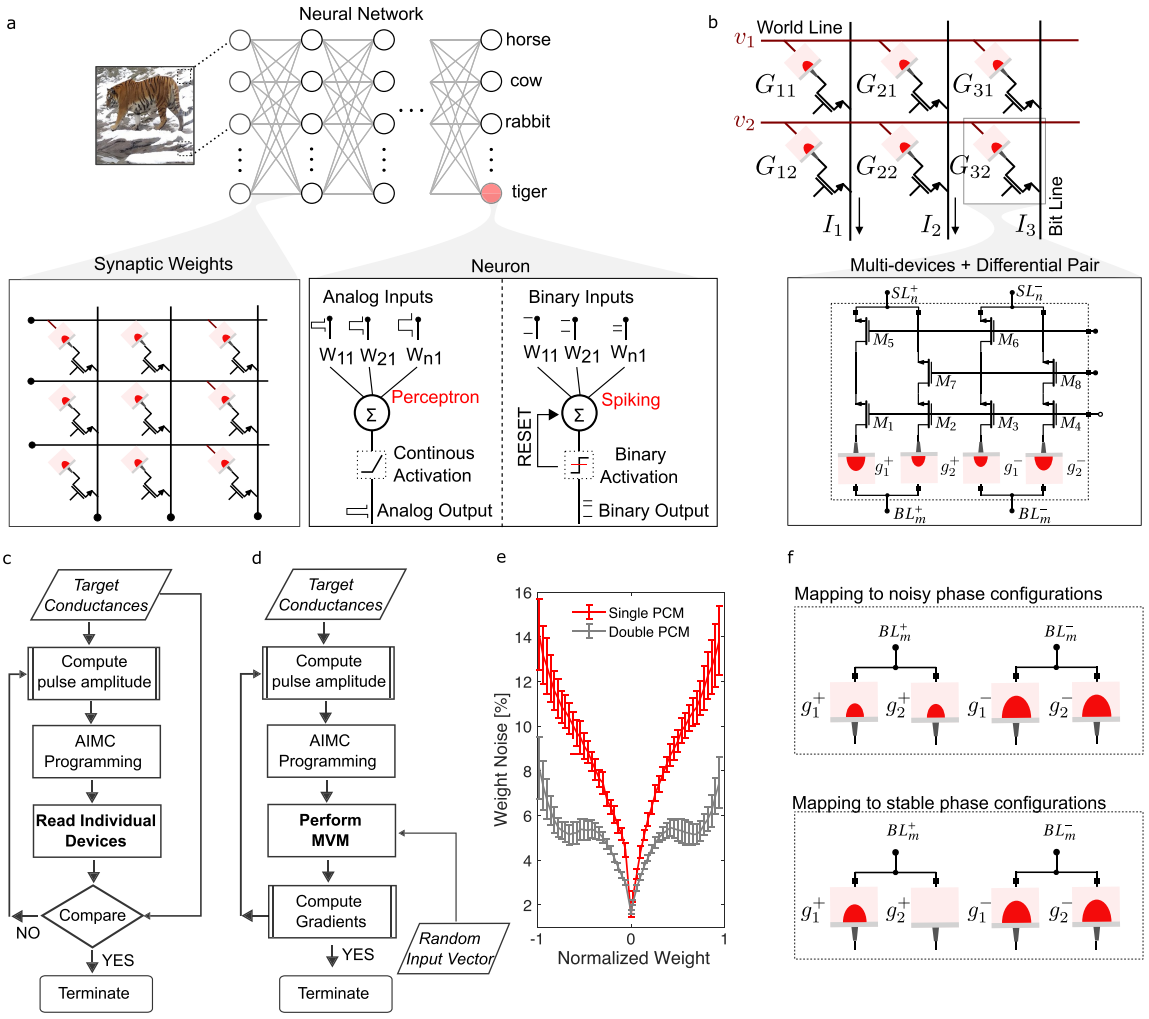


Figure 14. PCM for deep neural network inference. An illustration of an artificial neural network comprising multiple layers of synaptic weights and neurons. In a classifier network, as shown here, the task during inference is to classify input data (images) into predefined classes. The bottom panel highlights the key components of the network, where PCMs represent the synaptic weights, and the neurons can be CMOS-based, either spiking or perceptron types, depending on the network. Spiking neural networks, which are more brain-inspired, operate on spikes of asynchronous pulses for binarized data flow (with pulses that are rate or temporally coded). Perceptron-based networks, used in most neural network demonstrations, employ clock-synchronized data flow and typically operate within the analogue domain. (b) An illustration of a synaptic unit-cell. Each cell features a differential pair architecture, utilizing a pair of complementary PCM devices to encode positive and negative numbers. Multiple devices can be employed to represent a signed number, thereby increasing the numeric range for encoding and improving the signal-to-noise ratio by averaging out temporal fluctuations. (c) A flow diagram illustrating the closed-loop iterative programming scheme for mapping synaptic weight values into the conductance values of PCM devices, with error minimization performed at the level of individual devices. (d) A flow diagram depicting the closed-loop gradient-descent programming scheme for mapping synaptic weight values into the conductance values of PCM devices, where error minimization is achieved through the simultaneous read-out of a collection of devices. (e) Mapping synaptic weight values using amplitude-controlled pulses is prone to errors due to device conductance fluctuations. Employing two PCM devices per synapse helps to reduce this error. (f) Visual depiction of weight mapping. The top panel illustrates the use of intermediate states in mapping a numeric value. This approach is more susceptible to noisy fluctuations. Maximizing the number of devices in the SET and RESET states provides a means to avoid the noisy intermediate states, as is shown in the bottom panel.

diodes, similar in structure, differ primarily in junction doping and the number of contacts/terminals. Two-terminal polysilicon diodes offer the smallest layout, occupying a $4F^2$ area, where F is the minimum lithographic feature size . While being three-terminal, the base contacts in BJTs can be shared across multiple cells, enabling a reduced effective device footprint of approximately $5.5F^2$ (see Figure 12b–d).

In terms of floor area, MOSFET selectors are less area-efficient and scale directly with the technology node. Their size typically approaches 20 F^2 due to the need for additional selector metal lines, and their channel width—or the number of fins in FinFET architectures—is generally large, dictated by the required programming currents for melt-quenching. As a first approximation, the transistor's drive capability must satisfy the condition $I_{sat}\ F > J_{melt}\ \eta F^2$, where I_{sat} is the available drive current per unit width, J_{melt} is the critical current density required for RESET via melt quenching, and η $(0 < \eta \leq 1)$ is an area efficiency factor. This ensures sufficient current is available to overcome the RESET threshold current . Notably, the inequality shows that it is critical to aggressively reduce J_{melt}, to keep up with the scaling down in the CMOS technology node. For IMC, low programming currents

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163–5194

decrease the power budget needed for programming the crossbar (thus also improving the overall system efficiency) as well as decrease the footprint of the peripheral circuitries, such as the sizes of the driver units. Note that maximizing integration density can be achieved not only by minimizing the synaptic cell footprint with ηF^2, but also by stacking multiple crossbar layers (L) vertically to achieve $\eta F^2/L$. However, front-end selectors create challenges in achieving the 3D stackability of PCM crossbars. Consequently, there is a strong interest in BEOL selectors like the $4F^2$ ovonic threshold switches, which can be integrated with PCM using chalcogenide glasses and self-aligned techniques.[151] The market already offers memory chips featuring up to four layers of integrated PCM.[152] The main drawback of BEOL selector-based configurations is the difficulty in achieving precise current control, which complicates analogue programming and restricts device operation to a purely binary regime. Moreover, increasing the layer count to levels seen in flash memory is challenging due to the complexities associated with patterning and encapsulating high aspect ratio structures. Furthermore, as feature sizes shrink to around 10 nm, metal lines can suffer from increased resistivity due to edge effects,[153] and significant thermal interference, manifesting as write disturbance, can occur among adjacent cells.[154,155]

6. PART 5: APPLICATION PORTFOLIO OF PCM-BASED IMC

PCM devices have been applied to a wide range of application domains, ranging from computing that requires high precision to computing that is enabled by imprecision and randomness. These applications include, for example, mixed-precision linear equation solvers for scientific computing, signal processing such as image compression, and machine learning, including deep neural network inference. A high-level overview of the main applications that are being researched for PCM-based IMC is shown in Figure 13. While computational precision is one approach to categorize the application use cases, another is the mode in which the devices are operated to perform the computations. Broadly, these include computing by collective data retrieval using read-out pulses, computing by dynamically changing the conductance states of the devices using programming pulses, and a mixture of both. In all cases, IMC reduces the complexity of a problem as well as reduces the amount of data being accessed by performing computations inside the memory arrays.

Scalar multiplication and addition of many partial scalar products to create matrix-vector multiplication operation is the core computation performed by a crossbar array of PCM devices under the read-out scheme. In performing the matrix operation, the conductance values of the PCM devices are viewed as the elements of the matrix, while voltage signals are read as inputs. Notably, each scalar computation is enabled by the Ohms law, while the additions between the scalar products are enabled by the current summation law. The same crossbar can be utilized to perform multiplication with the transposition of the matrix. Take the example of DNN inference. A DNN can be mapped onto multiple crossbar arrays of PCM devices that communicate with each other as illustrated in Figure 14a. A layer of the DNN can be implemented on (at least) one crossbar, in which the weights of that layer are stored in the conductance state of the PCM devices at the crosspoints. The propagation of data through that layer is performed in a single step by inputting the data into the crossbar rows and deciphering the results in the columns. The results are then passed through the neuron nonlinear function, and the resultant output is fed as input to the next layer. The neuron nonlinear function is typically implemented at the crossbar periphery using analogue or digital circuits. Using PCM devices to store the weights ensures that they will be retained when the power supply is turned off, unlike with volatile memory such as SRAM. Also, the multilevel storage capability of PCM can be exploited to implement nonbinary networks, which yield higher accuracy and are easier to train than binary weight networks. Other applications, where read-out based schemes are used include database query,[156] image compression,[157] associative search,[158−161] and optimizers.[58] In all these applications, PCM devices represent a predefined matrix, that is used either to manipulate input vectors or store searchable attributed vectors. However, in most of these read-out based applications implemented with PCM, programming noise, conductance drift, read noise, and conductance polarity dependence were found to degrade the accuracy. Besides using the optimized designs presented in a previous section to reduce these nonidealities, they can also be mitigated at the software level and at the circuit level. At the software level, custom hardware-aware training schemes that make DNNs more robust to noise have been successful in improving the accuracy of inference.[121,162] Furthermore, compensation methods for conductance drift have been developed, which primarily involve scaling the digitized MVM output by predetermined factors that are recalculated over time.[163] In brief, first a calibration phase computes the summed current of L columns in an array representing a network layer, with the devices initially set to known conductance values, $G_{mn}(t_0)$. By periodically measuring the column currents (I_n) under an applied voltage (V_{cal}) across all or a subset of rows, the global conductance shifts can be corrected during inference. The crossbar output (that is the MVM result) during inference is scaled by $\frac{1}{\hat{\alpha}}$, where $\hat{\alpha} = \frac{\sum_{n=1}^{L} I_n}{V_{cal} \sum_{m=1}^{N} \sum_{n=1}^{L} G_{mn}(t_0)}$ This method is straightforward, as L can be kept small enough to provide sufficient statistics, and $\hat{\alpha}$ is computed directly from device data without requiring any assumptions or additional timing information.

Note, however, that while such correction methods can mitigate these effects, they do not fully eliminate them.[124] This is primarily due to two key factors: the state dependency of drift exponents and the variability between devices, which linear correction schemes have failed to address. As a result, the accuracy of MVMs tends to degrade over time.[121,162] Material and device innovations could help to address this issue. For instance, low mean drift exponents can minimize absolute variance across arrays, while state invariance, where the drift coefficient remains consistent across all conductance states, can improve drift compensation.[164] A good example is projected-type devices, where the metallic liner provides both of these beneficial effects,[132,165] or materials with inherently low drift coefficients.[149,166] Like drift, the ambient temperature is another crucial state variable that can impact inference accuracy. At the device level, drift and temperature have opposing effects: while drift causes a decrease in conductance over time, an increase in temperature leads to higher conductance. However, since $\hat{\alpha}$ is an empirical measure of conductance changes, it inherently accounts for both effects. As a result, variations in ambient temperature can be corrected using the calibration procedure described above.[124,125] Never-

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

theless, state dependency in activation energies and their variability lead to similar challenges, necessitating further material and device innovation such as projected devices. By employing a similar scheme that utilizes linear factors obtained during calibration, the dependence of conductance on polarity can be addressed at the circuit level, although not entirely, for the same underlying reasons. This can be achieved by using different read voltages based on the polarity of the input or by tuning the column voltages.[124,167]

Programming noise can also be addressed at the circuit level through optimized iterative algorithms[122] and by mapping weights onto multiple devices,[53,168] which will be discussed next. The precision of weight programming plays a vital role in determining the achievable accuracy in AIMC. This requires finely controlled phase configurations during programming. This control is influenced by device geometry (for eg. bridge-type devices exhibit a linear dependence on the size of the amorphous mark, while mushroom-type devices show a nonlinear response) as well as by current regulation through the programming circuitry's ability to generate precisely adjustable current pulses. Conventional chips typically produce box-shaped pulses, controlling the amplitude and width rather than trailing edges. Additionally, current control can be implemented through the gate modulation of current-limiting selectors. As described earlier, the standard algorithm for programming crossbar arrays relies on iteratively reading and correcting the individual unit-cell conductance until the read-out value is within a predefined error margin of the target weight (see Figure 14d). In current demonstrations, floating-point numeric values W_{mn} are mapped to the continuous conductance levels of the PCM device G_{mn} without quantization, according to $G_{mn} = W_{mn} \cdot \frac{G_{max}}{W_{max}}$ where W_{max} is the maximum absolute value to be programmed, and G_{max} is the maximum reliably programmable unit-cell conductance (for example, 90 μS). Convergence is reached when the error is below a specified threshold, typically 5% of the value being mapped. For example, a weight $W_{mn} = 0.5$ in the range of $-1 \leq W \leq 1$ can be mapped to a value $G = 45 \pm 2.25$, where the ± 2.25 represents the permissible error margin. The iterative programming approach is taken from data storage technology. It is becoming clear, however, that for AIMC, this approach presents some challenges. Specifically, once a device is considered converged, it is typically disregarded for the rest of the programming procedure. Hence, the device conductance drifts away from its target value, while the other devices are programmed, leading to additional uncorrected errors. Moreover, accurately reading the conductance of an individual device requires a highly precise read circuitry as well as long integration times, since sub-μA currents are involved. Programming algorithms specifically curated for AIMC have therefore been developed. One of them is gradient descent-based programming (GDP), which relies on directly optimizing the MVM accuracy per core instead of the individual device conductance values.[122] In GDP, after initializing the unit-cell conductances, batched MVMs are performed on-chip with randomly generated input vectors (Figure 14d). The resulting MVM error is quantified as a loss function, and the gradients of this loss function are used to set the amplitudes of the programming pulses. This procedure is iterated until a satisfactory MVM error is reached.

In addition, as discussed earlier, it has been observed that it is advantageous to use multiple devices to encode the positive or negative weight components. This has led to the adoption of *diff*-N unit cells, which use N devices per weight polarity, comprising a total of 2N devices. This configuration allows for more optimal weight mapping. In *diff*-N unit cells, G_{max} can be programmed in a variety of possible ways (i.e., phase configurations) since the total conductance is a linear summation of the conductance states of individual devices. It has been observed that the most optimal mapping scheme for compute precision is one that selects the smallest number of devices needed to accommodate G_{max}, through the use of true SET and RESET states[168] (see Figure 14f). This ensures that devices in the nonideal intermediate states are avoided. Reiterating the earlier discussion: in IMC, there is a strong incentive to maintain low conductance values to enable larger array sizes. At the same time, it is advantageous for G_{min} to approach zero, representing the smallest achievable conductance. Unlike digital systems, where a numerical zero results in no output, a small but nonzero G_{min} can still contribute to the computation—particularly as the sparsity in the inputs and weights decreases.

In computations that make use of programming pulses, the device's conductance gets modified during the computation from the material's crystallization (or amorphization) dynamics. For example, in the so-called accumulation scheme, the device's conductance evolves in accordance with the number of (constant amplitude) crystallization pulses that encode a computational problem. Moreover, the result of the computation is stored in place. Such a compute primitive is particularly attractive for performing temporal signal analysis, such as unsupervised correlation detection. In the performing correlation, each process is assigned to a PCM device. The sum of instantaneous events across all processes is used to modulate the width or amplitude of the crystallization pulses that are applied to the devices. By monitoring the conductance of the memory devices, both correlated and uncorrelated groups can be identified, with clustering accuracy improving in proportion to the correlation coefficients.[169] The programming pulses furthermore add a knob of the computational stochasticity associated with the switching behavior. For example, the SET operation exhibits a delay time with significant cycle-to-cycle statistical variations attributed to the threshold switching dynamics. Additional stochasticity arises from small variations in the atomic configurations of the amorphous volume created upon preceding RESET. This results in variability associated with the number of pulses that are needed to fully crystallize the amorphous volume. These form a use case for creating compact and efficient true random number generators.[170] However, progressive changes in the PCM device switching characteristics due to elemental segregation upon cycling cause issues implementing reliable random number generators. This would likely require additional feedback circuits to adjust the pulse amplitude accordingly, which consume additional area and energy.[170] In addition, the finite endurance of PCM devices imposes limitations on how many random numbers can be generated with a single device. Other applications that make use of conductance switching for computations include uncloneable functions,[171] arithmetic factorization,[172] and stateful boolean logic.[173]

PCM devices have also been used in the context of supervised training of DNNs with backpropagation.[174−177] This training involves three stages: forward propagation of labeled data through the network, backward propagation of the error gradients from the output to the input of the network,

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

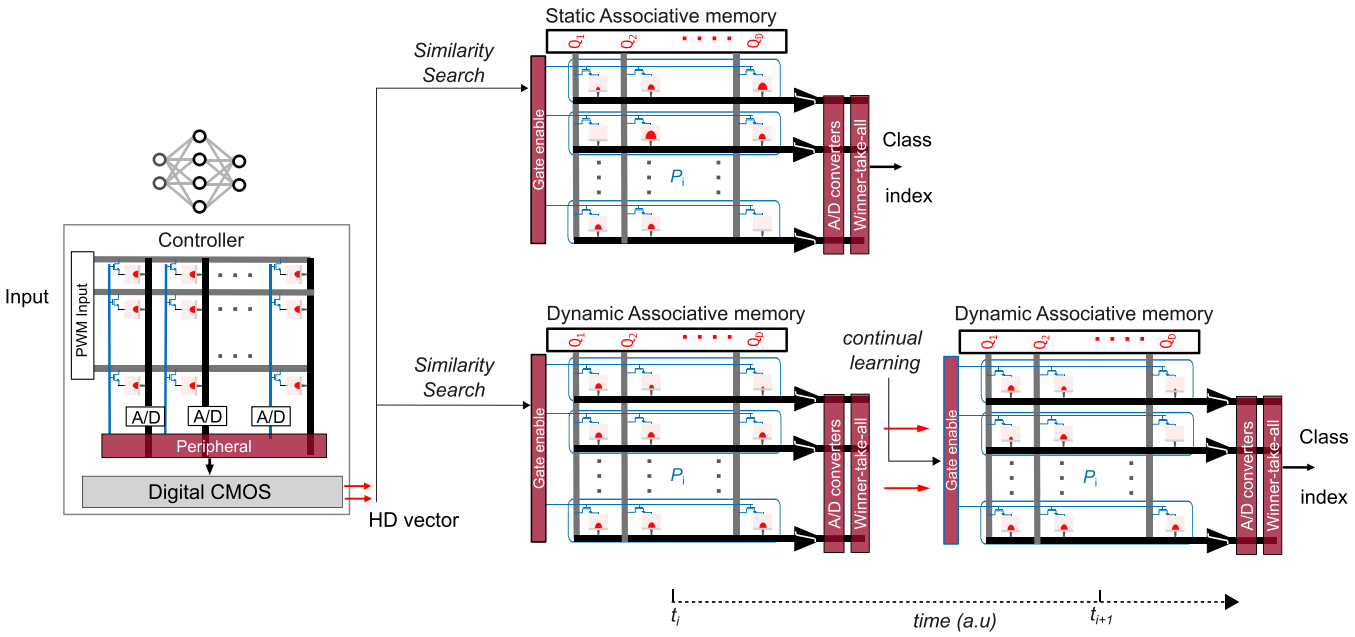


Figure 15. IMC for memory augmented neural networks. Schematic of a MANN architecture implementing associative and continual learning using a hyperdimensional representation of data. The controller network is mapped into the computational memory, which learns features during training to assign quasi-orthogonal vectors to new classes in the associative memory, also mapped into computational memory. During inference, classification is performed through a similarity search across all classes, selecting the top-scoring entries. The top panel illustrates this process on a static associative memory that remains unchanged once preprogrammed. Hardware similarity search is achieved by applying read voltage pulses and digitizing the accumulated currents. The bottom panel shows a dynamic memory that evolves during inference, becoming richer and expanding in classes. This is accomplished by applying crystallization pulses to a column of devices.

and weight update based on the computed gradients with respect to the weights of each layer. When performing training of a neural network encoded in crossbar arrays, forward propagation is performed in the same way as for the inference described above, where the mode of operation is read-out based. The only difference is that all of the activations of each layer have to be stored locally or externally if the on-chip digital storage is not sufficient. Next, backward propagation is performed by inputting the error gradient from the subsequent layer onto the columns of the current layer and deciphering the result from the rows. Finally, the weight update is performed based on the outer product of activations and error gradients of each layer, effectively making use of the programming-mode for computation. However, the high stochasticity, nonlinearity and granularity of conductance updates in PCM makes it very challenging to perform the small and precise weight updates needed to obtain accurate training convergence on DNNs.[174] One solution that was proposed to mitigate this issue is to use a separate array of 3T1C capacitor-based unit-cells to accumulate the small weight updates precisely, and transfer them periodically to the PCM devices via iterative programming.[175] Although this approach could demonstrate accurate training of a multilayer perceptron on the MNIST benchmark, the weight updates performed in the 3T1C unit-cells are still too imprecise for training larger networks on more complex tasks, and the periodic reprogramming of the entire PCM array is costly in terms of time and energy. Another approach is to accumulate the weight updates with high-precision digital computing in a separate digital memory, and update the PCM devices via single-shot pulses when the accumulated weight updates reach a given threshold.[176] This approach was shown to successfully train medium-sized DNNs accurately at the cost of additional digital computing and memory requirements. These training approaches are still at the stage of functionality demonstration, and additional algorithmic work is being pursued to overcome the device-related challenges.[178]

PCM arrays can also be coupled with neurons that compute with asynchronous spikes that are temporally precise in a so-called spiking neural network (SNN). Such SNNs are ideally suited for processing spatiotemporal event-driven information from neuromorphic sensors.[179−182] Unsupervised learning with local learning rules is also generally used in neuromorphic engineering, as opposed to the global supervised back-propagation algorithm employed in deep learning. Implementations of local learning using PCM as synaptic weight have been mainly based on overlapping voltage pulses to implement potentiation and depression.[180,181,183,184] However, STDP-based learning rules have still not been able to reach the accuracy of conventional DNNs trained with backpropagation, despite ongoing progress.[185,186]

Another interesting application of PCM devices is in memory-augmented neural networks. Memory-augmented neural networks (MANN) enhance neural networks with explicit memory to overcome slow training procedures in traditional DNNs. Crucially, new information can be offloaded to explicit memory, where it does not endanger the previously learned information to be overwritten. MANN architectures are composed of a controller, which is a neural network model, followed by a structured memory as the explicit memory[187] (see Figure 15). The entries in the explicit memory are not accessed by stating a discrete address but by comparing a query from the controller's side with all entries. This means that access to the memory occurs via soft read operations, which involve every individual memory entry instead of a single discrete entry, creating the so-called associative type search. Both components of the MANN system can be implemented

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

on PCM crossbars. In the learning stage, the controller is trained to assign hyper-dimensional quasi-orthogonal, and thus dissimilar, vectors to novel classes in the explicit memory. During inference, a hyper-dimensional projection of an input can be compared in parallel across all entries in the memory and scored for classification. The explicit memory, furthermore, can be transformed to dynamically evolve during inference, which is made to better learn existing information or to learn new information altogether during inference. Here, the controller interacts with the memory via programming operations. Although the controller remains stationary, the memory dynamically updates its contents with new examples and grows its size by storing new classes.[188] Crucially, this feature enables one-/few-shot learning, where new classes can be rapidly assimilated from a few training examples of never-seen-before classes to be written in explicit memory, making it an architectural solution to continually learning new classes from a few training examples without forgetting previous old classes. Further work is nonetheless needed to demonstrate a fully hardware-implemented MANN showing end-to-end benefits over a digital implementation.

Recent advancements have harnessed the semiconductivity inherent in phase-change materials for temporal signal analysis.[58] This intrinsic property enables the electrostatic tuning of electrical conductivity, without necessitating alterations in the material's atomic arrangement. The resultant field-effect modulation is transient yet notably volatile. Consequently, the phase-change memtransistor, a singular device, embodies mixed-plasticity attributes. The dynamic transitions between amorphous and crystalline phases, coupled with the nonvolatility of the ensuing phase configurations, facilitate long-term plasticity governed weight modifications. The volatile electronic Fermi level shifts facilitate short-term plasticity-governed weight adjustments. This unique combination has been demonstrated to effectively integrate information, thereby enabling the utilization of past event data to predict current and future occurrences. This capability lays the foundation for sequential learning, vital for optimal modeling of dynamically changing environments.[182] In this context, the weights of an SNN are adjusted, allowing the network to classify static inputs by using the long-term plasticity feature. In the inference phase, the combination of long-term and transient plasticities in the synapses results in compensatory effects. These effects are significant to the extent that transient plasticity, which does not rely on prior training, can facilitate precise inference, even for inputs that only partially align with the trained weights. There are also reports of exploiting PCM device nonidealities for computational purposes. For instance, the stochasticity associated with their cumulative behavior can create biorealistic randomly spiking neurons,[189] and structural relaxation can be used to implement eligibility traces for reinforcement learning.[190] The conductance fluctuations in PCM have also been utilized in an in-memory factorizer to disentangle visual attributes[191] and solve combinatorial optimization problems.[58] Nevertheless, these device concepts present some challenges. In memtransistive devices, the need for a third terminal complicates large-scale integration, requiring a re-evaluation of traditional memristive peripheral circuits and crossbar layouts. Additionally, a strong electrostatic effect in memtransistive materials is crucial given the stringent output voltage limits of CMOS peripheral circuits. In phase-change neurons, where computations involve repetitive programming operations, exceptionally high device endurance and programming efficiency become a fundamental requirement.

There is also increasing interest in areas such as PCM-based IMC designed for operation at ultralow temperatures, targeting applications in deep space and cryogenic electronics.[192] By and large, the principles of IMC at and above room temperature remain applicable to ultralow temperatures as well. It could also be expected that at ultra low temperatures, metrics such as compute precision improve as conductance drift and noise can be arrested. However, a deeper understanding of the computational workloads suited to these applications as well as a comprehensive exploration of the effects of cryogenic temperatures (approaching 4 K and below) on device performance is required. These include gathering insights on temperature (radiation)-dependent conductance values, state-dependent readout, and programming characteristics. There have been also efforts to leverage device nonidealities to enhance the efficiency of certain machine learning tasks, such as reinforcement learning. For example, conductance drift can be utilized to implement eligibility traces,[190,193] allowing synapses to retain a memory of past activities over extended periods. Unlike in the Hebbian learning, learning in this framework is not solely dependent on present neuronal activity but also on past activity and the presence of a reward or prediction signal.[194,195] Implementing eligibility traces in-memory eliminates the need for complex analogue/digital circuitry required for numerical integration. However, such an approach still relies on accumulative phase transitions and operates within a constrained window. The former results from the requirement of write pulses, increasing power consumption, and being constrained by endurance limits. The latter is restricted by inherently small drift coefficients.

Another recent development involves leveraging PCM devices to enhance computational efficiency and data privacy by enabling processing capabilities directly within sensor units.[196] For instance, PCM devices embedded in active image sensor units can perform real-time scalar multiplication operations on photogenerated currents, effectively transforming pixel outputs into computational results. Convolution operations can be performed with the accumulation step, adding the products of multiple pixels, by summing the outputs of neighboring pixels (determined by the kernel size) in parallel along the sensor's crossbar interconnects. Such an approach decouples the sensing and computation elements, facilitating dense integration and more manufacturable computational sensors. However, further investigation is needed to understand the challenges and strategies for effectively integrating PCM devices with sensor systems effectively. There is also interest in exploring the implementation of cryptography using IMC based on PCM devices. A range of operations from the prototypical advanced encryption standard (AES), a widely adopted algorithm in symmetric key cryptography (that is, the same key is used for encryption and decryption), have been successfully implemented on PCM devices though novel in-memory operators. These operations include encryption, key expansion, and stochastic key generation, all of which are supported by the PCM devices operating in the binary mode, and the peripherals of the crossbar units.[197,198] Beyond AES, stochastic key generation within memory arrays is a compelling concept with the potential to greatly enhance security guarantees. However, ensuring sufficiently reliable stochastic key generation remains a challenge, necessitating further research. Lastly, applications

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

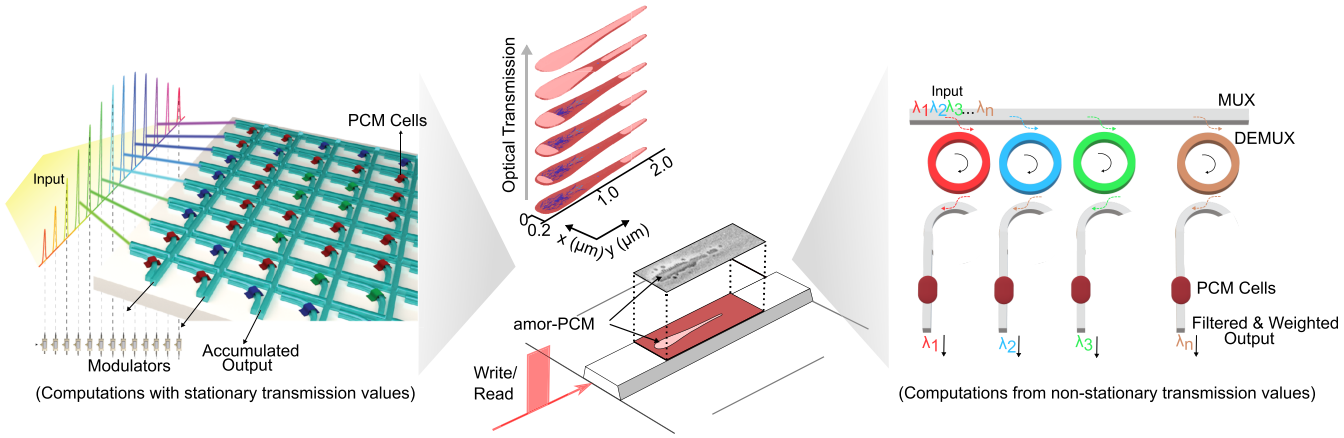


Figure 16. Photonic computational memory. A conceptual illustration of an integrated photonic computational memory. The central panel depicts various phase configurations in the functional phase change material layer achieved through optical programming pulses. The left panel illustrates a photonic tensor core executing in-memory matrix-vector multiplication with multiple parallel wavelength-encoded inputs. The right panel showcases a photonic circuit performing correlation detection on multiple wavelength-encoded inputs.

that require frequent programming of PCM devices are still being actively researched. Examples include solving combinatorial optimization problems[199] and Neural Architecture Search[200] operations. As before, in these applications, high programming efficiency and endurance are critical factors. There have been proposals to use AIMC to enhance computational efficiency when working with large data sets, such as in clustering tasks. In this approach, data (input vectors) are stored in the crossbar, and randomly initialized cluster vectors iteratively operate until convergence is reached.[201] Such methods reduce data movement and can perform clustering faster than traditional approaches, such as Lloyd's algorithm. Additionally, there have been proposals to implement kernel approximations (which approximate nonlinear operations) using linear operations on AIMC. Kernel functions capture nonlinear relationships in the input data while operating within the original space. However, they become computationally expensive as the data set grows. In-memory kernel approximation leverages AIMC to explicitly map inputs into a higher-dimensional space, allowing similarity to be computed through dot products, which are computationally cheaper than directly applying nonlinear kernels. These methods also enable the implementation of other operations, such as the attention mechanisms, which is required for encoder and decoder models in transformers.[202]

The functional properties of phase-change materials in the optical domain makes it possible to use them to implement computational memory on photonic integrated circuits.[203−205] Integrated photonics enhances computational memory by enabling parallel data transfers via wavelength division multiplexing (WDM) and providing extremely high data modulation speeds (Figure 16). Recently, these features have been utilized in photonic tensor cores to accelerate deep learning. In the first category of photonic circuits, phase-change materials serve as attenuating elements, absorbing a specified amount of light based on their amorphous crystalline fraction (phase configuration). The programming of phase-change materials to multiple transmissive states is achieved through partial amorphization/crystallization pulses. Massively parallel MVM operations, as well as event-based spike aggregation, have been demonstrated in single-time steps using WDM.[206,207] The accumulative behavior arising from the crystallization dynamics was further leveraged in the optical domain. Here, the device's transmission evolves by the number of (constant amplitude) crystallization pulses that encode a computational problem. Using this property, unsupervised correlation detection on real-time data streams for big data analytics has been demonstrated.[208] In the second category of photonic circuits, phase modulation is used as a means to perform computations. This category utilizes phase-change materials that exhibit large changes in the real part of the refractive index at minimal losses across all phase configurations.[33,209] Nevertheless, current photonic accelerators present some technological challenges. At the level of devices, there is interest in developing photonic circuits integrated with microheaters, utilizing electrical Joule heating to program larger areas of the phase-change material, in order to achieve a higher memory window for analogue programming.[210,211] Another equally important research focus is improving the device endurance, which is currently limited to a few thousand cycles.[212] Unlike the electronic counterpart, currently, the photonic circuits cannot support the extensive number of optical components required for neural network-based accelerators, and down scaling the device footprint is constrained by the diffraction limit of light. The fabrication errors necessitate the development of corrective optics schemes.[12,213] Maintaining a constant signal-to-noise ratio across network layers requires on-chip amplifiers and powerful broadband lasers. While III−V lasers show promise, their integration with photonic and electronic circuits remains under investigation. Furthermore, high-performance on-chip photodiodes and modulators are needed, with energy efficiency and scalability being critical. The functional components will also require power supplies, control systems, and CMOS integration, which are still being developed.[214] New optical components for nonlinear data processing and input/output data transfers, such as fiber-to-package connectivity and environmental resilience, will also prove essential. Additionally, exploring computational workloads that minimize electro-optical conversions by keeping a significant portion of processing within the photonics domain and requiring fewer unit cells could help address many existing challenges.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

Examples include the implementation of physical solvers that perform optimization tasks through recursive loops within the same photonic circuit.[215] Another example is similarity search devices, such as content-addressable memories, used in applications like switches, where a single layer of photonic hardware can carry out search and logic operations in superposition.[216]

7. OUTLOOK

The material science and device physics of PCM are well understood, and large-scale manufacturability using back-end-of-line processing on advanced CMOS technology nodes is well-established.[70,74] These advancements position PCM as potentially the most advanced memristive technology for IMC. Currently, PCM-based IMC chips incorporate up to 35 million PCM devices, giving them the largest weight capacity among IMC chips.[217] In contrast, other emerging memristive devices, such as advanced ReRAM, still face challenges. These include overcoming high forming voltages that impede further scaling and integration density, as well as array-level device variability that limits large-scale integration.[218−221]

While progress has been made in PCM-based IMC, there remains significant room for improvement—particularly in enhancing compute density and compute precision. Enhancing compute density critically depends on higher integration levels, achievable by minimizing both the size and count of selector devices—such as transistors—per PCM unit. This, in turn, requires reducing programming currents and threshold voltages, which calls for innovation at both the material and device levels. Ideally, each unit cell would incorporate a single selector operating at an advanced technology node. Advancing integration density further will also require the development of 3D-stacked PCM architectures. When discussing compute precision, current solutions like global software-based compensation schemes aim to correct device nonidealities, but they are fundamentally limited. These largely stem from state-dependent material properties and array level variabilities—for instance, the drift exponent and activation energy for carrier transport can vary with conductance and from device to device—making accurate correction difficult. To further enhance compute precision, it is therefore essential to develop new devices with lower intrinsic drift, reduced noise, and minimal temperature sensitivity.[121,124] To that end, projected-type devices, can be an excellent candidate,[141,222] which achieve these characteristics without compromising on other critical properties like retention and endurance. Devices with low conductance values across all programmable states are also essential, as they help minimize *IR* drops on interconnects. This strategy, furthermore, would enable PCM devices to be integrated beneath the first metal layer, positioning them closest to the selectors.[223]

Furthermore, PCM-IMC has reached a level of maturity where it should be viewed within the broader context of the entire technology stack. This includes circuits, architectures, integration, packaging, algorithms, and software. At the peripheral circuitry level, there is a need to rethink both circuits and microarchitectures. For example, data converters, such as ADCs, can be significant in terms of energy consumption and area. The design of the ADC is closely tied to the areal footprint of unit cells and the conductance of PCM devices. Ideally, each bit-line in the crossbar should have its own ADC. When multiple columns share an ADC, efficient multiplexing schemes must be developed. Additionally, the efficiency of utilizing crossbar arrays is determined by specific application requirements. For example, the fixed size of crossbar (tile) poses challenges, whether dealing with a large layer that exceeds a single tile or a small layer that underutilizes a tile, both scenarios can lead to reduced performance. To optimize end-to-end workloads, IMC cores should be enhanced with either small custom digital blocks for simpler auxiliary operations or larger custom digital logic for more complex tasks.[224] These units may include digital accelerators, such as SRAM-based IMCs, to efficiently handle MVM workloads, especially in cases where amortization in PCM IMC tiles are not optimal, such as small neural network layers. Additionally, general-purpose processors like RISC-V can manage these workloads, execute nonlinear operations like sigmoidal activations and batch normalization, and act as host subsystem managing communication, memory access, and execution coordination.[225,226] Innovations are also needed to develop fine-grained pipelines for data transfers, tailored to various execution patterns.[227] Additionally, a comprehensive software stack is necessary, including hardware-aware training libraries,[228] hardware-optimized compilers, and kernels that support various deep learning frameworks such as PyTorch.

The impact of external perturbations, such as ambient temperature variations, will need to be more carefully accounted for. While software compensation schemes can partially address compute inaccuracies, they do not yet account for real-time temperature fluctuations, and prototype chips require downtime to cool down. Therefore, adaptive circuitry and corresponding methods for real-time calibration of PCM arrays and circuits must be explored.

Current analyses suggest that the first commercial application of PCM-based IMC is likely to be in on-chip neural processing units for deep neural network inference. In this context, PCM could be utilized both as IMC units and as high-density analogue memory units supporting digital IMCs. These on-chip accelerators have the potential to significantly broaden the scope of edge-AI applications. For instance, modern microcontrollers used in edge devices typically operate at frequencies from tens to hundreds of MHz, with memory capacities ranging from hundreds of kilobytes to a few megabytes. These metrics are insufficient for the demands of advanced AI algorithms in areas such as computer vision and natural language processing. If on-chip neural processing units are commercially successful, then standalone accelerators could subsequently be developed for more demanding applications, including those in cloud-based environments. At the same time, there is increasing interest in identifying a broader range of application use cases for PCM-based IMC, extending beyond current neural networks and inference. These applications could include areas such as cryogenic electronics, in-memory kernel approximation, computational sensors, and cryptography, all of which require further exploration.

AUTHOR INFORMATION

Corresponding Author

Ghazi Sarwat Syed − *IBM Research−Europe, 8803 Rüschlikon, Switzerland*; Email: ghs@zurich.ibm.com

Authors

Manuel Le Gallo − *IBM Research−Europe, 8803 Rüschlikon, Switzerland*

Abu Sebastian − *IBM Research−Europe, 8803 Rüschlikon, Switzerland*; orcid.org/0000-0002-2054-4887

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemrev.4c00670

Notes

The authors declare no competing financial interest.

Biographies

Ghazi Sarwat Syed: Dr. Ghazi Sarwat Syed is a Research Staff Member at IBM Research−Zurich, where he conducts research on phase-change materials and memory devices for processing-in-memory accelerators across both electronic and optical platforms. He holds a Bachelor of Engineering from PSG College of Technology, India, and a Ph.D. in Metallurgical Engineering and Materials Science from the University of Oxford, United Kingdom. Dr. Syed has authored over 35 publications, delivered 15 invited talks/tutorials, and holds 28 international patents in the field. He has also contributed to the development of IBM's inference server. In 2023, he was recognized as a Forbes 30 Under 30 honoree and also awarded the honorary title of IBM Master Inventor.

Manuel Le Gallo: Manuel Le Gallo joined IBM Research Europe in 2013, where he is currently employed as a Staff Research Scientist in the In-Memory Computing Group of the Zurich laboratory. His main research interest is in using phase-change memory devices for non-von Neumann computing. He has coauthored more than 100 scientific papers in journal and conferences, holds 35 granted patents and has given 15 invited talks. He was appointed IBM Master Inventor in 2019 and 2024 for significant contributions to intellectual property and is a recipient of the MIT Technology Review's 2020 Innovators Under 35 award.

Abu Sebastian: Dr. Abu Sebastian is a Distinguished Scientist and technical manager at IBM Research−Zurich. He is one of the technical leaders of IBM's research efforts towards next-generation AI Hardware and manages the in-memory computing group at IBM Research−Zurich. He is the author of over 250 publications in peer-reviewed journals/conference proceedings and holds over 100 US patents. In 2015, he was awarded the European Research Council (ERC) consolidator grant and in 2020, he was awarded an ERC Proof-of-Concept grant. He was an IBM Master Inventor and was named Principal and Distinguished Research Scientist in 2018 and 2020, respectively. In 2019, he received the Ovshinsky Lectureship Award for his contributions to "Phase-change materials for cognitive computing". In 2023, he was conferred the title of Visiting Professor in Materials by University of Oxford. He is a distinguished lecturer and fellow of the IEEE.

ACKNOWLEDGMENTS

We thank the members of the In-Memory Computing Device Group at IBM Research − Europe for useful discussions on the manuscript.

REFERENCES

(1) Ovshinsky, S. R. Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.* **1968**, *21*, 1450.
(2) Lanza, M.; Sebastian, A.; Lu, W. D.; Le Gallo, M.; Chang, M.-F.; Akinwande, D.; Puglisi, F. M.; Alshareef, H. N.; Liu, M.; Roldan, J. B. Memristive technologies for data storage, computation, encryption, and radio-frequency communication. *Science* **2022**, *376*, No. eabj9979.
(3) Le Gallo, M.; Sebastian, A. An overview of phase-change memory device physics. *J. Phys. D: Appl. Phys.* **2020**, *53*, 213002.
(4) Sarwat, S. G. Materials science and engineering of phase change random access memory. *Mater. Sci. Technol.* **2017**, *33*, 1890−1906.
(5) Raoux, S. Phase change materials. *Annu. Rev. Mater. Res.* **2009**, *39*, 25−48.
(6) Pellizzer, F.; Redaelli, A. *Semiconductor Memories and Systems*; Elsevier, 2022; pp 253−276.
(7) Cheng, H.-Y.; Carta, F.; Chien, W.-C.; Lung, H.-L.; BrightSky, M. J. 3D cross-point phase-change memory for storage-class memory. *J. Phys. D: Appl. Phys.* **2019**, *52*, 473002.
(8) Fantini, P. Phase change memory applications: the history, the present and the future. *J. Phys. D: Appl. Phys.* **2020**, *53*, 283002.
(9) Cappelletti, P.; Annunziata, R.; Arnaud, F.; Disegni, F.; Maurelli, A.; Zuliani, P. Phase change memory for automotive grade embedded NVM applications. *J. Phys. D: Appl. Phys.* **2020**, *53*, 193002.
(10) Kato, T.; Tanaka, K. Electronic properties of amorphous and crystalline Ge2Sb2Te5 films. *Japanese journal of applied physics* **2005**, *44*, 7340.
(11) Aryana, K.; Zhang, Y.; Tomko, J. A.; Hoque, M. S. B.; Hoglund, E. R.; Olson, D. H.; Nag, J.; Read, J. C.; Rios, C.; Hu, J.; Hopkins, P. E. others Suppressed electronic contribution in thermal conductivity of Ge2Sb2Se4Te. *Nat. Commun.* **2021**, *12*, 7187.
(12) Ghazi Sarwat, S.; Cheng, Z.; Youngblood, N.; Sharizal Alias, M.; Sinha, S.; Warner, J.; Bhaskaran, H. Strong opto-structural coupling in low dimensional gese3 films. *Nano Lett.* **2019**, *19*, 7377−7384.
(13) Nazeer, H.; Bhaskaran, H.; Woldering, L. A.; Abelmann, L. Young's modulus and residual stress of GeSbTe phase-change thin films. *Thin solid films* **2015**, *592*, 69−75.
(14) Sebastian, A.; Le Gallo, M.; Khaddam-Aljameh, R.; Eleftheriou, E. Memory devices and applications for in-memory computing. *Nature Nanotechnol.* **2020**, *15*, 529−544.
(15) Sebastian, A.; Le Gallo, M.; Burr, G. W.; Kim, S.; BrightSky, M.; Eleftheriou, E. Tutorial: Brain-inspired computing using phase-change memory devices. *J. Appl. Phys.* **2018**, *124*, 111101.
(16) Syed, G. S.; Le Gallo, M.; Sebastian, A. *Phase Change Materials-Based Photonic Computing*; Elsevier, 2024; pp 11−35.
(17) Gildart, L. Electrical, optical and thermal properties of M_2 N_3 semiconductors. *Final Technical Report on Contract AF 49(638)-90*; Armed Forces Technical Information Agency: Arlington, VA, 1961.
(18) Dewald, J. F.; Northover, W. R.; Pearson, A. D. *Multiple Resistance Semiconductor Elements*. ; US Patent 3,241,009, 1966,
(19) Sie, C. H. *Memory cell using bistable resistivity in amorphous As-Te-Ge film*. Ph.D. dissertation. Iowa State University, 1969.
(20) Chen, M.; Rubin, K. A.; Barton, R. Compound materials for reversible, phase-change optical data storage. *Applied physics letters* **1986**, *49*, 502−504.
(21) Yamada, N.; Ohno, E.; Akahira, N.; Nishiuchi, K.; Nagata, K.; Takao, M. High speed overwritable phase change optical disk material. *Jpn. J. Appl. Phys.* **1987**, *26*, 61.
(22) Yamada, N.; Ohno, E.; Nishiuchi, K.; Akahira, N.; Takao, M. Rapid-phase transitions of GeTe-Sb2Te3 pseudobinary amorphous thin films for an optical disk memory. *J. Appl. Phys.* **1991**, *69*, 2849−2856.
(23) Wuttig, M.; Yamada, N. Phase-change materials for rewriteable data storage. *Nature materials* **2007**, *6*, 824−832.
(24) Iwasaki, H.; Harigaya, M.; Nonoyama, O.; Kageyama, Y.; Takahashi, M.; Yamada, K.; Deguchi, H.; Ide, Y. Completely erasable phase change optical disc II: Application of Ag-In-Sb-Te mixed-phase system for rewritable compact disc compatible with CD-velocity and double CD-velocity. *Japanese journal of applied physics* **1993**, *32*, 5241.
(25) Borg, H. J.; Van Schijndel, M.; Rijpers, J. C.; Lankhorst, M. H.; Zhou, G.; Dekker, M. J.; Ubbens, I. P.; Kuijper, M. Phase-change media for high-numerical-aperture and blue-wavelength recording. *Jpn. J. Appl. Phys.* **2001**, *40*, 1592.
(26) Horii, H.; Yi, J.; Park, J.; Ha, Y.; Baek, I.; Park, S.; Hwang, Y.; Lee, S.; Kim, Y.; Lee, K.; Chung, U.-I.; Moon, J. A novel cell technology using N-doped GeSbTe films for phase change RAM. 2003 Symposium on VLSI Technology. *Digest of Technical Papers*; IEEE Cat. No.03CH37407; IEEE, 2003; pp 177−178.
(27) Zhou, X.; Xia, M.; Rao, F.; Wu, L.; Li, X.; Song, Z.; Feng, S.; Sun, H. Understanding phase-change behaviors of carbon-doped Ge2Sb2Te5 for phase-change memory application. *ACS Appl. Mater. Interfaces* **2014**, *6*, 14207−14214.

(28) Matsuzaki, N.; Kurotsuchi, K.; Matsui, Y.; Tonomura, O.; Yamamoto, N.; Fujisaki, Y.; Kitai, N.; Takemura, R.; Osada, K.; Hanzawa, S.; *Oxygen-doped gesbte phase-change memory cells featuring 1.5 V/100-/spl mu/A standard 0.13/spl mu/m CMOS operations*; IEDM Technical Digest; IEEE International Electron Devices Meeting, 2005; pp 738−741.
(29) Ryu, S. W.; Oh, J. H.; Lee, J. H.; Choi, B. J.; Kim, W.; Hong, S. K.; Hwang, C. S.; Kim, H. J. Phase transformation behaviors of SiO_2 doped $Ge_2Sb_2Te_5$ films for application in phase change random access memory. *Appl. Phys. Lett.* **2008**, *92*, 142110.
(30) Raoux, S.; Jordan-Sweet, J. L.; Kellock, A. J. Crystallization properties of ultrathin phase change films. *J. Appl. Phys.* **2008**, *103*, 114310.
(31) Salinga, M.; Kersting, B.; Ronneberger, I.; Jonnalagadda, V. P.; Vu, X. T.; Le Gallo, M.; Giannopoulos, I.; Cojocaru-Mirédin, O.; Mazzarello, R.; Sebastian, A. Monatomic phase change memory. *Nature materials* **2018**, *17*, 681−685.
(32) Zhang, W.; Mazzarello, R.; Wuttig, M.; Ma, E. Designing crystallization in phase-change materials for universal memory and neuro-inspired computing. *Nature Reviews Materials* **2019**, *4*, 150−168.
(33) Zhang, Y.; Chou, J. B.; Li, J.; Li, H.; Du, Q.; Yadav, A.; Zhou, S.; Shalaginov, M. Y.; Fang, Z.; Zhong, H.; Roberts, C.; Robinson, P.; Bohlin, B.; Rios, C.; Lin, H.; Kang, M.; Gu, T.; Warner, J.; Liberman, V.; Richardson, K.; Hu, J. Broadband transparent optical phase change materials for high-performance nonvolatile photonics. *Nat. Commun.* **2019**, *10*, 4279.
(34) Delaney, M.; Zeimpekis, I.; Lawson, D.; Hewak, D. W.; Muskens, O. L. A new family of ultralow loss reversible phase-change materials for photonic integrated circuits: Sb2S3 and Sb2Se3. *Adv. Funct. Mater.* **2020**, *30*, 2002447.
(35) Guerin, S.; Hayden, B.; Hewak, D. W.; Vian, C. Synthesis and screening of phase change chalcogenide thin film materials for data storage. *ACS combinatorial science* **2017**, *19*, 478−491.
(36) Kusne, A. G.; Yu, H.; Wu, C.; Zhang, H.; Hattrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; Davydov, A. V.; Agarwal, R.; Bendersky, L. A.; Li, M.; Mehta, A.; Takeuchi, I. others On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **2020**, *11*, 5966.
(37) Ovshinsky, S.; Fritzsche, H. Amorphous semiconductors for switching, memory, and imaging applications. *IEEE Trans. Electron Devices* **1973**, *20*, 91−105.
(38) Wong, H.-S. P.; Raoux, S.; Kim, S.; Liang, J.; Reifenberg, J. P.; Rajendran, B.; Asheghi, M.; Goodson, K. E. Phase change memory. *Proceedings of the IEEE* **2010**, *98*, 2201−2227.
(39) Neale, R.; Nelson, D.; Moore, G. E. Nonvolatile and reprogrammable, the read-mostly memory is here. *Electronics* **1970**, *43*, 56−60.
(40) Lai, S.; Lowrey, T. OUM−A 180 nm nonvolatile memory cell element technology for stand alone and embedded applications. *Proc. of the IEEE International Electron Devices Meeting*, 2001, *2001*, 36.5.1, ,
(41) Hwang, Y.; Lee, S.; Ahn, S. J.; Lee, S. Y.; Ryoo, K. C.; Hong, H. S.; Koo, H. C.; Yeung, F.; Oh, J. H.; Kim, J. H.; Jeong, W. C.; Park, J.; Horii, H.; Ha, Y.; Yi, J. H.; Koh, G. H.; Jeong, G. T.; Jeong, H.; Kim, K., et al. Writing current reduction for high-density phase-change RAM. *Proc. IEEE International Electron Devices Meeting.*, 2003; *2003*, 37.1.1.,
(42) Lankhorst, M. H.; Ketelaars, B. W.; Wolters, R. A. Low-cost and nanoscale non-volatile memory concept for future silicon chips. *Nat. Mater.* **2005**, *4*, 347−352.
(43) Breitwisch, M. et al. Novel Lithography-Independent Pore Phase Change Memory. *IEEE Symposium on VLSI Technology*. 2007; pp 100−101.
(44) Servalli, G. A 45 nm generation Phase Change Memory Technology. *IEEE International Electron Devices Meeting (IEDM)*. 2009; pp 1−4.
(45) Xiong, F.; Liao, A. D.; Estrada, D.; Pop, E. Low-power switching of phase-change materials with carbon nanotube electrodes. *Science* **2011**, *332*, 568−570.
(46) Yang, Z.; Li, B.; Wang, J.-J.; Wang, X.-D.; Xu, M.; Tong, H.; Cheng, X.; Lu, L.; Jia, C.; Xu, M.; Miao, X.; Zhang, W.; Ma, E. Designing Conductive-Bridge Phase-Change Memory to Enable Ultralow Programming Power. *Advanced Science* **2022**, *9*, 2103478.
(47) Wang, X.; Song, S.; Wang, H.; Guo, T.; Xue, Y.; Wang, R.; Wang, H.; Chen, L.; Jiang, C.; Chen, C.; Shi, Z.; Wu, T.; Song, W.; Zhang, S.; Watanabe, K.; Taniguchi, T.; Song, Z.; Xie, X. Minimizing the Programming Power of Phase Change Memory by Using Graphene Nanoribbon Edge-Contact. *Advanced Science* **2022**, *9*, 2202222.
(48) Sarwat, S. G.; Gehring, P.; Rodriguez Hernandez, G.; Warner, J. H.; Briggs, G. A. D.; Mol, J. A.; Bhaskaran, H. Scaling limits of graphene nanoelectrodes. *Nano Lett.* **2017**, *17*, 3688−3693.
(49) Kau, D.; Tang, S.; Karpov, I. V.; Dodge, R.; Klehn, B.; Kalb, J. A.; Strand, J.; Diaz, A.; Leung, N.; Wu, J.; others A stackable cross point phase change memory. *2009 IEEE International Electron Devices Meeting (IEDM)* IEEE, 2009; pp 1−4.
(50) Ikuma, Y.; Shoji, Y.; Kuwahara, M.; Wang, X.; Kintaka, K.; Kawashima, H.; Tanaka, D.; Tsuda, H. Small-sized optical gate switch using Ge2Sb2Te5 phase-change material integrated with silicon waveguide. *Electronics letters* **2010**, *46*, 368−369.
(51) Pernice, W. H.; Bhaskaran, H. Photonic non-volatile memories using phase change materials. *Appl. Phys. Lett.* **2012**, *101*, 171101.
(52) Hosseini, P.; Bhaskaran, H.; Broughton, B. *Optical device with thermally switching phase change material*. US Patent 11215852, 2022.
(53) Boybat, I.; Le Gallo, M.; Nandakumar, S.; Moraitis, T.; Parnell, T.; Tuma, T.; Rajendran, B.; Leblebici, Y.; Sebastian, A.; Eleftheriou, E. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **2018**, *9*, 2514.
(54) Suri, M.; Bichler, O.; Querlioz, D.; Cueto, O.; Perniola, L.; Sousa, V.; Vuillaume, D.; Gamrat, C.; DeSalvo, B. Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction. *International Electron Devices Meeting (IEDM)*. 2011, *2011*, 6131488.,
(55) Kuzum, D.; Jeyasingh, R. G.; Lee, B.; Wong, H.-S. P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* **2012**, *12*, 2179−2186.
(56) Koelmans, W. W.; Sebastian, A.; Jonnalagadda, V. P.; Krebs, D.; Dellmann, L.; Eleftheriou, E. Projected phase-change memory devices. *Nat. Commun.* **2015**, *6*, 8181.
(57) Giannopoulos, I.; Sebastian, A.; Le Gallo, M.; Jonnalagadda, V. P.; Sousa, M.; Boon, M.; Eleftheriou, E. 8-bit precision in-memory multiplication with projected phase-change memory. *2018 IEEE International Electron Devices Meeting (IEDM)*; IEEE, 2018; pp 27−7.
(58) Sarwat, S. G.; Kersting, B.; Moraitis, T.; Jonnalagadda, V. P.; Sebastian, A. Phase-change memtransistive synapses for mixed-plasticity neural computations. *Nat. Nanotechnol.* **2022**, *17*, 507−513.
(59) Simpson, R. E.; Fons, P.; Kolobov, A. V.; Fukaya, T.; Krbal, M.; Yagi, T.; Tominaga, J. Interfacial phase-change memory. *Nature Nanotechnol.* **2011**, *6*, 501−505.
(60) Syed, G. S.; Philicelli, T.; Luchtenveld, J.; Jonnalagadda, V. P.; Brew, K.; Kooi, B. J.; Narayanan, V.; Saulnier, N.; Sebastian, A. *Disc-type Phase-Change Memory Devices for Low Power and High Density Analog In-memory computing*; Proceedings of the European Phase-Change and Ovonic Symposium (E): Leipzig, Germany, 2024.
(61) Hong, S.; Choi, H.; Park, J.; Bae, Y.; Kim, K.; Lee, W.; Lee, S.; Lee, H.; Cho, S.; Ahn, J.; Kim, S.; Kim, T.; Na, M.; Cha, S. Extremely high performance, high density 20nm self-selecting cross-point memory for Compute Express Link. *IEEE International Electron Devices Meeting (IEDM)* **2022**, *2022*, 10019415.
(62) De Sandre, G.; Bettini, L.; Pirola, A.; Marmonier, L.; Pasotti, M.; Borghi, M.; Mattavelli, P.; Zuliani, P.; Scotti, L.; Mastracchio, G.; A 90 nm 4 Mb embedded phase-change memory with 1.2 V 12 ns read access time and 1MB/s write throughput. *2010 IEEE International Solid-State Circuits Conference-(ISSCC)*, 2010; pp 268−269.
(63) Hanzawa, S.; Kitai, N.; Osada, K.; Kotabe, A.; Matsui, Y.; Matsuzaki, N.; Takaura, N.; Moniwa, M.; Kawahara, T. A 512 kB

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

Embedded Phase Change Memory with 416 kB/s Write Throughput at 100μA Cell Write Current. *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 2007; pp 474−616.
(64) Clarke, P. Exclusive: micron drops phase-change memory−for now. *Electronics 360* **2014**.
(65) Close, G. F.; Frey, U.; Morrish, J.; Jordan, R.; Lewis, S.; Maffitt, T.; Breitwisch, M.; Hagleitner, C.; Lam, C.; Eleftheriou, E. A 512 Mb phase-change memory (PCM) in 90 nm CMOS achieving 2b/cell. *2011 Symposium on VLSI Circuits - Digest of Technical Papers*, 2011; pp 202−203.
(66) Close, G. F.; Frey, U.; Morrish, J.; Jordan, R.; Lewis, S. C.; Maffitt, T.; BrightSky, M. J.; Hagleitner, C.; Lam, C. H.; Eleftheriou, E. A 256-Mcell Phase-Change Memory Chip Operating at 2{+} Bit/Cell. *IEEE Transactions on Circuits and Systems I: Regular Papers* **2013**, *60*, 1521−1533.
(67) Clarke, P. Patent search supports view 3D XPoint based on phase-change. *EE Times*, Jul **2015**, *31*.
(68) Marks, H. Why Intel Optane SSDs have 'failed'. *Tech Target*, Jun **2018**.
(69) Yi, J.; Kim, M.; Seo, J.; Park, N.; Lee, S.; Kim, J.; Do, G.; Jang, H.; Koo, H.; Cho, S. The chalcogenide-based memory technology continues: beyond 20 nm 4-deck 256 gb cross-point memory. *2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2023; p 363.
(70) Arnaud, F.; Zuliani, P.; Reynard, J.; Gandolfo, A.; Disegni, F.; Mattavelli, P.; Gomiero, E.; Samanni, G.; Jahan, C.; Berthelon, R.; Weber, O.; Richard, E.; Barral, V.; Villeret, E.; Kohler, S.; Grenier, J.; Ranica, R.; Gallon, C.; Souhaite, A.; Risfolu, D.; Favennec, L.; Caubet, V.; DelMedico, S.; Cherault, N.; Beneyton, R.; Chouteau, S.; Sassoulas, P.; Vernhet, A.; Friec, Y. L.; Domengle, F.; Scotti, L.; Pacelli, P.; Ogler, J.; Boucard, F.; Lagrasta, S.; Benoit, D.; Clement, L.; Bolvin, P.; Ferreira, P.; Annunziata, R.;et al.Truly innovative 28nm FDSOI technology for automotive micro-controller applications embedding 16MB phase change memory*IEEE International Electron Devices Meeting (IEDM)*201818.4.1.
(71) Khaddam-Aljameh, R.; Stanisavljevic, M.; Mas, J. F.; Karunaratne, G.; Braendli, M.; Liu, F.; Singh, A.; Müller, S. M.; Egger, U.; Petropoulos, A.; et al. HERMES Core−A 14nm CMOS and PCM-based In-Memory Compute Core using an array of 300ps/LSB Linearized CCO-based ADCs and local digital processing. *2021 Symposium on VLSI Circuits* **2021**, 1−2.
(72) Narayanan, P.; Ambrogio, S.; Okazaki, A.; Hosokawa, K.; Tsai, H.; Nomura, A.; Yasuda, T.; Mackin, C.; Lewis, S. C.; Friz, A.; Ishii, M.; Kohda, Y.; Mori, H.; Spoon, K.; Khaddam-Aljameh, R.; Saulnier, N.; Bergendahl, M.; Demarest, J.; Brew, K. W.; Chan, V.; Choi, S.; Ok, I.; Ahsan, I.; Lie, F. L.; Haensch, W.; Narayanan, V.; Burr, G. W. Fully on-chip MAC at 14 nm enabled by accurate row-wise programming of PCM-based weights and parallel vector-transport in duration-format. *IEEE Trans. Electron Devices* **2021**, *68*, 6629−6636.
(73) Khwa, W.-S.; Chiu, Y.-C.; Jhang, C.-J.; Huang, S.-P.; Lee, C.-Y.; Wen, T.-H.; Chang, F.-C.; Yu, S.-M.; Lee, T.-Y.; Chang, M.-F. A 40 nm, 2M cell, 8b precision, hybrid SLC-MLC PCM computing-in-memory macro with 20.5−65.0 TOPS/W for tiny-Al edge devices. *International Solid-State Circuits Conference (ISSCC)*, 2022; pp 1−3.
(74) Le Gallo, M.; Khaddam-Aljameh, R.; Stanisavljevic, M.; Vasilopoulos, A.; Kersting, B.; Dazzi, M.; Karunaratne, G.; Brandli, M.; Singh, A.; Muller, S. M.; Buchel, J.; Timoneda, X.; Joshi, V.; Rasch, M. J.; Egger, U.; Garofalo, A.; Petropoulos, A.; Antonakopoulos, T.; Brew, K.; Choi, S.; Ok, I.; Philip, T.; Chan, V.; Silvestre, C.; Ahsan, I.; Saulnier, N.; Narayanan, V.; Francese, P. A.; Eleftheriou, E.; Sebastian, A. A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference. *Nature Electronics* **2023**, *6*, 680−693.
(75) Ambrogio, S.; Narayanan, P.; Okazaki, A.; Fasoli, A.; Mackin, C.; Hosokawa, K.; Nomura, A.; Yasuda, T.; Chen, A.; Friz, A.; Ishii, M.; Luquin, J.; Kohda, Y.; Saulnier, N.; Brew, K.; Choi, S.; Ok, I.; Philip, T.; Chan, V.; Silvestre, C.; Ahsan, I.; Narayanan, V.; Tsai, H.; Burr, G. W. An analog-AI chip for energy-efficient speech recognition and transcription. *Nature* **2023**, *620*, 768.
(76) Syed, G. S.; Brew, K.; Vasilopoulos, A.; Jonnalagadda, V. P.; Kersting, B.; Philip, T.; Bragaglia, V.; Ambrogio, S.; Büchel, J.; Giannopoulos, J.; et al. In-Memory Compute Chips with Carbon-based Projected Phase-Change Memory Devices. *IEEE International Electron Devices Meeting (IEDM)* **2023**, 1−4.
(77) Orava, J.; Greer, A. á.; Gholipour, B.; Hewak, D.; Smith, C. Characterization of supercooled liquid Ge2Sb2Te5 and its crystallization by ultrafast-heating calorimetry. *Nature materials* **2012**, *11*, 279−283.
(78) Pries, J.; Weber, H.; Benke-Jacob, J.; Kaban, I.; Wei, S.; Wuttig, M.; Lucas, P. Fragile-to-Strong Transition in Phase-Change Material Ge3Sb6Te5. *Adv. Funct. Mater.* **2022**, *32*, 2202714.
(79) Zalden, P.; Quirin, F.; Schumacher, M.; Siegel, J.; Wei, S.; Koc, A.; Nicoul, M.; Trigo, M.; Andreasson, P.; Enquist, H.; Shu, M. J.; Pardini, T.; Chollet, M.; Zhu, D.; Lemke, H.; Ronneberger, I.; Larsson, J.; Lindenberg, A. M.; Fischer, H. E.; Hau-Riege, S.; Reis, D. A.; Mazzarello, R.; Wuttig, M.; Sokolowski-Tinten, K. Femtosecond x-ray diffraction reveals a liquid−liquid phase transition in phase-change materials. *Science* **2019**, *364*, 1062−1067.
(80) Sakka, S.; Mackenzie, J. High pressure effects on glass. *J. Non-Cryst. Solids* **1969**, *1*, 107−142.
(81) Redaelli, A.; Petroni, E.; Annunziata, R. Material and process engineering challenges in Ge-rich GST for embedded PCM. *Materials Science in Semiconductor Processing* **2022**, *137*, 106184.
(82) Salinga, M.; Carria, E.; Kaldenbach, A.; Bornhöfft, M.; Benke, J.; Mayer, J.; Wuttig, M. Measurement of crystal growth velocity in a melt-quenched phase-change material. *Nat. Commun.* **2013**, *4*, 2371.
(83) Wei, S.; Evenson, Z.; Stolpe, M.; Lucas, P.; Angell, C. A. Breakdown of the Stokes-Einstein relation above the melting temperature in a liquid phase-change material. *Science advances* **2018**, *4*, No. eaat8632.
(84) Sosso, G. C.; Behler, J.; Bernasconi, M. Breakdown of Stokes−Einstein relation in the supercooled liquid state of phase change materials. *physica status solidi (b)* **2012**, *249*, 1880−1885.
(85) Amini, N.; Pries, J.; Cheng, Y.; Persch, C.; Wuttig, M.; Stolpe, M.; Wei, S. Thermodynamics and kinetics of glassy and liquid phase-change materials. *Materials Science in Semiconductor Processing* **2021**, *135*, 106094.
(86) Lee, B.-S.; Shelby, R. M.; Raoux, S.; Retter, C. T.; Burr, G. W.; Bogle, S. N.; Darmawikarta, K.; Bishop, S. G.; Abelson, J. R. Nanoscale nuclei in phase change materials: Origin of different crystallization mechanisms of Ge2Sb2Te5 and AgInSbTe. *J. Appl. Phys.* **2014**, *115*, 063506.
(87) Sebastian, A.; Le Gallo, M.; Krebs, D. Crystal growth within a phase change memory cell. *Nat. Commun.* **2014**, *5*, 4314.
(88) Mocanu, F. C.; Konstantinou, K.; Lee, T. H.; Bernstein, N.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Modeling the phase-change memory material, Ge2Sb2Te5, with a machine-learned interatomic potential. *J. Phys. Chem. B* **2018**, *122*, 8998−9006.
(89) Abou El Kheir, O.; Bonati, L.; Parrinello, M.; Bernasconi, M. Unraveling the crystallization kinetics of the Ge2Sb2Te5 phase change compound with a machine-learned interatomic potential. *npj Computational Materials* **2024**, *10*, 33.
(90) Kolobov, A.; Fons, P.; Tominaga, J. Understanding phase-change memory alloys from a chemical perspective. *Sci. Rep.* **2015**, *5*, 13698.
(91) Yamada, N.; Matsunaga, T. Structure of laser-crystallized Ge 2 Sb 2+ x Te 5 sputtered thin films for use in optical memory. *J. Appl. Phys.* **2000**, *88*, 7020−7028.
(92) Bragaglia, V.; Arciprete, F.; Zhang, W.; Mio, A. M.; Zallo, E.; Perumal, K.; Giussani, A.; Cecchi, S.; Boschker, J. E.; Riechert, H.; Privitera, S.; Rimini, E.; Mazzarello, R.; Calarco, R. others Metal-insulator transition driven by vacancy ordering in GeSbTe phase change materials. *Sci. Rep.* **2016**, *6*, 23843.
(93) Kooi, B. J.; Wuttig, M. Chalcogenides by design: functionality through metavalent bonding and confinement. *Adv. Mater.* **2020**, *32*, 1908302.
(94) Guarneri, L.; Jakobs, S.; von Hoegen, A.; Maier, S.; Xu, M.; Zhu, M.; Wahl, S.; Teichrib, C.; Zhou, Y.; Cojocaru-Miredin, O.;

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

Raghuwanshi, M.; Schon, C.-F.; Drogeler, M.; Stampfer, C.; Lobo, R. P. S. M.; Piarristeguy, A.; Pradel, A.; Raty, J.-Y.; Wuttig, M. others Metavalent bonding in crystalline solids: how does it collapse? *Adv. Mater.* **2021**, *33*, 2102356.
(95) Jones, R. O.; Elliott, S. R.; Dronskowski, R. The Myth of "Metavalency" in Phase-Change Materials. *Adv. Mater.* **2023**, *35*, 2300836.
(96) Lee, T. H.; Elliott, S. R. Chemical bonding in chalcogenides: the concept of multicenter hyperbonding. *Adv. Mater.* **2020**, *32*, 2000340.
(97) Turnbull, D. Under what conditions can a glass be formed? *Contemporary physics* **1969**, *10*, 473−488.
(98) Zhong, L.; Wang, J.; Sheng, H.; Zhang, Z.; Mao, S. X. Formation of monatomic metallic glasses through ultrafast liquid quenching. *Nature* **2014**, *512*, 177−180.
(99) Kersting, B.; Sarwat, S. G.; Le Gallo, M.; Brew, K.; Walfort, S.; Saulnier, N.; Salinga, M.; Sebastian, A. Measurement of Onset of Structural Relaxation in Melt-Quenched Phase Change Materials. *Adv. Funct. Mater.* **2021**, *31*, 2104422.
(100) Pirovano, A.; Lacaita, A. L.; Pellizzer, F.; Kostylev, S. A.; Benvenuti, A.; Bez, R. Low-field amorphous state resistance and threshold voltage drift in chalcogenide materials. *IEEE Trans. Electron Devices* **2004**, *51*, 714−719.
(101) Le Gallo, M.; Krebs, D.; Zipoli, F.; Salinga, M.; Sebastian, A. Collective structural relaxation in phase-change memory devices. *Advanced Electronic Materials* **2018**, *4*, 1700627.
(102) Peng, S.-X.; Cheng, Y.; Pries, J.; Wei, S.; Yu, H.-B.; Wuttig, M. Uncovering β-relaxations in amorphous phase-change materials. *Science Advances* **2020**, *6*, No. eaay6726.
(103) Fugazza, D.; Ielmini, D.; Lavizzari, S.; Lacaita, A. L. Random telegraph signal noise in phase change memory devices. *IEEE International Reliability Physics Symposium* **2010**, 743−749.
(104) Nardone, M.; Kozub, V.; Karpov, I.; Karpov, V. Possible mechanisms for 1/f noise in chalcogenide glasses: A theoretical description. *Phys. Rev. B* **2009**, *79*, 165206.
(105) Le Gallo, M.; Athmanathan, A.; Krebs, D.; Sebastian, A. Evidence for thermally assisted threshold switching behavior in nanoscale phase-change memory cells. *J. Appl. Phys.* **2016**, *119*, 025704.
(106) Ielmini, D.; Zhang, Y. Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices. *J. Appl. Phys.* **2007**, *102*, 054517.
(107) Kaes, M.; Le Gallo, M.; Sebastian, A.; Salinga, M.; Krebs, D. High-field electrical transport in amorphous phase-change materials. *J. Appl. Phys.* **2015**, *118*, 135707.
(108) Fantini, P.; Ferro, M.; Calderoni, A.; Brazzelli, S. Disorder enhancement due to structural relaxation in amorphous Ge2Sb2Te5. *Appl. Phys. Lett.* **2012**, *100*, 213506.
(109) Le Gallo, M.; Kaes, M.; Sebastian, A.; Krebs, D. Subthreshold electrical transport in amorphous phase-change materials. *New J. Phys.* **2015**, *17*, 093035.
(110) Lee, S.; Nathan, A.; Ye, Y.; Guo, Y.; Robertson, J. Localized tail states and electron mobility in amorphous ZnON thin film transistors. *Sci. Rep.* **2015**, *5*, 13467.
(111) Wahid, S.; Daus, A.; Khan, A. I.; Chen, V.; Neilson, K. M.; Islam, M.; Chen, M. E.; Pop, E. Lateral electrical transport and field-effect characteristics of sputtered p-type chalcogenide thin films. *Appl. Phys. Lett.* **2021**, *119*, 232106.
(112) Rütten, M.; Geilen, A.; Sebastian, A.; Krebs, D.; Salinga, M. Localised states and their capture characteristics in amorphous phase-change materials. *Sci. Rep.* **2019**, *9*, 6592.
(113) Krebs, D.; Raoux, S.; Rettner, C. T.; Burr, G. W.; Salinga, M.; Wuttig, M. Threshold field of phase change memory materials measured using phase change bridge devices. *Appl. Phys. Lett.* **2009**, *95*, 082101.
(114) Noé, P.; Verdy, A.; d'Acapito, F.; Dory, J.-B.; Bernard, M.; Navarro, G.; Jager, J.-B.; Gaudin, J.; Raty, J.-Y. Toward ultimate nonvolatile resistive memories: The mechanism behind ovonic threshold switching revealed. *Science Advances* **2020**, *6*, No. eaay2830.
(115) Novielli, G.; Ghetti, A.; Varesi, E.; Mauri, A.; Sacco, R. Atomic migration in phase change materials. *2013 IEEE International Electron Devices Meeting*, 2013; pp 22−23.
(116) Crespi, L.; Lacaita, A.; Boniardi, M.; Varesi, E.; Ghetti, A.; Redaelli, A.; D'Arrigo, G. Modeling of atomic migration phenomena in phase change memory devices. *IEEE International Memory Workshop (IMW)* **2015**, *2015*, 1−4.
(117) Padilla, A.; Burr, G. W.; Rettner, C. T.; Topuria, T.; Rice, P. M.; Jackson, B.; Virwani, K.; Kellock, A. J.; Dupouy, D.; Debunne, A.; Shelby, R. M.; Gopalakrishnan, K.; Shenoy, R. S.; Kurdi, B. N. others Voltage polarity effects in $Ge_2Sb_2Te_5$-based phase change memory devices. *J. Appl. Phys.* **2011**, *110*, 054501.
(118) Xie, Y.; Kim, W.; Kim, Y.; Kim, S.; Gonsalves, J.; BrightSky, M.; Lam, C.; Zhu, Y.; Cha, J. J. Self-healing of a confined phase change memory device with a metallic surfactant layer. *Adv. Mater.* **2018**, *30*, 1705587.
(119) Zhou, X.; Dong, W.; Zhang, H.; Simpson, R. E. A zero density change phase change memory material: GeTe-O structural characteristics upon crystallisation. *Sci. Rep.* **2015**, *5*, 11150.
(120) Zhou, X.; Wu, L.; Song, Z.; Rao, F.; Ren, K.; Peng, C.; Guo, X.; Liu, B.; Feng, S. Study on interface adhesion between phase change material film and SiO2 layer by nanoscratch test. *Jpn. J. Appl. Phys.* **2011**, *50*, 091402.
(121) Joshi, V.; Le Gallo, M.; Haefeli, S.; Boybat, I.; Nandakumar, S. R.; Piveteau, C.; Dazzi, M.; Rajendran, B.; Sebastian, A.; Eleftheriou, E. Accurate deep neural network inference using computational phase-change memory. *Nat. Commun.* **2020**, *11*, 2473.
(122) Büchel, J.; Vasilopoulos, A.; Kersting, B.; Lammie, C.; Brew, K.; Philip, T.; Saulnier, N.; Narayanan, V.; Le Gallo, M.; Sebastian, A. Programming weights to analog in-memory computing cores by direct minimization of the matrix-vector multiplication error. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **2023**, *13*, 1052−1051.
(123) Syed, G. S. Analog Device Non-Idealities Tutorial. Thesis. 2025 (accessed: 2025-01-14).
(124) Sarwat, S. G.; Le Gallo, M.; Bruce, R. L.; Brew, K.; Kersting, B.; Jonnalagadda, V. P.; Ok, I.; Saulnier, N.; BrightSky, M.; Sebastian, A. Mechanism and Impact of Bipolar Current Voltage Asymmetry in Computational Phase-Change Memory. *Adv. Mater.* **2023**, *35*, 2201238.
(125) Boybat, I.; Kersting, B.; Sarwat, S. G.; Timoneda, X.; Bruce, R.; BrightSky, M.; Le Gallo, M.; Sebastian, A. Temperature sensitivity of analog in-memory computing using phase-change memory. *2021 IEEE International Electron Devices Meeting (IEDM)*, 2021; pp 28.
(126) Papandreou, N.; Pozidis, H.; Pantazi, A.; Sebastian, A.; Breitwisch, M.; Lam, C.; Eleftheriou, E. Programming algorithms for multilevel phase-change memory. *International Symposium on Circuits and Systems (ISCAS)* **2011**, *2011*, 329−332.
(127) Nandakumar, S.; Boybat, I.; Han, J.-P.; Ambrogio, S.; Adusumilli, P.; Bruce, R. L.; BrightSky, M.; Rasch, M.; Le Gallo, M.; Sebastian, A. Precision of synaptic weights programmed in phase-change memory devices for deep learning inference. *IEEE International Electron Devices Meeting (IEDM)* **2020**, *2020*, 29.4.1.
(128) Boniardi, M.; Redaelli, A.; Cupeta, C.; Pellizzer, F.; Crespi, L.; D'Arrigo, G.; Lacaita, A. L.; Servalli, G. Optimization metrics for Phase Change Memory (PCM) cell architectures. *2014 IEEE International Electron Devices Meeting*, **2014**; *2014*, 29.1.1.
(129) Park, S.-O.; Hong, S.; Sung, S.-J.; Kim, D.; Seo, S.; Jeong, H.; Park, T.; Cho, W. J.; Kim, J.; Choi, S. Phase-change memory via a phase-changeable self-confined nano-filament. *Nature* **2024**, *628*, 293−298.
(130) Rajendran, B.; Breitwisch, M.; Lee, M.-H.; Burr, G. W.; Shih, Y.-H.; Cheek, R.; Schrott, A.; Chen, C.-F.; Joseph, E.; Dasaka, R. Dynamic resistance—A metric for variability characterization of phase-change memory. *IEEE electron device letters* **2009**, *30*, 126−129.
(131) Ha, Y.; Yi, J.; Horii, H.; Park, J.; Joo, S.; Park, S.; Chung, U.-I.; Moon, J. An edge contact type cell for phase change RAM featuring very low power consumption. *Digest of Technical Papers of the Symposium on VLSI Technology*, 2003; pp 175−176.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

(132) Koelmans, W. W.; Sebastian, A.; Jonnalagadda, V. P.; Krebs, D.; Dellmann, L.; Eleftheriou, E. Projected phase-change memory devices. *Nat. Commun.* **2015**, *6*, 8181.
(133) Ghazi Sarwat, S.; Philip, T. M.; Chen, C.-T.; Kersting, B.; Bruce, R. L.; Cheng, C.-W.; Li, N.; Saulnier, N.; BrightSky, M.; Sebastian, A. Projected Mushroom Type Phase-Change Memory. *Adv. Funct. Mater.* **2021**, *31*, 2106547.
(134) Bruce, R. L.; Sarwat, S. G.; Boybat, I.; Cheng, C.-W.; Kim, W.; Nandakumar, S.; Mackin, C.; Philip, T.; Liu, Z.; Brew, K.; others Mushroom-type phase change memory with projection liner: An array-level demonstration of conductance drift and noise mitigation. *2021 IEEE International Reliability Physics Symposium (IRPS)*. 2021; pp 1−6.
(135) Kim, W.; Bruce, R. L.; Masuda, T.; Fraczak, G.; Gong, N.; Adusumilli, P.; Ambrogio, S.; Tsai, H.; Bruley, J.; Han, J.-P.; others Confined PCM-based analog synaptic devices offering low resistance-drift and 1000 programmable states for deep learning. *2019 Symposium on VLSI Technology*. 2019; pp T66−T67.
(136) Khan, A. I.; Kwon, H.; Chen, M. E.; Asheghi, M.; Wong, H.-S. P.; Goodson, K. E.; Pop, E. Electro-thermal confinement enables improved superlattice phase change memory. *IEEE Electron Device Lett.* **2022**, *43*, 204−207.
(137) Okabe, K. L.; Sood, A.; Yalon, E.; Neumann, C. M.; Asheghi, M.; Pop, E.; Goodson, K. E.; Wong, H.-S. P. Understanding the switching mechanism of interfacial phase change memory. *J. Appl. Phys.* **2019**, *125*, 184501.
(138) Chong, T.; Shi, L.; Qiang, W.; Tan, P.; Miao, X.; Hu, X. Superlattice-like structure for phase change optical recording. *Journal of applied physics* **2002**, *91*, 3981−3987.
(139) Khan, A. I.; Perez, C.; Wu, X.; Won, B.; Kim, K.; Kwon, H.; Ramesh, P.; Neilson, K. M.; Asheghi, M.; Saraswat, K.; others First Demonstration of GST-Based Superlattice Phase Change Memory with Low Reset Current Density and Low Resistance Drift. *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2022; pp 310−311.
(140) Ding, K.; Wang, J.; Zhou, Y.; Tian, H.; Lu, L.; Mazzarello, R.; Jia, C.; Zhang, W.; Rao, F.; Ma, E. Phase-change heterostructure enables ultralow noise and drift for memory operation. *Science* **2019**, *366*, 210−215.
(141) Ghazi Sarwat, S.; Philip, T. M.; Chen, C.-T.; Kersting, B.; Bruce, R. L.; Cheng, C.-W.; Li, N.; Saulnier, N.; BrightSky, M.; Sebastian, A. Projected Mushroom Type Phase-Change Memory. *Adv. Funct. Mater.* **2021**, *31*, 2106547.
(142) Prili, S.; Bragaglia, V.; Jonnalagadda, V. P.; Luchtenveld, J.; Arciprete, F.; Kooi, B. J.; Sebastian, A.; Syed, G. S. Understanding the Growth and Properties of Sputter-Deposited Phase-Change Superlattice Films. *arXiv* 2024, arXiv:2411.19343
(143) Wu, X.; Khan, A. I.; Lee, H.; Hsu, C.-F.; Zhang, H.; Yu, H.; Roy, N.; Davydov, A. V.; Takeuchi, I.; Bao, X.; Wong, H.-S. P.; Pop, E. Novel nanocomposite-superlattices for low energy and high stability nanoscale phase-change memory. *Nat. Commun.* **2024**, *15*, 13.
(144) Zhao, J.; Khan, A. I.; Efremov, M. Y.; Ye, Z.; Wu, X.; Kim, K.; Lee, Z.; Wong, H.-S. P.; Pop, E.; Allen, L. H. Probing the melting transitions in phase-change superlattices via thin film nanocalorimetry. *Nano Lett.* **2023**, *23*, 4587−4594.
(145) Lee, B. C.; Ipek, E.; Mutlu, O.; Burger, D. Architecting phase change memory as a scalable dram alternative. *Proceedings of the 36th annual international symposium on Computer architecture*. 2009; pp 2−13.
(146) Noé, P.; Hippert, F. Chapter 10: Phase Change Memory: Device Physics, *Reliability and Applications*. 2018.
(147) Neumaier, D.; Pindl, S.; Lemme, M. C. Integrating graphene into semiconductor fabrication lines. *Nature materials* **2019**, *18*, 525−529.
(148) Zuliani, P.; Varesi, E.; Palumbo, E.; Borghi, M.; Tortorelli, I.; Erbetta, D.; Dalla Libera, G.; Pessina, N.; Gandolfo, A.; Prelini, C.; Ravazzi, L.; Annunziata, R. Overcoming Temperature Limitations in Phase Change Memories With Optimized GexSbyTez, Electron Devices. *IEEE Trans. Electron Devices*, 2013; Vol. *60*, p 4020.
(149) Khan, A. I.; Daus, A.; Islam, R.; Neilson, K. M.; Lee, H. R.; Wong, H.-S. P.; Pop, E. Ultralow−switching current density multilevel phase-change memory on a flexible substrate. *Science* **2021**, *373*, 1243−1247.
(150) Syed, G. S.; Sebastian, A. *Device with reconfigurable short term data retention*. 2024; US Patent *12*,046,282.
(151) Cheng, H.-Y.; Carta, F.; Chien, W.-C.; Lung, H.-L.; BrightSky, M. J. 3D cross-point phase-change memory for storage-class memory. *J. Phys. D: Appl. Phys.* **2019**, *52*, 473002.
(152) *Intel and Micron Produce Breakthrough Memory Technology 3D XPoint*; Intel, 2015; https://en.wikipedia.org/wiki/3D_XPoint.
(153) Liang, J.; Yeh, S.; Wong, S. S.; Wong, H.-S. P. Scaling Challenges for the Cross-Point Resistive Memory Array to Sub-10nm Node - An Interconnect Perspective. *4th IEEE International Memory Workshop* **2013**, *9*, 1−14.
(154) Yoo, S.; Lee, H. D.; Lee, S.; Choi, H.; Kim, T. Electro-thermal model for thermal disturbance in cross-point phase-change memory. *IEEE Trans. Electron Devices* **2020**, *67*, 1454−1459.
(155) Lee, S.; Kim, M.; Do, G.; Kim, S.; Lee, H.; Sim, J.; Park, N.; Hong, S.; Jeon, Y.; Choi, K.; Programming disturbance and cell scaling in phase change memory: For up to 16 nm based 4f 2 cell. *2010 Symposium on VLSI Technology*. 2010; pp 199−200.
(156) Giannopoulos, I.; Singh, A.; Le Gallo, M.; Jonnalagadda, V. P.; Hamdioui, S.; Sebastian, A. In-Memory Database Query. *Advanced Intelligent Systems* **2020**, *2*, 2000141.
(157) Le Gallo, M.; Sebastian, A.; Cherubini, G.; Giefers, H.; Eleftheriou, E. Compressed sensing recovery using computational memory. *International Electron Devices Meeting (IEDM)* **2017**, 28−3.
(158) Rahimi, A.; Datta, S.; Kleyko, D.; Frady, E. P.; Olshausen, B.; Kanerva, P.; Rabaey, J. M. High-dimensional computing as a nanoscalable paradigm. *IEEE Transactions on Circuits and Systems I: Regular Papers* **2017**, *64*, 2508−2521.
(159) Wu, T. F.; Li, H.; Huang, P.; Rahimi, A.; Rabaey, J. M.; Wong, H. P.; Shulaker, M. M.; Mitra, S. Brain-inspired computing exploiting carbon nanotube FETs and resistive RAM: Hyperdimensional computing case study. *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018; pp 492−494.
(160) Li, H. et al. Hyperdimensional computing with 3D VRRAM in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition. *2016 IEEE International Electron Devices Meeting (IEDM)*. 2016; Vol. *2016*, p 16.1.1.
(161) Karunaratne, G.; Le Gallo, M.; Cherubini, G.; Benini, L.; Rahimi, A.; Sebastian, A. In-memory hyperdimensional computing. *Nature Electronics* **2020**, *3*, 327−337.
(162) Rasch, M. J.; Mackin, C.; Le Gallo, M.; Chen, A.; Fasoli, A.; Odermatt, F.; Li, N.; Nandakumar, S. R.; Narayanan, P.; Tsai, H.; Burr, G. W.; Sebastian, A.; Narayanan, V. others Hardware-aware training for large-scale and diverse deep learning inference workloads using in-memory computing-based accelerators. *Nat. Commun.* **2023**, *14*, 5282.
(163) Le Gallo, M.; Sebastian, A.; Cherubini, G.; Giefers, H.; Eleftheriou, E. Compressed Sensing With Approximate Message Passing Using In-Memory Computing. *IEEE Trans. Electron Devices* **2018**, *65*, 4304−4312.
(164) Cohen, G.; Majumdar, A.; Cheng, C.-W.; Ray, A.; Piatek, D.; Gignac, L.; Lavoie, C.; Grun, A.; Cheng, M.; Liu, Z.-L.; lan Lung, H.; BrightSky, M. 1T1R mushroom PCM cell device using epitaxial GeSbTe. *Proceedings of the European Phase-Change and Ovonic Symposium (E). 2023; Presented at the European Phase-Change and Ovonic Symposium*, 2023.
(165) Giannopoulos, I.; Sebastian, A.; Le Gallo, M.; Jonnalagadda, V.; Sousa, M.; Boon, M.; Eleftheriou, E. 8-bit Precision In-Memory Multiplication with Projected Phase-Change Memory. *2018 IEEE International Electron Devices Meeting (IEDM)* **2019**, *52*, 27.
(166) Li, C.; Hu, C.; Wang, J.; Yu, X.; Yang, Z.; Liu, J.; Li, Y.; Bi, C.; Zhou, X.; Zheng, W. Understanding phase-change materials with

unexpectedly low resistance drift for phase-change memories. *Journal of Materials Chemistry C* **2018**, *6*, 3387−3394.
(167) Le Gallo, M.; Hrynkevych, O.; Kersting, B.; Karunaratne, G.; Vasilopoulos, A.; Khaddam-Aljameh, R.; Syed, G. S.; Sebastian, A. Demonstration of 4-quadrant analog in-memory matrix multiplication in a single modulation. *npj Unconventional Computing* **2024**, *1*, 11.
(168) Vasilopoulos, A.; Buchel, J.; Kersting, B.; Lammie, C.; Brew, K.; Choi, S.; Philip, T.; Saulnier, N.; Narayanan, V.; Le Gallo, M.; Sebastian, A.; others Exploiting the state dependency of conductance variations in memristive devices for accurate in-memory computing. *IEEE Trans. Electron Devices*, 2023, Vol. *70*, p 6279.
(169) Sebastian, A.; Tuma, T.; Papandreou, N.; Le Gallo, M.; Kull, L.; Parnell, T.; Eleftheriou, E. Temporal correlation detection using computational phase-change memory. *Nat. Commun.* **2017**, *8*, 1115.
(170) Le Gallo, M.; Tuma, T.; Zipoli, F.; Sebastian, A.; Eleftheriou, E. Inherent stochasticity in phase-change memory devices. *2016 46th European Solid-State Device Research Conference (ESSDERC)* **2016**, *37*, 1238.
(171) Go, S.-X.; Wang, Q.; Lim, K. G.; Lee, T. H.; Bajalovic, N.; Loke, D. K. Ultrafast Near-Ideal Phase-Change Memristive Physical Unclonable Functions Driven by Amorphous State Variations. *Advanced Science* **2022**, *9*, 2204453.
(172) Wright, C. D.; Hosseini, P.; Diosdado, J. A. V. Beyond von-Neumann computing with nanoscale phase-change memory devices. *Adv. Funct. Mater.* **2013**, *23*, 2248−2254.
(173) Hoffer, B.; Wainstein, N.; Neumann, C. M.; Pop, E.; Yalon, E.; Kvatinsky, S. Stateful logic using phase change memory. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* **2022**, *8*, 77−83.
(174) Burr, G. W.; Shelby, R. M.; Sidler, S.; di Nolfo, C.; Jang, J.; Boybat, I.; Shenoy, R. S.; Narayanan, P.; Virwani, K.; Giacometti, E. U.; Kurdi, B. N.; Hwang, H. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element. *IEEE Trans. Electron Devices* **2015**, *62*, 3498−3507.
(175) Ambrogio, S.; Narayanan, P.; Tsai, H.; Shelby, R. M.; Boybat, I.; di Nolfo, C.; Sidler, S.; Giordano, M.; Bodini, M.; Farinha, N. C. P.; Killeen, B.; Cheng, C.; Jaoudi, Y.; Burr, G. W. others Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **2018**, *558*, 60.
(176) Nandakumar, S. R.; Le Gallo, M.; Piveteau, C.; Joshi, V.; Mariani, G.; Boybat, I.; Karunaratne, G.; Khaddam-Aljameh, R.; Egger, U.; Petropoulos, A.; Antonakopoulos, T.; Rajendran, B.; Sebastian, A.; Eleftheriou, E. Mixed-Precision Deep Learning Based on Computational Memory. *Frontiers in Neuroscience* **2020**, *14*, 406.
(177) Nandakumar, S.; Le Gallo, M.; Boybat, I.; Rajendran, B.; Sebastian, A.; Eleftheriou, E. Mixed-precision architecture based on computational memory for training deep neural networks. *International Symposium on Circuits and Systems (ISCAS)* **2018**, *2018*, 1−5.
(178) Rasch, M. J.; Carta, F.; Fagbohungbe, O.; Gokmen, T. Fast offset corrected in-memory training. *arXiv* 2023, arXiv:2303.04721.
(179) Bichler, O.; Suri, M.; Querlioz, D.; Vuillaume, D.; DeSalvo, B.; Gamrat, C. Visual pattern extraction using energy-efficient 2-PCM synapse neuromorphic architecture. *IEEE Trans. Electron Devices* **2012**, *59*, 2206−2214.
(180) Tuma, T.; Le Gallo, M.; Sebastian, A.; Eleftheriou, E. Detecting correlations using phase-change neurons and synapses. *IEEE Electron Device Lett.* **2016**, *37*, 1238−1241.
(181) Pantazi, A.; Woźniak, S.; Tuma, T.; Eleftheriou, E. All-memristive neuromorphic computing with level-tuned neurons. *Nanotechnology* **2016**, *27*, 355205.
(182) Moraitis, T.; Sebastian, A.; Eleftheriou, E. The role of short-term plasticity in neuromorphic learning: Learning from the timing of rate-varying events with fatiguing spike-timing-dependent plasticity. *IEEE Nanotechnology Magazine* **2018**, *12*, 45−53.
(183) Kuzum, D.; Jeyasingh, R. G.; Lee, B.; Wong, H.-S. P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* **2012**, *12*, 2179−2186.
(184) Kim, S.; Ishii, M.; Lewis, S.; Perri, T.; BrightSky, M.; Kim, W.; Jordan, R.; Burr, G.; Sosa, N.; Ray, A. NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning. *International Electron Devices Meeting (IEDM)*, 2015, Vol. 17, p. 1.
(185) Kheradpisheh, S. R.; Ganjtabesh, M.; Thorpe, S. J.; Masquelier, T. STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks* **2018**, *99*, 56−67.
(186) Mozafari, M.; Ganjtabesh, M.; Nowzari-Dalini, A.; Thorpe, S. J.; Masquelier, T. Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks. *Pattern recognition* **2019**, *94*, 87−95.
(187) Karunaratne, G.; Schmuck, M.; Le Gallo, M.; Cherubini, G.; Benini, L.; Sebastian, A.; Rahimi, A. Robust high-dimensional memory-augmented neural networks. *Nat. Commun.* **2021**, *12*, 2468.
(188) Karunaratne, G.; Hersche, M.; Langeneager, J.; Cherubini, G.; Le Gallo, M.; Egger, U.; Brew, K.; Choi, S.; Ok, I.; Silvestre, C.; et al. In-memory realization of in-situ few-shot continual learning with a dynamically evolving explicit memory. *arXiv* **2022**, arXiv:2207.06810−108.
(189) Tuma, T.; Pantazi, A.; Le Gallo, M.; Sebastian, A.; Eleftheriou, E. Stochastic phase-change neurons. *Nature Nanotechnol.* **2016**, *11*, 693−699.
(190) Lu, Y.; Li, X.; Yan, B.; Yan, L.; Zhang, T.; Song, Z.; Huang, R.; Yang, Y. In-memory realization of eligibility traces based on conductance drift of phase change memory for energy-efficient reinforcement learning. *Adv. Mater.* **2022**, *34*, 2107811.
(191) Langenegger, J.; Karunaratne, G.; Hersche, M.; Benini, L.; Sebastian, A.; Rahimi, A. In-memory factorization of holographic perceptual representations. *Nat. Nanotechnol.* **2023**, *18*, 479.
(192) Quintino Palhares, J. H.; Garg, N.; Mouny, P.-A.; Beilliard, Y.; Sandrini, J.; Arnaud, F.; Anghel, L.; Alibart, F.; Drouin, D.; Galy, P. 28 nm FDSOI embedded PCM exhibiting near zero drift at 12 K for cryogenic SNNs. *npj Unconventional Computing* **2024**, *1*, 8.
(193) Demirağ, Y.; Moro, F.; Dalgaty, T.; Navarro, G.; Frenkel, C.; Indiveri, G.; Vianello, E.; Payvand, M. PCM-trace: scalable synaptic eligibility traces with resistivity drift of phase-change materials. *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021; pp 1−5.
(194) Gerstner, W.; Lehmann, M.; Liakoni, V.; Corneil, D.; Brea, J. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits* **2018**, *12*, 53.
(195) Sarwat, S. G.; Moraitis, T.; Wright, C. D.; Bhaskaran, H. Chalcogenide optomemristors for multi-factor neuromorphic computation. *Nat. Commun.* **2022**, *13*, 2247.
(196) Syed, G. S.; Kersting, B.; Egger, U.; Sebastian, A. Phase change computational sensor. *npj Unconventional Computing* **2025**, *2*, 1.
(197) Kovats, T.; Rameshan, N.; Karunaratne, K. G.; Giannopoulos, I.; Sebastian, A. In-memory encryption using the advanced encryption standard. *Philosophical Transactions A* **2025**, *383*, 20230396.
(198) Xie, M.; Li, S.; Glova, A. O.; Hu, J.; Wang, Y.; Xie, Y. AIM: Fast and energy-efficient AES in-memory implementation for emerging non-volatile main memory. *2018 Design, Automation & Test in Europe Conference Exhibition (DATE)* **2018**, *2018*, 625−628.
(199) Kang, Y. G.; Ishii, M.; Park, J.; Shin, U.; Jang, S.; Yoon, S.; Kim, M.; Okazaki, A.; Ito, M.; Nomura, A.; Hosokawa, K.; BrightSky, M.; Kim, S. others Solving Max-Cut Problem Using Spiking Boltzmann Machine Based on Neuromorphic Hardware with Phase Change Memory. *Advanced Science* **2024**, *11*, 2406433.
(200) Yan, L.; Wu, Q.; Li, X.; Xie, C.; Zhou, X.; Li, Y.; Shi, D.; Yu, L.; Zhang, T.; Tao, Y.; Yan, B.; Zhong, M.; Song, Z.; Yang, Y.; Huang, R. Neural Architecture Search with In-Memory Multiply−Accumulate and In-Memory Rank Based on Coating Layer Optimized C-Doped $Ge_2Sb_2Te_5$ Phase Change Memory. *Adv. Funct. Mater.* **2024**, *34*, 2470083.
(201) Syed, G. S.; Rahimi, A.; Sebastian, A. Efficiently clustering data points with an in memory computing system. US Patent18/176684, 2024.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194

(202) Büchel, J.; Camposampiero, G.; Vasilopoulos, A.; Lammie, C.; Le Gallo, M.; Rahimi, A.; Sebastian, A. Kernel approximation using analogue in-memory computing. *Nature Machine Intelligence* **2024**, *6*, 1605−1615.
(203) Ríos, C.; Youngblood, N.; Cheng, Z.; Le Gallo, M.; Pernice, W. H.; Wright, C. D.; Sebastian, A.; Bhaskaran, H. In-memory computing on a photonic platform. *Science Adv.* **2019**, *5*, aau5759.
(204) Shastri, B. J.; Tait, A. N.; Ferreira de Lima, T.; Pernice, W. H.; Bhaskaran, H.; Wright, C. D.; Prucnal, P. R. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **2021**, *15*, 102−114.
(205) Ríos, C.; Stegmaier, M.; Hosseini, P.; Wang, D.; Scherer, T.; Wright, C. D.; Bhaskaran, H.; Pernice, W. H. Integrated all-photonic non-volatile multi-level memory. *Nat. Photonics* **2015**, *9*, 725−732.
(206) Feldmann, J.; Youngblood, N.; Karpov, M.; Gehring, H.; Li, X.; Stappers, M.; Le Gallo, M.; Fu, X.; Lukashchuk, A.; Raja, A. S.; Liu, J.; Wright, C. D.; Sebastian, A.; Kippenberg, T. J.; Pernice, W. H. P.; Bhaskaran, H. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **2021**, *589*, 52−58.
(207) Wu, C.; Yu, H.; Lee, S.; Peng, R.; Takeuchi, I.; Li, M. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **2021**, *12*, 96.
(208) Ghazi Sarwat, S.; Brückerhoff-Plückelmann, F.; Carrillo, S. G.-C.; Gemo, E.; Feldmann, J.; Bhaskaran, H.; Wright, C. D.; Pernice, W. H.; Sebastian, A. An integrated photonics engine for unsupervised correlation detection. *Science Advances* **2022**, *8*, eabn3243.
(209) Youngblood, N.; Ríos Ocampo, C. A.; Pernice, W. H.; Bhaskaran, H. Integrated optical memristors. *Nat. Photonics* **2023**, *17*, 561−572.
(210) Kato, K.; Kuwahara, M.; Kawashima, H.; Tsuruoka, T.; Tsuda, H. Current-driven phase-change optical gate switch using indium−tin-oxide heater. *Applied Physics Express* **2017**, *10*, 072201.
(211) Zhou, W.; Dong, B.; Farmakidis, N.; Li, X.; Youngblood, N.; Huang, K.; He, Y.; David Wright, C.; Pernice, W. H.; Bhaskaran, H. In-memory photonic dot-product engine with electrically programmable weight banks. *Nat. Commun.* **2023**, *14*, 2887.
(212) Martin-Monier, L.; Popescu, C. C.; Ranno, L.; Mills, B.; Geiger, S.; Callahan, D.; Moebius, M.; Hu, J. Endurance of chalcogenide optical phase change materials: a review. *Opt. Mater. Express* **2022**, *12*, 2145−2167.
(213) Varri, A.; Taheriniya, S.; Bruckerhoff-Pluckelmann, F.; Bente, I.; Farmakidis, N.; Bernhardt, D.; Rosner, H.; Kruth, M.; Nadzeyka, A.; Richter, T.; Wright, C. D.; Bhaskaran, H.; Wilde, G.; Pernice, W. H. P. Scalable non-volatile tuning of photonic computational memories by automated silicon ion implantation. *Adv. Mater.* **2024**, *36*, 2310596.
(214) Prabhathan, P.; Sreekanth, K. V.; Teng, J.; Ko, J. H.; Yoo, Y. J.; Jeong, H.-H.; Lee, Y.; Zhang, S.; Cao, T.; Popescu, C.-C.; Mills, B.; Gu, T.; Fang, Z.; Chen, R.; Tong, H.; Wang, Y.; He, Q.; Lu, Y.; Liu, Z.; Yu, H.; Mandal, A.; Cui, Y.; Ansari, A. S.; Bhingardive, V.; Kang, M.; Lai, C. K.; Merklein, M.; Muller, M. J.; Song, Y. M.; Tian, Z.; Hu, J.; Losurdo, M.; Majumdar, A.; Miao, X.; Chen, X.; Gholipour, B.; Richardson, K. A.; Eggleton, B. J.; Sharda, K.; Wuttig, M.; Singh, R. Roadmap for phase change materials in photonics and beyond. *Iscience* **2023**, *26*, 108396.
(215) Syed, G. S.; Sebastian, A. Solving optimization problems with photonic crossbars. US Patent 17/643256, 2023.
(216) Syed, G. S.; Sebastian, A. Photonic content-addressable memory and applications thereof. US Patent 17/938373, 2024.
(217) Ambrogio, S.; Narayanan, P.; Okazaki, A.; Fasoli, A.; Mackin, C.; Hosokawa, K.; Nomura, A.; Yasuda, T.; Chen, A.; Friz, A.; Ishii, M.; Luquin, J.; Kohda, Y.; Saulnier, N.; Brew, K.; Choi, S.; Ok, I.; Philip, T.; Chan, V.; Silvestre, C.; Ahsan, I.; Narayanan, V.; Tsai, H.; Burr, G. W. An analog-AI chip for energy-efficient speech recognition and transcription. *Nature* **2023**, *620*, 768−775.
(218) Rao, M.; Tang, H.; Wu, J.; Song, W.; Zhang, M.; Yin, W.; Zhuo, Y.; Kiani, F.; Chen, B.; Jiang, X.; Liu, H.; Chen, H.-Y.; Midya, R.; Ye, F.; Jiang, H.; Wang, Z.; Wu, M.; Hu, M.; Wang, H.; Xia, Q.; Ge, N.; Li, J.; Yang, J. J. Thousands of conductance levels in memristors integrated on CMOS. *Nature* **2023**, *615*, 823−829.
(219) Song, W.; Rao, M.; Li, Y.; Li, C.; Zhuo, Y.; Cai, F.; Wu, M.; Yin, W.; Li, Z.; Wei, Q.; Lee, S.; Zhu, H.; Gong, L.; Barnell, M.; Wu, Q.; Beerel, P. A.; Chen, M. S.-W.; Ge, N.; Hu, M.; Xia, Q.; Yang, J. J. Programming memristor arrays with arbitrarily high precision for analog computing. *Science* **2024**, *383*, 903−910.
(220) Zhang, W.; Yao, P.; Gao, B.; Liu, Q.; Wu, D.; Zhang, Q.; Li, Y.; Qin, Q.; Li, J.; Zhu, Z.; Cai, Y.; Wu, D.; Tang, J.; Qian, H.; Wang, Y.; Wu, H. Edge learning using a fully integrated neuro-inspired memristor chip. *Science* **2023**, *381*, 1205−1211.
(221) Wen, T.-H.; Hung, J.-M.; Huang, W.-H.; Jhang, C.-J.; Lo, Y.-C.; Hsu, H.-H.; Ke, Z.-E.; Chen, Y.-C.; Chin, Y.-H.; Su, C.-I; Khwa, W.-S.; Lo, C.-C.; Liu, R.-S.; Hsieh, C.-C.; Tang, K.-T.; Ho, M.-S.; Chou, C.-C.; Chih, Y.-D.; Chang, T.-Y. J.; Chang, M.-F. Fusion of memristor and digital compute-in-memory processing for energy-efficient edge computing. *Science* **2024**, *384*, 325−332.
(222) Giannopoulos, I.; Gallo, M. L.; Jonnalagadda, V. P.; Eleftheriou, E.; Sebastian, A. Temperature Compensation Schemes for In-Memory Computing using Phase-Change Memory. *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2020; pp 286−290
(223) Chen, A. A comprehensive crossbar array model with solutions for line resistance and nonlinear device characteristics. *IEEE Trans. Electron Devices* **2013**, *60*, 1318−1326.
(224) Jain, S.; Tsai, H.; Chen, C.-T.; Muralidhar, R.; Boybat, I.; Frank, M. M.; Woźniak, S.; Stanisavljevic, M.; Adusumilli, P.; Narayanan, P.; Hosokawa, K.; Ishii, M.; Kumar, A.; Narayanan, V.; Burr, G. W. A Heterogeneous and Programmable Compute-In-Memory Accelerator Architecture for Analog-AI Using Dense 2-D Mesh. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **2023**, *31*, 114−127.
(225) Ferro, E.; Vasilopoulos, A.; Lammie, C.; Le Gallo, M.; Benini, L.; Boybat, I.; Sebastian, A. A Precision-Optimized Fixed-Point Near-Memory Digital Processing Unit for Analog In-Memory Computing. *arXiv* **2024**, arXiv:2402.07549.
(226) Boybat, I.et al.; Heterogeneous Embedded Neural Processing Units Utilizing PCM-based Analog In-Memory Computing. *International Electron Devices Meeting (IEDM)*, 2024; Vol. *2024*, p 1.
(227) Dazzi, M.; Sebastian, A.; Benini, L.; Eleftheriou, E. Accelerating inference of convolutional neural networks using in-memory computing. *Frontiers in Computational Neuroscience* **2021**, *15*, 674154.
(228) Le Gallo, M.; Lammie, C.; Buchel, J.; Carta, F.; Fagbohungbe, O.; Mackin, C.; Tsai, H.; Narayanan, V.; Sebastian, A.; El Maghraoui, K.; Rasch, M. J. Using the IBM analog in-memory hardware acceleration kit for neural network training and inference. *APL Machine Learning* **2023**, *1*, 041102.

https://doi.org/10.1021/acs.chemrev.4c00670
Chem. Rev. 2025, 125, 5163−5194