Phase change memory technology

Geoffrey W. Burr, Matthew J. Breitwisch, Michele Franceschini, Davide Garetto, Kailash Gopalakrishnan, Bryan Jackson, Bülent Kurdi, Chung Lam, Luis A. Lastras, Alvaro Padilla, Bipin Rajendran, Simone Raoux, and Rohit S. Shenoy

REVIEW ARTICLE

Phase change memory technology

Geoffrey W. Burr[a]
IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120

Matthew J. Breitwisch and Michele Franceschini
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598

Davide Garetto, Kailash Gopalakrishnan, Bryan Jackson, and Bülent Kurdi
IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120

Chung Lam and Luis A. Lastras
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598

Alvaro Padilla and Bipin Rajendran
IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120

Simone Raoux
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598

Rohit S. Shenoy
IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120

The authors survey the current state of phase change memory (PCM), a nonvolatile solid-state memory technology built around the large electrical contrast between the highly resistive amorphous and highly conductive crystalline states in so-called phase change materials. PCM technology has made rapid progress in a short time, having passed older technologies in terms of both sophisticated demonstrations of scaling to small device dimensions, as well as integrated large-array demonstrators with impressive retention, endurance, performance, and yield characteristics. They introduce the physics behind PCM technology, assess how its characteristics match up with various potential applications across the memory-storage hierarchy, and discuss its strengths including scalability and rapid switching speed. Challenges for the technology are addressed, including the design of PCM cells for low reset current, the need to control device-to-device variability, and undesirable changes in the phase change material that can be induced by the fabrication procedure. They then turn to issues related to operation of PCM devices, including retention, device-to-device thermal cross-talk, endurance, and bias-polarity effects. Several factors that can be expected to enhance PCM in the future are addressed, including multilevel cell technology for PCM (which offers higher density through the use of intermediate resistance states), the role of coding, and possible routes to an ultrahigh-density PCM technology. © *2010 American Vacuum Society.*
[DOI: 10.1116/1.3301579]

I. MOTIVATION FOR PHASE CHANGE MEMORY

A. Case for a next-generation memory

As with many modern technologies, the extent to which nonvolatile memory (NVM) has pervaded our day-to-day lives is truly remarkable. From the music on our MP3 players, to the photographs on digital cameras, the stored e-mail and text messages on smart phones, the documents we carry on our USB thumb drives, and the program code that enables everything from our portable electronics to cars, the NVM known as Flash memory is everywhere around us. Both NOR and NAND Flash began humbly enough, as unappreciated side projects of a Toshiba DRAM engineer named Masuoka.[1] However, from his basic patents in 1980 and 1987,[1] Flash has grown in less than 3 decades to become a $20 billion/year titan of the semiconductor industry.[2,3]

This market growth has been made possible by tremendous increases in the system functionality (e.g., more gigabytes) that can be delivered in the same size package. These improvements are both a byproduct *of* and the driving force *for* the relentless march to smaller device dimensions known as Moore's law.[4] The history of the solid-state memory industry, and of the semiconductor industry as a whole, has been dominated by this concept: Higher densities at similar cost lead to more functionality, and thus more applications, which then spur investment for the additional re-

---

[a]Electronic mail: burr@almaden.ibm.com

search and development needed to implement the "next size smaller" device. Throughout this extensive history, extrapolation from the recent past has proven to be amazingly reliable for predicting near-future developments. Thus the memory products that will be built in the next several years have long been forecast.[5]

Beyond the near future, however, while the *planned* device sizes may be sketched out, for the first time in many years it is not clear exactly how achievable these goals might be. This uncertainty is present in many portions of the semiconductor industry, primarily due to the increasing importance of device-to-device variations, and to the common dependence on continued lithographic innovation. New patterning techniques will almost certainly be needed to replace the 193 nm immersion and "double patterning" techniques now being used to implement the 32 nm and even 22 nm nodes.[6,7] In addition to such issues common to the larger semiconductor industry, however, the Flash industry faces additional uncertainties specific to its technology.

Over the past few years, Flash has been wrestling with unpleasant tradeoffs between the scaling of lateral device dimensions, the need to maintain coupling between the control and floating gates, the stress-induced leakage current (SILC) that is incurred by programming with large voltages across ultrathin oxides, and the cell-to-cell parasitic interference between the stored charges in closely packed cells.[3,8–10] Many alternative cell designs were proposed, typically involving replacement of the floating polysilicon gate by some type of charge-trapping layer, such as the silicon nitride at the center of the silicon-oxide-nitride-oxide-semiconductor (SONOS) cell structure.[11] While early SONOS memory devices used extremely thin tunnel and blocking oxides for acceptable write/erase performance, and thus suffered from data retention issues,[12] recent work seems to have migrated to tantalum nitride-alumina-nitride-oxide-semiconductor (TANOS) structures.[13–16] These structures offer improved immunity to both SILC and parasitic interference between cells,[16] while also allowing any defects to gracefully degrade signal-to-noise ratio rather than serve as avenues for catastrophic charge leakage.[9,16] TANOS data retention has improved to acceptable levels,[9] and the reduced programming efficiency is now understood.[16]

However, TANOS structures cannot help to scale NOR Flash, because the charge injected at one edge of such devices by channel hot-electron injection[17] must be redistributed throughout the floating gate after programming.[10] For NAND Flash, the finite and fairly modest number of discrete traps in each TANOS cell have accelerated the onset of new problems, ranging from device-to-device variations in V_t,[9] stochastic or "shot-noise" effects,[9] random telegraph noise,[18,19] and a significant reduction in the number of stored electrons that differentiate one stored analog level from the next.[20] These issues are particularly problematic for multi-level cell (MLC) Flash, where multiple analog levels allow an increase in the effective number of bits per physical device by a factor of 2, 3, or even 4. Worse yet, such few-electron problems will only increase with further dimen-

sional scaling, leading Flash researchers to explore even more complicated schemes for FinFET Flash devices[21,22] or three-dimensional (3D) stacking of Flash memory.[23–26]

With these difficulties in scaling to future technology nodes, Flash researchers are already hard pressed to maintain specifications, such as write endurance, retention of heavily cycled cells, and write/erase performance, let alone improve them. As one indication of these pressures, some authors have pointed out that in cases such as digital photography, larger capacity formats can be expected to be tolerant of even more relaxed endurance specifications.[20] However, at the same time that Flash is struggling to maintain current levels of reliability and performance while increasing density, new applications are opening up for which these specifications are just barely adequate.

The solid-state drive (SSD) market—long dominated by high-cost, battery-backed DRAM for military and other critical applications—has grown rapidly since the introduction of Flash-based SSD drives, passing $400 million in revenues in 2007.[27] One reason for the time delay between the widespread use of Flash in consumer applications and its appearance in SSD applications was the need to build system controllers that could hide the weaknesses of Flash. Consider that each underlying block of Flash devices takes over a millisecond to erase, and if written to continuously, would start to exhibit significant device failures in mere seconds. Sophisticated algorithms have been developed to avoid unnecessary writes, to perform static or dynamic wear leveling, to pipeline writes, and to maintain pre-erased blocks in order to finesse or hide the poor write/erase performance.[28,29] Together with simple overprovisioning of extra capacity, these techniques allow impressive system performance. For instance, the Texas Memory Systems RamSan-500 can write at 2 Gbytes/s with an effective Flash endurance of >15 years.[30] However, it is interesting to note that despite the fact that MLC Flash costs much less than 1 bit/cell single-layer cell (SLC) Flash, for a long time only SLC Flash was used in SSD devices.[30] This is because MLC Flash tends to have ten times lower endurance and two times lower write speed than SLC Flash,[30] illustrating the importance of these specifications within SSD applications.

Thus there is a need for a new next-generation NVM that might have an easier scaling path than NAND Flash to reach the higher densities offered by future technology nodes. Simultaneously, there is a need for a memory that could offer better write endurance and input-output (I/O) performance than Flash, in order to bring down the cost while increasing the performance of NVM-based SSD drives. However the size of the opportunity here is even larger: The emergence of a nonvolatile solid-state memory technology that could combine high performance, high density, and low cost could usher in seminal changes in the memory/storage hierarchy throughout all computing platforms, ranging all the way up to high-performance computing. If the cost per bit could be driven low enough through ultrahigh memory density, ultimately such a storage-class memory (SCM) device could po-

tentially displace magnetic hard-disk drives (HDD) in enterprise storage server systems.

Fortunately, new NVM candidate technologies have been under consideration as possible Flash "replacements" for more than a decade.[31] These candidates range from technologies that have reached the marketplace after successful integration in real complementary metal oxide semiconductor (CMOS) fabs [ferroelectric (FeRAM) and magnetic (MRAM) random access memory (RAM)], to novel ideas that are barely past the proof-of-principle stage (racetrack memory and organic RAM), to technologies that are somewhere in between [phase change memory (PCM), resistance RAM, and solid-electrolyte memory].[31] Each of these has its strengths and weaknesses. In general, the farther along a technology has progressed toward real integration, the more that is known about it. Moreover, as research gives way to development, it is typically new weaknesses—previously hidden yet all too quickly considered to be obvious in hindsight—that tend to be revealed. In contrast, by avoiding these known pitfalls, fresh new technologies are immediately attractive, at least until their own unique weaknesses are discovered.

In this article, we survey the current state of PCM. This technology has made rapid progress in a short time, having passed older technologies such as FeRAM and MRAM in terms of sophisticated demonstrations of scaling to small device dimensions. In addition, integrated large-array demonstrators with impressive retention, endurance, performance, and yield characteristics[31] have been built.

The article is organized into seven sections, beginning with the current section titled "Motivation for PCM." Section I also includes a brief overview of PCM technology and an assessment of how its characteristics match up with various potential applications across the memory-storage hierarchy. Section II goes into the physics behind PCM in more depth, in terms of the underlying phase change materials and their inherent scalability, and the physical processes affecting the switching speed of PCM devices. This section concludes with a survey of PCM modeling efforts published to date, and a discussion of scalability as revealed by ultrasmall prototype PCM devices.

In Sec. III, we address factors that affect the design and fabrication of PCM devices, including cell design, variability, changes in the phase change material induced by the fabrication procedure, and the design of surrounding access circuitry. We then turn to issues related to operation of PCM devices in Sec. IV, including endurance, retention, and device-to-device cross-talk. Section V addresses several factors that can be expected to enhance PCM in the future, including multilevel cell technology for PCM, the role of coding, and possible routes to an ultrahigh-density PCM technology. The conclusion section (Sec. VI) is followed by a brief acknowledgments section.

B. What is PCM?

PCM exploits the large resistance contrast between the amorphous and crystalline states in so-called phase change
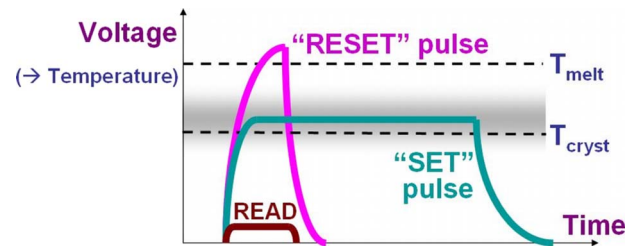

FIG. 1. (Color online) Programming of a PCM device involves application of electrical power through applied voltage, leading to internal temperature changes that either melt and then rapidly quench a volume of amorphous material (reset), or which hold this volume at a slightly lower temperature for sufficient time for recrystallization (set). A low voltage is used to sense the device resistance (read) so that the device state is not perturbed.

materials.[32] The amorphous phase tends to have high electrical resistivity, while the crystalline phase exhibits a low resistivity, sometimes three or four orders of magnitude lower. Due to this large resistance contrast, the change in read current is quite large, opening up the opportunity for the multiple analog levels needed for MLC operations.[32]

To set the cell into its low-resistance state, an electrical pulse is applied to heat a significant portion of the cell above the crystallization temperature of the phase change material. This set operation tends to dictate the write speed performance of PCM technology, since the required duration of this pulse depends on the crystallization speed of the phase change material (Sec. II B). Set pulses shorter than 10 ns have been demonstrated.[33–36] Because the crystallization process is many orders of magnitude slower at low temperatures (<120 °C), PCM is a NVM technology that can offer years of data lifetime.

In the reset operation, a larger electrical current is applied in order to melt the central portion of the cell. If this pulse is cutoff abruptly enough, the molten material quenches into the amorphous phase, producing a cell in the high-resistance state. The reset operation tends to be fairly current and power hungry, and thus care must be taken to choose an access device capable of delivering high current and power without requiring a significantly larger footprint than the PCM element itself. The read operation is performed by measuring the device resistance at low voltage so that the device state is not perturbed. These operations are summarized in Fig. 1.

Even though the principle of applying phase change materials to electronic memory was demonstrated as long ago as the 1960s,[37] interest in PCM was slow to develop compared to other NVM candidates. However, renewed interest in PCM technology was triggered by the discovery of fast (<100 ns) crystallizing materials such as $Ge_2Sb_2Te_5$ (GST) or Ag- and In-doped Sb_2Te (AIST) (Refs. 38 and 39) by optical storage researchers. Over the past few years, a large number of sophisticated integration efforts have been undertaken in PCM technology, leading to demonstration of high endurance,[40] fast speed,[41] inherent scaling of the phase change process out beyond the 22 nm node,[42] and integration at technology nodes down to 90 nm.[43] One important remaining unknown for the success of PCM technology is

whether the memory access device (diode,[43] transistor,[44] etc.) in a dense memory array will be able to supply sufficient current to reset the PCM cell. Already, in order to try to minimize the reset current, it is assumed that the dimension of the phase change material will be only 30% of the lithographic feature size F,[5] mandating the use of sublithographic techniques for accurate definition of this critical dimension (CD). However, even with this difficult integration task, the success of PCM technology may end up depending on advances in the access device as much as on the PCM cell itself.[5]

Important device characteristics for a PCM cell include widely separated set and reset resistance distributions (necessary for sufficient noise margin upon fast readout), the ability to switch between these two states with accessible electrical pulses, the ability to read/sense the resistance states without perturbing them, high endurance (allowing many switching cycles between set and reset), long data retention (usually specified as 10 year data lifetime at some elevated temperature), and fast set speed (the time required to recrystallize the cell from the reset state). Data retention usually comes down to the cell's ability to retain the amorphous reset state by avoiding unintended recrystallization. An additional aspect that can be of significant importance is the ability to store (and retain over time) more than 1 bit of data per cell since this allows one to increase effective density much like MLC Flash without decreasing the feature size.

A critical property of phase change materials is the so-called threshold switching.[45–48] Without this effect PCM would simply not be a feasible technology because in the high resistance-state, extremely high voltages would be required to deliver enough power to the cell to heat it above the crystallization temperature. However, when a voltage above a particular threshold V_t is applied to a phase change material in the amorphous phase, the resulting large electrical fields greatly increase the electrical conductivity. This effect is still not completely understood but is attributed to a complex interplay between trapped charge, device current, and local electrical fields.[45,49] With the previously resistive material now suddenly highly conducting, a large current flows—which can then heat the material. However, if this current pulse is switched off immediately after the threshold switching, the material returns to the highly resistive amorphous phase after about 30 ns,[50] with both the original threshold voltage V_t and reset resistance recovering slowly over time.[50,51] Only when a current sufficient to heat the material above the crystallization temperature, but below the melting point, is sustained for a long enough time does the cell switch to the crystalline state. The threshold switching effect serves to make this possible with applied voltages of a few volts, despite the high initial resistance of the device in the reset state.

C. Potential applications of PCM

The ultimate goal of researchers and developers studying emerging memory technologies is to devise a universal memory that could work across multiple layers of the existing memory hierarchy for modern computers. This memory hierarchy, shown in Fig. 2, is designed to bridge the performance gap between the fast central processing units and the slower (sometimes much slower) memory and storage technologies, while keeping overall system costs down. Figure 3 shows how PCM is expected to compare to the four major incumbent memory and storage technologies in terms of cost and performance. The enormous range of cost and performance spanned by these technologies makes a single universal memory—one capable of replacing all of these well-established memory and storage techniques—an aggressive goal indeed.

However, Fig. 4 shows that there is currently a gap of more than three orders of magnitude between the access time of off-chip dynamic random access memory (DRAM) (60
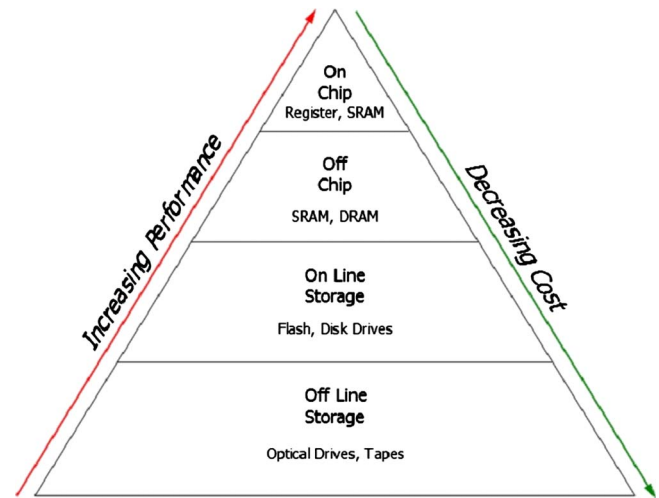

FIG. 2. (Color online) Memory hierarchy in computers spans orders of magnitude in read-write performance, ranging from small amounts of expensive yet high-performance memory sitting near the CPU to vast amounts of low cost yet very slow off-line storage.
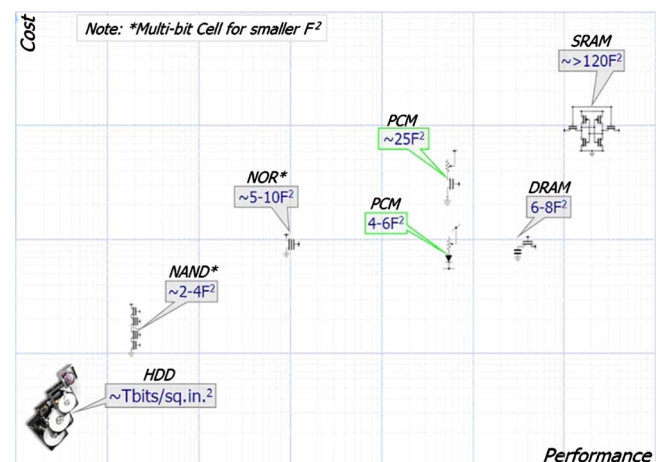

FIG. 3. (Color online) Qualitative representation of the cost and performance of various memories and storage technologies, ranging from extremely dense yet slow HDDs to ultrafast but expensive SRAM. F is the size of the smallest lithographic feature. A smaller device footprint leads to higher density and thus lower cost.
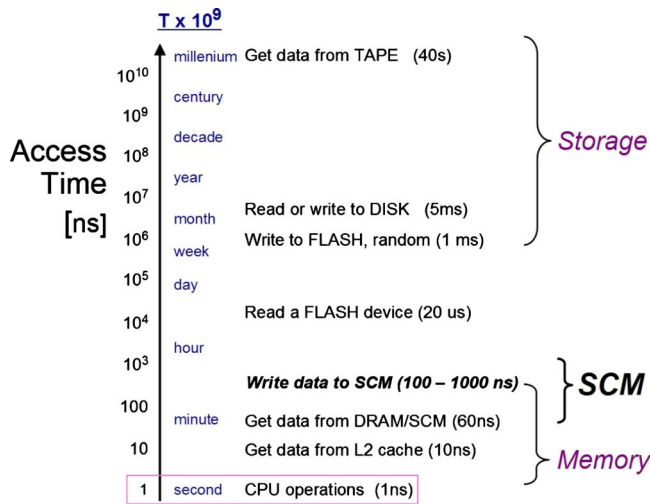
FIG. 4. (Color online) Access times for various storage and memory technologies, both in nanoseconds and in terms of human perspective. For the latter, all times are scaled by 10^9 so that the fundamental unit of a single CPU operation is analogous to a human making a 1 s decision. In this context, writing data to Flash memory can require more than "1 week" and obtaining data from an offline tape cartridge takes "1000 years" (Refs. 29 and 55).

ns) and the write-cycle time of Flash (1 ms). To set this into human perspective, this slow write-cycle time is equivalent to a person, who might be making data-based decisions analogous to a single central processor unit (CPU) operation every second, having to wait approximately 10 days to record a small block of information. An interesting region on this chart sits just above off-chip DRAM, where access times of 100–1000 ns could potentially be enabled by a "Storage Class Memory (SCM)" made possible by PCM.

In the remainder of this section, we examine the suitability of PCM for the layers of the memory hierarchy currently served by static random access memory (SRAM), DRAM (dynamic), and NOR and NAND Flash. We also discuss the emerging area of storage-class memory, for which Flash-based solid-state drives are just now becoming available. While the two principal integration metric are cost and performance, we also briefly examine critical reliability issues such as data retention and read/write endurance here (leaving more in-depth discussion to Sec. IV). We do not consider the relative merit of power consumption, assuming instead that all these technologies are roughly comparable within an order of magnitude. The nonvolatility of PCM does compare favorably to volatile memories, both in terms of standby power as well as by enabling easier recovery from system or power failures in critical applications.

1. PCM as SRAM

Much of the SRAM used in computers today is embedded close to the CPU, serving as high-performance level 1 (L1) and level 2 (L2) cache memories. Some off-chip level 3 (L3) cache memories also use SRAM. In consumer electronics, SRAM has been used in combination with NOR Flash in cell phones. A typical SRAM cell comprises of six CMOS tran-

sistors, two positive metal oxide semiconductor field effect transistors and four negative metal oxide semiconductor field effect transistors, and thus occupies more than $120F^2$ in chip real estate per bit. (Here F is the size of the smallest lithographic feature, so that this measure of device size is independent of the particular device technology used to fabricate the memory.) Embedded SRAM typically runs at the CPU clock speed, thus access times for these devices must be less than 10 ns. Commodity SRAM used in cell phones runs at slower clock speeds, allowing access times in the tens of nanoseconds.

While there is no problem for PCM to improve upon the large SRAM cell size, even if a large access device is used for the PCM cell, SRAM performance is hard to match. The performance limiter for PCM is the set speed, which in turn depends on the crystallization speed of the phase change material. As will be described in detail in Sec. II B, while some researchers demonstrated the use of set pulses shorter than 10 ns,[33–36] most of the realistically large array demonstrations tend to use set pulses that range from roughly 50 to 500 ns in length.[52]

In any case, the most stringent requirement for any emerging memory technology that seeks to replace SRAM is endurance. For all practical purposes, the read/write endurance of SRAM is infinite. While read endurance is not a likely problem for PCM, the required write endurance for SRAM replacement is probably 10^{18}—out of reach for nearly all NVM technologies. Storing data semipermanently with PCM and most other NVM technologies involves some form of "brute force" that alters an easily observable material characteristic of the memory device. For PCM, this brute force is the melt-quench reset operation, and at such elevated temperatures, it has been shown that the constituent atoms of a phase change material will tend to migrate over time,[40,53] as discussed in Secs. IV C and IV D.

Since nonvolatility is not a requirement for SRAM applications, one might be able to trade some data retention for improved endurance. Some remote evidence of this trade-off has been demonstrated by showing a strong correlation between the total energy in the reset pulse and the resulting PCM endurance.[40] The best case endurance, achieved for the lowest-energy reset pulses, was 10^{12} set-reset cycles.[40] Yet this is still six orders of magnitude away from the target specification for SRAM.

2. PCM as DRAM

DRAM is used in a more diversified set of applications than SRAM. Most of the characteristics discussed in Sec. I C 1 for the replacement of SRAM also apply to the replacement of DRAM, although in most cases the specifications are slightly relaxed. Access times of tens of nanoseconds would be acceptable for most computer and consumer electronics applications of DRAM. For embedded DRAM used as video RAM and L3 cache,[54] however, an access time of 10 ns or less is required. As for write endurance, the requirement can be estimated using the following equation:[55]

$$E = T_{\text{life}} \frac{B}{\alpha C}, \tag{1}$$

where E is endurance, T_{life} is the life expectancy of the system, B is the memory bandwidth, α is the wear-leveling efficiency, and C is the system memory capacity. Assuming a typical server with a 10 year life expectancy, 1 Gbyte/s bandwidth, 10% wear-leveling efficiency, and 16 Gbyte capacity, the endurance requirement is approximately 2×10^8—well within the reach of PCM.[44,56]

There is also a power argument to be made when discussing PCM as a potential DRAM replacement. This might seem to be a difficult case to make for a technology for which every write cycle involves heating to temperatures ranging from 400 to 700 °C. However, DRAM turns out to be a fairly power-hungry technology. This is not due to its periodic refresh, however, which takes place only infrequently, and is not too strongly related to the underlying physical storage mechanism of charging up a local capacitor. Instead, power inefficiency in DRAM is due to the simultaneous addressing of multiple banks within the chip. For every bit that passes into or out of a DRAM chip, 8 or even 16 devices are being internally accessed (read and then rewritten), somewhat as if your librarian knocked an entire row of books onto the floor each time you asked for a book. Low-power DRAM intended for mobile, battery-powered applications tends to have lower performance, although some developments have been made that can combine high performance with low power.[57] However, the inherent need to rewrite after each read access is unavoidable for a volatile memory such as DRAM. Thus simply by being nonvolatile, PCM could potentially offer a lower-power alternative to DRAM, despite the inherently power-hungry nature of PCM write operations.

For standalone memories, cost is directly proportional to memory cell size. State of the art DRAM cells occupy $6F^2$ in chip area. Thus for PCM to compete in the DRAM arena, PCM cell size would need to be this size or smaller with comparable "array efficiency" (the fraction of the chip area dedicated to memory devices rather than to peripheral circuitry). Fortunately, such small cell sizes have already been demonstrated using a diode select device.[56] PCM also competes favorably with DRAM in terms of forward scaling into future generations, as DRAM developers are quickly hitting various scaling limits associated with storage interference, device leakage, and challenges in integrating high aspect-ratio capacitors in tight spaces. Currently, DRAM has fallen behind NAND Flash and standard CMOS logic technologies in terms of scaling to the 45 nm technology node and preparation for the 32 nm node. However, DRAM is a proven, reliable technology that has been employed in modern computers since the early 1970s. It would be a long journey to displace such a stable technology.

3. PCM as Flash

There are two kinds of Flash memories, NOR and NAND. In (common-source) NOR memory architectures, each cell in a two-dimensional array is directly connected to its word- and bit-line input lines (with the source electrode of each cell sharing a common ground), whereas in NAND memory architectures, small blocks of cells are connected in series between a high input signal and ground. Thus, while NAND flash can inherently be packed more densely (due to its smaller unit cell size) than NOR flash, NOR flash offers significantly faster random access (since each cell in the array is directly connected to the input lines). However, since NOR memory requires large programming currents (to place charge on the floating gate via channel hot electrons), its programming throughput (measured in MB/s) is much slower than that of the block-based NAND memory architectures (which, by utilizing the Fowler–Nordheim tunneling, can utilize lower programming currents that permit many bits to be processed in parallel).[58] As a consequence, NOR memory offers significantly faster random access with low programming throughput, and thus is mainly used for applications such as embedded logic that require fast access to data that is modified only occasionally. In contrast, NAND memory is a high-density, block-based architecture with slower random access, which is mainly used for mass storage applications.

NOR Flash memory cells occupy about $10F^2$, with an access time upon read of a few tens of nanoseconds or more. However, the access time upon write for NOR Flash is typically around 10 μs, and the write/erase endurance (for both NOR as well as NAND) is only 100 000 cycles. These characteristics are well within the capabilities of current PCMs. NOR Flash with its floating gate technology has difficulties scaling below 45 nm, mainly due to difficulties in scaling the thickness of the tunnel oxide. It is thus no surprise that NOR Flash is the popular target for first replacement by most PCM developers.

NAND Flash, on the other hand, is a much harder target despite PCM's superiority in both endurance and read performance. Cost is the biggest challenge. A NAND Flash memory cell occupies only $4F^2$ of chip area, and as discussed earlier, NAND will be able to maintain this through at least 22 nm using trap storage technology[13] and possibly three-dimensional integration.[26] Furthermore, MLC NAND has been shipping 2 bits per physical memory cell for years, and is promising to increase this to 4 bits per cell.[59]

NAND Flash is mainly used in consumer electronic devices, where cost is the paramount concern, and in the emerging SSD market to replace magnetic HDDs, where both cost and reliability are important. The prerequisites for PCM to replace NAND Flash are $4F^2$ memory cell size, at least 2 bit MLC capability, and three-dimensional integration to further increase the effective number of bits per unit area of underlying silicon. A $4F^2$ cell dictates a memory element that can be vertically stacked over the select device, as shown in Fig. 5. Multilevel storage seems to be within reach of PCM given its inherently wide resistance range, and both 2 and 4 bits per cell have already been demonstrated in small-scale demonstrations.[60,61] Even though write operations are slow for NAND Flash, it tends to achieve an impres-
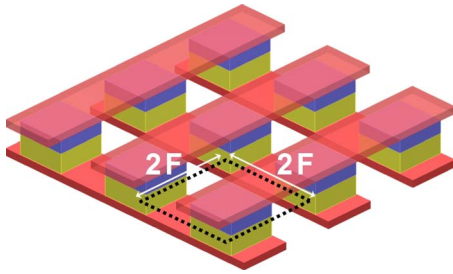
FIG. 5. (Color online) Semiconductor device technology node is commonly described by the minimum feature size F that is available via lithographic patterning. Thus the smallest device area that can be envisioned which is still accessible by lithographically defined wiring is $4F^2$. To increase *effective bit density* beyond this, either sublithographic wiring, multiple bits per device (analogous to MLC Flash technology), or multiple layers of stacked memory arrays are required, as described in Sec. V C.

sive write data rate because its low write power allows for programming of many bits in parallel. Thus to deliver equal or better write bandwidth, PCM developers will need to work on reducing the write power so that the data bus can be as wide as possible.

4. PCM as storage-class memory

In addition to the established segments of the memory hierarchy we have described (SRAM, DRAM, and Flash), the gap in access times between 1 ms and 100 ns shown in Fig. 4 opens up the possibility of SCM.[28,29] SCM would blur the traditional boundaries between storage and memory by combining the benefits of a solid-state memory, such as high performance and robustness, with the archival capabilities and low cost of conventional hard-disk magnetic storage. Such a technology would require a solid-state nonvolatile memory that could be manufactured at an extremely high effective areal density, using some combination of sublithographic patterning techniques, multiple bits per cell, and multiple layers of devices (Sec. V C). The target density probably needs to exceed current MLC NAND Flash densities by a factor of 2–8 times in order to bring the cost of SCM down close to the cost of reliable enterprise HDD.

The opportunity for SCM itself actually breaks into two segments. The slower variant, referred to as S-class SCM,[29] would act much like a Flash-based SSD except with better native endurance and write performance. Here access times of $1-3$ μs would be acceptable, but low cost via high density would be of paramount importance. The other variant, referred to as M-class SCM,[29] requires access times of 300 ns or less, with both cost and power as considerations. This threshold of 300 ns is considered to be the point at which an M-class SCM would be fast enough to be synchronous with memory operations, so that it could be connected to the usual memory controller.[29] In contrast, S-class SCM, SSD, and HDD would all be accessed through an I/O controller for asynchronous access. M-class SCM would likely not be as fast as main memory DRAM. However, by being nonvolatile, lower in power per unit capacity (via high density), and lower in cost per capacity, the presence of M-class SCM

could potentially allow the total amount of DRAM required to maintain ultrahigh bandwidth to be greatly reduced, thus reducing overall system cost and power.

II. PHYSICS OF PCM

A. Phase change materials and scalability

As discussed in Sec. I, the NVM industry faces the prospect of a costly and risky switch from a known and established technology (Flash) into something much less well known (either PCM or something else). Understandably, the industry wants to make such leaps rare.

The problem here is not that one might fail to create a successful first product. That would be unpleasant but not devastating, because this would happen during the early development stage, where the level of investment is small and multiple alternative approaches are still being pursued. Instead, the nightmare scaling scenario is one in which the new technology works perfectly well for the first generation, yet is doomed to failure immediately afterward. If only one or two device generations succeed, then the NVM industry, having just invested heavily into this new technology, will be forced to make yet another switch and start the learning process all over again.

Thus scaling studies are designed to look far down the device roadmap, to try to uncover the showstoppers that might bedevil a potential NVM technology at sizes much smaller than what can be built today. In the case of PCM technology, two aspects of scalability need to be considered: the scaling properties of the phase change materials and the scaling properties of PCM devices. In this section, we survey recent literature covering both of these considerations. In general, experiments have shown that PCM is a very promising technology with respect to scalability.

It is well known that the properties of nanoscale materials can deviate from those of the bulk material, and can furthermore be a strong function of size. For example, it is typical for nanoparticles to have a lower melting temperature than bulk material of the same chemical composition, because the ratio of surface atoms to volume atoms is greatly increased. A recurring theme in such studies is the larger role that surfaces and interfaces play as dimensions are reduced.

Phase change material parameters that are significant for PCM applications—and the device performance properties that are influenced by these parameters—are summarized in Table I. For optical applications the change in optical constants as a function of film thickness is also important, but for this article we restrict our considerations to material parameters relevant to electronic memory applications. As can be seen from Table I, there is a large set of material parameters that influence the PCM device, either affecting one of the two writing operations (set to low resistance; reset to high resistance) or the read operation.

A particularly important phase change material parameter is the crystallization temperature, T_x. This is not necessarily the temperature at which crystallization is most likely, but instead is the lowest temperature at which the crystallization

TABLE I. Some phase change material parameters and the device performance characteristics they influence.

Phase change material parameter	Influence on PCM device performance
Crystallization temperature and thermal stability of the amorphous phase	Data retention and archival lifetime
	Set power
Melting temperature	Reset power
Resistivity in amorphous and crystalline phases	On/off ratio
	Set and reset current
Threshold voltage	Set voltage and reading voltage
Thermal conductivity in both phases	Set and reset power
Crystallization speed	Set pulse duration (and thus power)
	Data rate
Melt-quenching speed	Reset pulse duration (and thus power)

process becomes "fast." It is typically measured by raising the temperature slowly while monitoring the crystallinity (either looking for x-ray diffraction from the crystalline lattice or the associated large drop in resistivity). Thus the crystallization temperature is a good measure of how hot a PCM cell in the reset state could be made before the data stored by an amorphous plug would be lost rapidly due to unwanted crystallization. While the crystallization temperature by itself does not reveal how "slowly" such data would be lost for slightly lower or much lower temperatures, it sets a definitive and easily measured upper bound on the retention versus temperature curve for a new phase change material.

The crystallization temperature of phase change materials tends to vary considerably as a function of material composition.[62–64] For example, some materials, such as pure Sb, crystallize below room temperature. Yet adding only a few at. % of Ge to Sb, creating the phase change material Ge_x-Sb_{1-x}, increases the crystallization temperature significantly above room temperature. In fact, T_x can reach almost 500 °C for GeSb alloys that are high in Ge content.[63,64] Studies of the crystallization temperature as a function of film thickness show an exponential increase as film thickness is reduced (for phase change materials sandwiched between insulating materials such as SiO_2 or $ZnS-SiO_2$).[65,66] However, for phase change materials sandwiched between metals, metal-induced crystallization can occur and the crystallization temperature can be reduced for thinner films.[67] It is known that for phase change materials the crystallization is typically heterogeneous, starting at defects that can be located in the bulk, but which tend to be more prevalent at surfaces and interfaces. As film thickness is reduced, the volume fraction of phase change material that is at or near an interface increases, leading to changes in the externally observable crystallization temperature.

Phase change nanowires are typically fabricated by the vapor-liquid-solid technique, and are crystalline as synthesized.[68] To measure crystallization behavior as a function of wire size, PCM devices were fabricated from single-crystalline, as-grown $Ge_2Sb_2Te_5$ nanowires using Pt contact pads.[68] The central section of the nanowire devices was reamorphized by electrical current pulses and the activation energy was determined by measuring the recrystallization temperature as a function of heating rate. Here, the activation

energy was found to fall from 2.34 eV for 190 nm diameter devices to 1.9 eV for 20 nm diameter devices, indicating a deterioration of data retention as the $Ge_2Sb_2Te_5$ nanowire diameter is reduced. However, Yu *et al.*[69] did not observe a dependence of the crystallization temperature on the device diameter for PCM devices fabricated by contacting GeTe and Sb_2Te_3 nanowires using Cr/Au contacts.

Figure 6 shows phase change nanoparticles fabricated by a variety of techniques including electron-beam lithography, solution-based chemistry, self-assembly-based lithography combined with sputter deposition, and self-assembly-based lithography combined with spin-on deposition of the phase change material. When the crystallization temperature of amorphous-as-fabricated nanoparticles was studied, it was found that larger phase change nanoparticles have a very similar crystallization temperature compared to bulk
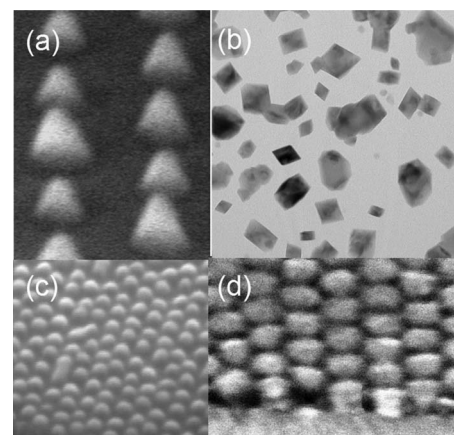

FIG. 6. (a) Phase change nanoparticles of Ge–Sb with 15 at. % Ge, fabricated by electron-beam lithography, diameter of about 40 nm. Reprinted with permission from S. Raoux *et al.*, J. Appl. Phys., 102, 94305, 2007. © 2007, American Institute of Physics. (b) GeTe nanoparticles synthesized by solution-based chemistry, diameter of about 30 nm (Ref. 73). (c) Nanoparticles of Ge–Sb with 15 at. % Ge, fabricated by self-assembly based lithography and sputter deposition, diameter of about 15 nm. Reprinted with permission from Y. Zhang *et al.*, Applied. Physics. Letters, 91, 13104, 2007. © 2007, American Institute of Physics. (d) Nanoparticles of Ge–Sb–Se, fabricated by self-assembly based lithography and spin-on deposition, diameter of about 30 nm. Reprinted with permission from D. J. Milliron *et al.*, Nature Mater., 6, 352, 2007. © 2007, Macmillan Publishers Ltd.

material,[62,70] whereas the smallest nanoparticles in the 10 nm range can show either decreased[71] or increased[72] crystallization temperature.

In terms of size effects, ultrathin films still can show crystallization down to thicknesses of only 1.3 nm,[66] and nanoparticles as small as 2–5 nm synthesized by solution-based chemistry have been found to be crystalline.[73] This is very promising for the scalability of PCM technology to future device generations.

Beyond crystallization temperature, the melting temperature is a parameter that can vary with composition and, at small dimensions, with size. In fact, a reduction in the melting temperature of phase change materials has been observed for very thin films,[74] nanowires,[75] and nanoparticles.[76] This is advantageous for device performance because a lower melting point implies a reduction in the power (and current) required to reset such a PCM cell. The electrical resistivity for thin films increases slightly for both phases when film thickness is reduced.[65] This is also beneficial for scaling because higher resistivities lead to higher voltage drop across the material and can thus reduce switching currents.

The threshold voltage is a phenomenological parameter of PCM devices that describes the applied voltage (typically around 1 V) required to induce an electrical breakdown effect. Such a sudden increase in electrical conductivity allows the PCM device to rapidly and efficiently attain a significantly lower dynamic resistance (typically three to ten times lower than the room temperature set resistance), allowing efficient heating with moderate applied voltages. Thus the presence of this electrical switching effect is an important component of PCM technology.

However, a more accurate description of the underlying physical process calls for a threshold electric field, rather than a threshold voltage, that must be surpassed for the amorphous material to become highly conductive. Studies of phase change bridge devices (described in Sec. II D) have shown that the threshold voltage scales linearly as a function of the length of the bridge along the applied voltage direction, confirming the role of an underlying material-dependent threshold field.[77,78]

No deviation from this linear behavior was observed for bridge devices as short as 20 nm. The value of the threshold field varied considerably, from 8 V/μm for Ge (15 at. %)–Sb devices to 94 V/μm for thin Sb devices. For nanowire devices with even smaller amorphous areas, however, Yu *et al.*[69] observed a deviation from this linear behavior. Once the amorphous volume spanned less than approximately 10 nm along the nanowire, the threshold voltage saturated at 0.8 and 0.6 V for GeTe and Sb_2Te_3 devices, respectively. This scaling behavior was explained with the impact ionization model previously developed to explain the threshold switching phenomenon.[79]

Such a saturation in the effective threshold voltage is actually desirable because for practical device performance a threshold voltage around 1 V is optimum. This places the switching point well above the typical reading voltage of about 50–100 mV, yet not so far that a transistor or diode access device would fail to easily deliver the required switching pulses with moderate supply voltages. (Note that exceeding the threshold voltage to produce breakdown is not the same as delivering sufficient power to heat the cell to achieve the reset condition). If the threshold voltage were to continue as a linear function of device size for sub-10-nm devices, then reading the cells without accidentally switching them out of the reset state could become problematic.

The thermal conductivity of phase change materials is important because it strongly influences the thermal response of a PCM device to an electrical current pulse. However, so far the materials that have been studied ($Ge_2Sb_2Te_5$, nitrogen-doped $Ge_2Sb_2Te_5$, Sb_2Te, and Ag- and In-doped Sb_2Te) show only a slight variation in the values for the thermal conductivities between 0.14 and 0.17 W/m K for the as-deposited amorphous phase, and values between 0.25 and 2.47 W/m K for the crystalline phase.[80] Reifenberg *et al.*[81] studied the thermal conductivity of $Ge_2Sb_2Te_5$ with thicknesses between 60 and 350 nm using nanosecond laser heating and thermal reflectance measurements. They found about a factor of 2 decrease in the thermal conductivity as film thickness is reduced—from 0.29, 0.42, and 1.76 W/m K in the amorphous, fcc, and hexagonal phases, respectively, for 350 nm thick films, to 0.17, 0.28, and 0.83 W/m K for 60 nm thick films. As with earlier results, such a trend leads to advantageous scaling behavior for PCM applications, by helping reduce the energy required for the power-intensive reset operation.

In addition to these changes in effective material properties as device sizes scale down, there are also simple yet powerful geometric effects that are associated with scaling. As we will discuss extensively in Sec. III A, scaling decreases the size of the limiting cross-sectional aperture within each PCM cell, thus driving down the reset current. However, at constant material resistivity, geometric considerations cause both the set and dynamic resistances to increase. As a result, the effective applied voltage across the device during the reset operation remains unchanged by scaling, at least to first order. These effects can be expected to eventually have adverse effects, as the decreasing read current (from the higher set resistance) makes it difficult to accurately read the cell state rapidly, and as the nonscaling voltages exceed the breakdown limits of nearby scaled-down access transistors.

To summarize scaling properties of phase change materials, it has been observed that the crystallization temperature is in most cases increased as dimensions are reduced (beneficial to retention), and melting temperatures are reduced as dimensions are reduced (beneficial to reset power scaling). Similarly, resistivities in both phases tend to increase (beneficial for reset power), threshold voltages are first reduced as dimensions are reduced but then level out around 0.6–0.8 V for dimensions smaller than 10 nm (beneficial for voltage scaling), and thermal conductivity seems to decrease as film thickness is reduced (beneficial for reset current scaling). As will be seen in more detail in the next section, the raw crystallization speed can either decrease (detrimental to write

performance) or increase (beneficial), and seems to depend strongly on the materials and their environment. Crystallization has been observed to reliably occur for films as thin as 1.3 nm, and crystalline nanoparticles as small as 2–3 nm in diameter have been synthesized. Overall, phase change materials show very favorable scaling behavior—from the materials perspective, this technology can be expected to be viable for several future technology nodes.

B. Speed of PCM

Of all the material parameters mentioned so far, crystallization speed is probably the most critically important for PCM because it sets an upper bound on the potential data rate. Moreover as discussed in Sec. I C, data rate and endurance are the two device characteristics that dictate what possible application spaces could potentially be considered for PCM. (Of course, cost and reliability are then critically important to succeed in that space, but without the required speed and endurance for that market segment, such considerations would be moot anyway.)

As mentioned in Sec. I B, the early discovery of electronic-induced phase change behavior by Ovshinsky[37] did not immediately develop into the current PCM field. Early phase change materials simply crystallized too slowly to be technologically competitive, with switching times in the microsecond to millisecond time regimes.[82,83] Phase change technology began to gain traction in the late 1980s with new phase change materials capable of recrystallization in the nanosecond time regime.[84,85] These discoveries both led to the widespread use of phase change materials in optical rewritable technology (DVDs, CDs, and now Blu-Ray), and fostered renewed interest in PCM.

In phase change devices, there are three steps that could determine the overall operating speed: read, reset (to high resistance), and set (to low resistance). The read operation depends on the speed with which two (or more) resistance states can be reliably distinguished, and thus is dominated by the circuit considerations (capacitance of the bit line being charged up, leakage from unselected devices). Although the resistance contrast and absolute resistance of the PCM cell do play a role, the read operation can generally be performed in 1–10 ns.[52] The set and reset steps, however, involve the physical brute force transformations between distinct structural states.

The energetically less-favorable amorphous phase—which gives a PCM cell in the reset state its high resistance—is attained by melting and then rapidly cooling the material. As the temperature falls below the glass transition temperature and molecular motion of the undercooled liquid is halted, a "kinetically trapped" phase results. This process can be separated into three steps: (1) current-induced heating above the melting temperature, (2) kinetics of melting,[86] and (3) kinetics of solidifying the molten material.[87] Steps (1) and (2) involve the rapid injection of energy to first heat and then melt the material; step (3) involves the rapid cooling to temperatures below those favor-

able to recrystallization (see below). Thus the most practical method to engineer step (3) is through design of the PCM cell's thermal environment.[88]

These steps are each quite fast: electrical pulses as short as 400 ps have been used to switch $Ge_2Sb_2Te_5$ into the amorphous state.[33] While the kinetics of phase change devices are strongly material dependent, the generation of the amorphous phase in any practical phase change material is necessarily faster than the speed of crystallization—because otherwise, the amorphous phase would simply never be observed.[87] Note that after a reset operation, the amorphous phase can continue to evolve extremely slowly at low temperature, undergoing both continued relaxation of the amorphous phase as well as electronic redistribution of the trapped charge that participates in the electrical breakdown phenomenon.[49,51,89–91] This drift can be an issue for MLC in PCM devices, as discussed in Sec. V A.

While the crystalline form is thermodynamically favorable, its kinetics are much slower than the formation of the high-resistance state,[92] by typically at least one order of magnitude. Thus the step that dictates the achievable data rate for PCM technology is the crystallization process associated with the set operation.

Formation of the crystalline phase involves as many as four steps: (1) threshold switching,[46] (2) current-induced heating to elevated temperatures (but below the melting point), (3) crystal nucleation,[93] and (4) crystal growth.[94] The latter two steps are the slowest, and realistically combine to determine the speed of the device. Not all of the steps will be encountered, however. Step (1) is relevant only if the device started in the high-resistance reset state so that a large portion of the applied voltage dropped across material in the amorphous phase.

If all of the contiguous PCM material being heated is in the amorphous phase, then step (3) must take place before step (4) can begin. An example is the first crystallization of materials (or devices) containing amorphous-as-deposited material, where no crystalline-amorphous interfaces are present. This nucleation step can be extremely slow in so-called growth-dominated materials, where nucleation is a highly unlikely event compared to the fast speed of crystal growth. In fact, frequently the crystallization of microns of surrounding material in such materials can be traced to the creation of a single nanoscopic critical nucleus.[95] In contrast, nucleation-dominated materials tend to have a lower barrier to nucleation so that a large region of crystalline material stems from the growth of numerous supercritical nuclei.[96]

In a typical PCM cell, only a portion of the phase change material is quenched into the amorphous state, meaning that for both types of materials, step (4) above mainly depends on the crystal growth speed at high temperature. The main difference between the two classes of phase change materials is that the recrystallization of amorphous nucleation-dominated material will occur both within the interior (nucleation) as well as from the edge (growth), while for a growth-dominated material only the propagation of the crystalline-amorphous boundaries matters. This can either be an advan-

tage or disadvantage, depending on whether this added nucleation is desirable (having seeded nuclei that can help speed up the set operation), or undesirable (by causing data to be lost faster at low to intermediate temperatures).

Examples of growth-dominated materials, where the rate of crystal nucleation < growth rate, include Ge-doped SbTe, GeSb, GeSnSb, and $Ge_3Sb_6Te_5$. In contrast, $Ge_2Sb_2Te_2$, $Ge_2Sb_2Te_5$, and $Ge_4Sb_1Te_5$ are considered nucleation-dominated materials, with a rate of crystal nucleation > growth rate. It should be noted that nucleation and growth kinetics have unique responses to temperature, so under certain conditions a material typically considered growth dominated may appear nucleation dominated, e.g., AIST.[97]

The crystallization time of phase change materials, even those intended for use in electrical devices, can be measured relatively easily using optical techniques. This is because most phase change materials of interest to PCM also have the same large optical contrast between the two phases that originally motivated the rewritable optical storage application. Somewhat like an optical storage device with a stationary disk, a static laser tester uses a low-power continuous-wave laser to constantly monitor the reflectivity while a high-power pulsed laser induces the desired phase changes. The pulsed laser heats the material above its crystallization temperature for set, or above the melting point for reset. The advantage of optical testing is that large-area, thin film samples of new phase change materials can be quickly prepared, without the need for any patterning or other steps required for fabrication of full PCM devices. Then a wide range of powers (e.g., temperatures) and times can be tested out rapidly.

As expected from the above discussion, it has been observed that the recrystallization time of a melt-quenched area in a crystalline matrix is typically orders of magnitude faster than the first-crystallization time of as-deposited amorphous films.[97] Because device operation hinges on the repeated cycling between the two phases, the relevant parameter to use for assessing the viability of a new material for PCM is the recrystallization time. However, since it is quite difficult to prepare isolated regions of melt-quenched material without introducing new interface effects, amorphous-as-deposited films remain the best way to study the physics of the nucleation process, to avoid the difficulty of deconvolving the entangled roles of nucleation and growth once a crystalline-amorphous boundary is present. For growth-dominated materials, it is the presence of these crystalline-amorphous boundaries which make the recrystallization speed so much faster than the initial nucleation from the amorphous-as-deposited state.

Figure 7 shows the change in reflectivity of a Ge–Sb thin film with 15 at. % Ge, as a function of laser power and duration measured by a static laser tester.[62] The film was first crystallized by annealing it in a furnace for 5 min at 300 °C, which is above this material's crystallization temperature.[62] A two pulse experiment was then performed. The first pulse, of fixed time and power (100 ns for 50 mW), created a small region of melt-quenched amorphous material at a previously
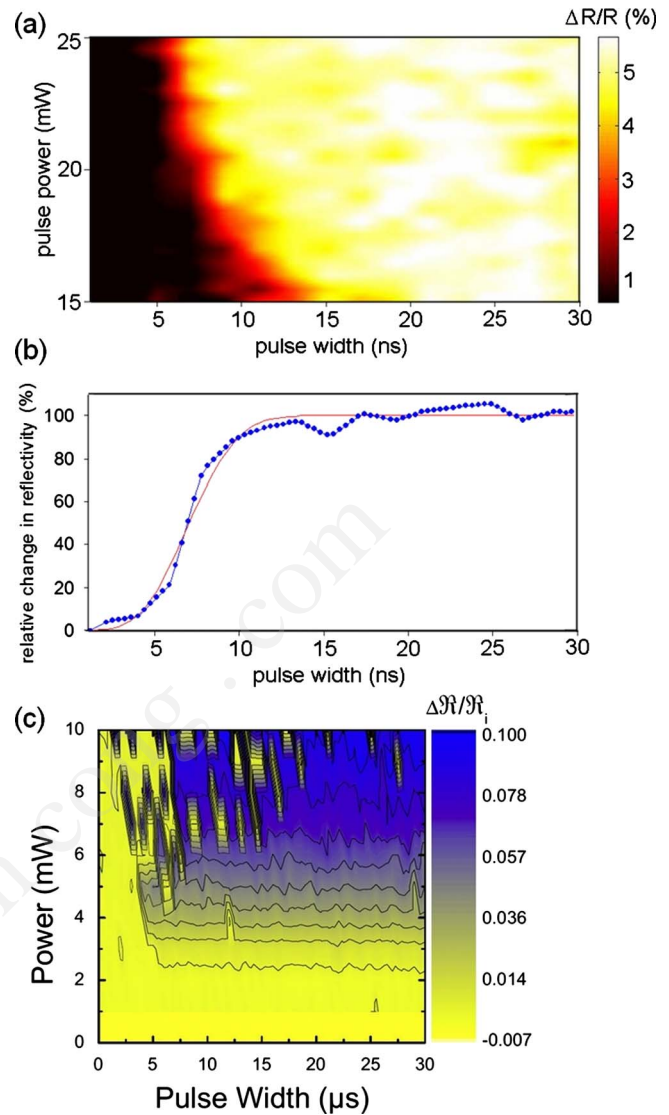

FIG. 7. (Color online) (a) Relative change in reflectivity $\Delta R/R$ in percent of a crystalline Ge–Sb thin film with 15 at. % Ge as a function of laser power and duration. The film was first crystallized by heating it in a furnace for 5 min at 300 °C. A first pulse of fixed time and power (100 ns, 50 mW) was applied to create melt-quenched spots in the crystalline film, and then a second laser spot of variable power and duration at the same location was used to recrystallize the amorphous spots. (b) Normalized change in reflectivity (in percent) integrated over a power range between 24 and 25 mW from (a) as a function of laser pulse length. The dots are experimental data, the line is a fit to $1-\exp-(t/\tau)^a$, with t being the time, $\tau=7$ ns, and $a=3$. (c) Relative change in reflectivity $\Delta R/R$ in percent of an amorphous Ge–Sb thin film with 15 at. % Ge as a function of laser power and duration. Note that much longer pulses are required. Reprinted with permission from D. Krebs *et al.*, J. Appl. Phys., **106**, 054308, 2009. © 2009, American Institute of Physics.

unused spot in the crystalline film. Then a second laser spot of variable power and duration, applied a few seconds later at the same location, was used to recrystallize the amorphous spot. Figure 7(b) plots the normalized change in reflectivity caused by this second pulse. Since the crystalline phase has a higher reflectivity than the amorphous, the increase in reflectivity observed for all pulses longer than 5–10 ns, independent of applied power, indicates extremely fast recrystalliza-
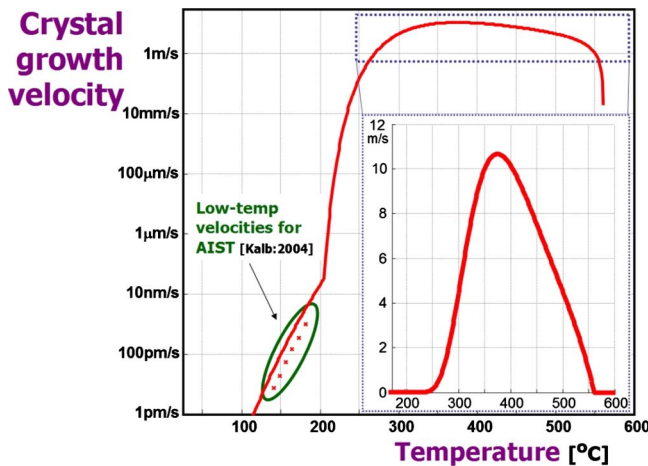
FIG. 8. (Color online) Crystal growth velocity (solid line) for GeSb as inferred by matching between simulation and empirical measurements. Low temperature crystal growth speed was measured by monitoring the slow growth of crystalline nuclei for growth-dominated (AIST) material (Ref. 93); high-temperature crystal growth speeds represent the best match between the measured optically induced recrystallization of amorphous marks on thin film GeSb and simulations of this process (Refs. 93 and 105).

tion. Because this material is crystallizing the amorphous spot predominantly by growth from the surrounding crystalline-amorphous border, this time scales with the size of the melt-quenched amorphous spot.[97] In contrast, Fig. 7(c) shows the much slower initial crystallization from the amorphous-as-deposited phase of Ge–Sb measured in a single-pulse experiment.[35]

Several factors contribute to nucleation and growth kinetics: temperature,[98] composition,[77,99,100] material interfaces,[101] device geometry,[77,102] device size,[33,103] material thickness,[42,66] polarity,[104] and device history.[96,97] Of these, the two most important factors governing switching speeds are temperature and local composition. In fact, most of the macroscopically observable nucleation effects associated with geometry, size, thickness, device history, polarity, and even material interfaces can be understood in terms of the effects of varying local composition on the delicate balance between surface and volume energies that drive crystallization.

For any given composition, the crystallization properties of a phase change material tend to be a strong function of temperature. As shown in Fig. 8, the crystal growth speed can vary by more than 15 orders of magnitude between room temperature (which is off scale in Fig. 8) and the melting point. The symbols in Fig. 8 correspond to growth speeds of less than 10 nm/s, measured by exhaustive atomic force microscopy (AFM) at temperatures below 180 °C.[93] (These data were taken for AIST, a growth-dominated material similar to GeSb, although both $Ge_2Sb_2Te_5$ and $Ge_4Sb_1Te_5$ were measured to have very similar low-temperature growth velocities[93]). The solid curve in Fig. 8 represents a simulation model built to match both these low-temperature experimental data as well as extensive recrystallization data for GeSb measured on a static laser tester. These experimental data included both reflectivity measurements similar to Fig. 7

as well as AFM measurements of the size of melt-quenched spots before and after laser pulses.[105] The sharp increase in crystal growth speed above the glass transition temperature at 205 °C is associated with a sharp drop in viscosity at these temperatures, characteristic of "fragile glass-forming" materials.[106,107]

This wide range of crystal growth speeds between moderate and high temperatures is one of the most important features of phase change materials. It allows the amorphous phase to remain unchanged for several years at temperatures near room temperature, while at programming temperatures crystallization can proceed in <100 ns. The kinetic response of phase change materials to temperature has been described by nucleation theory in great detail.[107–109] Perhaps the most influential parameters from nucleation theory involve the relation between the interfacial energy (energetic cost of adding material to a crystalline-amorphous interface) and the free energy of crystallization (thermodynamic driving force for crystallization). This interplay controls the size of the critical or smallest stable nucleus, which in turn influences the nucleation rate at which such nuclei can be incubated at lower temperatures (typically reaching a maximum rate near the glass transition temperature), and the growth rate at which large nuclei expand into the surrounding undercooled liquid (typically peaking at higher temperatures closer to the melting point). The presence of these subcritical nuclei can be detected by fluctuation transmission electron microscopy (FTEM),[110] which can explore medium-range spatial correlations beyond the nearest atomic neighbor. Using FTEM, it was observed that an increase in the population of subcritical nuclei led to a decrease in the incubation time before crystallization,[110] as predicted by classical nucleation theory.[107]

As discussed earlier, scaling is beneficial for PCM device speed. For example, it was observed that the set time (crystallization time) and reset times (melt quenching) were reduced for $Ge_2Sb_2Te_5$ material when device dimensions were reduced from 90/1.5 ns for set/reset operation for 470 nm diameter "pore" cell devices to 2.5/0.4 ns for 19 nm diameter devices.[33] Such results help move toward one important goal of phase change materials research: The quest for materials that can reliably switch at speeds comparable to RAM (approximately 10–50 ns) without sacrificing retention, endurance, or any other critical performance specification. Due to the large number of experimental variables that can contribute to switching kinetics (thermal environment, deposition conditions, changes, or damage induced during processing), extrapolating from simple thin film recrystallization experiments to PCM devices remains difficult. Crystallization from electrical pulses has been reported to range from 2.5 ns (Ref. 111) to 1 s (Ref. 112) for similar materials $Ge_2Sb_2Te_2$ and $Ge_2Sb_2Te_5$, respectively.

There are certainly phase change materials that crystallize at much higher speed than the widely used $Ge_2Sb_2Te_5$ alloy. For example, GeSb and Sb_2Te_3 are two high speed materials that have been demonstrated to crystallize in tenths of nanoseconds, comparable to the performance of consumer

SRAMs. Phase change memory devices fabricated from Ge–Te were shown to switch in resistance by nearly three orders of magnitude with set pulses of 1 ns.[36] Femtosecond laser pulses have been demonstrated to induce disorder-to-order transition in amorphous GeSb films,[42,113,114] indicating very high speed potential. However, pulse duration is often confused with material switching speed. This is an oversimplification: Crystallization and growth are thermally activated processes[115,116] and removal of stimulation is followed by a cooling period where additional crystal growth can contribute to observed phenomenon. Nevertheless, from a large aggregate of reports we can expect first generation devices to attain switching speeds of 20–200 ns. These phase change materials have the significant advantage of being highly nonvolatile at temperatures near room temperature, while retaining fast-switching speeds at high temperature. Extensive materials and device research continues to decrease crystallization times below these values,[117,118] offering hope of a bright future for PCM technology in application niches that call for rapid switching.

C. Modeling of PCM physics and devices

Because of the large number of factors influencing the performance of PCM devices, a number of groups have begun to perform predictive numerical simulations. Particularly for the consideration of reducing the reset current that can limit density by requiring a overly large access device, even straightforward electrothermal modeling of the temperature produced by a particular injected current can be highly revealing. Such electrothermal studies typically need to simultaneously solve the heat diffusion equation,

$$dC_p \frac{dT}{dt} = \nabla \cdot (\kappa \nabla T) + |\rho J|, \tag{2}$$

and Laplace's equation,

$$\nabla \cdot (\sigma \nabla V) = 0. \tag{3}$$

In these equations, temperature T and voltage V are each computed as a function of time t. Even inside each material of density d, parameters such as specific heat (C_p), thermal conductivity (κ), and electrical resistivity and conductivity (ρ and σ) are frequently functions of both position and temperature. The current density J and the temperature dependence of the electrical conductivity serve to intimately cross couple these two equations.

A number of studies have used analytical equations,[49,108,119–122] finite-element techniques,[123–131] and finite-difference techniques[42,132] to analyze either PCM cells or phase change material. Pirovano *et al.*[41] studied the reset current and the thermal proximity effect of scaled PCM by both simulation and experiment. Although analytical techniques are attractively simple and work well for explaining the incubation of new crystalline nuclei[108] or threshold switching,[49] it is difficult to include the effects of inhomogeneous temperature distributions and temperature-dependent resistivity, which critically affect the reset current through their effect on the dynamic resistance of the cell. Finite ele-

ment techniques can include these effects and work well for cylindrically symmetric cell designs, such as the conventional "mushroom" cell, since such structures can be reduced to a single (r, z) plane. However, because of the inherent computational difficulty in inverting matrices as they grow very large, these techniques are difficult to extend to three-dimensional cell designs. Finally, even though nucleation is unlikely to play an effect during the fast reset pulse, recrystallization at the end of a reset pulse does play an important role in the value of the reset current, especially for the fastest-crystallizing phase change materials that hold the most attraction for applications. The best case scenario would be to have a finite-difference simulation tool capable of handling large and arbitrary 3D structures, which could potentially be matched against fast electrical set and reset experiments, slow thin film crystallization experiments, and optical pulse experiments performed with the same material.

From our experience with such a simulation tool,[42] the reset condition is not dictated by the maximum temperature at the cell center, but by what happens at the edge of the cell. Typically, a voltage pulse just below the reset condition leaves a small portion of the limiting cross-sectional aperture remaining in the crystalline state, usually at the extreme edges of the cell.[42] In general, besides the obvious choice of reducing the diameter of this limiting aperture, the best way to reduce the reset current is to improve the efficiency with which injected electrical power heats the cell. In the best case scenario, this power would heat just the portion of the cell needed to block all of the cross-sectional apertures and produce a high-resistance state. However, in any practical case, the surrounding material is also heated to some degree. Optimization can be performed by ensuring that the heated volume is minimized and by reducing the heat loss through the thermally conductive electrodes as much as possible. Another popular way to decrease reset current is to increase the overall resistance of the cell by increasing the series resistance of the contact electrode,[133] although it is not clear how much of this benefit may be due to associated changes in thermal resistance.

As with any simulation, care must be taken to establish the boundary conditions correctly, because the computer memory available for simulation is inevitably finite. For instance, Dirichlet boundary conditions, which call for the edge of the cell to be held at room temperature, are frequently used[128] and are easy to program. However, in a simulation where the hot central region of the PCM cell is not very far from this boundary, then the effective heat transfer over this boundary can become unphysically large, skewing the results. In contrast, Neumann boundary conditions, in which the spatial derivative of temperature (or equivalently, outgoing heat flow) is held constant at the boundaries, allow a truncated simulation to act as if it is embedded within a large expanse of surrounding material.

Another important consideration is the optimization metric. It is conventional in the PCM community to discuss the importance of reset current. However, it is the dissipated power, not current, which leads to the heating of the cell. The

focus on current comes from the assumption that a colocated access transistor will have a current-voltage characteristic that saturates. If the applied current needs to be higher than this saturation value, there is no way for that particular transistor to supply it. This in turn implies that density will need to be sacrificed in order to provide this current. In actuality, the transistor and PCM device will interact in a complex manner, with the transistor supplying power to the PCM device as if it were a load resistor. The added complexity is that the dynamic resistance of the PCM device itself will in turn be a strong function of this supplied power.[122] An overly tight focus on reset current can end up optimizing into a shallow minimum in reset current, which in fact is quite disadvantageous in terms of reset power.[128]

D. Scalability of prototype PCM devices

Since so many interlocking parameters influence the performance of PCM devices, especially as they become ultrasmall, an important part of scalability studies is the fabrication and testing of prototype PCM devices. Because only a modest number of such prototype devices are typically fabricated and tested in research environments, these types of experiments cannot hope to predict the actual device reliability statistics (yield, endurance, and resistance distributions) that can be expected from full arrays. However, prototype devices are an extremely important test for the scalability of PCM—if you cannot get any ultrasmall devices to operate correctly, then this is a bad sign for the future of the technology. Here we will report on the properties of one such prototype device: the bridge cell.

The phase change bridge cell is a relative simple testing vehicle for studying novel phase change materials,[42] extending the line-device concept that had been introduced earlier[77] to ultrasmall dimensions. In these studies, two 80 nm thick TiN electrodes separated by a planarized dielectric layer were typically used as the contacts, with a thin phase change bridge fabricated to connect the two electrodes. Bridge devices have been fabricated from various materials including undoped and doped Ge–Sb with 15 at. % Ge, $Ge_2Sb_2Te_5$, Ag- and In-doped Sb_2Te, Ge–Te with 15 at. % Ge, and thin Sb phase change materials.[34,42,78]

After fabrication of these TiN bottom electrodes using KrF lithography and chemical mechanical polishing, a thin layer of phase change material (down to 3 nm thick) was deposited by sputter deposition and capped with a thin SiO_2 layer to prevent oxidation. Electron-beam lithography was used to define the phase change bridge itself. Bridge widths (set by the e-beam lithography) varied between 20 and 200 nm, and the length (determined by the spacing between the underlying electrodes) ranged from 20 to 500 nm. Figure 9(a) shows a scanning electron microscope image of the phase change bridge and the TiN electrodes, while Fig. 9(b) shows a cross-sectional transmission electron microscope image of a Ge–Sb bridge that is only 3 nm thick. Negative photoresist was used to define the bridge so that the resist did not need to be removed after the fabrication process. Ar ion milling was applied to transfer the exposed photoresist pat-
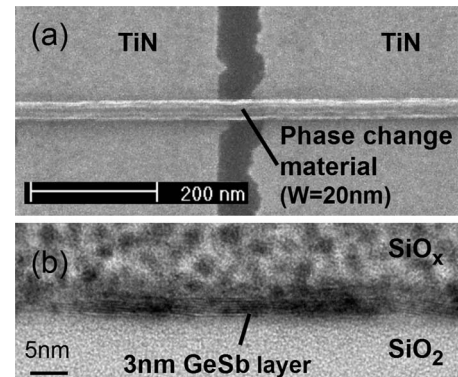

FIG. 9. (a) Scanning electron microscope image of a phase change bridge and its TiN electrodes. (b) Cross-sectional transmission electron microscope image of a 3 nm thick GeSb bridge. Reprinted with permission from Y. C. Chen *et al.*, Tech. Dig. - Int. Electron Devices Meet., 2006, S30P3. © 2006, IEEE.

tern into the phase change material, and a 50 nm thick layer of SiO_2 was subsequently deposited for protection.

These devices could be cycled through more than 30 000 set-reset cycles and the stored data were shown to survive temperature excursions up to 175 °C.[42] These ultrasmall devices—down to devices with cross-sectional apertures as small as 60 nm^2—correspond to effective switching areas that will not be encountered by mainstream device technology until the 22 nm node, which Flash is expected to reach in 2015.[5] Thus these bridge-device demonstrations show that PCM will remain not only functional but robust through at least the 22 nm technology node.

Several device parameters were measured as a function of device geometry. Current-voltage (*I*-*V*) curves revealed typical PCM behavior with threshold switching. By modifying the fabrication procedure so that the phase change material never experienced any temperatures over 120 °C, bridges that remained in the amorphous-as-deposited phase could be produced.[34,35] This allowed the precise measurement of threshold switching as a function of device length in a well-known geometry. Each material was found to have a unique threshold field, measured by determining the threshold voltage as a function of device length. These fields were 8, 19, 39, 56, and 94 V/μm for Ge–Sb with 15 at. % Ge, Ag- and In-doped Sb_2Te, Ge–Te with 15 at. % Ge, $Ge_2Sb_2Te_5$, and thin Sb phase change materials, respectively.[34,35] No leveling-off of the threshold voltage with length was observed, although the shortest bridge devices were 20 nm, as opposed to the nanowire devices where amorphous plugs shorter than 10 nm were studied (Sec. II A). Unfortunately, line edge roughness in the TiN electrodes [see Fig. 9(a)] led to shorts between the long TiN electrodes for separations smaller than 20 nm. Figure 10 demonstrates the scaling behavior of bridge devices in terms of the reset current. Both measured reset current (dots) and the predictions of numerical simulations (lines)[42] decrease linearly with the cross-sectional area of the bridge device. This indicates a favorable
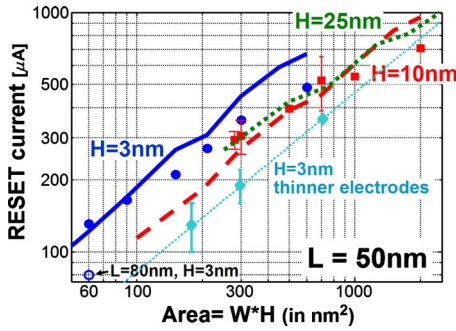
FIG. 10. (Color online) Reset current of doped-GeSb phase change bridge devices vs cross-sectional area defined by the lithographic bridge width *W* and the ultrathin film thickness *H*. Reprinted with permission from Y. C. Chen *et al.*, Tech. Dig. - Int. Electron Devices Meet., 2006, S30P3. © 2006, IEEE.

FIG. 11. (Color online) Back-of-the-envelope estimate for the expected pulse width and associated current density as function of the pitch 2*F* of the active volume of the phase change material. Also shown are the empirical current densities for a phase change bridge device [300 μA for a $H=10$ nm, $W=30$ nm bridge (Ref. 42), with the equivalent pitch for lithographic definition estimated to be $2\sqrt{(10 \times 30)} \sim 35$ nm pitch], and 160 μA for a 7.5 $\times$ 65 nm^2 dash-type cell (plotted for an equivalent 45 nm pitch).

scaling behavior because the required reset current determines the size of the access device which in turn determines the effective density of the PCM array.

It was possible to repeatedly cycle bridge devices fabricated from fast-switching Ge–Sb material with set and reset pulses of 10 ns,[34,35] confirming the observations of fast crystallization for this material seen in optical testing (Fig. 7). Very short switching times and reduced switching times with reduced device dimensions have also been observed for ultrascaled pore devices by Wang *et al.*[33]

III. DESIGN AND FABRICATION OF PCM

In this section, we discuss issues relevant to the design of PCM cells, such as cell structures and access circuitry, as well as those related to fabrication, such as the effects of variability and the deleterious effects of semiconductor processing on PCM materials and devices.

A. Cell structures

Over the next few technology generations, the most serious consideration for PCM is the large current needed to switch PCM cells. We can roughly estimate this required current using "back-of-the-envelope" calculations. The operation of the PCM cell relies on Joule heating, so the cell structure and operating conditions are dictated by the electrothermal diffusion equation shown earlier [Eq. (2)].

We can assume that the critical volume undergoing phase change within the cell in a closely packed memory array is laid out at a pitch of 2*F*. This immediately imposes the restriction that, to first order, the applied electric pulse width should be such that the thermal diffusion length should not exceed the half-pitch distance, *F*, in order to avoid cross-talk during programming. Thus

$$F > \sqrt{2D\tau}, \qquad (4)$$

where *D* is the diffusion constant defined as (κ/dC_p) and τ is the time duration of the applied electric pulse. This approximation is only good to first order, however, because thermal diffusion through a one-dimensional (1D) geometry is *not* identical to a 1D slice of diffusion through a 3D geometry.
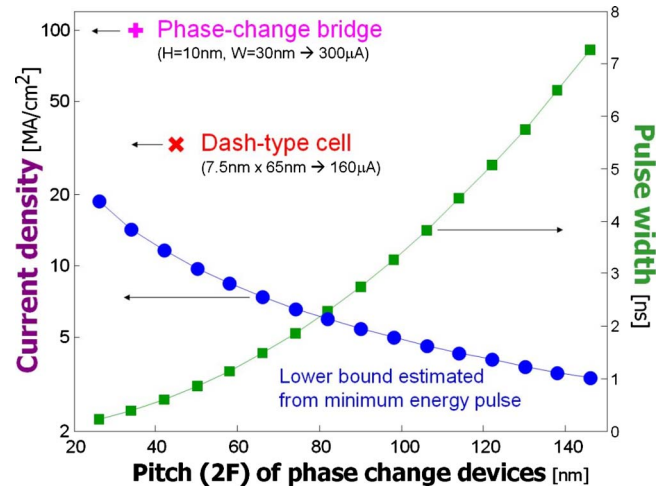
We address this issue again in Sec. IV B where we discuss cell-to-cell thermal cross-talk.

However, this computation allows us to set a lower bound for the required current density to achieve melting. We can use the expected pulse duration τ from Eq. (4) to satisfy the minimum condition that the supplied energy should be large enough to raise the temperature of the critical volume above the melting point,

$$J^2\rho > dC_p\frac{\Delta T}{\tau}, \qquad (5)$$

where ΔT is the difference between the melting point of the phase change material and ambient temperature. Note that the energy spent to heat neighboring material, as well as any inhomogeneous heating of the center of the critical volume beyond the melting point, is not included.

Figure 11 shows a plot of pulse duration τ and, more importantly, the lower bound on current density required to heat the minimum volume. Here typical material parameters for phase change materials have been used. This analysis thus suggests that we will need to supply a current density of *at least* 10^6 A/cm^2 (10^4 μA/μm^2) to be able to melt the critical volume within the cell for reset.

In contrast, the measurements and simulations shown in Fig. 10 would seem to indicate that the current density required is actually much larger, possibly as high as 300 μA for a 300 nm^2 aperture, or 10^8 A/cm^2 (10^6 μA/μm^2). However, the large expense of metallic electrodes in close proximity to these tiny prototype devices actually makes this number more pessimistic than is warranted. Other demonstrations, such as the 160 μA reset current shown for a 7.5 $\times$ 65 nm^2 dash-type cells,[134] seem to suggest a number such as $J_{PCM} \sim 3 \times 10^7$ A/cm$^2 = 3 \times 10^5$ μA/μm^2. Because the

estimated numbers plotted in Fig. 11 represent a lower bound, we will instead use this higher empirical value in the remainder of this discussion.

In a full memory array, an access device such as a diode or transistor must be included at each memory cell to ensure that the read and write currents on each bit line are interacting with one and only one memory device at a time. The amount of current that this access device can supply must comfortably exceed the required reset current of the worst-case PCM element. Unfortunately, the CMOS field effect transistors (FETs) often considered for use as access devices in a PCM memory array have limited current drive capability; most optimized devices can provide only about I_{acc} $\sim 800–1500$ $\mu A/\mu m$, where this current capability scales linearly with the effective gate width of the drive transistor. One way to solve this problem is to simply make the access device larger so that it can drive a larger current. However, since this immediately sacrifices memory density which subsequently drives up the cost per megabyte, such a move would be economic suicide for a prospective memory technology.

Given a transistor of width F, the available transistor drive-current, $I_{acc}F$, must exceed the required current for reset, $J_{PCM}\eta F^2$, where η is an area factor between 0 and 1. Although returning to the set state does involve exceeding the threshold voltage, the amount of power (and current) in the set pulse is typically 40%–80% that of the reset pulse. Thus it is almost always the reset pulse that must be considered when determining if the access device will supply sufficient current, while the set pulse is typically the factor that dictates the write speed of PCM technology. Inserting the numbers above produces

$$I_{acc}F > J_{PCM}\eta F^2,$$
$$(1.5 \times 10^3 \ \mu A/\mu m)F > (3 \times 10^5 \ \mu A/\mu m^2)\eta F^2,$$
$$5 \ nm > \eta F, \qquad (6)$$

which implies that transistor current scaling could only hope to catch up with the reset current of lithographically defined PCM devices at ultrasmall technology nodes. This is especially sobering given that PCM reset current, as shown in Fig. 10, has been empirically observed to scale with CD at a pace somewhere between $F^{1.5}$ and $F^{1.0}$, rather than as F^2.

This analysis implies that minimal-width FETs cannot supply the necessary current *if the dimensions of the phase change volume are determined lithographically*. However, use of a factor $\eta \sim 0.1$, corresponding to a sublithographic CD for the phase change element of roughly $F/3$, allows Eq. (6) to be satisfied for $F \sim 45$ nm, precisely where industry has been targeting first PCM products.[135] At this node, the required reset current of 61 μA (not that far from the demonstrated reset current in Ref. 134) could be supplied by the CMOS transistor capable of 67 μA.

Thus there are two parallel routes to ensuring sufficient reset current for PCM devices:

• Use access devices that have higher current drive capability, including either known devices such as bipolar junc-

tion transistors (BJTs) (Ref. 44) or diodes[43] or novel devices such as surrounding-gate transistor[136] or FinFETs;[137]
• locally increase the current density within the phase change element and decrease the switching volume by creating sublithographic features in the current path through the PCM element.

Both of these approaches have been pursued aggressively to demonstrate the basic operation of PCM technology. In this section, we will focus on the latter path: optimization of the phase change element itself by creation of a sublithographic aperture.

A typical PCM cell is designed so that the only current path through the device passes through a very small aperture. As this aperture shrinks in size, the volume of phase change material that must be melted (and quenched into the amorphous state) to completely block it is reduced. In turn, this decreases the power (and thus the current) requirements. If this current is low enough, then a minimum-size access device can provide enough power to switch the cell from the set state to the reset state.

In order to fabricate a PCM cell that will work even with these small currents, an innovative integration scheme is needed which creates a highly sublithographic yet controllable feature size. Subtle variations in cell design may have a large impact on critical device characteristics, including endurance, retention, set and reset resistance distributions, and set speed. These considerations will be the subject of Sec. III C. The cell design must be scalable as well as highly manufacturable since scaling implies not only a shrink in physical dimensions of the memory cell but also an increase in the number of memory cells per chip. Lastly, to maximize the number of bits per cell, a cell structure which allows multibit functionality is highly desirable.[60,138]

Depending on how this sublithographic aperture is implemented, PCM cell structures tend to fall into one of two general categories: those which control the cross section by the size of one of the electrical contacts to the phase change material [contact minimized, Fig. 12(a)][41,139–144] and those which minimize the size of the phase change material itself at some point within the cell [volume minimized, also known as confined, Fig. 12(b)].[41–43,77,103,124,138,145–147] The typical volume-minimized cell structures tend to be a bit more thermally efficient, offering the potential for lower reset current requirements compared to the contact-minimized structures.[41]

1. Contact-minimized cell

The most common contact-minimized cell structure is the mushroom cell, where a narrow cylindrical metal electrode contacts a thin film of phase change material. Figure 13 shows TEM images of mushroom cells in the set state (a) and in the reset state (b). In the reset state, an amorphous dome of the phase change material—resembling the cap of a mushroom, thus the name—plugs the critical current path of the memory cell, resulting in an overall high-resistance state for the cell. The bottom electrode contact (BEC), typically made of TiN, is the smallest element in this cell. It is com-

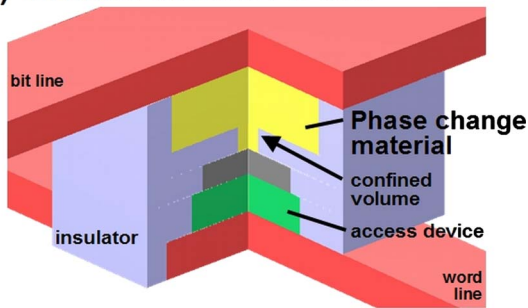## a) Contact-minimized cell
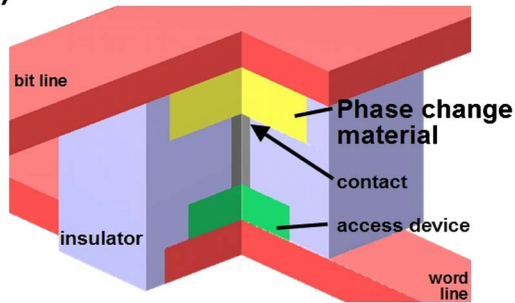
b) Volume-minimized cell

FIG. 12. (Color online) Phase change device archetypes: (a) A typical contact-minimized cell, the mushroom cell, forces current to pass through a small aperture formed by the intersection of one electrode and the phase change material. (b) A typical volume-minimized cell, the pore cell, confines the volume of the phase change material in order to create a small cross section within the PCM device. Reprinted with permission from S. Raoux *et al.*, IBM J. Res. Dev., 52, 465, 2008. © 2008, IBM.

mon to see this critical dimension (CD) described as the "heater," although the cell works most efficiently when the heat is mostly generated in the phase change material at the top of this BEC.

The sublithographic BEC can either be formed by a spacer process,[148] resist trimming,[60] or by the key-hole transfer process,[103] followed by chemical mechanical polishing (CMP) for planarization. The processing of phase change materials is discussed in more detail in Sec. III D. Prototype devices can use e-beam patterning to define the heater but
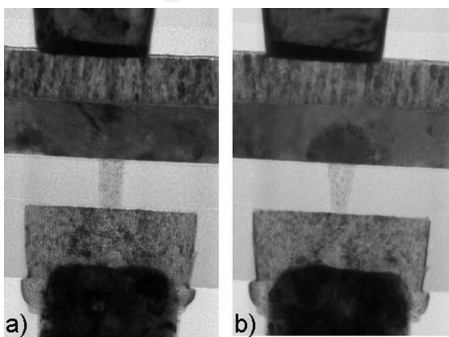

FIG. 13. TEM cross sections of a mushroom cell PCM element in the (a) set state and (b) reset state. In the set state, the phase change material is polycrystalline throughout. In the reset state, a "mushroom cap" of amorphous phase change material restricts the current flow through the bottom electrode. From Breitwisch, Phase Change Materials: Science and Applications. © 2009 by Springer.
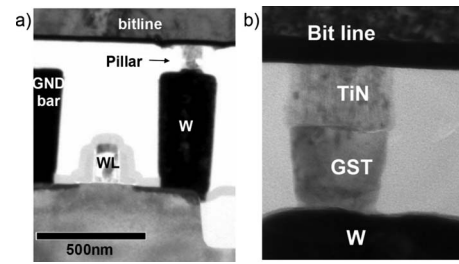
FIG. 14. (a) TEM cross section of a pillar cell with a FET access device. (b) Close-up TEM cross section of GST/TiN pillar. The simulated reset current dependence of this device is shown in Fig. 18(a). Reprinted with permission from T. D. Happ *et al.*, Tech. Dig. VLSI Symp., 2006, 120. © 2006, IEEE.

this is too slow to be used commercially. The thin film of the phase change material (or a stack of different phase change materials[149] with varying alloy concentrations) can then be deposited over the planarized feature using standard techniques such as physical vapor deposition (PVD) or chemical vapor deposition (CVD). The top electrode contact (TEC) is also deposited, usually without breaking vacuum. The simplicity of the phase change material portion of the process—and the ability to define the CD before any novel materials are introduced—represents two of the most attractive features of the mushroom cell.

The deposited films are then patterned into islands using conventional lithography to form individual cells, and isolated and encapsulated using thermally insulating dielectric materials such as Si_3N_4.[143] A variety of materials engineering techniques has been introduced to optimize the cell performance, especially the minimization of reset current. These include increasing the resistivity of either the electrode material[150] or the phase change material,[133] and decreasing the thermal diffusive losses through both the top and bottom electrode regions.[130,151]

Although cells with horizontal heater electrodes have been demonstrated,[152] the vast majority of contact-minimized cells closely resemble the mushroom cell. One popular variant is the ring-electrode mushroom cell, where the heater electrode consists of a thin ring of metal surrounding a center dielectric core.[153] The incentive here is to reduce both reset current and variability by decreasing the effective area, since compared to a normal heater the metal annulus has a smaller area, which also scales only linearly with CD.

2. Volume-minimized cell

Significant research efforts have been spent exploring a variety of volume-minimized cell structures, owing to their superior scaling characteristics. However, achieving such a structure can be a challenge, requiring the development of processing technologies that can successfully confine the phase change material within a sublithographic feature. The most obvious structure in this category is the pillar cell (Fig. 14), where a narrow cylinder of phase change material sits between two electrodes.[138] This cell is fabricated in a similar fashion to the mushroom cell, with a thin film stack of the phase change material and top electrode material deposited
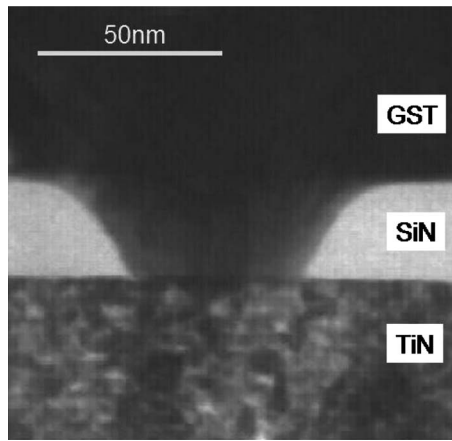
FIG. 15. TEM cross section of a 45 nm bottom CD low aspect-ratio pore cell filled with a PVD GST process. The simulated reset current dependence of this device is shown in Fig. 18(b). Reprinted with permission from M. Breitwisch *et al.*, Tech. Dig. VLSI Symp., 2007, 100. © 2007, IEEE.

atop a bottom electrode, and then patterned. However, in the pillar cell, the BEC is large, and it is the phase change material that must be successfully and reliably patterned into sublithographic islands. This patterning can be performed in various ways, including lithography followed by resist trimming.[138] In addition to the challenges in controlling the size of the patterned islands, this cell structure also suffers from the drawback that the reactive ion etch (RIE) of the phase change material can form a thin layer of altered alloy composition at the surface, strongly affecting the performance and yield of the cell.[154] This is discussed in further detail in Sec. III D.

A modified version of the pillar cell structure is the pore cell,[103] where a sublithographic hole formed in an insulating material atop the BEC is filled with the phase change material (Fig. 15). Conformal filling of nanoscale holes with high aspect ratio is difficult using conventional PVD processes; hence development of CVD or atomic layer deposition (ALD) technology for phase change materials will be neces-

sary to enable continued scaling of pore cell devices.[155] As with nearly all PCM cell designs, the dimensions and aspect ratio of the phase change material region critically influence the reset current.

3. Hybrid PCM cells

The most advanced scaling demonstration of PCM technology to date was realized using the volume-confined bridge cell,[42] which consists of a narrow line of ultrathin phase change material bridging two underlying electrodes (Fig. 9). The cross-sectional area of this device is determined by film thickness in one direction and by electron-beam lithography in the other, allowing the realization of functional cells with cross-sectional area of about 60 nm^2 and reset current requirement of about 80 μA.

The μ-trench cell[139] and the dash-confined cell[134] are examples of PCM cells that combine the contact-minimized and volume-minimized approaches. The μ-trench cell is an extension of the bridge concept, with a PCM element formed at the intersection of an underlying sidewall-deposited CVD TiN bottom electrode and a trench of phase change material formed at right angles across this electrode (Fig. 16). The dash-confined cell[134] is an extension of the pore cell idea, except that the bottom electrode contact is formed by a spacer process. A CVD process is then employed to fill in the rectangular sublithographic holes formed by the recess etch into the metal BEC (Fig. 17). In all three of these cases, one critical dimension associated with the limiting cross-sectional aperture is controlled by a thin film deposition process. As a result, only one dimension inherits the variability of the lithography process. Conversely, of course, only one dimension will enjoy the associated scaling benefits as F shrinks from one technology node to the next.

B. Access circuitry

In order to fully leverage the scalability of PCM and thus achieve the very high densities needed for SCM, the most ideal implementation of PCM would be a cross-point array
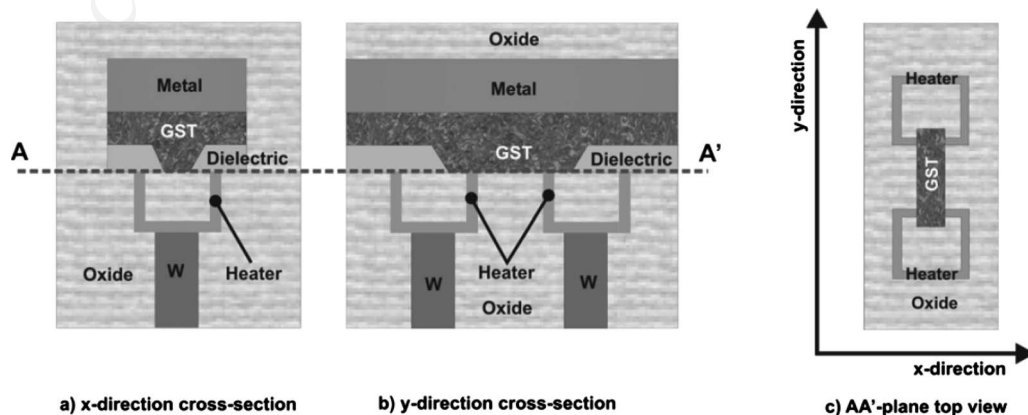

FIG. 16. Illustration of the μ-trench cell, showing two neighboring devices with a common top electrode (e.g., along the bit line). Current passes through an aperture which is limited in one dimension by the thickness of the underlying sidewall-deposited metal heater, and in the other dimension by the width of the narrow trench in which phase change material (here, GST) is deposited. Reprinted with permission from F. Pellizzer *et al.*, Tech. Dig. VLSI Symp., 2004, 18. © 2004, IEEE.
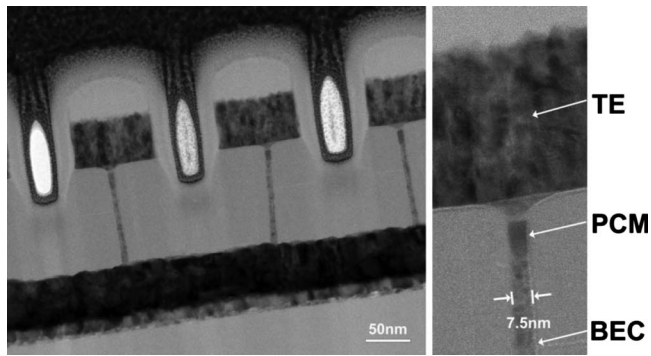
FIG. 17. TEM cross section of the dash-confined cell, showing devices fabricated by a spacer process that are only 7.5 nm wide, and 65 nm deep in the orthogonal direction. Reprinted with permission from D. H. Im *et al.*, Tech. Dig. - Int. Electron Devices Meet., 2008, 9. © 2008, IEEE.

architecture, where each memory element in the array is directly connected to two orthogonal lines (Fig. 5). In fact, a novel architecture has been proposed for such a direct cross-point memory, in which PCM devices are switched not between set and reset, but between the reset state and an over-reset state (strongly reset state).[156] Even though these resistances may differ, the difference in read current would be too low to support rapid read. However, because the threshold voltage varies between these two states, the device state can be sensed by detecting whether a "read" voltage intermediate between these levels produces an electrical switching event. While this scheme does cleverly avoid an access device, it has several serious issues. The margins between the read and the two threshold voltages are uncomfortably tight, and it would be difficult to detect the breakdown without potentially heating up the cell, which means that reads must be treated as destructive. Even so, typical reset resistances are still low enough that the leakage through "half-selected" devices (those that share either the same bit line or the same word line as the "selected" device) would be quite high, thus limiting the maximum array size (and the effective memory density) that could be built. Worst of all, the strong negative correlation between switching energy and endurance (which will be discussed in detail in Sec. IV C) means that improving the sense margin by increasing the resistance (and thus the threshold voltage) of the over-reset state will sharply reduce endurance.

Thus the integration of PCM into an array architecture seems to require the use of an access device: either a diode,[43,157] a field effect transistor,[158,159] or a bipolar junction transistor.[44,61] The main role of this device is to minimize the leakage current that would otherwise arise from the nonselected cells in the array. As has been mentioned, the most important unknown for the success of PCM technology is whether this memory access device is able to provide sufficient current to reset the PCM cell. While a diode can provide a current-to-cell size advantage over a planar transistor down to the 16 nm node,[160] the diode scheme is more vulnerable to write disturbs due to bipolar turn-on of nearest-neighbor cells.[43] A $5.8F^2$ PCM diode cell has been demonstrated using a 90 nm technology, in which the diode was able to supply 1.8 mA at 1.8 V.[43] In comparison, a 90 nm $10F^2$ trigate FET could only supply approximately half as much current.[43]

Although the raw footprint of the access device is of primary concern for memory density, other considerations can also come into play. Peripheral circuits, such as charge pumps to increase the voltage supply level or special read and write circuitry for MLC operation,[161] reduce the portion of the chip that can be dedicated to memory devices (the array efficiency of the chip). In addition, the effective area per cell can grow because of additional vias or wiring that may be required within the memory array. For instance, given the high currents required for PCM programming, the voltage drop along metal bit and word lines can become significant, further reducing the power that can be delivered to the actual PCM cell. Thus splitting an array in half to reduce the maximum line length is attractive because it reduces the worst-case line loss, but detrimental because chip real estate is now being used for redundant wiring rather than memory devices. Particularly problematic is the wiring that must sit under the PCM layer (such as common-source lines, as well as the word lines to transistor gates) since these lines must typically use high-resistance tungsten or polysilicon. Because of the significant voltage losses in such lines, a via must often be introduced every few (e.g., four or eight) cells in order to strap this line to an overlying low-resistance copper line. This extra via immediately increases the effective footprint per memory cell. Although copper cannot be introduced near the transistors, lest the CMOS devices become degraded, one solution would be to move the PCM devices higher up away from the silicon so that a layer of copper interconnect can be fabricated under the PCM devices.

C. Variability

Different types of variability can affect the operation, performance, endurance, and reliability of PCM devices, ranging from intercell variability introduced during processing, intercell variability as resistances change over time after programming, and cycle-to-cycle variation of the set and reset resistances of any given cell (intracell variability). While the read voltage, the reset pulse, and the set pulse can be optimized for the average memory cell, variations between cells must be minimized so that these same choices can successfully operate all the cells in the memory array. The same is true for the performance of the access device.

Any fabrication- or process-induced variability in the physical structure of the phase change element can result in devices which react differently to the same stimuli. Thus any large collection of phase change elements, which because of variability are in similar yet nonidentical states, may then propagate in time on different resistance trajectories. In addition to group behavior caused by intercell variability, there is also intracell variability produced by motion and rearrangement of the atoms within the active region of the phase change device during the programming of the phase change device. Understanding the variability of PCM devices is particularly important in the context of multibit stor-

age capability, due to the reduced margin between resistance levels that must be correctly sensed in order to successfully retrieve data. Finally, variability can be expected to play an increasingly important role in the scaling of PCM technology into future technology nodes.

1. Structural variability

Each step in the process of fabricating a wafer of PCM memory devices is typically associated with one or more physical attributes (e.g., thickness, CD, sidewall angle, etc.), each with a nominal target value. To illustrate the variability challenge inherent in wafer-level fabrication process, features on the order of nanometers (10^{-9} m) must be accurately controlled across the surface of a wafer which spans nearly a third of a meter (0.3 m). Exact control over such a large range of dimensions is impossible. Consequently, each process is associated with an acceptable range around some target value.

Furthermore, this fabrication process must construct a fully integrated set of devices comprising a CMOS technology (CMOS field effect transistors, diodes, resistors, capacitors, wiring levels, and the vias connecting them) in addition to the PCM devices (which usually reside on top of the CMOS devices, at the bottom of the wiring levels). Thus a wafer will undergo hundreds of processing steps (including photolithography, atomic implant, reactive ion etch, material deposition, chemical mechanical planarization, wet etches, etc.), each with associated variability, before the final processed wafer is ready for the dicing and packaging of the chips. Variability in processes that affect either the structure of the phase change element or the access device (transistor, diode, etc.) can contribute to the overall phase change device variability. These structural variations then translate into variations in electrical (device operation) properties of the device.

Most relevant to the operation of PCM are the structural physical properties, which affect the temperature profile (how much heat is generated and where), the critical limiting cross-sectional aperture (which must be fully blocked to get high resistance contrast), and the volume where the phase change material undergoes repeated melting and crystallization. Crucial features of the cell are the aperture size and shape, the thickness and uniformity of the phase change material (in both stoichiometry and doping), the resistivity and interface resistance of electrodes, the thermal conductivities of surrounding materials, and the stresses on the active switching volume introduced by surrounding material.

As discussed in Sec. III A, several cell structures have been proposed and developed in order to minimize the required power (current) for reset. Each of these structures employs a similar strategy for minimizing the reset power: At one and only one point within the cell, the electrical current is forced to pass through a small aperture. This aperture increases the current density, maximizing the thermal power which is generated, and reducing the volume of high resistivity material needed to significantly alter the external device resistance. Together with the electrical and thermal
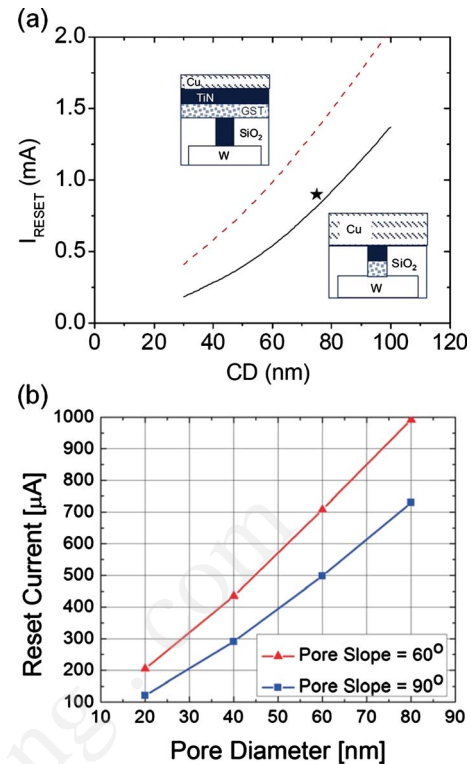

FIG. 18. (Color online) (a) Reset currents for the pillar cell and the mushroom cell both show a strong dependence on the critical aperture size. Reprinted with permission from T. D. Happ *et al.*, Tech. Dig. VLSI Symp., 2006, 120. © 2006, IEEE. (b) The pore cell reset current is both strongly dependent on the aperture size and shape (pore slope). (Figure 10 shows how the reset current of the bridge cell scales directly with the cross-sectional area of the phase change material.) Reprinted with permission from M. Breitwisch *et al.*, Tech. Dig. VLSI Symp., 2007, 100. © 2007, IEEE.

properties of the phase change material, electrodes, and surrounding materials, the size and shape of this aperture determines the temperature profile obtained within the phase change element for a given set or reset programming current pulse.

In order to estimate the impact of small variations of the aperture size on the cell operation, it is instructive to examine the functional dependence of the aperture size on the required reset programming current. We show the published dependence of the reset current on the aperture size for the bridge cell (Fig. 10), the pillar and mushroom cells [Fig. 18(a)], and the pore cell [Fig. 18(b)]. In each case, the required current is a steep function of critical dimension. Variability introduces a distribution of PCM cells of different sizes and thus different reset currents. In order to be sure to be able to reset all of the devices, it is thus the largest diameter cell that dictates the required current-driving capability of the access device. Variability directly reduces density by mandating a larger-area access device. At the same time, as we will see in Sec. IV C, switching cells with more power than is necessary has a strong and negative influence on device endurance. Thus this same variability may also reduce endurance in the smaller-area devices, which are being driven much harder than they really need to be.
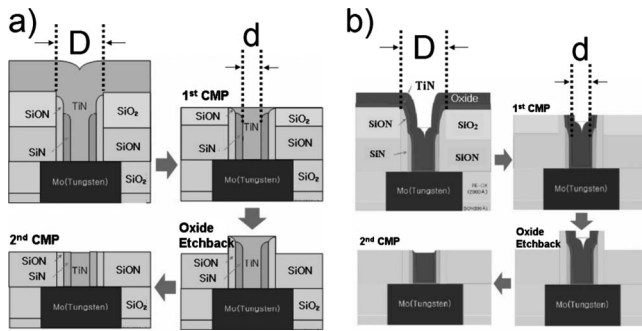
FIG. 19. (a) Collar process is used to create a sublithographically sized TiN bottom electrode. First, a lithographically defined hole of diameter D is etched into a SiON/SiN stack. A first collar is formed by depositing a conformal SiON layer followed by a collar RIE step. A second collar is formed in the same manner. Next, the CVD TiN is deposited to fill the hole. Finally, a series of CMP and oxide etch back processes are performed, resulting in a cylindrical TiN bottom electrode. (b) A TiN ring electrode is constructed in a similar manner except that only a thin layer of CVD TiN is deposited into the hole, and then the center of the hole is filled with oxide. Original figure from Ref. 140. From Breitwisch, Phase Change Materials: Science and Applications. © 2009 by Springer.
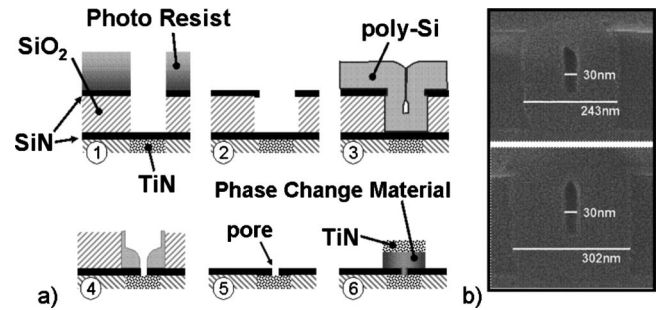

FIG. 20. (a) Sublithographic and lithography-independent feature is fabricated using the keyhole-transfer process: (1) A lithographically defined hole is etched and (2) the middle SiO$_2$ layer is recessed. (3) A highly conformal poly-Si film is deposited, producing a sublithographic keyhole whose diameter is equal to the recess of the SiO$_2$ layer. (4) The keyhole is transferred into the underlying SiN layer to define a pore, followed by (5) removal of the SiO$_2$ and poly-Si. (6) The phase change and top electrode (TiN) materials are then deposited and the cell is patterned for isolation. (b) A SEM cross section corresponding to step (3), showing keyholes for two different sized lithographically defined holes. Since the keyhole size does not depend on lithography, the phase change CD can be successfully decoupled from any lithographic variability. From Breitwisch, Phase Change Materials: Science and Applications. © 2009 by Springer.

2. Sources of structural variability

There are several ways a sublithographically sized aperture can be defined. Not surprisingly, the degree to which the aperture size can be controlled varies for each of these methods.[162] A direct way to reduce the size of a lithographically defined hole is to implement a collar process, as shown in Fig. 19. Similarly, a subtractive method can be used (as in the pillar cell scheme[138]), where an island of lithographically patterned photoresist is trimmed in size using RIE and then used as a mask to transfer the sublithographic pattern down into the underlying layers. Unfortunately, in both of these techniques, any variability in the diameter of the original hole (or photoresist pillar) introduced by either lithography or etch (or resist development) transfers directly to the final CD. Thus the fractional variability ($\Delta CD/CD$) can become uncomfortably large.

The μ-trench cell,[139] the ring bottom electrode mushroom cell,[153] the dash-confined cell,[134] and the bridge cell[42] define one of the dimensions of the cross-sectional area of the aperture through film deposition. The thickness of such deposited film can be tightly controlled, especially for CVD and even more so for ALD techniques. This thickness can easily be much thinner than the lithographic dimension F (at least for current and near-future technology nodes). This method of defining one dimension of the aperture by film thickness is inherently decoupled from any lithographic variability. However, for these schemes, lithography is still needed to define the "other" dimension of the aperture.

A keyhole-transfer process has been developed for PCM devices, which decouples both dimensions of the final aperture from lithography. Here, a keyhole is defined within a lithographically defined hole by film deposition. Typically, a keyhole is undesired and indicates a failure to fill the hole. However, the advantage is that the dimension of the keyhole itself can be tightly controlled by accurate control over the hole depth (through film deposition) and over the etched un-

dercut into the sidewalls of the hole, despite poor control over the actual lithographically defined diameter of the hole itself. Figure 20 describes this process and shows how the keyhole process can produce identical sublithographic holes (30 nm) despite significant variation in the much larger lithographically defined holes (243 and 302 nm).

The keyhole-transfer process has been experimentally demonstrated to decouple the final aperture size from the initial lithographically defined hole size.[103] Over a span of initial lithographically defined hole sizes, the distribution of reset current was found to be consistently narrower for pore cells fabricated through the keyhole-transfer method than for those fabricated with a collar process. Similar results have been demonstrated for mushroom cells.[122] Here, mushroom cells with heaters defined with the keyhole-transfer process were compared to otherwise identical mushroom cells with heaters defined by trimming of photoresist islands. This comparison was performed by examining the dynamic resistance R_{dyn} during programming, which tends to exhibit a dependence on programming current I as

$$R_{dyn} = \frac{A}{I} + B. \tag{7}$$

Here the term A depends only on material characteristics, while B incorporates both material and structure-dependent factors.[122] Thus the lower variability in B empirically observed for mushroom cells with heaters defined by the keyhole-transfer method (as compared to those defined by trimmed photoresist) is indicative of the tighter CD control offered by the keyhole-transfer method.

3. Impact of structural variability

Structural variability gives rise to variability in electrical response, which leads to broader resistance distributions after single-shot programming. Figure 21 shows a series of set
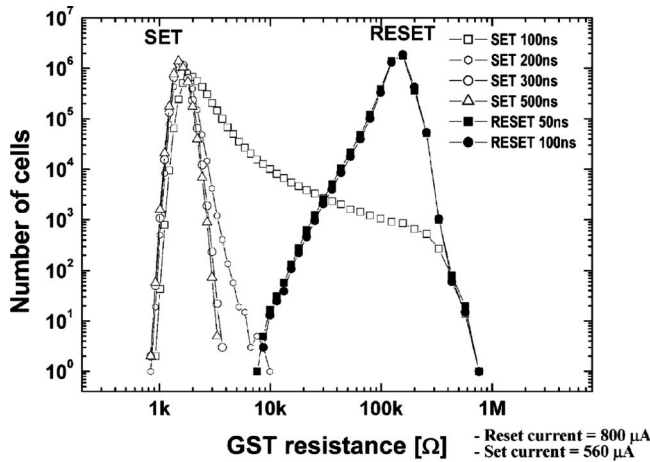
FIG. 21. Set resistance and reset resistance distributions as a function of the programming pulse width. In this example, while 100 ns is sufficient to set most of the cells to below 2 kΩ, many cells still have a resistance greater than 10 kΩ. However, extending the 50 ns reset pulse to 100 ns has no noticeable effect on increasing the resistance of the reset tail. GST refers to the phase change material used in this experiment, Ge$_2$Sb$_2$Te$_5$ (Ref. 249). Reprinted with permission from S. Kang *et al.*, IEEE J. Solid-State Circuits, 42, 210, 2007. © 2007, IEEE.

FIG. 22. (Color online) Reset tail modulation by the quenching time t_Q. A long quench time results in a partial set of a small fraction of devices within a large array. Reprinted from with permission from D. Mantegazza *et al.*, Solid-State Electron., 52, 584, 2008. © 2008, Elsevier.

resistance distributions, including one (for 100 ns set pulses) which is strikingly broad. Although the majority of cells can reach a low resistance of approximately 2 kΩ with a set pulse of 100 ns duration, for this collection of cells there is a subset of devices whose resistances after such a single set pulse extend all the way out to the fully reset resistance of several hundred kilo-ohms. For a given programming current amplitude, devices with different diameters will present different dynamic resistances during programming, thus dissipating different amounts of power despite the same drive voltages. This variable power will lead to a different maximum temperature, and even a different temperature distribution within the cell because variations in aperture or heater size affect the thermal resistances within the cell. For reset pulses, the size of the amorphous plug required to significantly affect the room temperature low-field resistance of the cell changes drastically with changes in the cross-sectional aperture of the cell. In terms of set pulses shown in Fig. 21, rapid set requires that the optimal temperatures for crystal growth be present at the crystalline-amorphous boundary (growth-dominated material) or within the cell interior (nucleation-dominated material). Since the crystallization speed is a strong function of temperature (Fig. 8), variability in aperture size can result in a variety of incompletely set cells if the pulse is too short. As Fig. 21 shows, increasing the duration of set pulses tends to overcome this effect. Another even more effective route is to ramp down the set pulse slowly, allowing each cell to pass through the temperature for maximum crystal growth.[163]

In addition to broadened resistance distributions, device variability also affects how these resistances evolve over time. It is well known that the resistance of cells in the reset state tends to increase slowly over time, an effect which has been attributed t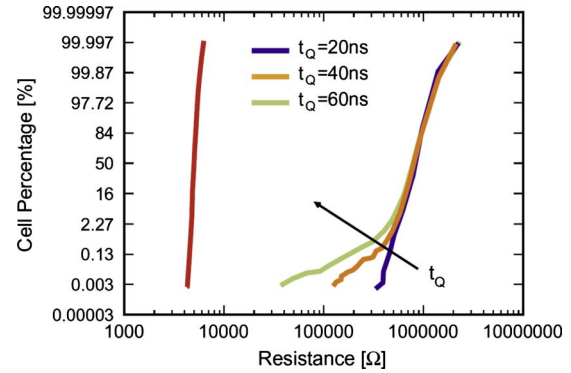o either mechanical relaxation forced by the reduced density associated with the amorphous state,[164,165] to the formation of electronic traps associated with lone-pair states which increases resistance by repositioning the Fermi level,[91] the annihilation of defects by trap filling which reduces transport and thus increases resistance,[89,166] or to some combination of these effects. Since this drift increases the already high resistance of the reset state, it is not an issue for binary PCM devices. However, drift can be particularly problematic in the context of multiple bits per cell, as discussed in Sec. V A.

Drift would be bad enough if all cells changed in resistance along the same trajectory. However, variability introduces different drift coefficients, so that cells with similar resistances immediately after programming tend to separate in resistance over time. In addition, some variability in crystallization speed has been observed over large arrays of PCM devices, so that some cells tend to set more easily (both at elevated temperature and after low-temperature anneals) than the average cell.[167] Figure 22 illustrates the signature of this effect, with reset tail bits emerging when a reset programming pulse with an insufficiently short quench time is used for programming. Although such anomalous devices can be reset to high resistance by using pulses with rapid quench, such tail-bit devices will still be associated with increased long-term retention loss and resistance distributions that broaden more rapidly over time. In general, combining the effects of drift and recrystallization, variable drift effects tend to broaden the resistance distribution of a large collection of cells over time: Some cells increase in resistance due to structural relaxation, some cells decrease due to partial recrystallization. While some may first increase and then only later decrease, the random walk nature of these effects typically leads to broader distributions over time.

4. Intradevice variability

The act of programming a PCM cell involves the rearrangement and movement of atoms within the cell. When a given cell, starting from the polycrystalline set state, is melt quenched and then crystallized back to the set state, the distribution of crystal grains within the cell is not identical. The
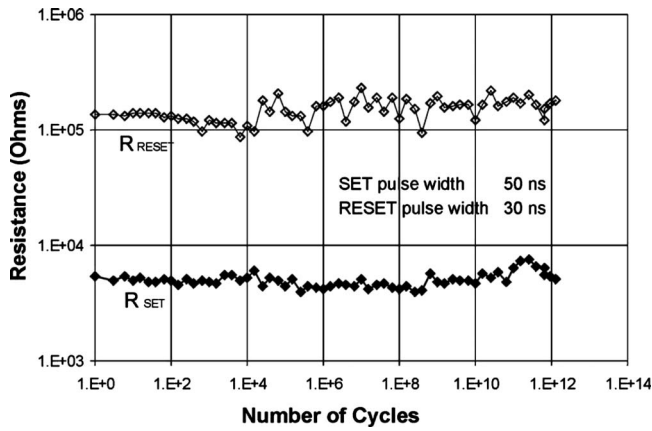
nucleation of these crystalline grains within the amorphous plug leads to a decrease in reset resistance, which accelerates at elevated temperature.[168,169] It has been shown that a small fraction of cells in the reset state will tend to show these effects more rapidly than the average cell, but that the location of these cells is random from cycle to cycle. A cell may participate in this tail of the resistance distribution on one cycle, yet show much improved resistance to the anneal upon the next, consistent with random nucleation of crystalline grains within the amorphous plug.[169] Thus we can expect that this random formation of crystalline grains is present throughout the portion of the cell that reaches elevated temperatures.

Upon the application of a subsequent programming pulse, the temperature distribution of the cell may not be exactly identical to the previous cycle, leading to variations in resistance. This intradevice programming variability can be readily observed in any single-cell endurance plot, as illustrated in Fig. 23.

5. Variability and MLC

Variability also forces the use of iterative write schemes for programming of multilevel resistance states, as illustrated in Fig. 24.[60] Here, a collection of cells is programmed with single current pulses, with the trailing edges controlled to produce intermediate resistance values. The highest and lowest resistance levels have the smallest variation in resistance—for these states the reset and set resistances tend to saturate, as overly large voltages produce little additional increase in reset resistance and overly long pulses produce little additional decrease in set resistance. The intermediate resistance levels, which represent hybrid states where a smaller portion of the cell is amorphous than in the full reset state,[170] are much more sensitive to variability.

Even so, since each given cell does respond in a somewhat reproducible and hence predictable manner, the desired resistance value can be produced by an iterative write scheme. As illustrated in Fig. 25, the same cells used in Fig. 24 can be programmed into relatively tight resistance distri-

butions using iterative write attempts, where the length of the trailing edge of the pulse is carefully controlled based on the just-accumulated prior experience with that cell. Such iterative write schemes and other considerations of multilevel cell programming are discussed in more detail in Sec. V A.
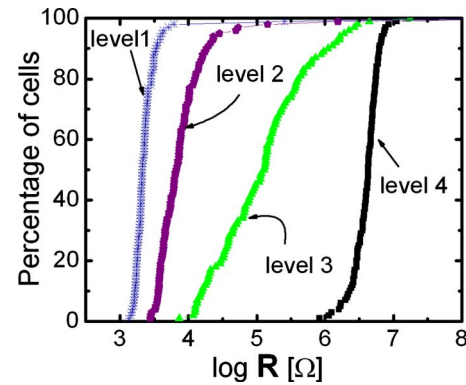
6. PCM scaling and variability

As PCM devices scale into future technology nodes, variability can be expected to become an even more important consideration. Difficulties in scaling lithography and processing techniques may lead to even looser specifications on relative CD uniformity in future technology nodes since the acceptable range of control will be driven by considerations for conventional CMOS devices and not by PCM technology. At future technology nodes, PCM cell designs which depend on control over film thickness may no longer be as attractive, if CD is shrinking while both the minimum thickness t and the achievable control of this variable, Δt, remain relatively static. As the number of atoms participating in the
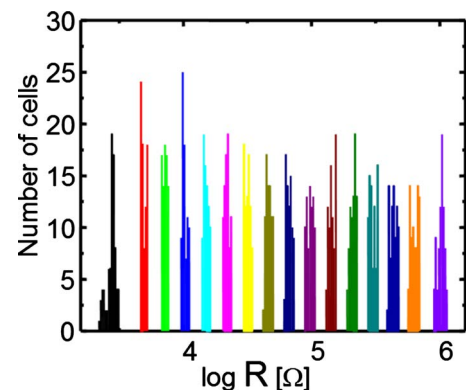
active portion of the PCM cell decreases, any variations in local doping or stoichiometry can be expected to affect individual cells, reducing yield and further broadening distributions. Eventually, in the same way that Poissonian effects have arrived for CMOS devices (e.g., as the number of dopants in the channel becomes countable),[171] such few-atom effects can be expected to be observable for PCM. That said, all logic and memory devices seeking to operate in this regime will have some variant of these effects—the winners will be those with not-yet-known physics that keeps those effects sufficiently unlikely, or those amenable to not-yet-invented engineering techniques that can finesse these issues to an acceptable degree.

D. Processing

Most studies related to the integration and processing of PCM technology focus on the popular variant, $Ge_2Sb_2Te_5$, here abbreviated simply as GST, along with various choices of dopant. This reflects the fact that GST possesses many favorable characteristics and has thus received the bulk of attention of technologists attempting to take PCM technology from research to development. Although these studies reveal numerous characteristics which might be considered less than perfect, the mere existence of such a large base of knowledge also represents a substantial hurdle for any alternative phase change material hoping to challenge GST.

1. Deposition

The vast majority of phase change materials reported in literature have been deposited by sputtering, i.e., PVD. Sputter deposition from multiple targets makes it very simple to try different compositions. However, sputtered films typically do not have good step coverage and cannot completely fill high aspect-ratio vias without keyhole formation.[103] A number of interesting PCM cell structures have been discussed (Sec. III A) that call for confining the phase change volume inside a contact hole (or "pore") formed in a dielectric, primarily to reduce the reset current. It is therefore essential to examine alternatives to sputter deposition for GST and other phase change materials that can successfully fill higher aspect-ratio vias.

Cho *et al.*[146] developed a novel GST sputter deposition process with *in situ* deposition/etch/deposition, in order to fill GST into high aspect-ratio ($>2:1$) pores of approximately 50 nm diameter. A number of groups[134,155,172,173] have investigated CVD of GST and related phase change materials. Kim *et al.*[172] deposited hexagonal phase GST by metal oxide chemical vapor deposition at temperatures in the range of $330–370\ °C$ using precursors that were bubbled at different temperatures, demonstrating complete filling into 120 nm diameter trenches with aspect-ratio larger than 1.6. However, it should be noted that in contrast to sputter deposition, CVD processes are far less flexible with respect to changes in material composition.

Lee *et al.*[155] used CVD to deposit GST from metal-organic precursors and hydrogen at $350\ °C$ into high aspect-ratio (3:1) 50 nm contact holes. X-ray diffraction (XRD) measurements showed that the GST had a hexagonal phase and was thermally stable up to $400\ °C$. CMP was used to planarize the GST, confining it to the contact holes (and removing it from everywhere else). Devices fabricated with this technique showed a significant reduction in reset current as well as good endurance. Im *et al.*[134] used CVD GST to fill high aspect-ratio (4:1) dash-type contacts of 7.5 nm width. As expected from device modeling, these cells showed very low (160 μA) reset currents. In addition, these devices also had fast set speeds and very good endurance.

Lee *et al.*[174] developed ALD of GST with the assistance of a H_2 plasma. They synthesized new precursors for the GST deposition and obtained good step coverage (90%) in 7:1 aspect-ratio holes. Finally, Milliron *et al.*[71] developed a solution-based deposition method for GeSbSe films using hydrazine and Se to form soluble precursors, so that film composition and properties were tunable through appropriate combination of these novel precursors in solution. XRD and laser pulse annealing were used to study the phase change and crystallization speed, and complete filling was demonstrated in 2:1 aspect-ratio vias patterned in thermal silicon oxide.

2. Etching

Etching of phase change materials has been explored using both wet as well as dry etching schemes. Such steps are important for electrical isolation in most cell designs, and are absolutely critical for cell designs such as the pillar cell that depend on subtractive processing to produce a confined volume of phase change material. Some important parameters that are studied in developing etch processes are etch rate, selectivity (i.e., how fast the desired PCM etches compared to the masking layer and other surrounding films), and anisotropy (etched sidewall angle; steeper profiles are usually desirable since they enable higher resolution patterning). It is also important to understand etch-induced material modification and other sidewall damage effects since these could impact device operation, especially if the damaged portion is close to or part of the active switching volume of the PCM.

A number of groups reported wet etching of phase change materials using alkaline[175–177] as well as acidic[178,179] solutions. Depending on the etchant and specific phase change material etched, in some cases the crystalline material was found to have etched faster than the amorphous phase while the opposite was true in other cases. Cheng *et al.*[178] used x-ray photoelectron spectroscopy (XPS) and inductively coupled plasma (ICP) to study the wet etch mechanism of GST. They explained their results, obtained with a 20% aqueous solution of nitric acid, as a chemical etching process that involves bond breakage, oxidation of the various constituents, and dissolution of those oxides. They inferred that the Sb component was hardest to etch. The authors used this wet etching process to fabricate large PCM cells (400 $\times 400\ \mu m^2$) and performed *I-V* measurements showing successful switching of devices from the high- to low-resistance states.

There have been a number of studies of plasma etching of phase change materials such as GST and its doped variants. A variety of gas chemistries has been reported in literature, including Cl$_2$/Ar,[180–185] CHF$_3$/Ar,[180] CF$_4$/Ar,[185,186] HBr/Ar,[187] CHF$_3$/O$_2$,[188,189] and CHF$_3$/Cl$_2$/Ar.[154] Some of these studies have examined the effects of varying process parameters such as reactant gas concentration fraction, chamber pressure, coil power, and dc bias on the etch rate and anisotropy (i.e., etched sidewall profile).

Yoon *et al.*[180] indicated that the GST removal mechanism in high-density helicon plasma etching using Cl$_2$/Ar chemistry was due to ion-assisted chemical etching. They reported GST etching selective to SiO$_2$ and did not notice any significant change in GST composition after etching. In the same work, the authors also studied the CHF$_3$/Ar chemistry and observed that the GST removal in this case was by physical etching due to ion bombardment. They noticed a small decrease in the Ge/Te ratio after this etching process. Min *et al.*[184] also performed a systematic study of GST etching in Cl$_2$/Ar using an ICP. Using a model that analyzed etch kinetics, they explained the GST removal mechanism as a combination of spontaneous and ion-assisted chemical etching. They suggested that TeCl$_4$, not being as volatile as the Ge and Sb chlorides, accumulated on the surface and was removed by ion-stimulated desorption. In their etching studies of GST films using HBr/Ar in an ICP tool, Lee *et al.*[187] pointed to the role of hydrogen in passivating the etched surface as well as the sidewalls, leading to more anisotropic etching and lower etch rates at higher HBr concentrations. Their XPS analysis showed that as the etch progressed, the surface became Te deficient and Ge and Sb rich implying that Te reacted most readily with the bromine and was preferentially etched.

Attempting to form nanoscale features into GST with SiO$_2$ hard mask and e-beam lithography (HSQ resist), Yoon *et al.*[185] reported that using the Cl$_2$/Ar chemistry to etch the GST resulted in undercut and collapse of small GST features due to the isotropic component of the etching. By switching to a TiN hard mask process and using a CF$_4$/Ar chemistry, they were able to successfully pattern sub-100-nm features in the GST. Joseph *et al.*[154] developed etch processes to pattern small features in N-doped GST films deposited on and capped with TiN. After using trimming processes to shrink the CD of 248 nm photolithography-patterned resist down to 100 nm, ICP RIE in an Ar/Cl$_2$/CHF$_3$ gas mixture was used to first etch the TiN and then the N-GST. The authors pointed to the importance of carefully choosing the amount of Cl$_2$ so as to minimize undercut and obtain anisotropic profiles.

In the same work,[154] the authors also reported the presence of a uniform 10 nm thick damaged layer on all plasma-exposed surfaces of the N-GST, suggesting that the material modification was chemically driven and not enhanced by ion bombardment. TEM-electron-energy-loss spectroscopy and energy dispersive x-ray spectrometry (EDS) analyses were used to show that the damaged layer had depletion of Sb and/or Te along with potential oxidation. Depth profile XPS analysis on blanket films subjected to etch (and etch fol-

lowed by a strip process in an oxygen plasma) revealed selective loss of N and metallic Sb, and increased amount of Ge oxide. Finally, they also showed that the damaged layer could be removed selectively to the undamaged N-GST.

3. Chemical-mechanical polishing

CMP is now a ubiquitous process in CMOS technology, especially in the back-end-of-the-line (BEOL) flow. It enables the formation of inlaid, or "damascene," structures that are used in copper interconnect fabrication.[190] Such structures are created by first forming a hole (or trench) in a dielectric, then filling the hole, usually with a metal, followed by CMP to planarize the metal with the surrounding dielectric. Thus CMP allows patterning of the metal without needing an explicit metal etching step. There have been a number of publications on the CMP of phase change materials such as GST.

Liu *et al.*[191] made arrays of damascene PCM cells using CMP process. Scanning electron microscopy (SEM) and EDS were used to verify that the GST was properly filled in the contact holes and that the material composition of the GST was not changed by the CMP process. Slow current sweeps on fabricated cells showed successful switching from the high to low-resistance state.

Zhong *et al.*[192] used CMP to fabricate damascene-type PCM cells. An alkaline slurry was used to polish GST deposited into 300 nm wide vias. AFM measurements confirmed that a very smooth (0.8 nm rms roughness) surface was obtained as a result of the CMP. A thin TiN layer was placed between the top surface of the GST and the top electrode. The programming endurance of such devices was an order of magnitude better than devices built without CMP. In addition, set and reset state resistance fluctuations along the cycling sequence were greatly reduced in the CMP-processed devices. The authors attributed this to the smooth surface and good quality TiN/GST interface leading to lower and more uniform contact resistance. They also suggested that better confinement of the Joule heating and enhanced heat flux in the damascene structure were responsible for more uniform temperature distribution across the GST, and thus a more homogeneous material composition distribution as the cycling progressed.

The effect of adding oxidants such as H$_2$O$_2$ to the GST CMP slurry was studied by Zhong *et al.*[193] This study showed that while CMP of GST in a pure acidic silica slurry resulted in a rough surface with microscratches, the addition of 2 wt % H$_2$O$_2$ allowed them to obtain a smooth surface that was free of damage. XPS analysis on an unpolished GST sample dipped into the oxidant for 10 min showed that the surface was oxidized. Although the component elements were oxidized by different amounts, i.e., Ge > Sb > Te, the CMP process was believed to be able to remove all of these oxides. The authors suggest that the GST CMP mechanism is similar to that of metals, i.e., oxidation followed by removal of this oxide due to friction with the abrasives in the slurry. The same group also compared "RIE" cleaning in an Ar plasma post-CMP of GST to conventional ultrasonic

cleaning.[194] An alkaline silica slurry was used to polish GST into submicron contact holes. AFM measurements showed that low surface roughness (0.64 nm rms) was obtained after the CMP and RIE cleaning. Slow *I-V* sweep measurements on fabricated devices also showed a significant reduction in the threshold switching voltage as a result of the RIE cleaning method.

Lee *et al.*[155] used CMP of GST in their demonstration of a scalable confined PCM cell concept. CVD GST deposited into high aspect-ratio 50 nm diameter contact holes was planarized by CMP, thus fully confining the phase change material in the contact. SEM images showed that the GST CMP was successful with no scratches, excessive dishing, or void formation in the GST. Electrical tests on these devices showed low reset currents and good cycling endurance.

4. Process-induced damage

A number of cell failures observed in fully integrated PCM chips have been attributed to process-related damage. Both the phase change material itself and its interfaces with the top and bottom electrodes are susceptible to degradation as a result of steps in the integration flow.

Lee *et al.*[195] introduced suitable interface cleaning processes in order to obtain good contact resistance in the TEC/GST/BEC current path and thus reduce the write current. They also noticed that edge damage in small GST cells could lead to increased initial cell resistance. Ahn *et al.*[159] used nitrogen-doped GST so as to increase the dynamic resistance and lower the writing current. However, they found that higher nitrogen doping increased the cell's contact resistance and broadened its distribution across cells. They suggested that this was due to instability between the BEC and GST caused by the nitrogen doping and exacerbated by interface defects and the subsequent thermal processing. In addition, they observed that smaller cells exhibited a wider set resistance distribution. This was attributed to the effects of contaminants and GST etch-related damage. By appropriate interface treatment, optimization of the GST etch process, minimization of process damage, and thermal budget reduction, they were able to obtain sharp set and reset resistance distributions.

As part of "product-level reliability verification" for 64 Mbyte PCM chips, Kim and Ahn[196] reported that the activation energy for loss of data retention (i.e., due to crystallization of amorphous volume) in fully processed PCM cells (2.1 eV) was lower than that measured on as-deposited films of the phase change material (2.46 eV). They suggested that the lower retention time in the fully processed cells could be attributed to higher nucleation probability resulting from processing damage on the GST or defects at the BEC/GST interface. In addition to developing an optimized etch chemistry for patterning the GST/TE stack, Oh *et al.*[43] used a line-type GST layout (as opposed to an island-type GST layout) thereby maximizing the size of the patterned GST region and thus reducing the effect of RIE damage on the GST switching volume.

There have been multiple reports on the development of confined PCM cell structures.[146,155] While the most striking advantage of such cells is the reduced reset current, the authors also point out that these cells are also more resilient to sidewall edge damage during BEOL processes since the volume of material undergoing phase change is farther away from the damaged edge regions of the cell.

In particular, in their "on-axis confined cell" structure, Cho *et al.*[146] confirmed that there was no degradation, such as an increase in set resistance, as the cell size was reduced. Song *et al.*[143] pointed to the problem of increased set resistance in devices due to oxygen penetration during BEOL processing steps that degraded the BE/GST interface through oxidation. In order to address this issue, they developed encapsulation processes for the GST cell. Operating under the constraints that the oxygen-blocking encapsulation layers should be scalable and should not chemically react with GST films or degrade their electrical properties during deposition, they investigated a variety of encapsulation schemes. They found that a "double-capping" technique worked best—resulting in reduced set resistance while still being able to achieve low reset currents. Oh *et al.*[43] also used encapsulation in their 90 nm diode-accessed high-density PCM demonstration.

While studying set and reset resistance distributions in 4 Mbyte level PCM cell arrays, Mantegazza *et al.*[197] found two kinds of anomalous cells that contributed to low-resistance tails in the reset distributions. One of those anomalous cell types showed saturation of reset resistance at lower-than-desired values even if higher reset currents were applied. These cells also had a distinctly different (higher) slope in the plot of threshold voltage (V_t) versus cell resistance. The authors found that these cells could be modeled as having a conducting path in parallel with the amorphous plug. They pointed to contamination or impurities in the active PCM volume as likely causes. Solving this problem required a modified GST integration scheme that involved reducing the number of contamination sources, improving cell encapsulation, and cleaning the heater/PCM interface.

Modified PCM cell structures sometimes lead to distinct problems with GST integration and related failure modes. A case in point is the ring-type bottom electrode cell, introduced by Ahn *et al.*[140] in order to reduce the GST/bottom electrode contact area as well as its dependence on the patterned contact diameter. Subsequent researchers[143,144] pointed out that a failure mechanism in these ring bottom electrode cells involved the formation of a recess in the core dielectric (that is surrounded by the ring electrode) due to the wet cleaning solution used in the CMP process. This led to increased effective GST/BE contact area and resulted in reduced reset resistance. The authors solved this problem by optimizing the CMP process and using a more resilient core dielectric.

In summary, addressing yield loss in fully integrated PCM chips requires that adequate care be taken to minimize process-related damage to the phase change material and its various interfaces. While the specific details of optimized
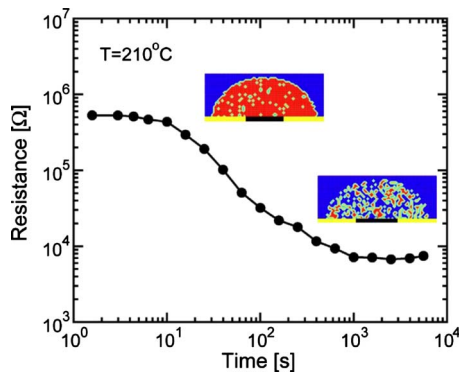
FIG. 26. (Color online) Accelerated failure of a μ-trench PCM cell, showing a decrease in reset resistance as a function of time at 210 °C. Reprinted with permission from U. Russo, IEEE Trans. Electron Devices, 53, 3032, 2006. © 2006, IEEE.

integration flows are not easily gleaned from the published literature, some of techniques that are usually mentioned include contamination reduction, GST/BEC interface optimization, minimizing the phase change material sidewall etch damage and/or keeping such damage as far away as practical from the active switching volume, and the use of encapsulating layers as protection from oxidation during BEOL processing.

IV. PCM RELIABILITY

A. Retention

In order for a nonvolatile memory candidate to be considered a viable next-generation memory option, it must demonstrate long-term retention of stored data. The typical criterion is 10 years (or 100 000 h) at 85 °C, with fewer than 1 ppb (part per 10^9) retention failures at the array level over this entire lifetime, independent of the previous cycling history of the memory array.[168,198] Since the crystalline set state is a stable low-resistance state, it is the stability of the quenched high-resistance reset phase that dominates retention issues. Not surprisingly, the stability of this phase has been investigated widely. As discussed earlier, the amorphous phase suffers from two independent resistance-altering processes: resistance drift and spontaneous crystallization. The drift process is a steady increase in the resistivity of the amorphous phase, related to structural rearrangement of the amorphous chalcogenide and the dynamics of intrinsic traps. Since this process increases the on/off ratio, it does not cause any data loss for binary PCM devices.

On the other hand, thermally activated crystallization of the amorphous material eventually leads to significant reduction in the resistance of the active layer, causing eventual retention failures for both binary and MLC storage. Data retention measurements typically involve monitoring the resistance of the cell that has been set in the reset state, as shown in Fig. 26. When the resistance of the cell falls below a threshold resistance (between the set and reset resistance values), the cell is said to have suffered a retention failure. Similar to accelerated tests used on other nonvolatile

memory candidates, PCM retention measurements are done at higher temperatures to speed up this crystallization process and the subsequent resistance change. Using measurements done at a number of high temperatures (typically near 160 °C) and a reasonable activation model, the retention properties of PCM at 85 °C can be predicted fairly well. Of course, the validity of the activation model and of the extrapolation from higher temperature measurements is also strongly dependent on the nucleation and grain-growth properties of the specific phase change material being used in the memory cell.

Early work[40] used the activation energy of blanket GST (3.5 eV) to estimate that a 10 year lifetime at 120 °C should be possible for PCM cells. Using data on mushroom PCM cells, Pirovano et al.[147] showed a retention activation energy of 2.6 eV, adequate for more than 300 years of data retention at 85 °C.

Later Redaelli and co-workers[199–201] proposed a percolation model for retention on 180 nm μ-trench PCM cells and analyzed this model using a temperature-dependent percolation effect. This is the most widely accepted model for PCM retention. By showing that repeated retention measurements on the same device resulted in widely varying retention times, they demonstrated the stochastic component to GST crystallization discussed earlier. They proposed that this variation has its origin in the random spatial configuration of the as-nucleated grains in the amorphous region. As nucleation proceeds with time, the cell resistance decreases significantly when a percolation path finally appears through the amorphous layer. Since the crystalline state is so much more conductive, the occurrence of even a partial path through the amorphous plug can strongly reduce the overall device resistance.

Due to these percolation effects, the measured retention times of an ensemble of cells tend to obey the Weibull statistics, where the cumulative distribution of retention times represent a line of constant slope β on a Weibull plot (which plots the logarithm of the fraction of failed cells against the logarithm of elapsed time). It was also experimentally observed that β increases (e.g., distributions become tighter) at lower measurement temperatures. This temperature dependency was explained using a Monte Carlo model that randomly generated crystalline nuclei and let them grow in the thin amorphous region (note that heterogeneous nucleation was neglected). By comparing this model to experimental data, they also showed that the effective grain size r at most temperatures is higher than the as-nucleated grain radius, with the difference being attributed to grain growth. By assuming a temperature-activated Arrhenius model for failure times, retention exceeding 10 years at 105 °C was predicted using this model. Below 120 °C, the model postulated that grain growth is negligible, so that the size r of each grain is no larger than the nucleation-limited fcc GST-crystal monomer radius. Thus nucleation becomes solely responsible for retention failures, and this causes the Weibull slope β to increase significantly. This also implies that the distributions of retention times could become very tight in the temperature

regime of interest for retention (around 85 °C), which led them to conclude that ppb retention failures should still exceed 10 years at 103 °C. Modeling and simulations done on different PCM cell structures also showed that the retention properties are identical when the effective amorphous layer thickness is the same.

The previous papers also assumed that failure times have an Arrhenius temperature dependency and that defect mechanisms played no role in tail-bit (ppb) failures. Russo *et al.*[202] showed that a pure Arrhenius extrapolation[199,201] would be pessimistic in its estimation of retention and concluded that GST-based PCM cells should show data retention exceeding 10 years at 118 °C (instead of 10 years at 105 °C shown earlier). This is largely because the energy barrier for nucleation is larger at higher temperatures (as the driving force for crystallization reduces), causing the increase in nucleation rate with temperature to be less than Arrhenian.

Gleixner *et al.*[168] studied large-sized PCM arrays fabricated with 180 and 90 nm technology in order to study new defect failure modes and retention loss at the ppb level. They noticed that even at the lowest times used to measure retention (at elevated temperatures), a small fraction of the bits (<ppm) had already failed, indicating that the time to failure for the first cell is not accurately predicted by the failure curve. It was also shown that these tail bits show similar initial resistances to the nominal bits and similar activation energy of the time to failure (2.4 eV). Furthermore these tail bits show a Weibull distribution consistent with a weak-link failure mechanism, in which rare combinations of closely set nuclei can lead to rapid retention loss. Although the same fraction of bits fail (to within 10%) during each test, it was different bits that occupy this tail (and thus fail) each time. This indicates that manufacturing defects were not responsible for these tail-bit failures. It was therefore postulated that the most likely cause of this failure comes from a scenario where the nucleation sites are arranged such that when thermal energy is applied, very little growth is required before a quick resistance decrease can be observed. Data also showed that cycling has no impact on retention—if anything a slight improvement was observed. Gleixner *et al.*[168] also showed that optimization of the process and the write scheme could be used to significantly suppress these tail bits, although the exact nature of this optimization was not discussed in detail.

Work done by Shih *et al.*,[169] although with slightly worse data retention, concluded that a large fraction of the bits fail due to grain growth from the amorphous/crystalline boundary. While this retention loss mechanism has not been observed before (presumably due to a reduction in grain-growth velocity at lower temperatures), these differences could be attributed to differences in nucleation and growth properties of the materials used in different experiments. Clearly, reducing the crystal growth component at retention temperatures would be an important step in improving the retention times.

Finally, Redaelli *et al.*[203] showed that the value of the threshold resistance that is chosen to define the retention fail-

ure criterion directly impacts the calculation of the activation energy (E_A) extracted for failure times. This might explain the wide variability in previously reported E_A values, ranging from 1.9 to 3.5 eV. They concluded that reliable extraction of the activation energy of crystallization can be obtained by using a critical threshold resistance that is close to the crystalline resistance, or by performing all resistance measurements at room temperature.

Finally, it should be mentioned that while much of the work described here deals with GST, the exact composition and doping of the GST that have been explored by different groups are very likely different. This itself could alter the value of the measured activation energy. A number of groups have also investigated alternate PCM materials and doping as a way to increase the activation energy for crystallization and to increase retention margins. Matsuzaki *et al.*[204] showed that oxygen doping of GST films results in a much higher activation energy due to a smaller grain size—which leads to improved retention. Kim and Ahn[196] also showed that N-doped GST single-cell PCM devices show retention lifetimes exceeding 10 years at 85 °C. They also concluded that the activation energy is lower for devices compared to blanket films because of process-induced damage and defects at the interface between the bottom electrode and the GST—effects that have to be minimized in order to maximize retention. Morikawa *et al.*[205] also explored In–Ge–Te as a phase change material and have shown that its retention properties are significantly better than GST, ranging from 10 years at 122–156 °C depending on the indium fraction. Other materials explored for improved PCM retention include doped-GeSb,[42] Si-doped Sb_2Te_3,[206] and Si_xSb_{1-x}.[207,208] However, only very basic materials studies have been carried out for most of these new materials, and detailed array data such as the study of tail bits are lacking. Further materials development will be key for improved retention as the size of the amorphous plug shrinks with scaling, as well as for improved tail-bit retention, MLC capability, and proximity disturb performance.

B. Cross-talk

From the discussion in Sec. III A, it is clear that during the reset operation the peak temperature within a PCM device exceeds the melting point of the phase change material. In fact, since the extent of the amorphous plug quenched from the molten state must be larger than the critical dimension of the limiting aperture, the material at the edge of this aperture is just over the melting point, and the peak temperature in the center of the cell is well over the melting temperature. As an example, $Ge_2Sb_2Te_5$ melts at ~630 °C (Ref. 38) and other phase change materials have similar melting temperatures.[209] However, from Sec. IV A, any significant exposure of this amorphous plug (once it has been formed) to temperatures exceeding 150 °C or so will lead to the recrystallization of enough crystalline nuclei within the plug to induce data loss.

It is thus not surprising that upon learning these two facts, many engineers introduced to PCM immediately ask about
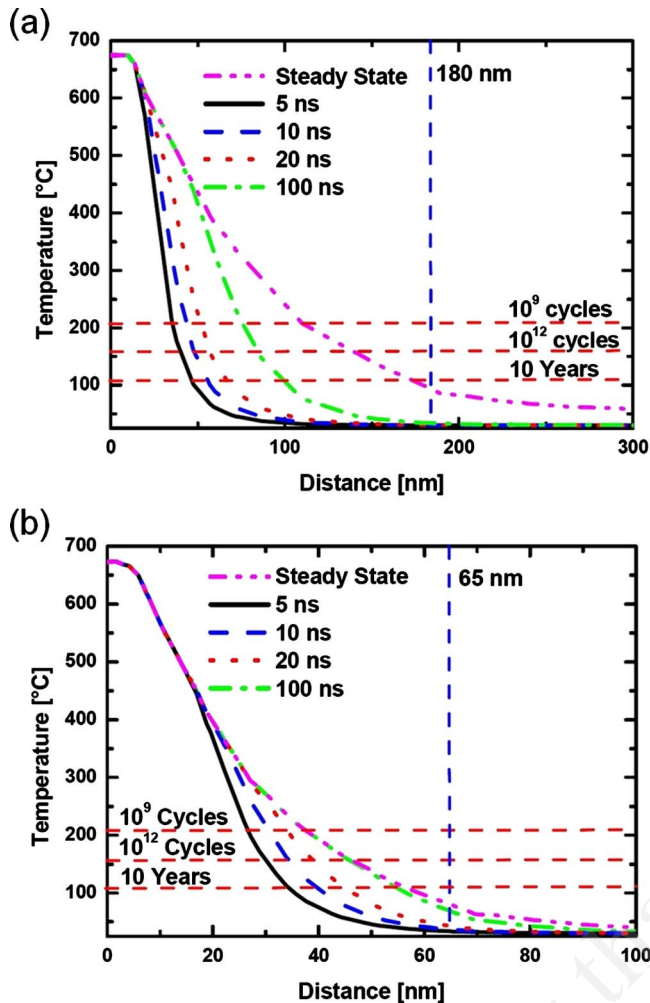
FIG. 27. (Color online) Simulated temperature profiles for PCM devices (microtrench-type devices) for the 180 and 65 nm technology nodes. Note that while the transient temperatures become close to the steady-state temperature, the expected temperature rise at the neighboring device remains much lower than 100 °C. Reprinted with permission from A. Pirovano, Tech. Dig. - Int. Electron Devices Meet., 2003, 29.6.1. © 2003, IEEE.

the thermal cross-talk between cells. In particular, the worst-case scenario is the effect on an amorphous plug encoding the reset state when one or more of the immediately neighboring cells is programmed through multiple reset-set cycles. What is truly surprising is that for most researchers working in the PCM field, thermal cross-talk or "proximity disturb" is considered to be a second-order rather than first-order problem. In fact, the small amount of empirical data available on proximity is focused primarily on the absence of any cross-talk effects.[147,210]

Pirovano *et al.*[41] showed with careful thermal simulations that at least out to the 65 nm node, the thermal cross-talk between cells should remain low enough that 10 year lifetime will not be significantly reduced by write disturbs from neighboring cells. Figure 27 shows their simulated temperature profiles for the 180 and 65 nm nodes, assuming microtrench-type PCM cells. The expected temperature rise at the position of the neighboring cell remains well below 100 °C. Russo *et al.*[129] did a similar study extending down

to the 16 nm node. They found that isotropically scaled devices, where all cell dimensions scale with the technology node, can be expected to experience no thermal cross-talk problems. However, because their cell designs had been optimized to emphasize low reset current without any consideration to efficiency, reset power, or proximity effects,[128] the hottest point in their cells is deep within the highly resistive heater electrode (both for the mushroom cell, which they refer to as a "lance" cell, and to a lesser extent for the microtrench cell). As a result, the temperature at the neighboring cell grows rapidly if only the lateral dimensions are scaled (nonisotropic scaling). Despite these highly power-inefficient design points, though, Russo *et al.*[129] found it straightforward to avoid thermal cross-talk simply by choosing a mixed scaling approach that decreased both the lateral and thickness dimensions.

One observation made by Russo *et al.*[129] is that as the spacing between devices drops, the time dimension becomes much less effective as an avenue for avoiding proximity issues. As the size of the heated volume decreases, the thermal time-constant τ_{th} decreases, but the characteristic thermal diffusion length L_{th} drops more slowly (scaling as $\sqrt{\tau_{th}}$). Thus as the technology node scales down, the neighboring device eventually moves inside the characteristic thermal diffusion length. This effect is noticeable even in the data of Pirovano *et al.*—in Fig. 27(a), one can avoid experiencing the maximum steady-state temperature change by simply using a very short reset pulse, while in Fig. 27(b), the difference between the temperature rise due to a short reset pulse and the steady-state temperature rise has become much smaller.

To a certain extent, knowledge of the temperature to which neighboring cells will be heated during reset, can be combined with retention measurements which reveal a failure time [curves of t_X versus $1/k_BT$ (Refs. 199, 201, and 202)] in order to estimate the effect of proximity disturb. The additive effects of retention and proximity both need to be considered to account for data integrity loss in high-density PCM-based cross-point memory arrays. In particular, the presence of even modest proximity effects can push the temperature for which acceptable retention is required significantly higher than the actual maximum average operating temperature of the memory chip.

However, it is important to keep in mind that the temperature distribution during a retention failure measurement is completely homogeneous across the PCM cell. In contrast, in proximity disturb situations, the temperatures are strong functions both of position within the neighboring PCM cell and of time due to the fast thermal transients during reset pulses and any slow ramp downs used at the end of long set pulses. This is particularly relevant when considering that early-to-fail retention problems are attributed to the unpleasantly rapid generation of a percolation path that connects a statistically rare chain of crystalline nuclei. For proximity disturb, this unlucky combination of closely set nuclei could only lead to trouble if it were also located at the extreme edge of the amorphous plug, where the maximum temperature excursion generated by the frequently cycled neighbor
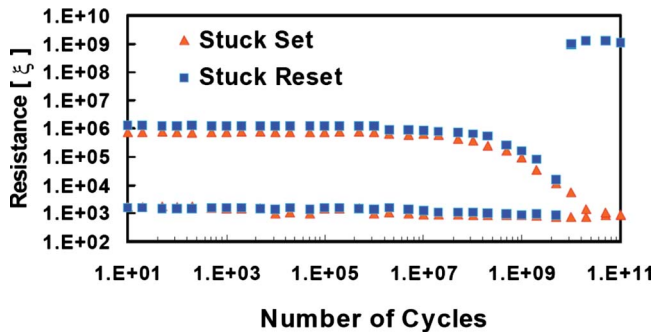
FIG. 28. (Color online) Set and reset resistances during cycling, illustrating the differences between failure by stuck set and by stuck reset. Reprinted with permission from B. Gleixner, NVSMW 2007. © 2007, IEEE.
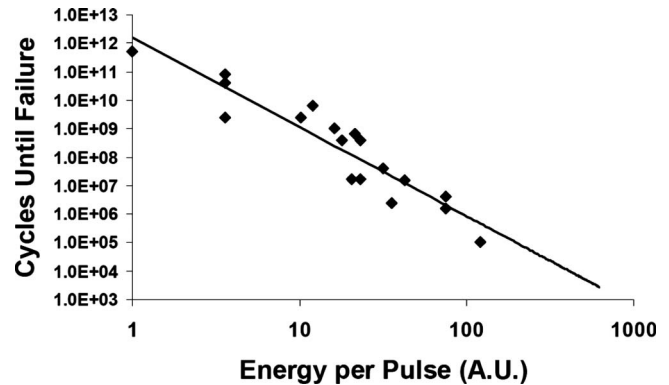

FIG. 29. Cycling endurance as a function of pulse energy, showing that device endurance drops rapidly with prolonged exposure to high temperatures. Reprinted with permission from S. Lai, Tech. Dig. - Int. Electron Devices Meet., 2003, 10.1.1. © 2003, IEEE.

could then rapidly drive retention failure. This additional criterion will further suppress the likelihood of such an initial failure, but it also complicates the incorporation of such early-to-fail retention data together with modeled temperature distributions (within the "neighboring" PCM cell) for the accurate prediction of proximity disturb.

C. Endurance

Cycling endurance has long been one of the strengths of PCM, especially in comparison to established Flash technologies, where SILC frequently limits device endurance to $10^4 - 10^5$ program-erase cycles. The demonstration, as early as 2001, of 10^{12} set-reset cycles in PCM devices without any significant degradation of resistance contrast, as shown in Fig. 23,[141] was almost certainly a significant factor in the surge of interest in PCM technology that followed. Of course, while it is telling that single devices could be operated reliably for so many cycles, the more critical question is what happens to the worst-case device in a large array. Subsequent large-scale PCM integration experiments have tended to show endurance numbers in the range of $10^8 - 10^{10}$ cycles (Refs. 196 and 198)—still easily exceeding the endurance of Flash, but coming somewhat short of what would be necessary for DRAM replacement without wear leveling [Eq. (1)].

Two different failure modes have been observed to occur after cycling, termed "stuck-reset" and "stuck-set" failures,[196,198] as illustrated by Fig. 28. In a stuck-reset failure, the device resistance suddenly and irretrievably spikes, entering a "blown-fuse" state that is much more resistive than the reset state. This sometimes occurs after some degradation in resistance contrast (as in Fig. 28), but can also suddenly occur, with no prior indication that failure is imminent. These failures are typically attributed to void formation or delamination that catastrophically severs the electrical path through the device, typically at a material interface such as the heater-to-GST contact in a mushroom cell.

In contrast, in a stuck-set failure, a gradual degradation of resistance contrast is typically observed, as if the cell were being slowly but inexorably altered by the set-reset cycling. The cell seems to change its characteristics so much that the original reset pulse is, at some later point in time, somehow

less effective at creating an amorphous plug in the device than before. Eventually, the reset step of the cycling fails to induce any change in the device resistance, and the device becomes "stuck" in the set state.

Typically, for a cell in this stuck-set condition, a larger amplitude reset pulse proves sufficient to reset the device and cycling can resume, although with a larger reset power. However, this continued operation inevitably hastens the onset of a stuck-reset failure. Figure 29 shows the strong correlation between pulse energy and device failure: Every order-of-magnitude increase in pulse energy implies three orders-of-magnitude lower endurance.[40] Unfortunately, it is not clear from Fig. 29 whether it is pulse amplitude or pulse duration that is critical, nor does the plot differentiate between stuck-set and stuck-reset failures.

Fortunately, Goux *et al.*[211] carefully studied endurance failure due to stuck set in phase change bridge cells. By measuring the resistance-versus-current curves at various points during cycling, they clearly demonstrated that stuck-set failure is due to a change in the reset condition (i.e., the pulse amplitude required for reset) that is induced by cycling. They also observed that while pulse amplitude had a fairly minor impact on device endurance, pulse duration had a strong effect. By using pulses ranging from 10 ns to 10 μs in length, they were able to show that the time spent melting their PCM material (in their case, Ge-doped SbTe) was the critical factor. Their endurance data suggest that endurance scales inversely with $t_m^{3/2}$, where t_m is the time spent melting during each reset pulse. The way to reach a very large number of set-reset cycles is thus to minimize the time spent melting during each reset pulse. This data also fit with observations that repeated cycling with only set pulses shows greatly extended endurance ($>10^{12}$ cycles) over reset-set cycling (10^{10}).[196] Together these results imply that the gradual cell degradation associated with stuck set is strongly correlated with the melting inherent in each reset operation.

A number of groups have been using techniques such as EDS, secondary ion mass spectrometry, and energy dispersive x-ray spectrometry to perform elemental analysis on failed cells.[53,212–223] Measurements on mushroom cells built
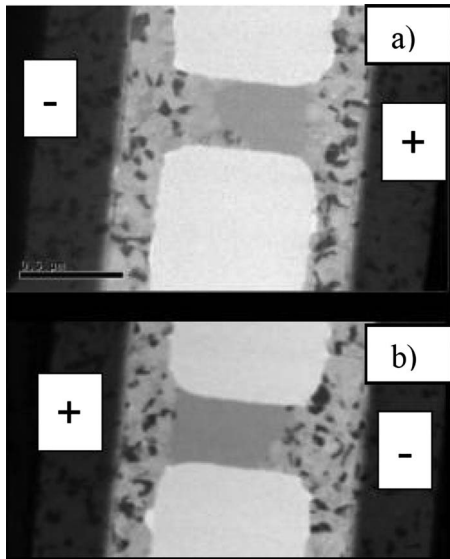
FIG. 30. Top-down TEM images of large phase change bridge devices (L =740 nm, M=300 nm bridges of 20 nm thick Ge-doped SbTe material), showing an ~100 nm polarity-dependent shift of the amorphous plug toward the anode (+). Reprinted with permission from D. Tio Castro *et al.*, Tech. Dig. - Int. Electron Devices Meet., 2007, 12.5. © 2007, IEEE.

from $Ge_2Sb_2Te_5$ material[212,213] tend to show agglomeration of antimony (Sb) at the bottom electrode at the expense of tellurium (Te). [The tendency of germanium (Ge) is not clear—Ref. 213 shows it clearly depleting from the repeatedly molten mushroom cap over the heater, while Ref. 212 indicates no motion]. Sarkar and Gleixner[224] used these results to explain why their $Ge_2Sb_2Te_5$ material actually improves slightly, in characteristics such as resistance contrast and reset current, over the first 10 000 set-reset cycles. They surmised that as the phase change material evolves in composition through cycling, the volume melted by each reset pulse increases (increasing reset resistance) while the inherent lower resistivity of the more Sb-rich GST material provides a lower set resistance.

Eventually, however, it would appear that such compositional changes driven by cycling will steadily decrease the dynamic resistance of the active region, shifting the required reset current to larger values. This leads to stuck-set failure if the reset pulse is not adaptively increased, or to stuck-reset failure if the reset pulse energy is increased to compensate.

D. Polarity issues

A number of recent measurements of phase segregation, in various types of phase change bridge devices, have forced a reinterpretation of the failure analysis results just described within the context of two previously unknown bias-polarity-dependent effects.

Tio Castro *et al.*[104] showed convincing top-down TEMs of bridge devices in the reset state (Fig. 30) that demonstrate that the amorphous plugs in their devices were shifted by the polarity of the applied bias, by as much as 100 nm. They attributed this effect to the thermoelectric Thomson effect, in which the overlap of temperature gradients with electrical

current can lead to additional heat generation or absorption. Because the hot spot in the center of a phase change device is surrounded by temperature gradients of opposite signs but the current flow is unidirectional, the Thomson effect acts to shift the centroid of the hot spot depending on the polarity of the applied voltage. Tio Castro *et al.*[104] estimated from their observations that the Thomson coefficient in their material might be in the range of -100 $\mu V/K$.

In those same experiments, it was observed that for intentionally asymmetric "dog-bone" bridge devices (somewhat like a high-aspect-ratio pore device on its side), there was a "bad" polarity of operation (large-area electrode negative and small-area electrode positive) for which subsequent set operations were unable to return the device to low resistance. In symmetric bridge devices, other researchers reported that the most reliable set operations can only be produced by alternating the polarity between set and reset, with little dependence on the absolute sign of the bias polarity.[225] Tellingly, these results were only observed for bridges fabricated from GST, and not for ultrathin bridges fabricated from doped-GeSb.[42]

Similarly, other researchers reported bias-dependent operation of pore devices where only the "good" choice of polarity (positive on large-area electrode) can be used to produce low set resistances, while the bad polarity is associated with "hard-to-set" operation.[226] In these experiments, only a narrow voltage window could be used for the set operation, which also required longer pulses and which never produced the same low set resistances as the good polarity.[226]

Fortunately, this good polarity corresponds exactly to the typical operation of integrated PCM devices, where the positive voltage is applied to top of the PCM device built over the underlying transistor.[138] In fact, operation of such integrated devices in the bad polarity is difficult to study since in that configuration, the gate-to-source voltage changes dynamically during each pulse as the PCM device resistance changes. Thus it is not surprising that such effects have not been widely reported for PCM devices integrated together with access transistors.

The Thomson effect would seem to be inadequate by itself to explain all of these bias-polarity effects. However, a few groups have been performing bias-dependent failure analysis experiments on various types of phase change bridge devices. Early versions of these experiments were affected by lingering uncertainties related to the role of metallic electrodes at the failure point,[219] and to the difficulty in understanding material desegregation during long-term cycling by studying the aftermath of a single-pulse blown-fuse failure.[53,219,220] However, the most recent experiments have investigated the cycling failure of tapered bridge structures located far from any metal electrodes,[221] and the controlled fast melting of large symmetric bridge devices.[222,223]

The results of these experiments, together with the earlier failure analysis data on mushroom devices,[212,213] sketch out a convincingly consistent story for $Ge_2Sb_2Te_5$ devices: Te moves toward the positive electrode (anode), while Sb moves toward the negative electrode

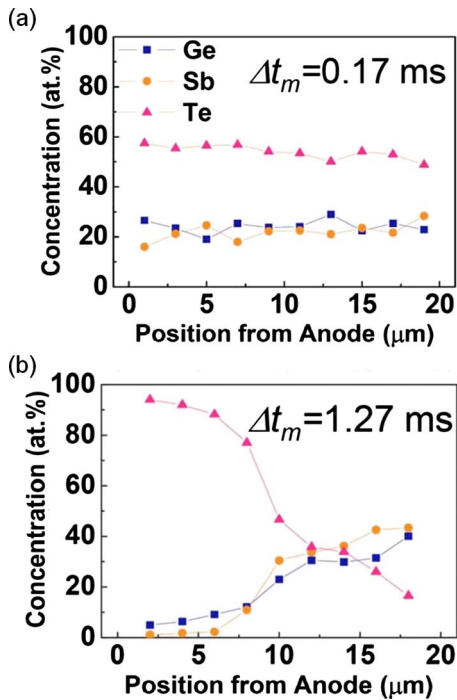FIG. 31. (Color online) WDS profiles of elemental concentration (Te, Sb, and Ge) along the length of a 20 μm long $Ge_2Sb_2Te_5$ bridge at (a) 0.17 ms and (b) 1.27 ms after melting was initiated by a voltage pulse, showing rapid desegregation of elements in the molten state. Reprinted with permission from T. Y. Yang *et al.*, Appl. Phys. Lett., 95, 032104, 2009. © 2009, American Institute of Physics.
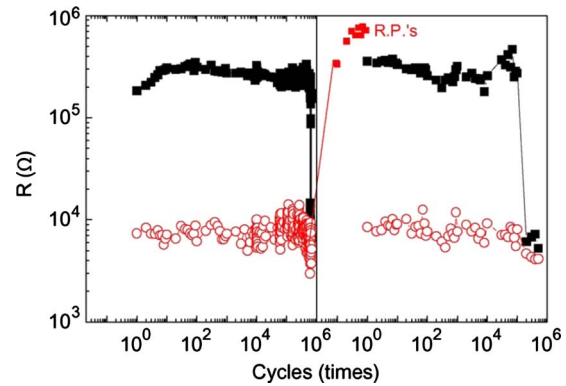

FIG. 32. (Color online) Cycling of a pore-PCM GST device, showing a stuck-set failure after $\sim 5 \times 10^5$ cycles, followed by ten pulses of reverse polarity (and of slightly higher magnitude), which were proven sufficient to allow cycling to continue for another 10^5 set-reset cycles. Reprinted with permission from S. Lee *et al.*, EPCOS 2008. © 2009, IEEE; S. Lee *et al.*, IEEE Electron Device Lett., 30, 448, 2009. © 2009, IEEE.

(cathode).[53,212,213,219–223] This motion is attributed to the higher electronegativity (5.49 eV) of Te compared to Ge and Sb (4.6 and 4.85 eV).[223] Most of the data seem to indicate that Ge moves together with the Sb toward the cathode, but as mentioned earlier, there are some data which indicate otherwise. In connecting these interpretations of bridge and mushroom devices, we assume that the mushroom cycling was performed in the good polarity direction, with positive voltage on the large-area top electrode.

One particularly interesting study was performed by Yang *et al.*[223] on very large symmetric devices (20 μm long by 2 μm wide bridges of 300 nm thick $Ge_2Sb_2Te_5$), where both high-amplitude millisecond-long pulses and day-long exposure to low-amplitude 10 MHz pulsed dc were used to explore elemental segregation through wavelength dispersive spectroscopy (WDS). They were able to show that the material segregation is very rapid in the molten state, observing nearly complete desegregation along a 10 μm length of bridge after a 1.5 ms pulse, as shown in Fig. 31. This works out to an effective diffusion coefficient of $(1-2) \times 10^{-5}$ cm^2/s,[223] which roughly corresponds to a field- or current-driven migration of 1 nm/ns. In contrast, their long-term measurements of bias-induced elemental desegregation through the crystalline state seemed to suggest diffusion coefficients nine orders of magnitude lower,[223] implying that the drift of elements through that same 1 nm would take 1 s.

While it is not yet clear how these observed effects (the Thomson effect and polarity-dependent elemental segrega-

tion via electromigration) can be combined into a unified theory that quantitatively explains PCM polarity and cycling endurance, it is already quite clear that such bias-polarity effects and cycling endurance are intimately related. It has already been independently observed that switching from one cycling polarity to the other can be used to continue cycling of bridges after a stuck-set failure.[211] In fact, as shown in Fig. 32, even as few as ten aggressive pulses applied in the opposite polarity direction can allow cycling to not only resume after a stuck-set failure, but to continue for another 10^5 cycles with the original reset pulse conditions.[227,228]

Continued understanding of what differentiates the good and bad polarity should allow researchers to continue to improve PCM cycling endurance, through a combination of creative use of the bad polarity,[211,227,228] improved cell design, and new materials that show greater resistance to elemental segregation. As with Sec. III D, most of the detailed data available are limited to GST because of its ubiquity. However, it has been observed that the phase change material GeSb phase segregates quite readily,[229] implying that simply reducing the number of atomic species involved is not necessarily the best approach.

V. FUTURE OF PCM

A. MLC

Judging from the recent history of the aggressive nonvolatile memory market currently dominated by Flash, it is clear that all available directions for improving effective density (e.g., the average number of information bits that can be stored per unit area) will be exploited. One direction, both promising and challenging, is that of the so-called MLC technology, which exploits the intrinsic capability of a memory cell to store analog data in order to encode more than 1 bit of digital data per cell. The feasibility of MLC for PCM has already been shown,[60,61,161,210] including the demonstration of programming into both 4 and 16 sharply dis-
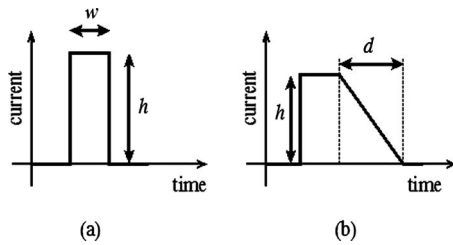
FIG. 33. Two example families of electrical signals that can be used for MLC programming: (a) rectangular pulses and (b) variable slope pulses.

tinct analog levels, corresponding to 2 and 4 bits/cell, respectively. Such intermediate resistance levels are obtained by properly modulating the electrical signals used to program the PCM element. In Fig. 33, two examples of such signals are shown: (a) rectangular current pulses of different height h and width w (Refs. 61 and 161) and (b) variable slope pulses, with different durations d of the trailing edge of a trapezoidal pulse.[60] By controlling these parameters carefully, one can control the analog resistance of the PCM element and thus enable MLC operation.

The degree of success of such a MLC writing scheme can be characterized by the resistance distributions over a large ensemble of PCM devices. Figure 34 illustrates this concept: each of the four levels labeled as "00," "01," "10," and "11" is associated with a resistance distribution. In a perfect world, these distributions would be delta functions, simplifying the classification process into a straightforward thresholding operation. If the distributions overlap, however, then there is a nonzero probability of level misdetection at the receiver, resulting in the retrieval of erroneous data. The use of the logarithm of the resistance is expedient to obtain more uniform shapes of the distributions across all levels. However, it should be pointed out that since these levels are classified using read current, the optimal configuration may not necessarily call for spacings between levels that are uniform in either resistance or log(resistance).

There are several factors that can limit the number of effective levels, which can be reliably stored in a PCM cell. Among them are as follows:

- the intrinsic randomness associated with each write attempt, or write noise;
- Resistance drift, which we will refer to as *short-term drift*;
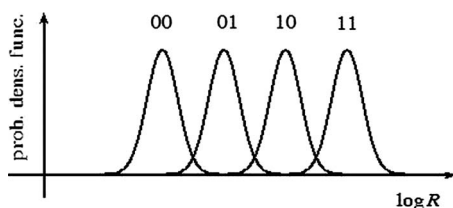

FIG. 34. Example distribution of the logarithm of the resistance for each of four possible stored levels, implementing 2 bit MLC. The distributions shown here would suggest a non-negligible probability of classification error.
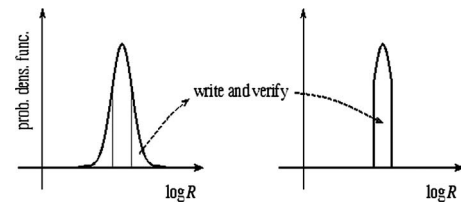

FIG. 35. Distribution tightening by means of a write-and-verify procedure.

- array variability, which includes any variability during the lifetime of the PCM array;
- crystallization of the amorphous phase, which we will refer to as *long-term drift*.

Some of these factors, such as short and long-term drift, represent a fundamental limitation to the *storage capacity*—the maximum number of bits that can be stored in the average PCM cell. As such, these factors cannot be overcome but only mitigated, as we will discuss later.

In contrast, factors such as write noise and, to a certain extent, array variability, do not directly limit the storage capacity but instead make it harder to achieve a given storage capacity. In order to deal with these kinds of limiting factors, resistance distribution tightening techniques have been developed based on *write-and-verify* procedures. These iterative techniques consist of applying programming pulses and verifying that a specified precision criterion is met, along the lines of what is currently done in Flash memories.[230] These methods have been used to demonstrate 16 level PCM,[60] 4 level PCM,[60,61,161,210] and to tighten the distribution of set state resistances for binary PCM.[167]

The effect of a write-and-verify technique is to reshape the conditional probability density function, as shown in Fig. 35. This can be obtained by successively refining[231] the write procedure until the verify step finds the resistance value within the desired range around the nominal resistance target. When properly used, a write-and-verify algorithm produces tighter resistance distributions (compare Fig. 24 to Fig. 25), and therefore allows the packing of more MLC levels into the same resistance range. This increase in the number of levels obtained with write-and-verify reflects an actual increase in the information-theoretic storage capacity. There exists an intricate tradeoff between storage capacity and the average number of write-and-verify iterations. In particular, Refs. 228 and 229 show that for a simple cell model affected by write noise, the achievable storage capacity tends to increase logarithmically with the number of write-and-verify iterations.[232,233] This logarithmic increase is expected to hold even for more realistic cell models at a sufficiently large number of write iterations.

Among the factors representing a fundamental limitation to the storage capacity, resistance drift plays an important role. Short-term drift manifests itself as a slow but steady increase in the resistivity of the amorphous material (see Ref. 51 and references therein).

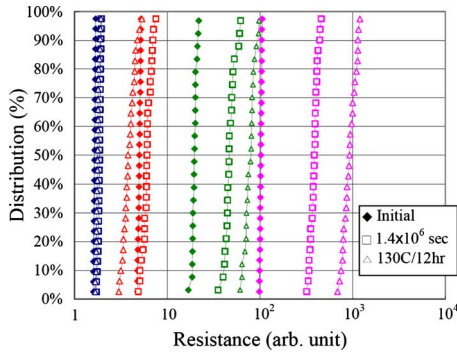The resistance drift has been shown to follow a power law,

FIG. 36. (Color online) Distributions of four resistance levels immediately after programming after 400 h at room temperature and after an additional thermal annealing at 130 °C for 12 h. Reprinted with permission from D.-H. Kang *et al.*, Tech. Dig. VLSI Symp., 2008, 10. © 2008, IEEE.

$$R(t) = R_0 \left(\frac{t}{t_0} \right)^{\nu}, \tag{8}$$

where $R(t)$ denotes the resistance at time t, R_0 denotes the initial resistance at time t_0, and ν is a drift coefficient.[90] Typical values for ν for thin amorphous GST layers are on the order of 0.05–0.1.[51] This phenomenon has been explained in terms of structural relaxation in the amorphous material influencing a Poole or Poole–Frenkel conduction,[49,90] and in terms of kinetics of electrically active defects in the amorphous GST material.[91] The drift process is fairly predictable in the case of thin films of amorphous GST, which would suggest that a few cells of known state in a block that evolved in time together might serve to identify the needed shifts in threshold bias. However, short-term drift appears to be a random process, which can be expected to vary from cell to cell.[89,91,210,234] By introducing yet another source of unpredictability, this short-term drift reduces the effective storage capacity of PCM. The phenomenon can be perceived as a broadening of the resistance distributions over time. Figure 36 compares the cumulative distributions for four resistance levels measured immediately after programming to the distributions after programmed cells have been drifting, both at room temperature and then at elevated temperature.[210]

A number of techniques have been proposed for coping with drift. These include changing the write target resistances to take into account the expected broadening of the resistance distributions due to drift[210] and compensation techniques at read time, where pulses are used upon readout to return the device to its initial as-written resistance (presumably without accidentally reprogramming the cell).[235] We remark that these topics are the subject of current active research in the PCM research community.

B. Role of coding

Although little or no published literature yet exists on coding techniques designed expressly for PCM, the success that these techniques have had in established memory and storage technologies would imply that coding will be exten-

sively used in PCM. Moreover, careful adaptation of existing codes as well as development of new coding technologies suited to the physical characteristics of PCM could prove essential to unlocking much of its inherent potential.

Although there are many different types of coding, the most prevalent technique is error correcting coding (ECC), which allows for the detection and correction of bit or symbol errors. Other coding techniques, called modulation codes, are designed to ensure that the patterns stored in memory are adapted to particular characteristics of the physical medium. Here we discuss how both of these approaches might be incorporated into PCM, with emphasis on ECC.

ECC technology is now a standard feature in most storage and high-end computing systems, finding applications in caches and buffers built using SRAM, main memory which generally uses DRAM, and hard disks, solid-state drives, and tape. The sophistication of the coding technology employed at each layer is often inversely related to the proximity of a memory technology to the computing element. Much of this depends on the speed with which such coding can be implemented—close to the processor, a few extra nanoseconds spent on decoding may represent a significant (and unacceptable) delay, while far from the processor those same nanoseconds represent a tiny rounding error. Processor caches tend to utilize regular or extended binary Hamming codes,[236] which are arguably among the simplest codes that can be found in a pervasive manner. In contrast, main memory uses symbol-based codes such as Reed–Solomon codes,[237] while disks use extremely sophisticated constructions involving error correcting codes, modulation codes, and advanced signal-detection and signal processing technology. Since PCM holds promise both as a storage and as a memory device, it is reasonable to expect that it will draw coding and signal processing ideas from all of these technologies.

To begin with a relatively simple example, consider a single-bit PCM cell. Here, a zero or a one correspond to a cell being set or reset, with the dramatic difference in resistance that accompanies these two states. Detection of a zero or one can be accomplished by a simple "hard decision" based on a resistance threshold in between these two resistance states based on read current. In PCM, one of the complications is that the measured resistance may change over time (short-term drift, see Sec. V A) or with the instantaneous temperature of the cell, requiring a threshold that can be shifted with time and temperature. However, since the short-term drift is associated with the amorphous phase, if the set state resistance is sufficiently dominated by the crystalline resistivity, then its resistance can be considered relatively independent of elapsed time. (Note that temperature-dependent thresholds can be produced by circuits designed around inherent temperature-dependent changes in the silicon underlying the PCM devices.) Thus for binary storage, simple thresholding may prove adequate. Another solution is to have "pilot cells" that are known to be in the reset and/or set states. These pilot cells are programmed and read at the same time as the data-bearing cells in the same block, and thus can convey the information necessary to construct a

suitable threshold for discriminating between a zero and a one. This is where device-to-device variability in this short-term drift proves to be the real issue.

Errors in data that have been retrieved can have their origin in either a failure of the cell to switch to the desired state, an erroneous reading caused by noise, quantization, or resistance fluctuations, or in the state of a cell switching over time. As discussed in Sec. IV C, the most likely such state-change event is the gradual transition from reset to set caused by crystallization of the phase change material. Although it is difficult to place a definite bound on the number of errors that these problem sources will cause, it is likely a safe statement to say that standard coding and decoding techniques, such as Reed–Solomon codes based on the Bose, Ray-Chaudhuri, and Hocquenghem (BCH) code family[238–240] will be adequate to address these problems for a good number of applications.

Having stated this, it is entirely possible that more sophisticated error control coding techniques may find their way even in single-bit PCM. These techniques would allow a greater number of errors to be corrected for a given number of redundant ("check") bits. The benefits of this include extending of the lifetime of PCM. In fact, as currently happens in Flash memory, the error rate of PCM devices can be expected to increase gradually during the lifetime of the memory due to wear out mechanisms whose physics are currently not completely understood. A higher error correction capability would imply, in this case, a longer lifetime for the memory. Other benefits include the possibility of more relaxed engineering requirements on a cell's expected physical behavior, which could greatly accelerate the introduction of PCM in the marketplace. Examples of such relevant coding techniques include enhancing traditional algebraic coding techniques with soft decoding capabilities, as well as using powerful coding mechanisms such as low density parity check codes with iterative decoding methods.[241] The additional complexity demanded by these newer techniques may be well within reach given the significant progress that has been achieved both from the algorithmic and logic device technology fronts.

The development of multibit PCM is a significant engineering challenge, in many ways similar to the challenges faced by multibit NAND Flash manufacturers. However, unlike block-based Flash memories, single PCM bits can be erased and reprogrammed. As discussed previously, due to variability in the response of different cells to the same input signal, as well as the smaller yet still significant variability in the response of the same cell to repeated applications of the same input signal, it appears impractical, at least presently, to attain a desired resistance level in a PCM cell by the application of a single write pulse. Instead, write-and-verify techniques will be necessary to sharpen the distribution of the outcome of each write procedure. A write-and-verify procedure is associated with a probability of failure because even after exhausting the allowed resources (in terms of time, iterations, energy, etc.), the resistance of a cell may still not be within the desired range around the target value. This prob-

ability of failure generally decreases as more resources are devoted to the write procedure, but in general, will not be negligible.

Error control coding can be pivotal in the management of these iterative write failures, provided that a good upper estimate can be established for how often they are expected to happen. One possibility for handling these errors is to employ nonbinary codes, that is, codes that detect and correct errors in symbols with more than two states. For example, a 4 bit/cell PCM may employ BCH codes defined over 4 bit symbols. If reliable information on the shape of the output distribution of an iterative write procedure is available, soft decoding procedures can be employed to improve the likelihood of successful decoding of data.

These changes in programmed resistance due to short and long-term drifts are further exacerbated in the case of multibit PCM, and constitute the fundamental limitation in storage capacity. The effects of both of these phenomena are limited whenever the PCM devices have access to a reliable power source that allows them to do regular refreshes (known as "scrubbings") over time. This is often the case for PCM devices employed in a memory context. In the case of a device intended for storage applications, no such guarantee of refresh power can be assumed. Thus the problem of recovering the stored information is markedly harder. Fortunately, the relatively relaxed bandwidth and latency requirements demanded by storage applications allow for the possibility of more complex processing at a decoder. Such more complex processing can, in general, include signal processing to recover levels that have drifted (for short and long-term drifts) as well as advanced error control coding techniques that might incorporate soft information from the drift recovery layer to enhance decoding success.

From the perspective of a read operation, after adjustments for drift have taken place, a multibit PCM cell appears to be an analog write medium with some noise around a written level, with restrictions on the minimum and maximum values we might write on the medium. Many techniques can be borrowed from communication system theory which are relevant to this setting. For example, a technique that might prove relevant is the notion of trellis coded modulation (TCM). At a very high level, TCM is a method for obtaining highly reliable multibit cells that exploits the idea of writing coded levels in the memory at a precision higher than that ultimately intended. The specific manner in which this coding is designed is one of the cornerstone successes of communication theory.[242]

A complementary idea is the concept of adding cells with redundant content, rather than writing more precise signals, which is a paradigm that is much more accepted in the memory community as it matches established methods for designing reliable memories. The correct balance between these two kinds of redundancies in a memory system design will ultimately depend on the design point for the memory, including expected density, power, bandwidth, latency, etc. One example of the options available here is "endurance" coding, in which effective PCM device endurance can be

greatly extended by using full reset pulses which melt the PCM material only sparingly,[232] although at the cost of reduced storage capacity. Once the design guidelines are established and available technologies for doing iterative programming and analog to digital conversion are set forth, it is possible to objectively identify the best method for designing the various redundancies that will be necessary for the attainment of an extremely reliable and high-density PCM system.

C. Routes to ultrahigh density

As discussed earlier (Sec. I C), the cost of a semiconductor technology depends strongly on its device density. Even though PCM technology already appears to have a good chance of matching or exceeding Flash technology in terms of performance and endurance, neither of these will matter if the cost of PCM does not (eventually) match or improve upon Flash. For instance, one possible scenario might find PCM perpetually more expensive than Flash, either because of cost issues related to large cell size or reliability issues that dampen achievable yield. Along this path, the future of PCM is dim—few customers will be willing to pay more for the better performance and/or endurance of what is essentially a new, unproven, and low-volume technology. However, in the alternative scenario, where PCM can pass Flash in terms of cost, then not only would PCM be able to compete in all the markets that Flash now occupies, but it would be immediately more suitable for solid-state disk devices and other not-yet-existing storage-class memory applications that may develop.

Thus the eventual cost of PCM technology is absolutely a key. While somewhat dependent on high-yield processing of robust PCM memory devices in high-volume manufacturing, a significant component of the cost equation depends on implementation of ultrahigh density. In particular, it is already clear that even 1 bit/$4F^2$ will not catch up with NAND flash since MLC Flash is already two times better than this and moving toward four times higher densities.[5]

Thus other techniques must be invoked in order to achieve the ultrahigh memory densities that PCM will need, both in order to succeed as a successor to Flash and to enable new storage-class memory applications. We have already extensively discussed one of these, that is, multiple bits per cell using MLC techniques in Sec. V A. Two other approaches that have been discussed are the implementation of a sublithographic cross-bar memory to go beyond the lithographic dimension, F,[243,244] and 3D integration of multiple layers of memory, currently implemented commercially for write-once solid-state memory.[245]

A sublithographic cross-bar memory requires a scheme for connecting the ultrasmall memory devices laid out at tight pitch to the "larger" wiring created at the tightest pitch offered by lithography. One scheme that was proposed used a micro-to-nanoaddressing block, in which current injected into a lithographically defined via was steered into one of several sublithographic wires using either precise control over depletion regions[244] or binary gating by overlying control gates.[246] These schemes work because lithography is

typically capable of overlay errors that are five to ten times smaller than the minimum-size feature. Thus overlying control gates can be placed to cover two but not three sublithographic wires, even though the control gate cannot possibly be made as narrow as the sublithographic wire.

The weakness of such a sublithographic cross-bar scheme is that it requires the creation and careful placement of sublithographic wire arrays of nontrivial complexity. Next-generation techniques such as imprint lithography may soon be capable of delivering small arrays at roughly the same pitch as cutting-edge lithography,[246] but unfortunately pitches that are four times denser than cutting-edge lithography is what would be required. Intriguingly, large portions of such sublithographic wire arrays would resemble simple grating patterns, suggesting the use of techniques such as interferometric lithography. Unfortunately, the addressing schemes require that wires at the edges of such arrays terminate precisely yet nonuniformly along the edge, thus complicating the task greatly for an interferometric exposure scheme.[244,246]

A more flexible approach is to build layers of PCM memory devices, stacking the memory in 3D above the silicon wafer. This is not the same as 3D packaging, where devices originally fabricated on separate silicon wafers are connected together using vias that punch through the upper silicon layers to connect to the underlying circuitry. Instead, the entire memory is built above a single layer of silicon just as the multiple wiring levels of a conventional semiconductor product are built in the "back end" of a CMOS process.

This approach has several constraints, including the need to tolerate a significant BEOL temperature budget (an example might be ~400 °C for >1 h) and the need to implement an access device for the PCM devices, which can be produced in the metal-and-dielectric layers above the original silicon wafer. Given the difficulty of growing single-crystal silicon without a seed layer, this implies that the access device must be implemented with either a polysilicon or nonsilicon device. (Note of course that the ability to easily grow multiple layers of high-quality silicon would likely enable a straightforward path to multilayer Flash memory.)

One example of such a stacked memory is the write-once antifuse memory technology developed by Matrix semiconductor (now part of SanDisk), which uses a high-performance polysilicon diode.[245] However, it is difficult to obtain the high currents and current densities needed for PCM from such diodes. This is the case even after accounting for the consideration that the lithographically defined polysilicon diode can be ten times larger in area than the sublithographic PCM device without increasing the $4F^2$ footprint.

Thus the advent of 3D-in-the-BEOL PCM technology depends on either dramatic improvements in the current-carrying capability of polysilicon diodes or on the development of a high-performance nonsilicon access device. Such a device would need to be BEOL-compatible yet not require any temperatures higher than ~400 °C, would need to readily pass the high currents ($50-150$ μA) needed for

PCM, yet must provide ultralow leakage for all nonselected devices. For instance, in a half-selected scheme implemented across a 1000×1000 device array, at the same instant that a selected device is receiving its reset current, there are ~ 2000 devices that share either the same word line or bit line with the selected device. Although these devices are each "seeing" half the voltage across the selected device, the total leakage through all these devices must remain much lower than the reset current value, implying that the required on-off ratio should be significantly in excess of 2000. However, if such an access device could be developed, since PCM itself has been proven to be BEOL compatible, the path to four- to eight-fold increases in effective areal density would be available. In combination with 2–4 bits of MLC, this would provide an extremely attractive density (and thus cost) differential over even 4 bit MLC Flash.

VI. CONCLUSIONS

PCM has made great strides over the past decade. Ten or so years ago, PCM was merely a long-dead technology that had, before expiring, helped point the way to fast-crystallizing materials and the mass-market success of read-write optical storage. At that point, any worries about the future of Flash technology were safely covered by the promise of ferroelectric and magnetic RAM. Since that time, both FeRAM and MRAM have proven to be less scalable than had been hoped,[31] although the original MRAM concept has since been mostly replaced by the more promising spin-transfer torque (RAM) (Ref. 247) and racetrack memory[248] concepts. In addition, as is often the case, the large and talented body of engineers working on Flash technology managed to hold off its "impending" demise, and successfully scaled their technology to smaller and smaller technology nodes.

However, these continuing worries about ultrascaled Flash devices have not gone away—and the rapid increases in the size of the NAND Flash market, as driven by consumer-oriented devices such as cell phones and MP3 players, now means that significant financial implications are associated with such worries. Thus PCM was given another opportunity, which it seized by quickly demonstrating better endurance than Flash and near-DRAM switching speeds using those new materials,[141] and later CMOS-compatible integration,[158] scalability to future technology nodes,[42] and the capability for robust MLC operation.[60] All this despite needing to perform high-temperature melting or recrystallization on every writing step.

That said, there remain significant hurdles standing between PCM and its success in the NVM market. These high temperatures force the associated transistor or diode used as an access device to supply a significant amount of current, and lead to PCM cell designs built around aggressively sublithographic features. In turn, the need to define such tiny features with high yield yet low variability, when coupled with the sensitivity of phase change materials to process-related damage, leads to fabrication processes that are difficult to perfect. MLC performance is frustrated by long-term drift of resistance in the amorphous phase, while PCM retention is bedeviled by early failure of the amorphous plugs in a few "unlucky" reset cells. Cycling endurance is affected by slow yet steady separation of the constituent atoms, which may be dependent on bias polarity, leading to void formation (stuck reset) or to significant shifts of the cell's operating characteristics (stuck set).

It remains to be seen if PCM researchers and developers will be able to successfully navigate these hurdles, allowing the strengths of PCM technology (its high endurance and performance relative to Flash) to shine through in marketable products. Given these strengths, one can surmise that PCM will either succeed in the long run or will fail completely, but will not be condemned to serving a few niche markets. Instead, if PCM fails, it will be on a cost basis: either tricky processes proved too difficult to implement, delivered unacceptable yields even after many months of effort, or designs were constrained to large cell sizes and thus uninteresting density points. In the cutthroat memory and storage landscape, few customers can be expected to be interested in paying significantly more for the better endurance and performance characteristics of PCM. However, if researchers can finesse the issues of resistance drift and deliver high-current-capable nonsilicon access devices, and if developers can take these advances and implement robust, high-yielding processes that combine MLC and multiple layers, then the resulting ultrahigh memory densities will put PCM in a highly advantageous position. It would be well positioned to compete directly with Flash, while simultaneously creating new applications ranging from "storage-type" storage-class memory (high-performance PCM-based SSDs for HDD replacement) to "memory-type" storage-class memory (synchronously accessed fast PCM that could bring down the cost and power of DRAM-based systems).

[1] B. Fulford, Unsung hero, Forbes, 24 June 2002).

[2] iSuppli Corporation, http://isuppli.com.

[3] S. Lai, Tech. Dig. - IElectron Devices Meet. **2008**, 01.

[4] G. E. Moore, Electronics **38**, 114 (1965).

[5] International Technical Roadmap for Semiconductors, www.itrs.net/Links/2008ITRS/Up-date/2008 Update.pdf, 2008.

[6] A. J. Hazelton *et al.*, J. Microlithogr., Microfabr., Microsyst. **8**, 011003 (2009).

[7] B. J. Lin, Microelectron. Eng. **86**, 442 (2009).

[8] K. Kim and S. Y. Lee, Microelectron. Eng. **84**, 1976 (2007).

[9] K. Prall, Non-Volatile Semiconductor Memory Workshop (NVSMW), 2007 (unpublished), pp. 5–10.

[10] S. K. Lai, IBM J. Res. Dev. **52**, 529 (2008).

[11] M. H. White, D. A. Adams, and J. K. Bu, IEEE Circuits Devices Mag. **16**, 22 (2000).

[12] W. J. Tsai *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2001**, 32.6.1.

[13] C. H. Lee, K. I. Choi, M.-K. Cho, Y. H. Song, K. C. Park, and K. Kim, Tech. Dig. - Int. Electron Devices Meet. **2003**, 26.5.1.

[14] Y. Park *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, 2.1.

[15] J. S. Sim *et al.*, Non-Volatile Semiconductor Memory Workshop (NVSMW), 2007 (unpublished),,pp. 110–111.

[16] A. Mauri, C. Monzio Compagnoni, S. Amoroso, A. Maconi, F. Cattaneo, A. Benvenuti, A. S. Spinelli, and A. L. Lacaita, Tech. Dig. - Int. Electron Devices Meet. **2008**, 22.

[17] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, Proc. IEEE **91**, 489 (2003), special issue on flash technology.

[18] H. Kurata *et al.*, Proceedings of the VLSI Circuits Symposium, 2006 (unpublished), p. 140.

[19] A. Ghetti, C. Monzio Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, A. S. Spinelli, and A. Visconti, Tech. Dig. - Int. Electron Devices Meet. **2008**, 34.3.

[20] K. Kim and J. Choi, Non-Volatile Semiconductor Memory Workshop (NVSMW), 2006 (unpublished), pp. 9–11.

[21] T.-H. Hsu *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2008**, 35.1.

[22] S. Lombardo *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2007**, 35.3.

[23] S.-M. Jung *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, 2.3.

[24] E.-K. Lai *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, 2.4.

[25] H. Tanaka *et al.*, Tech. Dig. VLSI Symp. **2007**, 14.

[26] Y. Fukuzumi *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2007**, 17.

[27] IDC, Worldwide solid state drive 20082012 forecast and analysis: Entering the no-spin zone quoted on www.STORAGEsearch.com/ssd-analysts.html.

[28] R. Freitas and W. Wilcke, IBM J. Res. Dev. **52**, 439 (2008).

[29] R. Freitas, W. Wilcke, B. Kurdi, and G. W. Burr, Storage class memory, technology, and uses, FAST'09 February 2009.

[30] W. Hutsell, J. Bowen, and N. Ekker, Flash solid-state disk reliability, Texas Memory Systems White paper, http://www.texmemsys.com/files/f000252.pdf, November 2008.

[31] G. W. Burr, B. N. Kurdi, J. C. Scott, C. H. Lam, K. Gopalakrishnan, and R. S. Shenoy, IBM J. Res. Dev. **52**, 449 (2008).

[32] S. Raoux *et al.*, IBM J. Res. Dev. **52**, 465 (2008).

[33] W. J. Wang, L. P. Shi, R. Zhao, K. G. Lim, H. K. Lee, T. C. Chong, and Y. H. Wu, Appl. Phys. Lett. **93**, 043121 (2008).

[34] D. Krebs, S. Raoux, C. T. Rettner, G. W. Burr, M. Salinga, and M. Wuttig, Appl. Phys. Lett. **95**, 082101 (2009).

[35] D. Krebs, S. Raoux, C. T. Rettner, G. W. Burr, R. M. Shelby, M. Salinga, C. M. Jefferson, and M. Wuttig, J. Appl. Phys. **106**, 054308 (2009).

[36] G. Bruns, P. Merkelbach, C. Schlockermann, M. Salinga, M. Wuttig, T. D. Happ, J. B. Philipp, and M. Kund, Appl. Phys. Lett. **95**, 043108 (2009).

[37] S. R. Ovshinsky, Phys. Rev. Lett. **21**, 1450 (1968).

[38] N. Yamada, E. Ohno, K. Nishiuchi, N. Akahira, and M. Takao, J. Appl. Phys. **69**, 2849 (1991).

[39] J. Tominaga, T. Kikukawa, M. Takahashi, and R. T. Phillips, J. Appl. Phys. **82**, 3214 (1997).

[40] S. Lai, Tech. Dig. - Int. Electron Devices Meet. **2003**, 10.1.1.

[41] A. Pirovano, A. L. Lacaita, A. Benvenuti, F. Pellizzer, S. Hudgens, and R. Bez, Tech. Dig. - Int. Electron Devices Meet. **2003**, 29.6.1.

[42] Y. C. Chen *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, S30P3.

[43] J. H. Oh *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, 2.6.

[44] F. Pellizzer *et al.*, Tech. Dig. VLSI Symp. **2006**, 122.

[45] A. Pirovano, A. L. Lacaita, A. Benvenuti, F. Pellizzer, and R. Bez, IEEE Trans. Electron Devices **51**, 452 (2004).

[46] D. Adler, H. K. Henisch, and N. Mott, Rev. Mod. Phys. **50**, 209 (1978).

[47] D. Adler, M. S. Shur, M. Silver, and S. R. Ovshinsky, J. Appl. Phys. **51**, 3289 (1980).

[48] A. Redaelli, A. Pirovano, E. Pellizzer, A. L. Lacaita, D. Ielmini, and R. Bez, IEEE Electron Device Lett. **25**, 684 (2004).

[49] D. Ielmini and Y. G. Zhang, J. Appl. Phys. **102**, 054517 (2007).

[50] D. Ielmini, A. L. Lacaita, and D. Mantegazza, IEEE Trans. Electron Devices **54**, 308 (2007).

[51] A. Pirovano, A. L. Lacaita, F. Pellizzer, S. A. Kostylev, A. Benvenuti, and R. Bez, IEEE Trans. Electron Devices **51**, 714 (2004).

[52] S. Kang *et al.*, ISSCC Technical Digest, 2006 (unpublished), pp. 140–141.

[53] S. W. Nam *et al.*, Electrochem. Solid-State Lett. **12**, H155 (2009).

[54] S. S. Iyer, J. E. Barth, Jr., P. C. Parries, J. P. Norum, J. P. Rice, L. R. Logan, and D. Hoyniak, IBM J. Res. Dev. **49**, 333 (2005).

[55] W. Wilcke (private communication).

[56] C. W. Oh *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2006**, 37.3.

[57] S. Adee, www.spectrum.ieee.org/feb09/7629, February 2009.

[58] P. Cappelletti, C. Golla, P. Olivo, and E. Zanoni, *Flash Memories* (Springer, New York, 1999).

[59] N. Shibata, ISSCC, 2009 (unpublished), Vol. 13.6.

[60] T. Nirschl *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2007**, 17.5.

[61] F. Bedeschi *et al.*, ISSCC Technical Digest, 2008 (unpublished), Vol. 23.5.

[62] S. Raoux, C. T. Rettner, J. L. Jordan-Sweet, A. J. Kellock, T. Topuria, P. M. Rice, and D. C. Miller, J. Appl. Phys. **102**, 094305 (2007).

[63] S. Raoux *et al.*, J. Appl. Phys. **105**, 064918 (2009).

[64] T. Okabe, S. Endo, and S. Saito, J. Non-Cryst. Solids **117–118**, 222 (1990).

[65] X. Q. Wei, L. P. Shi, T. C. Chong, R. Zhao, and H. K. Lee, Jpn. J. Appl. Phys., Part 1 **46**, 2211 (2007).

[66] S. Raoux, J. L. Jordan-Sweet, and A. J. Kellock, J. Appl. Phys. **103**, 114310 (2008).

[67] S. Raoux, H.-Y. Cheng, J. L. Jordan-Sweet, B. Muñoz, and M. Hitzbleck, J. Appl. Phys. **94**, 183114 (2009).

[68] S. H. Lee, Y. W. Jung, and R. Agarwal, Nano Lett. **8**, 3303 (2008).

[69] D. Yu, S. Brittman, J. S. Lee, A. L. Falk, and H. Park, Nano Lett. **8**, 3429 (2008).

[70] Y. Zhang *et al.*, Appl. Phys. Lett. **91**, 013104 (2007).

[71] D. J. Milliron, S. Raoux, R. M. Shelby, and J. Jordan-Sweet, Nature Mater. **6**, 352 (2007).

[72] Y. Zhang *et al.*, J. Appl. Phys. **104**, 074312 (2008).

[73] M. Caldwell, S. Raoux, H.-S. P. Wong, and D. Milliron, 234th Meeting of the American Chemical Society, August 2007 (unpublished).

[74] S. Raoux *et al.*, Microelectron. Eng. **85**, 2330 (2008).

[75] X. Sun, B. Yu, G. Ng, T. D. Nguyen, and M. Mayyappan, Appl. Phys. Lett. **89**, 233121 (2006).

[76] X. H. Sun, B. Yu, G. Ng, and M. Meyyappan, J. Phys. Chem. C **111**, 2421 (2007).

[77] M. H. R. Lankhorst, B. W. S. M. M. Ketelaars, and R. A. M. Wolters, Nature Mater. **4**, 347 (2005).

[78] D. Krebs, S. Raoux, C. Rettner, Y.-C. Chen, G. W. Burr, and M. Wuttig, Proceedings of the Materials Research Society, 2008 (unpublished), Vol. 1072, pp. G06–G07.

[79] D. Ielmini, Phys. Rev. B **78**, 035308 (2008).

[80] W. P. Risk, C. T. Rettner, and S. Raoux, Appl. Phys. Lett. **94**, 101906 (2009).

[81] J. P. Reifenberg *et al.*, Appl. Phys. Lett. **91**, 111904 (2007).

[82] J. Feinleib, J. deNeufville, S. C. Moss, and S. R. Ovshinsky, Appl. Phys. Lett. **18**, 254 (1971).

[83] S. R. Ovshinsky and H. Fritzsch, IEEE Trans. Electron Devices **20**, 91 (1973).

[84] N. Yamada, E. Ohno, N. Akahira, K. Nishiuchi, K. Nagata, and M. Takao, Jpn. J. Appl. Phys., Part 1 **26**, 61 (1987).

[85] M. Chen, K. A. Rubin, and R. W. Barton, Appl. Phys. Lett. **49**, 502 (1986).

[86] V. Weidenhof, N. Pirch, I. Friedrich, S. Ziegler, and M. Wuttig, J. Appl. Phys. **88**, 657 (2000).

[87] D.-S. Suh, K. H. P. Kim, J.-S. Noh, W.-C. Shin, Y.-S. Kang, C. Kim, Y. Khang, and I. K. Yoo, Tech. Dig. - Int. Electron Devices Meet. **2007**, S30P2.

[88] D. H. Kang *et al.*, Tech. Dig. VLSI Symp. **2007**, 6B.

[89] D. Ielmini, S. Lavizzari, D. Sharma, and A. L. Lacaita, Tech. Dig. - Int.

Electron Devices Meet. **2007**, 36.1.

[90]D. Ielmini, D. Sharma, S. Lavizzari, and A. L. Lacaita, 2008 IEEE International Reliability Physics Symposium (IRPS), 2008 (unpublished), pp. 597–603.

[91]A. Redaelli, A. Pirovano, A. Locatelli, and F. Pellizzer, Non-Volatile Semiconductor Memory Workshop, 2008 (unpublished), pp. 39–42.

[92]W. Welnic, A. Pamungkas, R. Detemple, C. Steimer, S. Blugel, and M. Wuttig, Nature Mater. **5**, 56 (2006).

[93]J. Kalb, F. Spaepen, and M. Wuttig, Appl. Phys. Lett. **84**, 5240 (2004).

[94]L. van Pieterson, M. H. R. Lankhorst, M. van Schijndel, A. E. T. Kuiper, and J. H. J. Roosen, J. Appl. Phys. **97**, 083520 (2005).

[95]S. Ziegler and M. Wuttig, J. Appl. Phys. **99**, 064907 (2006).

[96]J. H. Coombs, A. P. J. M. Jongenelis, W. Vanesspiekman, and B. A. J. Jacobs, J. Appl. Phys. **78**, 4906 (1995).

[97]S. Raoux *et al.*, EPCOS 2008, 2008 (unpublished).

[98]J. A. Kalb, F. Spaepen, and M. Wuttig, J. Appl. Phys. **98**, 054910 (2005).

[99]J. H. Coombs, A. P. J. M. Jongenelis, W. Vanesspiekman, and B. A. J. Jacobs, J. Appl. Phys. **78**, 4918 (1995).

[100]M. Wuttig, R. Detemple, I. Friedrich, W. Njoroge, I. Thomas, V. Weidenhof, H. W. Woltgens, and S. Ziegler, J. Magn. Magn. Mater. **249**, 492 (2002).

[101]N. Ohshima, J. Appl. Phys. **79**, 8357 (1996).

[102]L. C. Wu, Z. T. Song, F. Rao, Y. F. Gong, B. Liu, L. Y. Wang, W. L. Liu, and S. L. Feng, Appl. Phys. Lett. **93**, 103107 (2008).

[103]M. Breitwisch *et al.*, Tech. Dig. VLSI Symp. **2007**, 100.

[104]D. Tio Castro *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2007**, 12.5.

[105]G. W. Burr, M. Salinga, S. Raoux, R. M. Shelby, and M. Wuttig, Materials Research Society Spring Meeting, 2008 (unpublished).

[106]C. A. Angell, K. L. Ngai, G. B. Mckenna, P. F. Mcmillan, and S. W. Martin, J. Appl. Phys. **88**, 3113 (2000).

[107]M. Salinga, "Phase change materials for non-volatile electronic memories," Ph.D. thesis, RWTH-Aachen, 2008.

[108]S. Senkader and C. D. Wright, J. Appl. Phys. **95**, 504 (2004).

[109]J. A. Kalb, in *Phase Change Materials: Science and Applications*, edited by S. Raoux and M. Wuttig (Springer Verlag, New York, 2009), pp. 125–148.

[110]B.-S. Lee *et al.*, Science **326**, 980 (2009).

[111]F. Wang, T. Zhang, C. L. Liu, Z. T. Song, L. C. Wu, B. Liu, S. L. Feng, and B. Chen, Appl. Surf. Sci. **254**, 2281 (2008).

[112]S.-M. Yoon *et al.*, IEEE Electron Device Lett. **27**, 445 (2006).

[113]K. Sokolowski-Tinten, J. Solis, J. Bialkowski, J. Siegel, C. N. Afonso, and D. von der Linde, Phys. Rev. Lett. **81**, 3679 (1998).

[114]J. P. Callan, A. M. T. Kim, C. A. D. Roeser, E. Mazur, J. Solis, J. Siegel, C. N. Afonso, and J. C. G. de Sande, Phys. Rev. Lett. **86**, 3650 (2001).

[115]M. Wuttig, D. Lusebrink, D. Wamwangi, W. Welnic, M. Gillessen, and R. Dronskowski, Nature Mater. **6**, 122 (2007).

[116]J. A. Kalb, C. Y. Wen, F. Spaepen, H. Dieker, and M. Wuttig, J. Appl. Phys. **98**, 054902 (2005).

[117]C.-M. Lee *et al.*, International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA), 2007 (unpublished), pp. 1–2.

[118]S. Y. Lee, S. M. Yoon, Y. S. Park, B. G. Yu, S. H. Kim, and S. H. Lee, J. Vac. Sci. Technol. B **25**, 1244 (2007).

[119]S.-B. Kim and H.-S. Wong, Non-Volatile Semiconductor Memory Workshop (NVSMW), 2006 (unpublished), pp. 92–94.

[120]C. D. Wright, K. Blyuss, and P. Ashwin, Appl. Phys. Lett. **90**, 063113 (2007).

[121]K. Sonoda, A. Sakai, M. Moniwa, K. Ishikawa, O. Tsuchiya, and Y. Inoue, IEEE Trans. Electron Devices **55**, 1672 (2008).

[122]B. Rajendran *et al.*, IEEE Electron Device Lett. **30**, 126 (2009).

[123]A. L. Lacaita, A. Redaelli, D. Ielmini, F. Pellizzer, A. Pirovano, A. Benvenuti, and R. Bez, Tech. Dig. - Int. Electron Devices Meet. **2004**, 911.

[124]Y. T. Kim *et al.*, Jpn. J. Appl. Phys., Part 1 **44**, 2701 (2005).

[125]G. Wicker, Proc. SPIE **3891**, 2 (1999).

[126]T. Gille, L. Goux, J. Lisoni, K. De Meyer, and D. J. Wouters, MRS 2006, 2006 (unpublished), p. G07–02.

[127]Y. Yin, H. Sone, and S. Hosaka, Jpn. J. Appl. Phys., Part 1 **45**, 8600 (2006).

[128]U. Russo, D. Ielmini, A. Redaelli, and A. L. Lacaita, IEEE Trans. Electron Devices **55**, 506 (2008).

[129]U. Russo, D. Ielmini, A. Redaelli, and A. L. Lacaita, IEEE Trans. Electron Devices **55**, 515 (2008).

[130]T. Zhang, Z. T. Song, Y. F. Gong, Y. Lin, C. Xu, Y. F. Chen, B. Liu, and S. L. Feng, Appl. Phys. Lett. **92**, 113503 (2008).

[131]Y. Liu, Z. T. Song, Y. Ling, Y. F. Gong, and S. L. Feng, Jpn. J. Appl. Phys. **48**, 024502 (2009).

[132]D. H. Kim, F. Merget, M. Forst, and H. Kurz, J. Appl. Phys. **101**, 064512 (2007).

[133]W. Czubatyj, T. Lowrey, S. Kostylev, and I. Asano, EPCOS 2006, 2006 (unpublished).

[134]D. H. Im *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2008**, 9.2.

[135]R. Krishnamurthy, inside the numonyx pcm chip, www.semiconductor.net/blog/Chipworks_Inside_Angle/13698-Inside_the_Numonyx_PCM_Chip.php, May 2009.

[136]V. Schmidt, H. Riel, S. Senz, S. Karg, W. Riess, and U. Gosele, Small **2**, 85 (2006).

[137]X. Huang *et al.*, Tech. Dig. - Int. Electron Devices Meet. **1999**, 67.

[138]T. D. Happ *et al.*, Tech. Dig. VLSI Symp. **2006**, 120.

[139]F. Pellizzer *et al.*, Tech. Dig. VLSI Symp. **2004**, 18.

[140]S. J. Ahn *et al.*, Tech. Dig. VLSI Symp. **2005**, 98.

[141]S. Lai and T. Lowrey, Tech. Dig. - Int. Electron Devices Meet. **2001**, 3651.

[142]C. W. Jeong *et al.*, Jpn. J. Appl. Phys., Part 1 **45**, 3233 (2006).

[143]Y. J. Song *et al.*, Tech. Dig. VLSI Symp. **2006**, 118.

[144]K. C. Ryoo *et al.*, Jpn. J. Appl. Phys., Part 1 **46**, 2001 (2007).

[145]S. Tyson, G. Wicker, T. Lowrey, S. Hudgens, and K. Hunt, *2000 IEEE Aerospace Conference Proceedings* (IEEE, Piscataway, NJ, 2000), Vol. 5, pp. 385–390.

[146]S. L. Cho *et al.*, Tech. Dig. VLSI Symp. **2005**, 96.

[147]A. Pirovano, A. Redaelli, F. Pellizzer, F. Ottogalli, M. Tosi, D. Ielmini, A. L. Lacaita, and R. Bez, IEEE Trans. Device Mater. Reliab. **4**, 422 (2004).

[148]H. Horii *et al.*, Tech. Dig. VLSI Symp. **2003**, 177.

[149]Y. F. Lai, J. Feng, B. W. Qiao, Y. F. Cai, Y. Y. Lin, T. A. Tang, B. C. Cai, and B. Chen, Appl. Phys. A: Mater. Sci. Process. **84**, 21 (2006).

[150]S. Lee *et al.*, MRS Proceedings, 2005 (unpublished), Vol. 830, p. D7.9.1.

[151]F. Rao, Z. T. Song, Y. F. Gong, L. C. Wu, B. Liu, S. L. Feng, and B. M. Chen, Appl. Phys. Lett. **92**, 223507 (2008).

[152]Y. H. Ha, J. H. Yi, H. Horii, J. H. Park, S. H. Joo, S. O. Park, U.-I. Chung, and J. T. Moon, Tech. Dig. VLSI Symp. **2003**, 12B.

[153]C. W. Jeong *et al.*, Extended Abstracts of the 2005 International Conference on Solid State Devices and Materials, 2005 (unpublished), pp. 1048–1049.

[154]E. A. Joseph *et al.*, International Symposium on VLSI Technology, Systems and Applications, 2008 (unpublished).

[155]J. I. Lee *et al.*, Tech. Dig. VLSI Symp. **2007**, 102.

[156]Y.-C. Chen, C. F. Chen, C. T. Chen, J. Y. Yu, S. Wu, S. L. Lung, R. Liu, and C.-Y. Lu, Tech. Dig. - Int. Electron Devices Meet. **2003**, 37.4.

[157]Y. Zhang, S. B. Kim, J. P. McVittie, H. Jagannathan, J. B. Ratchford, C. E. D. Chidsey, Y. Nishi, and H.-S. P. Wong, Tech. Dig. VLSI Symp. **2007**, 98.

[158]Y. N. Hwang *et al.*, Tech. Dig. VLSI Symp. **2003**, 173.

[159]S. J. Ahn *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 907.

[160]H. L. Lung, M. Breitwisch, T. Happ, and C. Lam, International Conference on Memory Technology and Design, Giens, France, 2007 (unpublished).

[161]F. Bedeschi *et al.*, IEEE J. Solid-State Circuits **44**, 217 (2009).

[162]M. J. Breitwisch, in *Phase Change Materials: Science and Applications*, edited by S. Raoux and M. Wuttig (Springer, New York, 2009), pp. 381–408.

[163]H.-R. Oh *et al.*, ISSCC Technical Digest, 2005 (unpublished), Vol. 2.3, pp. 48–50.

[164]I. V. Karpov, M. Mitra, D. Kau, G. Spadini, Y. A. Kryukov, and V. G. Karpov, J. Appl. Phys. **102**, 124503 (2007).

[165]S. Braga, A. Cabrini, and G. Torelli, Appl. Phys. Lett. **94**, 092112 (2009).

[166]D. Ielmini, S. Lavizzari, D. Sharma, and A. L. Lacaita, Appl. Phys. Lett. **92**, 193511 (2008).

[167]D. Mantegazza, D. Ielmini, A. Pirovano, A. L. Lacaita, E. Varesi, F. Pellizzer, and R. Bez, Solid-State Electron. **52**, 584 (2008).

[168]B. Gleixner, A. Pirovano, J. Sarkar, F. Ottogalli, E. Tortorelli, M. Tosi, and R. Bez, Proceedings of the 45th Annual International Reliability Physics Symposium, 2007 (unpublished), pp. 542–546.

[169]Y. H. Shih *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2008**, 9.1.

[170]A. Itri, D. Ielmini, A. L. Lacaita, A. Pirovano, F. Pellizzer, and R. Bez, International Reliability Physics Symposium, 2004 (unpublished), pp. 209–215.

[171]Y.-Y. Liao and S.-F. Horng, Jpn. J. Appl. Phys. **48**, 021207 (2009).

[172]R. Y. Kim, H. G. Kim, and S. G. Yoon, Appl. Phys. Lett. **89**, 102107

(2006).

[173]R. Y. Kim, H. G. Kim, and S. G. Yoon, J. Electrochem. Soc. **155**, D137 (2008).

[174]J. Lee, S. Choi, C. Lee, Y. Kang, and D. Kim, Appl. Surf. Sci. **253**, 3969 (2007).

[175]T. Shintani, Y. Anzai, H. Minemura, H. Miyamoto, and J. Ushiyama, Appl. Phys. Lett. **85**, 639 (2004).

[176]Y. Anzai, T. Shintani, H. Minemura, H. Miyamoto, and J. Ushiyama, EPCOS 2004, 2004 (unpublished).

[177]Y. Lin, M. H. Hong, G. X. Chen, C. S. Lim, L. S. Tan, Z. B. Wang, L. P. Shi, and T. C. Chong, J. Mater. Process. Technol. **192–193**, 340 (2007).

[178]H. Y. Cheng, C. A. Jong, R. J. Chung, T. S. Chin, and R. T. Huang, Semicond. Sci. Technol. **20**, 1111 (2005).

[179]H.-Y. Cheng, C.-A. Jong, C.-M. Lee, and T. S. Chin, IEEE Trans. Magn. **41**, 1031 (2005).

[180]S. M. Yoon, N. Y. Lee, S. O. Ryu, Y. S. Park, S. Y. Lee, K. J. Choi, and B. G. Yu, Jpn. J. Appl. Phys., Part 2 **44**, L869 (2005).

[181]I. H. Park, J. W. Lee, and C. W. Chung, Integr. Ferroelectr. **80**, 207 (2006).

[182]N. K. Min, M. Kim, K. H. Kwon, A. Efremov, H.-W. Lee, and S. Kim, J. Korean Phys. Soc. **51**, 1686 (2007).

[183]K. Y. Yang, S. H. Hong, D. K. Kim, B. K. Cheong, and H. Lee, Microelectron. Eng. **84**, 21 (2007).

[184]N. K. Min, A. Efremov, Y. H. Kim, M. Kim, H. H. Park, H. W. Lee, and K. H. Kwon, J. Vac. Sci. Technol. A **26**, 205 (2008).

[185]S. M. Yoon, K. J. Choi, Y. S. Park, S. Y. Lee, N. Y. Lee, and B. G. Yu, Jpn. J. Appl. Phys., Part 2 **45**, L1080 (2006).

[186]G. M. Feng, B. Liu, Z. T. Song, S. L. Feng, and B. M. Chen, Microelectron. Eng. **85**, 1699 (2008).

[187]J. W. Lee, H. N. Cho, S. R. Min, and C. W. Chung, Integr. Ferroelectr. **90**, 95 (2007).

[188]G. M. Feng, B. Liu, Z. T. Song, S. L. Feng, and B. Chen, Electrochem. Solid-State Lett. **10**, D47 (2007).

[189]C. Xu, B. Liu, Z. T. Song, S. L. Feng, and B. M. Chen, Thin Solid Films **516**, 7871 (2008).

[190]C. Andricacos, C. Uzoh, J. O. Dukovic, J. Horkans, and H. Deligianni, IBM J. Res. Dev. **42**, 567 (1998).

[191]Q.-B. Liu, Z. T. Song, K. L. Zhang, L. Y. Wang, S. L. Feng, and B. M. Chen, Chin. Phys. Lett. **23**, 2296 (2006).

[192]M. Zhong, Z. T. Song, B. Liu, L. Y. Wang, and S. L. Feng, Electron. Lett. **44**, 322 (2008).

[193]M. Zhong, Z. T. Song, B. Liu, S. L. Feng, and B. Chen, J. Electrochem. Soc. **155**, H929 (2008).

[194]M. Zhong, Z. T. Song, B. Liu, S. L. Feng, and B. Chen, Chin. Phys. Lett. **25**, 762 (2008).

[195]S. H. Lee *et al.*, Tech. Dig. VLSI Symp. **2004**, 20.

[196]K. Kim and S. J. Ahn, IEEE International Reliability Physics Symposium, 2005 (unpublished), pp. 157–162.

[197]D. Mantegazza, D. Ielmini, A. Pirovano, B. Gleixner, A. L. Lacaita, E. Varesi, F. Pellizzer, and R. Bez, Tech. Dig. - Int. Electron Devices Meet. **2006**, p. S2P7.

[198]B. Gleixner, NVSMW 2007, 2007 (unpublished).

[199]A. Redaelli, D. Ielmini, A. L. Lacaita, F. Pellizzer, A. Pirovano, and R. Bez, Tech. Dig. - Int. Electron Devices Meet. **2005**, 742.

[200]U. Russo, D. Ielmini, A. Redaelli, and A. L. Lacaita, IEEE Trans. Electron Devices **53**, 3032 (2006).

[201]A. Redaelli, D. Ielmini, U. Russo, and A. L. Lacaita, IEEE Trans. Electron Devices **53**, 3040 (2006).

[202]U. Russo, D. Ielmini, and A. L. Lacaita, IEEE Trans. Electron Devices **54**, 2769 (2007).

[203]A. Redaelli, A. Pirovano, I. Tortorelli, D. Ielmini, and A. L. Lacaita, IEEE Electron Device Lett. **29**, 41 (2008).

[204]N. Matsuzaki *et al.*, Tech. Dig. VLSI Symp. **2005**, 738.

[205]T. Morikawa *et al.*, Tech. Dig. VLSI Symp. **2007**, 307.

[206]Y. Y. Lin *et al.*, Non-Volatile Semiconductor Memory Workshop (NVSMW), 2007 (unpublished), pp. 61–62.

[207]T. Zhang, Z. T. Song, F. Wang, B. Liu, S. L. Feng, and B. Chen, Jpn. J. Appl. Phys., Part 2 **46**, L602 (2007).

[208]T. Zhang, Z. T. Song, B. Liu, and S. L. Feng, Semicond. Sci. Technol. **23**, 055010 (2008).

[209]*Phase Change Materials: Science and Applications*, edited by S. Raoux and M. Wuttig (Springer Verlag, New York, 2009).

[210]D.-H. Kang *et al.*, Tech. Dig. VLSI Symp. **2008**, 10.

[211]L. Goux, D. T. Castro, G. A. M. Hurkx, J. G. Lisoni, R. Delhougne, D. J. Gravesteijn, K. Attenborough, and D. J. Wouters, IEEE Trans. Electron Devices **56**, 354 (2009).

[212]J.-B. Park, G.-S. Park, H.-S. Baik, J.-H. Lee, H. Jeong, and K. Kim, J. Electrochem. Soc. **154**, H139 (2007).

[213]B. Rajendran *et al.*, Tech. Dig. VLSI Symp. **2008**, 96.

[214]S. O. Ryu *et al.*, J. Electrochem. Soc. **153**, G234 (2006).

[215]L. Krusin-Elbaum *et al.*, Appl. Phys. Lett. **90**, 141902 (2007).

[216]S. M. Yoon, K. J. Choi, N. Y. Lee, S. Y. Lee, Y. S. Park, and B. G. Yu, Jpn. J. Appl. Phys., Part 2 **46**, L99 (2007).

[217]S. M. Yoon, K. J. Choi, N. Y. Lee, S. Y. Lee, Y. S. Park, and B. G. Yu, Appl. Surf. Sci. **254**, 316 (2007).

[218]S. Y. Lee, Y. S. Park, S. M. Yoon, S. W. Jung, and B. G. Yu, J. Electrochem. Soc. **155**, H314 (2008).

[219]S. W. Nam *et al.*, Appl. Phys. Lett. **92**, 111913 (2008).

[220]D. M. Kang, D. Lee, H. M. Kim, S. W. Nam, M. H. Kwon, and K.-B. Kim, Appl. Phys. Lett. **95**, 011904 (2009).

[221]C. Kim *et al.*, Appl. Phys. Lett. **94**, 193504 (2009).

[222]T. Y. Yang, I. M. Park, H. Y. You, S. H. Oh, K. W. Yi, and Y. C. Joo, J. Electrochem. Soc. **156**, H617 (2009).

[223]T. Y. Yang, I. M. Park, B. J. Kim, and Y. C. Joo, Appl. Phys. Lett. **95**, 032104 (2009).

[224]J. Sarkar and B. Gleixner, Appl. Phys. Lett. **91**, 233506 (2007).

[225]Y. Y. Lin *et al.*, 2008 International Conference on Solid State Devices and Materials, 2008 (unpublished), pp. 462–463.

[226]S. Y. Lee, J. H. Jeong, T. S. Lee, W. M. Kim, and B. K. Cheong, Appl. Phys. Lett. **92**, 243507 (2008).

[227]S. Lee, J.-H. Jeong, Y.-W. Park, W. Zhe, T. S. Lee, and B.-K. Cheong, EPCOS 2008, 2008 (unpublished).

[228]S. Lee, J. Jeong, T. S. Lee, W. M. Kim, and B. Cheong, IEEE Electron Device Lett. **30**, 448 (2009).

[229]C. Cabral, L. Krusin-Elbaum, J. Bruley, S. Raoux, V. Deline, A. Madan, and T. Pinto, Appl. Phys. Lett. **93**, 071906 (2008).

[230]M. Grossi, M. Lanzoni, and B. Ricco, Proc. IEEE **91**, 594 (2003).

[231]J. B. Philipp and T. Happ, U.S. Patent No. 7,372,725, 13 May 2008.

[232]L. A. Lastras-Montano, M. Franceschini, T. Mittelholzer, and M. Sharma, International Symposium on Information Theory and its Applications (ISITA 2008), 2008 (unpublished).

[233]T. Mittelholzer, M. Franceschini, L. A. Lastras-Montano, I. Elfadel, and M. Sharma, International Conference on Communications (ICC 2009), 2009 (unpublished).

[234]D. Mantegazza, D. Ielmini, E. Varesi, A. Pirovano, and A. L. Lacaita, Tech. Dig. - Int. Electron Devices Meet. **2007**, 311.

[235]S. A. Kostylev, W. Czubatyj, and T. Lowrey, U.S. Patent No. 6,914,801, 5 July 2005.

[236]R. W. Hamming, Bell Syst. Tech. J. **29**, 147 (1950).

[237]I. S. Reed and G. Solomon, J. Soc. Ind. Appl. Math. **8**, 300 (1960).

[238]R. C. Bose and D. K. Ray-Chaudhuri, Inf. Control **3**, 68 (1960).

[239]R. C. Bose and D. K. Ray-Chaudhuri, Inf. Control **3**, 279 (1960).

[240]A. Hocquenghem, Chiffres **2**, 147 (1959).

[241]R. G. Gallager, *Low Density Parity Check Codes* (MIT Press, Cambridge, MA, 1963).

[242]G. Ungerboeck, IEEE Trans. Inf. Theory **28**, 55 (1982).

[243]Z. H. Zhong, D. L. Wang, Y. Cui, M. W. Bockrath, and C. M. Lieber, Science **302**, 1377 (2003).

[244]K. Gopalakrishnan *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2005**, 471.

[245]M. Johnson *et al.*, IEEE J. Solid-State Circuits **38**, 1920 (2003).

[246]M. Hart, EIPBN, June 2007 (unpublished).

[247]M. Hosomi *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2005**, 459.

[248]S. S. P. Parkin, Tech. Dig. - Int. Electron Devices Meet. **2004**, 903.

[249]S. Kang *et al.*, IEEE J. Solid-State Circuits **42**, 210 (2007).