INVITED PAPER

Magnetic Racetrack Memory: From Physics to the Cusp of Applications Within a Decade

This review evaluates major breakthroughs in racetrack memory technology from a physics and computer architecture perspective and provides an outlook on its future

By Robin Bläsing, Asif Ali Khan, Panagiotis Ch. Filippou, Chirag Garg, Fazal Hameed, Jeronimo Castrillon, *Senior Member, IEEE*, and Stuart S. P. Parkin, *Fellow, IEEE*

ABSTRACT | Racetrack memory (RTM) is a novel spintronic memory-storage technology that has the potential to overcome fundamental constraints of existing memory and storage devices. It is unique in that its core differentiating feature is the movement of data, which is composed of magnetic domain walls (DWs), by short current pulses. This enables more data to be stored per unit area compared to any other current technologies. On the one hand, RTM has the potential for mass data storage with unlimited endurance using considerably less energy than today's technologies. On the other hand, RTM promises an ultrafast nonvolatile memory competitive with static random access memory (SRAM) but with a much smaller footprint. During the last decade, the discovery of novel physical mechanisms to operate RTM has led to a major enhancement in the efficiency with which nanoscopic, chiral DWs can be manipulated. New materials and artificially atomically engineered thin-film structures have been found to increase the speed and lower the threshold current with which the data bits can be manipulated. With these recent developments, RTM has attracted the attention of the computer architecture community that has evaluated the use of RTM at various levels in the memory stack. Recent studies advocate RTM as a promising compromise between, on the one hand, power-hungry, volatile memories and, on the other hand, slow, nonvolatile storage. By optimizing the memory subsystem, significant performance improvements can be achieved, enabling a new era of cache, graphical processing units, and high capacity memory devices. In this article, we provide an overview of the major developments of RTM technology from both the physics and computer architecture perspectives over the past decade. We identify the remaining challenges and give an outlook on its future.

KEYWORDS | Curved wires; domain wall memory (DWM); DW motion; nonvolatile memory; racetrack memory (RTM); threshold current.

I. INTRODUCTION

Conventional data storage and memory technologies are highly constrained by fundamental technology limits. As a result, a number of nonvolatile storage and memory technologies have emerged recently. The uninterrupted scaling and 3-D integration of NAND flash technology have enabled it to outperform hard disk drives (HDDs) in terms of volume and planar storage density [1]. However, its limited write endurance and higher erase and write latencies limit its applicability in future computing systems. Similarly, in memory technologies, phase change and magnetic memories have been proposed as candidate replacements for static random access memory (SRAM) and dynamic random access memory (DRAM) [2]. However, phase change memory (PCM) suffers from durability issues and its write

Manuscript received September 11, 2019; revised December 23, 2019; accepted January 10, 2020. This work was supported in part by the German Research Council (DFG) through the TraceSymm Project under Project CA 1602/4-1 and the Collaborative Research Center under Grant SFB 762, in part by the Cluster of Excellence Center for Advancing Electronics Dresden (CfAED), and in part by the European Research Council (ERC) through the European Union's Horizon 2020 Research and Innovation Program under Grant 670166. *(Robin Bläsing and Asif Ali Khan contributed equally to this work.) (Corresponding author: Stuart S. P. Parkin.)*

Robin Bläsing and **Stuart S. P. Parkin** are with the Max Planck Institute of Microstructure Physics, 06120 Halle, Germany (e-mail: stuart.parkin@mpi-halle.mpg.de).

Asif Ali Khan, Fazal Hameed, and **Jeronimo Castrillon** are with the Department of Computer Science, TU Dresden, 01069 Dresden, Germany.

Panagiotis Ch. Filippou and **Chirag Garg** are with IBM Research–Almaden, San Jose, CA 95120 USA.

Digital Object Identifier 10.1109/JPROC.2020.2975719

Table 1 RTM Comparison With Other Memory Technologies [1], [18]–[22]

	SRAM	DRAM	STT-RAM	RRAM	PCM	MRAM	V-NAND	RTM	HDD
Cell Size (F^2)	$120-200$	$4-8$	$6-50$	$4-10$	$4-12$	$10-60$	$1-5$	≤ 2	0.5
Write Endurance	$\geq 10^{16}$	$\geq 10^{16}$	$4\ x\ 10^{12}$	10^{11}	10^{9}	$> 10^{12}$	10^3-10^5	$\geq 10^{16}$	$\geq 10^{16}$
Read Time (ns)	$1-100$	30	$3-15$	$10-20$	$5-20$	$3-20$	$25\ x\ 10^3$	$3-250$*	$2\ x\ 10^6$
Write /Erase Time (ns)	$1-100$	50	$3-15$	20	> 30	$10-20$	$(0.1-1)\ x10^6$	$3-250$*	$2\ x\ 10^6$
Read Energy	*Low*	*Medium*	*Low*	*Low*	*Medium*	*Low*	*Medium*	*Low*	*Medium*
Write Energy	*Low*	*Medium*	*High*	*High*	*High*	*High*	*High*	*Low*	*Medium*
Leakage Power	*High*	*Medium*	*Low*	*Low*	*Low*	*Low*	*Low*	*Low*	*Low*
Retention Period	*As long as voltage applied*	$64-512\ ms$	*Variable*	*Years*	*Years*	*Years*	*Years*	*Years*	*Years*

*including the shift latency.

latency is an order of magnitude higher compared to SRAM [2]. The spin-orbitronics-based magnetic racetrack memory (RTM) combines the best of all worlds, simultaneously offering endurance of magnetic HDDs, the density of 3-D vertical NAND flash, with the attractive latency rates of SRAM and DRAM [3]–[5]. A summary of qualitative and quantitative comparison of RTM with other technologies is presented in Table 1, which shows tradeoffs among various parameters that include latency, area, power, and retention characteristics.

RTM was first proposed in 2002 [6] and its fundamental underlying principle was first demonstrated in 2008 [4], [7]. Research studies over the past decade have led to unexpected physical mechanisms to operate RTMs. The information in RTM is stored in a magnetic track in which magnetic regions serve as bits, similar to HDDs. In contrast to the latter, RTM is neither limited to a 2-D design nor relies on the mechanical motion for operation. Instead, the magnetic bits are moved by electrical currents in which spin-polarized electrons interact with magnetic moments. As the motion is always along the electrical current, arbitrary pathways can be structured, making it possible to move bits in curved or even vertical wires [8]. Advances in spin-orbit mechanisms have led to different generations of RTM (see RTM 1.0–4.0 [3]), each characterized by leaps in the motion efficiency of the bits.

In this article, we review major breakthroughs and recent advances in the RTM technology starting from fundamental physics and materials science to the overall memory architecture. We focus on demonstrated experimental work of racetrack domain wall (DW) motion and materials used. We explain the data sensing and read/write mechanisms of the RTM access ports, organization of racetracks into arrays, data storage, and access ports management in order to give a comprehensive picture of the overall functionality of the technology. We review critical technology parameters such as threshold current densities and their impact on the velocities of magnetic DWs and movement of DW in curved wires and its associated challenges. We discuss the impact of different magnetic materials such as Heusler structures and ferrimagnetic bi-layers on DW motion. In these, novel DW driving mechanisms allow faster and more efficient DW motion reducing the power consumption of RTM. This article also includes a survey of prominent applications of RTM and its evaluation at various levels in the memory subsystem. We then discuss hardware/software (HW/SW) optimizations required to mitigate the cost of shifting domains and potential errors inherent to RTM technologies. This article closes with insights into future research directions, concerning materials, circuits and design methods, and future reconfigurable memory hierarchies based on RTM.

II. RTM PRELIMINARIES

This section provides a background on the RTM cell structure, read/write mechanism, access ports management, array architecture, and data organization.

A. RTM Cell Structure and Data Representation

An RTM consists of magnetic nanowires—magnetic racetracks—which are organized horizontally or vertically on a silicon wafer as depicted in Fig. 1 [4], [6]. In many magnetic materials, grown as a thin film, the magnetization can take two states, for example, pointing up or down. These states can serve as bits representing “0”s or “1”s, respectively, which can be stored with unprecedented density. By sending an electrical current along the wire, the bits can be shifted, synchronously to another location on the racetrack [4]. In that way, the information can be moved to a readout unit, referred to as an access port, which determines the state of the magnetization (read operation). The access port could also switch the magnetization state by sending a larger current (write operation), as explained in Section II-B.

The fundamental 2-D arrangement can also be extended to a 3-D design in which the information is shifted

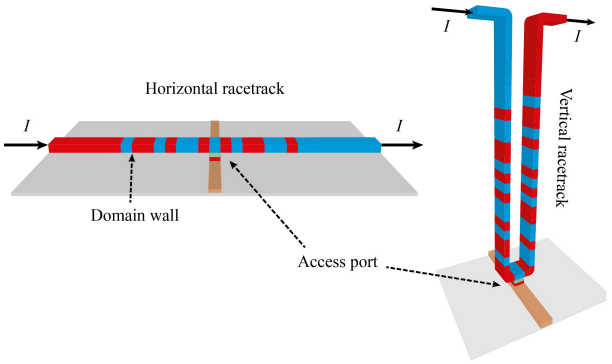

Fig. 1. ***Horizontal and vertical racetrack with one access port. The current flows through the device along the bit motion direction. Overflow bits at the ends of the wire can be reduced by increasing the number of access ports [4], [6].***

vertically, thereby further increasing the storage capacity per feature size [4]. The motion of the magnetic bits is derived from the interaction of current at the boundaries between oppositely magnetized regions. These boundaries are the magnetic DWs, within which the magnetic moment gradually rotates from one direction to the other direction. Typically, the DWs are just a few nanometers wide. By sending an electrical current, the local magnetic moments rotate such that the center of the DW moves either along or against the current flow direction.

The access latency and energy consumption of RTM largely depend on the number of shift operations required. While a track could be equipped with multiple access ports, the number of ports per track is always lower than the number of domains. Therefore, ports are shared among multiple domains. Increasing the degree of sharing improves the area efficiency, but significantly increases the number of needed shifts which, in turn, reduces the RTM average access latency.

Typically, an access port is made up of an access transistor and a magnetic tunnel junction (MTJ). The access transistor, controlled by the word-line, enables read/write operations, and controls the current density. The transistor size in the access port is typically much larger than the track size [9]. As a result, it dominates the die-area as schematically depicted in Fig. 2.

A simple adjacent placement of tracks on a horizontal surface leads to significant die-area wastage. To avoid this, recent designs overlap access transistors by grouping tracks together and placing them adjacently. Groups of racetracks are referred to as macrounits [9]–[11] or DW block clusters (DBCs) [12]–[17] in the literature (Fig. 2).

B. Read/Write Mechanism in RTM

As mentioned above, RTM is equipped with access ports. Bits can be shifted to the access port locations for data reading or writing. In a conventional HDD, a read/write sensor moves mechanically to the location of the magnetic bit on the rotating disk in order to engage in a read/write operation. In contrast, the RTM access ports are fixed at particular locations on the track and instead, the bits are moved electrically to the port location for read/write operations. The magnetic state readout can be realized via magneto-resistive effects. Giant magnetoresistance (GMR) [23]–[26] and tunneling magnetoresistance (TMR) [27] are two such phenomena that occur when two magnetic layers are separated by a nonmagnetic conductive layer or an insulator, respectively. Originally, it was proposed by Chen and Parkin in [28] that the read operation can be performed by affixing a magneto-resistive sensor in proximity to the track in order to use the emanating fringing fields for distinguishing between the magnetic states, or by integrating an MTJ sensor directly onto the racetrack [4]. Recent developments in MTJs [29], [30] allow for CMOS integration and scaling to feature sizes compatible with RTM applications. Furthermore, TMR values can far exceed those of GMR reaching values upward of 600% at room temperature [31], [32]. Thus, the key element of an access port which can perform both read and write is an MTJ. In Fig. 3, an access port is shown where the MTJ interfaces the racetrack at the top but could also be at the bottom. In the MTJ, one of the magnetic layers is engineered to have a fixed orientation [33], [34], and the other magnetic layer is formed by the section of the magnetic track which is in contact with the insulating layer (most commonly used MgO or Al_2O_3). The resistive state of the junction can be read by flowing a small reading current perpendicular to the junction. The parallel or antiparallel orientation of the magnetic bit relative to the fixed magnetic layer corresponds to two distinct resistance states, “0” or “1.” Thus, the magnetic bit within the access port can be read depending on its

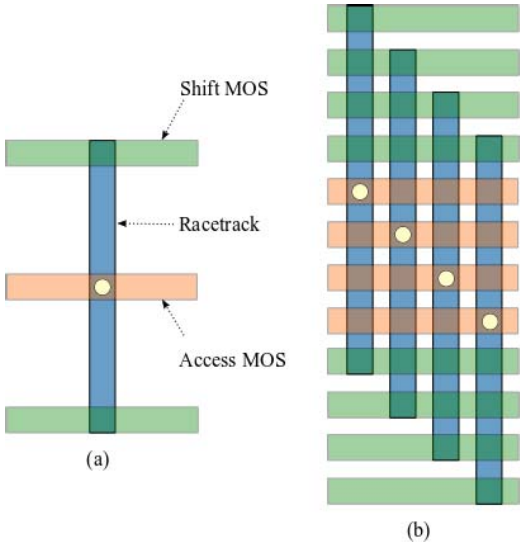

Fig. 2. ***Macro-unit/DBC. (a) Single-cell DBC (top view). (b) Four-cell DBC with an overlapped transistor area [9], [12].***

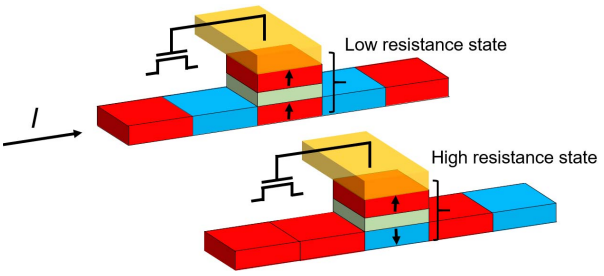

Fig. 3. ***Part of the RTM showing the access port. The access port consists of an MTJ with a fixed upper magnetic layer, an intermediary insulating layer (green), and a section of the racetrack. Shifting of the magnetic track is accomplished upon application of an electric pulse and readout is carried through the MTJ at the access port. In actual devices, long-range dipole fields emanating from the magnetic layers need to be eliminated using, for example, an SAF structure that was originally devised by one of the authors in 1989 [38], [39].***

individual orientation. The MTJ can also be used as a writing device when larger currents are used [35]–[37]. As the tunneling currents derived from the magnetic layers are spin-polarized, they result in a strong interaction that can reorient the magnetic bit in the access port via spin transfer torques (STTs). The orientation of the bit to be written can be determined by the polarity of the applied current through the MTJ.

An MTJ device for reading and writing is perhaps the simplest solution but there are other possible solutions for both reading and writing. There can be significant advantages in having two distinct devices, one for reading and one for writing since these devices can be individually optimized for their respective functions. For example, an optimized MTJ for reading can have a thicker tunnel barrier which provides for higher TMR and better endurance against junction breakdown. Other ways of writing have been demonstrated which involve, for example, the use of spin-orbit torques (SOTs) that are derived from the nonmagnetic underlayers showing spin Hall effect [40], [41]. In this case, the current is applied through the magnetic track instead of applying it across the tunnel junction. This prevents the deterioration of the insulating layer that may occur with the use of high writing currents through the MTJ but may require fringing fields generated through the addition of electrical contact lines [42]. In contrast, an in-line injector has been demonstrated that involves the flow of current through the track but without the requirement of adding any further electrical lines. The fringing fields, in this case, are generated by the creation of a 90° DW through local irradiation of the magnetic film that has perpendicular magnetic anisotropy (PMA). The passage of currents in the track results in the nucleation of DWs through the generation of STT in the presence of these fringing fields [43].

In the RTM access port, the operation to read or write is fully electrical with no need for moving parts, leading to high-performance and high-density memory storage.

C. Access Ports Management

The shift-controller maintains and manages the status of the access ports in an RTM. At each memory access, the shift-controller decides which access port will access the data, computes the number of shifts required for aligning the port position to the requested data, and updates the status of the access ports. The selection of access ports and the number of shifts required before accessing the requested domains depends upon the port access policy which can either be static or dynamic [12], [15]. Similarly, updating the port positions after completion of a memory request also depends upon the port access policy.

In static port access policies, ports are statically assigned to domains. For instance, if a racetrack stores 64 domains and has two access ports, one possible static assignment is to dedicate the first 32 domains (i.e., 0–31) to port 0 and the remaining domains (i.e., 32–63) to port 1. In dynamic port access policies, the access port that is closest to the requested domain accesses it. This implies that any access port can access any domain in the racetrack depending on the data access pattern and the positions of the current ports.

While a static port access policy makes the implementation of the shift controller a lot simpler, it can lead to significant increases in the number of shifting operations. For instance, if the initial positions of the access ports are set to 0 and 63, consecutive accesses to domains 31 and 32 require 31 shifts each and will be accessed by different ports (ports 0 and 1, respectively). In a dynamic port access policy, both accesses will be performed by port 0 and will incur a total of 32 $(31 + 1)$ shifting operations. In rare situations, a dynamic port access policy can still increase the number of shifts compared to a static port access policy. To illustrate such a scenario, consider the above assumptions of initial port positions and the following domain access pattern, 31, 45, 52, 57, and 25. In a dynamic port access policy, all accesses are performed by port 0 because it is always closer to the next requested domain and the total number of shifts incurred sums up to 89. On the contrary, in the sample static policy mentioned above, port 0 serves the first and the last requests and port 1 serves the remaining three requests, incurring altogether 67 shifts.

Another important design aspect of the shift controller is the port update policy. After accessing a domain, the position of the access port can be either restored to its default location (incurring twice as many shifts as required for aligning) or updated to the location of the current access. The former is known as “eager” while the latter is referred to as the “lazy” port update policy [12], [15]. Finally, the port access policy also affects the number of overflow bits per track. A static port access policy requires less overflow bits compared to a dynamic policy. Most RTM designs adopt dynamic policies for the port access and the lazy policy for the port update.

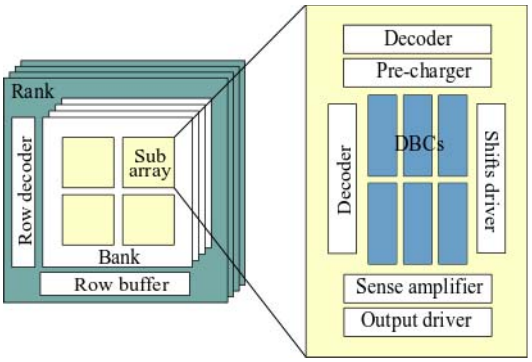

Fig. 4. *Overview of the overall architecture of an integrated RTM. A DBC serves as a basic building block of an RTM array. Like other memory technologies, one or more arrays are then combined to form independent banks.*

D. RTM Architecture and Data Organization

A DBC is the basic building block of an RTM array. It consists of M tracks where each track is equipped with P access ports and has N domains. Although the RTM cell structure is fundamentally different than existing memory technologies, recent RTM designs maintain the same I/O interface and memory hierarchy to ease technology adoption [9], [17], [44]. For example, the entire memory unit is hierarchically decomposed into ranks, banks, and subarrays. One such widespread architecture is shown in Fig. 4.

A subarray being the smallest component in the architecture needs to be carefully designed. The subarray design substantially affects the RTM's performance, energy, and area efficiency [44], where area efficiency refers to the area ratio of the data array and the peripheral circuitry. Although most of the peripheral circuitry in an RTM subarray is similar to existing memory technologies, the shift-controller is specific to RTM subarrays.

RTMs are inherently sequential in nature. A track in an RTM contains multiple DWs which can accommodate an entire data word. However, storing a single word in the track serially leads to significant performance degradation. To completely eliminate the shift operation, a track can store a single DW. However, single DW RTMs have a negative impact on density.

To keep both the performance and the density benefits intact, recent designs store data in DBCs in an interleaved fashion and move the DWs in a lockstep fashion [9]–[17], [44]–[46]. An M-bit memory object is distributed across the M tracks of a DBC as schematically shown in Fig. 5. Large size variables can be further distributed across multiple DBCs. Accessing a variable requires shifting and aligning the access port position to all required domains at the same time and all bits of the requested data can be accessed in parallel.

III. PHYSICAL AND MATERIAL DEVELOPMENTS IN RTM

This section overviews the development of RTM from version 1.0 to 4.0 in which especially the mechanisms of DW motion evolved to highly efficient driving torques. These also apply to ferrimagnetic systems in which very low threshold current densities to move DWs have been discovered. Finally, an overview of recent advances in epitaxially grown materials is provided in which fast DW motion and extremely low threshold current densities have been reported.

A. Development of RTM

RTM relies on the motion of magnetic DWs by an electrical current. This was first demonstrated in permalloy nanowires in which the DWs moved at about 100 m s^{-1} by the use of volume STT [47]. This was the driving mechanism in the first prototype of RTM. In a second version, the magnetic materials were improved so that the magnetization did not lie in the wire plane but instead pointed out of the wire plane. Such magnetic materials exhibit a strong PMA which makes the DWs narrower and more robust against annihilation. As a result, a higher packing density can be achieved. DW motion in materials with PMA was first demonstrated in Co/Ni multilayers [48], [49]. The motion of DWs by volume STT for RTM versions 1.0 and 2.0 is depicted in Fig. 6.

In 2011, a much faster DW motion was reported in a system consisting of an ultrathin magnetic layer which exhibits PMA by virtue of its interface with a heavy metal underlayer such as Pt [50]. Interestingly, the direction of DW motion was now observed to be opposite to the electron flow direction. To account for that, a new, much more efficient mechanism was introduced—the chiral spin torque (CST) [51]. In these systems, there is the generation of the Dzyaloshinskii–Moriya interaction (DMI) inside the magnetic film due to symmetry breaking in the

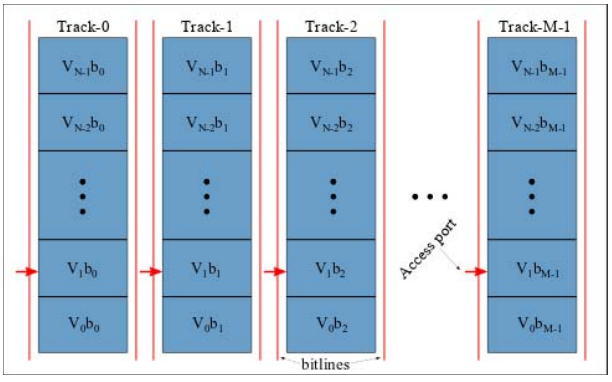

Fig. 5. *Data organization in a DBC (v: variable, b: bit). N variables each of size M are stored in an M-cell DBC in a bit-interleaved fashion. If access ports of all M-cells point to the same location (as shown), all bits of the variable can be read in parallel.*

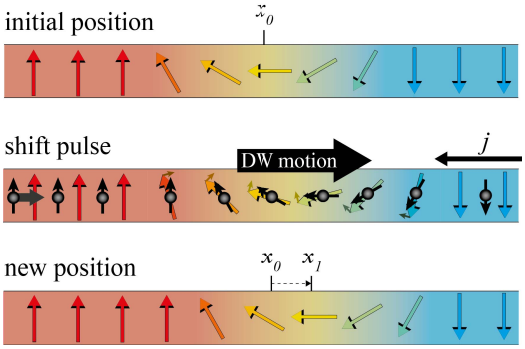

Fig. 6. *Magnetic DWs are shifted by current pulses which rotate the local magnetization (indicated by colored arrows). In the 1.0 and 2.0 RTM versions, motion is governed by a volume STT in which the electrons (black arrows) transfer their angular momentum to the localized magnetic moments. The DW motion is generally in the electron flow direction.*

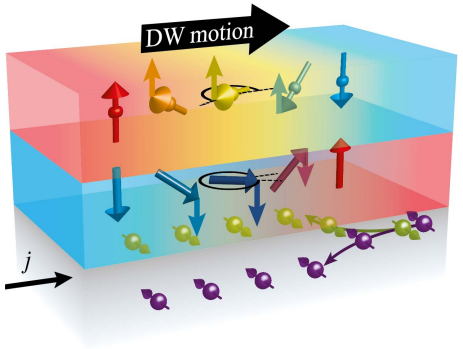

Fig. 8. *Two AF coupled layers in which the DW motion is governed by an ECT. Spin current from the underlayer turns magnetic moments toward the spin polarization. Due to the rotation out of the antiparallel alignment an exchange coupling field is created which applies an ECT on the magnetic moments, moving the DW into the current flow direction [54].*

presence of spin–orbit coupling at the heavy metal/magnetic layer interface. This exchange favors a canting of the magnetic moments with a fixed chirality which is observed through the formation of Néel DWs as shown in Fig. 7. The rotation of magnetization at the DW boundary in the Pt/Co system features a counterclockwise (CCW) direction. In addition to the DMI, the heavy metal exhibits a spin Hall effect which creates a spin current perpendicular to the current flow direction illustrated in Fig. 7(b). This spin current flows into the magnetic layer and exerts an STT on the magnetic moments. This CST is much more efficient than the volume STT, resulting in much larger DW velocities of almost 400 m s^{-1} [51], [52]. This effect has also been demonstrated in other heavy metal underlayers besides Pt [52].

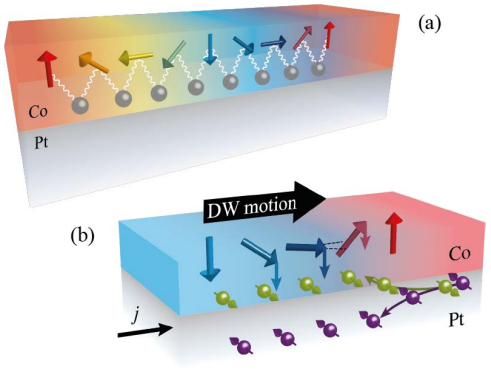

Fig. 7. *Magnetic DWs in a ferromagnetic material (e.g., Co). (a) DW chirality in subsequent DWs is conserved due to the DMI at the interface to a heavy metal layer (such as Pt). (b) The electrical current in the heavy metal layer creates a spin current due to the spin Hall effect which diffuses into the ferromagnetic layer. The spins are polarized such that they exert a torque on the magnetization, rotating them out of the DMI-favored orientation. Hence, an effective DMI field is created which exerts a CST on the magnetic moments which finally moves the DW along the current flow direction [51], [53].*

Lastly, a great step toward application was achieved by using antiferromagnetically (AF) coupled structures [54] where the orientation of magnetization in two magnetic layers is antiparallel to each other as shown in Fig. 8. In a synthetic antiferromagnet (SAF) two magnetic layers are in indirect contact through a spacer layer such as Ru which mediates the AF exchange coupling [26]. The magnetic sublattices can also couple AF without the need of a spacer layer as discussed further in Section III-B. As in the previous version of RTM, the structure is grown on top of a heavy metal underlayer. Due to the AF coupling, an exchange coupling torque (ECT) derived from the exchange field of a much higher magnitude than the DMI field [53], [54]. When the magnetization of the two magnetic layers is equal, the ECT is maximized. Due to the ECT, the DW mobility, which is the increase of DW velocity with respect to increasing current density, is also high at high current densities. For an SAF structure, a DW velocity of >750 m s^{-1} has been reported [54]. In this fourth version of RTM, the antiferromagnetic coupling not only allows a higher DW mobility due to the ECT but also makes the DWs highly stable against external magnetic fields. Most importantly, the SAF structure eliminates magnetostatic stray fields that would otherwise emanate from the magnetic layers in the racetrack and lead to unwanted interactions between DWs within and between racetracks.

B. DW Motion in Ferrimagnetic Systems

After the discovery of the ECT in SAF structures, renewed interest in research on DW motion in AF coupled systems emerged [55]–[62]. Besides SAF structures, ferrimagnetic alloys or multilayers that are composed of rare earth (RE) metals and transition metals (TMs) exhibit an antiferromagnetic coupling between the magnetic moments of the RE and TM materials. The exchange coupling between these elements can be stronger than the coupling in SAF structures as these elements couple AF without the need for a spacer layer.

When two magnetic sublattices of an RE and a TM couple together, the respective magnetizations per unit volume, m_{RE} and m_{TM}, can become compensated when $m_{\mathrm{RE}} = m_{\mathrm{TM}}$ such that the net magnetization is zero. This can be achieved either by varying the composition or by varying the temperature. As the magnetism in REs is carried by the inner shell 4*f* electrons instead of the 3*d* conduction electrons as in TMs, the dynamics of the respective magnetic moments are distinct. This is embedded in the gyromagnetic ratio $\gamma = g(\mu_{\mathrm{B}}/\hbar)$ with the Bohr magnetron μ_{B}, the reduced Planck constant $\hbar$ and the material-dependent Landé g-factors g_{RE} and g_{TM}. As a result, in dynamic processes like DW motion, the response of each magnetic sublattice to spin currents is different. However, there exists a compensation point where $m_{\mathrm{RE}}/\gamma_{\mathrm{RE}} = m_{\mathrm{TM}}/\gamma_{\mathrm{TM}}$ where m/γ is the respective angular momentum. Recent studies have shown that the DW mobility in ferrimagnetic systems is maximized at this angular momentum compensation point [55], [56], [61]. This has allowed for the motion of DWs with speeds at least as fast as those of SAF structures [56]. This effect is likely to originate from the comparably low magnetization in the REs which are highly temperature sensitive [55]. Because of this temperature dependence, Joule heating can influence the DW motion greatly. It has been shown that a single current pulse of 10-ns length at a density of 1×10^8 A cm^{-2} can easily heat up the device by ∼75 K [55]. Hence, lowering the threshold current to at least 1×10^6 A cm^{-2} but keeping a large DW mobility at the same time is of major interest for applications. Ferrimagnetic systems are a step toward fulfilling both requirements but their extreme temperature dependence makes them less appealing than SAFs.

C. Threshold Current Density

The minimum energy for shifting DWs is determined by the threshold current density that needs to be applied to overcome DW pinning. In ferromagnetic systems, e.g., consisting of a Co/Ni multilayer, the DW is driven by CST (see Section III-A) and the threshold current is of the order of 0.5×10^8 A cm^{-2} [51]. In SAF structures, the DW mobility is increased but the threshold current density is not significantly reduced [54]. Table 2 summarizes the measured threshold current density for nanosecond long pulses for various magnetic material systems on a Pt underlayer, as of today. It shows that films containing RE metals show a lower threshold current density. Considering a 20-nm-wide racetrack, the energy required for one shift at the given current densities in these materials is of the order of a few fJ.

Several proposals for the origin of the threshold current density have been made. Depending on the DW driving mechanism, the pinning is either intrinsic [64] or extrinsic arising from defects and roughness of the sample [65]. For the mechanisms in RTM versions 3.0 and 4.0, no intrinsic mechanism has been identified which could explain the large threshold currents which appear in the experiments. Hence, extrinsic pinning is a likely explanation for the appearance of a threshold current density. While edge roughness of patterned nanowires is difficult to avoid, especially in nanometer wide wires [66], atomic defects and inhomogeneities also have to be taken into account [67]. The density and strength of these defects determine the threshold current density.

Table 2 Comparison of Threshold Current Densities for Different Magnetic Materials

Material system	Threshold current density* (10^8 A cm^{-2})
Pt / Co	1 [51]
Pt / Co / Ni / Co	∼0.5 [52]
Pt / Co / Ni / Co / Ru / Co / Ni /Co	∼0.5 [55]
[Co / Tb]$_9$ / Pt	0.15 [64]
Pt / Co / Gd	∼0.3 (at 200 K) [56]
Pt / $Co_{44}Gd_{56}$	∼0.3 (at 314 K) [57]
Pt / $Co_{74}Tb_{26}$	0.2 [62]

* Pulse length 1 to 100 ns

To obtain a lower threshold current density, the most straightforward approach is the reduction of defects and roughness in the sample. Although most samples are of good crystallinity, a further improvement, for example, of the interface roughness could be achieved by using different underlayers or utilizing various growth methods. To describe homogenously distributed defects in a sample, for example, the dry friction model [65] is in good agreement with the experiments [55]. In such a model, besides the defect distribution, there are two other parameters that can be tuned in order to reduce the threshold current [55]. One is the spin Hall angle of the underlayer which, if larger, can produce the same spin current into the magnetic layer at a lower electrical current density. The other parameter is the magnetization of the magnetic layer. By decreasing it, the DW is effectively lighter and hence, easier to move. Continued efforts in material engineering at the atomic scale are needed to achieve a combination of a low threshold current while maintaining a high DW velocity at a particular current density.

D. Influence of Curvature on the Operation of RTM

The dynamics of DW motion have been well studied for magnetic nanowires that are straight. It has been found that irrespective of the underlying mechanism of DW motion—whether the torque is derived from STT, CST, or ECT–DWs move in a synchronistic fashion, a key requirement of RTM. This is not the case for the motion of chiral DWs in curved nanowires as shown in Fig. 9. Instead, the curvature of the wire can significantly alter the motion of DWs [68]. Two adjacent DWs in a curved nanowire travel with very different speeds, leading to a difference in speed that was observed to be up to an order of magnitude. This difference results from a speeding up and a slowing down of the DW pair, relative to their motion in a straight wire. It was also found that whether a DW

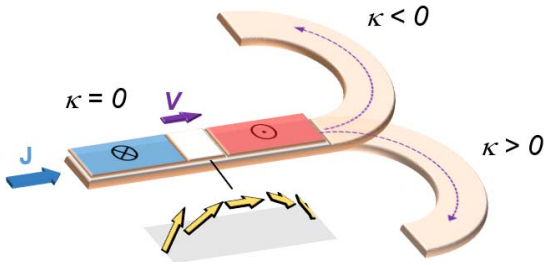

Fig. 9. *Chiral DW in a ferromagnetic track traveling through a curvature speeds up or slows down depending on the sign of the curvature (κ).*

speeds up/slows down depends on its direction (clockwise (CW) or CCW) of motion along a curvature. When the difference in speeds causes the separation between the DWs to shrink sufficiently, the DW pair can annihilate leading to a loss of data. Although horizontal racetracks that are made exclusively from straight nanowires do not suffer from this shortcoming, vertical racetracks conceived as U-shaped wires or other designs such as the ring memory which incorporate bends are likely to be affected.

An analytical model based on the motion of a DW along a curved wire revealed that the difference in speeds arises from the disparate tilting behavior of the adjacent DWs during motion in a curved wire. A DW that travels orthogonally to the wire experienced greater driving torques and moved faster in contrast to a DW that accumulates a tilt during its motion. Remarkably, this problem was found to be eliminated in curved nanowires that were composed of SAF structures [68], [69]. The driving torques in such magnetic structures are largely derived from an ECT that is insensitive to the tilting behavior of the DWs. DWs in such structures move at the same speeds in both the curved and straight sections of the wire. Thus, the SAF removes an unanticipated but critical hurdle to the implementation of RTM in two and three dimensions.

From the architecture perspective, some curved wires, such as the ring-shaped, may be more favorable compared to the traditional stripe-shaped RTM which is open-ended and suffers from data overflow issues. Overflow happens when a bit is shifted beyond the end of the track which causes data loss. This problem can be addressed by using additional peripheral registers that latch the overflow information [70], or by increasing the number of ports [71], or by employing extra domains in the track to avoid data loss. However, these techniques degrade device density, performance, and energy consumption [70], [72]. The ring-shaped RTM has already been demonstrated as a possible option to overcome this issue [70], [72], [73] and ensure end-to-end information storage by avoiding the data overflow caused by shifting. In addition, a ring-shaped RTM also reduces the worst case shifts from $(N-1)$ to $N/2$ for an N-bit racetrack [73]. The latter shift reduction property of ring-shaped further reduces the latency and energy consumption compared to stripe-shaped RTM.

Similarly, some recent works have also demonstrated the implementation of a two-bit de-multiplexer using a Y-shaped RTM where DWs created in its input branch are sorted into one of the two output branches of a Y-shaped magnetic nanostructure based on their chiralities [74], [75].

E. Epitaxial Racetracks

Recent material developments and techniques have allowed for the exploration of RTM on high-quality epitaxial ferrimagnetic oxides [76], [77] and Heusler compounds with a wide range of fascinating properties. The latter were shown to grow in ultrathin form on technologically relevant silicon substrates using novel chemical templating layers (CTL) [78].

Efficient chiral DW motion was demonstrated in thin (5–8 nm) ferrimagnetic iron garnets. Layers of $Tm_3Fe_5O_{12}$ (TmIG) and $Tb_3Fe_5O_{12}$ (TbIG), having PMA, were epitaxially grown on (111) $Gd_3Ga_5O_{12}$ [76] or $Gd_3Sc_2Ga_3O_{12}$ [77] substrates, paired with a 4–5-nm-thick Pt overlayer. In these systems, besides the interface with the heavy metal Pt, a strong interfacial DMI is found at the oxide interface between the substrate and the ferrimagnetic TmIG or TbIG. Thus, the combined DMI contributions of the top and the bottom interface set the DW chirality. The fastest DW velocities are up to $\sim$800 m s^{-1} with low threshold current densities of 0.04–0.05 $\times$ 10^8 A cm^{-2}. Yet, reducing the high temperatures needed for the growth of these layers (650 °C) and developing methods for epitaxial growth on silicon substrates is a prerequisite for their integration with CMOS electronics.

Another epitaxial system—Heusler materials and compounds—has shown efficient chiral DW motion in their ultrathin form, with growth at room temperature [78]. A CTL method allows even single unit cell thick layers of the low moment ferrimagnetic binary Heusler alloys, Mn_3Z, where Z = Ge, Sn, Sb, to be grown on amorphous underlayers. The two magnetic sublattices, formed from alternating Mn–Mn and Mn–Z atomic layers, are coupled AF. Since these layers are formed from TMs the net (low) moment is weakly dependent on temperature when compared to RE-TM materials.

DW motion in Mn_3Z Heusler racetracks shows a rather complex mechanism where both a volume STT and a chiral SOT drive the DW motion. A bulk derived DMI field H_{DM} sets the chirality of the DWs whose handedness (CW or CCW) is dependent upon the choice of the Heusler compound. The main driving mechanism of the DWs in these Heusler materials is the volume STT which also defines the direction of the DW motion that is determined by the sign of the spin polarization in these materials. A spin current originating from the neighboring nonmagnetic (at room temperature) CTL layer exerts an STT on the magnetic moments of the chiral DWs. The SOT contribution can be tuned and, furthermore, can either be additive or subtractive to the main DW driving mechanism.

The result of the STT on the chiral Néel DWs creates a damping torque that acts on the magnetic moments of the

Table 3 Summary of Spin Polarization Direction, DMI Direction, Spin Hall Angle, and Sot Contribution to the DW Motion Driven Mainly by STT for Mn_3Ge, Mn_3Sn, and Mn_3Sb. Note that H_{DM} is Along the Nanowire Axis

Mn_3Z Heusler	STT direction	chirality	H_{DM} (Oe)	CTL θ_{SH}	SOT contribution
Mn_3Ge	Current flow	CW	1400	positive	unfavorable
Mn_3Sn	Current flow	CCW	-1000	positive	favorable
Mn_3Sb	Electron flow	CW	350	positive	favorable

DW, causing thereby a precessional motion and the DW motion along the electron spin polarization direction. The spin current from the CTL layer (or adjacent nonmagnetic overlayers [78]), owing to the DMI, and the damping torque results in a torque that is always out of the plane of the racetrack layers. The direction of this torque depends on the handedness of the chirality and the spin accumulation of the material-dependent CTL. Thus, the DW experiences a chiral SOT whose direction is influenced by the direction of H_{DM} since the spin Hall angle (θ_{SH}) is set by the CTL. This leads to either an additive or subtractive contribution of the SOT to the DW motion driven mainly by STT. The direction of DW motion in different Mn_3Z Heusler racetracks is summarized in Table 3.

The lowest critical current density that is required to initiate the DW motion of 0.028×10^8 A cm^{-2} was found for Mn_3Sb, which also showed the highest DW velocities among the Mn_3Z Heuslers. Tuning the composition of Mn_xSb, the bulk DMI effective field, can be tuned from ~50 Oe for $Mn_{2.0}Sb$ to ~750 Oe for $Mn_{3.3}Sb$ demonstrating the high tunability of the Heusler materials. Additionally, appropriate CTLs can help meet application needs by tuning the SOT contribution.

IV. RTM APPLICATION IN THE MEMORY SUBSYSTEM

As depicted in Table 1, RTM has a significantly higher capacity and reduced leakage power benefits compared to volatile DRAMs and SRAMs. Compared to emerging nonvolatile memories, i.e., STT-RAM, resistive random access memory (RRAM), PCM and magnetic RAM, RTM has a higher capacity with similar or better energy and access latency behavior. Similarly, RTM is significantly faster than dense vertical NAND (V-NAND) and HDD. Since RTMs have the potential to outperform existing memory technologies in terms of endurance, energy consumption, and storage density, they have received much attention and many research studies have advocated employing them at different levels in the memory hierarchy and for different application domains. This section provides an overview of such proposals.

A. RTM Caches

Previous works investigated caches implemented with RTMs for performance, density, and energy improvements [79]–[83]. Venkatesan *et al.* [79] proposed a prominent RTM-based cache design that provides circuit and architectural level optimizations. The circuit-level design presents two types of DW memory (DWM) bit cells namely 1-b DWM and multibit DWM. The 1-b DWM cell design is optimized for latency as it does not require shift operations. The architectural optimization employs a 1-b DWM cell design for the latency-critical tag array design. The data array is further partitioned into a latency-optimized fast region with a 1-b DWM cell design and a capacity-optimized slow region with multibit DWM cell design. A data migration policy is proposed to dynamically migrate data between fast and slow regions.

Cross-layer optimizations are provided at the last level cache using RTM-based caches that include a cell design, array organization, and application-aware data allocation policy. The cell and array designs mitigate the area gap between RTM storage elements and the large access transistor. The application-aware data allocation policy places frequently accessed data near the access port [80]. A combination of circuit and architectural techniques is proposed in [81] to achieve simultaneous performance, density, and energy enhancements. The proposed circuit-level methods include merged read/write head design to provide high density, flipped-bit cell, and shift gating design for energy improvement, and wordline strapping to optimize latency. To ensure low energy and high reliability, architectural optimization dynamically adapts the shift and write currents while considering the application memory access pattern.

Another cross-layer optimization study performs design space exploration of RTM-based caches at the physical and architectural levels [84]. The physical design includes hybrid-port and uniform-port array designs. The hybrid-port array design contains many narrower read ports and few wider read/write ports. The uniform-port array design only contains read/write ports with different physical layouts. The impact of these aforementioned layouts on different components of energy (e.g., read, write, shift, and leakage) and latency is evaluated. At the architectural level, a combination of mixed array organization is presented which comprises hybrid-port and mixed-port array regions. The regular and read-intensive data are steered to the hybrid-port array region, whereas the mixed port array region is suitable for write-intensive data. To achieve energy efficiency with minimal performance degradation, way-based cache reconfiguration is applied which adapts the cache size based on the application runtime cache requirements. Similarly, set-based cache reconfiguration is applied in [85] to achieve improved energy efficiency.

Highly dense RTM magnetic nanowire storage elements are employed to integrate multiple cache levels in a single cache array and the RTM shift operation is leveraged to switch between different levels [82], [83]. The FusedCache provides a unified L1/L2 cache architecture that stores L1 cache lines exactly at the access port position, whereas the L2 cache lines are not aligned to the access port [82]. As a result, FusedCache architecture provides

constant access latency for L1 cache because access to L1 cache lines does not require any shift operations. In contrast, it provides variable access latency for L2 cache because its access latency depends on the distance of the desired L2 cache line from the port position. The multilane racetrack cache (MRC) architecture synergistically combines the benefits of lightweight compression and fine-grained shifting to mitigate the negative impact of shift operations [83]. The MRC architecture compresses multiple cache lines and stores them in the same DBC, requiring less data storage as well as fewer accesses to domains within a racetrack compared to a conventional uncompressed design. In addition, this article adjusts the starting location of the compressed cache lines within the racetrack which not only reduces the number of shift operations but also allows concurrent accesses to multiple cache lines that belongs to different racetracks.

B. RTM GPU Register File

The immense storage requirement of GPU applications makes RTMs a preferable alternative to be employed as a GPU register file. To this end, various proposals propose RTM-based GPU register files to alleviate the high leakage and scalability problems of conventional SRAM-based register files [86]–[89]. These proposals are based on the tenet of reducing the shift overhead via different techniques that include smart register renaming [86], [87], [89], proactive preshifting [87]–[90], and intelligent thread scheduling [86]. The register renaming technique assigns likely accessed registers closer to the access port to reduce the shift costs. The preshifting policy reduces the shift cost by exploiting the data locality at interthread, intra-SM (SM: streaming multiprocessor), and inter-SM levels. The thread scheduling policy schedules request to register file only when the relevant registers are aligned to their corresponding RTM access ports. Using the RTM GPU register file, the performance gain compared to an iso-area SRAM GPU register file lies in the range 4%–30% (via high density), whereas the energy gain translates to 2–3 times (via reduced leakage).

C. RTM as OFF-Chip Memory

Sun *et al.* [91] provide a cross-layer RTM framework for off-chip memory that explores the design space at device and circuit levels. The device-level design space exploration evaluates the impact of racetrack resistivity by varying the nanowire length for three different materials (CoFe, NiFe, and CoFeB). Similarly, the influence of metal line thickness and distance between magnetic nanowires on the generation of the magnetic field is investigated. The circuit-level design space exploration analyzes the impact of varying the number of ports, cell overlapping, and array partitioning on latency, energy, and the shift distance. Another similar study explores the main memory design based on RTM technology at the circuit and architectural levels [10]. The design space exploration investigates the impact of a number of racetracks, number of domains in each racetrack, number of access ports, subarray size, and cell size on overall area, latency, and energy.

The memory performance critically depends on fast access to metadata. In particular, it is extremely important to provide quick access to the page table, which records virtual-to-physical address mapping information. To reduce page table access latency, recent work [92] rethinks the page table layout in an RTM-based memory which mitigates the number of shifts compared to existing layouts. The new layout places highly accessed fields of a page table entry (PTE) close to the access port. In addition, the RTM shift-aware optimization takes into account different states of a PTE to further reduce the page table latency. This reduction is made possible by proactively preshifting the port position to the desired PTE field in advance based on PTE state prediction. The next state predictor accurately predicts the future PTE state based on the current PTE state. An intelligent RTM-based page table outperforms conventional DRAM-based implementation by 84% and 98% in terms of latency and energy improvements, respectively.

D. RTM as a Disk Replacement

The traditional magnetic disk technology faces many limitations that include speed, durability, and rewritability. For applications with high capacity and speed requirements, RTM memories are a key enabling technology due to their scalability and ultrahigh storage density with the additional advantages that no mechanical parts are necessary [3]. RTMs can be very dense with the usage of 3-D vertical racetrack technology which can be constructed by, for example, atomic layer deposition on the patterned side-walls of deep trenches. The 3-D vertical racetrack technology will realize its true potential by enabling the fabrication of vertical racetracks storing more than 100 bits each of which could enormously increase storage capacity. Therefore, RTM technology has the potential to show improvements in speed, durability, capacity, and cost-per-bit which can be realized with multilayer materials [26], [93], [94] and 3-D vertical racetrack technology. This implies that RTMs can provide much better performance than HDDs. An RTM-based disk substitute may fit into a lapel pin with gigabytes of information storage capability [95]. Recent research replaces traditional magnetic disks by RTMs for graph processing, not only thereby expediting graph processing ($\sim$40%–87% improvement) but also attaining higher energy efficiency ($\sim$13% saving) [96].

E. Processing in Memory (PIM) Using RTM

PIM is a concept in which data computations are performed within memory, directly where the data are stored. The idea is to preprocess the data within memory or near memory (using a computation unit close to memory) instead of transferring a large amount of raw data to an external processor. PIM thus significantly minimizes the data movement penalty by involving the processor only

for summarized data. Additionally, PIM reduces the number of operands transferred to the processor, significantly improving the performance and energy efficiency of the computing system. RTM-based PIM has been demonstrated for lookup table (LUT) and simple logical functions, including XOR, addition, and multiplication [97]. Furthermore, the machine learning operations can be mapped to a PIM architecture by employing a computation unit near memory that provides intermediate results to the processor. DW- and skyrmion-based adders and multipliers for complex convolutional neural networks (CNNs) have been proposed in [98].

Recently, reconfigurable in-memory logic gates are proposed which are based on RTMs [149]–[154]. Employing the basic in-memory logic gates, a multibit magnetic adder design is introduced in [149]. The inputs and the output of the adder are stored in RTMs which act as nonvolatile registers. The use of small nonvolatile RTM cells enables negligible leakage power and small die area compared to a CMOS-based adder architecture. A PIM-based reconfigurable architecture is presented by unifying memory and logical functions using four-terminal RTM cells which exploit the spin Hall effect [150], [151]. In this architecture, the reconfigurable platform is divided into data and logic blocks. The data block simply performs the basic bitwise-XOR operation on the stored data. The logic block performs both bitwise-XOR as well as complex in-memory logic functions. Fast reconfigurable logic gates are realized by storing the computation results in magnetic domains during initial configuration [153]. The DWs are then shifted to implement the desired logical function based on the input data values. A nonvolatile LUT design is presented in [154] by combining the RTM storage unit and CMOS circuit. The LUT enables fast reconfiguration and is composed of a configuration module, multiplexer, and sense amplifier units.

In the Reconfigurable Dual-Mode In-Memory Processing Architecture (RIMPA), the spintronic-based RTM cells can operate in two modes, namely, memory and compute modes [152]. In the memory mode, the RTM cell acts as a normal storage cell. The computing mode enables in-memory logic computations where the RTM cell performs basic logic (i.e., bitwise-AND and bitwise-OR) functions within memory. Similarly, domain-specific in-memory logical functions (bitwise-XOR, sum, carry, and LUT) and HW accelerators are implemented using DW-based nanowires for image processing systems [155]. These HW accelerators are employed near data storage in a distributed fashion to perform frequent compute-intensive operations. In addition, the in-memory HW accelerators enable parallel access to distributed data which significantly improves data parallelism.

V. HW/SW OPTIMIZATIONS FOR RTM

From the architecture perspective, fast and accurate shifting of DWs is the biggest challenge that not only impacts RTM's latency and energy but may also lead to reliability issues [72]. In this section, we discuss HW/SW optimizations that minimize the impact of the shifting operations on RTM performance and energy and improve its reliability.

A. Hardware Techniques for Minimizing Shifts

The straightforward solution to minimize the number of shifts in RTM is to increase the number of access ports. However, this solution quickly becomes impractical due to the additional HW complexity and die-area overhead. The number of shifts can also be reduced with an efficient data-to-port mapping by taking into account the application reuse behavior. For instance, storing frequently accessed data elements near the access ports can significantly reduce the number of shifts [80]. The reuse behavior of different data elements can be predicted using HW monitors in the RTM controller. At runtime, the RTM controller swaps the blocks with the highest frequency with those closer to the access ports.

As mentioned in Section IV-A, RTM caches place some data close to the access port and others further away. Closer DWs have thus a relatively lower latency compared to those farther away. This disparity in RTM latency can be exploited to reduce the number of shifts in an application-specific manner. For instance, some applications demand more cache space compared to other applications. For applications with lesser cache demands, the DWs that are far away from the access ports are disabled and only those closer to the access ports are used which minimizes the total number of shifts without significantly degrading the performance. The cache can be resized based on the application runtime cache demand by turning off/on the active/inactive DWs [84]. In a similar manner, a dynamically reconfigurable cache is proposed in [85].

Literature suggests that the most established technique to improve RTM performance without increasing the number of access ports is preshifting [71], [87], [88]. The concept of preshifting is analogous to prefetching which consists in fetching the data of the next likely accessed element in advance. In the case of shifts, preshifting consists in aligning the access ports to the next likely accessed element. Preshifting can be applied within and across RTM subarrays. Although a DBC is busy serving a memory request, other DBCs can be preshifted proactively.

Other techniques to mitigate shifting overhead include: 1) data compression to reduce the number of bits stored in a racetrack and thereby the shifts overhead [83]; 2) efficient data mapping and dynamic prioritization of the memory requests closer to the access ports [86]; and 3) data swapping and data migration [44], [82]. Although all these techniques improve the overall performance, the total number of shifting operations and energy consumption are rarely affected. For instance, preshifting improves the RTM access latency but may increase its energy consumption. Additionally, all these techniques require additional HW support which not only increases

the HW complexity but also the area utilization and energy consumption.

B. Software Techniques for Minimizing Shift

The most prominent SW solution for RTM shift reduction is a compiler guided intelligent data and instruction placement [13]–[15], [99]. By static code analysis and profiling, the compiler constructs an internal model of the applications' memory access pattern. Based on this model, different techniques are employed to find the best possible mapping of the memory objects to RTM with the objective to minimize the total number of shifts. Exact solutions have been proposed using integer linear programming (ILP) and integer nonlinear programming (INLP) [14], [15], [100]. More computational tractable solutions include meta-heuristics like genetic algorithms and custom heuristics, which deliver near-optimal solutions in considerably less time [15], [45].

SW-controlled scratchpad memory (SPM) is an alternative to caches known for predictable memory access patterns. SPMs feature better performance and energy efficiency at a reduced small chip area and predictable performance. In the context of RTM-based SPM, recent work proposed three heuristics and a genetic algorithm to reduce the RTM shift overhead [45]. The first naïve heuristic adopts the first-come-first-store allocation strategy, which does not perform well for loop accesses. To overcome this problem, the second heuristic allocates frequently accessed data closer to the access port which is located in the middle of the racetrack. To further reduce the shift overhead, the third heuristic applies a greedy algorithm for data allocation where the least frequently accessed data are stored on one end of the racetrack while the most frequently accessed data are placed on the other end. The improved genetic algorithm starts with the results of the three heuristics as the initial population of data mapping (i.e., initial solutions) and applies mutation and crossover with carefully selected mutation elements and crossover points. Experimental results show similar performance to that of an exhaustive search.

While genetic algorithms can take hours and days to compute, heuristic solutions have been reported to effectively minimize the number of shifts in less than a few hundred seconds. The group-based heuristics for data placement in RTM maintain a group of memory objects where a new object is added to the group based on its adjacency with previously added elements in the group [13], [14]. The order of assignment to the group is actually the memory offset assigned to an object. The total shifts are minimized because highly consecutively accessed elements are assigned adjacent positions in the group. The Chen heuristics for data placement in RTM scratchpad finds an ordering of the data items in an access sequence that maximize the likelihood that two consecutive references have minimal shift distance between them [14]. The heuristic models the data placement problem by an undirected edge-weighted access graph, and it exploits the temporal locality of data items to reduce the shift overhead through data grouping. However, the aforementioned heuristics do not effectively reduce the number of end-to-end shifts that are required to move the DW from one end of the track to the other. The heuristic presented in [15] introduces 2-D grouping which further reduces the end-to-end shifts in long racetracks.

The work in [138] investigates the layouts of high-dimensional data structures such as tensors in RTM-based SPMs. For the tensor contraction operation, an optimized data layout reduces the number of shifts by 50% compared to a naïve layout. This improves the performance and energy consumption of the RTM-based SPM by 24% and 74%, respectively, compared to an iso-capacity SRAM. The work in [99] explores RTM as an instruction memory and proposes layouts that best suit the sequential reads/writes of RTM and that of the instruction stream.

C. Improving RTM Reliability

There exists no mechanism in RTM that ensures that DWs are correctly shifted and aligned to the access ports when a shift current is applied. The misalignment of DWs to the access port positions are referred to as position errors [101]–[112]. The typical position error rate in RTM is in the range 10^{-4}–10^{-5} compared to the minimum standard 10^{-19} required for satisfying the required ten-year mean time to failure (MTTF) [111].

Depending on the shift current density and homogeneity of the racetrack, the DWs may be over- or under-shifted. These errors are known as out-of-step/deletion and stop-in-the-middle/insertions errors [109]–[111]. The stop-in-the-middle position errors can be completely eliminated by applying a sub-threshold-shift (STS). An STS consists in applying a shift current (J) with a density less than the critical current (J_0) to the racetrack. The idea is to apply a normal shift current followed by a subsequent STS. If the DWs, for whatever reason, have stopped in the middle, the STS operation enables them to reach the notch regions, otherwise the pinned DW remains unaffected [111].

To detect a single bit out-of-step error, techniques analogous to the parity check can be adopted by employing redundant domains and access ports. Two extra read ports, two guard domains, and $(L-1)$[1] extra domains are needed to correct a single step error and detect two-step errors. In general, $2m$ guard domains and $2m+1$ extra read ports are needed to correct an m-step position error. The position errors are corrected by applying shift current with reverse polarity [111]. Although the position error correction scheme in [111] significantly improves the RTM MTTF, it does not consider the possibility of position errors inside the position error correction code (p-ECC) bits. A slightly improved version of the previous scheme eliminates such

[1] L represents the length of a data segment which is a set of domains that are accessed via a single access port.

errors without incurring any overhead by changing only the mapping of p-ECC bit to the racetrack [110].

The aforementioned ECC techniques [110], [111] suffer from significant area and performance overheads. Every access is performed twice and additional ports are introduced which causes substantial area increase. The codes introduced in [109] completely eliminate the area overhead arises from the additional access ports. The required encoder and decoder consume little power and the codes are easy to implement. By decoupling the error detection from correction, the error correction mechanism is activated only when an error is detected. This decoupling of error detection and correction allows for faster accesses to RTM. The adopted Varshamov–Tenengolts (VT) codes in combination with blocks of delimiter bits can detect up-to two and correct one position errors.

The two types of errors can also be modeled as deletion (out-of-step) and sticky-insertion (stop-in-the-middle) errors. Assuming that each racetrack uses more than one access port, each domain is accessed more than once where the additional reads are used to detect and correct the position errors [101]–[105]. For correcting d deletions, a single extra domain and $d+1$ extra ports are required.

It is worth mentioning that the shift operation in the latest RTM version is much more controlled and accurate compared to earlier versions. In RTM 3.0, the DWs tilt during motion due to the combination of DMI and SHE [113]. This tilting gives rise to inertia of the DW within the first nanoseconds of a shift pulse. Additionally, friction can cause a residual tilt angle after the pulse. If a subsequent pulse into the opposite current direction is applied, the DW would tilt back first before moving. Hence, an asymmetric pulse pattern would be required to avoid shifting errors. In contrast, in RTM 4.0 the tilting in the two AF coupled layers exactly compensates [69]. Consequently, shifting bits comes with almost no inertia and is symmetric for the positive and negative shift direction.

VI. OUTLOOK

To exploit the full potential of RTMs, it is important to consider optimizations in many different directions. This section provides an outlook on potential future studies.

A. Material Research and CMOS Integration

Ferrimagnetic systems have attracted much attention [114] due to their low magnetization which in turn leads to fast magnetization dynamics while they are more robust against perturbations and exhibit efficient DW motion. In RE-TM ferrimagnetic systems, due to the antiferromagnetic coupling of TMs (such as Co, Ni, and Fe) with RE metals (especially Gd and Tb), the ECT mechanism can be used to drive DWs. To maximize the efficiency of this mechanism, the magnetic moments of the two magnetic sublattices need to be such that the overall magnetic layer is at angular momentum compensation ($m_{\mathrm{RE}}/\gamma_{\mathrm{RE}} = m_{\mathrm{TM}}/\gamma_{\mathrm{TM}}$) at the RTM operating temperature. Usually, REs do not exhibit a ferromagnetically ordered state at room temperature but the close interaction with TMs can induce ferromagnetism also at 300 K [115]. Consequently, the use of alloys might be favorable over multilayer structures because the magnetic moments are more intermixed in the former. The thermal stability in these systems remains to be proven for technological applications.

In epitaxial RTMs, well-defined crystalline interfaces in oxides provide a template for a broad range of functionalities and emergent electronic and magnetic properties [116]. CMOS compatibility of the ferrimagnetic iron garnets would require strict specifications on their growth and integration as aforementioned. On the other hand, thermodynamically stable CTLs provide compatibility of Heusler integration with CMOS technologies. Heusler materials are a large family of materials with a wide range of properties that can display low damping and high spin polarization at room temperature and have tunable properties that can be simply varied by changing the Heusler alloy composition. They can exhibit large values of PMA, as shown in their tetragonally distorted forms [117]. Although, the higher resistivities in Heuslers like the Mn_3Z, Z = Ge, Sn, Sb (compared to conventional ferro/ferrimagnetic RTMs) could be a limiting factor for RTM design.

There exist many challenges in the fabrication of 3-D vertical racetracks. Therefore, to ensure the adoption of a 3-D vertical racetrack into commercial products, research into multilayer materials is required that are compatible with silicon and 3-D stacking. However, the 2-D design can already provide many advantages compared to other existing technologies, as discussed in Section IV. In addition to those presented here, another interesting field of application is neuromorphic computing. By using an MTJ which provides readout over the entire racetrack, a multilevel memristor can be realized [118]. In such a design, a DW can be moved to one of several intermediate positions inside the track. As the TMR depends on the relative orientation between the two magnetic layers, the output resistance depends on the position of the DW, which can potentially serve as a magnetic synapse in a neural network, allowing a gradual adjustment of the synaptic weight [119]–[121]. A proof of concept has already been provided [122], [123].

Emerging proposals also suggest the use of skyrmions instead of DWs in an RTM [124], [125]. Skyrmions can be viewed as point-like perturbations in a region of uniform magnetization that exist within a swirl of rotating spins [126]. The direction of rotation has a chirality that is defined by the DMI in the magnetic system. In comparison to DWs, it is expected that skyrmions do not interact with the edges of the wire and are therefore immune to any pinning arising from the edge roughness of the track. The injection and motion of skyrmions have already been demonstrated at room temperature [127], [128]. However, the lateral drift in their motion due to the skyrmion Hall effect and their instability warrants further work in solving these challenges [127], [129].

B. Reducing Threshold Current Density

The inception of RTM research into the materials and physical mechanisms of DW motion has led to a significant reduction in the threshold current density to move DWs. On the one hand, this has been made possible through the discovery of new physical mechanisms that have led to new generations of efficient torques to drive the DW at a higher velocity for the same current density. On the other hand, optimization of the material parameters such as Gilbert damping, gyromagnetic ratio, anisotropy, spin Hall angle, magnetization or the exchange coupling constant of RTM systems remains a promising route in this direction. Finally, improving the quality of the device by reducing edge roughness and crystal defects should give rise to lower threshold current densities. For that, new methods of growing underlayers have to be developed.

The latest research on Heusler structures and ferrimagnetic systems has paved the way for materials with low threshold current densities. One main driving factor in the systems studied to date is the relatively low magnetization at room temperature which decreases the pinning barrier. Depending on the model, a quadratic scaling of the threshold current density with the magnetization is predicted [55]. However, engineering toward lower magnetization materials has to be treated with caution because a lower magnetization also causes a decrease in the thermal stability and consequently the retention period of the device. Instead, finding new spin Hall materials that have a significantly larger spin Hall angle can allow for higher torques while retaining thermal stability. For example, tungsten in the β-phase exhibits a spin Hall angle of almost 50% which is about three times larger than the spin Hall angle of Pt [130].

Beyond heavy metals and their alloys, recently it has been reported that topological insulators [131]–[134] and layered van der Waals TM dichalcogenides [135], [136] give rise to much more efficient SOTs than those observed to date using conventional heavy metals, thereby allowing for the possibility of more efficient magnetization control by electrical currents. Such exotic materials have been reported to exhibit charge to spin current conversions that are an order of magnitude larger than conventional metals. Whether such materials can be readily integrated needs to be further studied along with the experimental demonstration of DW motion from incorporated magnetic layers.

C. Device- and Circuit-Level Investigations

For an earlier version of RTMs, some design space exploration has been carried out at the device level to analyze the impact of various parameters (e.g., nanowire length and resistivity, number and spacing of bits, distance between nanowires, and influence of stray magnetic fields) on different performance metrics (e.g., shift current, energy, area, and speed). However, there is a need to carry out a comprehensive device level investigation for RTM 4.0 that will facilitate the designer to meet the system-level optimization goals and design requirements. Similarly, circuit-level optimizations need to rethought by considering the physics of RTM 4.0. This is required to analyze the influence of various circuit level parameters (e.g., cell, subarray, port, bitline, and wordline layouts) and peripheral circuitry (row/column decoder, sense amplifiers, and write drivers) on overall latency, area, and energy consumption. Finally, existing position-error correction schemes appear (see Section V-C) to be effective but energy consuming. Therefore, exploring new materials such as multiferroic heterostructures [137] could help in improving the reliability of RTMs with lesser energy consumption.

D. HW/SW Codesign

To efficiently exploit the inherent potential of RTM via HW-SW codesign it is necessary to build bridges between: 1) RTM storage; 2) shift-aware memory controller; 3) runtime system (to facilitate data allocation and mapping); and 4) SW layers (i.e., how to abstract RTM characteristics to the application programmer). To realize an effective RTM architecture, it is necessary to explore techniques that exploit the interesting tradeoff between speed and density that can be guided by application, compiler, AND/OR operation system layers. Therefore, the HW–SW codesign is very important for RTM design in order to achieve simultaneous performance and energy efficiency.

In the past, many techniques have been proposed to reduce the shift cost of DW stripe-shaped RTM [45], [79]–[83]. However, there is a lack of the architectural investigation of the skyrmion, ring-shaped, and Y-shaped RTM. Therefore, it is necessary to devise efficient topology-aware (DW- or skyrmion-based) and structure-aware (stripe-shaped, ring-shaped, or Y-shaped) techniques to leverage its true potential. For instance, different RTM topologies and structures differ in their error patterns which need to be analyzed at the architectural level. Similarly, at the compiler level, the memory access patterns of applications can be reordered from higher compiler abstractions, e.g., from a polyhedral model or by additional semantic information from domain-specific languages [138]. There is a need to investigate a runtime system that is flexible to adapt to various flavors of the racetrack (single DW versus multiple DWs; horizontal versus vertical racetrack) memories and different application characteristics (latency versus bandwidth-sensitive applications).

E. Tools for Design Space Exploration

A detailed RTM design space exploration to carry out aforementioned optimizations (see Sections VI-C and VI-D) requires the availability of accurate open-source device-circuit-architecture codesign simulation tools [17] which allow system architects to analyze the limiting parameters and issues of RTM-based memory. Accurate open-source

simulation tools will allow one to analyze the impact of RTM in terms of its functionality, performance, energy, and reliability characteristics before its integration into product systems.

F. RTM as Solid State Drives (SSDs) Replacement

RTM is a promising alternative to existing traditional and emerging memory technologies. Recent research demonstrates that RTM outperforms other technologies at lower levels in the memory hierarchy. However, its potential at the disk level is relatively less explored. Considering its high density, it is extremely important to also study RTM as a possible replacement for SSDs.

At present, SDD-based NAND flash technology is the most prominent alternative to conventional HDDs. After NAND flash was conceived in the latter half of the 1980s [139], it has undergone fundamental breakthroughs in the last two decades. From single bit per cell (b/cell) (SLC) to 2 b/cell (MLC), 3 b/cell (TLC), and now 4 b/cell (QLC), the technology has maintained its scaling pace. The feature size has been reduced from $\sim$100 nm down to $\sim$1 nm and the gross bit storage density (GBSD) has increased by a factor of 2$\times$every two years [1]. However, further reduction in its feature size will lead to processing and reliability challenges. Therefore, research efforts since 2015 have mainly turned to vertical stacking of the planar NAND flash arrays. This 3-D architecture, it is forecast, will drive the growth rate of the technology with the same pace through the next decade [1], [140].

Despite the technological advances, NAND flash memory cannot fulfill the multifaceted requirements of the next-generation data-intensive applications demanding expanded capacity, improved reliability, and lower latencies [141]. As per data published by technology manufacturers at the IEEE ISSCC, the read latency of NAND flash is of the order of tens of microseconds [149]–[155]. A nontrivial increase in the NAND flash latency is observed when going from SLC all the way to QLC technologies. As a result, random accesses to individual cells are extremely costly, thus necessitating sequential accesses to large chunks of data (pages) which typically are in the range of kilobyte sizes (8 and 16 kB in the latest technologies). By contrast, RTM is byte addressable and the read latency of RTM lies in the range of a few nanoseconds to a few hundreds of nanoseconds.

Similarly, the program time of the NAND flash ranges from a few hundred microseconds (in SLC technology) to a few milliseconds (in QLC technologies). The erase operation is performed at the block granularity with typical block sizes of 4 MB. The erase time lies in the millisecond range. In contrast to extreme nonsymmetrical flash technology, RTM does not exhibit significant variation in read/write latencies. In addition, reliability is still the biggest concern in the NAND flash technology. The array endurance in state-of-the-art flash technologies is still in the range of a few program/erase cycles [1]. In contrast, the endurance of RTM technology is equivalent to that of SRAMs and DRAMs.

As mentioned in Section IV-D, the 3-D vertical racetrack technology is a promising candidate to replace SSDs. However, the efficiency of such RTM disk replacement critically depends on its architecture. Such an architecture may hierarchically decompose the data into sector, pages, word, and bytes which can be synchronously read or written. An RTM controller needs to manage different operations that include read, write, and in particular shift operations. Other responsibilities of the RTM controller may include mapping of logical (sector, page, word, byte, etc.) data to the physical (Bank, DBC, racetrack, port, domain, etc.) RTM organization. RTMs allow for an interesting tradeoff between latency and density since the number of DWs in a racetrack can be dynamically varied from 1 (for minimum latency at the cost of density) to maximum (for maximum capacity at the cost of latency). The performance-critical frequently accessed address translation table may be stored in a latency-optimized racetrack with less DWs per racetrack, whereas the data may be stored in capacity-optimized racetrack with more DWs per racetrack. The RTM disk controller may also get useful information via compiler or operating system hints for hot/cold data migration between latency and capacity optimized racetracks.

VII. CONCLUSION

The discoveries of novel current-induced DW motion mechanisms using chiral spintronic phenomena within the last decade have paved the way for bringing RTM to the cusp of application. These developments in spintronics have enabled an order of magnitude improvement in the efficiency with which RTM magnetic bits can be moved. In particular, a recent work on SAFs has realized significantly lower threshold current densities with much higher DW mobilities. Reducing the threshold current further remains an important goal that could be solved, for example, with new materials that give rise to large spin Hall effects, atomic engineering to optimize the fundamental properties of the magnetic layer or entirely new mechanisms.

From the architectural perspective, the development of circuit and architecture level simulators have enabled and expedited RTM research and its exploration at different levels in the memory stack. The intrinsic shift operations in RTM appear to be the biggest challenge and performance bottleneck. However, HW/SW techniques can be employed to minimize the number of shifts or at least mitigate their impact on the overall system's performance. Recent research has demonstrated that RTM, with a carefully designed memory controller for efficient handling of the RTM shifts, can be as fast as SRAM and DRAM while being highly energy efficient. It has been shown that the memory access patterns in various applications can be reordered from higher programming abstractions to minimize the number of RTM shifts.

The many advances in experimental physics and computer architectures highlight the very positive prospects of RTM for imminent technological applications. Key challenges include the reduction of power consumption and device testing on the nanometer scale with the development of racetracks that might include artificial pinning sites to allow for thermally stable and robust DW bits as well as for reliable shifting of trains of closely spaced DW bits. Realizing a 3-D design of RTM is a major technological challenge. However, 2-D RTMs augur a major step forward in memory-storage technology, either as a single layer 2-D RTM or as multiple horizontal racetracks stacked one on top of each other. RTM has applications that range from an ultrafast single DW racetrack that could replace SRAM to ultradense multi-DW, single or multilayer horizontal racetracks that have the potential to replace DRAM and V-NAND. ■

REFERENCES

[1] C. M. Compagnoni, A. Goda, A. S. Spinelli, P. Feeley, A. L. Lacaita, and A. Visconti, "Reviewing the evolution of the NAND flash technology," *Proc. IEEE*, vol. 105, no. 9, pp. 1609–1633, Sep. 2017.

[2] M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable high performance main memory system using phase-change memory technology," *ACM SIGARCH Comput. Archit. News*, vol. 37, no. 3, p. 24, Jun. 2009.

[3] S. Parkin and S.-H. Yang, "Memory on the racetrack," *Nature Nanotechnol.*, vol. 10, no. 3, pp. 195–198, Mar. 2015.

[4] S. S. P. Parkin, M. Hayashi, and L. Thomas, "Magnetic domain-wall racetrack memory," *Science*, vol. 320, no. 5873, pp. 190–194, Apr. 2008.

[5] S. Mittal, "A survey of techniques for architecting processor components using domain-wall memory," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 2, pp. 1–25, Nov. 2016.

[6] S. S. Parkin, "Shiftable magnetic shift register and method of using the same," U.S. Patent 6 834 005, Dec. 21, 2004.

[7] M. Hayashi, L. Thomas, R. Moriya, C. Rettner, and S. S. P. Parkin, "Current-controlled magnetic domain-wall nanowire shift register," *Science*, vol. 320, no. 5873, pp. 209–211, Apr. 2008.

[8] C. Garg, S.-H. Yang, T. Phung, A. Pushp, and S. S. P. Parkin, "Dramatic influence of curvature of nanowire on chiral domain wall velocity," *Sci. Adv.*, vol. 3, no. 5, May 2017, Art. no. e1602804.

[9] C. Zhang, G. Sun, W. Zhang, F. Mi, H. Li, and W. Zhao, "Quantitative modeling of racetrack memory, a tradeoff among area, performance, and power," in *Proc. 20th Asia South Pacific Design Autom. Conf.*, Jan. 2015, pp. 100–105.

[10] Q. Hu, G. Sun, J. Shu, and C. Zhang, "Exploring main memory design based on racetrack memory technology," in *Proc. 26th Great Lakes Symp. VLSI (GLSVLSI)*, 2016, pp. 397–402.

[11] Y. Zhang *et al.*, "Perspectives of racetrack memory for large-capacity on-chip memory: From device to system," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 5, pp. 629–638, May 2016.

[12] R. Venkatesan, "TapeCache: A high density, energy efficient cache based on domain wall memory," in *Proc. ACM/IEEE Int. Symp. Low Power Electron. Design*, Redondo Beach, CA, USA, 2012, pp. 185–190.

[13] X. Chen, "Optimizing data placement for reducing shift operations on domain wall memories," in *Proc. 52nd ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Jun. 2015, pp. 1–6.

[14] X. Chen, E. H.-M. Sha, Q. Zhuge, C. J. Xue, W. Jiang, and Y. Wang, "Efficient data placement for improving data access performance on domain-wall memory," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 10, pp. 3094–3104, Oct. 2016.

[15] A. A. Khan, F. Hameed, R. Bläsing, S. S. Parkin, and J. Castrillon, "Shiftsreduce: Minimizing shifts in racetrack memory 4.0," *ACM Trans. Archit. Code Optim.*, vol. 16, no. 4, pp. 1–23, 2019.

[16] B. Li, "Design and data management for magnetic racetrack memory," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–4.

[17] A. A. Khan, F. Hameed, R. Blasing, S. Parkin, and J. Castrillon, "RTSim: A cycle-accurate simulator for racetrack memories," *IEEE Comput. Archit. Lett.*, vol. 18, no. 1, pp. 43–46, Jan. 2019.

[18] S. Mittal, J. S. Vetter, and D. Li, "A survey of architectural approaches for managing embedded DRAM and non-volatile on-chip caches," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1524–1537, May 2015.

[19] G. Sun, J. Zhao, M. Poremba, C. Xu, and Y. Xie, "Memory that never forgets: Emerging nonvolatile memory and the implication for architecture design," *Nature Sci. Rev.*, vol. 5, no. 4, pp. 577–592, Aug. 2017.

[20] B. K. Kaushik, *Next Generation Spin Torque Memories*. Singapore: Springer, 2017.

[21] J. Meena, S. Sze, U. Chand, and T.-Y. Tseng, "Overview of emerging nonvolatile memory technologies," *Nanosc. Res. Lett.*, vol. 9, no. 1, p. 526, 2014.

[22] T. Coughlin, "Crossing the chasm to new solid-state storage architectures [The art of storage]," *IEEE Consum. Electron. Mag.*, vol. 5, no. 1, pp. 133–142, Dec. 2016.

[23] M. N. Baibich, "Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices," *Phys. Rev. Lett.*, vol. 61, no. 21, p. 2472, 1988.

[24] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange," *Phys. Rev. B*, vol. 39, no. 7, pp. 4828–4830, Mar. 1989.

[25] S. S. P. Parkin, R. Bhadra, and K. P. Roche, "Oscillatory magnetic exchange coupling through thin copper layers," *Phys. Rev. Lett.*, vol. 66, no. 16, pp. 2152–2155, Apr. 1991.

[26] S. S. P. Parkin, N. More, and K. P. Roche, "Oscillations in exchange coupling and magnetoresistance in metallic superlattice structures: Co/ru, Co/cr, and Fe/cr," *Phys. Rev. Lett.*, vol. 64, no. 19, pp. 2304–2307, May 1990.

[27] S. Maekawa and U. Gafvert, "Electron tunneling between ferromagnetic films," *IEEE Trans. Magn.*, vol. 18, no. 2, pp. 707–708, Mar. 1982.

[28] T.-C. Chen and S. S. Parkin, "Method of fabricating data tracks for use in a magnetic shift register memory device," U.S. Patent 6 955 926, Oct. 18, 2005.

[29] T. Ohashi *et al.*, "Variability study with CD-SEM metrology for STT-MRAM: Correlation analysis between physical dimensions and electrical property of the memory element," *Proc. SPIE Metrology, Inspection, Process Control Microlithography*, vol. 10145, Mar. 2017, Art. no. 101450H.

[30] K.-T. Nam *et al.*, "Switching properties in spin transper torque MRAM with sub-5Onm MTJ size," in *Proc. 7th Annu. Non-Volatile Memory Technol. Symp.*, Nov. 2006, pp. 49–51.

[31] S. Ikeda *et al.*, "Tunnel magnetoresistance of 604% at 300K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Appl. Phys. Lett.*, vol. 93, no. 8, Aug. 2008, Art. no. 082508.

[32] S. S. P. Parkin *et al.*, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature Mater.*, vol. 3, no. 12, pp. 862–867, Oct. 2004.

[33] S. V. Pietambaram, "Synthetic antiferromagnet structures for use in MTJs in MRAM technology," U.S. Patent 6 946 697, Sep. 20, 2005.

[34] J. L. Leal and M. H. Kryder, "Spin valves exchange biased by Co/Ru/Co synthetic antiferromagnets," *J. Appl. Phys.*, vol. 83, no. 7, pp. 3720–3723, Apr. 1998.

[35] Y. Huai, F. Albert, P. Nguyen, M. Pakala, and T. Valet, "Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 84, no. 16, pp. 3118–3120, Apr. 2004.

[36] H. Kubota *et al.*, "Evaluation of spin-transfer switching in CoFeB/MgO/CoFeB magnetic tunnel junctions," *Jpn. J. Appl. Phys.*, vol. 44, no. 40, pp. L1237–L1240, Sep. 2005.

[37] J. Hayakawa *et al.*, "Current-driven magnetization switching in CoFeB/MgO/CoFeB magnetic tunnel junctions," *Jpn. J. Appl. Phys.*, vol. 44, no. 41, pp. L1267–L1270, Sep. 2005.

[38] S. S. Parkin and D. E. Heim, "Magnetoresistive spin valve sensor with improved pinned ferromagnetic layer and magnetic recording system using the sensor," U.S. Patent 5 465 185, Nov. 7, 1995.

[39] S. S. P. Parkin and D. Mauri, "Spin engineering: Direct determination of the Ruderman-Kittel-Kasuya-Yosida far-field range function in ruthenium," *Phys. Rev. B, Condens. Matter*, vol. 44, no. 13, pp. 7131–7134, Oct. 1991.

[40] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. A. Buhrman, "Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from the spin Hall effect," *Phys. Rev. Lett.*, vol. 109, no. 9, Aug. 2012, Art. no. 096602.

[41] O. J. Lee *et al.*, "Central role of domain wall depinning for perpendicular magnetization switching driven by spin torque from the spin Hall effect," *Phys. Rev. B, Condens. Matter*, vol. 89, no. 2, Jan. 2014, Art. no. 024418.

[42] O. Alejos, V. Raposo, L. Sanchez-Tejerina, and E. Martinez, "Efficient and controlled domain wall nucleation for magnetic shift registers," *Sci. Rep.*, vol. 7, no. 1, p. 11909, Sep. 2017.

[43] T. Phung *et al.*, "Highly efficient in-line magnetic domain wall injector," *Nano Lett.*, vol. 15, no. 2, pp. 835–841, Jan. 2015.

[44] Z. Sun, W. Wu, and H. Li, "Cross-layer racetrack memory design for ultra high density and low power consumption," in *Proc. 50th Annu. Design Autom. Conf. (DAC)*, May 2013, pp. 1–6.

[45] H. Mao, C. Zhang, G. Sun, and J. Shu, "Exploring data placement in racetrack memory based scratchpad memory," in *Proc. IEEE Non-Volatile Memory Syst. Appl. Symp. (NVMSA)*, Aug. 2015, pp. 1–5.

[46] H. Zhang, C. Zhang, Q. Hu, C. Yang, and J. Shu, "Performance analysis on structure of racetrack memory," in *Proc. 23rd Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2018, pp. 367–374.

[47] M. Hayashi, L. Thomas, C. Rettner, R. Moriya, Y. B. Bazaliy, and S. S. P. Parkin, "Current driven domain wall velocities exceeding the spin angular momentum transfer rate in permalloy nanowires," *Phys. Rev. Lett.*, vol. 98, no. 3, Jan. 2007.

[48] D. Chiba *et al.*, "Control of multiple magnetic domain walls by current in a Co/Ni nano-wire," *Appl. Phys. Express*, vol. 3, no. 7, Jul. 2010, Art. no. 073004.

[49] L. Thomas *et al.*, "Racetrack memory: A high-performance, low-cost, non-volatile memory based on magnetic domain walls," in *IEDM Tech. Dig.*, Dec. 2011, pp. 22–24.

[50] I. M. Miron *et al.*, "Fast current-induced domain-wall motion controlled by the Rashba effect," *Nature Mater.*, vol. 10, no. 6, pp. 419–423, May 2011.

[51] K.-S. Ryu, L. Thomas, S.-H. Yang, and S. Parkin, "Chiral spin torque at magnetic domain walls," *Nature Nanotechnol.*, vol. 8, no. 7, pp. 527–533, Jun. 2013.

[52] K.-S. Ryu, S.-H. Yang, L. Thomas, and S. S. P. Parkin, "Chiral spin torque arising from proximity-induced magnetization," *Nature Commun.*, vol. 5, no. 1, pp. 1–8, May 2014.

[53] S.-H. Yang and S. Parkin, "Novel domain wall dynamics in synthetic antiferromagnets," *J. Phys., Condens. Matter*, vol. 29, no. 30, Jun. 2017, Art. no. 303001.

[54] S. H. Yang, K. S. Ryu, and S. Parkin, "Domain-wall velocities of up to 750 ms^{-1} driven by exchange-coupling torque in synthetic antiferromagnets," *Nature Nanotechnol.*, vol. 10, no. 3, pp. 221–226, 2015.

[55] R. Bläsing *et al.*, "Exchange coupling torque in ferrimagnetic Co/Gd bilayer maximized near angular momentum compensation temperature," *Nature Commun.*, vol. 9, no. 1, pp. 1–8, Nov. 2018.

[56] L. Caretta *et al.*, "Fast current-driven domain walls and small skyrmions in a compensated ferrimagnet," *Nature Nanotechnol.*, vol. 13, no. 12, pp. 1154–1160, Sep. 2018.

[57] Y. Hirata *et al.*, "Correlation between compensation temperatures of magnetization and angular momentum in GdFeCo ferrimagnets," *Phys. Rev. B, Condens. Matter*, vol. 97, no. 22, Jun. 2018.

[58] J. Finley and L. Liu, "Spin-orbit-torque efficiency in compensated ferrimagnetic cobalt-terbium alloys," *Phys. Rev. Appl.*, vol. 6, no. 5, Nov. 2016.

[59] K.-J. Kim *et al.*, "Fast domain wall motion in the vicinity of the angular momentum compensation temperature of ferrimagnets," *Nature Mater.*, vol. 16, no. 12, pp. 1187–1192, Sep. 2017.

[60] R. Mishra, J. Yu, X. Qiu, M. Motapothula, T. Venkatesan, and H. Yang, "Anomalous current-induced spin torques in ferrimagnets near compensation," *Phys. Rev. Lett.*, vol. 118, no. 16, Apr. 2017.

[61] S. A. Siddiqui, J. Han, J. T. Finley, C. A. Ross, and L. Liu, "Current-induced domain wall motion in a compensated ferrimagnet," *Phys. Rev. Lett.*, vol. 121, no. 5, Jul. 2018, Art. no. 057701.

[62] K. Ueda, M. Mann, P. W. P. de Brouwer, D. Bono, and G. S. D. Beach, "Temperature dependence of spin-orbit torques across the magnetic compensation point in a ferrimagnetic TbCo alloy film," *Phys. Rev. B, Condens. Matter*, vol. 96, no. 6, Aug. 2017.

[63] D. Bang *et al.*, "Enhancement of spin Hall effect induced torques for current-driven magnetic domain wall motion: Inner interface effect," *Phys. Rev. B, Condens. Matter*, vol. 93, no. 17, May 2016.

[64] G. Tatara and H. Kohno, "Theory of current-driven domain wall motion: Spin transfer versus momentum transfer," *Phys. Rev. Lett.*, vol. 92, no. 8, Feb. 2004.

[65] A. Malozemoff and J. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, vol. 1. New York, NY, USA: Academic, 1979.

[66] A. Thiaville, Y. Nakatani, J. Miltat, and Y. Suzuki, "Micromagnetic understanding of current-driven domain wall motion in patterned nanowires," *Europhys. Lett.*, vol. 69, no. 6, pp. 990–996, Jan. 2007.

[67] W. Baltensperger and J. S. Helman, "Dry friction in micromagnetics," *IEEE Trans. Magn.*, vol. 27, no. 6, pp. 4772–4774, Nov. 1991.

[68] C. Garg, S.-H. Yang, T. Phung, A. Pushp, and S. S. P. Parkin, "Dramatic influence of curvature of nanowire on chiral domain wall velocity," *Sci. Adv.*, vol. 3, no. 5, May 2017, Art. no. e1602804.

[69] O. Alejos, V. Raposo, L. Sanchez-Tejerina, R. Tomasello, G. Finocchio, and E. Martinez, "Current-driven domain wall dynamics in ferromagnetic layers synthetically exchange-coupled by a spacer: A micromagnetic study," *J. Appl. Phys.*, vol. 123, no. 1, Jan. 2018, Art. no. 013901.

[70] J. H. Franken, H. J. M. Swagten, and B. Koopmans, "Shift registers based on magnetic domain wall ratchets with perpendicular anisotropy," *Nature Nanotechnol.*, vol. 7, no. 8, pp. 499–503, Jul. 2012.

[71] R. Venkatesan, "DWM-TAPESTRI—an energy efficient all-spin cache using domain wall shift based writes," in *Proc. Design, Autom. Test Eur. Conf. Exhib.*, Mar. 2013, pp. 1825–1830.

[72] Y. Zhang *et al.*, "Ring-shaped racetrack memory based on spin orbit torque driven chiral domain wall motions," *Sci. Rep.*, vol. 6, no. 1, Oct. 2016, Art. no. 35062.

[73] G. Wang *et al.*, "Ultra-dense ring-shaped racetrack memory cache design," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 1, pp. 215–225, Jan. 2019.

[74] T. Phung, A. Pushp, C. Rettner, B. P. Hughes, S.-H. Yang, and S. S. P. Parkin, "Robust sorting of chiral domain walls in a racetrack biplexer," *Appl. Phys. Lett.*, vol. 105, no. 22, Dec. 2014, Art. no. 222404.

[75] A. Pushp *et al.*, "Domain wall trajectory determined by its fractional topological edge defects," *Nature Phys.*, vol. 9, no. 8, pp. 505–511, Jul. 2013.

[76] C. O. Avci *et al.*, "Interface-driven chiral magnetism and current-driven domain walls in insulating magnetic garnets," *Nature Nanotechnol.*, vol. 14, no. 6, pp. 561–566, Apr. 2019.

[77] S. Vélez *et al.*, "High-speed domain wall racetracks in a magnetic insulator," 2019, *arXiv:1902.05639*. [Online]. Available: http://arxiv.org/abs/1902.05639

[78] P. C. Filippou *et al.*, "Chiral domain wall motion in unit-cell thick perpendicularly magnetized heusler films prepared by chemical templating," *Nature Commun.*, vol. 9, no. 1, p. 4653, Nov. 2018.

[79] R. Venkatesan *et al.*, "Cache design with domain wall memory," *IEEE Trans. Comput.*, vol. 65, no. 4, pp. 1010–1024, Apr. 2016.

[80] Z. Sun, X. Bi, W. Wu, S. Yoo, and H. Li, "Array organization and data management exploration in racetrack memory," *IEEE Trans. Comput.*, vol. 65, no. 4, pp. 1041–1054, Apr. 2016.

[81] S. Motaman, A. S. Iyengar, and S. Ghosh, "Domain wall memory-layout, circuit and synergistic systems," *IEEE Trans. Nanotechnol.*, vol. 14, no. 2, pp. 282–291, Mar. 2015.

[82] H. Xu, Y. Alkabani, R. Melhem, and A. K. Jones, "FusedCache: A naturally inclusive, racetrack memory, dual-level private cache," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 2, no. 2, pp. 69–82, Apr. 2016.

[83] H. Xu, Y. Li, R. Melhem, and A. K. Jones, "Multilane racetrack caches: Improving efficiency through compression and independent shifting," in *Proc. 20th Asia South Pacific Design Autom. Conf.*, Jan. 2015, pp. 417–422.

[84] Z. Sun, X. Bi, A. K. Jones, and H. Li, "Design exploration of racetrack lower-level caches," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, Aug. 2014, pp. 263–266.

[85] A. Ranjan, S. G. Ramasubramanian, R. Venkatesan, V. Pai, K. Roy, and A. Raghunathan, "DyReCTape: A dynamically reconfigurable cache using domain wall memory tapes," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Mar. 2015, pp. 181–186.

[86] M. Mao, W. Wen, Y. Zhang, Y. Chen, and H. Li, "Exploration of GPGPU register file architecture using domain-wall-shift-write based racetrack memory," in *Proc. 51st ACM/EDAC/IEEE Des. Autom. Conf. (DAC)*, Jun. 2014, pp. 1–6.

[87] S. Wang *et al.*, "Performance-centric register file design for GPUs using racetrack memory," in *Proc. 21st Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2016, pp. 25–30.

[88] E. Atoofian, "Reducing shift penalty in domain wall memory through register locality," in *Proc. Int. Conf. Compil., Archit. Synth. Embedded Syst. (CASES)*, Oct. 2015, pp. 177–186.

[89] Y. Liang and S. Wang, "Performance-centric optimization for racetrack memory based register file on GPUs," *J. Comput. Sci. Technol.*, vol. 31, no. 1, pp. 36–49, Jan. 2016.

[90] M. Moeng, H. Xu, R. Melhem, and A. K. Jones, "ContextPreRF: Enhancing the performance and energy of GPUs with nonuniform register access," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 1, pp. 343–347, Jan. 2016.

[91] G. Sun *et al.*, "From device to system: Cross-layer design exploration of racetrack memory," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Mar. 2015, pp. 1018–1023.

[92] H. A. Khouzani, P. Fotouhi, C. Yang, and G. R. Gao, "Leveraging access port positions to accelerate page table walk in DWM-based main memory," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Mar. 2017, pp. 1405–1455.

[93] S. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, and M. Samant, "Magnetically engineered spintronic sensors and memory," *Proc. IEEE*, vol. 91, no. 5, pp. 661–680, May 2003.

[94] S. S. P. Parkin, "Systematic variation of the strength and oscillation period of indirect magnetic exchange coupling through the 3d, 4d, and 5d transition metals," *Phys. Rev. Lett.*, vol. 67, no. 25, pp. 3598–3601, Dec. 1991.

[95] (2010). *IBM100. Racetrack Memory (The Future of Data Storage)*. Accessed: Apr. 15, 2019. [Online]. Available: https://www.ibm.com/ibm/history/ibm100/us/en/icons/racetrack/

[96] E. Park, S. Yoo, S. Lee, and H. Li, "Accelerating graph computation with racetrack memory and pointer-assisted graph representation," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, 2014, pp. 1–4.

[97] Y. Wang *et al.*, "An energy-efficient nonvolatile in-memory computing architecture for extreme learning machine by domain-wall nanowire devices," *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 998–1012, Nov. 2015.

[98] B. Liu *et al.*, "An efficient racetrack memory-based Processing-in-memory architecture for convolutional neural networks," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl., IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, Dec. 2017, pp. 383–390.

[99] J. Multanen, P. Jaaskelainen, A. A. Khan, F. Hameed, and J. Castrillon, "SHRIMP: Efficient instruction delivery with domain wall memory," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2019, pp. 1–6.

[100] S. Gu, E. H.-M. Sha, Q. Zhuge, Y. Chen, and J. Hu, "A time, energy, and area efficient domain wall memory-based SPM for embedded systems," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 35, no. 12, pp. 2008–2017, Dec. 2016.

[101] Y. M. Chee, R. Gabrys, A. Vardy, V. K. Vu, and E. Yaakobi, "Reconstruction from deletions in racetrack memories," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2018, pp. 1–5.

[102] Y. M. Chee, H. M. Kiah, A. Vardy, K. Van Vu, and E. Yaakobi, "Codes correcting limited-shift errors in racetrack memories," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 96–100.

[103] Y. M. Chee, H. M. Kiah, A. Vardy, V. K. Vu, and E. Yaakobi, "Codes correcting under-and over-shift errors in racetrack memories," in *Proc. NVMW*, Apr. 2019, pp. 1–2.

[104] Y. M. Chee, "Codes correcting position errors in racetrack memories," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 161–165.

[105] Y. M. Chee, H. M. Kiah, A. Vardy, V. K. Vu, and E. Yaakobi, "Coding for racetrack memories," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7094–7112, Nov. 2018.

[106] T. Luo, "A racetrack memory based in-memory booth multiplier for cryptography application," in *Proc. 21st Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2016, pp. 286–291.

[107] H. Mahdavifar and A. Vardy, "Asymptotically optimal sticky-insertion-correcting codes with efficient encoding and decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2017, pp. 2683–2687.

[108] R. Shibata, G. Hosoya, and H. Yashima, "Joint iterative decoding of spatially coupled low-density parity-check codes for position errors in racetrack memories," *IEICE Trans. Fundam. Electron.*

Commun. Comput. Sci., vol. 101, no. 12, pp. 2055–2063, 2018.
[109] A. Vahid, G. Mappouras, D. J. Sorin, and R. Calderbank, "Correcting two deletions and insertions in racetrack memory," 2017, *arXiv:1701.06478*. [Online]. Available: http://arxiv.org/abs/1701.06478
[110] X. Wang, C. Zhang, X. Zhang, and G. Sun, "Np-ECC: Nonadjacent position error correction code for racetrack memory," in *Proc. IEEE/ACM Int. Symp. Nanosci. Archit. (NANOARCH)*, Jul. 2016, pp. 23–24.
[111] C. Zhang *et al.*, "Hi-fi playback: Tolerating position errors in shift operations of racetrack memory," in *Proc. 42nd Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2015, pp. 694–706.
[112] H. Zhang, C. Zhang, X. Zhang, G. Sun, and J. Shu, "Pin tumbler lock: A shift based encryption mechanism for racetrack memory," in *Proc. 21st Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2016, pp. 354–359.
[113] O. Boulle *et al.*, "Domain wall tilting in the presence of the Dzyaloshinskii-Moriya interaction in out-of-plane magnetized magnetic nanotracks," *Phys. Rev. Lett.*, vol. 111, no. 21, Nov. 2013, Art. no. 217203.
[114] J. Sinova, T. Jungwirth, and O. Gomonay, "Antiferromagnetic spintronics," *Phys. Status Solidi*, vol. 11, no. 4, Apr. 2017, Art. no. 1770322.
[115] T. Ha Pham *et al.*, "Very large domain wall velocities in Pt/Co/GdOx and Pt/Co/Gd trilayers with Dzyaloshinskii-Moriya interaction," *Euro Phys. Lett.*, vol. 113, no. 6, p. 67001, Apr. 2016.
[116] J. Mannhart and D. Schlom, "Oxide interfaces-an opportunity for electronics," *Science*, vol. 327, no. 5973, pp. 1607–1611, 2010.
[117] S. V. Faleev, Y. Ferrante, J. Jeong, M. G. Samant, B. Jones, and S. S. P. Parkin, "Origin of the tetragonal ground state of heusler compounds," *Phys. Rev. Appl.*, vol. 7, no. 3, Mar. 2017, Art. no. 034022.
[118] X. Wang, Y. Chen, H. Xi, H. Li, and D. Dimitrov, "Spintronic memristor through spin-torque-induced magnetization motion," *IEEE Electron Device Lett.*, vol. 30, no. 3, pp. 294–297, Mar. 2009.
[119] J. Cai, B. Fang, C. Wang, and Z. Zeng, "Multilevel storage device based on domain-wall motion in a magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 111, no. 18, Oct. 2017, Art. no. 182410.
[120] J. Grollier, D. Querlioz, and M. D. Stiles, "Spintronic nanodevices for bioinspired computing," *Proc. IEEE*, vol. 104, no. 10, pp. 2024–2039, Oct. 2016.
[121] J. Münchenberger, G. Reiss, and A. Thomas, "A memristor based on current-induced domain-wall motion in a nanostructured giant magnetoresistance device," *J. Appl. Phys.*, vol. 111, no. 7, Apr. 2012, Art. no. 07D303.
[122] S. Lequeux *et al.*, "A magnetic synapse: Multilevel spin-torque memristor with perpendicular anisotropy," *Sci. Rep.*, vol. 6, no. 1, pp. 1–7, Aug. 2016.
[123] A. Chanthbouala *et al.*, "Vertical-current-induced domain-wall motion in MgO-based magnetic tunnel junctions with low current densities," *Nature Phys.*, vol. 7, no. 8, pp. 626–630, Apr. 2011.
[124] J. Sampaio, V. Cros, S. Rohart, A. Thiaville, and A. Fert, "Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures," *Nature Nanotechnol.*, vol. 8, no. 11, pp. 839–844, Oct. 2013.
[125] N. Nagaosa and Y. Tokura, "Topological properties and dynamics of magnetic skyrmions," *Nature Nanotechnol.*, vol. 8, no. 12, pp. 899–911, Dec. 2013.
[126] U. K. Rößler, A. N. Bogdanov, and C. Pfleiderer, "Spontaneous skyrmion ground states in magnetic metals," *Nature*, vol. 442, no. 7104, pp. 797–801, Aug. 2006.
[127] S. Woo *et al.*, "Observation of room-temperature magnetic skyrmions and their current-driven dynamics in ultrathin metallic ferromagnets," *Nature Mater.*, vol. 15, no. 5, pp. 501–506, Feb. 2016.
[128] W. Jiang *et al.*, "Blowing magnetic skyrmion bubbles," *Science*, vol. 349, no. 6245, pp. 283–286, Jun. 2015.
[129] W. Jiang *et al.*, "Direct observation of the skyrmion Hall effect," *Nature Phys.*, vol. 13, no. 2, pp. 162–169, Sep. 2016.
[130] K.-U. Demasius *et al.*, "Enhanced spin–orbit torques by oxygen incorporation in tungsten films," *Nature Commun.*, vol. 7, no. 1, Feb. 2016, Art. no. 10644.
[131] J. Han, A. Richardella, S. A. Siddiqui, J. Finley, N. Samarth, and L. Liu, "Room-temperature spin-orbit torque switching induced by a topological insulator," *Phys. Rev. Lett.*, vol. 119, no. 7, Aug. 2017, Art. no. 077702.
[132] A. R. Mellnik *et al.*, "Spin-transfer torque generated by a topological insulator," *Nature*, vol. 511, no. 7510, pp. 449–451, Jul. 2014.
[133] M. Dc *et al.*, "Room-temperature high spin-orbit torque due to quantum confinement in sputtered Bi_x $Se_{(1-x)}$ films," *Nature Mater.*, vol. 17, no. 9, pp. 800–807, Jul. 2018.
[134] H. Wu *et al.*, "Spin–orbit torque switching of a nearly compensated ferrimagnet by topological surface states," *Adv. Mater.*, vol. 31, no. 35, Jul. 2019, Art. no. 1901681.
[135] S. Shi *et al.*, "All-electric magnetization switching and Dzyaloshinskii–Moriya interaction in WTe_2/ferromagnet heterostructures," *Nature Nanotechnol.*, vol. 14, no. 10, pp. 945–949, Aug. 2019.
[136] M. H. D. Guimaräes, G. M. Stiehl, D. MacNeill, N. D. Reynolds, and D. C. Ralph, "Spin–orbit torques in $NbSe_2$/permalloy bilayers," *Nano Lett.*, vol. 18, no. 2, pp. 1311–1316, Jan. 2018.
[137] N. Lei *et al.*, "Strain-controlled magnetic domain wall propagation in hybrid piezoelectric/ferromagnetic structures," *Nature Commun.*, vol. 4, no. 1, p. 1378, Jan. 2013.
[138] A. A. Khan, "Optimizing tensor contractions for embedded devices with racetrack memory scratch-pads," in *Proc. 20th ACM SIGPLAN/SIGBED Int. Conf. Lang., Compil., Tools Embedded Syst.*, Phoenix, AZ, USA, 2019, pp. 5–18.
[139] F. Masuoka, M. Momodom, Y. Iwata, and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," in *IEDM Tech. Dig.*, Dec. 1987, pp. 552–555.
[140] N. Shibata, "13.1 A 1.33Tb 4-bit/cell 3D-flash memory on a 96-word-line-layer technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 210–212.
[141] G. Fettweis, K. Leo, B. Voit, U. Schneider, and L. Scheuvens, "Venturing electronics into unknown grounds," in *IEDM Tech. Dig.*, Dec. 2018, pp. 1–2.
[142] M. Helm, "19.1 A 128Gb MLC NAND-flash device using 16nm planar cell," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*. Feb. 2014, pp. 326–327.
[143] D. Nobunaga *et al.*, "A 50nm 8Gb NAND flash memory with 100MB/s program throughput and 200MB/s DDR interface," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2008, pp. 426–625.
[144] R. Zeng, "A 172mm^2 32Gb MLC NAND flash memory in 34nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2009, pp. 236–237.
[145] R. Yamashita *et al.*, "11.1 a 512Gb 3b/cell flash memory on 64-word-line-layer BiCS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 196–197.
[146] H. Maejima *et al.*, "A 512Gb 3b/Cell 3D flash memory on a 96-word-line-layer technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 336–338.
[147] D. Kang, "13.4 A 512Gb 3-bit/Cell 3D 6 th-generation V-NAND flash memory with 82MB/s write throughput and 1.2 Gb/s interface," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Paper*, Feb. 2019, pp. 216–218.
[148] C. Siau *et al.*, "13.5 A 512Gb 3-bit/cell 3D flash memory on 128-wordline-layer with 132MB/s write performance featuring circuit-under-array technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Paper*, Feb. 2019, pp. 218–220.
[149] H.-P. Trinh, W. Zhao, J.-O. Klein, Y. Zhang, D. Ravelsona, and C. Chappert, "Magnetic adder based on racetrack memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 6, pp. 1469–1477, Jun. 2013.
[150] S. Angizi, Z. He, and D. Fan, "Energy efficient in-memory computing platform based on 4-terminal spin Hall effect-driven domain wall motion devices," in *Proc. Great Lakes Symp. VLSI (GLSVLSI)*, New York, NY, USA, 2017, pp. 77–82.
[151] S. Angizi, Z. He, N. Bagherzadeh, and D. Fan, "Design and evaluation of a spintronic in-memory processing platform for nonvolatile data encryption," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 9, pp. 1788–1801, Sep. 2018.
[152] S. Angizi, Z. He, F. Parveen, and D. Fan, "RIMPA: A new reconfigurable dual-mode in-memory processing architecture with spin Hall effect-driven domain wall motion device," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, Bochum, Germany, 2017, pp. 45–50.
[153] K. Huang and R. Zhao, "Magnetic domain-wall racetrack memory-based nonvolatile logic for low-power computing and fast run-time-reconfiguration," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 9, pp. 2861–2872, Sep. 2016.
[154] W. Zhao, N. Ben Romdhane, Y. Zhang, J. Klein, and D. Ravelosona, "Racetrack memory based reconfigurable computing," in *Proc. IEEE Faible Tension Faible Consommation*, Paris, France, 2013, pp. 1–4.
[155] Y. Wang *et al.*, "An energy-efficient nonvolatile in-memory computing architecture for extreme learning machine by domain-wall nanowire devices," *IEEE Trans. Nanotechnol.*, vol. 14, pp. 998–1012, 2015, doi: 10.1109/TNANO.2015.2447531.

ABOUT THE AUTHORS

Robin Bläsing received the bachelor's and master's degrees in physics and the master's degree in management, business and economics from RWTH Aachen University, Aachen, Germany, in 2012, 2014, and 2015, respectively, and the Ph.D. (Dr.rer.nat.) degree (Honors) in physics from Martin-Luther-University Halle-Wittenberg, Halle, Germany, in 2019.

For his master thesis, which was related to racetrack memory, he served as an Intern at IBM Almaden Research Center, San Jose, CA, USA. He is currently a Research Scientist with the Max Planck Institute of Microstructure Physics, Halle, working on the development of racetrack memory.

Dr. Bläsing was listed on the RWTH Dean's List and received the Springorum-Coin. In 2011, he also received an ERASMUS Grant for an exchange at KTH, Stockholm, Sweden.

Asif Ali Khan received the bachelor's and master's degrees in computer systems engineering from the University of Engineering and Technology at Peshawar, Peshawar, Pakistan, in 2012 and 2015, respectively. He is working toward the Ph.D. degree at the Chair for Compiler Construction, Computer Science Department, TU Dresden, Dresden, Germany.

Most recently, he has started working on developing new simulation and compilation tools for racetrack memories that enable their exploration at various levels in the memory subsystems for different application domains. His research interests include computer architecture, heterogeneous memories, and compiler support for memory systems.

Panagiotis Ch. Filippou received the bachelor's degree in applied physics from the National Technical University of Athens, Athens, Greece, in 2010, the M.S. degree from the Institut Polytechnique de Grenoble, Grenoble, France, in 2013, and the Ph.D. degree (Honors) in physics from Martin-Luther-University Halle-Wittenberg, Halle, Germany, in 2019.

He is currently a Research Scientist at IBM Research–Almaden, San Jose, CA, USA. His research interests include material science for spintronics, Heuslers, racetrack, and magnetic random access memory (MRAM) applications for future generation nonvolatile memories.

Chirag Garg received the bachelor's and master's degrees in ceramic engineering from IIT (BHU) Varanasi, Varanasi, India, in 2014 (joint program), and the Ph.D. degree from the Max Planck Institute of Microstructure Physics, Halle, Germany, in 2019.

He is currently a Research Scientist at IBM Research–Almaden, San Jose, CA, USA, working on MRAM. His key focus is the development of a new generation of spintronic devices for high-speed memory and computing applications.

Fazal Hameed received the Ph.D. (Dr.Ing.) degree in computer science from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2015.

He joined the Chair for Compiler Construction, TU Dresden, Dresden, Germany, as Postdoctoral Researcher, in March 2016. Before that, he worked on a similar position at the Chair of Dependable and Nano Computing (CDNC), Karlsruhe Institute of Technology (KIT), where he mainly worked with the Architecture Group, with a focus on memories. At the Chair for Compiler Construction, he is currently working on the development of a simulation framework to evaluate the performance, energy, and reliability of heterogenous multicore system architecture.

Dr. Hameed was a recipient of the CODES+ISSS 2013 Best Paper Nomination for his work on DRAM cache management in multicore systems. He has served as an external reviewer for major conferences in embedded systems and computer architecture.

Jeronimo Castrillon (Senior Member, IEEE) received the Electronics Engineering degree from Pontificia Bolivariana University, Medellín, Colombia, in 2004, the master's degree from the Advanced Learning and Research Institute, Lugano, Switzerland, in 2006, and the Ph.D. degree (Dr.Ing.) (Honors) from RWTH Aachen University, Aachen, Germany, in 2013.

In 2014, he co-founded Silexica GmbH/Inc., Cologne, Germany, a company that provides programming tools for embedded multicore architectures. He is currently a Professor with the Department of Computer Science, TU Dresden, Dresden, Germany, where he is also with the Center for Advancing Electronics Dresden (CfAED). He has more than 80 international publications. His research interests lie in methodologies, languages, tools, and algorithms for programming complex computing systems.

Dr. Castrillon was a Founding Member from 2017 to 2019 of the Executive Committee of the ACM Future of Computing Academy (FCA). He is also a member of ACM. He has been a member of technical program and organization committees in international conferences and workshops (e.g., DAC, DATE, ESWEEK, CGO, LCTES, Computing Frontiers, ICCS, and FPL). He is also a regular reviewer for ACM and IEEE journals: the IEEE Transactions on Computer-Aided Design, the IEEE Transactions on Parallel and Distributed Systems, the *ACM Transactions on Design Automation of Electronic Systems*, and the *ACM Transactions on Embedded Computing Systems*.

Stuart S. P. Parkin is currently the Managing Director of the Max Planck Institute for Microstructure Physics, Halle, Germany, and an Alexander von Humboldt Professor with Martin-Luther-University Halle-Wittenberg, Halle. He also proposed magnetic random access memory based on magnetic tunnel junctions in 1995 which he and his colleagues at IBM Research–Almaden, San Jose, CA, USA, demonstrated in 1999, and which has recently become available as a mainstream foundry technology. In 2002, he also proposed Racetrack Memory that is the subject of this article. His research interests include spintronic materials and devices for advanced sensor, memory, and logic applications, oxide thin-film heterostructures, topological metals, exotic superconductors, and cognitive devices. His discoveries in spintronics enabled a more than 10 000-fold increase in the storage capacity of magnetic disk drives. He has published approximately 570 articles and 118 issued patents and an impact factor H = 112 (Google Scholar).

Dr. Parkin is a Fellow/Member of the Royal Society, the Royal Academy of Engineering, the National Academy of Sciences, the National Academy of Engineering, the German National Academy of Science—-Leopoldina, the Royal Society of Edinburgh, the Indian Academy of Sciences, and Third World Academy of Sciences (TWAS)—Academy of Sciences for the developing world. For this work, which thereby enabled the "big data" world of today, he was awarded the Millennium Technology Award from the Technology Academy Finland in 2014 (worth 1 000 000 Euro). His awards include the American Physical Society International Prize for New Materials in 1994, the Europhysics Prize for Outstanding Achievement in Solid State Physics in 1997, the 1999–2000 American Institute of Physics Prize for Industrial Application of Physics, the 2008 IEEE Daniel E. Noble Award, the 2009 IUPAP Magnetism Prize and Neel Medal for outstanding contributions to the science of magnetism, the 2012 David Adler Lectureship Award, the 2012 von Hippel Award from the Materials Research Society, the 2013 Swan Medal of the Institute of Physics, London, the Alexander von Humboldt Professorship—International Award for Research in 2014, and the European Research Council (ERC) Advanced Grant—SORBET in 2015.