From Device to System: Cross-layer Design Exploration of Racetrack Memory

Guangyu Sun*, Chao Zhang*, Hehe Li†, Yue Zhang§, Weiqi Zhang*, Yizi Gu†,
Yinan Sun†, J.-O. Klein§, D. Ravelosona§, Yongpan Liu†, Weisheng Zhao‡§, Huazhong Yang†
*CECA, Peking University, Beijing, China
†EE Department, Tsinghua University, Beijing, China
‡Spintronics Interdisciplinary Center, Beihang University, Beijing, China
§Institut d'lectronique Fondamentale, Univ. Paris-Sud/UMR 8622 CNRS, Orsay, France
Email: * gsun@pku.edu.cn, † ypliu@tsinghua.edu.cn, ‡ weisheng.zhao@u-psud.fr

***Abstract*—Recently, Racetrack Memory (RM) has attracted more and more attention of memory researchers because it has advantages of ultra-high storage density, fast access speed, and non-volatility. Prior research has demonstrated that RM has potential to replace SRAM for large capacity on-chip memory design. At the same time, it also addressed that the design space exploration of RM could be more complicated compared to traditional on-chip memory technologies for several reasons. First, a single RM cell introduces more device level design parameters. Second, considering these device-level design factors, the layout exploration of a RM array demonstrates trade-off among area, performance, and power consumption of RM circuit level design. Third, in the architecture level, the unique "shift" operation results in an extra dimension for design exploration. In this paper, we will review all these design issues in different layers and try to reveal the relationship among them. The experimental results demonstrate that cross-layer design exploration is necessary for racetrack memory. In addition, a system level case study of using RM in a sensor node is presented to demonstrate its advantages over SRAM or STT-RAM.**

I. Introduction

As the number of process elements integrated on a single chip keeps increasing, on-chip memory design is in urgent demand of improving storage density to cache enough data for processing. Thus, various emerging memory technologies have been proposed as potential candidates of replacing traditional SRAM technology for future on-chip memory design. They include Spin-Transfer Torque Random-Access Memory (STT-RAM) [1], Resistive Random-Access Memory (ReRAM) [2], Conductive Bridging Random-Access Memory (CBRAM) [3], etc. Compared to SRAM, these emerging memory technologies have advantages of non-volatility, high storage density, and low leakage power [1], [4].

Recently, Racetrack Memory (RM) has attracted more and more attention of memory researchers because it can achieve even higher storage density than the other emerging non-volatile memory technologies (NVMs) introduced above. RM can be considered as a new generation of spintronic based memory technology. It can achieve ultra-high storage density by integrating many domains in a nanowire [5]. Thus, a RM storage cell is in the form of a tape-like structure. All these domains in a storage cell share several access ports for read and write operations. To this end, a domain needs to be shifted to the position of an access port before being accessed. Though the shift operation induces overhead of latency and energy consumption, we can still benefit from the ultra-high storage density of RM. This conclusion has been proved in previous research [6], [7], [8].

Though previous research has demonstrated benefits of using RM for on-chip memory design, we believe the advantages of RM is not full exploited. It is mainly because the design space of a RM design is so huge that the design trade-off must be carefully considered for different design goals. In fact, the design space of a RM is much larger than those of existing technologies for several reasons. First, a single RM cell introduces more device level design parameters. Second, considering these device-level design factors, the layout exploration of a RM array demonstrates trade-off among area, performance, and power consumption of RM circuit level design. Third, in the architecture level, the unique shift operation results in an extra dimension for design exploration.

In this work, we will explore the design space of a RM design in different layers ranging from the device level to the system level. The important design issues in different layers are discussed and quantitatively evaluated to reflect their impacts on design space. In addition, we try to reveal the interaction among them and argue that a cross-layer design exploration is critical to find a proper RM design for different goals. The rest of this paper is organized as follows. The device level, circuit level, and architecture level design exploration are presented in Section II, III, and IV, respectively. In Section V, we further provide a case study of using RM in a ultra-low power NV-processor, followed by conclusions in the last section.

II. Device Level Basics of Racetrack Memory

The structural concept of a racetrack memory cell is shown in Fig. 1. It is composed of three basic parts work for different operations. Write and read heads are used for data input and output. Magnetic nanowires is used for data storage and transfer [9], [10] In order to integrate with peripheral CMOS circuits, the write and read heads are always designed through magnetic tunnel junctions (MTJs). Such a design can also help to make fast operation and low power feasible [11]. The data transfer is based on current-induced DW motions [12]. The mainstream controlling strategy is to create notches or constrictions for DW pinning. The extremely scaled distance among these artificial pinning defects can result in considerable device storage density. In addition, three currents, Iw, Ir and Ish, flowing in three different paths, which are applied to perform DW nucleation, sensing, and shifting respectively. This could improve the reliability of the global device compared with the traditional two-terminal MTJ structure. The first prototype fabricated in 2011 confirmed its feasibility, although a lot of breakthroughs in terms of material and technique remain to overcome [13]. For example,

978-3-9815370-4-8/DATE15/©2015 EDAA

its limited thermal stability due to the material NiFe with in-plane magnetic anisotropy used in the prototype hinders its further miniaturization. The racetrack memory based on CoFeB/MgO with perpendicular magnetic anisotropy (PMA) was then proposed, [14] which enables high thermal stability and unprecedented performances on DW writing and sensing.

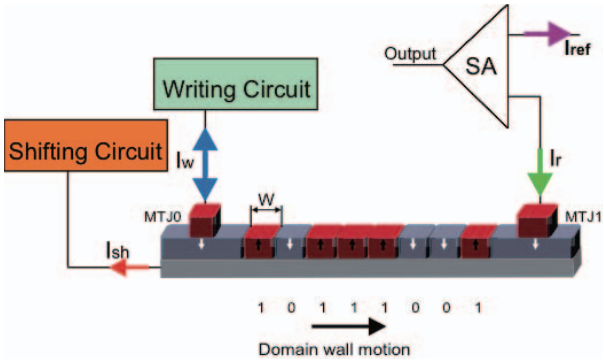

Fig. 1. Racetrack memory: write head (MTJ0), read head (MTJ1) and one magnetic nanowire. Writing circuit generates Iw to input data, shifting circuit generates Ish to induce DW motion and sense amplifier (SA) generates Ir to output storage data.

The capacity is one of the most critical issues for racetrack memory design. As the current-induced DW motions in racetrack memory are traditionally described by spin transfer torque (STT). The DW velocity is related to the applied current, which can be given as follows:

$$V = \frac{\beta}{\alpha} u \tag{1}$$

$$u = \frac{\mu_B P j}{e M_s} \quad (j > j_c), \tag{2}$$

where j is the applied current density, j_c is the critical current density, u is spin current velocity, α is the Gilbert damping constant, β is the dissipative correction to the STT, μ_B is the Bohr magneton, M_s is the saturation magnetization, e is the electron mass, P is the spin polarization percentage of tunnel current.

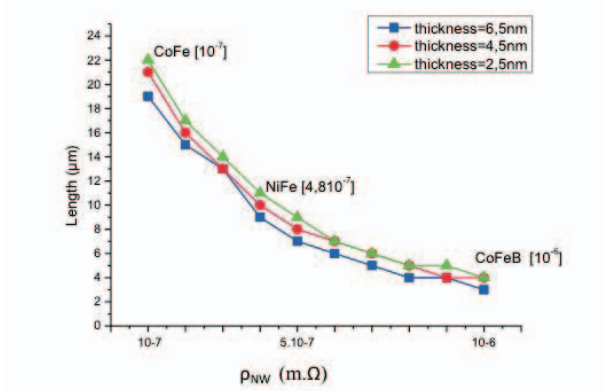

Fig. 2. Maximum nanowire length for racetrack memories with different material resistivity [14].

However, as the critical current density is still relatively high (order of $1012A/m^2$), the material resistivity of magnetic nanowire plays a crucial role in the capacity feature. In Fig.2, we studied the influence of resistivity on magnetic nanowire length for three types of magnetic alloys [15]. We can find that the maximum length for CoFeB with PMA can only be achieved to 4 um. The case for CoFe can reach up to more than 20 um, however its tunnel magnetorensistance (TMR) ratio cannot rival that of CoFeB. As a consequence, the racetrack memory with short magnetic nanowire is more suitable for CoFeB/MgO structure.

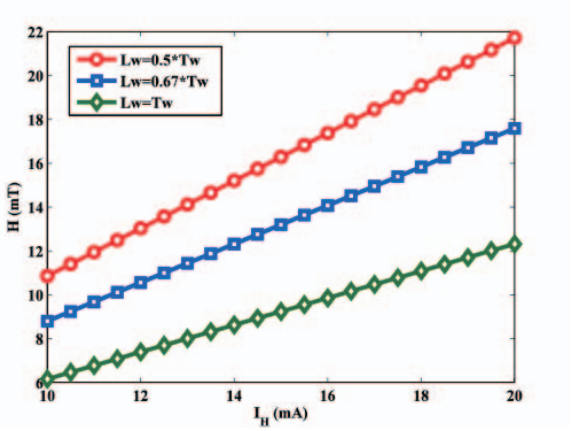

Fig. 3. Generation of magnetic field by current for different distance between metal line and racetrack memory.

On the other hand, a lot of alternative solutions for lowering the required current density have been proposed [16], [17]. For instance, an external magnetic field can be integrated to assist DW motions. The first prototype also benefits from this scheme. From the viewpoint of implementation, the magnetic field is often generated by the current flowing in high-level metal lines. According to the Biot-Savart-Laplace law, as shown in Fig. 3, 10-20 mT magnetic field requires 10-20 mA current. Tw and Lw represent the thickness of metal line and the distance between magnetic nanowire and metal line. We can also find that the distance has a great impact to the generation of magnetic field. For example, 0.5Tw shows a favorable performance compared with Tw and 0.67Tw. Meanwhile, the generation of magnetic field will cause additional energy consumption. This offers a tradeoff relation between the capacity and energy consumption, i.e. capacity improvement requires more energy.

Recently, an emerging current-induced phenomenon, called chiral DW motions, attracts more and more attentions. This phenomenon arises from the spin orbit torque (SOT), which combines the interactions of spin Hall effect (SHE) and Dzyaloshinskii-Moriya interaction (DMI) [18], [19]. A charge current flowing through heavy metals instead of ferromagnetic materials can induce the DW motions. This could solve the problem of high resistivity for conventional STT based racetrack memory. Furthermore, its high switching efficiency allows an enhanced energy economization. Nevertheless, the chiral DW motions in PMA materials still need the assistance of magnetic field. This DW dynamic behavior can be elucidated by one-dimension (1D) model considering the factor of a longitudinal magnetic field Hx [20].

$$\alpha \dot{X} + \triangle \dot{\varphi} = -\beta u - \frac{\pi}{2} \gamma \triangle H_{SHE} \sin(\varphi) \tag{3}$$

$$\dot{X} + \alpha \triangle \dot{\varphi} = \frac{\gamma \triangle H_K}{2} \sin(2\varphi) - u + \frac{\pi}{2} \gamma \triangle (H_X + H_{DMI}) \sin(\varphi), \tag{4}$$

where X is the position of a DW and φ is the angle that the DW magnetization forms with the easy plane, $\triangle$ is the DW width,

γ is the geromagnetic ratio. $H_{SHE} = u\frac{\theta_{SHE}}{\gamma Pt}$ is the effective field describing SHE, where t is the thickness of ferromagnetic layer, θ_{SHE} is the spin hall angle. H_K is the anisotropy field. $H_{DMI} = \frac{D}{\mu_0 M_S \triangle}$ is the effective field describing DMI, where μ_0 is the permeability in free space. D is the DMI parameter. It is noteworthy that the most recent observation shows that the structural asymmetry can allow eliminating the assistance of magnetic field, which greatly enhances the feasibility of this promising technology [21].

TABLE I
COMPARISON OF DIFFERENT TECHNOLOGIES BASED RACETRACK MEMORIES.

Type	STT based RT	Field assisted RT	Chiral DW RT
Capacity	Low	High	High
Energy	Low	High	Low

We summarized the performance of the aforementioned three racetrack memory structures. It is found that the applications requiring low energy consumption but low capacity is the most feasible path for the conventional STT based PMA racetrack memory. Aiming to achieve both low power and high capacity, alternative technologies, such as chiral DW motions, should be under further investigation.

III. Circuit Level Exploration

A. Overview

The cells of racetrack memory can be arranged as conventional memories [6], [22]. However, it may be more efficient to overlap the racetrack memory cells [8]. We explore those two layouts in circuit level, and demonstrate the circuit level design tradeoff on latency, energy and area.

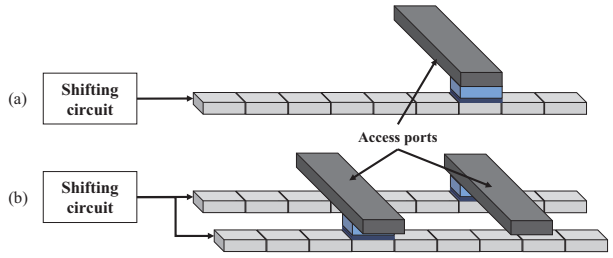

Fig. 4. The floor plan of racetrack memory cells. (a) conventional (b) 2-cell overlapped.

Both layouts for conventional and overlapped methods are briefed in Fig. 4(a) and (b), respectively. The access ports are used to perform data reading and programming. Shift current is supplied by shift circuit. For conventional method, the floor-plan area for one cell is exclusively occupied by the cell. For overlapped method, the floor-plan area for one cell is shared by multiple cells. Since a racetrack memory stripe is narrower and longer compared to a transistor, multiple stripes can be arranged atop of several transistors to fully exploit the floor-plan area. Thus overlap layout is proposed to increase the storage density.

B. Circuit Design Exploration

Number of Ports: The access port is used to read and program the bits aligned at the port. If more ports are formed on a stripe, the average shift distance can be reduced. Shorter shift distance means fewer overhead bits, and more domains can be used to save data bits. Thus, increasing the number of ports may reduce the average access latency. But it's not area efficient to connect each domain in racetrack memory. Too many ports will also degrade the storage density. Thus, the number of ports performs a tradeoff between area and access latency.

Cell Overlapping: In order to utilize the layout efficiently, multiple racetrack memory cells can be overlapped with each other in as in Fig. 4 (b). All transistors in access ports of those cells can be aligned to each other. Obviously, we can increase the number of racetrack memory cells overlapped to further improve area efficiency. However, with a fixed RM stripe length, overlapping more cells reduces the space for each cell to allocate their access ports. Thus, without expanding the overlapped cell area, fewer ports per racetrack stripe leads to longer shift distance.

Array Partitioning and Peripheral Circuitry: The latency, energy, and area of the array are also affected by the design optimisation targets [23]. The variation of targets leads to different number and size of transistors, buffers and wires in periphery circuitry. In addition, partitioning the RM array into sub-arrays also has a significant impact on the target. Thus, these issues also need to be explored.

C. Quantitative Analysis

Cell Overlapping: Fig. 5 shows the difference on area, latency and energy of different layouts. We take the conventional layout as a relative baseline. With the increased number of overlapped cells, the efficiency of the layout improves. The 4-cell overlapped layout can achieve 66% read latency reduction, and occupies only 29% area of the conventional one. The energy consumed in the 4-cell layout for read and shift are both shrunk to 59% and 48% of those for the conventional layout. Thus, we use the 4-cell overlapped layout to perform further analysis in the rest of this paper.

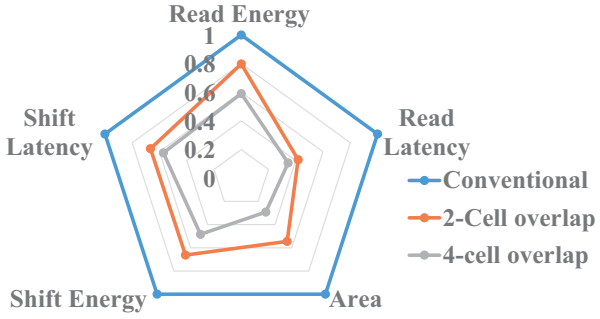

Fig. 5. Relative performance of overlapped layout.

Number of Ports: Fig. 6 shows the effect of increasing number of ports to the overlapped 4-cell, which builds up a 32MB racetrack memory cache. The shift distance decreases with the increase of port number. However, both latency for read and shift increase, due to the enlarged area. And for area, increasing the number of ports leads to the increase of area. When the number of ports is small (< 32), the increase of the area comes from word line routing; and it is dominated by the transistor size, when it is large (> 32).

Array Partitioning and Peripheral Circuitry: We analyze the circuit level performance based on different optimization targets. We assume the 32MB racetrack memory cache use 4-cell overlapped cells. Each cell has 64 data bits and 8

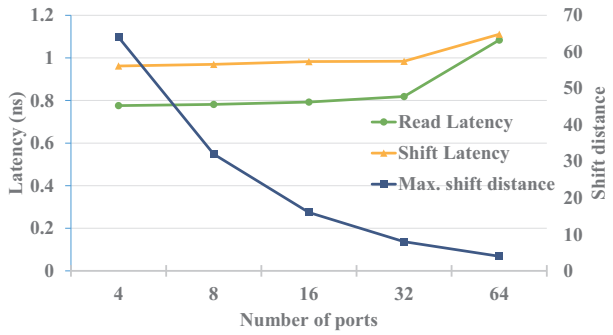

Fig. 6. Effect on the latency and area induced by number of ports .

access ports. The area changes from $2.66mm^2$ to $5.03mm^2$, by changing the optimization target from area to shift latency. The read latency and energy can be as low as $0.72ns$ and $0.19nJ$. The shift latency and energy can be as low as $0.93ns$ and $0.15nJ$. The detailed comparison is shown in Table II.

TABLE II
A 32MB RACETRACK MEMORY CACHE UNDER DIFFERENT OPTIMIZATION TARGETS.

Target	Area	Read Latency	Read Energy	Shift Latency	Shift Energy
Area(mm^2)	**2.66**	4.34	2.70	5.03	2.87
Read Latency (ns)	6.30	**0.72**	1.16	0.74	8.37
Read Energy (nJ)	0.27	0.53	**0.19**	0.58	0.24
Shift Latency (ns)	5.87	0.96	6.11	**0.93**	5.87
Shift Energy (nJ)	0.45	0.51	0.40	0.56	**0.15**

IV. Architecture Level Evaluation

In this section, we demonstrate that circuit level design optimization has impact on architectural level characters. We simulate PARSEC benchmarks [24] on gem5 simulation platform [25]. The system has 4 Alpha cores, 32KB split I/D cache for each core, a 2MB shared l2 cache, and a 32MB racetrack memory l3 cache. The main memory is DDR3-like simple DRAM with 100ns response time. The energy and latency numbers are based on the previous sections. The different optimization targets for area, latency and energy for read/shift are labeled as 'Area', 'Read latency', 'Read energy', 'Shift latency', and 'Shift energy'.

The overall system execution time is used to evaluate the system performance, as shown in Fig. 7. We take 'Shift energy' as baseline. The 'Read latency', 'Read energy', and 'Shift latency' reduce the overall time compared to 'Area' and 'Shift energy', due to their smaller read and shift latency. The performance difference among these targets can be up to 55.1%, and it is 17.5% on average.

The overall memory subsystem energy is used to evaluate the system energy consumption, as shown in Fig. 8. It includes the dynamic read and shift energy and the static leakage power from L1 to main memory. We take 'Shift latency' as baseline. Because the solutions optimized for latency use larger transistors and stronger drivers, the energy consumption both on dynamic operations and leakage power is higher. On average, area, read latency, read energy, and shift energy targets consume 33.5%, 70.1%, 28.6%, 42.8% of that of 'Shift latency', respectively.

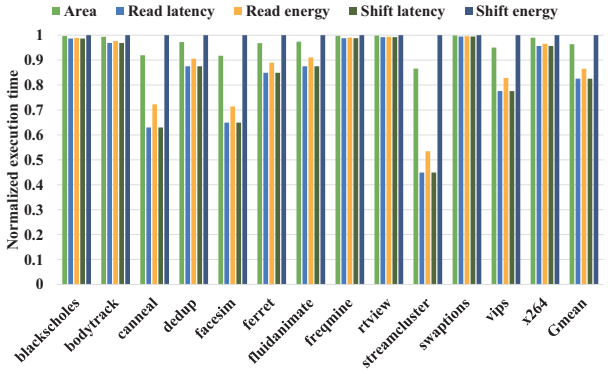

Fig. 7. The relative overall execution time.

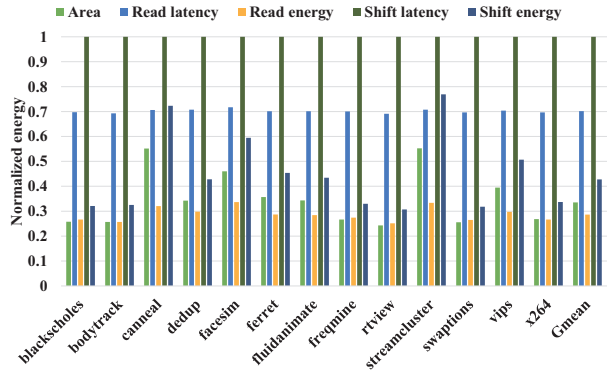

Fig. 8. The relative overall memory subsystem energy consumption.

V. RM based nonvolatile sensor nodes

In this section, we present a case study of using RM in non-volatile sensor nodes. We first introduce the overall architecture of non-volatile sensor nodes, followed by the structure of our proposed 11T1R RM based nvSRAM cell structure. Then, the performance and power evaluation results are provided and compared with other memory technologies.

A. Nonvolatile sensor node architecture

Energy harvesting sensor nodes have been widely used in habitat monitoring, volcano monitoring and structural monitoring because of its ultra-long operating time without maintenance. In order to ensure full system states retention in intermittent power nodes, racetrack memory is employed to replace both on-chip and off-chip volatile memory. The ultra-high storage density of racetrack memory can help reduce the fingerprint of a non-volatile processor to satisfy emerging demand of modern portable sensors. For example, it is feasible to be applied in some specific scenarios, such as smart patches.

A typical racetrack memory based non-volatile sensor node architecture is depicted in Fig. 9. It consists of an energy harvesting module, a non-volatile processor (NVP) [26], peripheral sensors, a backup capacitor, multi-bit RM based data storage and wireless transceivers. In order to avoid shift overhead of RM when used in high level memory hierarchy, we use RM in the form of nvSRAM. NvSRAM integrates a SRAM cell and a one-bit RM element in cell levels, forming a direct bit-to-bit connection [27]. In addition, to further improve efficiency of updating data in RM, a shift based write is employed. Therefore, it provides comparable power and

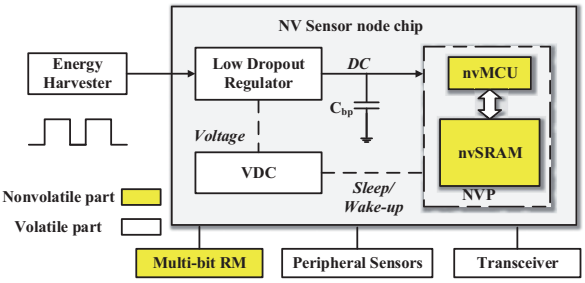

Fig. 9. RM based energy harvesting nonvolatile sensor node architecture

performance metrics as SRAM, while keeping the non-volatile capability when power failures happen. Supposing a square waveform is generated by a vibration based energy harvester, the voltage detection circuit (VDC) detects a power drop when a power failure arrives. The VDC generates a backup signal to the NVP, which starts the backup operation in nvSRAM and NVP. The off-chip storage is fabricated by multi-bit racetrack memory to make use of its characteristic of high density.

B. RM based nonvolatile SRAM structure

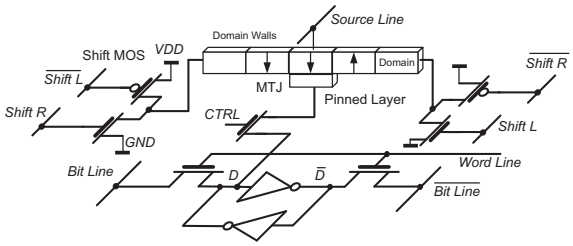

Fig. 10. Cell structure of proposed 11T1R RM based nvSRAM

We refer to the 7T1R nvSRAM in [28] and designs the 11T1R RM based nvSRAM cell structure. The circuit scheme of the proposed 11T1R RM based nvSRAM is shown in Fig. 10. It comprises a standard 6T SRAM cell and a multi-bit RM cell is connected by a CTRL MOS. Fig. 11 shows the backup and restore operations of the RM based 11T1R cell. At T1, the word line is applied to write the SRAM cell to '1' state. At T2, the CTRL signal is applied to enable the current from bit line to source line via the MTJ to write '1' (low resistance state) to the RM. During T3 to T4, the power is cut off. At T5 and T6, the CTRL signal and source line is set to high to drive D to '1' through the low resistance MTJ. After that, the SRAM cell is restored to '1'.

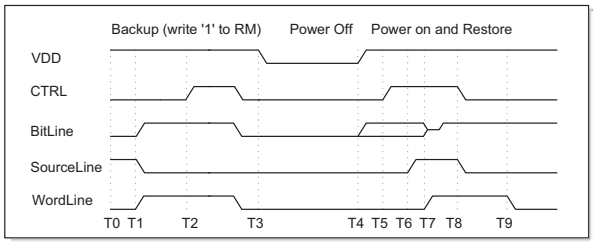

Fig. 11. The waveform of store and restore operations

C. Performance and energy evaluation

We simulate the sensor node model in gem5 [29] with the configurations listed in table III. We first compare the

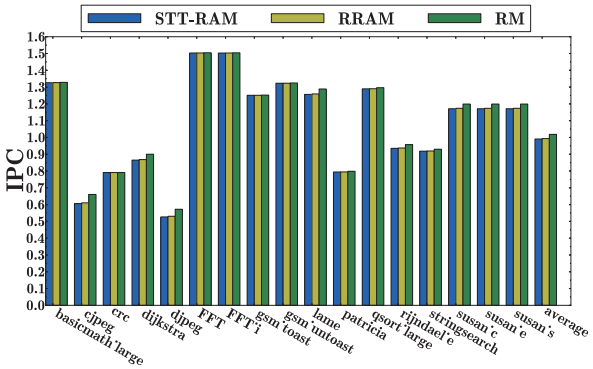

Fig. 12. Comparison of instructions per cycle (IPC)

energy and performance of non-volatile sensor nodes with various types of non-volatile off-chip memories. All important parameters are listed in table IV. Fig. 12 compares the IPC of benchmarks from Mibench [30]. We execute 30M instructions of a single benchmark. It can be inferred from Fig. 12 that using RM based storage achieves 2.8% IPC improvement over STT-RAM and 2.5% IPC improvement over RRAM because RM has shorter access time. Therefore, the overall cache miss penalty is reduced.

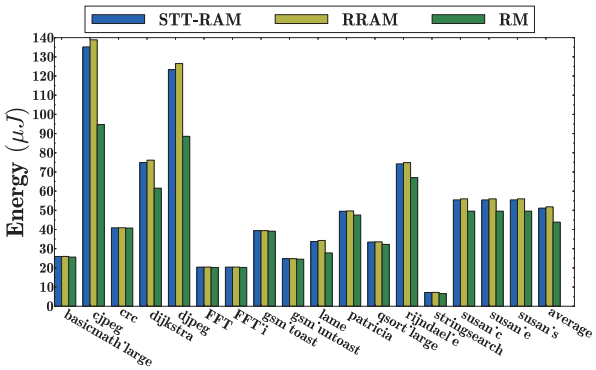

Fig. 13. Comparison of energy consumption

Fig. 13 shows the energy consumption comparison of various non-volatile off-chip memories. RM reduces the average energy consumption by 14.3% compared with STT-RAM based off-chip memory and 15.4% compared with RRAM based off-chip memory. The reason is in two-folds. First, the read energy of RM is lower than STTRAM and RRAM because of its smaller fingerprint. Second, shift based write operation of RM is also more efficient than STTRAM and RRAM.

TABLE III
SIMULATION SETUP

Component	Configuration
Processor (nvMCU)	Single core, 1GHz with 2-width issue
On-chip cache (nvSRAM)	I&D size: 8kB&8kB (4-way association) Block size: 64B
Off-chip memory	I&D size: 1MB&1MB

Table V compares the energy and timing parameters of RM based nvSRAM with the RRAM based nvSRAM [27]. The RM based nvSRAM can reduce store/restore energy significantly compared with RRAM based nvSRAM. In addition, the store/restore performance is also improved. It eases the backup capacitor volume requirements with/without store eliminations. Therefore, the area and in-rush current overhead

TABLE V
COMPARISON OF RRAM BASED NVSRAM AND RM BASED NVSRAM FOR A 8KB L1 CACHE

Ram Type	Store Energy	Restore Energy	Store time	Restore time	Backup capacitor	Backup capacitor with store elimination [31]
RRAM based nvSRAM [27]	1.714nJ/2kb	1.06nJ/2kb	10ns	10ns	34nF	23nF
RM based nvSRAM	1.07nJ/2kb	0.22nJ/2kb	6.95ns	2.47ns	25nF	18nF
Reduction rate	38%	79%	31%	75%	26%	22%

TABLE IV
PARAMETERS FOR NVSRAM AND VARIOUS NVM BASED OFF-CHIP MEMORY

Memory type		Read Latency	Write Latency	Read Energy	Write Energy
nvSRAM based L1 Cache		1ns	1ns	4.64pJ	2.173pJ
Off-chip memory	RM	21ns	22ns	21pJ	105pJ
	STT-RAM	22ns	31ns	34pJ	188pJ
	RRAM	21ns	32ns	37pJ	192pJ

is also reduced. The store and restore time penalty are also reduced. It can help to achieve fast checkpointing and fine grained power management.

VI. CONCLUSION

From the device layer to the architecture layer, a lot of design factors and issues should be considered in racetrack memory design. These design factors interact with each other and have impact on performance, energy, and area of RM design. For different design goals, the design space of RM should be well explored to find proper configurations. A system level case study of applying RM in sensor node demonstrates its advantages over SRAM and STTRAM.

VII. ACKNOWLEDGMENTS

The authors wish to acknowledge financial support from the European FP7 program MAGWIRE (257707) and the French national projects ANR-MARS, ANR-DIPMEM. This work is also supported in part by NSF CNS-1116171, Huawei shannon Lab and High-Tech Research and Development (863) Program under contract 2013AA01320 and the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions under contract YETP0102.

REFERENCES

[1] A. Mishra, X. Dong, G. Sun, Y. Xie, N. Vijaykrishnan, and C. Das, "Architecting on-chip interconnects for stacked 3d stt-ram caches in cmps," in *Computer Architecture (ISCA), 2011 38th Annual International Symposium on*, June 2011, pp. 69–80.

[2] S.-S. Sheu, M.-F. Chang, and K.-F. L. et al., "A 4mb embedded slc resistive-ram macro with 7.2ns read-write random-access time and 160ns mlc-access capability," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, Feb 2011, pp. 200–202.

[3] C. Gopalan, Y. Ma, T. Gallo, J. Wang, E. Runnion, J. Saenz, F. Koushan, P. Blanchard, and S. Hollmer, "Demonstration of conductive bridging random access memory (cbram) in logic {CMOS} process," *Solid-State Electronics*, vol. 58, no. 1, pp. 54 – 61, 2011.

[4] C. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, and M. Stan, "Relaxing non-volatility for fast and energy-efficient stt-ram caches," in *High Performance Computer Architecture (HPCA), 2011 IEEE 17th International Symposium on*, Feb 2011, pp. 50–61.

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

[6] R. Venkatesan, V. Kozhikkottu, C. Augustine, A. Raychowdhury, K. Roy, and A. Raghunathan, "Tapecache: a high density, energy efficient cache based on domain wall memory," in *Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. 2333707: ACM, pp. 185–190.

[7] R. Venkatesan, M. Sharad, K. Roy, and A. Raghunathan, "Dwm-tapestri - an energy efficient all-spin cache using domain wall shift based writes," in *Proceedings of the Conference on Design, Automation and Test in Europe*. 2485718: EDA Consortium, pp. 1825–1830.

[8] Z. Sun, W. Wu, and H. Li, "Cross-layer racetrack memory design for ultra high density and low power consumption," in *Proceedings of the 50th Annual Design Automation Conference*. 2488799: ACM, pp. 1–6.

[9] 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, 2008. [Online]. Available: http://www.sciencemag.org/content/320/5873/209.abstract

[10] S. S. P. Parkin, M. Hayashi, and L. Thomas, "Magnetic domain-wall racetrack memory," *Science*, vol. 320, no. 5873, pp. 190–194, 2008. [Online]. Available: http://www.sciencemag.org/content/320/5873/190.abstract

[11] Y. Zhang, W. Zhao, Y. Lakys, J.-O. Klein, J.-V. Kim, D. Ravelosona, and C. Chappert, "Compact modeling of perpendicular-anisotropy cofeb/mgo magnetic tunnel junctions," *Electron Devices, IEEE Transactions on*, vol. 59, no. 3, pp. 819–826, March 2012.

[12] C. Chappert, "The emergence of spin electronics in data storage," *Nature Publishing Group*, vol. 6, no. 11, pp. 513–823, Nov 2007.

[13] A. Annunziata, M. Gaidis, L. Thomas, C. Chien, C.-C. Hung, P. Chevalier, E. O'Sullivan, J. Hummel, E. Joseph, Y. Zhu, T. Topuria, E. Delenia, P. Rice, S. Parkin, and W. Gallagher, "Racetrack memory cell array with integrated magnetic tunnel junction readout," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec 2011, pp. 24.3.1–24.3.4.

[14] Y. Zhang, W. Zhao, D. Ravelosona, J.-O. Klein, J. Kim, and C. Chappert, "Perpendicular-magnetic-anisotropy cofeb racetrack memory," *Journal of Applied Physics*, vol. 111, no. 9, pp. 093 925–093 925–5, May 2012.

[15] N. Ben-Romdhane, W. Zhao, Y. Zhang, J.-O. Klein, Z. Wang, and D. Ravelosona, "Design and analysis of racetrack memory based on magnetic domain wall motion in nanowires," in *Nanoscale Architectures (NANOARCH), 2014 IEEE/ACM International Symposium on*, July 2014, pp. 71–76.

[16] Y. Zhang, W. Zhao, J.-O. Klein, C. Chappert, and D. Ravelosona, "Current induced perpendicular-magnetic-anisotropy racetrack memory with magnetic field assistance," *Applied Physics Letters*, vol. 104, no. 3, pp. 032 409–032 409–5, Jan 2014.

[17] Y. Z. et al., "Implementation of magnetic field assistance to current-induced perpendicular-magnetic-anisotropy racetrack memory," *Journal of Applied Physics*, vol. 115, no. 17, pp. 17D509–17D509–3, May 2014.

[18] S. Emori, U. Bauer, S.-M. Ahn, E. Martinez, and G. S. D. Beach, "Current-driven dynamics of chiral ferromagnetic domain walls," *Nat Mater*, vol. 12, no. 7, pp. 611–616, 2013.

[19] P. P. J. Haazen, E. Mur, J. H. Franken, R. Lavrijsen, H. J. M. Swagten, and B. Koopmans, "Domain wall depinning governed by the spin hall effect," *Nat Mater*, vol. 12, no. 4, pp. 299–303, 2013, 10.1038/nmat3553.

[20] Y. Z. et al., "Peristaltic perpendicualr-magnetic-anisotropy racetrack memory based on chiral domain wall motions," *submitted to J. of Phys D: Appl. Phys*.

[21] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L.-T. Chang, Y. Jiang, M. Lang, J. Tang, Y. Wang, Y. Tserkovnyak, P. K. Amiri, and K. L. Wang, "Switching of perpendicular magnetization by spin-orbit torques in the absence of external magnetic fields," *Nat Nano*, vol. 9, no. 7, pp. 548–554, 2014.

[22] R. Venkatesan, S. Ramasubramanium, S. Venkataramani, K. Roy, and A. Raghunathan, "Stag: Spintronic-tape architecture for gpgpu cache hierarchies," in *International Symposium on Computer Architecture (ISCA)*, June 2014.

[23] X. Dong, C. Xu, Y. Xie, and N. Jouppi, "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 31, no. 7, pp. 994–1007, July 2012.

[24] C. Bienia, S. Kumar, J. P. Singh, and K. Li, "The parsec benchmark suite: Characterization and architectural implications," Princeton University, Tech. Rep. TR-811-08, January 2008.

[25] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, and D. A. Wood, "The gem5 simulator," *SIGARCH Comput. Archit. News*, vol. 39, no. 2, pp. 1–7, Aug. 2011.

[26] Y. Wang and et al., "A 3us wake-up time nonvolatile processor based on ferroelectric flip-flops," in *ESSCIRC (ESSCIRC), 2012 Proceedings of the*. IEEE, 2012, pp. 149–152.

[27] P.-F. Chiu, M.-F. Chang, S.-S. Sheu, K.-F. Lin, P.-C. Chiang, C.-W. Wu, W.-P. Lin, C.-H. Lin, C.-C. Hsu, F. T. Chen *et al.*, "A low store energy, low vddmin, nonvolatile 8t2r sram with 3d stacked rram devices for low power mobile applications," in *VLSI Circuits (VLSIC), 2010 IEEE Symposium on*. IEEE, 2010, pp. 229–230.

[28] K. Chen, J. Han, and F. Lombardi, "On the non-volatile performance of flip-flop/sram cells with a single mtj."

[29] Binkert and et al., "The gem5 simulator," *ACM SIGARCH Computer Architecture News*, vol. 39, no. 2, pp. 1–7, 2011.

[30] Guthaus and et al., "Mibench: A free, commercially representative embedded benchmark suite," in *Workload Characterization, 2001. WWC-4. 2001 IEEE International Workshop on*, Dec 2001, pp. 3–14.

[31] H. Li, Y. Liu and et al., "An energy efficient backup scheme with low inrush current for nonvolatile sram in energy harvesting sensor nodes," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*. IEEE, 2015.