

# Voltage-Controlled MRAM for Working Memory: Perspectives and Challenges

Wang Kang, Liang Chang, Youguang Zhang, and Weisheng Zhao

Fert Beijing Research Institute, BDBC, School of Electronic and Information Engineering  
Beihang University  
Beijing, 100191, China  
{wang.kang, liang.chang, zyg, weisheng.zhao}@buaa.edu.cn

**Abstract**—Magnetic random access memory (MRAM) has been widely studied for future nonvolatile working memory candidate. However, the mainstream current (spin transfer torque, STT or spin Hall effect, SHE) driven MRAMs (STT-MRAM or SHE-MRAM) face intrinsic problems in terms of high write power and long latency, significantly limiting the applications for low-power and high-speed working memories. The recently-developed new-generation MRAM, named VCMA-MRAM, which exploits the voltage-controlled magnetic anisotropy (VCMA) effect to write (or assist to write) data information into magnetic tunnel junctions (MTJs), holds the promise to efficiently overcome these problems. Despite the impressive possibility of improving write power and speed, this technology, however, is currently under intensive research and development (R&D), and some challenges still await answers. In this paper, we investigate the perspectives and challenges of VCMA-MRAM for working memories from a cross-layer (device/circuit/architecture) design point of view. We demonstrate that VCMA-MRAM outperforms STT-MRAM and SHE-MRAM in terms of area, speed, energy consumption and instruction-per-cycle (IPC) performance, benefiting from the low-power and high-speed VCMA-driven data writing mechanism. On the other hand, challenges in terms of device fabrication and circuit design should be efficiently addressed before practical applications.

## I. INTRODUCTION

Nonvolatile memory (NVM) has been widely considered as a promising technology to help tackle the “leakage currents” or “power wall” problem of conventional complementary-metal-oxide-semiconductor (CMOS) technology. Among all the NVM technologies, magnetic random access memory (MRAM), which aims eventually to replace SRAM and/or DRAM as a working memory candidate (e.g., cache or embedded memory) in future computer architectures, has attracted considerable attention in both academia and industry [1-3]. Until now, the research and development (R&D) of MRAM has gone through three primary generations depending on the anisotropy property of magnetic tunnel junction (MTJ, one of the core devices of MRAM) and the data writing mechanism. Current mainstream MRAM relies on the perpendicular MTJ (p-MTJ) device and spin transfer torque (STT) effect, through which only a bi-directional charge current flowing through the p-MTJ can write the desired data into the memory cell [4, 5]. Recently, Everspin has announced the first 256Mb p-MTJ based STT-MRAM chip. Another current driven MRAM generation under intensive R&D is the spin Hall effect (SHE) driven MRAM, in which data is written into the MTJ device by a charge current flowing through a heavy metal strip under the MTJ [6].

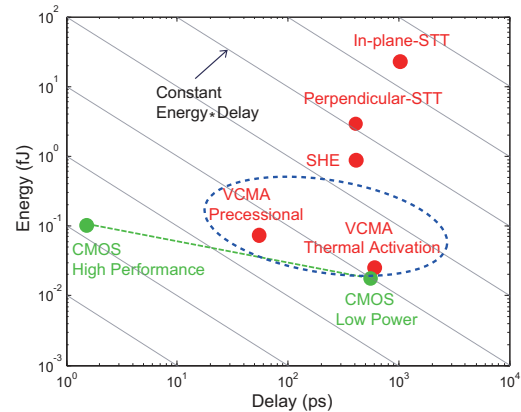


Fig. 1. Performance comparison in terms of energy and delay, among various MRAM generations and conventional CMOS technology [9].

Both STT-MRAM and SHE-MRAM have also been widely studied by computer architects as one potential NVM candidates in working memory applications [7, 8]. Unfortunately, based on our investigations, current STT-MRAM and SHE-MRAM still have a performance gap (at least 1-2 orders of magnitude) in terms of dynamic write energy and access delay, compared with conventional CMOS-based SRAM, as shown in Fig. 1 [9].

Recently, the discovery of the voltage-controlled magnetic anisotropy (VCMA) effect enables us a new possibility to utilize voltage (or electric field) for switching an MTJ device [10-12]. Based on this effect, a new-generation of MRAM is developed, named VCMA-MRAM, which utilizes the VCMA effect to write (or assist STT to write) data information into an MTJ and enables to narrow the performance gap between MRAM and SRAM (see also Fig. 1). The utilization of a voltage instead of a charge current for writing data into an MTJ allows for much lower energy dissipation (down to fJ/bit), as Ohmic loss or Joule heating can be greatly reduced. Furthermore, the VCMA effect enables rather fast precessional switching of an MTJ (down to hundreds of ps) by lowering the energy barrier between the two magnetization states of the MTJ. In addition, the voltage-driven mechanism brings other advantages, e.g., high density, since the size of the access transistors can be reduced, thanks to the decrease of the required driving current density for data writing operations. By merging the merits in terms of low-power, high-speed and high-density etc., VCMA-MRAM is expected to be a promising NVM candidate for working memory applications.

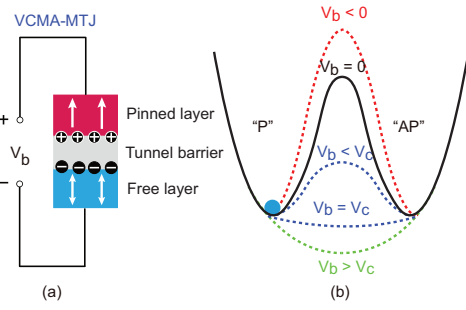


Fig. 2. (a) Schematic of a VCMA-MTJ device; (b) Illustration of the impacts of various bias voltages ( $V_b$ ) on the energy barrier ( $E_b$ ) of a MTJ device.

In this paper, we investigate the perspectives and challenges of VCMA-MRAM for low-power and high-speed working memories with a cross-layer (device/circuit/architecture) design point of view. In device-level investigations, the magnetization dynamics of an MTJ under the VCMA effect was described by solving a modified Landau-Lifshitz-Gilbert (LLG) equation. Then a physics-based VCMA-MTJ compact model, in which both precessional VCMA and STT-assisted VCMA effects were included, was then developed. In circuit-level investigations, the bit-cell and array of VCMA-MRAM were designed with the developed VCMA-MTJ model and a CMOS design-kit. Two data writing strategies, including precessional VCMA and STT-assisted VCMA, were studied respectively. Afterwards, system-level investigations were carried out to evaluate the potential of VCMA-MRAMs in cache memories. Finally, design challenges are discussed before employing VCMA-MRAM as a complete industrial NVM technology.

The rest of this paper is organized as follows. In Section II, we briefly introduce the fundamentals, device modeling and circuit design of VCMA-MRAM. Then Section III presents the perspectives of VCMA-MRAM in working memory application. Design challenges are discussed in Section IV. Finally, Section V concludes the paper.

## II. VCMA-MRAM TECHNOLOGY

### A. Fundamentals of VCMA Effect

The physics behind VCMA effect can be explained by the “orbital” theory, i.e., the voltage (electric field)-induced change of occupancy of the atomic orbitals at the ferromagnet|oxide interface, which in conjunction with the spin-orbit interaction, results in a change of perpendicular magnetic anisotropy (PMA) at the interface [12, 13]. By applying a voltage on an MTJ (e.g., CoFeB|MgO) device, an electric field is formed through the MgO tunnel barrier. The accumulation of electron charges at the CoFeB|MgO interface will induce a change in the occupation of atomic orbitals that promotes or represses the interfacial PMA depending on the polarity of the applied voltage [11, 12]. In specific, a positive (negative) voltage reduces (increases) the interfacial PMA for the CoFeB|MgO-based MTJ devices. For magnetization switching of an MTJ in the presence of the VCMA effect, it is equivalent to lower (or heighten) the energy barrier ( $E_b$ ) between the two stable magnetization states (“P” and “AP”) under a positive (or negative) voltage, as shown in Fig. 2. Depending on the amplitude of the applied positive

voltage on an MTJ, the MTJ switching behavior can be divided into two regimes: (a) if the positive voltage is insufficient to fully eliminate the energy barrier, then thermal activation or STT is required to switch the magnetization of the free layer of the MTJ, named thermal activation regime; (b) otherwise, if the positive voltage is sufficiently large to fully eliminate the energy barrier, then the magnetization of the free layer of the MTJ will become precessionally unstable and will walk back and forth between the “P” and “AP” states, named precessional regime. Herein, the amplitude of the positive voltage required to fully eliminate the energy barrier of an MTJ is defined as the critical voltage ( $V_c$ ) of the VCMA effect. For simplicity, the VCMA effect can also be quantified by a macrospin description from a technological perspective. As discussed above, the interfacial PMA of an MTJ can be modulated through an electric field, thereby the effective PMA ( $K_{eff}$ ) of an MTJ can be mathematically modeled as a function of the applied voltage ( $V_b$ ) on the MTJ in the presence of the VCMA effect [14], i.e.,

$$K_{eff}(V_b) = \frac{M_s H_{eff}(V_b)}{2} = \frac{K_i(0) - K_i(V_b)}{t_f} - 2\pi M_s^2, \quad (1)$$

where  $M_s$  is the saturation magnetization of the MTJ,  $H_{eff}(V_b)$  is the effective magnetic field under a voltage of  $V_b$ ,  $t_f$  is the thickness of the free layer of the MTJ,  $K_i(0)$  and  $K_i(V_b)$  are the interfacial PMA energies under applied voltages of zero and  $V_b$ , respectively. Furthermore, as most of the experimental results show a linear dependence between the interfacial PMA energy and the electric field,  $K_i(V_b)$  can then be simplified as [10, 15],

$$K_i(V_b) = \xi \frac{V_b}{d_{ox}}, \quad (2)$$

where  $\xi$  is a linear VCMA coefficient and  $d_{ox}$  is the thickness of the tunnel barrier of the MTJ.

### B. Device Modeling of VCMA-MRAM

The magnetization dynamics of the free layer of an MTJ in the presence of the VCMA effect can be described by a modified Landau-Lifshitz-Gilbert (LLG) equation [14, 16],

$$\frac{d\vec{m}}{dt} = -\gamma \vec{m} \times \vec{H}_{eff}(V_b) + \alpha \vec{m} \times \frac{d\vec{m}}{dt} - \rho_{stt} \vec{m} \times (\vec{m} \times \vec{m}_r), \quad (3)$$

where  $\vec{m} = \{m_x, m_y, m_z\}$  is the magnetization vector of the free layer,  $\vec{m}_r$  is a polarization vector,  $\gamma$  is the gyromagnetic ratio,  $\alpha$  is the Gilbert damping factor,  $\vec{H}_{eff}(V_b)$  is the effective magnetic field under a voltage  $V_b$ ,  $\rho_{stt} = \gamma \hbar P J_{stt} / 2e\mu_0 t_f M_s$  is the STT factor, in which  $\hbar$  is the reduced Planck constant,  $P$  is the STT polarization factor,  $J_{stt}$  is the driving current density inducing STT,  $e$  is the elementary charge, and  $\mu_0$  is the vacuum permeability. The first, second, and third terms on the right side of Eq. (3) are the precessional torque, damping torque, and STT, respectively. The effective magnetic field  $\vec{H}_{eff}(V_b)$  includes the external field ( $\vec{H}_{ext}$ ), the demagnetization field ( $\vec{H}_{dem}$ ), the thermal field ( $\vec{H}_{th}$ ), and the voltage-dependent anisotropy field ( $\vec{H}_{ani}(V_b)$ ). For solving the above Eq. (3), we define  $\vec{e}_x$ ,  $\vec{e}_y$ , and  $\vec{e}_z$  as the unit vectors along the  $x$ -,  $y$ - and  $z$ -axes of the Cartesian coordinate system. We further define  $\theta$  and  $\varphi$  as the polar and azimuthal angles of  $\vec{m}$ , respectively. We can then derive that  $\vec{m} = \sin\theta \cos\varphi \vec{e}_x + \sin\theta \sin\varphi \vec{e}_y + \cos\theta \vec{e}_z$  and  $\vec{m}_r = \vec{e}_z$ . By substituting the above parameters into Eq. (3), then

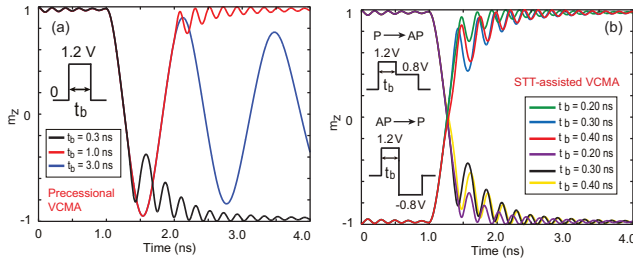


Fig. 3. Time-resolved evolutions of the magnetization of the free layer in the presence of the VCMA effect. The VCMA-MTJ switching operates in the (a) precessional regime; and (b) thermal-activation regime with STT assistance.

Table I Parameters of the VCMA-MTJ model

Parameter	Description	Default Value
$K_i$	Interfacial PMA	0.32 mJ/m <sup>2</sup>
$M_s$	Saturation magnetization	$0.625 \times 10^6$ A/m
$\xi$	VCMA coefficient	60 fJ/V · m
$\alpha$	Gilbert damping factor	0.05
$d_{ox}$	Oxide barrier thickness	1.5 nm or 1.2 nm
$H_x$	External magnetic field	$4.8 \times 10^4$ A/m

the differential equations on the time-dependent  $\theta$  and  $\varphi$  can be derived eventually. With the parameters listed in Table I, Fig. 3 shows the evolutions of  $\vec{m} = \{m_x, m_y, m_z\}$  of the free layer in the presence of the VCMA effect (here only  $m_z$  is plotted). If the applied voltage ( $V_b$ ) on the MTJ is larger than the critical voltage ( $V_c$ ) of the VCMA effect, the MTJ switching operates in the precessional regime, thus resulting in an oscillation of the magnetization of the free layer (see Fig. 3(a)). In this case, the final magnetization state of the MTJ depends mainly on the pulse duration of the voltage. In order to achieve deterministic MTJ switching, a precisely-controlled voltage pulse duration is required, which, however, is rather difficult in practice when taking into consideration the process, voltage, and temperature (PVT) variations. Alternatively, if  $V_b$  is smaller than  $V_c$ , then the MTJ switching operates in the thermal activation regime. In this case, either a current or a magnetic field is generally needed to switch the magnetization of the free layer.

In practice, to realize fast and deterministic MTJ switching operation, we can firstly apply a voltage pulse with amplitude higher than  $V_c$  to induce precessional magnetization oscillation (owing to the VCMA effect) of the free layer, then followed by a second voltage pulse with amplitude smaller than  $V_c$  to induce STT effect for determining the final magnetization orientation of the free layer, named STT-assisted VCMA (see Fig. 3(b)). The physics behind this phenomenon can be theoretically explained by the competition between the VCMA and STT effects, which has also been proved by experimental results [17]. Specifically, during applying the first voltage pulse, the VCMA effect is dominant and it lowers (or eliminates) the energy barrier of the MTJ. In this case, the magnetization of the free layer will precessionally oscillate around the effective field, as shown in Fig. 3(b). Afterwards, during applying the second voltage pulse, the STT effect will be dominant. Because the STT torque acts as a damping or anti-damping torque (depending on the polarity) on the present magnetization configuration, no oscillatory behavior is expected once after that the switching probabilities take a maximum close to a unity for the case of STT switching.

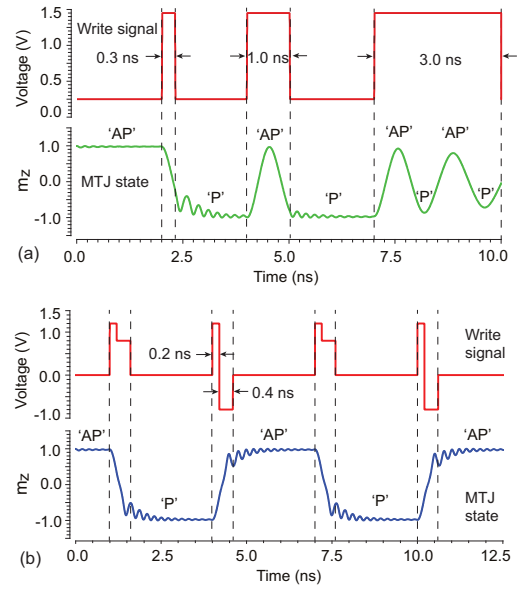


Fig. 4. Electrical simulation waveforms (write operations) of an MTJ under (a) the precessional VCMA strategy; and (b) the STT-assisted VCMA strategy.

In this strategy, we need no precise control of the duration of the first voltage pulse and the final magnetization orientation of the free layer depends on the polarity of the second voltage pulse.

After the derivations of the magnetization dynamics (i.e., time-dependent  $\theta$  and  $\varphi$ ) of the free layer of the MTJ, we further develop a VCMA-MTJ electrical model, in conjunction with the Slonczewski STT model [18], Julliere resistance model [19], and voltage-dependent TMR model [20]. The electrical model is written in Verilog-A language and is fully spice-compatible with the CMOS design-kit. The default parameters of the electrical model are listed in Table I. To verify the functionalities of the developed VCMA-MTJ electrical model, we performed hybrid MTJ/CMOS simulations at the 40 nm technology node. Fig. 4(a) and Fig. 4(b) show the simulation waveforms (only for write operations), when the MTJ is driven by the precessional VCMA and the STT-assisted VCMA strategies, respectively, which are consistent with the results of the magnetization dynamics of the free layer of the MTJ, as shown in Fig. 3(a) and Fig. 3(b).

### C. VCMA-MRAM Circuit Design

Similar to STT-MRAM, we utilize the 1T1MTJ (i.e., one MTJ connected in series with one CMOS access transistor) cell structure for VCMA-MRAM, as shown in Fig. 5. The main difference between STT-MRAM and VCMA-MRAM is the method for writing data information into MTJs. We investigate two types of VCMA-MRAM technologies based on the data writing strategies, including precessional VCMA-MRAM and STT-assisted VCMA-MRAM. Using the developed VCMA-MTJ compact model, we performed hybrid MTJ/CMOS simulations to compare the performance among STT-MRAM, SHE-MRAM and VCMA-MRAMs. Table II lists the results in terms of write power, write delay, read power, and read delay.

The write power and write delay may vary under different working conditions, including access transistor size, amplitude

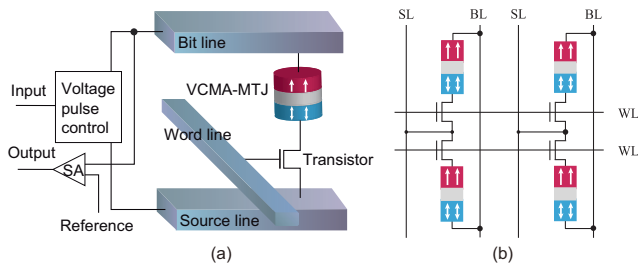


Fig. 5. Schematic of the (a) bit-cell and (b) array structures of VCMA-MRAM.

Table II Performance comparison

MRAM Technology	STT-MRAM	SHE-MRAM	Precessional VCMA	STT-assisted VCMA
Write power (fJ/bit)	~ 80-200	~ 40-120	~ 6.14-7.5	~ 6.68-10
Write delay (ns/bit)	~ 1-10	~ 1-5	~ 0.3-0.5	~ 0.6-1.0
Read power (fJ/bit)	~ 1.33	~ 1.35	~ 1.24	~ 1.25
Read delay (ps/bit)	~ 80	~ 85	~ 220	~ 130

and width of the driving voltage (or current) pulse, etc. Therefore, we give a reasonable distribution interval of these performance in the table. It is evident that both VCMA-MRAMs outperform STT-MRAM and SHE-MRAM in terms of write power and write delay. The read power of the four MRAM technologies are similar when using the same read methods. The read delay of the two VCMA-MRAMs increase, compared with STT-MRAM and SHE-MRAM, because the former have much larger MTJ resistances. The relatively high read delay per bit is a limitation of VCMA-MRAMs. However, the total read delay will be balanced when taking into account the signal transmission latency in a memory chip since VCMA-MRAMs have higher densities and the corresponding signal transmission latencies will be greatly reduced, compared with STT-MRAM and SHE-MRAM. In short, considering all the factors in terms of design complexity, density, write power, write delay, read power, and read delay, STT-assisted VCMA-MRAM may be the best candidate for working memories.

### III. PERSPECTIVES OF VCMA-MRAM IN CACHE MEMORY

In this section, we show the perspectives of VCMA-MRAM in cache memory from a system-level design perspective.

#### A. Simulation Setup

For the system-level simulations, we first employed a nonvolatile memory simulator, NVSim [21], to simulate various cache configurations for exploring the design space. Afterwards, a popular full-system simulator, i.e., Gem5 [22], was utilized to simulate a multiple-cache memory hierarchy in the processor. We performed an iso-capacity replacement for the L2 and/or L1 caches with STT-MRAM, SHE-MRAM, precessional VCMA-MRAM, and STT-assisted VCMA-MRAM, respectively. The processor configurations used in our simulations are provided in Table III. All benchmarks were from SPEC CPU 2006 and two billion instructions for each benchmark were performed.

Table III System-level simulation set-ups and configurations

Component	Configuration
CPU	Single core, 2 GHz, out-of-order
L1	Inst./Data: 32K Bytes/32K Bytes, Block: 64 Bytes, Line: 2-way, 1 Bank, Write-back SRAM: Lat.: 2 Cycle STT-MRAM: Lat.: R/W, 2/12 Cycle SHE-MRAM: Lat.: R/W, 2/4 Cycle Precessional VCMA-MRAM: Lat.: R/W, 2/2 Cycle STT-assisted VCMA-MRAM: Lat.: R/W, 2/3 Cycle
L2	1M Bytes, 64 Bytes, Line 8-way, 1 Bank, Write-back SRAM: Lat.: 10 Cycle STT-MRAM: Lat.: R/W, 10/20 Cycle SHE-MRAM: Lat.: R/W, 10/12 Cycle Precessional VCMA-MRAM: Lat.: R/W, 10/8 Cycle STT-assisted VCMA-MRAM: Lat.: R/W, 10/9 Cycle
Execution Unit	2x ALU, 2x CALU, 2x FPU
Main Memory	8GB, DDR3, 1600 MHz, 120 cycle, 12.8 GB/s.
Benchmarks	400.perlbenc, 401.bzip2, 429.mcf, 445.gobmk, 456.hammer, 458.sjeng, 462.libquantum, 433.milc, 435.gromacs, 437.Leslie3d, 444.namd, 450.soplex, 453.povray, 470.lbm

#### B. Cache Characteristics

Using NVSim, we investigated the area, write energy, write latency, read energy, read latency, and leakage energy for various cache capacities, as shown in Fig. 6, in which all the values were normalized to that of the base-line SRAM.

**Area:** SRAM-based cache has a minimum area when the cache capacity is below 256K bytes. However, as capacity increases, the MRAM-based caches significantly outperform SRAM in terms of area. In particular, VCMA-MRAMs have the smallest cache area when capacity is more than 256K bytes (Fig. 6(a)).

**Write Energy:** As shown in Fig. 6(b), the write energies of both VCMA-MRAMs are always lower than those of SRAM, STT-MRAM, and SHE-MRAM, owing to the voltage-induced MTJ switching mechanism.

**Write Latency:** As shown in Fig. 6(c), both STT-MRAM and SHE-MRAM have longer write latencies than SRAM, even if capacity is greater than 4M bytes. However, VCMA-MRAMs outperform SRAM in terms of write latency at a capacity of approximately 256K bytes.

**Read Energy:** As shown in Fig. 6(d), the read energies of both VCMA-MRAMs are always lower than those of SRAM, STT-MRAM, and SHE-MRAM, benefiting from the smaller cache area with reduced signal transmission energy.

**Read Latency:** As shown in Fig. 6(e), all the MRAMs have longer read latencies at a capacity of 32K bytes. However, the normalized latencies of MRAMs quickly decrease as capacity increases. In particular, STT-assisted VCMA-MRAM has a minimum read latency when capacity is greater than 1M bytes.

**Leakage Energy:** As expected, all the MRAMs significantly outperform SRAM in terms of leakage energy, as shown in Fig. 6(f), owing to the nonvolatility of the MTJ devices. Among them, VCMA-MRAMs achieve the minimum leakage energies for the same capacity, because they have the minimum area.

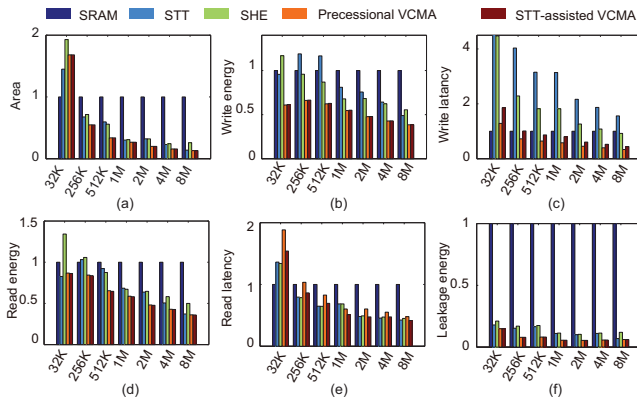


Fig. 6. Performance comparison in terms of area, write energy, write latency, read energy, read latency, and leakage energy, normalized to SRAM.

### C. Architectural Evaluation

Using Gem5, we performed an iso-capacity replacement for L2 (1M bytes) and/or L1 (32K + 32K bytes) caches with STT-MRAM, SHE-MRAM, precessional VCMA-MRAM, and STT-assisted VCMA-MRAMs, respectively. We compared the performances in terms of instruction-per-cycle (IPC) and cache energy consumption with those of the base-line SRAM-based cache through various benchmarks.

First, we consider a typical scenario wherein only the SRAM in L2 cache is replaced with MRAMs. Fig. 7 compares the IPC performance. It is evident that VCMA-MRAMs have similar IPCs to that of SRAM for most benchmarks. This suggests that VCMA-MRAMs can fulfill the performance requirements of the L2 cache. Further, Fig. 8 compares the energy consumption for L2 cache. MRAM technologies significantly reduce the leakage energy, owing the intrinsic nonvolatility of MTJ. Therefore, the total energy consumption for all MRAMs has a significant reduction across all benchmarks compared with that of SRAM. In particular, VCMA-MRAMs achieve the minimum total cache energy consumption, owing to the low-power voltage-induced MTJ switching mechanism. We can calculate that the average energy reduction with VCMA-MRAMs can achieve  $\sim 43.0\%$ ,  $\sim 28.2\%$ , and  $\sim 18.0\%$ , in comparison with that of SRAM, STT-MRAM, and SHE-MRAM, respectively. As a result, VCMA-MRAMs enable us to achieve maximum benefits in terms of IPC and energy consumption, when employed in the L2 cache.

Further, we consider another scenario wherein the SRAMs in both L1 and L2 caches are replaced with MRAM technologies. Fig. 9 compares the IPC performance for both L1 and L2 caches. As can be seen, STT-MRAM has a considerable decrease ( $\sim 4.2\%$ ) in IPC compared with that of SRAM, which indicates that STT-MRAM is not preferably employed for the L1 cache. For SHE-MRAM, the reduction of the IPC is approximately  $\sim 0.5\%$ . Nevertheless, it is evident that VCMA-MRAMs have comparable (or even larger) IPC, in comparison with that of SRAM, which implies that it is a potential nonvolatile memory candidate for the L1 cache. Fig. 10 compares the total energy consumption for both L1 and L2 caches. As can be seen that VCMA-MRAMs achieve a reduction of  $\sim 59.8\%$ ,  $\sim 17.6\%$ , as well as  $\sim 40.7\%$  in total energy consumption, compared with those of SRAM, STT-MRAM and SHE-MRAM, respectively.

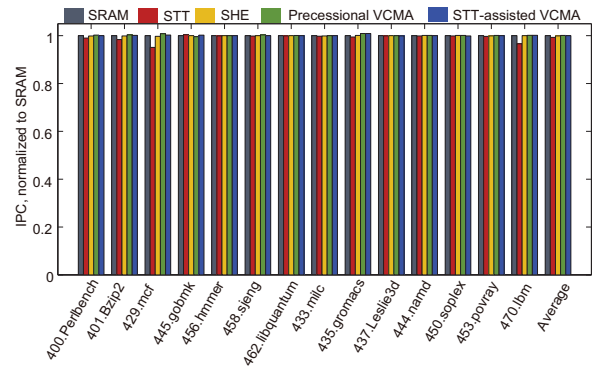


Fig. 7. Comparison of IPC among different memory technologies for L2 cache.

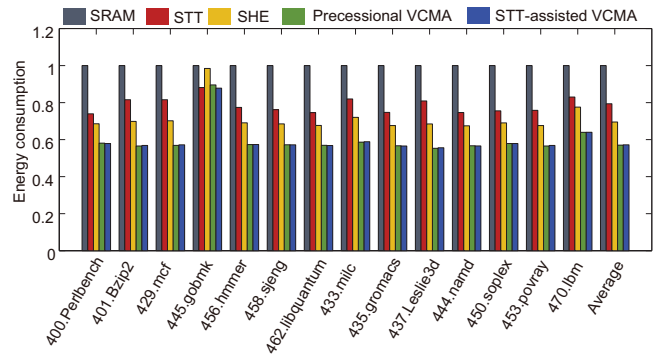


Fig. 8. Comparison of total energy consumption among different memory technologies for the L2 cache.

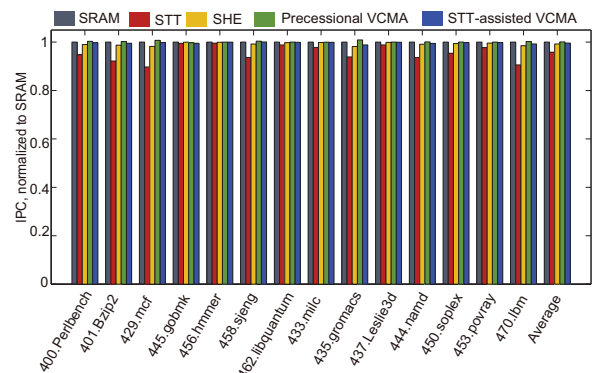


Fig. 9. Comparison of IPC among different memory technologies used for both L1 and L2 caches.

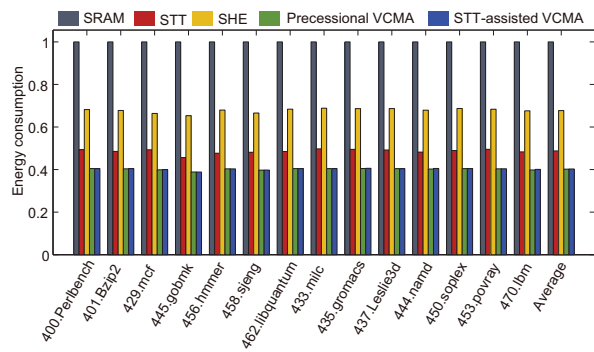


Fig. 10. Comparison of total energy consumption among different memory technologies used for both L1 and L2 caches.

In summary, our system-level investigations reveal that VCMA-MRAMs can fulfill the performance requirements of the L2 cache and bring benefits in reduction of cache area and energy consumption. Moreover, in addition to the capability of being employed for the L2 cache, VCMA-MRAMs have the potential to be utilized in the L1 cache.

#### IV. CHALLENGES OF VCMA-MRAM

This section discusses the challenges of VCMA-MRAM. As discussed above, the VCMA coefficient at the current stage is relatively small for effective voltage scaling, which is an issue in nanoscale technology nodes to meet the CMOS compatibility. Therefore further effort on developing novel material systems and device structures to improve this effect is strongly required. Fortunately, a recent study has reported that a rather large  $\xi = 290 \text{ fJ/V} \cdot \text{m}$  can be achieved with an ultrathin Fe layer in the MgO-based MTJ devices [23]. In this case, the required voltage amplitude can be greatly reduced. In circuit level design, we have shown that a precisely-controlled voltage pulse duration is required for the precessional VCMA-MRAM to induce deterministic MTJ switching, which, however, is rather difficult in practice when taking into consideration the PVT variations. To deal with this problem, a write-verify algorithm or circuit is generally required to ensure the data robustness [24], which adds the circuit complexity and degenerates the performance. For the STT-assisted VCMA-MRAM, although the MTJ switching is deterministic, the two voltage pulses for inducing the VCMA and STT effects should be carefully designed depending on the device parameters (e.g., oxide barrier thickness). In addition, it should be noted that the resistances of the MTJs in the VCMA-MRAMs are much larger than those of the CMOS transistors. How to make the MTJ devices be electrically compatible with the CMOS transistors is a key point in circuit design and chip fabrication. In system level design point of view, to exploit and benefit from the advantages features of VCMA-MRAM, novel architectures are preferable instead of a drop in replacement of SRAM or/and DRAM. For example, in VCMA-MRAM, the write/read asymmetry is inversed compared with that of STT-MRAM. Therefore, write-intensive data are more preferable to be stored in VCMA-MRAM for architecture design.

#### V. CONCLUSION

This paper studies the potential of using VCMA-MRAMs for cache design, through a cross-layer perspective from device to system. First, a physics-based VCMA-MTJ electrical model was developed and verified. Then, hybrid MTJ/CMOS circuit designs and evaluations of VCMA-MRAMs were performed by using the developed VCMA-MTJ model as well as a CMOS design-kit. Finally, system-level investigations were carried out to explore the prospects of VCMA-MRAMs in cache memories. Our simulation results demonstrate that VCMA-MRAMs offer great potential in improving the area, energy, and performance profiles for both L2 and L1 caches. However, challenges also exist and should be efficiently addressed before applications.

#### REFERENCES

- [1] International Technology Roadmap for Semiconductors (ITRS), 2013. Available: <http://www.itrs.net/>.
- [2] W. Kang, Y. Zhang, Z. Wang, et al., "Spintronics: Emerging ultra-low-power circuits and systems beyond MOS technology," *ACM J. Emerg. Technol. Comput.*, vol. 12, no. 2, article. 16, 2015.
- [3] C. Chappert, A. Fert, and F. N. Dau, "The emergence of spin electronics in data storage," *Nature Mater.*, vol. 6, no. 11, pp. 813–823, 2007.
- [4] D. Apalkov, A. Khvalkovskiy, S. Watts, et al., "Spin-transfer torque magnetic random access memory (STT-MRAM)," *J. Emerg. Technol. Circ. Syst.*, vol. 9 no. 13, pp.13:1, 2013
- [5] W. Kang, L. Zhang, J. O. Klein, et al., "Reconfigurable codesign of STT-MRAM under process variations in deeply scaled technology," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1769-1777, 2015.
- [6] Y. Seo, X. Fong, K. W. Kwon, and K. Roy, "Spin-Hall magnetic random-access memory with dual read/write ports for on-chip caches," *IEEE Magn. Lett.*, vol. 6, no. , pp. 1-4, 2015.
- [7] F. Oboril, R. Bishnoi, M. Ebrahimi, and M. B. Tahoori, "Evaluation of hybrid memory technologies using SOT-MRAM for on-chip cache hierarchy," *IEEE Trans. Comp. Aided. Design Integr. Circ. Syst.*, vol. 34, no. 3, pp. 367-380, March 2015.
- [8] S. P. Park, S. Gupta, N. Mojumder, A. Raghunathan, and K. Roy, "Future cache design using STT MRAMs for improved energy efficiency: devices, circuits and architecture," in *DAC*, New York, NY, USA, 2012, 492-497.
- [9] D. E. Nikonov and I. A. Young, "Benchmarking spintronic logic devices based on magnetoelectric oxides," *J. Mater. Res.*, vol. 29, no. 18, pp. 2109-2115, 2014.
- [10] W. G. Wang, M. Li, S. Hageman, and C. L. Chien, "Electric-field-assisted switching in magnetic tunnel junctions," *Nature Mater.*, vol. 11, no. 1, pp. 64-68, 2012.
- [11] J. G. Alzate, P. K. Amiri, P. Upadhyaya, et al., "Voltage-induced switching of nanoscale magnetic tunnel junctions," in *IEEE IEDM*, , pp. 29-5.1-4, 2012.
- [12] K. L. Wang, H. Lee, and P. K. Amiri, "Magnetoelectric random access memory-based circuit design by using voltage-controlled magnetic anisotropy in magnetic tunnel junctions," *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 992-997, 2015.
- [13] C. G. Duan, J. P. Velev, R. F. Sabirianov, et al., "Surface magnetoelectric effect in ferromagnetic metal films," *Phys. Rev. Lett.*, vol. 101, pp. 137201, 2008.
- [14] J. G. Alzate, "Voltage-controlled magnetic dynamics in nanoscale magnetic tunnel junctions," Ph.D. dissertation, *Dept. Elect. Eng. California Univ. LA*, 2014.
- [15] P. Khalili Amiri, J. G. Alzate, X. Q. Cai, et al., "Electric-field-controlled magnetoelectric RAM: progress, challenges, and scaling," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1-7, Nov. 2015.
- [16] D. V. Berkov and J. Miltat, "Spin-torque driven magnetization dynamics: micromagnetic modeling," *J. Magn. Magn. Mater.*, vol. 320, no. 7, pp. 1238-1259, 2008.
- [17] S. Kanai, Y. Nakatani, M. Yamanouchi, et al., "Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field-effect," *Appl. Phys. Lett.*, vol. 104, no. 21, pp. 212406, 2014.
- [18] J. C. Slonczewski, "Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier," *Phys. Rev. B.*, vol. 39, no. 10, pp. 6995, 1989.
- [19] W. F. Brinkman, R. C. Dynes, and J. Rowell, "Tunneling conductance of asymmetrical barriers," *J. Appl. Phys.*, vol. 41, no. 5, pp. 1915-1921, 1970.
- [20] S. Yuasa, T. Nagahama, A. Fukushima, et al., "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature Mater.*, vol. 3, no. 12, pp. 868-871, 2004.
- [21] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, "NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," *IEEE Trans. Comput. Aid. Design Integr. Circ. Syst.*, vol. 31, no. 7, pp. 994-1007, 2012.
- [22] N. Binkert, B. Beckmann, G. Black, et al., "The gem5 simulator," *SIGARCH Comput. Archit. News*, vol. 39, no. 2, pp. 1–7, 2011.
- [23] T. Nozaki, A. Kozioł-Rachwał, W. Skowroński, et al., "Large voltage-induced changes in the perpendicular magnetic anisotropy of an MgO-based tunnel junction with an ultrathin Fe layer," *Phys. Rev. Appl.*, vol. 5, no.4, pp. 044006, 2016.
- [24] H. Lee, J. G. Alzate, R. Dorrance, et al., "Design of a fast and low-power sense amplifier and writing circuit for high-speed MRAM," *IEEE Trans. Magn.*, vol. 51, no. 5, pp. 1-7, 2015.