

A Word Line Pulse Circuit Technique for Reliable Magnetoelectric Random Access Memory

Hochul Lee, Albert Lee, Shaodi Wang, Farbod Ebrahimi, Puneet Gupta, Pedram Khalili Amiri, *Member, IEEE*, and Kang L. Wang, *Fellow, IEEE*

Abstract—A word line pulse (WLP) circuit scheme is proposed toward the implementation of magnetoelectric random access memory (MeRAM). The circuit improves the write error rate (WER) and cell area efficiency by generating a better write pulse compared to conventional bitline pulse (BLP) techniques in terms of the pulse slew rate and amplitude. For the voltage-controlled magnetic anisotropy-induced precessional switching of the magnetic tunnel junction (MTJ), the write pulse shape has a large impact on the switching probability. Typically, a square shape pulse results in higher switching probability compared to that of a triangular shape pulse with long rise and falling edges, since the square shape pulse causes a more stable precessional trajectory of the free layer magnetization by providing a relatively constant in-plane-dominant effective field. Compared to the BLP scheme, the WLP can generate a better square shape pulse by eliminating discharge paths under the pulse condition, using the gain of the access transistor, and effectively diminishing the capacitive loading which needs to be driven. A macrospin compact model of voltage-controlled MTJ shows that the WLP can improve WER by 10^7 times and allow MeRAM to have four-time improvement in area efficiency of driver circuits compared to the BLP.

Index Terms—Magnetic tunnel junction (MTJ), magnetoelectric random access memory (MeRAM), voltage controlled magnetic anisotropy (VCMA), word line pulse (WLP), write error rate (WER).

I. INTRODUCTION

MAGNETIC tunnel junctions (MTJs) are promising next-generation memory devices which provide two distinct resistance states by changing their magnetic orientation. The spin transfer torque (STT) effect has been widely used to switch an MTJ state by creating a spin-polarized current [1]–[4]. However, the STT-based switching method intrinsically causes a significant ohmic loss (~ 100 fJ/bit), because a relatively large amount of charge current (>2 MA/cm²) is required to generate spin torque for switching.

Manuscript received October 18, 2016; revised January 9, 2017; accepted February 14, 2017. This work was supported by the NSF Phase II Small Business Innovation Research Project.

H. Lee and P. K. Amiri are with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095 USA, and also with Inston.Inc., Los Angeles, CA 90095 USA (e-mail: chul0524@ucla.edu; pedramk@gmail.com).

A. Lee, S. Wang, P. Gupta, and K. L. Wang are with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095, USA (e-mail: oncefriends9206@gmail.com; shaodiwang@ucla.edu; puneet@ee.ucla.edu; wang@ee.ucla.edu).

F. Ebrahimi is with Inston.Inc., Los Angeles, CA 90095 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVLSI.2017.2670502

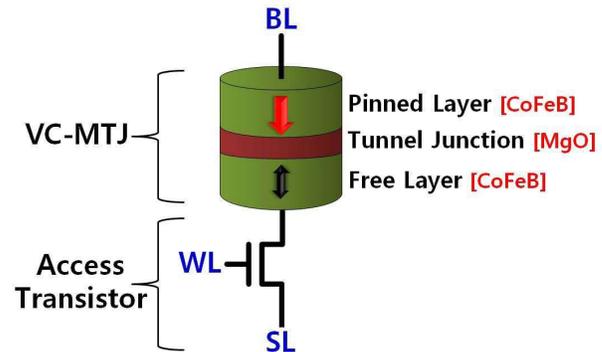


Fig. 1. MeRAM cell structure with a transistor as the access device and a VC-MTJ as the memory element. A BL and a source line are connected to the pinned layer of the VC-MTJ and the source of the access transistor, respectively. A WL controls the gate of the access transistor.

Recently, there has been increasing interest in precessional (i.e., resonant) switching via the voltage-controlled magnetic anisotropy (VCMA) effect, which ideally does not involve any Ohmic dissipation, resulting in low switching energy (down to ~ 1 fJ/bit) [5]–[9]. Also, the precessional switching of the perpendicularly magnetized voltage-controlled MTJ (VC-MTJ) offers the advantage of high-switching speed (down to ~ 100 ps).

Magnetoelectric random access memory (MeRAM) uses VC-MTJ as a memory element in its one transistor and one MTJ cell structure with an access transistor as shown in Fig. 1, taking advantage of high speed and low energy switching. Moreover, MeRAM also offers an advantage in terms of enhanced bit density, $\sim 8F^2$ unit cell area based on a special layout, because the VCMA effect allows the MeRAM cell to have a minimum size access transistor [10], and VC-MTJs can be fabricated on top of access transistors via back-end-of-line processes. Furthermore, MeRAM has very high endurance ($\sim 10^{16}$) compared to other nonvolatile memories, since the operation mechanism does not require physical displacement of atoms or ions, hence preventing permanent physical damage [11].

These characteristics allow MeRAM to be a promising candidate for main memory (e.g., dynamic RAM) and embedded (e.g., static RAM) memory applications. In addition to fulfilling the key performance requirements (high speed, high density, unlimited endurance, and low write energy) for embedded applications, MeRAM provides nonvolatility (retention time >10 years), reducing the static power of memory systems further, which is one of the primary concerns in volatile memory in advanced CMOS processes [12], [13].

However, to enable MeRAM to be realized in practical embedded system memory applications, low write error rate (WER) needs to be achieved. Write error in MeRAM is mainly caused by a degraded write pulse (e.g., slew rate and duration), and can limit its applications in high-speed memories. As an example, if WER is relatively high (e.g., $\sim 10^{-3}$), multiple write operations are required to achieve an acceptable bit error rate (BER) (i.e., $< 10^{-9}$) [14]; hence, the total write access time could become too long to meet the speed requirement of embedded system memory. Although the WERs based on the VCMA effect assisting STT or spin Hall effect switching methods are less sensitive to the write pulse shape and duration, they require additional time and energy compared to pure VCMA driven precessional switching [15]–[17].

The main contribution of this paper is to address the WER challenge as described above by using a new scheme to improve the write pulse shape for a significant reduction of WER based on four-time smaller word line (WL) and bitline (BL) drivers compared to a conventional method.

The remainder of this paper is organized as follows. Section II briefly explains the physics of the VC-MTJ, its precessional switching process, and the macrospin VC-MTJ compact model simulations. Section III introduces the proposed word line pulse (WLP) scheme and its simulation results. The evaluation of the WLP is discussed in Section IV, the scalability of MeRAM is analyzed in Section V, and the conclusions are drawn in Section VI.

II. COMPACT MODEL OF MAGNETIC TUNNEL JUNCTION

A VC-MTJ consists of two ferromagnetic layers (e.g., CoFeB) separated by a tunneling barrier (e.g., MgO) as shown in Fig. 1 where the magnetic moment of one layer (pinned layer) is fixed, and the other layer (free layer) can freely switch its magnetization via electrical or magnetic bias condition. Two energetically stable states exist in this device based on the magnetization of the free layer. The parallel state (P) occurs when the magnetic moments of the both layers are aligned in the same direction giving rise to a low resistance (R_P); on the other hand, the antiparallel state (AP) occurs when the magnetic moment of the free layer is magnetized in the opposite direction to that of the pinned layer giving rise to a high resistance (R_{AP}).

To make perpendicularly magnetized VC-MTJs, the interfacial perpendicular magnetic anisotropy (PMA) should be enhanced by adjusting the thickness of the ferromagnetic layers and using suitable materials for fabrication [9]. Perpendicularly magnetized devices typically have ultrathin (<2 nm) ferromagnetic layers. Since electric fields cannot be screened by the conductivity of the ferromagnetic material, the magnetic properties (e.g., PMA) are modulated by the applied electric field due to interface effects [11].

The switching dynamics of the free layer magnetic moment (\vec{m}) is described via a Landau–Lifshitz–Gilbert (LLG) equation in the presence of an effective field (\vec{H}_{eff}) [18]

$$\frac{d\vec{m}}{dt} = -\gamma' \vec{m} \times (\vec{H}_{\text{eff}} + \vec{H}_{\text{th}}) - \alpha \gamma' \vec{m} \times (\vec{m} \times \vec{H}_{\text{eff}}) \quad (1)$$

where α is the damping factor, \vec{m} is a unit vector in the direction of the magnetization, and γ' is the absolute value of the gyromagnetic ratio and the relative permeability of the free layer ($\gamma_e \mu_0$) divided by $(1 + \alpha^2)$. \vec{H}_{th} is the thermal noise term. Although the LLG equation in the compact model is expanded to account for the STT effect and field-like torque, (1) does not show them because their effects are negligible in the VC-MTJ due to a relatively thick MgO thickness (> 1.5 nm). The first term in the (1) is responsible for precessional motion, the circular rotation around the unit sphere, and the second term provides a damping torque that makes \vec{m} align with \vec{H}_{eff} . \vec{H}_{eff} can be represented by several magnetic anisotropies

$$\vec{H}_{\text{eff}} = \vec{H}_{\text{PMA}} - \vec{H}_{\text{VCMA}} + \vec{H}_{\text{ext}} + \vec{H}_{\text{dem}}. \quad (2)$$

\vec{H}_{PMA} is the PMA, which is observed in the interface between an MgO and a CoFeB layer, and \vec{H}_{VCMA} is the VCMA effect that modulates \vec{H}_{PMA} based on the amplitude and polarity of the applied voltage. These two components are modeled as follows [8], [11]:

$$\vec{H}_{\text{PMA}} = \frac{2K_i}{t_{\text{fl}} \mu_0 M_S} m_z \hat{z} \quad (3)$$

$$\vec{H}_{\text{VCMA}} = \frac{2\xi V_{\text{MTJ}}}{t_{\text{fl}} \mu_0 M_S d_{\text{MgO}}} m_z \hat{z} \quad (4)$$

where K_i is the PMA coefficient (interfacial anisotropy), ξ is the VCMA coefficient, V_{MTJ} is the voltage across the VC-MTJ, t_{fl} and d_{MgO} are thickness of the free layer and the tunnel oxide, respectively, and M_S is the saturation magnetization. \vec{H}_{ext} is the external magnetic field, which is represented with a constant in-plane field (\hat{x} -axis) in the compact model. \vec{H}_{dem} is the demagnetization field, described by (5), originating from the shape anisotropy where N_z and $N_{x,y}$ are the demagnetizing tensors for (z -axis) and (x - and y -axis), respectively

$$\vec{H}_{\text{dem}} = (N_z - N_{x,y}) M_S \hat{z}. \quad (5)$$

The thermal noise, which creates random fluctuations of the free layer magnetization, is included in the precessional term of LLG equation and modeled by

$$\vec{H}_{\text{th}} = \vec{\sigma} \sqrt{\frac{2k_B T \alpha}{\mu_0 M_S \gamma' v \Delta t}} \quad (6)$$

where k_B is the Boltzmann's constant [J/K], T is the temperature [K], v is the volume of the free layer [m^3], and Δt is the simulation time step. $\vec{\sigma}$ is a unit vector whose \hat{x} , \hat{y} , and \hat{z} components are independent Gaussian random variables with a mean of 0 and a standard deviation of 1. These components are produced using the built in Verilog-A random number generator functions.

Precessional switching is done via the use of a short pulse across the VC-MTJ, as shown in Fig. 2. At the zero bias condition (at t_0), the magnetic moment \vec{m} of the free layer is aligned with the out-of-plane \hat{z} axis, since the PMA (\vec{H}_{PMA}) is the dominant component of \vec{H}_{eff} . However, actual \vec{m} is slightly tilted (13.5°) to one side due to the in-plane (\hat{x} axis) \vec{H}_{ext} . When the voltage bias condition is supplied ($V_{\text{MTJ}} = V_P$) between t_1 and t_2 , the VCMA effect cancels out the PMA,

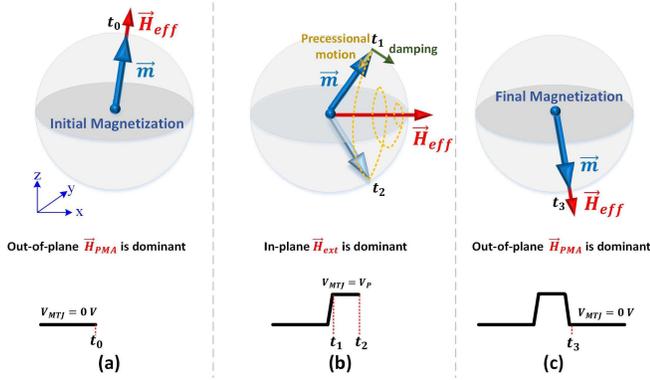


Fig. 2. Illustration of VCMA-induced precessional switching mechanism in the free layer of a perpendicularly magnetized VC-MTJ. (a) Under zero electric bias condition ($V_{MTJ} = 0V$ at $t < t_0$), the free layer is aligned with the out-of-plane direction because the PMA \vec{H}_{PMA} is a dominant component in \vec{H}_{eff} . (b) When an applied voltage across the device reduces \vec{H}_{PMA} via the VCMA effect, the magnetic moment starts to precess around the in-plane direction. (c) If the width of the applied pulse is designed to coincide with half the precession period, a full 180° switching can be achieved. Note that voltage with opposite polarity cannot switch the device because it enhances \vec{H}_{PMA} .

TABLE I
PARAMETERS OF THE MACROSPIN COMPACT MODEL

Parameters	Symbol	Value	Unit
MTJ diameter	l	60	nm
MgO thickness	d_{MgO}	1.62	nm
Free layer thickness	t_{fl}	1.1	nm
TMR	TMR	100	%
Temperature	T	300	K
Damping factor	α	0.02	
Saturation magnetization	M_s	1.2×10^6	A/m
PMA coefficient	K_i	1.06×10^{-3}	J/m ²
VCMA coefficient	ξ	61	fJ/V · m
Demagnetizing tensor (z)	N_z	0.96	
Demagnetizing tensor (x,y)	$N_{x,y}$	0.02	

temporarily changing \vec{H}_{eff} into an in-plane dominant field. This in-plane dominant \vec{H}_{eff} causes precessional and damping motion of the magnetic moment. If the timing of the applied voltage is well controlled (e.g., the pulse is removed at t_3 , near the half cycle of the precessional motion), the free layer of the VC-MTJ achieves a 180° reorientation.

In addition to the timing of the applied pulse, the dynamics of the magnetic moment is largely affected by the pulse shape, in particular, the rising and falling time of the pulse. To create a stable precessional motion, \vec{H}_{eff} needs to be a constant field, pointing in the in-plane direction, during the electric bias condition. Otherwise, the trajectory of the magnetic moment deviates from the precessional route. Fig. 3 shows two simulation results of the magnetization dynamics based on the macrospin model where a square shaped pulse [see Fig. 3(a)] and a triangular shaped pulse [see Fig. 3(b)] are applied. Fig. 3(c) and (d) shows 3-D magnetic moment trajectories induced by a square shaped pulse and a triangular shaped pulse, respectively. Parameters of the macrospin compact model are shown in Table I. In the case of applying a square pulse, the initial state of free layer is the P state ($m_z = -1$) and it starts to

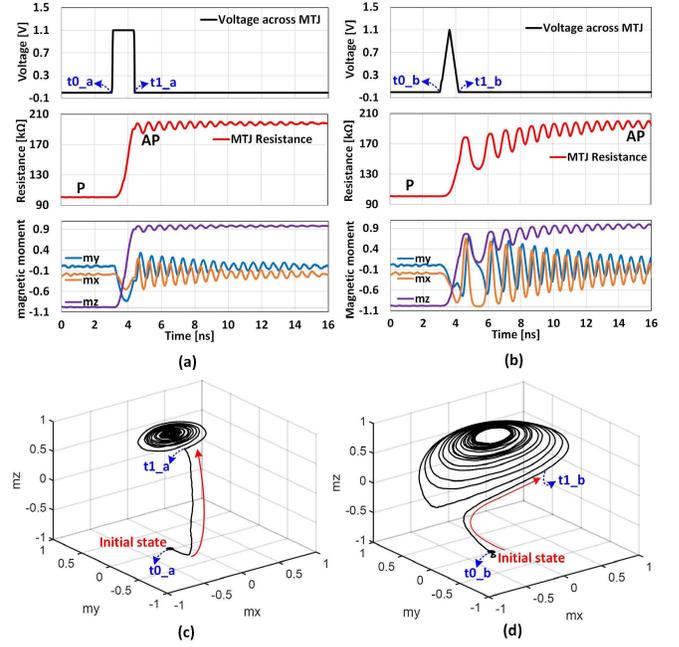


Fig. 3. Macrospin compact model transient simulations for (a) square shape write pulse (b) triangular shape write pulse. (c) Magnetic moment based on the square pulse shows a more stable precessional trajectory compared to (d) triangular pulse based one. This allows the square pulse driven switching to have a low WER by reducing the susceptibility to noise.

precess around the \hat{x} axis at $t = t0_a$. Since the PMA is abruptly reduced by the VCMA, which gives rise to a relatively constant in-plane component of \vec{H}_{eff} , the magnetic moment of the free layer can have a stable precessional trajectory and switch to the AP state ($m_z \approx 1$) at $t = t1_a$. However, in the case of applying a triangular pulse, the direction of \vec{H}_{eff} is no longer in-plane. Instead, \vec{H}_{eff} gradually changes its direction from out-of-plane to in-plane as a function of time, which in turn causes an unstable precessional motion. At the endpoint of the triangular pulse ($t = t1_b$), the magnetic moment cannot reach 180° reorientation ($m_z \approx 0.72$). After removing the pulse, the magnetic moment converges to the AP state via the damping and precessional motion driven by the PMA. During this process, the device becomes susceptible to thermal noise, which can produce a switching fail and increase the WER.

III. PROPOSED WL PULSE SCHEME

A. Method

Instead of applying the write pulse to the BL, called BLP, we propose a method of applying the write pulse to the WL, which is referred to as WLP. The WLP can create a better square shaped write pulse across the VC-MTJ, which in turn improves switching probability, and minimize the area overhead (e.g., driver size). There are three reasons why the WLP can have a better pulse shape compared to the conventional bitline pulse (BLP) scheme [14]: 1) eliminating discharge path during applying a pulse on the WL; 2) using the gain of the access transistor; and 3) reducing the capacitive loading which needs to be driven. Further explanation of the reasons will be discussed in the following.

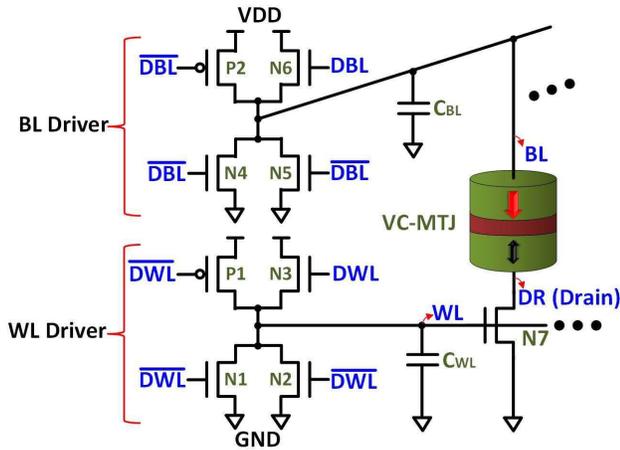


Fig. 4. Schematic of cell array architecture, including the BL driver and WL driver. The number of access transistors connecting the WL and the WL length determines its capacitive loading C_{WL} . The number of VC-MTJs connecting the BL and the BL length decide its capacitive loading C_{BL} . The size of the drivers are carefully chosen based on the magnitude of the capacitance of each line to generate a suitable pulse shape.

Fig. 4 shows a schematic design where MeRAM cells are connected to the BL driver and the WL driver. In this simulation, we implement the macrospin compact model into the circuit design simulation environment (e.g., Cadence Virtuoso). To achieve a fair performance comparison between the WLP and the BLP, both drivers consist of the same size transistors. We assume that the bitline capacitance C_{BL} is equal to the WL capacitance C_{WL} . Note that there are n-channel transistors (N3 and N6) in the pull-up path of each driver. These n-channel transistors can supply a large amount of charge at the beginning of the charge up for writing, compared to that of the same sized p-channel transistors, due to higher mobility, resulting in a better square pulse shape. But they gradually turn OFF as the potential of the target node (WL or BL) increases.

The control signals of the conventional BLP are shown in Fig. 5(a) where the Driver's input for Word Line (DWL) and the Driver's input for Bit Line (DBL) enable the WL driver and the BL driver, respectively. The \overline{DWL} and \overline{DBL} are their complementary signals. We assume that the rising and falling time of the control signals are 100 ps. For the BLP, the WL driver is enabled first at $t = t_{WL,a}$, which charges up the selected WL to Voltage Drain-to-Drain (VDD), turning on the access transistor (N7). Then, the DBL triggers the BL driver that starts to charge up the BL at $t = t_{BL,a}$. However, this scheme deforms the write pulse shape because the BL driver directly drives the entire BL capacitance loading C_{BL} , and some portion of the electric charge leaks through the unselected MeRAM cells, which prohibits the BL from reaching VDD within a 1 ns period and increases the rising time of the write pulse.

By contrast, in the case of the WLP, the waveform of the control signals (DWL and DBL) are shown in Fig. 5(b). The BL is charged first up to VDD, and the drain (DR) of the access transistor (N7) is also charged up to VDD since the N7 turns OFF at $t = t_{BL,b}$. Then, the WL driver is enabled and

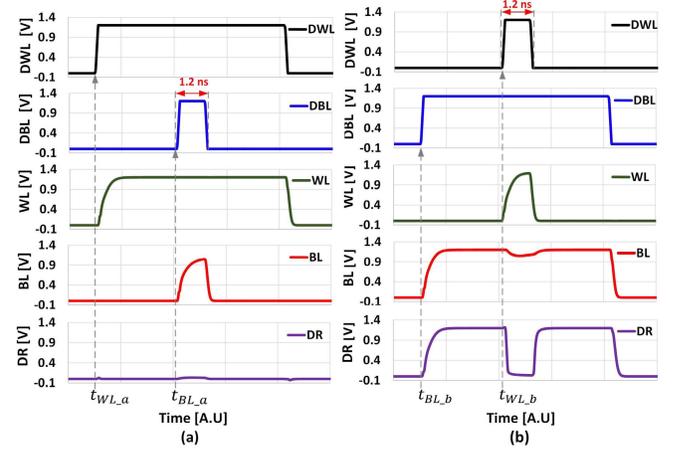


Fig. 5. (a) Conventional BLP scheme. After the WL driver completes charging the WL up to VDD, the BL driver applies a write pulse to the BL. (b) Proposed WLP scheme. The BL and DR are precharged to VDD, and then, a write pulse is applied to the WL via the WL driver. The WLP can make a better square shape write pulse based on the same size driver compared to the write pulse from the BLP.

starts to increase the WL potential at $t = t_{WL,b}$. The slew rate of the WL is improved by 20% compared to that of the BLP since the gate of the access transistor (N7) provides a high input resistance, eliminating a discharge path. Furthermore, the WLP can efficiently utilize the current gain of the access transistor N7 through a common-source stage. Even below the threshold of the N7, the current flowing through N7 exponentially increases as a function the WL voltage. Above the threshold, the provided current increases quadratically as the WL voltage increases further.

Last but foremost, the WL voltage rapidly discharges the DR node to ground via the N7 transistor, since the capacitance loading on the DR node consists only of the VC-MTJ and the access transistor (N7) itself, which is significantly smaller than the C_{BL} . The effects create a better square shape pulse across the VC-MTJ, allowing the circuit to achieve more reliable write operation.

B. Simulation Results

The resistance and capacitance of the BL and the WL in an array level can be calculated via the cell dimension, the sheet resistance, and the capacitance per unit length. We assumed that the sheet resistances of the metal layer for the BL and WL is $0.14 \Omega/\square$, and the capacitance per unit length is $0.2 \text{ fF}/\mu\text{m}$ when the metal width is equal to $0.1 \mu\text{m}$. Based on 28-nm technology node with $25F^2$ cell size (F is the minimum feature size), the dimension of the unit cell is $0.14 \mu\text{m} \times 0.14 \mu\text{m}$. If the width and length of the access transistor (a standard logic transistor) are 100/30 nm, its gate capacitance and junction capacitance are 57 and 48 aF, respectively. Table II shows the estimated values of RC loading on the WL and the BL in the array.

The voltage across the VC-MTJ (V_{MTJ}) is the potential difference between the BL and the DR nodes ($V_{BL} - V_{DR}$). Fig. 6 shows V_{MTJ} with a corresponding VC-MTJ resistance change based on the BLP (black) and the WLP (red) as a function of the capacitive loading on the BL and the WL.

TABLE II
RESISTIVE AND CAPACITIVE LOADS ON THE BL AND THE WL

# of cells	BL			WL		
	C [fF]	R [Ω]	RC [ps]	C [fF]	R [Ω]	RC [ps]
128	8	45	0.4	9	45	0.4
256	17	90	1.5	19	90	1.7
512	33	179	5.9	38	179	6.8
1024	66	358	23.6	75	358	26.9

*28 nm node, $25F^2$ unit cell size

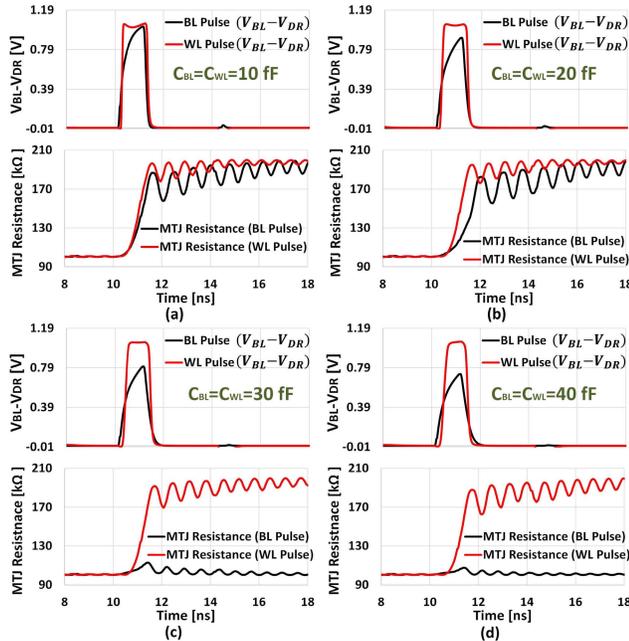


Fig. 6. Circuit simulation results of the BLP (black line) and the WLP (red line) based on the same size driver (the minimum size driver $W/L = 160/120$ nm) with the same magnitude of the BL and WL capacitances of (a) 10 fF; (b) 20 fF; (c) 30 fF; and (d) 40 fF. The write pulse from the BLP becomes degraded as the capacitance increases, and fails to switch the device beyond 30 fF. The WLP generates a square shape write pulse even under the largest loading of 40 fF and succeeds in switching the device.

As the capacitive loadings increase, the write pulse is severely degraded especially in the BLP case. Eventually, it fails to switch the VC-MTJ with $C_{BL} = 30$ fF, which is approximately equivalent to the number of 512 memory cells on the BL, since the pulse become a triangular shape and its amplitude also diminishes. In contrast, the WLP generates a square shape pulse regardless of the amount of the capacitive loading (within the capacitance range for the simulation), successfully switching the VC-MTJ state, although the slew rate is slightly increased. A quantitative evaluation of switching probability (or WER) will be discussed in the next section.

IV. EVALUATION

A. Write Error Rate

The WER, defined as the number of switching failures divided by a total number of write trials, is an important indicator to evaluate the performance of a write operation. Specifically, the WER influence the total access time of a memory system [14]. Because if a memory cell has a high WER at given write pulse, multiple write operations are nec-

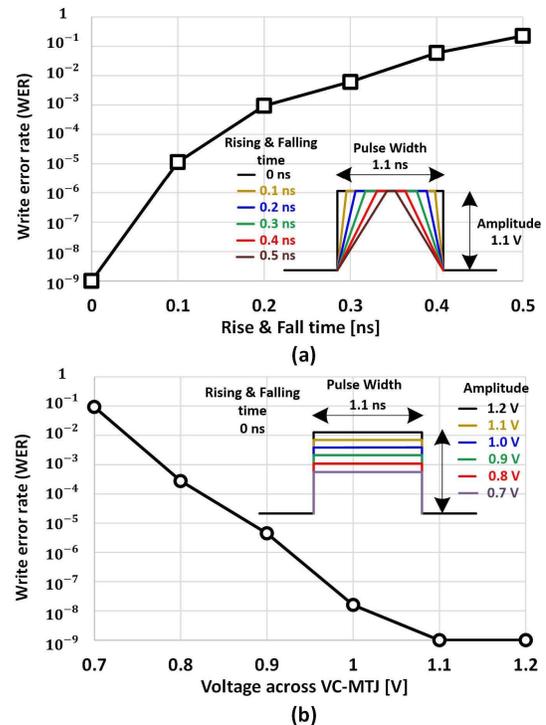


Fig. 7. WER simulation with an ideal voltage source (a) as a function of rise and fall time (slew rate); and (b) as a function of amplitude.

essary to achieve an acceptable BER, which is the maximum WER that can be successfully managed by an error correction code algorithm built in the memory system [19].

In order to understand which component is a dominant factor on the WER between the slew rate and the amplitude of the write pulse, we independently executed the WER via the macrospin compact model simulations with an ideal voltage source based on two conditions as shown in Fig. 7(a) and (b), respectively. Fig. 7 shows that the both components influence on the WER in an exponential manner since the energy barrier between the two states linearly decreases as a function of the applied voltage, and the slew rate decides the trajectory of the magnetic moment and the effective pulsewidth.

To quantitatively evaluate the performance of the BLP and the WLP, the WER of both cases are extracted via 10^{10} trials under the condition where both BL and WL drivers use the minimum size transistors for the fair comparison [20]. However, in the case of actual memory design, the size of drivers should be adjusted with respect to the capacitive loading to achieve an acceptable BER.

Fig. 8 shows the comparison result between the WER of the BLP and that of the WLP. Since the BLP fails to generate a proper write pulse in terms of slew rate and amplitude, the WER of the WLP is on average seven orders of magnitude lower than that of the BLP through the given capacitive loading (10–40 fF) based on the minimum size driver.

Note that the WER of the BLP with low capacitive loading (10 fF) is mainly due to the slew rate (rising time > 0.3 ns) compared to that of the WLP because both schemes reach the same amplitude (~ 1.1 V), as shown in Fig. 6(a). However,

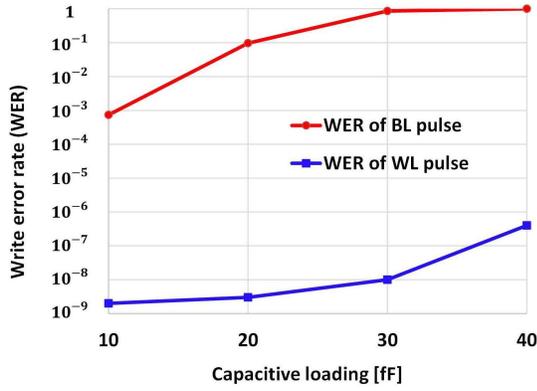


Fig. 8. WERs of the BLP and the WLP with respect to the capacitive loading via 10^{10} transient simulation trials with the minimum size driver. The WLP achieves on average seven orders of magnitude lower WER as compared to that of the BLP under the same condition (e.g., driver size, loading).

as the capacitive loading increases, the amplitude of the write pulse becomes the main reason for such high WER of the BLP. Because the amplitude decreases with a faster rate compared to that of the slew rate (see Fig. 6), which exponentially increases the WER [see Fig. 7(b)].

B. Cell Area Efficiency

A simple way to improve the write pulse shape is to increase the size of the transistor in the drivers associated with the loadings on the BL or the WL. However, the increase in the driver size might limit the memory capacity in a given die area, resulting in low cell area efficiency. The cell area efficiency is typically used as a target parameter to compare the compactness of memory designs, which is defined as follows:

$$\begin{aligned} \text{CellAreaEfficiency} &= \frac{\text{Area}(\text{cellarray})}{\text{TotalDieSize}} \\ &= \frac{\text{Area}(\text{cellarray})}{\text{Area}(\text{cellarray}) + \text{Area}(\text{logicandanalogsicuit})}. \end{aligned} \quad (7)$$

To achieve a high cell area efficiency the logic and analog circuits should be minimized while fulfilling the required performances such as speed and power. In a typical MeRAM macro design, the BL drivers and WL drivers occupy 14% and 4% in the total die area, respectively. Such a high area occupancy rate for the BL drivers results from the fact that the BL drivers consist of relatively large size transistors, which need to drive a significant amount of capacitive loads within a nanosecond [21]. If we put the area of the logic and analog circuits into the (7), the cell area efficiency is 67.8%.

A reduction in the driver size may improve the cell area efficiency. The proposed WLP allows reducing the driver size while generating a good square shape pulse. Fig. 9(a) shows a required write pulse shape that provides an acceptable BER ($< 10^{-9}$). Fig. 9(b) presents a normalized driver size, which can generate the required write pulse shape, in a given capacitive loading. Compared to the driver size of the BLP, the WLP can produce the required write pulse by using a four-time smaller driver. Therefore, the WLP has the potential to reduce the occupancy rate of the drivers in the total die size, resulting in 76.8% cell area efficiency.

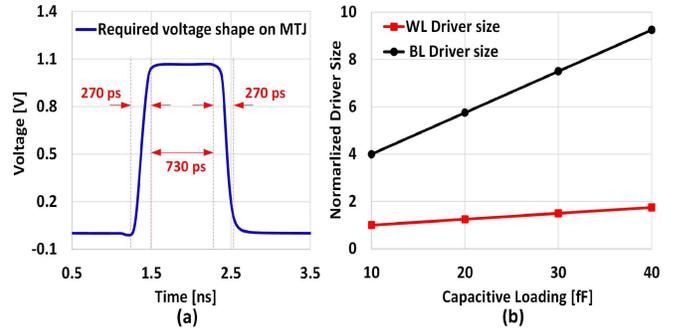


Fig. 9. (a) Required pulse shape that achieves an acceptable BER ($< 10^{-9}$). (b) Normalized driver size associated with a given capacitive loading to generate the required pulse shape. Note that the WLP allows a chip to have four times the area efficiency of drivers as compared to the BLP.

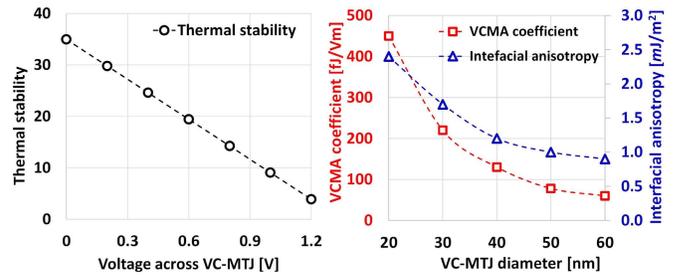


Fig. 10. (a) Thermal stability as a function of voltage across the VC-MTJ. (b) Required VCMA coefficient and interfacial anisotropy as the VC-MTJ size is scaled while maintain the same value of the critical voltage ($V_C/d_{MgO} = 1$ V/nm) and thermal stability ($\Delta_0 = 40$).

V. SCALABILITY OF MERAM

The thermal stability Δ is one of the most significant metrics that evaluate a memory cell characteristics, especially, for retention time, and can be calculated via the following equation:

$$\Delta = \frac{E_b}{k_B T} = \frac{\mu_0 M_s H_{k,\text{eff}} V}{2k_B T} \quad (8)$$

where E_b is the energy barrier between the two stable state of the VC-MTJ, V is the volume of the free layer, and μ_0 is the permeability. $H_{k,\text{eff}}$ is the sum of the three anisotropies: the PMA, VCMA, and demagnetization. Since the $H_{k,\text{eff}}$ is a function of voltage across the VC-MTJ (V_{MTJ}), the thermal stability Δ also changes with respect to V_{MTJ} as shown in Fig. 10(a). At zero bias condition with given VC-MTJ parameters in Table I, the thermal stability is 35, which is suitable for working memory applications [22]. However, for the storage class applications, the thermal stability should be ranging from 40 to 60 depending on the capacity of the memory chip.

The ratio between critical switching current and the thermal stability (I_C/Δ_0) is an indicator of the scalability of STT-RAM where Δ_0 is thermal stability at zero bias. Similarly, the scalability of MeRAM can be analyzed by the analogous critical voltage over thermal stability (V_C/Δ_0), meaning that any voltage larger than V_C can reconfigure the magnetic easy-axis to in-plane at a given thermal stability [23]. As shown in the (8), the thermal stability of the VC-MTJ is a function of the energy barrier, which is proportional

to $K_i AV_{\text{MTJ}}$ where the A is the device area. Hence, the interfacial anisotropy K_i needs to be increased as the VC-MTJ size scales while maintaining the same level of thermal stability.

The ratio of critical voltage over thermal stability V_C/Δ_0 can be represented as $d_{\text{MgO}}kT/\zeta A$. Therefore, as the device area scales down, the VCMA coefficient needs to be increased quadratically to compensate for the reduction of the VC-MTJ area A to keep the same rate of the energy barrier controllability by using the applied voltage. Fig. 10(b) shows the required VCMA coefficient and interfacial anisotropy as the VC-MTJ size is scaled while maintaining the same value of the critical voltage ($V_C/d_{\text{MgO}} = 1$ V/nm) and thermal stability ($\Delta_0 = 40$).

VI. CONCLUSION

We have proposed a WLP scheme that allows MeRAM to generate a better square shape write pulse compared to that of a conventional BLP scheme. The macrospin compact model simulation shows that a square pulse provides a stable precessional trajectory of the free layer magnetization, allowing a VC-MTJ to have greater immunity to thermal fluctuations. Under the same schematic condition (e.g., driver size, magnitude of loading), the WLP significantly outperforms the BLP in terms of WER and cell area efficiency.

ACKNOWLEDGMENT

The authors would like to thank the NSF Phase II Small Business Innovation Research Project for performing this work at Inston Inc.

REFERENCES

- [1] 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, p. 3118, Apr. 2004.
- [2] J. A. Katine and E. E. Fullerton, "Device implications of spin-transfer torques," *J. Magn. Magn. Mater.*, vol. 320, no. 7, pp. 1217–1226, Apr. 2008.
- [3] S. Ikeda *et al.*, "A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction," *Nature Mater.*, vol. 9, no. 9, pp. 721–724, Sep. 2010.
- [4] D. C. Worledge *et al.*, "Spin torque switching of perpendicular Ta/CoFeB/MgO-based magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 98, no. 2, p. 22501, Jan. 2011.
- [5] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, "Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses," *Nature Mater.*, vol. 11, no. 1, pp. 39–43, Jan. 2012.
- [6] 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, Jan. 2012.
- [7] S. Kanai, M. Yamanouchi, S. Ikeda, Y. Nakatani, F. Matsukura, and H. Ohno, "Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 101, no. 12, p. 122403, Sep. 2012.
- [8] J. G. Alzate *et al.*, "Voltage-induced switching of nanoscale magnetic tunnel junctions," in *Proc. Int. Electron Devices Meeting*, Dec. 2012, pp. 29.5.1–29.5.4.
- [9] K. L. Wang, J. G. Alzate, and P. K. Amiri, "Low-power non-volatile spintronic memory: STT-RAM and beyond," *J. Phys. D, Appl. Phys.*, vol. 46, no. 7, p. 74003, Feb. 2013.
- [10] 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, Nov. 2015.
- [11] P. K. Amiri *et al.*, "Electric-field-controlled magnetoelectric RAM: Progress, challenges, and scaling," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–7, Nov. 2015.
- [12] J. A. Mandelman *et al.*, "Challenges and future directions for the scaling of dynamic random-access memory (DRAM)," *IBM J. Res. Develop.*, vol. 46, nos. 2–3, pp. 187–212, Mar. 2002.
- [13] M. Qazi, M. Sinangil, and A. Chandrakasan, "Challenges and directions for low-voltage SRAM," *IEEE Design Test Comput.*, vol. 28, no. 1, pp. 32–43, Jan. 2011.
- [14] H. Lee *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, May 2015, Art. no. 3400507.
- [15] S. Kanai *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, p. 212406, May 2014.
- [16] W. Kang, Y. Ran, W. Lv, Y. Zhang, and W. Zhao, "High-speed, low-power, magnetic non-volatile flip-flop with voltage-controlled, magnetic anisotropy assistance," *IEEE Magn. Lett.*, vol. 7, pp. 1–5, Aug. 2016, Art. no. 3106205.
- [17] H. Lee, F. Ebrahimi, P. K. Amiri, and K. L. Wang, "Low-power and high-density spintronic programmable logic (SPL) Using Voltage-Gated Spin Hall Effect in Magnetic Tunnel Junctions," *IEEE Magn. Lett.*, to be published.
- [18] L. D. Landau and E. M. Lifshits, "On the theory of the dispersion of magnetic permeability in ferromagnetic bodies," *Phys. Zeitsch. Sow.*, vol. 8, pp. 153–169, Jan. 1935.
- [19] M. Fukuda, K. Higuchi, and K. Takeuchi, "Non-volatile random access memory and NAND flash memory integrated solid-state drives with adaptive codeword error correcting code for 3.6 times acceptable raw bit error rate enhancement and 97% power reduction," *Jpn. J. Appl. Phys.*, vol. 50, no. 4, p. 04DE09, Apr. 2011.
- [20] S. Wang, H. Lee, F. Ebrahimi, P. K. Amiri, K. L. Wang, and P. Gupta, "Comparative evaluation of spin-transfer-torque and magnetoelectric random access memory," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 2, pp. 134–145, Sep. 2016.
- [21] J. Kim and M. C. Papaefthymiou, "Constant-load energy recovery memory for efficient high-speed operation," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, 2004, pp. 240–243.
- [22] K. Ikegami *et al.*, "MTJ-based 'normally-off processors' with thermal stability factor engineered perpendicular MTJ, L2 cache based on 2T-2MTJ cell, L3 and last level cache based on 1T-1MTJ cell and novel error handling scheme," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2015, pp. 25.1.1–25.1.4.
- [23] P. K. Amiri, P. Upadhyaya, J. G. Alzate, and K. L. Wang, "Electric-field-induced thermally assisted switching of monodomain magnetic bits," *J. Appl. Phys.*, vol. 113, no. 1, p. 13912, 2013.



circuit.

Hochul Lee (S'13) received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2005 and the M.S. degree from the Semiconductor Material Device Laboratory, Seoul National University, Seoul, South Korea. He was with the Samsung Electronics Flash Memory Circuit Design Team until 2012.

He is currently pursuing the Ph.D. degree the Device Research Laboratory, University of California at Los Angeles, Los Angeles, CA, USA, with a focus on exploring MTJs-based hybrid CMOS



Albert Lee received the B.S. and M.S. degrees in electrical engineering from National Tsing-Hua University, Hsinchu, Taiwan. He is currently pursuing the Ph.D. degree at the Device Research Laboratory, University of California at Los Angeles, Los Angeles, CA, USA.

His current research interests include nonvolatile logics, emerging nonvolatile memory, and neuromorphic circuits.



Shaodi Wang (S'12) received the B.S. degree from the Division of Microelectronic, Electronics Engineering and Computer Science Department, Peking University, Beijing, China and the M.S. degree in electrical engineering from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, where he is currently pursuing the Ph.D. degree with the NanoCAD Laboratory, Department of Electrical Engineering, under the supervision of Prof. P. Gupta.

His current research interests include emerging memory and device technology circuit- and system-level design, evaluation and optimization, and modeling for manufacturing.



Farbod Ebrahimi received the B.E.E. degree and the M.S. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2010 and 2012, respectively.

In 2013, he joined the Device Research Laboratory (DRL), University of California at Los Angeles, Los Angeles, CA, USA, as a Visiting Scientist. In 2014, alongside his position at DRL, he joined Inston Inc., a startup working toward the realization of MeRAM. He is currently exploring voltage controlled magnetic tunnel junction (MTJ) for memory

and logic applications, and spin-wave systems for RF and logic applications.



Puneet Gupta (M'07) received the B.Tech. degree in electrical engineering from IIT Delhi, New Delhi, India, in 2000 and the Ph.D. degree from the University of California at San Diego, La Jolla, CA, USA, in 2007.

He is currently a Faculty Member of the Electrical Engineering Department at the University of California at Los Angeles (UCLA), Los Angeles, CA, USA. He co-founded Blaze DFM Inc., in 2004 and served as product architect until 2007.

He has authored over 130 papers and holds 16 U.S.

patents. His current research interests include building high-value bridges across application-architecture and implementation-fabrication interfaces for lowered cost and power, increased yield and improved predictability of integrated circuits and systems.

Dr. Gupta was a recipient of the NSF CAREER Award, the ACM/SIGDA Outstanding New Faculty Award, and the IBM Faculty Award.



Pedram Khalili Amiri (M'05) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2004 and the Ph.D. degree (*cum laude*) in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2008.

He joined the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA, in 2009, where he is currently an Assistant Adjunct Professor.

Dr. Amiri has served as a Guest Editor for Spin and on the Technical Program Committee of the Joint MMM/Intermag Conference.



Kang L. Wang (F'92) received the B.S. degree from the National Cheng Kung University, Taiwan, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, USA.

He is currently a Distinguished Professor and holds the Raytheon Chair Professor in physical science and electronics with the Electrical Engineering Department, University of California at Los Angeles, Los Angeles, CA, USA. His current research interests include nanoscale physics and materials, topological insulators, and spintronics and

devices.

Dr. Wang is a member of the American Physical Society. He has served as an Editor of Artech House and other publications.