Delivering on the Promise of Universal Memory for Spin-Transfer Torque RAM (STT-RAM)

[1]Anurag Nigam, [2]Clinton W. Smullen, IV, [2]Vidyabhushan Mohan, [3]Eugene Chen, [2]Sudhanva Gurumurthi, and [1]Mircea R. Stan

[1]ECE Dept. and [2]CS Dept., University of Virginia, 351 McCormick Road, Charlottesville, VA 22904 USA

[3]Grandis, 1123 Cadillac Court, Milpitas, CA 95035 USA

[1, 2]{an2z, cws3k, vm9u, gurumurthi, mircea}@virginia.edu, [3]Eugene.Chen@grandisinc.com

***Abstract*—Spin-Transfer Torque RAM (STT-RAM) has emerged as a potential candidate for Universal memory. However, there are two challenges to using STT-RAM in memory system design: (1) the intrinsic variation in the storage element, the Magnetic Tunnel Junction (MTJ), and (2) the high write energy. In this paper, we present a physically based thermal noise model for simulating the statistical variations of MTJs. We have implemented it in HSPICE and validated it against analytical results. We demonstrate its use in setting the write pulse width for a given write error rate. We then propose two write-energy reduction techniques. At the device level, we propose the use of a low-M_S ferromagnetic material that can reduce the write energy without sacrificing retention time. At the architecture level, we show that Invert Coding provides a 7% average reduction in the total write energy for the SPEC CPU2006 benchmark suite without any performance overhead.**

Keywords: STT-RAM, Magnetic tunnel junction, Thermal noise model and Invert coding

I. PROMISE OF UNIVERSAL MEMORY

Designing the memory hierarchy for microprocessors has grown significantly more challenging in the past decade. The conventional approach has been to use SRAM for caches, DRAM for main memory, and rotating disks (and now Flash memory) for storage. Each of these technologies has scalability limitations with regard to power, performance, or reliability. An alternative approach is to use a single "Universal Memory" that embodies the ideal properties of each layer: high performance, high density, high endurance, low power, and storage-class non-volatility. Spin-Transfer Torque RAM (STT-RAM), built using Magnetic Tunnel Junctions (MTJs), is a promising universal memory candidate. However, delivering on this promise requires addressing two key design challenges.

The first challenge is the stochastic nature of the MTJ, which is the storage element of an STT-RAM cell. The transient behavior of the MTJ is non-deterministic due to random thermal "kicks" acting on its magnetization during switching activity. It is important to model this phenomenon accurately in order to determine the appropriate write voltage and pulse width necessary to meet the required write error rate. We introduce a physical thermal noise model that simulates these statistical variations in SPICE. We then propose a methodology to determine the write voltage and pulse width necessary to achieve the desired write error rate.

The second challenge is that STT-RAM requires extremely high write energy as compare to SRAM to provide access times of 5 ns or less, as needed for on-die caches [1] [2] [3]. We show that the technique previously proposed by Smullen et al. reduces the write energy at the cost of reducing the MTJ's non-volatility, thereby increasing the error rate [4]. In order to address this issue, we propose the use of a low-M_S ferromagnetic material to reduce the write energy without compromising the non-volatility of the MTJ. We also evaluate the use of Invert Coding [5] to exploit the asymmetry in the energy required to write a logic "0" vs. a logic "1" into an STT-RAM cell. We show that this provides a 7% average reduction in the total write energy for the SPEC CPU 2006 benchmark suite without impacting performance.

This paper is organized as follows. Section II provides an overview of the design parameters for a STT-RAM bit-cell. Section III describes the thermal noise model and its validation with analytical results. Section IV evaluates the write error rate in a STT-RAM bit-cell as a function of the write pulse width and voltage. Section V presents the device and architecture level techniques for write energy reduction in STT-RAM, and Section VI concludes the paper.

II. DESIGN KNOBS IN STT-RAM

An STT-RAM cell consists of a Magnetic Tunnel Junction (MTJ) combined with an access transistor. The MTJ, which is the storage element, has an oxide layer sandwiched between two ferromagnetic (FM) layers, as shown in Figure 1(a). The

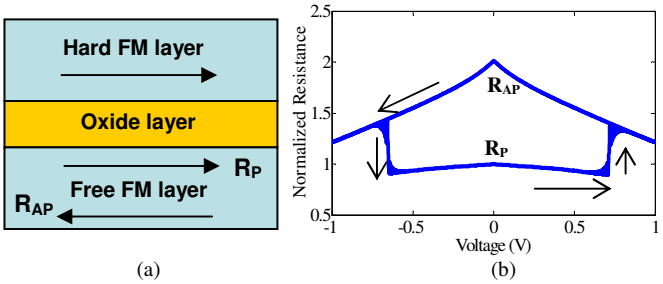

Figure 1: a) MTJ Structure b) R-V characteristic of MTJ

first FM layer has a fixed magnetic orientation and is called the Hard Layer, while the other has a variable magnetic orientation and is called the Free Layer. Data is stored in a MTJ as the relative orientation of the magnetization of the two FM layers. The MTJ exhibits low resistance (R_P) when the magnetic alignment of the FM layers is parallel (P). But, when the magnetizations are anti-parallel (AP), the MTJ exhibits a high resistance (R_{AP}). Figure 1(b) shows the resistance-voltage (R–V) characteristic of a MTJ. Writes are performed using a directional spin polarized current [6]. While writing a "1" (R_{AP}), a positive voltage is applied and the MTJ changes its state from P to AP. During write "0" (R_P), a negative voltage is applied across the MTJ to change its state from AP to P.

While STT-RAM has properties that make it a viable universal memory, it requires careful tuning to meet the requirements and constraints for each layer of the memory hierarchy. We now discuss how to tune STT-RAM to meet the requirements of the lowest and highest levels of the memory hierarchy.

1) Storage class memory

For storage class memory systems, data retention and endurance are the most important metrics [7]. The data retention time of an STT-RAM bit-cell depends on the thermal stability (Δ) of the MTJ and is given by Equation 1:

$$\tau = \tau_0 \exp(\Delta),\ \tau_0 = 1 \text{ ns and } \Delta = \frac{H_K M_S A_r t}{2 k_B T} \quad (1)$$

Where, k_B: Boltzmann constant
T: System temperature
H_K: Uni-axial anisotropy
M_S: Saturation magnetization
A_r: Area of the MTJ
t: Thickness of the free layer

The thermal stability of the MTJ should be greater than 75 to provide a storage-class retention time [8]. High thermal stability can be achieved by using materials with high H_K or M_S values, by increasing the area or the thickness of the free layer, or some combination thereof.

2) On-Die caches

For on-die caches, combining high performance with low dynamic energy is the primary objective. Various MTJ structures [9] and materials with high spin polarization [10] have been proposed for this purpose. However, as data retention time is less important for caches, retention time can be decreased for improved energy and performance characteristics [4]. Data retention time in the order of seconds will not add any overhead in the refresh policy of the caches [4]. Either the area or the thickness can be reduced to decrease the non-volatility of the MTJ, though doing so will increase the statistical variation in the MTJ, as we will discuss next.

III. VARIABILITY IN THE MTJ

The variability of CMOS circuits arises primarily due to parametric variation in the lithography and to random dopant fluctuations [11]. Additional variability in the MTJ comes from the precession of the free layer magnetization vector. This is produced by Langevin random field torque acting on the free layer magnetization, for which we next present a thermal noise model to simulate.

A. Thermal Noise Model

The transient behavior of an MTJ is governed by the torques acting on its free layer magnetization vector [12]. The dynamics of these torques can be simulated by solving the Landau-Lifshitz-Gilbert (LLG) differential equation. The switching speed of the free layer magnetization is largely dependent on the initial angle of magnetization (θ) with respect to the easy axis. In the presence of a thermal field, θ will be thermally distributed and the switching speed will vary from measurement to measurement. There will also be variation in the trajectory of the free layer magnetization vector. At a finite temperature, a Langevin random field torque term can be introduced to the LLG equation to simulate the effect of this thermal field [13]. The Langevin random field relates to the system temperature as follows:

$$H_{L,i} = \sqrt{\frac{2\alpha k_B T}{\gamma M_s V}} X_i(t),\ i = (x,y,z) \quad (2)$$

Where, γ: Gyromagnetic ratio
α: Damping constant
$X_i(t)$: Zero mean, unit variance Gaussian random noise. Each of the x, y, z components of the noise has its own uncorrelated X (t).

The torque on the free layer magnetization vector (n_m) due to the thermal field (H_L) will be given by: $\Gamma_L = n_m \times H_L$. Suppose n_m makes an angle θ with the easy axis, z, and an angle φ with the x-axis of easy plane, x-y. Using the x-y-z coordinate system, the thermal field torque can be simplified as follows:

$$\Gamma_L = (\sin\theta\cos\phi x + \sin\theta\sin\phi y + \cos\theta z) \times (H_{L,x}x + H_{L,y}y + H_{L,z}z)$$

Substituting this torque term into the LLG equation, the rate of change of the free layer magnetization vector due to the thermal field will be given by:

$$\frac{d\theta}{dt} = \alpha[H_{L,z}\sin\theta - H_{L,x}\cos\theta\cos\phi - H_{L,y}\cos\theta\sin\phi] + [H_{L,y}\cos\phi - H_{L,x}\sin\phi] \quad (3)$$

$$\frac{d\phi}{dt} = [\alpha[H_{L,x}\sin\phi - H_{L,y}\cos] + [H_{L,z}\sin\theta - H_{L,x}\cos\theta\cos\phi - H_{L,y}\cos\theta\sin\phi]] / \sin(\theta)$$

Equation 3 captures the effect of the thermal field on n_m, which we validate against analytical results in the next section.

B. Model Validation

The MTJ write speed varies according to the thermally-distributed initial magnetization angle (θ) for the free layer [14]. Using our thermal noise model, we calculate the initial angle distribution by solving the LLG differential equation under the excitation of thermal field torque and uni-axial anisotropy torque. A demonstration of the variation is shown in Figure 2, which shows the histogram for an example initial angle distribution. To validate our model, we compare the

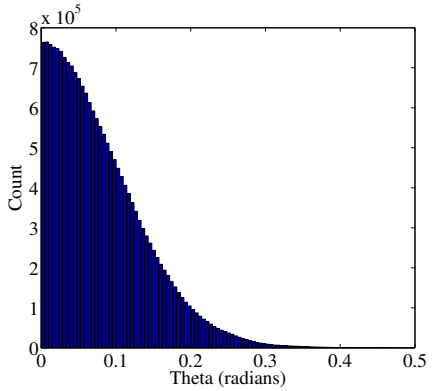

Figure 2: Histogram plot for Theta distribution. Parameters: H_K = 100 Oe, M_S = 1050 emu/cc, V=200x100x2 nm^3, T=300K and simulation time = 200 us

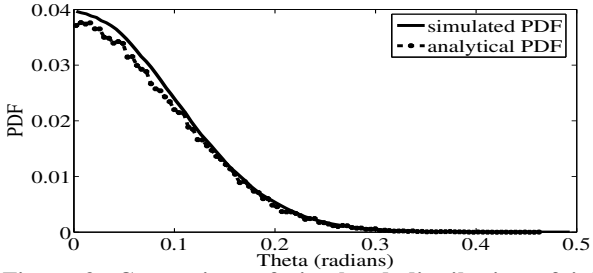

Figure 3: Comparison of simulated distribution of initial magnetization angle with analytical expression

simulated initial angle distribution with the distribution obtained using known analytical results. The thermally activated initial magnetization angle distribution can be estimated using Equation 4 [14].

$$P(\theta) = \frac{2H_K M_S}{k_B T} \exp(-\frac{H_K M_S}{k_B T} \sin^2 \theta) \quad (4)$$

Figure 3 shows the comparison of the simulated probability density function (PDF) of the initial magnetization angle with Equation 4. The simulated PDF of the initial angle is nearly identical to that obtained from Equation 4.

C. *HSPICE Implementation*

Analytical model for initial magnetization angle distribution doesn't capture the variation in magnetization trajectory due to the thermal noise. We have implemented our thermal noise model in HSPICE to simulate the statistical variation in MTJ performance due to the thermal noise. Equation 3 describes the free layer magnetization vector as a function of time under the influence of thermal field torque. In HSPICE, this differential equation is solved using the capacitor current equation: I = C dV/dt. The right-hand side of Equation 3 is implemented as a behavioral current source connected in series to a capacitor. Voltage at the capacitor node represents the θ and φ coordinates of free layer magnetization vector (n_m) and gives the solution of the differential equation. In HSPICE, noise is generated using the thermal noise of a resistor. The resistance value is set to match the variance of the thermal noise, as given in Equation 2. We use transient noise analysis for a set of Monte-Carlo simulations [15].

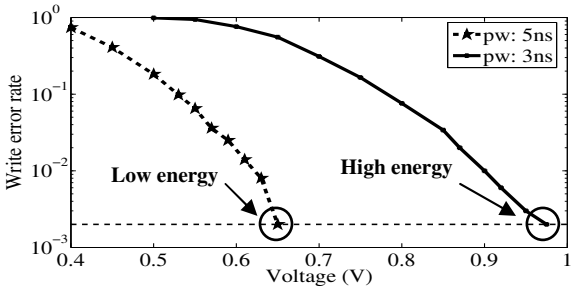

Figure 4: Write error rate as a function of write voltage and pulse width (pw) for AP→P write operation.

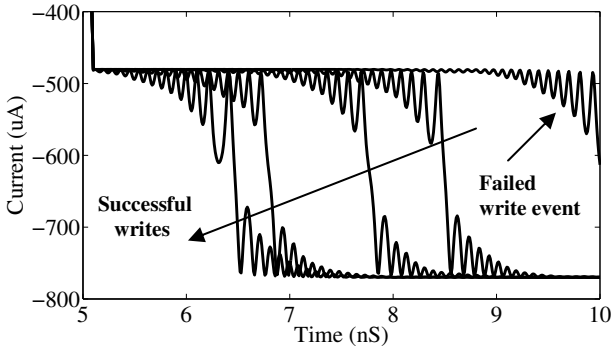

Figure 5: Snapshot of write events during AP→P write operation for V_{MTJ} = 0.65V and pulse width = 5ns.

IV. THE MTJ WRITE OPERATION

The write operation of an MTJ is probabilistic due to the thermal field torque acting on the free layer magnetization. To study the variation in the MTJ write operation, we have combined our thermal noise model with the MTJ model developed by Nigam et al. [16]. The MTJ model takes the write voltage and the write pulse width and outputs the write error rate.

A. *Write Error Rate vs. Pulse width*

The write speed of a MTJ is determined by the switching behavior of its free layer, which is acted upon by various torques (spin, thermal field, easy plane anisotropy, and uni-axial anisotropy) [12] [16]. In our circuit simulation, these torque terms are taken as inputs to the LLG equation and the free layer magnetization vector (n_m) is calculated as a function of time. The switching speed of the free layer magnetization is also dependent upon the initial angle of the free layer magnetization vector. To set the initial angle, the voltage across the MTJ is turned off for the initial 5 ns of simulation, after which the write voltage pulse is applied across the MTJ.

A thousand Monte-Carlo simulation runs are used to estimate the write error rate. The magnitude and width of the voltage pulse is varied to obtain the Write Error Rate (WER) as a function of the pulse width and write voltage, as shown in Figure 4 for the AP→P write operation. For a given pulse width, the WER decreases as the voltage is raised due to the increase in spin torque acting on the free layer magnetization. This result can be used to selecting different pairs of write voltages and pulse widths to achieve a desired WER.

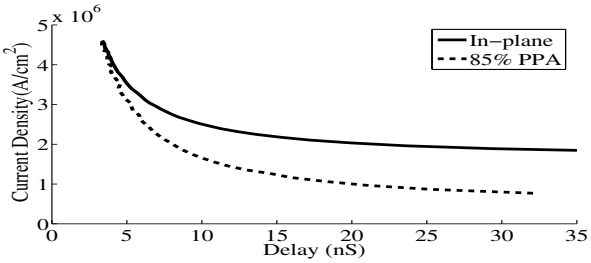

Figure 6: Comparison of current density vs. switching delay for in-plane and PPA. Parameters: M_S= 800 emu/cc, H_K =1050 Oe, α = 0.01 and t=1nm

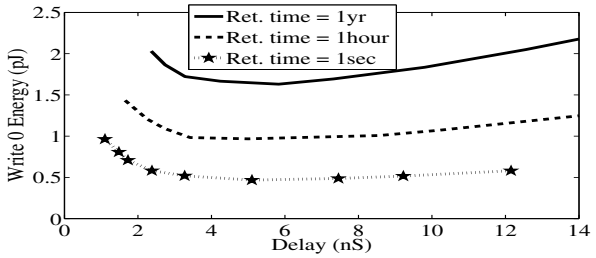

Figure 7: MTJ write "0" energy vs. delay comparison for different retention times. M_S=800 emu/cc, H_K=500 Oe, α=0.01

B. Case Study: Selecting a voltage and pulse width pair

We now examine a case where the WER specification for the MTJ is 2e-3. Figure 4 shows that this WER can be achieved for two voltage/pulse width pairs: 0.65V, 5ns and 0.975V, 3ns. Figure 5 shows a snapshot of write events for the first pair, where the write speed varies due to thermal field torque. As before, an initial 5 ns of time is used to randomize the initial angle of the free layer magnetization vector. The energy requirements for these two pairs are 2 pJ and 3.12 pJ, respectively. Depending on the speed and energy requirements of the application, we can choose either of these two pairs to obtain this write error rate, though the latency requirement for on-chip caches is typically less than 5 ns. We next explore different techniques for write energy reduction.

V. WRITE ENERGY OPTIMIZATION

Recent research in STT-RAM has focused on reducing the write energy by using device, circuit, and architecture level techniques [9] [10] [18]. At the device level, a significant reduction in write energy can be achieved by changing the MTJ structure (in-plane and Partial Perpendicular Anisotropy) or by decreasing the data retention time of the MTJ [9] [4]. In Section V.A, we first analyze the recently proposed Partial Perpendicular Anisotropic MTJ structure [9] using our MTJ circuit model. We then look at the proposal to reducing the data retention time to lower the write energy at the cost of increasing the write error rate [4]. To avoid this problem, we propose a Low-M_s ferromagnetic material . At the architecture level, we propose a novel application of the existing Invert Coding scheme to reduce the total write energy in STT-RAM without impacting performance.

A. Device level Optimization

High write energy in STT-RAM is due to the large intrinsic current density (J_{c0}) required for the write operation. For an in-plane MTJ, J_{c0} is given by following equation:

$$J_{c0} = \frac{2e\alpha M_s t(H_K + 2\pi M_s)}{\hbar\eta} \quad (5)$$ Where, e: electronic charge,

η: spin-transfer efficiency, and α: damping constant

At the device level, we have studied a range of techniques that can significantly reduce the value of J_{c0}.

1) Reducing J_{c0} using different MTJ structures

The large value of J_{c0} for in-plane MTJs is due to the extremely large out-of-plane demagnetization field, since the magnetization of the FM layers is oriented within the plane. One promising solution is to build a fully perpendicular MTJ (PMTJ) [19]. In a PMTJ, the magnetization direction is perpendicular to the plane of the FM layers, which cancels out the effect of the demagnetization field and reduces the value of J_{c0}. Though PMTJs reduce J_{c0}, they require high perpendicular anisotropic materials that require epitaxial growth at elevated temperatures and are harder to integrate with CMOS processes.

Chen et al. discovered a method to induce perpendicular anisotropy for an in-plane free layer, thereby cancelling most of the demagnetization field and thus reducing J_{c0} [9]. These in-plane ferromagnetic materials, called Partial Perpendicular Anisotropic materials (PPA), do not require in-situ heating during deposition and integrate more easily with CMOS processes. Figure 6 shows the comparison of an in-plane MTJ and a PPA MTJ with 85% partial perpendicular anisotropy. The free layer parameters (H_K, M_s, t and α) are the same for both structures. PPA provides a significant reduction in the required current density (J_c) at longer pulse widths. Thus PPA resolves the integration issues of PMTJ while reducing the required J_c for switching.

2) Reducing J_{c0} by Decreasing Data Retention Time

The required data retention time is different across the layers of the memory hierarchy. While storage systems have stringent data retention requirements, the caches of a processor do not. Prior work has demonstrated that the interval of time that data is resident in caches is on the order of microseconds [20]. Smullen et al. presented a methodology to reduce the STT-RAM write energy by trading retention time for improved performance and reduced write energy [4]. The area of the MTJ is used as a knob to reduce the retention time by decreasing the thermal stability, as given in Equation 1. By reducing the MTJ area, the write current decreases and the potential for significant write energy savings was demonstrated for the caches of a high-performance multicore processor.

As J_{c0} is independent of MTJ area (Equation 5), the MTJ's write time will remain unaffected by a reduction in its area [14]. To reduce both the MTJ write time and write energy, we propose to use the free layer thickness (t) as a design knob to reduce the retention time. We have assumed the material parameters (H_K, M_S, α) remain fixed. Figure 7 shows the write "0" (AP→P) energy of the MTJ as a function of delay for different data retention times. Decreasing the data retention time reduces the write energy.

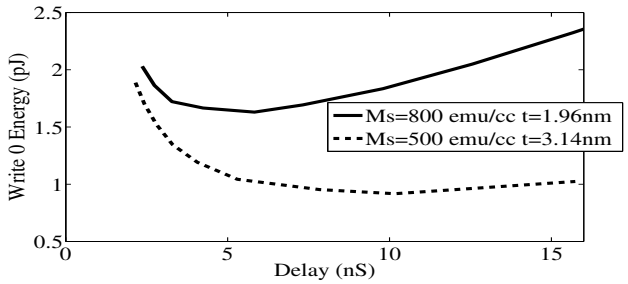

Figure 8: Comparison of write "0" energy for different M_S values. Retention time = 1year, H_K = 500 Oe, α = 0.01

The two previously proposed techniques (using PPA MTJ structure and reducing data retention time) reduce J_{c0} and the write energy significantly. However, reducing the data retention time will significantly increase the WER [9], which may require complex error correction to mitigate. We next propose a technique to reduce the write energy without compromising the thermal stability of the MTJ.

3) Reducing J_{c0} by using Low-Ms ferromagnetic material

We propose a write energy reduction technique by tuning the saturation magnetization (M_S). J_{c0} is proportional to the square of the saturation magnetization (M_S^2), so lowering M_S can reduce the write energy. However, this will also decrease the thermal stability of the MTJ, as Equation 1 shows. To maintain the same thermal stability, we increase the thickness of the free layer (t) proportionate to the reduction in M_S. As J_{c0} is proportional to tM_s^2, the current density (and thus write energy) will decrease while retaining the same thermal stability. Figure 8 compares the write "0" energy for two different M_S values but with same retention time of one year. The MTJ with the low-M_s free layer shows a significant reduction in write energy. However, the benefits diminish as the write latency is reduced, so this technique provides larger energy savings for the low-speed STT-RAM designs. We have reduced the value of M_S to 500 emu/cc to lower the write energy. Such a value of M_S has been demonstrated in previous work [17].

B. Architecture level Optimization

At the architecture level, different schemes have been proposed to lower the write energy by reducing the number of bits written [18] [21]. Zhou, et al. developed an early write termination technique which uses the initial phase of write cycle to compare the bits already stored in a word to the bits being written and only overwrites the flipped bits [18]. For short write pulse widths, the parallel read cannot be done correctly as some cells may switch early on. In this scenario, the scheme will limit the write performance as the read must occur before the write can be started.

Joo et al. developed a data inversion scheme to improve the write endurance for Phase Change Memory caches [21]. This scheme calculates the Hamming distance (HD) between the current word and the previously stored word. If the calculated HD is larger than half of the word size, the data word is inverted before being written. This could also be applied to STT-RAM caches, but it requires a serialized read before every write operation. In this work, we propose a novel application of Invert Coding for STT-RAM that elides the Hamming distance calculation to eliminate the performance overhead. We use Invert Coding to decrease the overall write energy by exploiting the asymmetry between the write "0" and write "1" energies for STT-RAM cells, which we will discuss next.

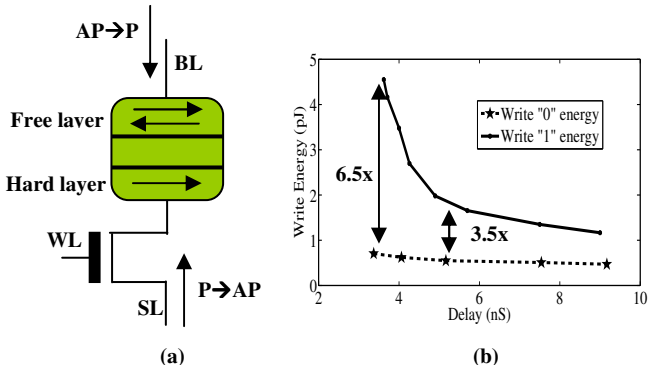

Figure 9: (a) STT-RAM bit-cell (b) Comparison of write "0" and "1" energy of bit-cell as a function of delay. MTJ Parameters: M_S=800 emu/cc, H_K=500 Oe, α = 0.01, Retention time = 1sec.

1) Asymmetry in write "0" and write "1" energy

At the MTJ level, P→AP switching is harder than AP→P switching [6]. Figure 9(a) shows the schematic of a STT-RAM bit-cell. During a write "0" (AP→P) operation, current flows from BL to SL and the MTJ is connected to the drain side of the NMOS. For a write "1" (P→AP) operation, current flows from SL to BL and the MTJ is connected to the source side of the NMOS. The MTJ acts a resistor connected to the source of the NMOS. This increases the voltage at the NMOS source above ground, resulting in lower NMOS current due to reduced V_{GS} and the body effect.

The combination of more difficult MTJ switching for write "1" and lower driving strength of the NMOS results in write "1" energy being larger than the write "0" energy in the STT-RAM bit-cell. Figure 9(b) shows the write "0" and write "1" energy of a bit-cell as a function of delay. For a 5ns pulse width, the write "1" energy is 3.5x the write "0" energy, but this difference increases to 6.5x for the 3.5ns pulse width. This demonstrates that the asymmetry in write energy increases as the write latency is improved.

2) An Invert Coding based Scheme

Due to the difference in the energy required for a write "0" vs. a write "1", Invert Coding can be applied to decrease the overall write energy. In this scheme, the number of ones in the word will be counted before the write is performed. If the number of ones is greater than half the word size, then inverting the bit pattern can decrease the total write energy. An invert bit is added to indicate this for later reads, and if the number of ones in the original word is not more than 50%, the invert bit will be set to zero. The modified word is then sent to the data bus to be written into the cache. As the asymmetry in the write "0" and write "1" energy increases with the decrease in delay, the Invert Coding scheme will provide large energy savings for STT-RAM caches designed for high performance processors.

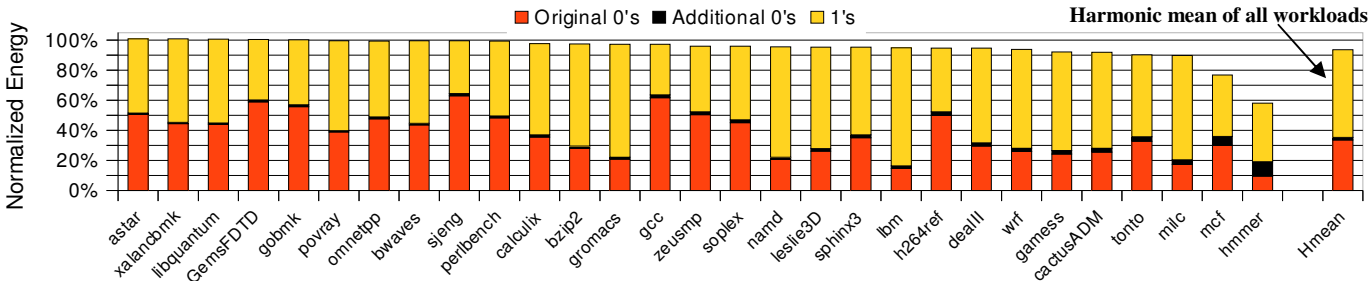

Figure 10: Write energy savings across different workloads of SPEC CPU 2006 benchmark suite using the Invert coding scheme.

a) Overheads in Invert Coding Scheme

An Invert Coding block must be added to the memory peripheral circuitry to count the number of ones and invert the word as needed. This will be done in parallel to address decoding, so there will be no performance overhead. There is an approximately a 1.5% area overhead as we use one invert bit per eight bytes of data.

b) Architectural Evaluation

To evaluate the benefits of the Invert Coding on real workloads, we calculated the data bit pattern written to memory for all 29 workloads of the SPEC CPU 2006 benchmark suit [22]. We have used one invert bit per each eight-byte data word written to memory. For a STT-RAM cache with a 4ns write pulse width, the write "0" energy of a bit-cell is 0.62pJ while the write "1" energy is 3.48pJ. We have combined this data to produce Figure 10, which shows the total write energy for each workload using Invert Coding normalized to the original total energy. For the first ten workloads, there is either a small increase (by ~0.5%) or a negligible decrease in the total energy, due to the additional energy required to write the invert bit. For the other 19 workloads, we see energy savings ranging from 2% to 42%. Overall, we observe an average energy savings of 7% due to Invert Coding scheme.

VI. CONCLUSION

In this work we address two major challenges of STT-RAM memory technology: modeling the intrinsic variation within the MTJ and mitigating the high write energy. We present a physics based thermal noise model to explore the variation in MTJ, from which we obtain the write voltage and pulse width for given specification of the write error rate, energy, and performance requirements. To address the write energy issue, we propose two write energy reduction techniques. Using a low-M_S ferromagnetic material reduces the write energy and eliminates the reduction in the data retention time of previous technique. This scheme provides large energy savings for low-speed designs. At the architecture level, we propose a novel application of Invert Coding to reduce the total write energy by 7%, on average, for the SPEC 2006 benchmark suite. This scheme provides large benefits for high-speed designs.

ACKNOWLEDGEMENT

We thank Kamaram Munira, Avik Ghosh, Jiwei Lu and Stu Wolf from the University of Virginia for their valuable inputs. This research has been supported by DARPA's STT-RAM program, NSF CAREER Award 0643925 and Google.

REFERENCES

[1] X. Dong, et al., "Circuit and Microarchitecture Evaluation of 3D Stacking Magnetic RAM (MRAM) as a Universal Memory Replacement," DAC, pp. 554-559, 2008.

[2] G. Sun, et al., "A Novel Architecture of the 3D Stacked MRAM L2 Cache for CMPs," HPCA, pp. 239-249, 2009.

[3] International Technology Roadmap for Semiconductor, 2009.

[4] C. W. Smullen, IV et al., "Relaxing Non-Volatility for Fast and Energy-Efficient STT-RAM Caches," HPCA, pp. 50-61, 2011.

[5] M. R. Stan, et al., "Bus-invert coding for low-power I/O," VLSI Systems, vol.3, no.1, pp.49-58, 1995.

[6] J. C. Slonczeski, "Currents, torques, and polarization factors in magnetic tunnel junctions," Phys. Rev. B71, 024411, 2005.

[7] R. F. Freitas and W. W. Wilcke, "Storage-class memory: The next storage system technology," IBM J. Res. Dev., vol.52, no.4.5, pp.439–447, 2008.

[8] A. Driskill-Smith, et al., "Non-volatile spin-transfer torque RAM (STT-RAM): Data, analysis and design requirements for thermal stability", VLSI Technology, pp.51-52, 2010.

[9] E. Chen, et al., "Advances and Future Prospects of STT-RAM", IEEE Transactions on Magnetics, vol.46, no.6, pp.1873-1878, 2010.

[10] Y. Huai, "Spin-Transfer Torque MRAM (STT-MRAM): Challenges and Prospects," AAPPS Bulletin, vol.18, no.6, pp.33-40, 2008.

[11] L.-T. Pang and B. Nikolic, "Measurements and analysis of process variability in 90 nm CMOS," IEEE J. Solid-State Circuits, vol.44, no.5, pp.1655–1663, 2009.

[12] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," Phys. Rev. B, vol.62, pp.570-578, 2000.

[13] J. Z. Sun, "Spin angular momentum transfer in current-perpendicular nanomagnetic junctions," IBM J. Res. Dev., vol.50, no.1, pp.81-100, 2006.

[14] A. Raychowdhury, et al., "Design Space and Scalability Exploration of 1T-1STT MTJ Memory Arrays in the Presence of Variability and Disturbances," IEDM, pp.1-4, 2009.

[15] HSPICE-RF user guide.

[16] A. Nigam, et al., "Self Consistent Parameterized Physical MTJ Compact Model for STT-RAM," CAS, vol.02, pp.423-426, 2010.

[17] Y. Kitamoto, et al., "Co ferrite films with excellent perpendicular magnetic anisotropy and high coercivity deposited at low temperature," Journal of Applied physics, vol.85, no.8, pp.4708-4710, 1999.

[18] P. Zhou, et al., "Energy Reduction for STT-RAM using Early Write Termination", ICCAD, pp.264-268, 2009.

[19] U. K. Klostermann, et al., "A Perpendicular Spin Torque Switching based MRAM for the 28 nm Technology Node," IEDM, pp.187-190, 2007.

[20] X. Liang, et al., "Process Variation Tolerant 3T1D-Base Cache Architectures," MICRO, pp.15-26, 2007.

[21] Y. Joo, et al., "Energy- and Endurance-Aware Design of Phase Change Memory Caches," DATE, pp.136-141, 2010.

[22] SPEC – http://www.spec.org/cpu2006