Scalability and Design-Space Analysis of a 1T-1MTJ Memory Cell

Richard Dorrance, Fengbo Ren, Yuta Toriyama, Amr Amin, C.-K. Ken Yang, Dejan Marković
Department of Electrical Engineering, University of California, Los Angeles, CA, USA

***Abstract*—This paper introduces a design-space feasibility region as a function of MTJ characteristics and memory target specifications. The sensitivity of the design space is analyzed for scaling of both MTJ and underlying transistor technology. Design points for improved yield, density, and memory performance can be extracted for 90nm down to 32nm processes based on measured MTJ devices. To achieve flash-like densities in upcoming 22nm and 16nm technology nodes, scaling of the critical switching current density is required.**

***Index Terms*—STT-RAM, Design Space, Magnetic Tunnel Junction, Variability**

I. Introduction

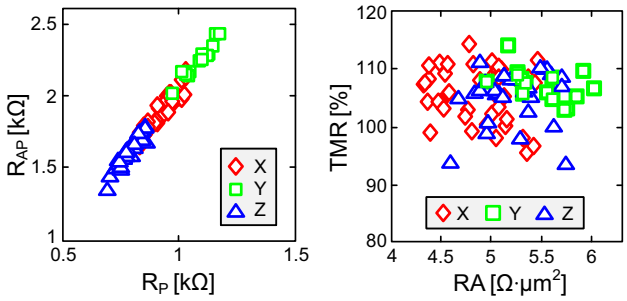
Spin-Torque Transfer RAM (STT-RAM) is a candidate for next-generation memory, with Magnetic Tunnel Junctions (MTJs) as the non-volatile storage element. However, questions remain unanswered for such a memory to succeed. The design of the MTJ depends considerably on the underlying transistor technology since a given CMOS technology constrains the design space due to the overhead and impact of the access transistor in each memory cell. Previous work, such as [1], has not addressed this dependency or provided the necessary framework with measured CMOS and MTJ characteristics. The feasibility and yield of the memory depend on the design space and the variation of the MTJs [2]. This paper presents such a design space as a function of the variation for technologies down to 32nm to show the scalability of STT-RAM.

II. Modeling MTJ Variability and Scaling

This section describes the MTJ model and characteristics that is used in the subsequent sections to explore the design space for several scaled CMOS technologies.

A. MTJ Device Variability

While statistical variation of CMOS is generally well understood, similar characteristics for MTJs have not been well documented. This paper uses a combination of fundamental equations and measured device characteristics to model the statistical behavior of MTJs. Fig. 1(a) contains a plot of measured R_P vs. R_{AP} for 105 MTJ nanopillars of varying size and target resistance-area (RA) products. Variations in resistance and tunnel magnetoresistance (TMR) are due to a combination of lithographic variations in the physical dimensions of the nanopillar and minute variations in the thicknesses of the up to 20 different layers in state-of-the-art MTJ processes [3]. The cumulative effects of these variations on RA and TMR can be calculated [4], as shown in Fig. 1(b) and Table I.



(a) R_{AP} vs. R_P (measured). **(b)** TMR vs. RA (calculated).

Fig. 1: Measured (a) R_{AP} vs. R_P and (b) TMR vs. RA for MTJ nanopillars measuring $150 \times 45nm^2$ (X), $130 \times 50nm^2$ (Y), and $170 \times 45nm^2$ (Z) at room temperature (300K).

TABLE I: Measured Device Statistics

	X	Y	Z
TMR [%]	105.7	107.3	105.3
σ_{TMR} [%]	4.7	2.7	4.6
RA [$\Omega \cdot \mu m^2$]	4.88	5.51	5.22
σ_{RA} [$\Omega \cdot \mu m^2$]	0.342	0.297	0.311

B. Scaling of MTJ Current and Resistance

The resistance and switching current can be modeled using a precessional-based switching model, modified to include thermally-activated switching [5]. The switching current of an MTJ, for a constant pulse of duration τ, is given by:

$$I_C = I_{C0}\left[1 - \frac{\ln(\tau/\tau_0)}{\Delta}\right], \quad (1)$$

where Δ is the thermal stability of the MTJ, τ_0 is the natural time constant, and I_{C0} is the critical switching current. This critical switching current [6] is given by:

$$I_{C0} = \frac{\alpha 4\pi e}{\eta\hbar} M_S^2 V, \quad (2)$$

where α is the Gilbert damping constant, η is the factor of spin polarization, $\hbar$ is the reduced Planck's constant, e is the elemental charge of an electron, M_S is the magnetization saturation of the free layer, and V is the volume of the free layer.

For an MTJ with free layer dimensions $l > w >> d$ [7], as shown in Fig. 2, the thermal stability of an MTJ is

978-1-4577-0994-4/11/$26.00 ©2011 IEEE

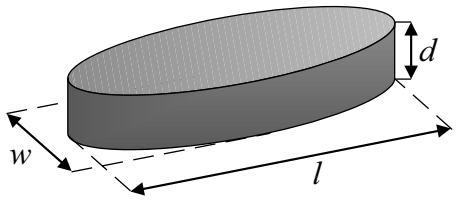
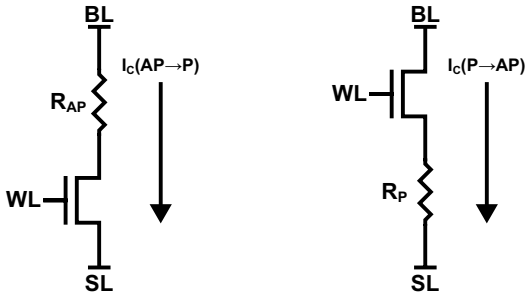


Fig. 2: MTJ free layer dimensions.



(a) $I_C(AP \rightarrow P)$. **(b)** $I_C(P \rightarrow AP)$.

Fig. 3: 1T-1MTJ cell architecture showing MTJ switching current for (a) AP to P and (b) P to AP.

approximately:

$$\Delta = \frac{E}{k_B T} = \frac{H_K M_S}{2 k_B T} V \approx d \left(\frac{1}{w} - \frac{1}{l} \right) \frac{M_S^2}{k_B T} V, \qquad (3)$$

where k_B is Boltzmann's constant, T is the absolute temperature in Kelvin, H_K is the out-of-plane uniaxial anisotropy, and E is the energy of anisotropy [8], [9].

In this paper, dimensional scaling is performed to maintain a constant Δ in order to ensure the long-term non-volatility of the STT-RAM. Therefore, dimensions l and w of the MTJ are scaled by a factor λ to manipulate I_{C0} and $R_{P/AP}$, then to keep Δ constant, d must scale by $\lambda^{-1/2}$. This results in $I_{C0} \propto lwd \rightarrow \lambda^{3/2}$ and $R_{P/AP} \propto l^{-1} w^{-1} \rightarrow \lambda^{-2}$.

III. Design Space

The analysis in this paper uses a conventional 1T-1MTJ cell architecture as shown in Fig. 3. The writing currents for flipping the cell resistance are defined as $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. The design space of a single STT-RAM memory cell can be illustrated using an R_{AP} vs. R_P plot as is shown in Fig. 4. The feasibility region is indicated by the shaded region. It contains all points (R_P, R_{AP}) in the design space so that a memory cell made with such an MTJ is functional. In the design space, the two lower bounds are set by the read margin of the cell, while the two upper bounds are set by the write margin of the cell.

The lower bound $R_{P,MIN}$ is dependent on the implementation of the sense amplifier, and represents the minimum resistance required for reliable circuit operation. Additionally, $R_{AP,MIN}$ is determined by TMR_{MIN} (Fig. 5), the minimum TMR required for the read amplifier to differentiate between R_P and R_{AP}. Regardless of the specifics of the implementation, all sense amplifiers are either a voltage- or current-

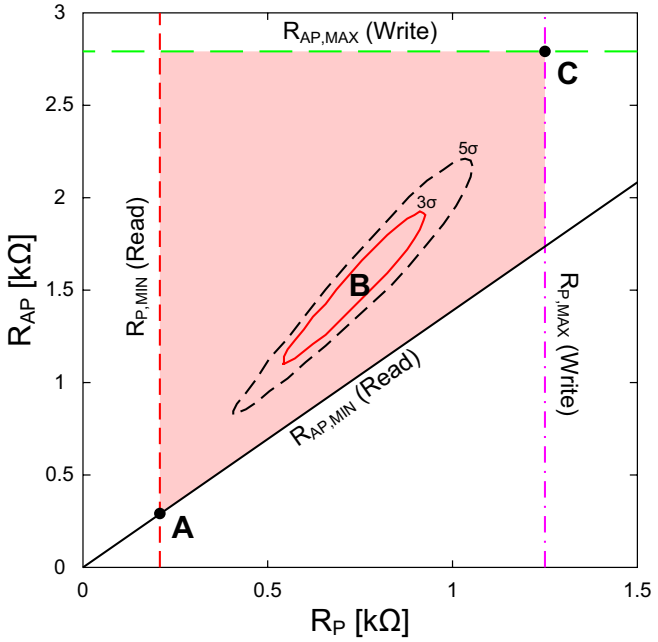


Fig. 4: Design space, in a 65nm process, for $W_N = 2.0\mu m$, $I_C(P \rightarrow AP) = 500\mu A$, $I_C(AP \rightarrow P) = 375\mu A$, with an overlay of device X from Table I.

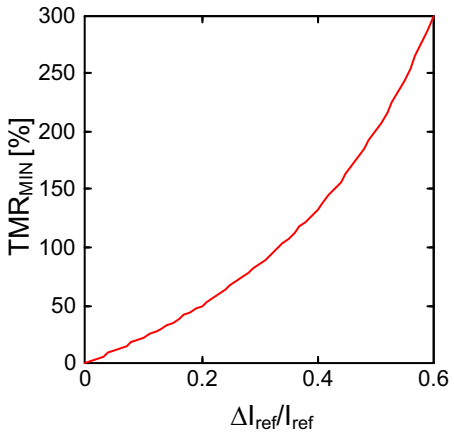
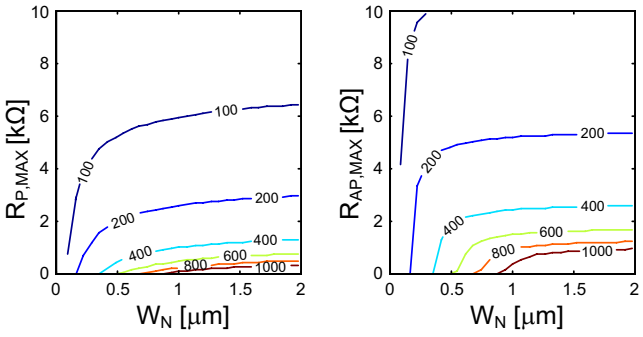


Fig. 5: Design space lower bound TMR_{MIN} vs. $\Delta I_{ref}/I_{ref}$ for a current-sensing read circuit with ideal reference resistance $2(R_P \parallel R_{AP})$.

sensing topology. For a generic current-sensing read circuit, TMR_{MIN} can be expressed as:

$$TMR_{MIN} = \frac{2\Delta I_{ref}/I_{ref}}{1 - \Delta I_{ref}/I_{ref}}. \qquad (4)$$

For I_{ref} flowing through the reference resistance R_{ref}, $I_{ref} + \Delta I_{ref,1}$ flows through R_P and $I_{ref} - \Delta I_{ref,2}$ through R_{AP}. When $\Delta I_{ref,1} = \Delta I_{ref,2} = \Delta I_{ref}$, TMR_{MIN} is minimized. Under this condition, $R_{ref} = 2(R_P \parallel R_{AP})$ and we can express TMR_{MIN} as a function of the normalized fractional sensing current ($\Delta I_{ref}/I_{ref}$). In Eq. 4, ΔI_{ref} must be chosen so that the read amplifier correctly evaluates across



(a) $R_{P,MAX}$ vs. W_N. **(b)** $R_{AP,MAX}$ vs. W_N.

Fig. 6: (a) $R_{P,MAX}$ and (b) $R_{AP,MAX}$ at nominal V_{DD} for a 65nm process (I_C contours are in μA).

all transistor PVT variations. Similarly, TMR_{MIN} can be derived for voltage-sensing topologies. It should be noted that the lower bounds $R_{P,MIN}$ and $R_{AP,MIN}$, while critical to the readability of the cell, are almost completely independent of the MTJs used. The only requirement is that sensing time and current (I_{ref}) be small enough so as not to disturb the cell during the read operation.

The upper bounds, $R_{P,MAX}$ and $R_{AP,MAX}$, are the maximum allowable resistances such that the access transistor, in a 1T-1MTJ configuration, is still able to provide the minimum critical writing currents $I_C(P \rightarrow AP)$ and $I_C(AP \rightarrow P)$. These upper bounds are subsequently very sensitive to the specific characteristics of the MTJ device used. Transistor-level simulations are used to determine the relationship between R_{MAX}, I_C, and cell size (transistor width W_N) for a technology. Fig. 6 shows an example of such a simulation in a 65nm process. Using the conventional configuration from Fig. 3, W_N is swept along with R_{MAX}. The contours of the simulated current are shown.

Fig. 4 shows a specific MTJ cell and its associated statistical variation (the concentric ovals around point B) overlaid on the design space. The design-space margin can be defined as the number of σ's of MTJ variation before crossing any of the previously defined bounds. Defining design-space margin (DSM) in terms of σ simplifies feasibility characterization to a single variable and thus allows yield to be quickly calculated. To a first order, 3σ, 4σ, 5σ, and 6σ of design margin roughly correspond to being able to reliably produce 1kbit, 32kbit, 4Mbit, and 1Gbit memory arrays.

Fig. 7 highlights the effects of CMOS variability on the design space bounds. To illustrate the effects more clearly, a $35F^2$ cell in a 65nm process is chosen. As expected, the more stringent constraints of the SS corner causes the design space to shrink. This shift is caused by an increase in the threshold voltage of the access transisor. Enviormental variables, like temperature, also have a significant effect in the design space. A consumer grade STT-RAM is expected to operate over a range of more than 100°C in which TMR can drop by more than 30% [10], degrading the DSM for readability. These sources of technological and environmental variability must also be considered in the design process.

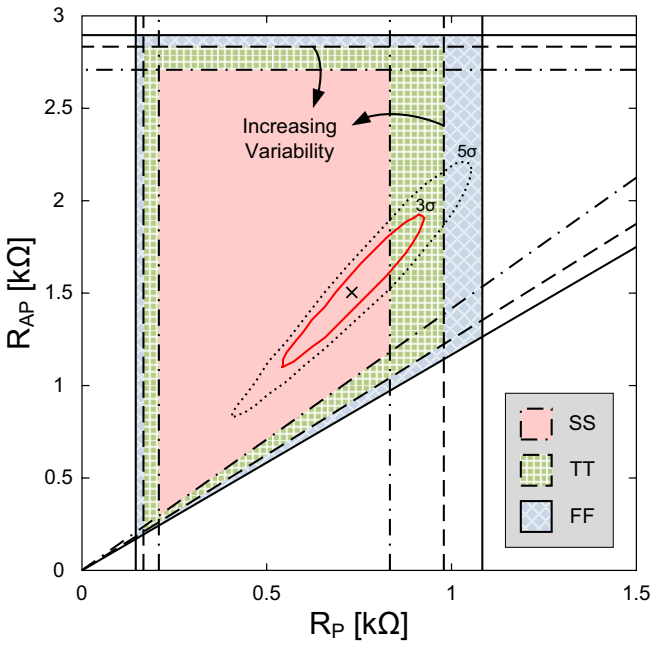


Fig. 7: Design space, in a 65nm process, for $W_N = 750nm$, $I_C(P \rightarrow AP) = 300\mu A$, $I_C(AP \rightarrow P) = 300\mu A$, with an overlay of device X from Table I, for SS, TT, and FF corners.

IV. Sensitivity Analysis and Design Example

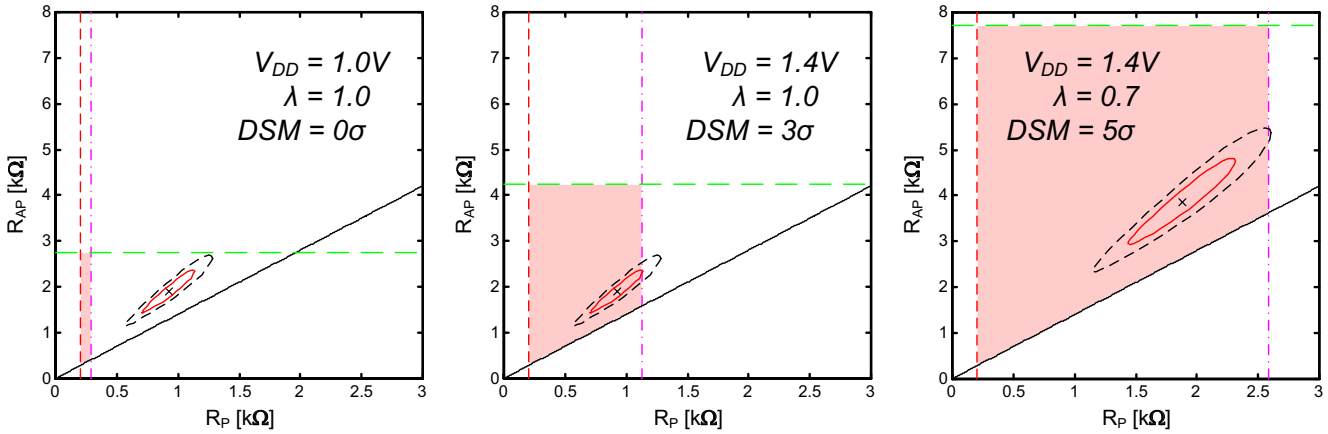
Many variables, at both the circuit and device levels, affect the design space. In order to optimize a variable for a target memory specification, we must determine how such a variable impacts the design space. This section introduces a design-space sensitivity (DSS) as a metric to quantify the behavior of the change in design space as a function of various design parameters (V_{DD}, λ, J_C, RA, TMR, W_N, etc.).

A. Design-Space Sensitivity Analysis

First consider the points A, B, and C in Fig. 4. Points A and C correspond to the corner values of R_P and R_{AP} in the feasible design space. Point B represents the nominal MTJ at the center of the MTJ device distribution. For a positive design margin to exist, point B must fall somewhere between points A and C.

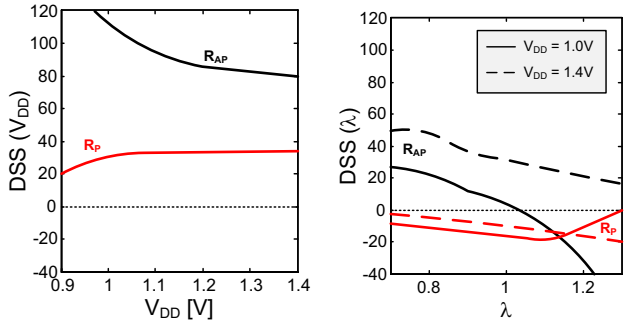
A "better" design space can be achieved from altering a design parameter, when a larger distribution of the MTJs (the number of σ) falls within the feasible region. Note that the improved design space is not simply increasing the area of the feasibility region, since the motion of point B must be considered as well. Recall that point A depends only slightly on the MTJ parameters. Therefore, the improvement (or deterioration) of the design space depends mostly on the change in DSM between points B and C as a function of a particular design variable.

Therefore, we define the design-space sensitivity to the



(a) $V_{DD} = 1.0V$ and $\lambda = 1.0$. 0σ design margin. **(b)** $V_{DD} = 1.4V$ and $\lambda = 1.0$. 3σ design margin. **(c)** $V_{DD} = 1.4V$ and $\lambda = 0.7$. 5σ design margin.

Fig. 8: Design space, in a 65nm process, for a $30F^2$ cell ($W_N = 0.65\mu m$) for device X from Table I: $I_C(P \rightarrow AP) = 450\mu A$, $I_C(AP \rightarrow P) = 300\mu A$. Inner red oval represents 3σ of MTJ device variation. Dashed, black oval corresponds to 5σ of MTJ variation.



(a) Sensitivity to V_{DD} vs. V_{DD}. **(b)** Sensitivity to λ vs. λ.

Fig. 9: Design-space sensitivity of parameters (a) V_{DD} and (b) λ in a 65nm technology.

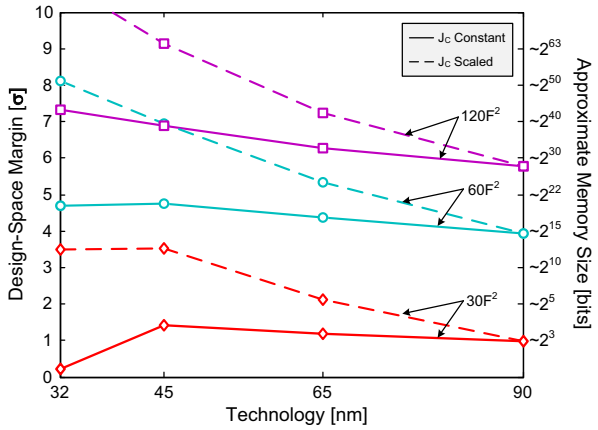


Fig. 10: Design margin vs. technology node. For constant $J_c(P \rightarrow AP) = 6 \times 10^6\ A/cm^2$ and $J_c(P \rightarrow AP)$ scaling by 4.5% each technology generation.

parameter X as:

$$DSS(X) = \frac{\partial(\frac{R_C - R_B}{\sigma})_{P/AP}}{\partial X}, \quad (5)$$

where R_B and R_C are taken as either R_P or R_{AP} at points B and C, thus defining the DSS along each dimension of the design space. $\frac{R_C - R_B}{\sigma}$ is the normalized distance between points B and C in the design space along the $R_{P/AP}$ dimension. Intuitively, the $DSS(X)$ describes the instantaneous rate of change in DSM to a particular design parameter X. The derivative loses positional information, and so we used the DSS in conjunction with the original plot of the design space to determine the benefit of tuning the design parameter X. For both the R_P and R_{AP} dimensions, if $DSS(X) > 0$, then the DSM is improved by increasing X, and if $DSS(X) < 0$, then DSM is improved by decreasing X. When the design-space sensitivities for the two dimensions conflict, the size of the design space in each dimension should then be taken into account.

B. Design Example

In this section we use the sensitivity analysis to design a 4Mbit STT-MRAM with a $30F^2$ cell size (comparable to eDRAM) in a 65nm technology. Device X from Table I with $I_C(P \rightarrow AP) = 450\mu A$ and $I_C(AP \rightarrow P) = 300\mu A$ is the nominal MTJ and can be scaled by λ. Also, approximately 5σ of design margin is required for reasonable yield.

Fig. 8(a) shows the design space for a nominal $V_{DD} = 1.0V$ and $\lambda = 1.0$. The inner red oval is the 3σ variation of the MTJ, while the dashed, black oval represents the 5σ variation of the MTJ. Clearly, with nominal V_{DD} and λ, the memory is not functional. Fig. 9 shows that the design space is much more sensitive to V_{DD} than it is to λ. Therefore, we choose to scale V_{DD} to $1.4V$. Fig. 8(b) shows the new design space, with the 3σ bound at the edge of the design boundary.

Scaling V_{DD} alone proves insufficient to meet the 5σ design margin required, and so we simultaneously scale λ. Fig. 9(b) shows that scaling λ results in conflicting DSS. The R_{AP} margin improves more by scaling λ up, while the R_P margin

TABLE II: $J_C(P \rightarrow AP)$ for an RA of 5 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	2.72	2.62	2.65	2.20	2.04	1.86
eDRAM ($30F^2$)	4.60	4.58	4.57	4.31	4.23	4.12
SRAM ($120F^2$)	6.67	6.76	6.86	7.07	7.17	7.30

‡Predicted

TABLE III: $J_C(P \rightarrow AP)$ for an RA of 10 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	2.22	2.14	2.19	1.82	1.69	1.56
eDRAM ($30F^2$)	3.27	3.28	3.33	3.20	3.18	3.16
SRAM ($120F^2$)	4.27	4.33	4.44	4.59	4.68	4.78

‡Predicted

TABLE IV: $J_C(P \rightarrow AP)$ for an RA of 15 $\Omega \cdot \mu m^2$

Equivalent Cell Size	J_C vs. Technology Node (10^6 A/cm^2)					
	90nm	65nm	45nm	32nm	22nm‡	16nm‡
FLASH ($6F^2$)	1.83	1.81	1.82	1.67	1.62	1.56
eDRAM ($30F^2$)	2.56	2.59	2.63	2.58	2.59	2.61
SRAM ($120F^2$)	3.17	3.21	3.33	3.47	3.55	3.65

‡Predicted

improves by scaling λ down. However, Fig. 8(b) indicates that R_{AP} has considerable margin and we can trade off some of that margin for improved margin in R_P. Therefore, we choose to scale λ down to 0.7. As we can see in Fig. 8(c), the desired 5σ bound on MTJ variation is essentially enclosed within the design space.

V. Future Scalability

Scalability is an important feature for the success of a memory technology. Fig. 10 shows how scaling of the transistor technology impacts the design margin when using an MTJ that has a current density (J_C) of 6×10^6 A/cm^2 for 10ns P to AP switching and RA of $5\Omega \cdot \mu m^2$. The figure shows that SRAM equivalent sizes ($120F^2$) scale well, with a design margin more than sufficient to construct gigabit memories ($> 6\sigma$). However, as we decrease cell size, the design margin begins to degrade below 45nm. The design margin practically disappears once we reach an eDRAM-equivalent cell size ($30F^2$). However, if the MTJ current density scales with technology, also shown in Fig. 10, then this trend is reversed. By scaling J_C by as little as 4.5%, a constant design margin can be achieved between each technology node below 45nm.

Tables II, III, and IV contain critical switching current densities for flash-, eDRAM-, and SRAM-equivalent cell sizes for RAs of 5, 10, and $15\Omega \cdot \mu m^2$, respectivly. All table values correspond to 5σ of design margin and sub-10ns switching times. It should be noted that while larger RAs require smaller current densities (in order to meet voltage headroom constraints), they scale much better between successive technology nodes. We can also see that flash-like cell sizes ($6F^2$) requires current densities below 3×10^6 A/cm^2 in current technologies and less than 2×10^6 A/cm^2 in upcoming 22nm and 16nm technology nodes. Aggressive scaling of MTJ currents will be required to achieve flash-like densities in future technologies.

VI. Conclusion

In this paper we have shown that the joint optimization of multiple design parameters is essential in the design of a STT-RAM memory array. We have derived the necessary framework to allow for such a systematic design procedure. Additionally, the analytical methodology presented in this work has been utilized to show that mild scaling of MTJ J_C is required to enable flash-like memory densities in upcoming CMOS technologies. Such densities, coupled with low write energies and its non-volatility, make STT-RAM a possible contender for next-generation memories.

Acknowledgments

This work was supported by the DARPA STT-RAM program. The authors would also like to thank Pedram Khalili, Henry Park, and Henry Chen of UCLA and Ilya Krivorotov of UCI for their contributions.

References

[1] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, "Design Space and Scalability Exploration of 1T-1STT MTJ Memory Arrays in the Presence of Variability and Disturbances," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, December 2009, pp. 1–4.

[2] M. Motoyoshi *et al.*, "A study for 0.18 μm high-density MRAM," in *Symposium on VLSI Technology, 2004. Digest of Technical Papers.*, June 2004, pp. 22–23.

[3] S. R. Min *et al.*, "Etch Characteristics of Magnetic Tunnel Junction Stack with Nanometer-Sized Patterns for Magnetic Random Access Memory," *Thin Solid Films, Proceedings of the International Symposium on Dry Process, 2006. (DPS 2006).*, vol. 516, no. 11, pp. 3507–3511, November 2008.

[4] R. Beach *et al.*, "A Statistical Study of Magnetic Tunnel Junctions for High-Density Spin Torque Transfer-MRAM (STT-MRAM)," in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, December 2008, pp. 1–4.

[5] T. Moriyama *et al.*, "Tunnel Magnetoresistance and Spin Torque Switching in MgO-based Magnetic Tunnel Junctions with a Co/Ni Multilayer Electrode," *Applied Physics Letters*, vol. 97, no. 7, p. 072513, 2010.

[6] J. Z. Sun, "Spin-Current Interaction with a Monodomain Magnetic Body: A Model Study," *Phys. Rev. B*, vol. 62, no. 1, pp. 570–578, July 2000.

[7] H. Chang and J. Burns, "Demagnetizing and Stray Fields of Elliptical Films," *Journal of Applied Physics*, vol. 37, no. 8, pp. 3240–3245, July 1966.

[8] V. Korenivski and R. Leuschner, "Thermally Activated Switching in Nanoscale Magnetic Tunnel Junctions," *IEEE Transactions on Magnetics*, vol. 46, no. 6, pp. 2101–2103, June 2010.

[9] J. Sun and D. Ralph, "Magnetoresistance and Spin-Transfer Torque in Magnetic Tunnel Junctions," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1227–1237, 2008.

[10] K. Lee and S. Kang, "Design Consideration of Magnetic Tunnel Junctions for Reliable High-Temperature Operation of STT-MRAM," *Magnetics, IEEE Transactions on*, vol. 46, no. 6, pp. 1537–1540, June 2010.