Analysis of STT-RAM Cell Design with Multiple MTJs Per Access

Henry Park, Richard Dorrance, Amr Amin, Fengbo Ren, Dejan Marković, and C.K. Ken Yang
Department of Electrical Engineering
University of California, Los Angeles
Los Angeles, CA 90095
henrypark@ucla.edu

Abstract—Density of STT-RAMs is limited by the area cost and width of the access device in a cell since it needs to support the programming currents. This paper explores a cell structure that shares each cell’s access transistor with multiple MTJ memory elements. Feasibility and limitations of such a cell structure is explored for both reading and writing of the memory. The analytical and simulation results indicate that only small amount of sharing is possible and having MTJs that can handle a high read current without disturbing the cell is needed.

Keywords-component; STT-RAM, Magnetic Tunnel Junction, 1T-3MTJ

I. INTRODUCTION

Aggressive integration of on-chip memory arrays raises the need for high-density memories. Advances in spin-torque transfer RAM (STT-RAM) enable storing data in a magnetic tunneling junction (MTJ) comprising a thin insulator and ferromagnetic sheets. The technology has shown promise as dense arrays of MTJs have been fabricated with very small cell sizes [1, 2]. However, as shown in Fig. 1, due to the required junction current density for switching, the access transistor must be sized to support such currents. For technology nodes down to 32nm, the actual cell size is determined by the access transistor. This paper analyzes the effectiveness and challenges in building an STT-RAM memory cell that shares multiple MTJs in a cell with a single access transistor.

Fig. 1. Required normalized NFET cell area at different technology nodes (RA = 4.9 μm², TMR = 120%, MTJ size = 70×120nm²)

III. DESIGN CONSIDERATIONS

Similar to most memories, an STT-RAM typically employs a structure with one MTJ switched/accessed by a single access transistor (1T-1MTJ). The organization of a conventional STT-RAM is shown in Fig. 2(a). The access transistor is sized [3] to meet the minimum write current requirement for the cell to switch either in the parallel to anti-parallel, min(Iw(AP→P)), or vice versa, min(Iw(P→AP)). Directions of each switching current are depicted in Figure 3. Switching from AP to P is equivalently moving operating point from A to B in Fig. 3(a) while switching in reverse direction (P to AP) is from C to D in Fig. 3(b). In both cases current density can be improved by boosting voltages above the nominal supply level [4]. Since each cell has a dedicated access device, a cell’s value is not perturbed by the writing of other cells. As long as the maximum read current, max(Ir(AP)) or max(Ir(P)), provides sufficient margin from the minimum write currents, the cell remains stable during a read operation (we assume minimum write current at 99.99% switching probability while maximum read current is at 0.01%). Such reliable read operation is traded with scalability of the memory cell due to minimum NFET size (Fig. 1). This paper explores possibility of sharing multiple MTJs with a single access device to reduce its impact on the cell size.

Sponsored by the DARPA/MTO STT-RAM program
cell area smaller than 6F² but they also revealed limitations on the array size and MTJ characteristics [5~7].

A. Impact of MTJ Sharing on Reads

Fig. 4 can be used to illustrate when a single MTJ in a cell is being read. The total amount of read current is the sum of the main MTJ current (I_MTJ through MTJ1 in the Fig. 4) and the parasitic current through all other MTJs attached to its bit-line (BL1). If the number of MTJ devices per cell is N_MTJ and the number of cells per column is NC, then the equivalent parasitic resistance between BL1 and VAC1 are expressed as [7]

\[
R_{PAR} = \frac{N_{MTJ} + N_C - 1}{N_{MTJ} - 1} \cdot R_{MTJ} = A_{PAR} \cdot R_{MTJ} \quad (1)
\]

R_{MTJ} is a discrete random variable bounded by R_P and R_AP. In comparison to a 1T-1MTJ cell, the parasitic resistive paths reduce the resistance difference when reading either R_P or R_AP of MTJ1. The parasitic paths effectively reduce the TMR seen by the read sense amplifier and an effective TMR can be expressed as

\[
\text{TMR}_{eff} = \frac{R_{AP} \parallel A_{PAR} \parallel R_P - R_P \parallel A_{PAR} \cdot R_{AP}}{R_P \parallel A_{PAR} \cdot R_{AP}} = \left( A_{PAR} - 1 \right) \cdot \text{TMR} \quad (2)
\]

Fig. 5(a) illustrates the impact of the parasitic resistance on current load lines from an MTJ resistance loading the nonlinear switching characteristics of an NFET. R_P and R_AP result in different load lines. Since R_AP has a higher resistance, the operating point during a read (point A) shows a lower current and lower voltage across the access device (V_{AC}). A region of uncertainty can be overlaid onto the load lines. When the regions overlap due to R_PAR, incorrect sensing can occur. The overlap condition occurs when min(I_P(P)) < max(I_P(AP)) which corresponds to A_PAR < 1. Fig. 5(b) shows that A_PAR becomes smaller as number of MTJs per cell or number of cells per column increases.
per column increases. Fig. 5(c) plots equation (2) to relate $\text{APAR}_{\text{MTJ}}$ and the MTJ’s TMR with the $\text{TMR}_{\text{eff}}$. Intuitively, $\text{APAR}>1$ can provide a lower bound depending on the ability to accurately set the individual node voltages. Section III addresses these cases. Analysis made here provides a lower bound.

B. Impact of MTJ Sharing on Writes

Designing a cell with multiple MTJs for programming has two primary issues. First, writing a cell can result in an unintentional write of a non-accessed MTJ due to the parasitic paths. Second, the parasitic paths can reduce the current through the accessed MTJ hence requiring a larger access transistor to provide the current.

A column of cells with multi-MTJ cells can be ideally visualized graphically in Fig. 8(a) when the $I_{\text{MTJ}}$ at point A is greater than the minimum write current while the leakage current through any non-accessed device should be smaller than the maximum read current. (b) Programming current of ‘0’ ($\text{R}_{\text{AP}}$ to $\text{R}_P$) and Leakage current.

Determining the minimum ratio for Fig. 6 is equivalent to a simplified model of 2 cells per column and 2 MTJs per cell as shown in Fig. 7. The worst-case conditions for writing $\text{P} \rightarrow \text{AP}$ and $\text{AP} \rightarrow \text{P}$ are shown. Writing a “1” (P→AP) can be visualized graphically in Fig. 8(a) when the $I_{\text{MTJ}}$ at point A > min($I_{\text{W}}(\text{P} \rightarrow \text{AP})$).

<table>
<thead>
<tr>
<th>$\text{R}_1$</th>
<th>$\text{R}_2$</th>
<th>$\text{R}_3$</th>
<th>$\text{R}_4$</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$I_{\text{R} \rightarrow \text{AP}}^{\text{MAX}} &lt; I_{\text{READ}}^{\text{MIN}}$</td>
</tr>
<tr>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$I_{\text{R} \rightarrow \text{AP}}^{\text{MAX}} &lt; I_{\text{READ}}^{\text{MIN}}$</td>
</tr>
<tr>
<td>$\text{R}_{\text{AP}}$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$I_{\text{R} \rightarrow \text{AP}}^{\text{MAX}} &lt; I_{\text{READ}}^{\text{MIN}}$</td>
</tr>
<tr>
<td>$\text{R}_{\text{AP}}$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$\text{R}_P$</td>
<td>$I_{\text{R} \rightarrow \text{AP}}^{\text{MAX}} &lt; I_{\text{READ}}^{\text{MIN}}$</td>
</tr>
</tbody>
</table>
With this $I_{MTJ}$ through the accessed MTJ, the worst-case parasitic current can be defined as listed in Table I where $R_2$ and $R_3$ in the first case and $R_4$ in the second case have a chance to be flipped if $I_{PAR,MTJ} > \max(I_R)$. When resistances through the parasitic path are equal, the ratio of $\max(I_R)/\min(I_W)$ is roughly 1/3. As shown in the Table I, the value can depend TMR since the programmed value of the non-accessed MTJ can vary between $R_P$ or $R_{AP}$. Fig. 8(b) and Table I show the worst case and ratio for writing a “0” (AP→P). Fig. 9 visualizes requirements listed in Table I. As $I_W$ becomes highly asymmetric, the minimum ratio of read and write current should be higher. This plot indicates that the worst-case condition is generally dominated by maximum read current of $R_{AP}$.

The current through the parasitic resistance paths determine the additional amount of current that must be sourced by the access transistor. The lower bound of roughly 1/3 indicates that the access device must be increased by at least 1/3 of $\min(I_W(P\rightarrow AP))$ for each additional MTJ/cell.

Depending on the ability to control the voltages at each node, a higher ratio may be required which implies that the given MTJ has very narrow state transition region over control current \cite{8}. The results in the next section assume the least complex implementation where nodes $V_{AC2}$ in Fig. 6 are not driven to a specific voltage when the cell is not accessed.

### IV. SIMULATION RESULTS

Feasibility and performance of using multi-MTJ cells that share a single access transistor depends on MTJ characteristics. Our analysis optimizes the area of the cell in a multi-MTJ design using an MTJ with characteristics as indicated in Table II. Values in the table are based on measured results of recent publications \cite{8, 9} with adjustable read/write current ratio. The cell area is minimized based on reducing the required access device size for writing. The minimum device size is determined by adjusting bit-line voltages to meet the design targets given in Table II. Maximum parasitic current of non-accessed MTJ is carefully tested under various worst case conditions to remain within the MTJ specification. For this simulation, a 45nm CMOS technology is used with HSPICE.

#### TABLE II. SIMULATED MTJ SPECS

<table>
<thead>
<tr>
<th>MTJ specifications</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_A$</td>
<td>4.9 $\Omega \mu m^2$</td>
</tr>
<tr>
<td>$R_P$</td>
<td>750 $\Omega$</td>
</tr>
<tr>
<td>TMR</td>
<td>120 (%)</td>
</tr>
<tr>
<td>Switching Current</td>
<td></td>
</tr>
<tr>
<td>$(I_W(P\rightarrow AP))$</td>
<td>500 $\mu A$</td>
</tr>
<tr>
<td>$(I_{AP}(AP\rightarrow P))$</td>
<td>375 $\mu A$</td>
</tr>
<tr>
<td>$\max(I_W(P)/\min(I_W(P\rightarrow AP))$</td>
<td>0.55</td>
</tr>
<tr>
<td>$\max(I_{AP}(AP\rightarrow P))/\min(I_W(AP\rightarrow P))$</td>
<td>0.6–0.8</td>
</tr>
</tbody>
</table>

Our analysis explores the impact of the shared design on cell area and effective TMR for various numbers of MTJs per cell and numbers of cells per column. The simulation results of the shared structure are normalized by the 1T-1MTJ cell area.

#### A. Normalized NFET Size

Figure 10 shows normalized device area of the access transistor as a function of # MTJ per cell with different ratios of $\max(I_{AP}(AP\rightarrow P))/\min(I_W(AP\rightarrow P))$. With a large number of MTJs per cell, the parasitic path has low impedance leading to higher parasitic currents. Hence, the bit-lines associated with the non-accessed cells should be driven as strongly as possible to avoid sizing up the access NFET unnecessarily. An inherent tradeoff is present with the $\max(I_{AP}(AP\rightarrow P))/\min(I_W(AP\rightarrow P))$ ratio...
and the amount of parasitic current. As shown in the figure, if the node at the access device of a non-accessed cell is not driven, the ratio must increase to be greater than 0.5 for the multi-MTJ structure to be feasible. The slope of each curve increases with larger numbers of MTJs/cell because the non-accessed nodes (nodes $V_{\text{AC2}}$ in Fig. 6) are not simultaneously driven. The larger number of parasitic paths in parallel with increasing number of MTJs/cell results in greater parasitic current through the non-accessed devices. Hence, the area efficiency degrades.

Fig. 11 sweeps the NFET size as a function of the number of cells per column (i.e. sub-array size). The sub-array size has little impact on the NFET size as long as net parasitic resistance is greater than the source impedance of bit-line driver. Once bit-line driver loses control over idle bit-lines, however, the performance worsens considerably where the parasitic current cannot be properly controlled. Equivalently, if the non-accessed bit-lines are floating, the amount of NFET current is roughly proportional to the number of MTJs per cell and is very sensitive to the number of cells per column.

B. Simulated Differential Read Current

As indicated by the effective TMR in Equation (2), parasitic resistors desensitize resistance variation of target MTJ. Fig. 12 shows that the differential read current is always less than 60% of 1T-1MTJ case. In this simulation, the access device size has been scaled as in Fig. 11. The results indicate that the differential current drops substantially with increasing number of MTJs per cell and hence realistically only 2 or 3 MTJs per cell is functionally feasible. With larger numbers of cells per column, we anticipate the differential read current to be <20% of the 1T-1MTJ case putting more demand on the sensing circuitry or longer time for sensing.

C. Area Saving

Assuming that access transistor size is a dominant factor of memory density, cell area reduction rate per MTJ is shown in Fig. 13. Note that this ratio is normalized to cell area of 1T-1MTJ case. This figure clearly indicates that higher memory density can be attained though the cost comes from reduced sensing margin and increased risk of data loss. Table III summarizes simulation results.

V. CONCLUSION

This paper describes the impact of using multiple MTJs that shares a single access device in a cell. Because of the resistive nature of the MTJs, device sharing has substantial limitations due to the parasitic resistive paths that degrade both the read differential current and write stability. The technique promises modest improvements in the array density but can only be applied to small number of MTJs per cell and limited column size. Furthermore, MTJ characteristics must have a high max($I_{\text{on}}$)/min($I_{\text{off}}$) ratio in order to avoid instability of non-accessed MTJs.
ACKNOWLEDGMENT

This work was supported by the DARPA/MTO STT-RAM program. The authors would like to thank UCLA STT-RAM team for helpful discussions and for providing measurement data.

REFERENCES