Thermal Magnetic Random Access Memory

IEEE International Conference on Computer Design
New Memory Technologies
San Jose, CA October 4, 2005

James Deak
NVE Corporation
Participants

• Jim Daughton - PI
• Art Pohm
• Dexin Wang
• Dave Brownell
• Pete Eames
• John Taylor

Supported in part by DARPA
Motivation

MRAM combines the non-volatility and infinite endurance of magnetic based memories with silicon processing technology.

<table>
<thead>
<tr>
<th>Storage Mechanism</th>
<th>SRAM</th>
<th>DRAM</th>
<th>Flash</th>
<th>FeRAM</th>
<th>CRAM</th>
<th>MRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Speed</td>
<td>Fast</td>
<td>Medium</td>
<td>Fast</td>
<td>Fast</td>
<td>Fast</td>
<td>Fast</td>
</tr>
<tr>
<td>Write Speed</td>
<td>Fast</td>
<td>Medium</td>
<td>Slow</td>
<td>Fast</td>
<td>Medium</td>
<td>Fast</td>
</tr>
<tr>
<td>Endurance</td>
<td>Infinite</td>
<td>Infinite</td>
<td>1E5</td>
<td>&gt;1E10</td>
<td>1E12</td>
<td>Infinite</td>
</tr>
<tr>
<td>Power</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Refresh</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Potential Cell Density</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Non-volatility</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Magnetoresistive Random Access Memory, By Saied Tehrani et al.

MRAM Issues:
• How to reliably select a random bit for writing
• How to combine low power and data retention
Outline

- MRAM Introduction
  - Status
  - MRAM Background
- MRAM Scaling Issues
- Magnetothermal MRAM
  - Modes – Curie, Neel/blocking temperature
  - Design Requirements
  - Thermal Writing Experiments
- Conclusions
MRAM Technology Status

256K x 16-Bit 3.3-V Asynchronous Magnetoresistive RAM

Introduction
The MR2A16A is a 256K x 16-bit magnetic random access memory (MRAM) device organized as 256,144 words of 16 bits. The MR2A16A is equipped with chip enable (CE), write enable (WE), and output enable (OE) pins, allowing for a significant amount of flexibility without bus contention. Because the MR2A16A has separate byte-select controls (I/O and I/OE), individual bytes can be written and read.

MRAM is a non-volatile memory technology that persists data in the event of power loss and does not require periodic refreshing. The MR2A16A is a high-quality solution for applications that must permanently store and retrieve critical data quickly.

The MR2A16A is available in a 440-mil 44-pin plastic small-outline TSOEP type-II package with an industry-standard 3.3-V power and ground (GND) plane.

Features
- Single 3.3-V power supply
- Commercial temperature range (0°C to 70°C)
- Symmetrical high-speed read and write with fast access times (25 ns)
- Full data bus control — 9-bit or 16-bit access
- Equal address and chip enable access times
- Automatic data protection from low-voltage inhibit circuitry to prevent write on power loss
- 94 inputs and outputs (74-pin version: 92 inputs and 2 outputs)
- Fully static operation
- Full nonvolatile operation with 10 years minimum data retention

Currently Sampling

Expected 2006
Magnetic Memories

- Hard drives
- Tape drives
- Floppy disks
- Magnetic Core Memory – 1948 thru the 1970s

53k bit in 512 cubic inches

Consider magnetic core memory

Here bits are read destructively by detecting a voltage pulse in a read line.

Voltage pulse created by writing a toroid at the intersection of two write conductors.

\[ V_{\text{sense}} \propto (\text{Volume}) \frac{dM}{dt} \]

• Read signal decreases with decreasing magnetic bit size.

• Does not scale well.
Tunneling Magnetoresistance

Julliere’s Model

*Phys. Lett. A 54, 225 (1975)*

- Spin conserved during tunneling
- Two independent spin channels
  - Parallel – Low Resistance
    - Majority $\rightarrow$ Majority
    - Minority $\rightarrow$ Minority
  - Anti-Parallel – High Resistance
    - Majority $\rightarrow$ Minority
    - Minority $\rightarrow$ Majority
- Room Temperature Max. TMR
  - $\text{AlO}_x \sim 70\%$
  - $\text{MgO} \sim 260\%$

Tunnel Conductance

Parallel

$$G_{\uparrow\uparrow} \propto n_L^\uparrow n_R^\uparrow + n_L^\downarrow n_R^\downarrow$$

Anti-Parallel

$$G_{\uparrow\downarrow} \propto n_L^\uparrow n_R^\downarrow + n_L^\downarrow n_R^\uparrow$$

Magnetoresistance

$$TMR = \frac{G_{\uparrow\uparrow} - G_{\uparrow\downarrow}}{G_{\uparrow\downarrow}} = \frac{R_{AP} - R_P}{R_P} = \frac{2P_L P_R}{1 - P_L P_R}$$

Spin Polarization

Left

$$P_L = \frac{n_L^\uparrow - n_L^\downarrow}{n_L^\uparrow + n_L^\downarrow}$$

Right

$$P_R = \frac{n_R^\uparrow - n_R^\downarrow}{n_R^\uparrow + n_R^\downarrow}$$

$n \sim$ spin dependent DOS
Typical MRAM Cell

**Read Mode**
- Sense Current
- Bit Line
- Free Magnetic Layer, Information Storage.
- Tunneling Barrier
- Fixed Magnetic Layer
- Digit Line
- Isolation Transistor “ON”

**Program Mode**
- Program Current $H_e$
- Program Current $H_h$
• Ix, Iy Alone Doesn’t Switch Cell
• Ix, Iy Together Switch Cell
Control of the distribution of switching fields makes the SW-MRAM implementation difficult to manufacture and scale.

Becomes more difficult as cell size is reduced
Stability and Thermal Activation

Magnetic Energy

\[ E = -\frac{1}{2} (N_y - N_x) M_s^2 V \sin^2 \theta - M_s \left( H_{\text{easy}} \cos \theta + H_{\text{hard}} \sin \theta \right) \]

where

\[ H_c = 2K / M_s \]

\[ E_b = \frac{1}{2} (N_y - N_x) M_s^2 V = K V \]

Probability for magnetization to be reversed by thermal activation in time \( t \) is

\[ P(t) = 1 - \exp\left( -\frac{t}{\tau} \right) \]

where

\[ \frac{1}{\tau} = f_0 \exp\left( \frac{-E_b(H)}{k_B T} \right) \]

Increasing \( E_b/k_B T \) increases the retention time but also increases programming field.
Energy Barrier and Error Rate in MRAM

Error Rate is related to the Barrier Height

\[
P(t) = \frac{N_{\text{rev}}}{N_{\text{bits}}} = 1 - e^{-t/\tau} = 1 - e^{-t f_0 e^{-E_b/k_B T}}
\]

\[-t f_0 e^{-E_b/k_B T} = \ln \left( 1 - \frac{N_{\text{rev}}}{N_{\text{bits}}} \right) \approx -\frac{N_{\text{rev}}}{N_{\text{bits}}} \quad \text{If } N_{\text{rev}} \ll N_{\text{bits}}\]

\[
\frac{N_{\text{rev}}}{t} = \frac{\text{err}}{s} = N_{\text{bits}} f_0 e^{-E_b/k_B T}
\]

\[
\frac{\text{err}}{hr} = (\text{WritFrac}) N_{\text{bits}} f_0 e^{-E_b/k_B T} \frac{-E_b(H)}{k_B T} (3600 \frac{s}{hr})
\]

Energy barrier can thus be determined as

\[
E_b = -\ln \left[ \frac{\text{err} / hr}{(\text{WrtFrac})(N)(f_0)(3600)} \right] k_B T
\]

- Required \( E_b \) is thus directly related to the desired error rate, \( T \), and \( N_{\text{bits}} \)
- \( E_b \) should be > 66 \( k_B T \sim 3.7e-19 \) J, for 1 Mbit, 1e-10 err/hr, 400 K
Write Current and Cell Size

- Small volume lowers thermal stability – need to increase thickness with decreasing size

- For small single layer free layers, increasing stability increases write field/current to impractical values
MRAM Scaling Issues

- Control of switching field distributions
- Large write current required for stability
Alternative Write Methods

- Toggle
- Spin Momentum Transfer
- Thermal Writing
  - Curie Point Write, $T_c$
  - Neel ($T_N$) or Blocking Temperature ($T_B$) Write
  - Thermal Spin-Momentum Transfer Write, T-SMT
Savtchenko “Toggle” Write

- Enhanced write selectivity due to pulse sequence
- SAF Stack FM/Ru/FM
Freescale 4 Mbit MRAM
• CMOS 0.18 \( \mu \text{m} \)
• 1T/1MTJ
• 25 ns Write/read
• 48 F\(^2\) cell (1.55 \( \mu \text{m}^2 \))

Results
• Excellent Selectivity
• High write current (~7 mA/line)

Thermal stability at small bit size still an issue \( \rightarrow \) high write current
• No external field required to reverse magnetization

• Angular momentum exchange between spin polarized current and ferromagnet induces torque and precession

\[
\text{Torque} \sim \pm J M_2 \times (M_1 \times M_2); \quad J \sim 10^7 \text{A/cm}^2
\]

Slonczewski (1996)

• At critical I precession large and magnetization precesses into reversed state

• Reversed magnetization stable since torque becomes zero
$E_b$ must remain high as bit size is scaled down – thickness must increase

Switching roughly proportional to $E_b$

At fixed $E_b$, smaller AR bits require more current

Smaller dimensions do not guarantee lower write current!
Principle of operation

• Use a very high stability storage layer to improve data retention.

• Apply current through the tunnel junction to heat the storage layer.

• Storage layer coercivity reduced at high temperature.

• Storage layer may be an AF/FM bilayer or a low T_c FM.

Goal: Design for high density → Heating current below 100 µA at 100 nm bit size.
Curie Point MRAM

Coercivity drops due to reduction of free layer magnetization

\[
\text{Si-SiN-CoFeNiSiB (16nm)-Ta (10nm)}
\]

\[
\begin{array}{c}
\text{Temperature (deg C)} \\
0 & 50 & 100 & 150 & 200 & 250 & 300 \\
\text{M_s (emu/cc)} \\
0 & 100 & 200 & 300 & 400 & 500 \\
\end{array}
\]

Low T

High T
Neel Point MRAM

Offset field ($H_{ex}$) drops to zero at high temperature ($T_B$)

Best mode since AF/FM bilayer easy to achieve high $E_B$ due to high FM/AF coupling and high AF anisotropy. Also has the best magnetic field immunity.
Previous $T_B$ MRAM Work

Low-Current Blocking Temperature Writing of Double-Barrier MRAM Cells

Jianguo Wang and P. P. Freitas

• Heating through barrier confirmed
• $T_B$ writing mechanism verified w/ 10 ns write pulse
• NOL layers used to increase RA and take stress off the barrier – at the expense of TMR – $V_{\text{lim}} = 2$ V
• Not expected to scale well due to stray field interaction in the FM/AF/FM storage layer
• Demonstrated $>10^{10}$ thermal writes

![Diagram of MRAM cell](image)
NVE Neel Point (or $T_B$) MRAM Operation

Write Sequence

- **initial state:** 1
- **heat**
- **field cool**
- **final state:** 0
Neel Point MRAM Considerations

- Need margin between high $T_b$ and Low $T_b$ layers (Easy)
- Low thermal conductivity vias to lower write current (several alternative materials such as PtMn and BiTe)
- Low thermal conductivity encapsulating material to lower write current (process requires some work)
- Barrier RA must be optimized to permit efficient heating at low write current, while maintaining a large TMR (Not too easy, but not impossible)
- AF must have $H_{ex} > H_c$, Low $H_c$ at $T_B$, $T_B > 150$ C (Easy)
Resistance Area Product Requirements

Junction temperature: \[ \Delta T = R_{th} P = \frac{L_v}{2K_v A} I^2 \langle RA \rangle \]

Heating current: \[ I(\Delta T) = A \sqrt{\frac{\Delta T 2 K_v}{L_v \langle RA \rangle}} \] Heating current proportional to bit area

Voltage: \[ V(\Delta T) = I(\Delta T) R = \sqrt{\frac{\Delta T 2 K_v \langle RA \rangle}{L_v}} \] Must limit < ~0.5 V for reliability

Maximum RA: \[ \langle RA \rangle_{\text{max}} \leq \frac{V_{\text{Lim}}^2 L_v}{2 \Delta T K_v} \] Optimal value depends on \( V_{\text{lim}}, R_{th}, \) and desired \( \Delta T \).
Encapsulating Material Optimization

- Heat diffuses into the encapsulating material \( \rightarrow \) Lowers heating efficiency

- Looks as if the thermal vias have a larger effective area

\[
A_{\text{eff}} = \left( \frac{2K_v \Delta T}{PL} \right) \frac{1}{A_{\text{via}}}
\]

Best if \( K_m < K_v/10 \)
Thermal Via Optimization

\[ \frac{K_v}{L_v} \leq \frac{V_{lim}^2}{2\Delta T \langle RA \rangle_{\text{max}}} \]

- Need minimum \( K_v \) and large \( L_v \)
- Can’t increase \( L_v \) too much
  - Thermal rise time becomes too long
  - Roughness becomes too large for low RA barrier
- Settle on 50 nm of PtMn
  - low \( K_v = 0.01 \) to 0.014 W/cm-K
  - Able to fabricate low RA junctions at this thickness
Test Structure Fabrication

Anneal Low $T_b$ AF +H
Anneal Low $T_b$ AF -H

E-beam and optical lithography

Unpatterned film measurement

PtMn(50)/Ru(4)/IrMn(4)/CoFeB(4)/AlOx(x)/CoFeB(4)/Ru(0.7)/FeCo(4)/PtMn(50)/Al(20)

Writing accomplished using external field

Bits patterned from 0.05 µm to 5 µm.

$T_B \sim 110 \degree C$
Slightly low

Post bit ion mill
Post top conductor etch
High RA Junction (13 Å Al, 28 MΩµm²)

Device burns out before IrMn Tₘ is exceeded

No TMR at 1000 nA
Check R(H) at lower current

Device burns out before IrMn Tₘ is exceeded

No change at 500 nA

Low RA Junction (6 Å Al, 5 Ωµm²)

Changes from SV to PSV behavior – Why?
Model Description

Pinned layer/barrier/storage layer

\[ \text{FM}_1(t_1,M_{s1},K_1)/\text{Ru}(J_{RU})/\text{FM}_2(t_2,M_{s2},K_2)/\text{AlO}_x(J_{AlO})/\text{FM}_3(t_3,M_{s3},K_3) \]

Strong coupling \hspace{1cm} Weak coupling

<table>
<thead>
<tr>
<th>Total Energy</th>
<th>( E_{tot} = E_z + E_{exc} + E_K + E_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>External Fields</td>
<td>( E_z = -B_x \left[ M_{s1} t_1 \cos(\phi_1) + M_{s2} t_2 \cos(\phi_2) + M_{s3} t_3 \cos(\phi_3) \right] ) ( -B_y \left[ M_{s1} t_1 \sin(\phi_1) + M_{s2} t_2 \sin(\phi_2) + M_{s3} t_3 \sin(\phi_3) \right] )</td>
</tr>
<tr>
<td>Interlayer Exchange</td>
<td>( E_{exc} = J_{RU} \cos(\phi_1 - \phi_2) + J_{AlO} \cos(\phi_2 - \phi_3) )</td>
</tr>
<tr>
<td>Anisotropy</td>
<td>( E_K = K_1 t_1 \sin^2(\phi_1) + K_2 t_2 \sin^2(\phi_2) + K_3 t_3 \sin^2(\phi_3) )</td>
</tr>
<tr>
<td>Demagnetization Fields</td>
<td>( E_d = \frac{1}{2} \mu_0 t_1 M_{s1} H_{d1} + \frac{1}{2} \mu_0 t_2 M_{s2} H_{d2} + \frac{1}{2} \mu_0 t_3 M_{s3} H_{d3} )</td>
</tr>
</tbody>
</table>

\[ H_{di} = \begin{bmatrix} H_{i}^x \\ H_{i}^y \end{bmatrix} = \sum_{j=1}^{3} \begin{bmatrix} N_{xx}(i,j) & N_{xy}(i,j) & M_{j}^x \\ N_{yx}(i,j) & N_{yy}(i,j) & M_{j}^y \end{bmatrix} \]
Simulated Response

- One layer pinned - high temperature
- Both layers pinned - low temperature

- Coercivity of the storage layer is too high at $T_B$!
- Apparently $H_c \text{ IrMn} > (H_c + H_{ex}) \text{ PtMn}$
- And pinned layer SAF not well matched
- $H_c < 50 \text{ Oe}$ Wanted this at $T_B$,
- $H_c > 100 \text{ Oe}$ Got this at $T_B$, $H_c > 100 \text{ Oe}$
IrMn $H_{ex}$ and $H_c$

Must increase IrMn layer to 60 nm

- Increasing thickness of IrMn decreases $H_c$ and increases $H_e$.
- $T_b$ would be increased 50 to 100 K.
- Increasing RA to 10 $\Omega \mu m^2$ would keep the write current at the present value and still permit safe writing

\[
I(\Delta T) = A \frac{\Delta T^2 K_v}{L_v \langle RA \rangle}
\]
Measured Programming Current

Current through barrier required to exceed $T_b$

\[ y = 8.3116x \]

\[
\pi \times (0.1/2) \times (0.1/2) \times 8.31 = 0.065 \text{ mA}
\]

Using only a single barrier and no heater layers

100 nm bit $\Rightarrow \pi \times (0.1/2) \times (0.1/2) \times 8.31 = 0.065 \text{ mA}$
Conclusions

- Demonstrated heating through the barrier, and potential for < 100 μA write currents at 100 nm dimensions
- IrMn thickness needs to be increased
- RA needs to be optimized, need roughly 50 Ωμm²
- Need to do a better job matching the SAF
- Lower thermal conductivity encapsulating material like BCB would further decrease write current
- If needed NOL heaters could be added to further decrease write current (at the expense of signal)