Embedded Memory Background

- System Memory Requirement is Exploding

- Average system memory in 2G cellphone is 10MB or 60Mb!! In 2.5G/3G this is exploding to 20-40MB!! Same is true for many other applications.
- NOVO is key memory type required in high growth applications.
- This has spawned the SIP (stacked die or stacked package) as a solution that meets BOM cost, time to market, footprint and flexibility goals of these applications.

Subsequent slides courtesy Randy McKee, TI
Embedded Memory Alternatives

Competitive Developments (late 2004)

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>Infineon</th>
<th>Intel</th>
<th>NEC</th>
<th>Samsung</th>
<th>STC</th>
<th>TI</th>
<th>Toshiba</th>
<th>TSMC</th>
<th>UMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDRAM</td>
<td>65nm</td>
<td>65nm</td>
<td>90nm</td>
<td>90nm</td>
<td>65nm</td>
<td>90nm</td>
<td>65nm</td>
<td>90nm</td>
<td>90nm</td>
<td>90nm</td>
</tr>
<tr>
<td>FB cDRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eFlash</td>
<td>90nm</td>
<td>90nm</td>
<td>90nm</td>
<td>TBD</td>
<td>65nm</td>
<td>90nm</td>
<td>130nm</td>
<td>90nm</td>
<td>130nm</td>
<td>130nm</td>
</tr>
<tr>
<td>FRAM</td>
<td>180nm</td>
<td></td>
<td>250nm</td>
<td>180/90nm</td>
<td></td>
<td>130nm</td>
<td>130nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUM</td>
<td>180nm</td>
<td></td>
<td>180nm</td>
<td>180nm</td>
<td></td>
<td>180nm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRAM</td>
<td>180nm</td>
<td>180nm</td>
<td>180/90nm</td>
<td>TBD</td>
<td>180nm</td>
<td>90nm</td>
<td>180nm</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FRAM Basics

Basic Principle
- Polar crystalline material which can be switched between 2 stable states by applying an external electric field
- DRAM like 1T-1C cell structure - plate driver added to sense switched charge

Key Developers
- Ramtron
- TI
- Samsung
- Toshiba
- Infineon

FERROELECTRIC MEMORY CELL: READ/WRITE OPERATION (1T-1C)

Simplified Memory Circuit

Simplified Timing Diagram

For “1” - \( Q = \Delta PA - (P_{off} + P_A)A \)
For “0” - \( Q = \Delta PA - (P_{on} - P_A)A \)

(Zurcher, Motorola, 1995)
TI 130nm FeRAM Structure

Masks
MET3
MET2
MET1
VIA0
FCAP
CONT
Parallel to Bit Line

BEOL
(No Changes)

eFRAM Module

FEOL
(No Changes)

23 Total Masks (including PO and ALCAP)

FABRICATION OF FERROELECTRIC CAPACITORS

Material Selection

Layer Structure

Hardmask
- Provides “fenceless” structure
- High selectivity relative to Ir

Top Electrode
- Oxidized to maximize endurance
- Must be stable during anneals

Ferroelectric – MOCVD Process
- Large polarization
- Low formation temperature
- Small grain size

Bottom Electrode
- Remains conductive after PZT dep.
- Diffusion barrier

Conducting Diffusion Barrier
- Adhesion
- Low oxidation rate
eFRAM Benefits and Issues

Potential reasons to use eFRAM

- 2x density advantage over SRAM by adding 2 masks
- Large memory with wide internal I/O like eDRAM, but with low leakage (cost vs eDRAM depends on cell size and mask count tradeoff)
- Low standby power ‘instant on’ capability due to non-volatile + fast write
- Lower cost embedded non-volatile memory – eFlash replacement
- Lower soft error rate

Problems with eFRAM

- Slow access time relative to SRAM (also slower than eDRAM due to need for plate drive)
- 2x density advantage over SRAM may decline in future nodes
- Endurance may be limited to ~10^{14} cycles (destructive read means limited read and write cycles - OK for many applications)

MRAM

3d-transition metals

Free atom configurations

A ferromagnetic substance contains permanent atomic magnetic dipoles that are spontaneously oriented parallel to one another, even in the absence of an external field. This seemingly unreasonable alignment of side-by-side dipoles is explained by a quantum mechanical effect known as the exchange interaction.

The magnetic properties of iron are thought to be the result of the magnetic moment associated with the spin of an electron in an outer atomic shell—specifically, the third d shell. The Pauli exclusion principle prohibits two electrons from having identical properties; for example, no two electrons can be in the same location and have spins in the same direction. This exclusion can be viewed as a “repulsive” mechanism for spins in the same direction. Its effect is opposite that required to align the electrons responsible for the magnetization in the iron domains. However, other electrons with spins in the opposite direction, primarily in the fourth s atomic shell, interact at close range with the magnetization electrons, and this interaction is attractive. Because of the attractive effect of their opposite spins, these s-shell electrons influence the magnetization elections of a number of the iron atoms and align them with each other.
Giant Magneto-Resistance (GMR) Effect

Figure 8. GMR sensor basics

From IBM article on web

Motorola Explanation of Tunneling MR Effect

The tunneling MR can be understood in terms of a two-band model in which the d-band is split into spin-up and spin-down bands with different density of states at the Fermi energy. When the magnetization of the layers is parallel, the majority-band electrons tunnel across to the majority band of the opposing electrode and the minority to the minority band. When they are antiparallel, the majority/minority band electrons are forced to tunnel into the minority/majority band of the opposing electrode. The reduced number of states available for tunneling between the ferromagnetic layers when the layers are antiparallel results in an increased tunneling resistance as compared to parallel.

From Motorola, JCM, 5/00
**1T/1-1MTJ MRAM**

Cell operation
- Free layer magnetic polarization is switched by combined field created by currents in the bit line and digit line (transistor is off)
- Data state is sensed by current flow in bitline with transistor on
- Tunnel barrier resistance is higher when free layer is polarized anti-parallel

Key Developers
- IBM
- Freescale (Motorola)
- Infineon

---

**MRAM Status and Issues**

Development status
- Current work is focused on stand-alone memories
- Freescale (Motorola) is sampling a 4Mb MRAM device
  - 1.55um^2 cell size in a 0.18um technology
- IBM is publishing work on a 16Mb device
  - 1.42um^2 cell size in a 0.18um technology, 30nsec read/write cycle

Problems with MRAM
- Ultra-thin tunnel barrier (tunnel resistance control and reliability)
- Several mA write currents (power, limited I/O width and EM concerns)
- Complex MTJ stack could be difficult to manufacture
- Coupling to adjacent cells – write disturb sensitivity
- Transition to super-paramagnetic behavior may limit scaling
- 250-300C processing temperature limitation
Ovonics Unified Memory

- Operation
  - Chalcogenide material alloys used in re-writable CDs and DVDs
  - Electrical energy (heat) converts the material between crystalline (conductive) and amorphous (resistive) phases
  - Cell reads by measuring resistance
  - Non-destructive read
  - $\sim 10^{12}$ write/erase cycles
  - Medium write power
  - Relatively easy integration with CMOS

Samsung 64Mb PRAM (or PCRAM)

- Unit cell components
  - 1 GST + 1 NMOS Transistor
- Characteristics
  - easy to integrate into conventional CMOS process
  - BEC size ~ 45nm
  - cell size: $0.56 \times 0.90 = 0.504 \mu m^2$ ($16F^2$, $D/R=0.18 \mu m$)

Samsung – ISSCC 2004
**PCRAM Status and Issues**

Development status

- Being studied by Ovonyx, Intel, ST, Samsung, others
- Like MRAM, work is focused on more mature technology nodes
- Samsung example – 64Mb, 0.18um, 3V, 0.5um², 60/120ns read/SET
- Much of the work is still focused on cell physics, high density array manufacturing and cell reliability

Problems with PCRAM

- Limited endurance – $10^{12}$ cycles? – reliability of heater/GST interface
- High write current/power (should improve with scaling)
- Voltage scaling may be a problem (currently at 3V)
- Increased resistance and reduced read current with scaling
- Sub-resolution bottom electrode contact hole scaling

---

**When Does SIP Make Sense?**

**Good Example:**

DSP chip + SDRAM chip

**SDRAM Package**

Size : 8.1x9.3mm
Pincount : 54
Pitch : 0.80mm

**DSP Package**

Size: 12x12mm
Pincount : 289
Pitch : 0.50mm

**SIP Version:**

100% internal connections

**SIP Package**

Size : 12x12mm
Pincount : 289-35 = 264
Pitch : 0.50mm

- **Area reduction** : 35%
- **Package pitch** : 0.50mm
- **Package costs** : Decrease
- **PCB** : Decrease
- **Test cost** : Similar
- **System Cost** : +/- Parity? Potential for saving

- Fewer Pins
- Smaller Area
- Lower power
- Higher performance
Some stacked solutions (TI)

Prototyped 2001

OMAP1513. In Production
1510 + 2x128Mbit Flash
3 die, 2 spacers = 5 ‘die’ total

OMAP733 - 3 die. Production 1Q04
OMAP730
2x128Mbit SDR

EMBEDDED MEMORY TECHNOLOGY COMPARISON

<table>
<thead>
<tr>
<th>Operating Voltage</th>
<th>SRAM</th>
<th>eFlash</th>
<th>eDRAM</th>
<th>Flash (EP)</th>
<th>eRAM (EP)</th>
<th>FeRAM</th>
<th>MRAM</th>
<th>PCRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vdd</td>
<td>10V</td>
<td>Vdd+boost</td>
<td>10V</td>
<td>Vdd+boost</td>
<td>Vdd+boost</td>
<td>Vdd</td>
<td>Vdd</td>
<td>3V</td>
</tr>
<tr>
<td>Non-volatile</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Endurance</td>
<td>unlimited</td>
<td>10^5</td>
<td>unlimited</td>
<td>10^5</td>
<td>unlimited</td>
<td>10^11</td>
<td>unlimited?</td>
<td>10^12</td>
</tr>
<tr>
<td>50nm cell size (um2)</td>
<td>0.97</td>
<td>0.05 (0.20 for automotive)</td>
<td>0.17</td>
<td>0.08</td>
<td>0.08</td>
<td>0.35</td>
<td>0.35</td>
<td>0.05</td>
</tr>
<tr>
<td>2004 cell size - in development (um2)</td>
<td>0.40</td>
<td>0.08 (0.33 for automotive)</td>
<td>0.11</td>
<td>0.04</td>
<td>0.04</td>
<td>0.54</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Mask area to single gate oxide logic</td>
<td>5</td>
<td>9</td>
<td>4-8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Random Access (ns)</td>
<td>2</td>
<td>25</td>
<td>5</td>
<td>50</td>
<td>50</td>
<td>25</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Standby Power (uA/Mb)</td>
<td>10</td>
<td>&lt;1</td>
<td>100</td>
<td>&lt;1</td>
<td>&lt;3</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Position</td>
<td>low cost, fast, meets most product needs</td>
<td>high cost, limited endurance, automotive application w/ relaxed cell size</td>
<td>high cost, high standby power, enabling large memory w/ wide I/O</td>
<td>low cost, used for some products</td>
<td>low cost, low standby power, used for some products</td>
<td>new materials, low cost, low standby power, fast write + non-volatile</td>
<td>new materials, low cost, low standby power, high write power, slow read</td>
<td>new materials, low cost, low standby power, high write power, limited endurance?</td>
</tr>
</tbody>
</table>