EE241 - Spring 2006
Advanced Digital Integrated Circuits

Lecture 22: Embedded Memory

---

Some Info

- **Rest of the semester plan**
  - Week 14 (this week): memory
  - Week 15: Test and Arithmetic
  - Week 16: (1 lecture on May 8): Future perspectives

- Project Presentations: Tuesday May 9 – Full afternoon
- Take home final: May 12-15
Semiconductor Memory Trends

Trends in Memory Cell Area

From [Itoh01]
Trends in RAM Developments (R&D)

Memory Capacity/Chip (bits)


16 G 1 G 4 M 64 M 256 K 1 G 64 M 4 M 16 K 1 K

Memory Cell Area (mm²)


10,000 1,000 100 10 1

Full CMOS

ISSCC/VLSI Circuits

ISSCC/VLSI Circuits

Stand-alone RAMs

Stand-alone RAMs

Full CMOS

Poly-Si load

TFT load

3-D capacitor

Planar capacitor

DRAM

SRAM

Embedded Memory Background

System Memory Requirement is Exploding

- Average system memory in 2G cellphone is 10MB or 80Mb!! 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

**FLASH**
- Control Gate
- Floating Gate
- Conduction-V

**SRAM**

**DRAM**

**FRAM**

**MRAM**

**PCRAM**

Data Storage Region
- Chalcogenide
- Phase Change Material
- Crystalline
- Heaters
- Resistive Electrode

Courtesy Randy McKee, TI

### EMBEDDED MEMORY TECHNOLOGY COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>SRAM</th>
<th>eFlash</th>
<th>eDRAM</th>
<th>Flash (SPD)</th>
<th>eRAM (SPD)</th>
<th>FeRAM</th>
<th>MRAM</th>
<th>PCRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Voltage</td>
<td>Vdd</td>
<td>10V</td>
<td>Vdd+boost</td>
<td>10V</td>
<td>Vdd+boost</td>
<td>Vdd</td>
<td>3V</td>
<td></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>10^12</td>
<td>unlimited?</td>
<td>10^12</td>
</tr>
<tr>
<td>40nm cell size (um²)</td>
<td>0.97</td>
<td>0.08 (0.20 for automotive)</td>
<td>0.17</td>
<td>0.18</td>
<td>0.35</td>
<td>0.35</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>2004 cell size (nm²)</td>
<td>2</td>
<td>0.98 (0.33 for automotive)</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.34</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>Mask adder to single gate oxide logic</td>
<td>0</td>
<td>5</td>
<td>4-8</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Random Access (ns)</td>
<td>2</td>
<td>25</td>
<td>6</td>
<td>50</td>
<td>25</td>
<td>30</td>
<td>100</td>
<td></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, low standby power, fast write, non-volatile</td>
<td>high cost, limited endurance, automotive application</td>
<td>high cost, high standby power, enable large memory size and 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</td>
<td>new materials, low standby power, high write power, slow read</td>
<td>new materials, low standby power, limited endurance, slow read</td>
</tr>
</tbody>
</table>
Competitive Developments (late 2004)

<table>
<thead>
<tr>
<th></th>
<th>IBM</th>
<th>Infineon</th>
<th>Intel</th>
<th>Motorola</th>
<th>NEC</th>
<th>Samsung</th>
<th>STM</th>
<th>TI</th>
<th>Toshiba</th>
<th>TSMC</th>
<th>UMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>eDRAM</td>
<td>65nm</td>
<td>65nm</td>
<td>90nm</td>
<td>90nm</td>
<td>65nm</td>
<td>90nm</td>
<td>65nm</td>
<td>65nm</td>
<td>90nm</td>
<td>90nm</td>
<td>90nm</td>
</tr>
<tr>
<td>FB eDRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130nm</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>
<td>130nm</td>
</tr>
<tr>
<td>FRAM</td>
<td>180nm</td>
<td></td>
<td>250nm</td>
<td>180/90nm</td>
<td>130nm</td>
<td>130nm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCM</td>
<td>180nm</td>
<td></td>
<td>180nm</td>
<td></td>
<td></td>
<td></td>
<td></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>
<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 (IT-1C)

Simplified Memory Circuit

Not Concerned with Capacitor Leakage - Only Polarization State of \( C_{\text{ferro}} \)

Simplified Timing Diagram

For “1” - \( Q = \Delta PA - (P_{\text{sat}} + P_{\text{t}})A \)
For “0” - \( Q = \Delta PA = (P_{\text{sat}} - P_{\text{t}})A \)

(Zurcher, Motorola, 1995)

---

TI 130nm FeRAM Structure

Masks

<table>
<thead>
<tr>
<th>BEOL (No Changes)</th>
<th>META1</th>
<th>META0</th>
<th>META0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VIA0</td>
<td>FCAP</td>
<td>CONT</td>
</tr>
<tr>
<td>eFRAM Module</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEOL (No Changes)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 – MO CVD 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
- Low 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 \( \sim 10^{14} \) cycles (destructive read means limited read and write cycles - OK for many applications)
MRAM

3d-transition metals

Free atom configurations

<table>
<thead>
<tr>
<th>Element</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc</td>
<td>4s^2 3d^1</td>
</tr>
<tr>
<td>Ti</td>
<td>4s^2 3d^2</td>
</tr>
<tr>
<td>V</td>
<td>4s^2 3d^3</td>
</tr>
<tr>
<td>Cr</td>
<td>4s^2 3d^4</td>
</tr>
<tr>
<td>Mn</td>
<td>4s^2 3d^5</td>
</tr>
<tr>
<td>Fe</td>
<td>4s^2 3d^6</td>
</tr>
<tr>
<td>Co</td>
<td>4s^2 3d^7</td>
</tr>
<tr>
<td>Ni</td>
<td>4s^2 3d^8</td>
</tr>
<tr>
<td>Cu</td>
<td>4s^2 3d^9</td>
</tr>
</tbody>
</table>

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 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 'repulsion' mechanism for spins in the same direction. Its effect is opposite that required to align the electron 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 electrons 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 band 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, JOM, 6/00

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

1-MTJ/1-Transistor Memory Cell

Motorola – 2002 VLSI Symposium
MRAM Status and Issues

Development status
- Current work is focused on stand-alone memories
- Freescale (Motorola) is sampling a 4Mb MRAM device
  - 1.55um² cell size in a 0.18um technology
- IBM is publishing work on a 16Mb device
  - 1.42um² 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
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 x 0.90 = 0.504µm² (16F², D/R=0.18µm)

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, 0.18µm, 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
- 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 = 254
- Pitch: 0.50

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area reduction</td>
<td>33%</td>
</tr>
<tr>
<td>Package pitch</td>
<td>0.50mm</td>
</tr>
<tr>
<td>Package costs</td>
<td>Decrease</td>
</tr>
<tr>
<td>PCB</td>
<td>Decrease</td>
</tr>
<tr>
<td>Test cost</td>
<td>Similar</td>
</tr>
<tr>
<td>System Cost</td>
<td>+/- Parity?</td>
</tr>
<tr>
<td>Potential for saving</td>
<td></td>
</tr>
</tbody>
</table>

- Fewer Pins
- Smaller Area
- Lower power
- Higher performance

---

Some stacked solutions (TI)

**Prototyped 2001**
- OMAP733
- 3 die
- Production 1Q04

**OMAP730**
- 2x128Mbit SDR

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

**OMAP1612**
- In Production
- 1611
- 256Mbit DDR
90 nm SRAM cell sizes

![Graph showing bit area and percentage larger than TI for different manufacturers.]
The art of making small cells

<table>
<thead>
<tr>
<th>Node</th>
<th>C15</th>
<th>C12</th>
<th>C10</th>
<th>C07</th>
<th>C05</th>
<th>C035</th>
<th>C027</th>
<th>C021</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litho Tool</td>
<td>I-Line</td>
<td>248nm DUV</td>
<td>193nm DUV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design Rules</td>
<td>Logic</td>
<td>SRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contacts</td>
<td>Borders</td>
<td>Borderless</td>
<td>Stretch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vias</td>
<td>Borders</td>
<td>Borderless</td>
<td>Damascene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Type</td>
<td>DRC Clean</td>
<td>Field Scalable</td>
<td>Small Aspect Ratio</td>
<td>Staggered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metal Lvs</td>
<td>DLM</td>
<td>TLM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6T SRAM Scaling Challenges

Stable operation
- RDF is a significant issue for small W/L SRAM transistors
- losing write and read margin as we scale voltages
- separate power rail for SRAM may be needed (bad for leakage)

Leakage
- transistor leakage components are increasing with every node
- standby leakage is the issue for low power applications -- active leakage is also an issue for high performance applications
- various power management methods are being used or considered (retention modes with reduced Vdd, raised Vss, Vddio on N-well, etc.)

SRAM cell specific design rules
- some key rules are contact-gate space, gate extension past active, gate-end to gate-end space, N+/P+ space, contact/active overlap
- manual tuning of layout/SLAF's/OPC and bit swap allow us to push rules in the SRAM cell
- need more innovations like SAR layout, stretch contacts, staggered poly ...

SRAM Yield Limitations

- **Read Stability**
  - Cell can flip due to increase in the “0” storage node above the trip voltage of the other inverter during a read.

- **Hold Stability**
  - Data retention current not able to compensate the leakage currents.

- **Access Time**
  - Time required to produce a pre-specified ΔV between the bit lines is higher than the maximum tolerable limit.

- **Write Stability**
  - “1” Storage node may not be reduced below the trip point of the other inverter before WL is discharged.

*Mukhopadhyay et al, 2004*
Read Stability

\[ F_1(V_L) = V_L + c \]
\[ F_2(V_L - S) = V_L - S + c \]
\[ S = F_2(V_L - S) - F_1(V_L) \]
\[ \frac{\partial F_1(V_L)}{\partial V_L} - \frac{\partial F_2(V_L - SNM)}{\partial V_L} = 0 \]

Hold Stability

- Similar to Read Stability analysis without access transistor.
- PR must provide enough leakage to compensate for leakage in NMOS pull-down and access transistors.
Read Access

Must provide $\Delta V$ between the bit lines within maximum tolerable time limit.

- Limited to floating bit-line implementation with voltage sensing amplifiers.
- Sum BL currents and integrate.

$$T_{ACCESS} = \left| \int_{V_{in}}^{V_{dd}-\Delta V_{BL,R}} C_{BL,L} dV_{BL,L} \right| = \left| \int_{V_{in}}^{V_{dd}-\Delta V_{BL,R}} C_{BL,R} dV_{BL,R} \right|$$

Mukhopadhyay et al, 2004

Write Stability

Need $V_{WRITE} < V_{TRIP}$ for writability.

$$T_{WRITE} = \left| \int_{V_{in}}^{V_{TRIP}} \frac{C_{r}(V_{R})dV_{R}}{I_{OUT}(V_{R}) - I_{OUT,R}(V_{R})} \right| < T_{WL} \text{ for write stability.}$$

Mukhopadhyay et al, 2004
Vboxlo Trip Voltage Simulations for C027 High Density SRAM Cell

Large variation of small W/L SRAM transistor Vt/Vds translates to large variation in SRAM bit stability (SNM and Vtrip) and Iread

Mean Beta for Rnd 7 lots is ~1.4

Mean Beta re-fitted model is 1.64

<table>
<thead>
<tr>
<th>Transistor Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transistor Location</td>
<td>Left Ut</td>
<td>Right Ut</td>
<td>Left PO</td>
<td>Right PO</td>
<td>Left LL</td>
<td>Right LL</td>
</tr>
<tr>
<td>Bias for low Vtrip bit</td>
<td>92</td>
<td>101</td>
<td>92</td>
<td>110</td>
<td>122</td>
<td>101 Vaux</td>
</tr>
<tr>
<td>Mean Iddq at 0.9V</td>
<td>146</td>
<td>132</td>
<td>136</td>
<td>146</td>
<td>182</td>
<td>182 Vaux</td>
</tr>
<tr>
<td>EU ideal</td>
<td>30.1</td>
<td>25.0</td>
<td>33.5</td>
<td>32.9</td>
<td>22.0</td>
<td>26.1</td>
</tr>
<tr>
<td>low Vtrip bit (tail)</td>
<td>0.63</td>
<td>0.75</td>
<td>0.42</td>
<td>0.74</td>
<td>0.56</td>
<td>0.62</td>
</tr>
<tr>
<td>Pred. err beta</td>
<td>0.95</td>
<td>1.14</td>
<td>0.32</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

High Density SRAM Cell Leakage Trend
Power Management Introduced at 90nm Node

SRAM Iddq by Tech Node

- .NUMBER
- .LETTER
- NUM with PM
- .LET with PM

PM=Power Management
Possible Solutions for 6T SRAM Scaling

Stable operation
- tune well voltage to adjust Vts (adds 1 mask for deep N-well to isolate Nch)
- separate power rail for SRAM may be needed (bad for leakage)
- separate array and WL Vdd to optimize SNM/Vtrip over process/temp
- reduced transistor mismatch by reduced channel doping or reduce EOT

Leakage
- even lower retention voltage enabled by repairing weak retention bits
- data retention for selective blocks enabled by system level design
- high-k gate dielectric to reduce/control the growing gate leakage component

SRAM cell specific design rules
- this area seems more difficult but there are still some ideas
- process additions such as self-aligned contacts and ‘poly cut’ processes
- several biasing and sensing circuit ideas that would allow smaller Ids/W

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: 289-36 = 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

OMAP1612 - 2 die, In Production
1611
256MbitDDR

---

## EMBEDDED MEMORY TECHNOLOGY COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>SRAM</th>
<th>eFlash</th>
<th>eRAM</th>
<th>Flash (EIP)</th>
<th>eSRAM (EIP)</th>
<th>FeRAM</th>
<th>MRAM</th>
<th>PCRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Voltage</td>
<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>3V</td>
</tr>
<tr>
<td>Non-volatile</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</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>10^10</td>
<td>unlimited?</td>
<td>10^12</td>
</tr>
<tr>
<td>100nm 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>0.35</td>
<td>0.35</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>2004 cell size - in development (um2)</td>
<td>0.45</td>
<td>0.08 (0.33 for automotive)</td>
<td>0.11</td>
<td>0.04</td>
<td>0.54</td>
<td>0.8</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Mask adder to single gate oxide logic</td>
<td>0</td>
<td>5</td>
<td>4</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</td>
<td>high cost, limited endurance</td>
<td>high cost, high standby power, fast write</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</td>
<td>new materials, low cost, low standby power, high write power</td>
<td>new materials, low cost, low standby power, high write power</td>
</tr>
</tbody>
</table>
Flash Memory

Slides adapted from Ken Takeuchi, Toshiba

Flash EEPROM

Control gate

Floating gate

Thin tunneling oxide

\( n^+ \) source

programming

\( p \)-substrate

\( n^+ \) drain

Many other options ...
Basic Operations in a NOR Flash Memory—Erase

Basic Operations in a NOR Flash Memory—Write
Basic Operations in a NOR Flash Memory—Read

History of Flash Memories

<table>
<thead>
<tr>
<th>84</th>
<th>85</th>
<th>86</th>
<th>87</th>
<th>⋯ ⋯ ⋯</th>
<th>91</th>
<th>92</th>
<th>93</th>
<th>⋯</th>
<th>PRESENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAND-type</td>
<td>SanDisk-type</td>
<td>NAND-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACEE-type</td>
<td>AND-type</td>
<td>AND-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOR-type</td>
<td>NOR-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Split-gate-type</td>
<td>SST-type</td>
<td>SST-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DiNOR-type</td>
<td>DiNOR-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FLASH MEMORY Invention

File-Storage

Code-Storage
## Flash Memory Comparison - Code vs File Storage -

<table>
<thead>
<tr>
<th>Applications</th>
<th>Performance</th>
<th>Type of Flash memory</th>
</tr>
</thead>
</table>
| **Code Storage** | Program storage for - Cellular Phone - DVD - Set TOP Box BIOS for - PC and peripherals | Important:  
- High speed random access  
- Byte programming  
Acceptable:  
- Slow programming  
- Slow erasing | NOR  
- Intel / Sharp  
- AMD / Fujitsu / Toshiba  
DINOR  
- Mitsubishi |
| **File Storage** | Small form factor card for - Digital Still Camera - Silicon Audio - PDA ... etc Mass storage as - Silicon Disk Drive | Important:  
- High speed programming  
- High speed erasing  
- High speed serial read  
Acceptable:  
- Slow random access | NAND  
- Toshiba / Samsung  
AND  
- Hitachi  
- SanDisk: NOR |

**Important**:  
- Acceptable:  
- High speed random access  
- Byte programming  
- Slow programming  
- Slow erasing  
- High speed serial read  
- High speed erasing  
- High speed serial read  
- High speed random access  
- Slow random access

**Requirements for File Storage Memory**

- Low Bit Cost  
  <$\cdot.2$/MByte
- High Density  
  >256MByte
- High Speed Programming  
  >6MByte/sec  
  and Erasing  
  <3msec/block
- High Speed Serial Read
- Low Power Consumption
- Good Program/Erase Endurance  
  >1 million cycles
Cell Array Comparison

<table>
<thead>
<tr>
<th>NOR</th>
<th>SanDisk</th>
<th>AND</th>
<th>NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="NOR Cell Array" /></td>
<td><img src="image2" alt="SanDisk Cell Array" /></td>
<td><img src="image3" alt="AND Cell Array" /></td>
<td><img src="image4" alt="NAND Cell Array" /></td>
</tr>
</tbody>
</table>

- Simplest wiring
- Smallest area

NAND Cell Array (Top view)

- Select transistor
- Word lines
- Active area
- STI
- Bit line contact
- Source line contact
NAND Cell Array (Cross sectional view)

Word line

A → [ ] → A'

Bit line

A → [ ] → A'

Word line

Select gate

Source line

Cell Size Shrink by Self-Aligned STI

LOCOS_NAND : 6F²+α

Floating Gate

Word Line

Current

STI_NAND : 4F²+α

2F
NAND Flash Cell Size Trend

Start of Mass Production

 LOCOS

 SA-STI

 MLC

 Cell Size (um²)

 LOCOS SA-STI Multi Level Cell

 Control Gate ONO Floating Gate Tunnel Oxide LOCOS

 0.1um 0.13um 0.175um 0.25um

 0.10um

 Jan- '93 Jan- '94 Jan- '95 Jan- '96 Jan- '97 Jan- '98 Jan- '99 Jan- '00 Jan- '01 Jan- '02 Jan- '03 Jan- '04

 0.01 0.1 1 10

 55