Lecture 14

- Advanced Technologies on SRAM
  - Fundamentals of SRAM
  - State-of-the-Art SRAM Performance
  - FinFET-based SRAM Issues
  - SRAM Alternatives

Reading: multiple research articles (reference list at the end of this lecture)
Static Random Access Memory (SRAM)

Design:
- Typically consists of 6 MOSFETs (6T), including 2 cross-coupled inverters ($M_{1-4}$) and 2 access transistors ($M_{5,6}$).
- Wordlines and (differential signal) Bitlines

Operation:
- **Standby**: $WL = 0$
- **Read**: $WL = 1$, $BL = \overline{BL} = 1$ then $\overline{BL}$ discharged to 0 by $M_1$
- **Write**: $WL = 1$
  - $BL = 1$, $\overline{BL} = 0 \rightarrow$ content = 0
  - $BL = 0$, $\overline{BL} = 1 \rightarrow$ content = 1

Performance:
- SNM
- Read Current
- Write Noise Margin (WNM)
- Write Current
6T SRAM Design Trade-Offs: Read vs. Write

READ - OPTIMIZED SYSTEM

WRITE - OPTIMIZED SYSTEM

H. Pilo, IEDM Short Course (2006)
SRAM Technology Scaling Challenges

Degraded Electrostatics

Reduced $V_{DD}$

Gate Leakage

Soft-Errors

Variability

Cosmic-ray neutron

Alpha-ray

Nuclear reaction

Retention SNM @ 1V

Retention SNM @ 0.6V

Stress aggravates impact

11/17/2013

Nuo Xu

EE 290D, Fall 2013
Impacts of Performance Variability
- Why variability is extremely harmful to SRAM?

• SRAM always uses minimum transistor size, to reduce cell area.
  → Poor immunity to random and systematic variability
• Read vs. Write conflicts; Voltage loss on PG during Read
• $V_{TH}$ mismatch results in significantly reduced SNM.
  → Lowers SRAM cell yield, and limits $V_{DD}$ scaling

J. Luo, SSDM (2010)

• Immunity to short-channel effects, as well as performance variations is needed to achieve high SRAM cell yield.
Impact of Technology Flavors

- Different technology favors are implemented by $V_{TH}$ engineering
- Fast switching device has less Read and Write stability, as well as larger cell leakage (standby power).

X. Wang, ESSDERC, 2012
Circuit Techniques to Improve SRAM Stability

Dynamic $V_{DD}$

Floating $V_{DD}$

K. Zhang, ISSCC (2005)


Pulsing WL

Negative BL

M.E. Sinangil, ISSCC (2011)

H. Pilo, ISSCC (2011)
45nm 6T SRAM Design Rules

PD NMOS: \( W/L_g = 85/34 \text{ nm} \)
PG NMOS: \( W/L_g = 65/39 \text{ nm} \)
PU PMOS: \( W/L_g = 65/34 \text{ nm} \)
Contact Size: 66nm
Gate-Cont. space: 35nm
Well Isolation: 100nm
P+/P+ Isolation: 70nm
DT Ratio: 1.50
DT Ratio(W): 1.31

Cell Size: \( 0.72 \times 0.345 = 0.248 \text{ um}^2 \)

- **Imposed by OPC**, SRAM cell layout evolved from arbitrary shapes to predominantly straight lines and holes.
- **Imposed by Double Patterning**, poly-gates are oriented in the same direction

State-of-the-Art SRAM Cell Area

Intel 22nm Tri-Gate
- HDC 0.092 \( \mu \)m\(^2\)
- LVC 0.108 \( \mu \)m\(^2\)

IBM 22nm FinFET

TSMC 20nm FinFET
- 0.094 \( \mu \)m\(^2\)
- 0.076 \( \mu \)m\(^2\)
- 0.063 \( \mu \)m\(^2\)

IBM 22nm PD-SOI

Samsung 20nm Bulk

Planar Platforms

Year of publication [IEDM, VLSI]

911/17/2013
# State-of-the-Art SRAM Performance

<table>
<thead>
<tr>
<th>Company</th>
<th>Node (nm)</th>
<th>Area (um²)</th>
<th>$V_{DD,1}$ (V)</th>
<th>SNM$_1$ (V)</th>
<th>$V_{DD,2}$ (V)</th>
<th>SNM$_2$ (V)</th>
<th>Ref.s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel</td>
<td>32</td>
<td>0.171</td>
<td>~0.9</td>
<td>~0.2</td>
<td></td>
<td></td>
<td>IEDM 09</td>
</tr>
<tr>
<td>TSMC</td>
<td>32</td>
<td>0.15</td>
<td>1.0</td>
<td>0.22</td>
<td>0.8</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>32</td>
<td>0.157</td>
<td>0.9</td>
<td>0.213</td>
<td></td>
<td></td>
<td>IEDM 08</td>
</tr>
<tr>
<td>UMC</td>
<td>28</td>
<td>0.124</td>
<td>0.9</td>
<td>0.179</td>
<td>0.7</td>
<td>0.144</td>
<td>VLSI 11</td>
</tr>
<tr>
<td>Samsung</td>
<td>20</td>
<td>0.9</td>
<td>0.255</td>
<td></td>
<td></td>
<td></td>
<td>IEDM 11</td>
</tr>
<tr>
<td>IBM</td>
<td>22</td>
<td>0.1</td>
<td>0.9</td>
<td>0.22</td>
<td>0.7</td>
<td>0.148</td>
<td>IEDM 08</td>
</tr>
<tr>
<td></td>
<td>PDSOI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>22</td>
<td>0.094</td>
<td>0.8</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>VLSI 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.063</td>
<td>0.8</td>
<td>0.15</td>
<td>0.4</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>TSMC</td>
<td>20</td>
<td>0.1</td>
<td>0.85</td>
<td>0.45</td>
<td>0.09</td>
<td></td>
<td>IEDM 10</td>
</tr>
<tr>
<td>Intel</td>
<td>22</td>
<td>0.092</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td>ISSCC 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FinFET Advantages and Challenges in 6T SRAM Design

TEM & SEM of a bulk FinFET-based SRAM cell

SNM of (left) planar control & (right) FinFET SRAMs

- **Pros:**
  - Improved SS → Lower $V_{\text{TH}}$ at given $I_{\text{OFF}}$ → Higher $I_{\text{Read}}, I_{\text{write}}$
  - Reduced DIBL → Larger output resistance → larger SNM
  - Reduced performance variability (LER & RDF-induced) → larger SNM

- **Cons:**
  - Effective width quantization
  - $V_{\text{TH}}$ engineering is difficult
  - Limited design space, have to use different $L_g$, hindering cell area scaling...
Planar FET vs. FinFET SRAM Design

Planar FET

- PassGate
- Pull-Up
- Pull-Down

FinFET

- PassGate
- Pull-Down
- Pull-Up

Graphs showing Read/Write Margin vs. Cell Beta Ratio and # Fins on NPD.
Impact of Performance Variability on FinFET-based SNM

Butterfly curves subject to statistical variability sources. Left tail is close to Gaussian.

Fin number affects SNM distribution

X. Wang, ESSDERC, 2012
FinFET-based SRAM SNM Enhancement

- Metal-Gate Granularity (MGG) causes large SNM variation.
- 2-fin PD design helps to increase SRAM SNM while reduces $\sigma V_{TH}$, with the cost of 20% cell area increase.

X. Wang, ESSDERC, 2012
Intel’s 22nm SRAM: Tri-Gate Technology

- High Density & Low Leakage Cell (0.092 um²)
  - 1 Pull Up
  - 1 Pass Gate
  - 1 Pull Down

- Low Voltage Cell (0.108 um²)
  - 1 Pass Gate
  - 2 Pull Down

- High Speed Cell (0.130 um²)
  - 2 Pass Gate
  - 3 Pull Down

Different SRAM families are implemented by using different # of PD and PG fins.

4-5 times standby power reduction due to supreme SCE control.

Intel’s 22nm SRAM: Collapsing $V_{DD}$ Technique
Independent-Gate FinFET-based SRAM Design

- Enhanced Read margin
- Reduced WL capacitance
- $I_{\text{READ}}$ nearly unaffected
1. Single fin and larger fin heights used for PD NMOS, which reduces over 20% SRAM cell area compared to a 2-fin PD design.
2. Extra lithography steps to pattern fins to 2 heights: 20nm and 40nm
SRAM Alternatives: 8T Cell

- $N_0, N_1$ separates Read and Write, to lower operation voltage, and hence power consumption.
- Demonstrations on a 14KB 8T-SRAM based on Intel’s 22nm Tri-Gate technology: $V_{DD,\text{MIN}}$ is lowered by 130-270mV with 27-46% less power consumption.
SRAM Alternatives: SDRAM & nvSRAM

Renasas’ SDRAM

8T2R nvSRAM

Technology:
1. Poly TFET as PU PMOS
2. FEOL PD NMOS
3. BEOL MIM Capacitors
4. FEOL PG NMOS as well as the Access Transistors in 1T1R

Features:
1. Ultra-low power
2. Immunity to soft errors

Source: Renasas (2012)

Source: Renasas (2012)

P.-F. Chiu, VLSI (2010)

RRAM Switching Characteristics

P.-F. Chiu, VLSI (2010)
SRAM Alternatives: Context Memory

Memory Hierarchy

<table>
<thead>
<tr>
<th>Access Time (ns)</th>
<th>Feature Size (F²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAM</td>
<td>0.5 ~ 1</td>
</tr>
<tr>
<td>eDRAM</td>
<td>5</td>
</tr>
</tbody>
</table>

• Higher density yet lower power embedded memories are needed to bridge the performance gap between “logic” and “memory” circuits, and eventually the “Von Neumann Bottleneck”?

Intel’s embedded DRAM at 22nm

R. Brain, VLSI (2013)
References

SRAM Technology


References

SRAM Circuits

Intel’s Tri-Gate Platform