Challenges to LV e-RAMs

**RAM Cells**
- Extend low-voltage limitation to sub-1 V
- Degraded S/N
- Increased leakage
- Reduce cell size

**Peripheral Circuits**
- Reduce leakage
  - Increased $I_{STB}$ & $I_{ACT}$
- Reduce speed variation
  - Unreliable operations

1. It is governed by soft-error of cells, or S/N of cells and cell-relevant circuits.

2. As long as ECC is used, it is governed by S/N.

   ECC: Error Checking and Correcting circuit

3. S/N is determined by
   • Signal charge & signal voltage of cells,
   • **Flip-flop circuits** that DRAMs use for sense amps, while SRAMs use for cells themselves.
Comparisons of Flip-Flop Circuits

<table>
<thead>
<tr>
<th>Circuit</th>
<th>DRAM SA</th>
<th>SRAM Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standby</strong></td>
<td>Off with all nodes at $V_{DD}/2$</td>
<td>On with static of all MOSTs</td>
</tr>
<tr>
<td><strong>Active</strong></td>
<td>Dynamic sensing of $V_s$ (SN on, then SP on)</td>
<td>$i$ by ratio of $M_1$ &amp; $M_5$ Static of other MOSTs</td>
</tr>
<tr>
<td><strong>Margin</strong></td>
<td>Sensitive to $V_T$ &amp; $\Delta V_T$ of only two MOSTs, $M_1$ &amp; $M_2$</td>
<td>Sensitive to $V_T$ &amp; $\Delta V_T$ of all MOSTs $\rightarrow$ Narrow margin</td>
</tr>
</tbody>
</table>
Signal Charge $Q_S$ of RAM Cells

$Q_S \equiv$ Soft-Error $Q_{crt}$. The larger the $Q_S$, the smaller the SER. $Q_S$ decreases with device and voltage scaling.

**DRAM**

$Q_S = \frac{C_S V_{DD}}{2}$

$C_S$; Intentionally added, large, and needs to be gradually decreased with device scaling for maintaining large $V_{SIG}$.

**SRAM**

$Q_S = C_S V_{DD}$, $C_S = (C_1 + 2C_2)$

$C_1$, $C_2$; parasitic, small, and needs to be rapidly decreased with device scaling.

SER is always larger than for DRAM.
Signal Charge ($Q_s$) of RAM Cells

- $Q_s$ reduced with capacity due to $V_{dd}$ & device scaling
  Smaller $Q_s$ of SRAM cell
- SER depends on $Q_s$
  DRAM; decreases with memory capacity due to large intentionally-added $C_s$ & spatial scaling that reduces charge collection.
  SRAM; increases with memory capacity due to rapidly-decreasing parasitic $C_s$ despite spatial scaling.

Solutions:
- Increase in $C_s$ (SRAM cells)
- Uses of triple well, redundancy, ECC etc.

Y. Nakagome et al., IBM J. R&D, Vol.47, No.5/6, Sep./Nov. 2003
E. Ibe, The Svedberg Laboratory Workshop on Applied Physics, Uppsala, May 3, 2001
Error Checking & Correcting (ECC)

ECC word = 128 data bits + 8 check bits, FIT = $10^{-9}$/ hour

SER without ECC (FIT)

SER with ECC (FIT)

- no correction during 10-year period
- periodic correction (1 ECC word/7.8 μs)
- one upset/1 k hours

K. Itoh, Hitachi

M. Horiguchi et al., IEEE J. SSC, 23, p. 27, Feb. 1988
Minimum $V_{DD}$ ($V_{min}$) of RAMs

**DRAM**

Cell

$V_S > \delta V_T_i$

$V_S \equiv (V_{DD}/2) C_S/C_D$

$\therefore V_{min} = 2\delta V_T C_D/C_S$

$= 10\delta V_T$ ($C_D/C_S \approx 5$)

**SRAM**

Cell

$V_G = V_{DD}/2 - (V_{TO} + \delta V_T)$

$\therefore V_{min} = 2(V_{TO} + \delta V_T)$

If only cross-coupled nMOSTs determine the voltage margin during read,

$V_G = V_{DD} - V_{TO} - \delta V_T \geq 0$

$\therefore V_{min} = V_{TO} + \delta V_T$

$\delta V_T$: $V_T$-mismatch between paired MOSTs, $V_{TO}$: Average $V_T$

K. Itoh, Hitachi
Lowest Necessary \( V_{TO} \) for SA

- Signal \((-v_S)\) is amplified, so \(DL\) is discharged to \(A\). After that, \(\overline{DL}\) is gradually discharged by \(i_{\text{sub}}\) (M2).
- \(SP\) is on before \(\overline{DL}\) reaches \(C\).
- \(V_T (n\text{MOST}) \geq 0.2 \text{ V} @\text{extra., } 25^\circ\text{C}, \text{ if } t_S = 5 \text{ ns, } \Delta v = 50 \text{ mV}, C_D = 100 \text{ fF} @ 120^\circ\text{C}\)
Cross-coupled MOSTs need a high $V_T$ to ensure a small retention current through reducing $i_L$.

$V_{TO}$ is the average in a chip, because it is the average that determines retention current of the chip.

$V_{TO}$ must be quite high and unscalable.

\[
V_{DD} - V_{TO} - \delta V_T
\]

\[
V_{DD} + \delta V_T
\]

1-Mb array retention current (A)

Extrapolated $V_T = V_T(nA/\mu m) + 0.3 V$

- $T_j = 125 \degree C$
- $L_g = 0.1 \mu m$
- $W(Q_T) = 0.20 \mu m$
- $W(Q_D) = 0.28 \mu m$
- $W(Q_L) = 0.18 \mu m$

- High speed (0.49)
- Low power (0.71)

1. $V_T$ Variation ($\Delta V_T$) as source of $\delta V_T$
   - **Extrinsic** $\Delta V_T$ due to implant non-uniformities & $\Delta (L, W)$
   - **Intrinsic** $\Delta V_T$ due to random microscopic fluctuations of dopant atoms in the channel area.

2. $\Delta V_T$ & $\delta V_T$ increase with reducing MOST size even for a fixed generation.
   - $\sigma_{int} \propto \frac{1}{\sqrt{LW}}$
   - $\sigma(\delta V_T) \equiv \sqrt{2} \sigma_{int}$

3. $\Delta V_T$ & $\delta V_T$ increase with device scaling.

$\Delta V_T$ in a chip has no room in time & area to be compensated for.

---

**$\delta V_T$**

M. Yamaoka et al., Symp. VLSI Circuits 2004
## Larger $\delta V_T$ of SRAM cell

<table>
<thead>
<tr>
<th>Circuit</th>
<th>DRAM SA</th>
<th>SRAM Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>$\overline{DL}$</td>
<td>$\overline{DL}$</td>
</tr>
</tbody>
</table>

### MOS Size
- **DRAM SA**: 10 - 20 $F^2$
- **SRAM Cell**: 1.8 - 2.8 $F^2$

### Circuit Count in a chip
- **DRAM SA**: $M$ (64 - 1024)
- **Relaxed SA layout**: 64-1024

### $\delta V_T$
- **Small**
- **Large**

$M$: memory capacity
Maximum $\delta V_T$ in a chip

$r$: repairable percentage

**DRAM SA**

$LW = 20F^2$, $t_{ox} = 1.9$ nm
64 cells/SA

**SRAM Cell**

$LW = 2F^2$, $t_{ox} = 1.9$ nm

$F$ (nm) | Max. $\delta V_T$ (mV)
---|---
0 | 0
90 | 100
65 | 80
45 | 60
32 | 40

$r = 0$

$F$ (nm) | Max. $\delta V_T$ (mV)
---|---
0 | 0
90 | 300
65 | 200
45 | 100
32 | 50

$F$ (nm) | Max. $\delta V_T$ (mV)
---|---
0 | 0
32 | 100
64 | 80
128 | 60
256 | 40

$F$ (nm) | Max. $\delta V_T$ (mV)
---|---
0 | 0
32 | 330
64 | 200
128 | 100
256 | 50
$V_{\text{min}} (r = 0.1\%)$

**DRAM**

- $LW = 20F^2$, $t_{\text{OX}} = 1.9\text{ nm}$
- 64 cells/SA, $V_{\text{DD}}/2$ DL pre.
- $V_{T0} = 0.2\text{ V}$

**SRAM**

- $LW = 2F^2$, $t_{\text{OX}} = 1.9\text{ nm}$
- $V_{T0} = 0.49\text{ V (HS)}, 0.71\text{ V (LP)}$

*Actual $V_{\text{min}}$ determined by all MOSTs in a cell.*

**Graphs**

- **DRAM**: Graph showing $V_{\text{min}}$ vs. $F$ (nm) for different memory sizes (512Mb to 256Gb).
- **SRAM**: Graph showing $V_{\text{min}}$ vs. $F$ (nm) for different memory sizes (32Mb to 256Gb).
Approaches to LV SRAMs

1. Use ECC & Redundancy.
2. Minimize $\Delta V_T$ & $\delta V_T$.
   - Large cells with large MOSTs despite losing bit density
   - Symmetric cell layout
3. Stay at a high $V_{DD} \geq 1$ V due to its still large $\Delta V_T$ & $\delta V_T$ of bulk CMOS.
   Even so, power-supply control cells needed for small subthreshold current.
4. Extend low-$V_{DD}$ limitation to sub-1-V with FD-SOI.

M. Khellah, et al., ISSCC Dig., pp. 624-625, 2006

K. Itoh, Hitachi
Symmetric Layout for Small $\delta V_T$

Sources of extrinsic $\delta V_T$ in the conventional cell:
- Pattern deformation after processing
- Mask misalignment
- Local size fluctuation

Solution: Lithographical symmetric cell ("Thin" Cell)
- Reduced $\delta V_T$ by simple patterns suitable for OPC
- DLs shielded by power lines

Conv. (2 cells)  LS cell (2 cells)  Dotted area: after processing

poly gate diffused

OPC: optical proximity correction

Power-Supply Control Cells
for small subthreshold currents

Boosted Power Supply

\[ V_{DD} + \delta V_D \]

- High \( V_T \) to reduce \( i_L \)
- \( \delta V_D \) to offset a high \( V_T \) & \( \delta V_T \)

Low leak, wide margin & low power with low- \( V_{DD} \) DL.
Unscalable MOSTs needed.

Source-Line Driving

\[ V_{DD} \]

- Raised source during STB to reduce \( i_L \) with increased \( V_T \) of off-MOST

Reduced margin during STB by \( \delta V_S (>0.3 \text{ V}) \)
Source-Line Driving

Along with reduced DL voltage at active-standby transition

- **Sub-S backbias**
  - 90% reduction in subthreshold leakage

- **G-S backbias**
  - 100% reduction in subthreshold leakage

- **Electric-field relaxation**
  - 90% reduction in gate leakage & GI DL

```
<table>
<thead>
<tr>
<th></th>
<th>DL</th>
<th>VSS</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standby</td>
<td>1.0 V</td>
<td>0.5 V</td>
<td>1.0 V</td>
</tr>
<tr>
<td>Active</td>
<td>1.5 V</td>
<td>0.0 V</td>
<td>1.5 V</td>
</tr>
</tbody>
</table>
```

K. Itoh, Hitachi

K. Osada et al. ISSCC2003 Dig. pp. 302-303
Measured Retention Current of Cell

25°C

<table>
<thead>
<tr>
<th></th>
<th>Sub. + GI DL</th>
<th>Tunnel</th>
<th>PMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv.</td>
<td>48.5</td>
<td>46.5</td>
<td></td>
</tr>
<tr>
<td>NMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop.</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>17 fA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ V_T (\text{extrap.}) = 0.7 \text{ V(N)}, -1 \text{ V(P)} \]
\[ t_{ox} (\text{electrical}) = 3.7 \text{ nm} \]

90°C

<table>
<thead>
<tr>
<th></th>
<th>Sub. + GI DL</th>
<th>Tunnel</th>
<th>PMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv.</td>
<td>1182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop.</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102 fA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subthreshold current sensitive to temp.

- Successful Application
  1.5-V 27-ns 6.42 x 8.76 mm² 16-Mb using ECC with 3.2-ns/9.7% speed/area penalties.

- Limitations and Challenges
  (1) Leakage still large 1.6 μA for 16 Mb despite high \( V_T \), thick \( t_{ox} \), and S-driving.
  (2) Reduced \( Q_S \) in standby mode
  The cell power-supply decreases by the raised source voltage.
  Further low-\( V_{DD} \) operation may be hazardous, even if ECC is used.
Source-Line Driving

to reduce leakage & its variation while retaining the data

\[ V_{DD} \]

\[ V_{\text{min}_\text{STB}} \]

\[ \text{SRAM array} \]

\[ V_{\text{REF}} = V_{DD} - V_{\text{min}_\text{STB}} \]

\[ \Delta V_T \]

\[ \delta V_S \]

\[ V_{T0} \]

\[ V_{T_{\text{max}_\text{STB}}} \]

\[ V_{DD} \]

\[ V_{\text{min}_\text{STB}} \]: Min. \( V_{DD} \) to retain the data of all cells in the array.

Lower \( V_{T0} \) → Larger \( \delta V_S \) → Higher \( V_{T0} \) with deeper body bias (\( \Delta V_T \)) → Lower \( i_L \)

K. Itoh, Hitachi

M. Khellah, et al., ISSCC Dig. pp. 624-625, 2006
Double-Gate FD-SOI

- Small $\Delta V_T$ & negligible $\delta V_T$ (ultra-thin & lightly-doped channel)
- Adjustable $V_T \rightarrow$ multi- $V_T$
- Large $V_T$ change (wide-range well-bias control)
- Reduced SER & small $i_{pn}$
- Dynamic $V_T$ MOST (e.g., G-well connection)

K. Itoh, Hitachi

SRAM Cells with Dynamic-\( V_T\) MOSTs

to widen the voltage margin

\[\text{M}_1: \text{decreased } V_T\]
\[\text{M}_2: \text{increased } V_T\]
\[\text{M}_3: \text{increased } V_T\]
\[\text{M}_4: \text{decreased } V_T\]

Write margin improved with decreased \( V_T \) for driver/transfer MOSTs & increased \( V_T \) for load MOSTs.

M. Yamaoka et al., A-SSCC Dig. pp. 109-112, Nov. 2005

K. Itoh, Hitachi
Challenges to LV e-RAMs

RAM Cells
- Extend low-voltage limitation to sub-1 V
- Degraded S/N
- Increased leakage
- Reduce cell size

Peripheral Circuits
- Reduce leakage
  - Increased $I_{STB}$ & $I_{ACT}$
- Reduce speed variation
  - Unreliable operations

K. Itoh, Hitachi

Leakage Currents of Periphery

**Gate tunneling current** (\(i_G\))
- Insensitive to \(V_G\) & temp.
- Sensitive to \(t_{ox}\)

1/ 10 \(i_G\)-reduction with \(t_{ox}\)-increment of only 2-3 Å for SiO₂, while the same reduction with \(V_G\)-decrement of as much as 0.5 V.
Such a large \(V_G\) control in low- \(V_{DD}\) region is risky.

→ Device designers are responsible for the reduction. (High- \(k\))

**Subthreshold current** (\(i_L\))
- Insensitive to device structures
- Sensitive to \(V_G\), \(V_T\) & temp that can be controlled by circuits.

1/ 10 \(i_L\)-reduction with \(V_T\)-increment, or \(V_G\)-decrement of only 100 mV.

→ Circuit designers are responsible for the reduction.
Subthreshold Current ($i_L$) of Periphery

Features of RAM Periphery

1. Input-Predictable Logic
   Designers can prepare the schemes in advance.

2. Slow Cycle ($t_{RC} = 25, 60$ ns)
   Each circuit is active for only a short period within “long” cycle, enabling additional time for $i_L$-control.

3. Iterative-Circuit Blocks
   Major $i_L$ sources.
   All circuits in each block are inactive, except selected one.

4. Robust Circuits
   $i_L$-immune NAND dec.
   (w/o $i_L$-sensitive NOR dec.)

5. On-Chip Power Supplies
   $V_{DH}$ & $V_{BB}$ utilized for dual-static $V_T$. 

K. Itoh, Hitachi

Y. Nakagome et al., IBM J. R&D, Vol.47, No.5/6, Sep./Nov. 2003
i_L-Increase in Periphery

- At present, $V_T$ is still so high that $i_L$ is small in active mode, though $i_L$ dominates in standby mode.
- In the future, with further reducing $V_T$, $i_L$ will dominate even in active mode. Leakage reduction for active mode is the key.

$$i_{ACT} = i_{AC} + i_{DC}$$
$$i_{AC} = \sum C_j V_{DD} f$$
$$i_{DC} = \sum i_{Lk} \infty \sum W_k 10^{-V_T/s}$$
$$V_T = aV_{DD}$$
$$k \gg j$$

Trends in DRAM Peripheral Current

<table>
<thead>
<tr>
<th>Capacity (bits)</th>
<th>$V_{DD}$ (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16M</td>
<td>0.53</td>
</tr>
<tr>
<td>64M</td>
<td>0.40</td>
</tr>
<tr>
<td>256M</td>
<td>0.32</td>
</tr>
<tr>
<td>1G</td>
<td>0.24</td>
</tr>
<tr>
<td>4G</td>
<td>0.19</td>
</tr>
<tr>
<td>16G</td>
<td>0.16</td>
</tr>
<tr>
<td>64G</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Cycle time=180 ns
$T = 75^\circ C$, $S = 97$ mV/dec.
Basic Concept of $i_L$-Reduction

- Use high-$V_T$ MOST achieved statically or dynamically with $V_{Th} = V_{Tl} + \Delta V_T$
- For static high-$V_T$ $\Delta V_T$ by ion impla. or static-$V_{BB}$ application.
- For dynamic high-$V_T$ $\Delta V_T$ by dynamic back-biasing schemes G-S back-bias is best due to large $\Delta i_L / \Delta \delta$, applicable even to active mode.

Y. Nakagome et al., IBM J. R&D, Vol.47, No.5/6, Sep./Nov. 2003
# Three Practical Reduction Circuits

## Applicable even to Active Mode

<table>
<thead>
<tr>
<th>Switched-Source Imp. (G-S Self-Backbiasing)</th>
<th>![Switched-Source Imp. Diagram]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Switch utilizing internal power supply (G-S Offset Driving)</td>
<td>![Power Switch Diagram]</td>
</tr>
<tr>
<td>Dual-Static $V_T$ utilizing internal power supply</td>
<td>![Dual-Static $V_T$ Diagram]</td>
</tr>
</tbody>
</table>
**SSI (G-S Self-Backbias)**

**for fast \(i_L\)-control of input predictable logic**

No matter how large \(i_1\) is, it is confined to const. current \(i_2'\) with self-adjusting \(\delta\).

**Stacking effects (\(\delta = 0.2\) V)**
- G-S backbias of Q1 (1/100)
- Sub-S backbias of Q1 (1/1.5)
- DI BL effect of Q2 (1/2)

**Applicable even to active mode**
- Fast \(i_L\)-control capability with small \(\delta\) & \(C_L\) and self-reduction
- Small area penalty if applied to iterative circuit blocks
- Capability of confining to min. active circuitry

\[i_1' = i_2'\]

\[\therefore \delta = (S/\ln10) \ln(W_1/W_2)\]

Reduction Ratio \(\gamma = i_1'/i_1 = 10^{-\delta/S} = W_2/W_1\)
- Smaller \(i_L = i_2'\) & larger \(\delta\) with smaller \(W_2\)
- \(\gamma = 1\) (no reduction) for \(W_2 = W_1\)

Other secondary effects reduce \(i_L\).

(Sub-S backbias & DI BL)

---

Small Area Penalty with $SSI$ Sharing

**Inverter Chain**

- For 0-V input, $i_L$ flows from an n-MOST in each 0-V input inverter, and accumulates into $SSI$.
- $SSI$ confines to its constant current $\left( = i_0 W_2 10^{-V\tau0/S} \right)$.
- $W_2 \equiv w_1 \equiv \ldots \equiv w_n$ without speed penalty because each inverter switches at different timing.
- Area penalty is negligible with increasing the number of inverters because $W_2 \ll \Sigma w_i$.

Such is the case for p-MOST $SSI$. 

---

K. Itoh, Hitachi

During non-selected periods \( i_L (= i') \) flows from each circuit, and accumulates into \( SSI \).

\( SSI \) confines to its constant current \( (= i_0 W_2 10^{-\nu t/S}) \).

\( W_2 \approx W \) without speed penalty because only one MOST is activated with \( SSI \) on.

Area penalty is negligible with \( n \gg 1 \) because \( W_2 \ll nw \).
Confining to Minimum Active Circuitry

Partial activation of multi-divided block

Block (selected)

\[ i = i_0 \cdot W \cdot 10^{-V_{th} / S} \]

Subblock (selected)

\[ i_2' = i_0 \cdot W_2 \cdot 10^{-V_{th} / S} \equiv i, \quad W_2 \equiv w \]
Low- $V_T$ switch (Q) shuts off the supply of low- $V_T$ core during standby. A raised $V_{DH}$ needed to cut off Q with G-S backbias.

**Problems:**

1. If $V_{DH}$ generated by charge pump,
   - Unregulated floating $V_{DH}$
     For well-regulated $V_{DH}$,
     $C_2 V_{DH} f < C_P V_{DD} f_P$, $C_1 \gg C_P$, $C_1 \gg C_2$, level monitor.
   - Increased pump power
     For low pump power with keeping the $V_{DH}$ level,
     $C_P V_{DD} f_P$ & $C_2 V_{DH} f$ reduced.
     - Smaller $C_2$ & slower $f$

2. Area penalty by large Q
3. Slow recovery of internal power node (←)
Power Switch with Level Holder

Applicable even to active mode, if the switch itself operates fast enough. After evaluating the input, the output level continues to be held by high-$V_T$ holder without leakage. Otherwise

- Floating output discharges, causing a large $i_L$ at pMOST in the succeeding circuit, in which the switch is still on.
- Unnecessary discharging prevents the output from quick recovery.
Dual Static $V_T$

Useful for active & standby modes

Selective use of a high $V_T$ to
- Off-MOSTs during standby
- Non-critical path, while using low $V_T$ to critical path.
  
  $\Rightarrow$ Low $i_L$ & high speed chip

The reduction is not remarkable because $V_T$-difference must be small.

A large $V_T$-difference may cause a racing problem:
  a pulse-timing imbalance between $V_{TI}$- & $V_{TH}$-circuits.

$i_L$ reduced to $1/5$ for uniform use of $V_{TI}$ with assumptions:

$W_{total}$ (critical path) = $10\%$ of $W_{total}$ (chip),

$V_{TI} = 0.21$ V, $V_{TH} = 0.31$ V, $S = 0.1$ V/dec.
256-Mb DRAM (Standby)

K. Itoh, Hitachi

M. Hasegawa et al., ISSCC Dig. pp. 80-81, 1998.
Standby Current Reduction (256 Mb)

1.28 mA

Conv. Peri. Drivers Y-dec Array control

SSI

Dual static $V_T$

$T = 75^\circ C$

$S = 109 \text{ mV/dec. (PMOST)}$

89 mV/dec. (NMOST)

$V_T = 0.03 \text{ V (PMOST, 0.7 nA/\mu m)}$

-0.02 V (NMOST, 0.7 nA/\mu m)

$V_{DD} = 2.0 \text{ V, } V_{DH} = 3.8 \text{ V, } V_{BB} = -1 \text{ V}$

Proposed

Power down self-refresh

0.21 mA

40 $\mu$A with dual $V_T$

K. Itoh, Hitachi

M. Hasegawa et al., ISSCC Dig. pp. 80-81, 1998.
# 1-V 16-Gb DRAM

## Active Current Reduction

<table>
<thead>
<tr>
<th>Conv.</th>
<th>AC 75</th>
<th>DC</th>
<th>(subthreshold)</th>
<th>1105</th>
</tr>
</thead>
<tbody>
<tr>
<td>word drivers</td>
<td>695</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>decoder</td>
<td>209</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA drv.</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>others</td>
<td>132</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed</th>
<th>75 41</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSI</td>
<td>116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power switch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>holder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $V_T = -0.12 \text{ V (2nA/ \mu m)}$, $S = 97 \text{ mV/dec.}$, $75^\circ C$
- $V_{DH} = 1.75 \text{ V}$, $V_{DD} = 1 \text{ V}$, $t_{RC} = 180 \text{ ns}$

K. Itoh, Hitachi

T. Sakata et al., 1993 Symp. VLSI Circuits.
**0.6-V 16-Mb e-DRAM**

**Y Circuits**

- Some int. voltages are controlled to compensate for $\Delta (i_L, \tau)$ caused by $\Delta (V_T, T, V_{DD})$.
- In sleep mode, 0.3-V G-S b.b. reduces SA’s $i_L$.

![Diagram of Y Circuits with voltage levels and circuit components](image)

---

K. Itoh, Hitachi

0.6-V 16-Mb e-DRAM

X Circuits (Active)

K. Itoh, Hitachi

K. Hardee et al, ISSCC2004 Dig. p. 200
0.6-V 16-Mb e-DRAM

X Circuits (Sleep)


K. Hardee et. al, ISSCC2004 Dig. p. 200

K. Itoh, Hitachi
1.2-V 1-Mb e-SRAM

Multi-Bank Architecture

Active Mode
4-bank arch. with one-bank activation confines active circuitry to 1/4, and reduces
- AC power of cont. signals
- $i_L$ in inactive banks if $SSI$ is applied to WD & cells.

$SSI_1$: Small $\delta_1$ & $C_L \rightarrow 0.3$ ns
$SSI_2$: Drawbacks;
  Large $\delta_2$ (0.4 V) & $C_L \rightarrow 3$ ns
  Cell-supply reduced by $\delta_2$

Sleep Mode
Periphery off with power switch off $\rightarrow 3$ns

M. Yamaoka et. al., ISSCC2004 Dig. p. 494

K. Itoh, Hitachi
Leakage of 1-Mb e-SRAM (Active)

1.2 V, room temp.

- Conv. (300 MHz): 460 μA (1)
- $SSI_1$ (300 MHz): 350 μA (0.75)
- $SSI_1$ & $SSI_2$: 150 μA (0.33)

$V_T = 0.4/0.3$ V

K. Itoh, Hitachi
M. Yamaoka et al., ISSCC2004 Dig. p. 494
Challenges to LV e-RAMs

RAM Cells
- Extend low-voltage limitation to sub-1 V
- Degraded S/N
- Increased leakage
- Reduce cell size

Peripheral Circuits
- Reduce leakage
  - Increased $I_{STB}$ & $I_{ACT}$
- Reduce speed variation
  - Unreliable operations

Ever-increasing $\Delta V_T$, and rapidly-lowering gate-over drive with device scaling.

They enhance speed variation of periphery $\Delta \tau / \tau \propto \Delta V_T / (V_{DD} - V_T)$

Solutions

- For inter-die $\Delta \tau$, Compensation with $V_{BB}$ generator. e.g., Speed improvement by 63%
- For intra-die $\Delta \tau$, FD-SOI
K. Itoh, Hitachi

\( I_{DS} (M1) \) is a good indicator of \( i_L \) & speed. \( V_{GS} (M1) = \frac{V_{DD}}{2} \approx V_T \). \( V_D \) is compared to \( \frac{V_{DD}}{2} + \Delta \) and \( \frac{V_{DD}}{2} - \Delta \) to determine if \( V_{NB} \) should be increased or decreased.

For low \( V_T \) (fast process or high \( T_j \))
\[ V_D < \frac{V_{DD}}{2} - \Delta. \]
The lower OP activates PUMP, so \( V_{NB} \) starts to decrease and \( V_T \) is increased to compensate.

For high \( V_T \) (slow process or low \( T_j \))
\[ V_D > \frac{V_{DD}}{2} + \Delta. \]
The upper OP discharges M2-gate for driving the body, allowing \( V_T \) to be reduced and compensated for.

Such is the case for \( V_{DD} \).

K. Hardee et. al, ISSCC2004 Dig. p. 200
Intra-Die Speed Variation

Low-Power CMOS LSIs

\[ \frac{\tau(V_{T0}+\Delta V_T)}{\tau(V_{T0})} \]

- \( L/W = 1F/6F \)
- \( V_{T0} = 0.3 \text{ V}, \Delta V_T = \pm 3\sigma(V_T) \)
- \( \tau(V_T) \propto V_{DD}/(V_{DD}-V_T)^{1.25} \)

<table>
<thead>
<tr>
<th>( F (\text{nm}) )</th>
<th>90</th>
<th>65</th>
<th>45</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{DD} (V) ) (ITRS'03)</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>( \sigma(V_T) ) bulk (ratio)</td>
<td>1</td>
<td>1.29</td>
<td>1.89</td>
<td>2.51</td>
</tr>
<tr>
<td>( \sigma(V_T) ) SOI (ratio)</td>
<td>-</td>
<td>0.23</td>
<td>0.38</td>
<td>0.50</td>
</tr>
</tbody>
</table>

K. Itoh, Hitachi
Challenges to LV e-RAMs

**RAM Cells**
- Extend low-voltage limitation to sub-1 V
- Degraded S/N
- Increased leakage
- Reduce cell size

**Peripheral Circuits**
- Reduce leakage
  - Increased $I_{STB} \& I_{ACT}$
- Reduce speed variation
  - Unreliable operations

---

### RAM Cells (DRAM)

Short DL allows a small $C_S$ & simple $C_S$-structure with small $C_D$. Planar-$C_S$ cell might replace e-SRAM cells. $v_{\text{sig}} \approx \frac{C_s}{C_d} \cdot \frac{V_{\text{DD}}}{2}$.

In addition, short DL enables low-$V_{\text{DD}}$ fast operation.

<table>
<thead>
<tr>
<th>Proposed (ISSCC2005)</th>
<th>Conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Diagram" /></td>
<td><img src="image2.jpg" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cells/ DL</th>
<th>32</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_S$</td>
<td>5 fF (Ta$_2$O$_5$; MIM)</td>
<td>$\geq$ 15 fF (MIS)</td>
</tr>
<tr>
<td>Additional wire</td>
<td>No</td>
<td>local wire (M0)</td>
</tr>
<tr>
<td>Thermal budget</td>
<td>no impact on logic</td>
<td>intolerable impact</td>
</tr>
<tr>
<td>Cell $RC$ delay</td>
<td>W storage cont.</td>
<td></td>
</tr>
<tr>
<td>Co-salicided S/ D</td>
<td>Co-salicided S/ T</td>
<td>Non-metalized cell</td>
</tr>
<tr>
<td>Cell contact $R$</td>
<td>10 Ω</td>
<td>10 kΩ</td>
</tr>
</tbody>
</table>

K. Itoh, Hitachi

M. Iida et al., ISSCC2005 Dig. p. 460, M. Shirahata et al., ISSCC2005 Dig. p. 462
Cell Size Reduction (6-T SRAM Cell)

Stacked TFT SRAM Cells

- **Single-crystal TFT**
  The highest density cell \((25F^2)\) comparable to DRAM cells.
  1.8-V 61.1-mm\(^2\) 144-MHz 256-Mb SRAM.

- **Drawbacks as e-SRAMs**
  Sophisticated process, High-\(V_{DD}\) operation due to TFT PMOST of \(S = 140\) mV/dec., \(I_{DS} = 2/3\) of the bulk.

Load p-TFTs & transfer n-TFTs double-stacked over bulk driver n-MOSTs in different levels of layers.

K. Itoh, Hitachi

Cell-Size Comparisons

SAC for stand-alone DRAM
No SAC for others
Future Prospects for RAMs

6-T SRAM Cell: Due to high necessary $V_T$ & large $\Delta V_T$
• Not suitable for sub-1-V $V_{DD}$
• Continue to be used for a high $V_{DD}$ ($\geq 1$ V).
  Challenge; Small-$\Delta V_T$ MOSTs. TFT cells for stand-alone SRAMs.

1-T DRAM Cell: • Suitable even for sub-1-V $V_{DD}$.
  Challenges; Planar capacitors, Small-$\Delta V_T$ MOSTs.

Peripheral Circuits:
• Subthreshold-currents will be reduced sufficiently with existing techniques even for active mode.
• Speed variations will continue to be serious.
  Challenges; $V_{BB}$ control, Small-$\Delta V_T$ MOSTs.

Two Approaches:
  High- $V_{DD}$ bulk-CMOS for low-cost RAMs,
  Low- $V_{DD}$ FD-SOI for high-speed low-power RAMs.
Conclusion

1. I discussed challenges and trends in LV RAMs.
2. I reviewed state-of-the-art LV RAM circuits.
3. I gave prospects of RAMs with emphasis on further needs for
   • Ultra-low voltage RAM cells,
   • Advanced devices & circuits to reduce speed variations.