IEEE SSCS DLP, Colorado, June 15, 2006

Ultra-Low Voltage Nano-Scale Embedded RAMs

Kiyoo Itoh, Hitachi Ltd.

OUTLINE

- 1. Introduction
- 2. Trends & Challenges
- 3. Low-Voltage Limitation of RAMs
- 4. Leakage & Speed Variation of Peripheral Logic Circuits
- 5. Memory Cell Size
- **6. Future Prospects**
- 7. Conclusion

Trends in RAM Developments (R&D)



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 ISTB & IACT
- Reduce speed variation -Unreliable operations



Low-Voltage Limitation of RAMs

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



Signal Charge *Q_S* of RAM Cells

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

DRAM



$Q_S = C_S V_{DD}/2$

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



 $Q_{S} = C_{S}V_{DD}, C_{S} = (C_{1} + 2C_{2})$

*C*₁, *C*₂; parasitic, small, and rapidly decrease 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 *Qs*

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)



M. Horiguchi et al., IEEE J. SSC, 23, p. 27, Feb. 1988

Minimum V_{DD} (V_{min}) of RAMs



 $V_{S} > \delta V_{T_{I}} \qquad V_{G} = V_{DD}/2 - (V_{TO})$ $V_{S} \cong (V_{DD}/2)C_{S}/C_{D} \qquad + \delta V_{T} > 0$ $\therefore V_{min} = 2\delta V_{T}C_{D}/C_{S} \qquad \therefore V_{min} = 2(V_{TO} + \delta V_{T})$ $= 10\delta V_{T}(C_{D}/C_{S} \cong 5)$

If only cross-coupled nMOSTs determine the voltage margin during read, $V_G = V_{DD} - V_{TO} - \delta V_T \ge 0$, $\therefore V_{min} = V_{TO} + \delta V_T$

SRAM

VDD

Cell

 $V_{TO} + \delta V_T$

 δV_T : V_T -mismatch between paired MOSTs, V_{TO} : Average V_T

Lowest Necessary V_{T0} for SA



- Signal (-*v_s*) is amplified, so DL is discharged to A. After that, DL is gradually discharged by *isub* (M2).
- SP is on before DL reaches C.
- V_{TO} (nMOST) $\ge 0.2 \text{ V}$ @extra., 25°C, if $t_S = 5 \text{ ns}$, $\Delta v = 50 \text{ mV}$, $C_D = 100 \text{ fF}$ @ 120°C



Lowest Necessary V_{T0} for SRAM Cell



- •Cross-coupled MOSTs need a high V_T to ensure a small retention current through reducing *i*_L.
- *V*_{T0} 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_T -Mismatch (δV_T)

1. V_T Variation (ΔV_T) as source of δV_T

- Extrinsic ∠//₇ due to implant non-uniformities & ∠ (L, W)
- Intrinsic △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.

 $\sigma_{int} \propto 1/\sqrt{LW}$ $\sigma(\delta V_T) \cong \sqrt{2} \ \sigma_{int}$ 3.\Delta V_T & \delta V_T increase with device scaling.

 ΔV_T in a chip has no room in time & area to be compensated for.



M. Yamaoka et al., Symp. VLSI Circuits 2004

Larger δV_T of SRAM cell



K. Itoh, Hitachi

M: memory capacity

Maximum δV_T in a chip

r : repairable percentage





$V_{min} (r = 0.1 \%)$



^{*}Actual V_{min} determined by all MOSTs in a cell.

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
- Stay at a high V_{DD} (≥1 V) due to its still large ΔV_T & δV_T of bulk CMOS.
 Even so, power-supply control cells needed for st



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

control cells needed for small subthreshold current.
4. Extend low- V_{DD} limitation to sub-1-V with FD-SOI.

Symmetric Layout for Small δV_T

Sources of extrinsic δV_T in the conventional cell:

- Pattern deformation after processing
- Mask misalignment
 Local size fluctuation
- **Solution:** Lithographical symmetric cell ("Thin" Cell)
 - Reduced δV_T by simple patterns suitable for OPC
 - DLs shielded by power lines



K. Osada et al., IEEE J. SSC, vol. 36, No. 11, pp. 1738-1744, Nov. 2001. M. Kanda et al., Symp. VLSI Tech. Dig. Tech. Papers, pp. 13-14, June 2003. F. Arnaud et al., Symp. VLSI Tech. Dig. Tech. Papers, pp. 65-66, June 2003.

Power-Supply Control Cells

for small subthreshold currents

Boosted Power Supply

Source-Line Driving



- High *V*₇ to reduce *i*_L
- δV_D to offset a high V_T
 & δV_T
 Low leak, wide margin &

low power with low-*V_{DD}* DL. Unscalable MOSTs needed.

K. Itoh, Hitachi K. Itoh, ICICDT2005 Dig.



 Raised source during STB to reduce *i*_L with increased *V*_T of off-MOST

Reduced margin during STB by δV_S (>0.3 V)

Source-Line Driving

Along with reduced DL voltage at active-standby transition



Measured Retention Current of Cell





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

 Leakage still large
 μA for 16 Mb despite high V_T, thick t_{ox}, and S-driving.
 Reduced Q_S in standby mode
 The cell power-supply decreases
 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



 $V_{min}(STB)$: Min. V_{DD} to retain the data of all cells in the array.

Lower $V_{TO} \rightarrow \text{Larger } \delta V_S$ \rightarrow Higher V_{TO} with deeper body bias (ΔV_T) \rightarrow Lower i_L

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

Double-Gate FD-SOI



M. Yamaoka et al., Symp. VLSI Circuits 2004, R. Tsuchiya et al., IEDM2004 Dig. pp. 631-634

SRAM Cells with Dynamic- V7 MOSTs

to widen the voltage margin



M. Yamaoka et al., SOI Conf. Dig. pp. 109-111, Oct. 2004



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

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



Leakage Currents of Periphery

Gate tunneling current (*i_G*)

- Insensitive to V_G & temp.
- Sensitive to tox

1/10 i_G -reduction with t_{ox} -increment of only 2-3 Å for S_iO₂, 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. \rightarrow 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.



 $i_L \propto W 10^{(V_G - V_T)/S}$

 $S \sim 100 \text{mV/dec.}@100^{\circ}\text{C}$ 1/10 with $\Delta V_T = 100 \text{mV}$ or $\Delta V_G = -100 \text{mV}$

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, VLSI Memory Chip Design, Springer-Verlag, March 2001 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} i will dominate even in active mode. Leakage reduction for active mode is the key.



Basic Concept of *i*_L-Reduction

• Use high- *V*₇ MOST achieved statically or dynamically with $V_{Th} = V_{TI} + \Delta V_T$ • For static high- V_T ΔV_T by ion impla. or static- V_{BB} application. • For dynamic high- V_T ΔV_T by dynamic backbiasing schemes G-S back-bias is best due to large $\Delta i_L / \Delta \delta_i$ applicable even to active mode.



Three Practical Reduction Circuits

Applicable even to Active Mode



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 δ .

Stacking effects ($\delta = 0.2$ V)

- G-S backbias of Q1 (1/100)
- Sub-S backbias of Q1 (1/1.5)
- DIBL 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

 $i_{1'} = i_{2'}$ $\therefore \delta = (S/\ln 10) \ln(W_1/W_2)$

Reduction Ratio $\gamma = i_1' / i_1 = 10^{-\delta/S} = W_2 / W_1$

- Smaller i_L (= i_2 ') & larger δ with smaller W_2
- $\gamma = 1$ (no reduction) for $W_2 = W_1$ Other secondary effects reduce i_L . (Sub-S backbias & DIBL)

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 $(=i_0 W_2 10^{-V_T/S})$.
- $W_2 \cong W_1 \cong ... \cong W_n$ 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*.

Small Area Penalty with SSI Sharing

Iterative Circuit Block (non-selected/standby)

- During non-selected periods
 *i*_L (= *i*') flows from each circuit,
 and accumulates into *SSI*.
- SSI confines to its constant current $(=i_0 W_2 10^{-V_T/S})$.
- W₂ ≅ w without speed penalty because only one MOST is activated with SSI on.
- Area penalty is negligible with n » 1 because W₂ « nw.

Confining to Minimum Active Circuitry Partial activation of multi-divided block subblock subblocks block (selected) (non-selected) (selected) #1 #*m*-1 V_{DD} – activated VDDiVDD n/m • • • n 0-AVDD Q $1,023i \rightarrow 134i$ for *n/m* = 1024/8 \bigcirc N_2 W_2 $\mathbf{Q}_{\mathbf{2}}$ V_{DD}-SSI (n/m-1)i≅ (*m*-1)*i* (*n*-1)*i* $i = i_0 w 10^{-V_T/S}$ $i_2' = i_0 W_2 10^{-V_T/S} \cong i, W_2 \cong W$

K. Itoh, Hitachi K. Itoh, VLSI Memory Chip Design, Springer-Verlag, March 2001

Power Switch (G-S Offset Drive)

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_1} $C_2V_{DH}f < C_PV_{DD}f_P$, $C_1 \gg C_P$, $C_1 \gg C_2$, level monitor.
- Increased pump power For low pump power with keeping the V_{DH} level, C_PV_{DD}f_P & C₂V_{DH}f reduced.
 Smaller C₂ & 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.

K. Itoh, Hitachi K. Itoh, VLSI Memory Chip Design, Springer-Verlag, March 2001

Dual Static V_T

*i*_L reduced to 1/5 for uniform use of $V_{T/}$ with assumptions: W_{total} (critical path)=10% of W_{total} (chip), $V_{T/}$ = 0.21 V, V_{Th} = 0.31 V, S = 0.1 V/dec.

Chip

256-Mb DRAM (Standby)

Standby Current Reduction (256 Mb)

K. Itoh, Hitachi M. Hasegawa et al., ISSCC Dig. pp. 80-81, 1998.

1-V 16-Gb DRAM

Active Current Reduction

0.6-V 16-Mb e-DRAM

Y Circuits 0.6 V **External** DL 0.3 V V_{DD} (0.6 V) $V_{DD I/O}$ (2.5 V) Some int. voltages are DL controlled to compensate DL DL on-chip converters for Δ (*i*_L, τ) caused by 256 $\Delta (V_{T_1} T_1 V_{DD}).$: 40 fF *V_{NSB}* 3 V 0.9 V **WLs** In sleep mode, 0.3-V *V_{PSB}* -0.3 V G-S b.b. reduces SA's i_L . V_{NB} V_{PB} PD .6 9 VDD 0.6 **VSP 0.9** 0.9 SSI2 $\rightarrow V_{PB}$ **P-switch** PG ⊣ → Vnn $0.6 \text{ V}-\delta$ $\rightarrow V_{PSB}$ $\rightarrow V_{PB}$ 0.6 PD 0 ND SA **SEN** $0.2 - V V_{T}$ 0 $\leftarrow V_{NSB}$ ÷ VNB $\leftarrow V_{NB}$ P-switch NG -0.3SSI1 777 -0.3 VSN ND P-switch driver (inverter chain)

K. Itoh, Hitachi

K. Hardee et. al, ISSCC2004 Dig. p. 200 K. Itoh et al., CICC 2004 Dig. Tech. Papers, pp. 339-344, Oct. 2004

0.6-V 16-Mb e-DRAM

X Circuits (Active)

0.6-V 16-Mb e-DRAM

X Circuits (Sleep)

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. SSI₁: Small $\delta_1 \& C_L \rightarrow 0.3$ ns **SSI**₂: Drawbacks; Large δ_2 (0.4 V) & $C_L \rightarrow 3$ ns Cell-supply reduced by δ_2 **Sleep Mode** Periphery off with power switch off \rightarrow 3ns

M. Yamaoka et. al., ISSCC2004 Dig. p. 494 K. Itoh et al., CICC 2004 Dig. Tech. Papers, pp. 339-344, Oct. 2004

Leakage of 1-Mb e-SRAM (Active)

1.2 V, room temp.

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

Speed Variation

 \blacksquare Ever-increasing $\Delta V_{T_{I}}$ 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 $\Delta \tau_{i}$ **Compensation with** V_{BB} generator. e.g., Speed improvement by 63% • For intra-die $\Delta \tau_{i}$ FD-SOI

V_{BB} Generator for NMOS Body

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

e.g.,

For low V_T (fast process or high T_j) $V_D < V_{DD}/2 - \Delta$.

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) $V_D > V_{DD}/2 + \Delta$.

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} .

Intra-Die Speed Variation

Low-Power CMOS LSIs

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 ISTB & IACT
- 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. $V_{sig} \cong C_S / C_D \cdot V_{DD} / 2$. In addition, short DL enables low- V_{DD} fast operation.

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 (25 F²) comparable to DRAM cells.
 1.8-V 61.1-mm² 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.

Cell-Size Comparisons

Future Prospects for RAMs

6-T SRAM Cell: Due to high necessary V_T & large ΔV_T •Not suitable for sub-1-V V_{DD_1}

- Continue to be used for a high V_{DD} ($\geq 1 \text{ V}$). Challenge; Small- ΔV_T MOSTS. TFT cells for stand-alone SRAMs.
- **1-T DRAM Cell:** Suitable even for sub-1-V V_{DD} . Challenges; Planar capacitors, Small- ΔV_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- ΔV_T MOSTs.
- **Two Approaches:**

High- VDD 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.