2008 CMOS ET, August 5, 2008

Impact of Variability on Voltage Scaling Limitations of Nano-Scale CMOS LSIs

Kiyoo Itoh
Hitachi Ltd.

Challenge of Nano-Scale CMOS LSIs

Reduce V_{DD} and thus V_{min} to sub-1 V.

1. To reduce V_{DD}, both V_{T0} & ΔV_T must be scaled down with the same factor as that for V_{DD}.
   V_{T0}: Lowest necessary V_T, ΔV_T : V_T-variation.
2. However,
   - **V_{T0} is un-scalable.** It must be constant and high to suppress i_{SUB},
   - **ΔV_T is un-scalable.** It increases with MOS scaling, causing larger speed-variation ($\Delta\tau$).
     To offset $\Delta\tau$, V_{DD} must be increased with MOS scaling, but causing increased Pd.
3. Such un-scalable parameters prevent V_{DD} and V_{min} from down-scaling.

Relationship between V_{DD} and V_{min}

Actual $V_{DD} > V_{min}$.

$$V_{DD} = V_{min} + V_1(V_T) + V_2(ps) + V_3(\tau).$$

V_{min} : Min. functional V_{DD} determined by V_{T0} and intrinsic ΔV_T (randam dopant fluctuations),

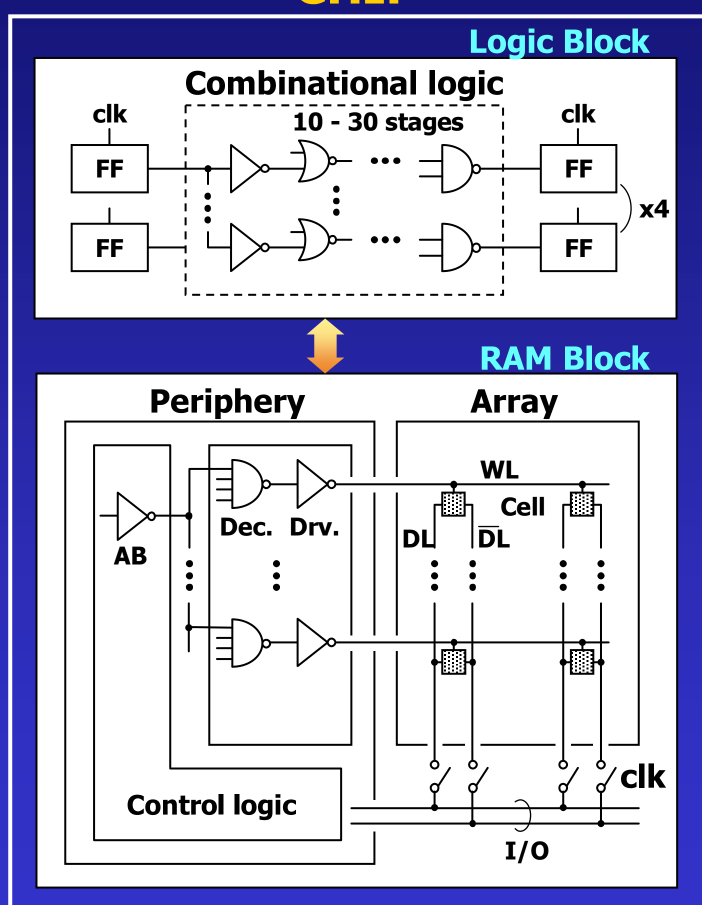
$V_1(V_T)$: To cope with other extrinsic ΔV_Ts,

$V_2(ps)$: To cope with power supply droop,

$V_3(\tau)$: To meet speed target.

- V_{min} is almost intrinsic. $V_1(V_T)$, $V_2(ps)$, and $V_3(\tau)$ depend on quality of process and design targets.
- $V_{min} \gg V_1(V_T)$, $V_2(ps)$, and $V_3(\tau)$ in nano-scale LSIs.
- V_{min} must be reduced as much as possible for low-V_{DD} operations.

The V_{min} of CMOS LSIs



1. Equal to the highest of the three V_{min} values;
   V_{min} (logic block),
   V_{min} (SRAM block),
   V_{min} (DRAM block).

- Each V_{min} depends on V_{T0} & ΔV_{Tmax} of the block.
- V_{T0} determined by subthreshold current spec. Almost constant.
- ΔV_{Tmax} determined by random ΔV_T of MOSFET & circuit count in the block.

2. Assumptions for calculating each V_{min}
   V_{min} is determined

- by CMOS inverter for logic block,
- by 6-T SRAM cell for SRAM block,
- by DRAM sense amp for DRAM block, if a twin cell and V_{DD} DL-precharge are used.

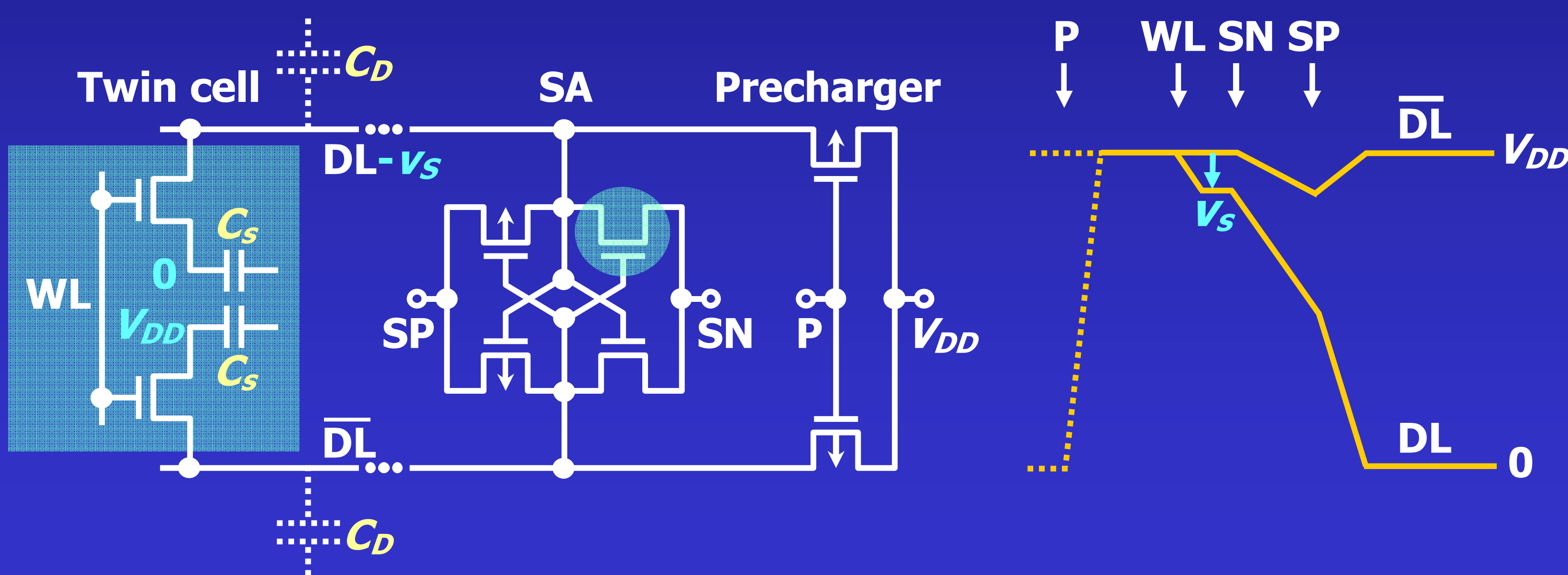
K. Itoh, Hitachi

2-T DRAM Cell

Ideal low-voltage cell, if combined with V_{DD} DL-precharging.

- The well-balanced structure minimizes noise with competitive cell size despite 2-T cell (< 30F^2).
- V_{DD} precharging halves V_{min} of conv. half-V_{DD} precharging.

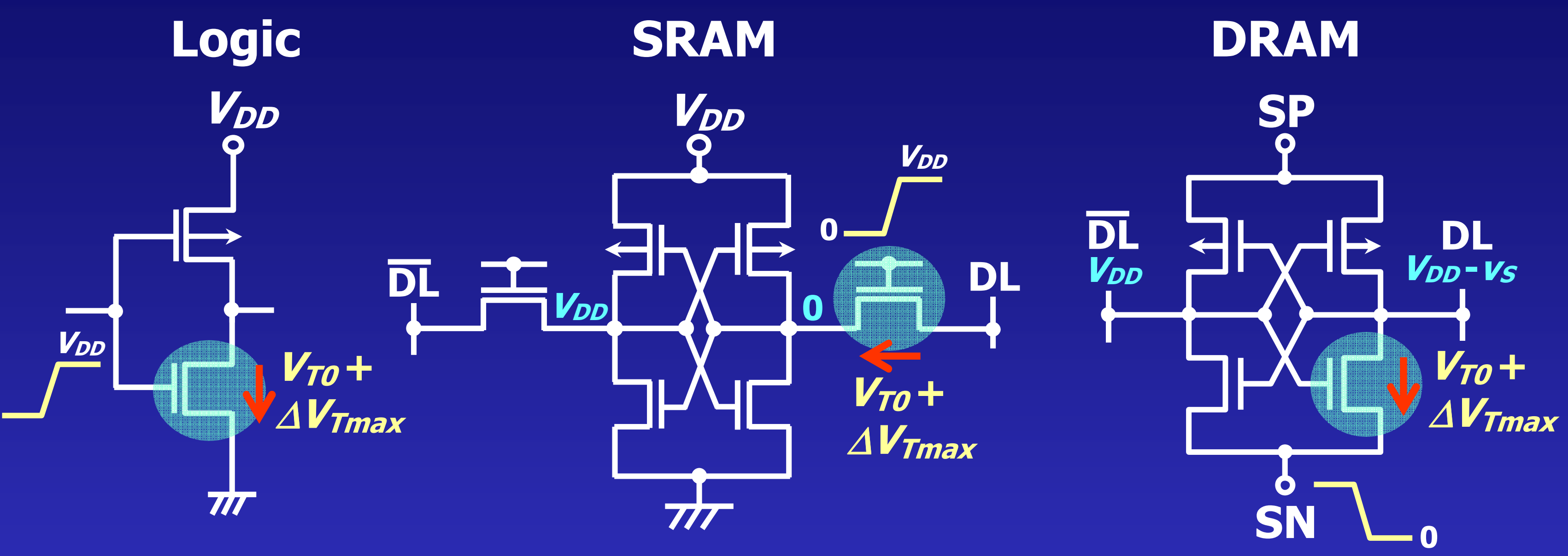
If C_S is large, V_{min} (DRAM block) is determined by nMOS SA.



K. Itoh, Hitachi

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

Definition of V_{min}



$\tau(V_T) \propto V_{DD}/(V_{DD}-V_T)^{1.2}$

$\Delta\tau = \tau(V_{T0}+\Delta V_{Tmax})/\tau(V_{T0})$

$= \{(V_{DD}-V_{T0})/(V_{DD}-V_{T0}-\Delta V_{Tmax})\}^{1.2}$

V_{T0}: The lowest necessary av. V_T for a tolerable leakage.

ΔV_{Tmax}: Max. V_T variation of the block.

$V_{min} = V_{DD}$ for a fixed $\Delta\tau$

$= V_{T0} + (1+\alpha)\,\Delta V_{Tmax}$

$\alpha = 1/(\Delta\tau^{1/1.2} - 1)$

6.1 ($\Delta\tau = 1.2$),

3.1 (1.4),

2.1 (1.6).

K. Itoh, Hitachi

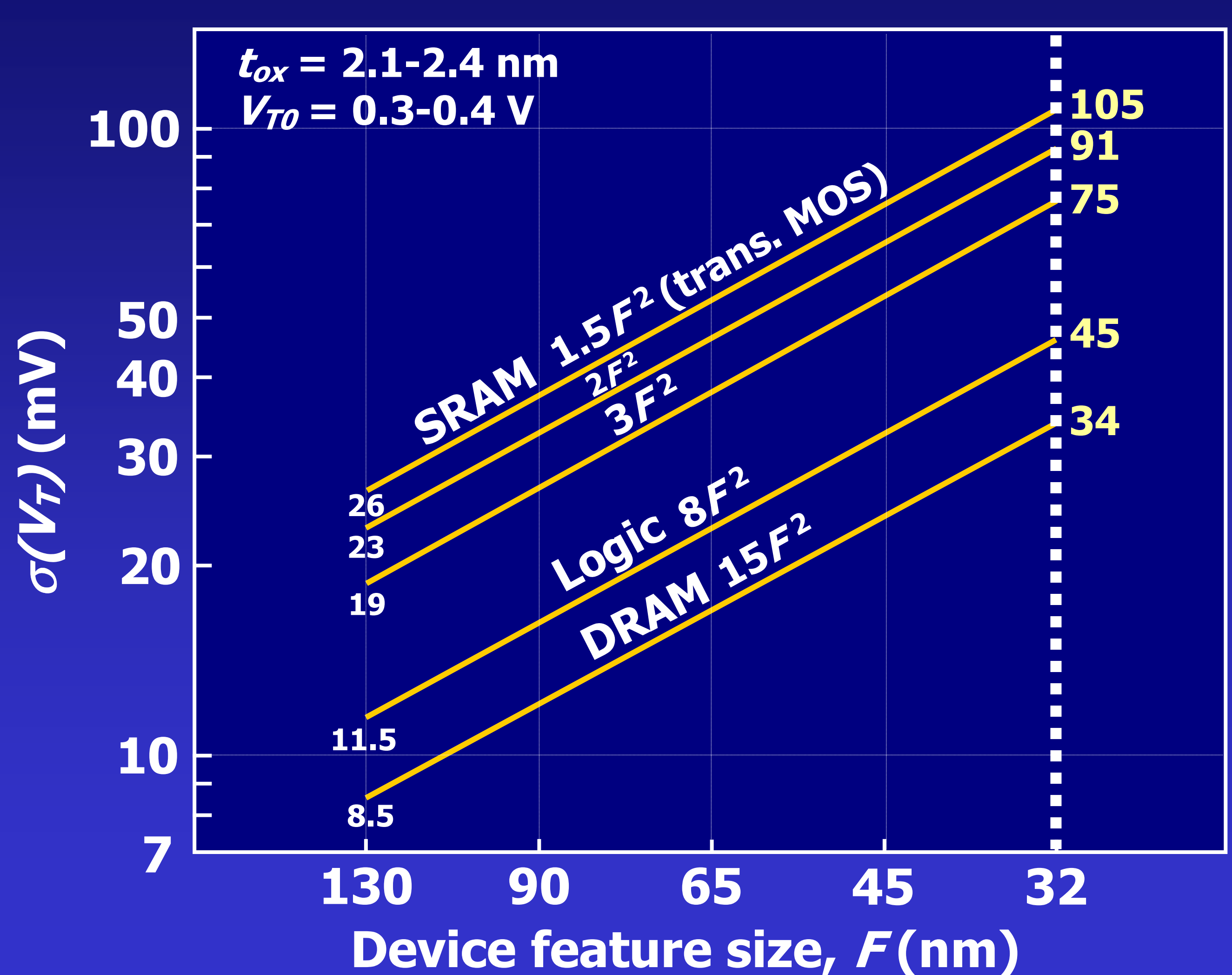
Assumptions

$$V_{min} = V_{T0} + (1 + \alpha)\,\Delta V_{Tmax}$$

1. V_{T0} = 0.3 V (logic), 0.4 V (SRAM), 0.2 V (DRAM).
   The highest V_{T0} of SRAM is due to many cells and severe spec. of i_{RET}.
2. $\Delta\tau = 1.6 \rightarrow \alpha = 2.1$.
3. $\Delta V_{Tmax} = m\sigma(V_T)$.
   - m depends on circuit count and repair for RAMs.
     3 (logic), ~6 (SRAM), ~5.5 (DRAM) w.o. repair
     3 (logic), ~3.3 (SRAM), ~2.9 (DRAM) w. repair
   - $\sigma(V_T) = A_{vt} / \sqrt{LW}$.
     A_{vt} depends on material & structure for MOSFETs.
     A_{vt} = 4.2 mVμm for poly-Si gate bulk MOSFET with
     V_{T0} = 0.3-0.4 V, t_{OX} = 2.1-2.4 nm.

$\sigma(V_T)$ vs. F for Poly-Si MOSFETs

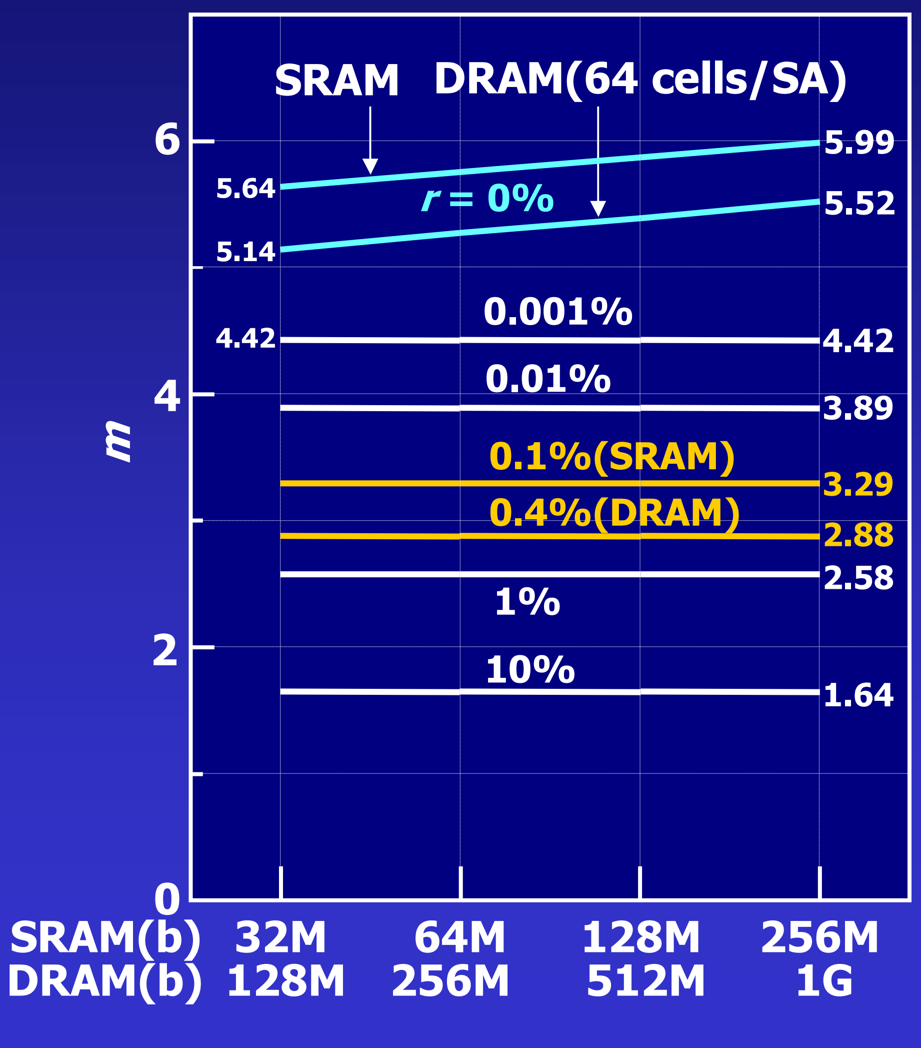
$\sigma(V_T) = A_{vt} / \sqrt{LW}$, A_{vt} = 4.2 mVμm

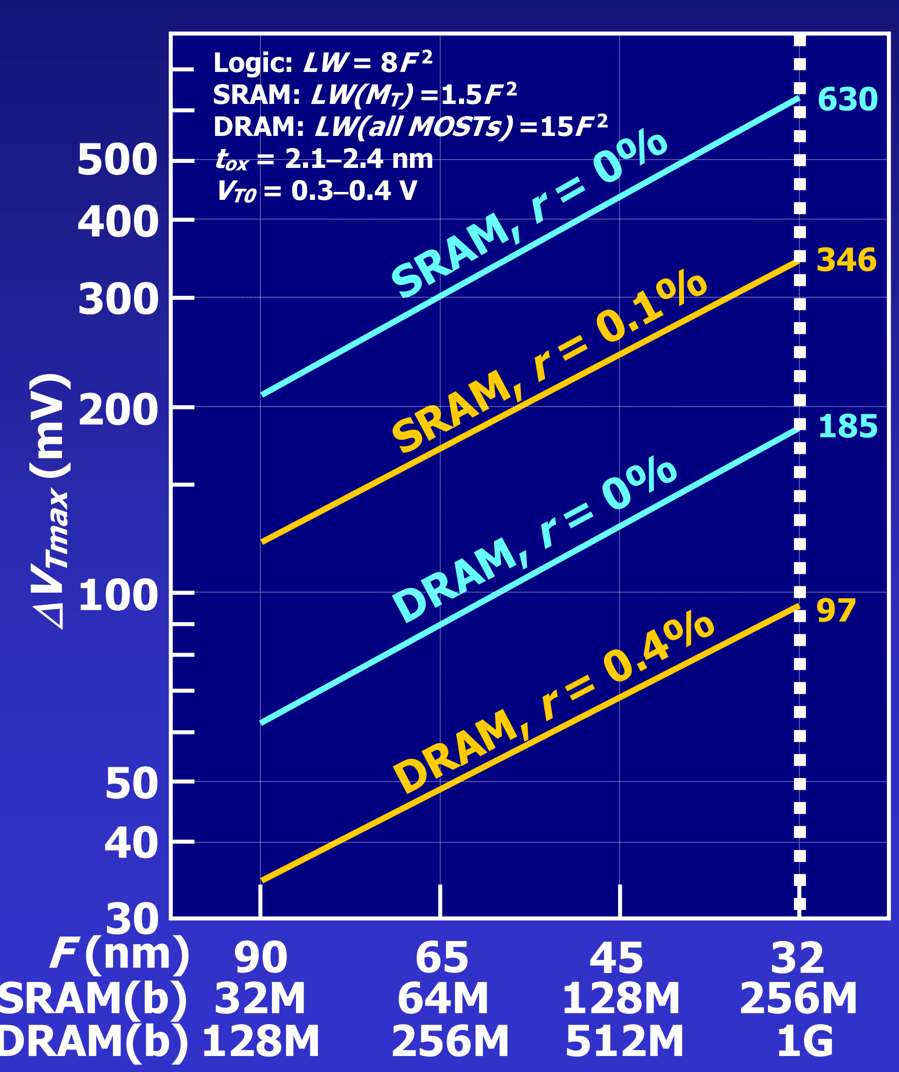


K. Itoh, Hitachi

Repair to halve ΔV_{Tmax} $(=m\sigma(V_T))$

with almost halving *m*.

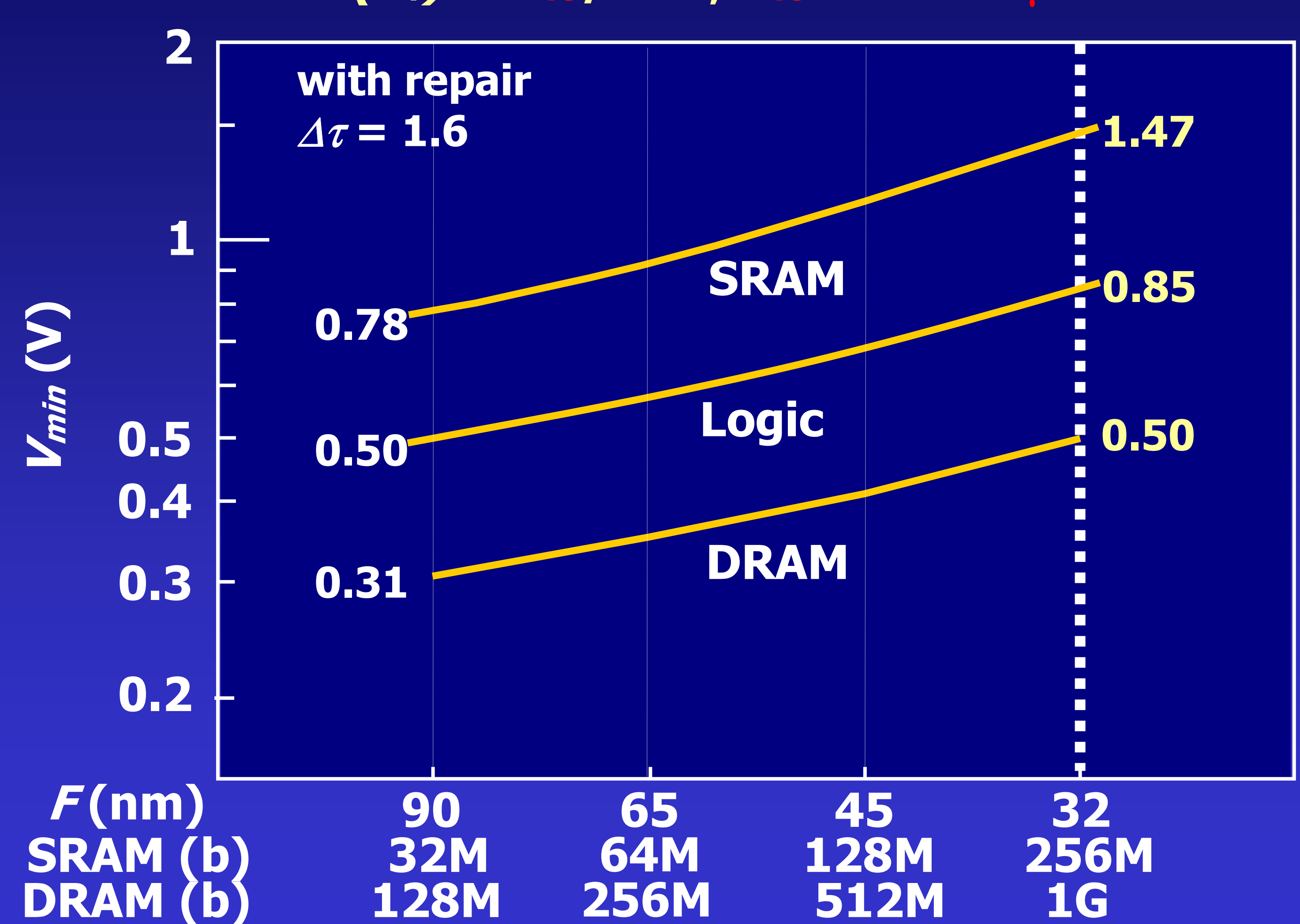




K. Itoh, Hitachi

V_{min} vs. F for Poly-Si MOSFETs

$\sigma(V_T) = A_{vt} / \sqrt{LW}$, $A_{vt} = 4.2$ mVμm



K. Itoh, Hitachi

V_{min} vs. F for Various A_{vt}s

poly-Si

A_{vt} = 4.2

A_{vt} = 2.1

A_{vt} = 1.5

A_{vt} =1.05

V_{min} (V)

2
1
0.5
0.4
0.3
0.2

SRAM 0.78 1.47
Logic 0.50 0.85
DRAM 0.31 0.50

SRAM 0.58 0.93
Logic 0.40 0.58
DRAM 0.25 0.35

SRAM 0.50 0.67
Logic 0.35 0.44
DRAM 0.23 0.27

F (nm)	90	65	45	32
SRAM (b)	32M	64M	128M	256M
DRAM (b)	128M	256M	512M	1G

K. Itoh, Hitachi

State-Of-The-Art 45-nm MOSFETs

$$\sigma(V_T) = A_{vt} / \sqrt{LW} \propto t_{OX}(inv)\, N_{SUB}^{0.25} / \sqrt{LW}$$

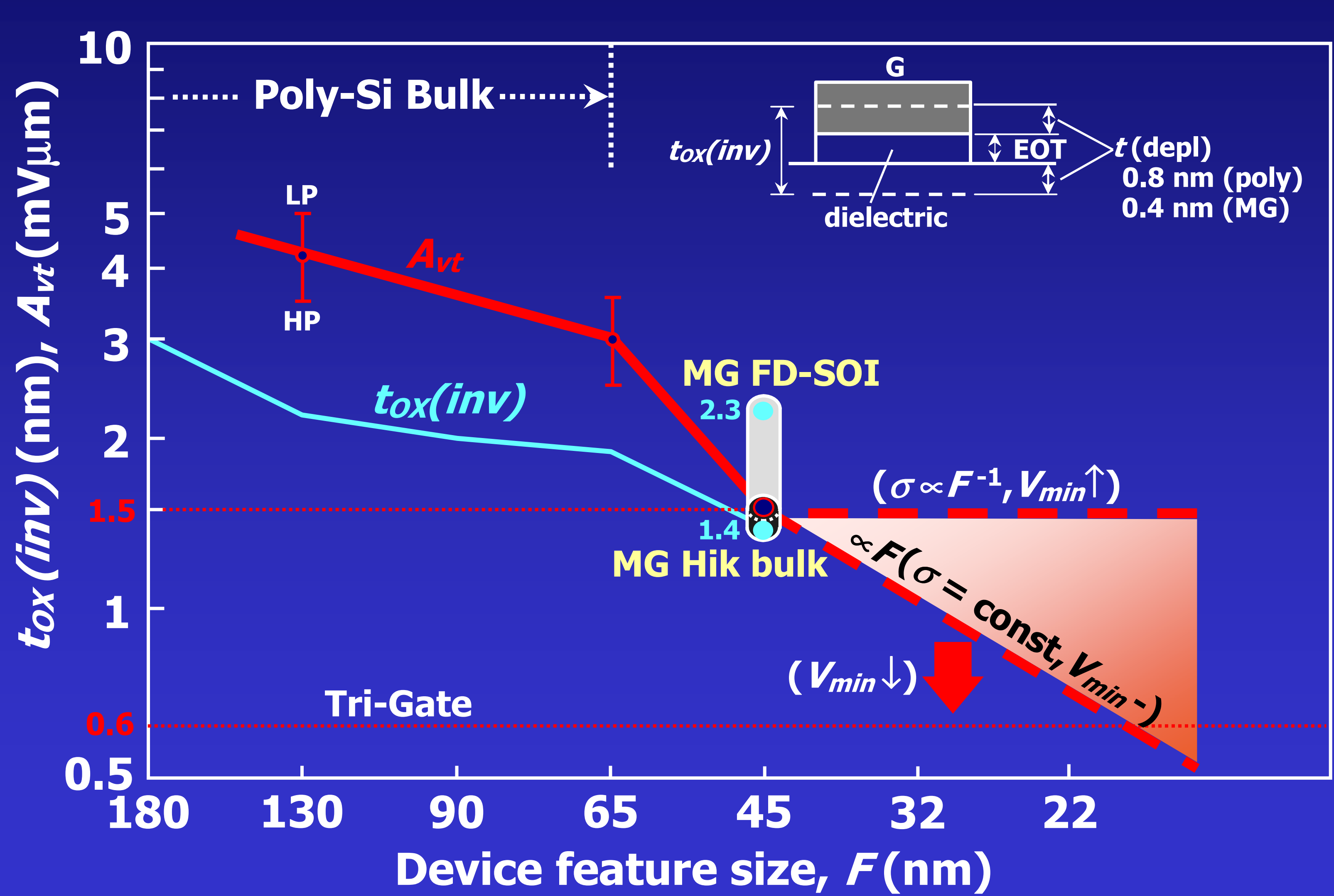
	MG Hi-k Bulk (HP)	MG DG-FD SOI (LP)
Structure	Metal, High-k, SiGe, SiGe, Sub.	SOI Channel, G, S, D, BOX, sub., 50nm
A_{vt}	1.5 (mVμm)	1.5 (mVμm) expected
$t_{OX}(inv)$	1.4 nm	2.3 nm
EOT	1.0 nm	1.9 nm
N_{SUB} (ratio)	7.2	1
I_{on}/ I_{off} (nMOS)	1.36 mA/μm@100 nA/μm (1 V)	0.55 mA/μm@20 pA/μm (1.2 V)
Gate dielectric	Hf based	S_iON
Others	Strained S_i channel N_i salicidation	Thin BOX (10 nm) N_i salicidation

K. Itoh, Hitachi

[1] K. J. Kuhn, IEDM2007 Dig. p. 471, [2] K. Mistry, IEDM2007 Dig. p. 247, [3] R. Tsuchiya, IEDM2007 Dig. p. 475

Trends in $t_{OX}(inv)$ and A_{vt}

$$\sigma(V_T) = A_{vt}/\sqrt{LW} \propto A_{vt}F^{-1} \propto t_{OX}(inv)\,N_{SUB}^{0.25}/\sqrt{LW}$$



K. Itoh, Hitachi

Pros and Cons of HiK MG MOSFETs

	Bulk	FD-SOI (FinFET)	FD-SOI (D-G)
Structure			
ΔV_T	large (N_{sub})	small	small
SER	large (depl.)	small	small
i_{pn}	large $\rightarrow \Delta V_{BB}$	small	small
Comp. for inter-die ΔV_T	possible only if ΔV_{BB} is small	difficult (no V_{BB})	easy by V_{BB}
Multi-V_T	easy	difficult (FD$\rightarrow$PD)	easy by impla.
Drive current	small	large (height↑)	small

K. Itoh, Hitachi

Future Prospects and Conclusion

Variability will continue to give voltage scaling limitations of CMOS LSIs, calling for technologies to reduce it.

- **Poly-Si bulk prevents CMOS LSIs from sub-1-V operations due to the ever-larger ΔV_T. SRAM is problematic most. Thus, in the 32-nm and beyond, high-*k* MG and/or FD-SOI will be needed for sub-1-V LSIs.**
- **For RAMs and even for logic, repair will be vital to reduce V_{min}.**
- **DRAM might replace SRAM due to its lower V_{min} and higher density.**