

Technology Scaling and Roadmap for 22nm CMOS and beyond

May 1, 2009

@University College of Dublin

Hiroshi Iwai

Tokyo Institute of Technology

Outline

1. Scaling

2. ITRS Roadmap

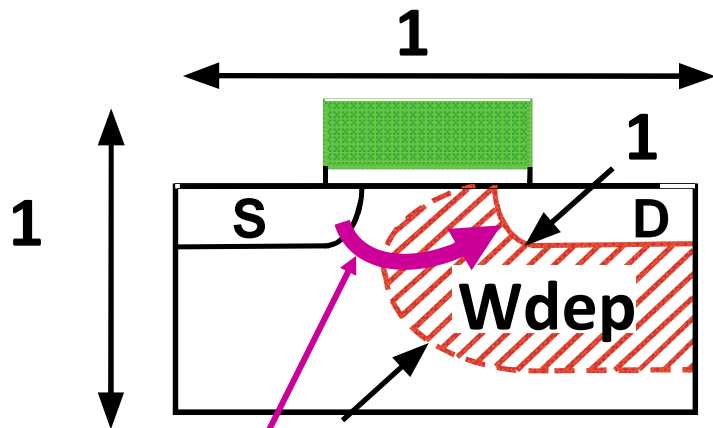
3. Voltage Scaling/ Low Power and Leakage

4. SRAM Cell Scaling

5. Roadmap for further future
as a personal view

1. Scaling

Scaling Method: by R. Dennard in 1974

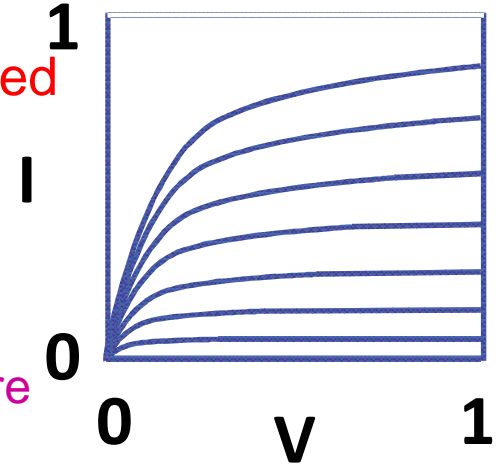


Wdep: Space Charge Region (or Depletion Region) Width

Wdep has to be suppressed
Otherwise, large leakage
between S and D

Leakage current

Potential in space charge region is high, and thus, electrons in source are attracted to the space charge region.



K=0.7
for
example

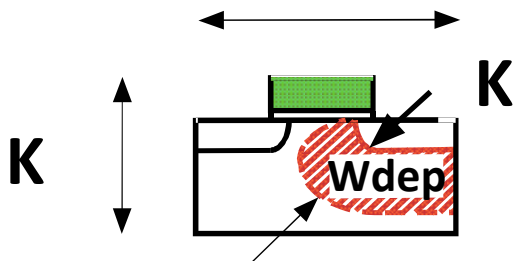


$$X, Y, Z : K, \quad V : K, \quad Na : 1/K$$

By the scaling, Wdep is suppressed in proportion,
and thus, leakage can be suppressed.

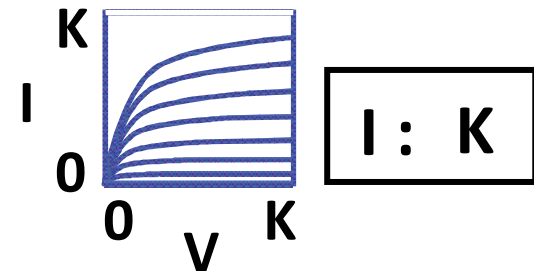
K

→ Good scaled I-V characteristics



$$W_{dep} \propto \sqrt{V / Na}$$

: K



$$I : K$$

Downscaling merit: Beautiful!

Geometry & Supply voltage	L_g, W_g T_{ox}, V_{dd}	K	Scaling K : K=0.7 for example
Drive current in saturation	I_d	K	$I_d = v_{sat} W_g C_o (V_g - V_{th})$ C_o : gate C per unit area $\rightarrow W_g (t_{ox}^{-1})(V_g - V_{th}) = W_g t_{ox}^{-1}(V_g - V_{th}) = KK^{-1}K = K$
I_d per unit W_g	$I_d/\mu m$	1	I_d per unit $W_g = I_d / W_g = 1$
Gate capacitance	C_g	K	$C_g = \epsilon_o \epsilon_{ox} L_g W_g / t_{ox} \rightarrow KK/K = K$
Switching speed	τ	K	$\tau = C_g V_{dd} / I_d \rightarrow KK/K = K$
Clock frequency	f	1/K	$f = 1/\tau = 1/K$
Chip area	A_{chip}	α	α: Scaling factor \rightarrow In the past, $\alpha > 1$ for most cases
Integration (# of Tr)	N	α/K^2	$N \rightarrow \alpha/K^2 = 1/K^2$, when $\alpha=1$
Power per chip	P	α	$fNCV^2/2 \rightarrow K^{-1}(\alpha K^{-2})K(K^1)^2 = \alpha = 1$, when $\alpha=1$

$k = 0.7$ and $\alpha = 1$

Single MOFET

$V_{dd} \rightarrow 0.7$

$L_g \rightarrow 0.7$

$I_d \rightarrow 0.7$

$C_g \rightarrow 0.7$

P (Power)/Clock

$\rightarrow 0.7^3 = 0.34$

τ (Switching time) $\rightarrow 0.7$

$k = 0.7^2 = 0.5$ and $\alpha = 1$

$V_{dd} \rightarrow 0.5$

$L_g \rightarrow 0.5$

$I_d \rightarrow 0.5$

$C_g \rightarrow 0.5$

P (Power)/Clock

$\rightarrow 0.5^3 = 0.125$

τ (Switching time) $\rightarrow 0.5$

Chip

N (# of Tr) $\rightarrow 1/0.7^2 = 2$

f (Clock) $\rightarrow 1/0.7 = 1.4$

P (Power) $\rightarrow 1$

N (# of Tr) $\rightarrow 1/0.5^2 = 4$

f (Clock) $\rightarrow 1/0.5 = 2$

P (Power) $\rightarrow 1$

- The concerns for limits of down-scaling have been announced for every generation.
- However, down-scaling of CMOS is still the ‘royal road’* for high performance and low power.
- Effort for the down-scaling has to be continued by all means.

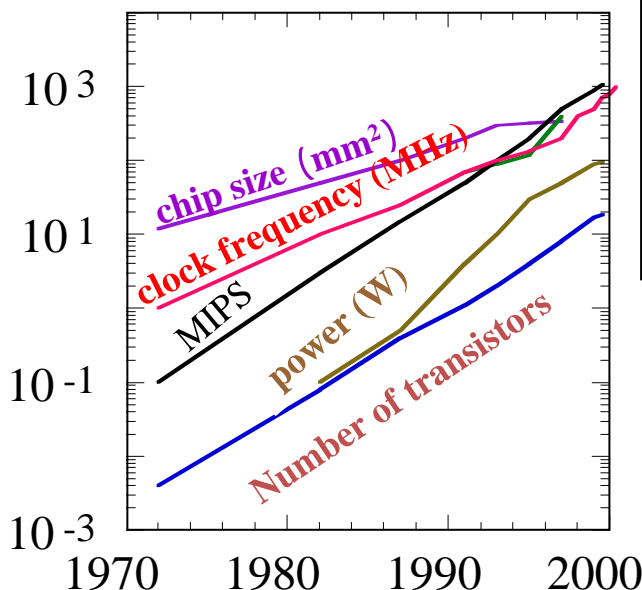
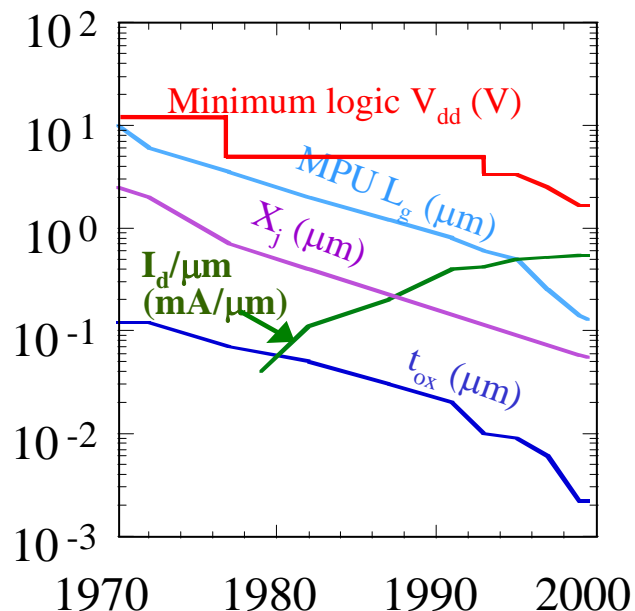
* Euclid of Alexandria (325BC?-265BC?)

‘There is no royal road to Geometry’

Mencius (Meng-zi), China (372BC?-289BC?)

孟子: 王道, 霸道 (Rule of right vs. Rule of military)

Actual past downscaling trend until year 2000



Past 30 years scaling
 Merit: N, f increase
 Demerit: P increase

V_{dd} scaling insufficient
 ↓
 Additional significant increase in I_d, f, P

Source: Iwai and S. Ohmi, Microelectronics Reliability 42 (2002), pp.1251-1268

Change in 30 years

	Ideal scaling	Real Change		Ideal scaling	Real Change		Ideal scaling	Real Change
L_g	K	10^{-2}	I_d	$K (10^{-2})$	10^{-1}	f	$1/K(10^2)$	10^3
t_{ox}	$K(10^{-2})$	10^{-2}	$I_d/\mu m$	1	10^1	P	$\alpha(10^1)$	10^5
V_{dd}	$K(10^{-2})$	10^{-1}	N	$\alpha/K^2(10^5)$	10^4	$= f\alpha NCV^2$		
A_{chip}	α	10^1						

V_d scaling insufficient, α increased $\rightarrow N, I_d, f, P$ increased significantly

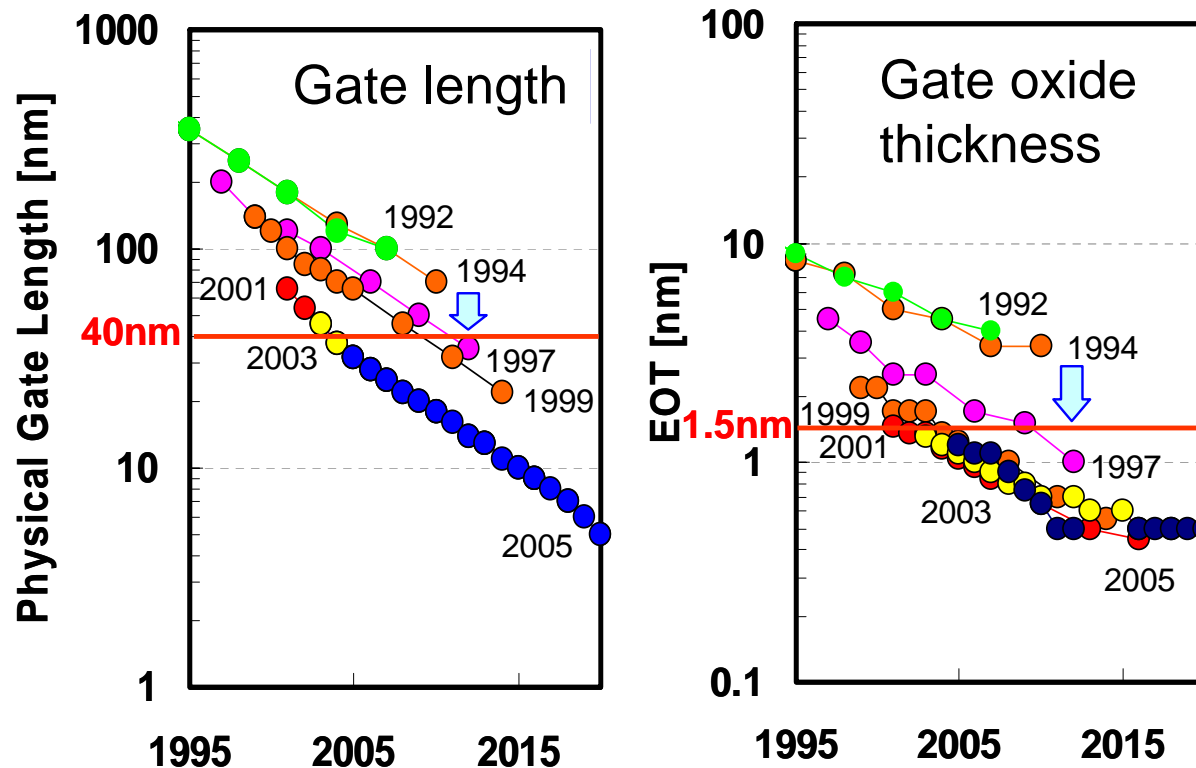
- Now, power and/or heat generation are the limiting factors of the down-scaling
- Supply voltage reduction is becoming difficult, because V_{th} cannot be decreased any more, as described later.
- Growth rate in clock frequency and chip area becomes smaller.

2. ITRS Roadmap (for 22 nm CMOS logic)

What is a roadmap? What is ITRS?

Roadmap: Prediction of future technologies

ITRS: International Technology Roadmap for Semico
made by SIA (Semiconductor Industry Associatio
with Japan, Europe, Korea and Taiwan

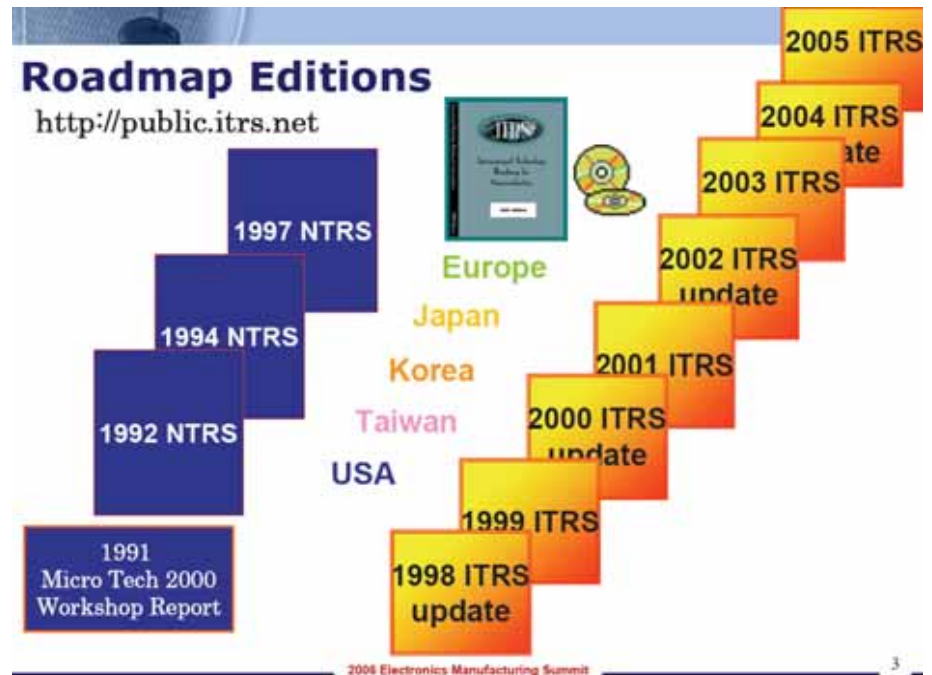


1992 -1997:NTRS (National Technology Roadmap)
1998 - : ITRS (International Technology Roadmap)

2008 ITRS
update

2007 ITRS

2006 ITRS
update

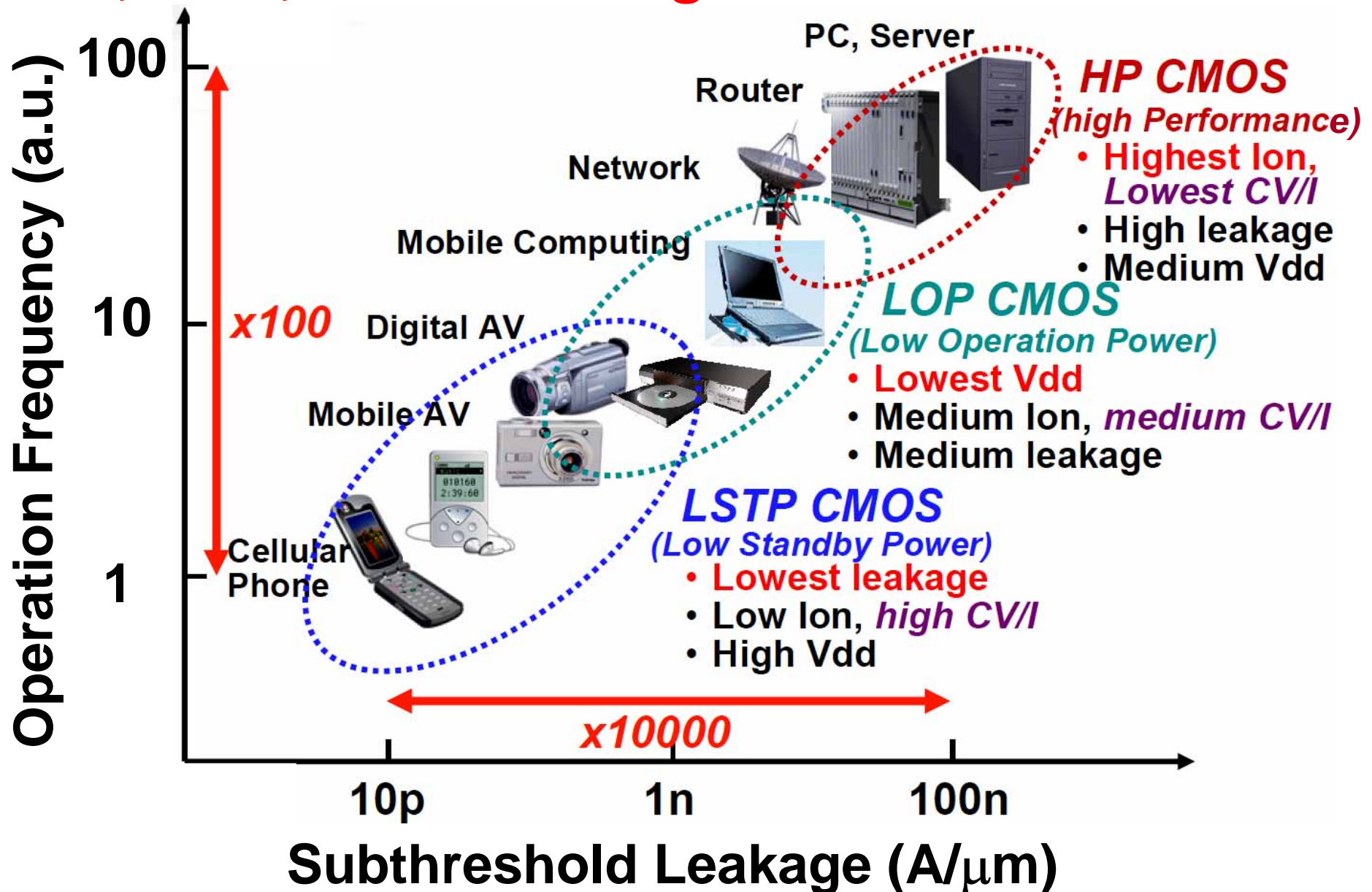


ITRS Roadmap does change every year!

<u>2007 Edition</u>	<u>2003 Edition</u>
<u>2006 Update</u>	<u>2002 Update</u>
<u>2005 Edition</u>	<u>2001 Edition</u>
<u>2004 Update</u>	<u>2000 Update</u>

<http://www.itrs.net/reports.html>

HP, LOP, LSTP for Logic CMOS



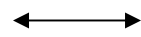
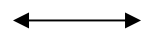
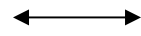
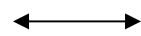
Source: 2007 ITRS Winter Public Conf.

What does '22 nm' mean in 22 nm CMOS Logic?

'XX nm CMOS Technology

Commercial Logic CMOS products

Technology name	Starting Year
45 nm	2007
32 nm	2009?
22 nm	2011?~ 2012?
16 nm	2013?~ 2014?



ITRS (Likely in 2008 Update)

for High Performance Logic

Year	Half Pitch (1 st Metal)	Physical Gate Length
2007	68 nm	32 nm
2008	59 nm	29 nm
2009	52 nm	27 nm
2010	45 nm	24 nm
2011	40 nm	22 nm
2012	36 nm	20 nm
2013	32 nm	18 nm
2014	29 nm	16 nm

Source: 2008 ITRS Summer Public Conf.

'XX nm' CMOS Logic Technology:

- In general, there is no common corresponding parameter with 'XX nm' in ITRS table, which stands for 'XX nm' CMOS.

What does '22 nm' mean in 22 nm CMOS Logic?

- 'XX nm' does not correspond to the 'Half Pitch' nor 'Physical Gate Length' of ITRS.
- 'XX nm' is now just a commercial name for CMOS Logic generation of size and its technology.
- Actual parameter values and starting years for commercial products are somewhat different from the above ITRS table, depending on semiconductor companies.
- In 22 and 16 nm technologies, physical gate lengths of high-performance logic device may be close to XX nm.

What does '22 nm' mean in 22 nm CMOS Logic?

8 μ m \rightarrow 6 μ m \rightarrow 4 μ m \rightarrow 3 μ m \rightarrow 2 μ m \rightarrow 1.2 μ m \rightarrow 0.8 μ m \rightarrow 0.5 μ m

- Originally, 'XX' means lithography resolution.
- Thus, 'XX' was the gate length, and half pitch of lines
- 'XX' had shrunk 0.7 in 3 years in average (0.5 in 6 years) those days.
- 'XX' value deviated among companies: **example: 1.5 μ m, 1.2 μ m, 1 μ m**

\rightarrow 350nm \rightarrow 250nm \rightarrow 180nm \rightarrow 130nm \rightarrow 90nm \rightarrow 65nm \rightarrow 45nm

- 'XX' values were established by NTRS* and ITRS with the term of 'Technology Node**' and 'Cycle***' using typical 'half pitch value'.

*NTRS: National Tech. Roadmap, **Term 'Technology Node' is not used now.

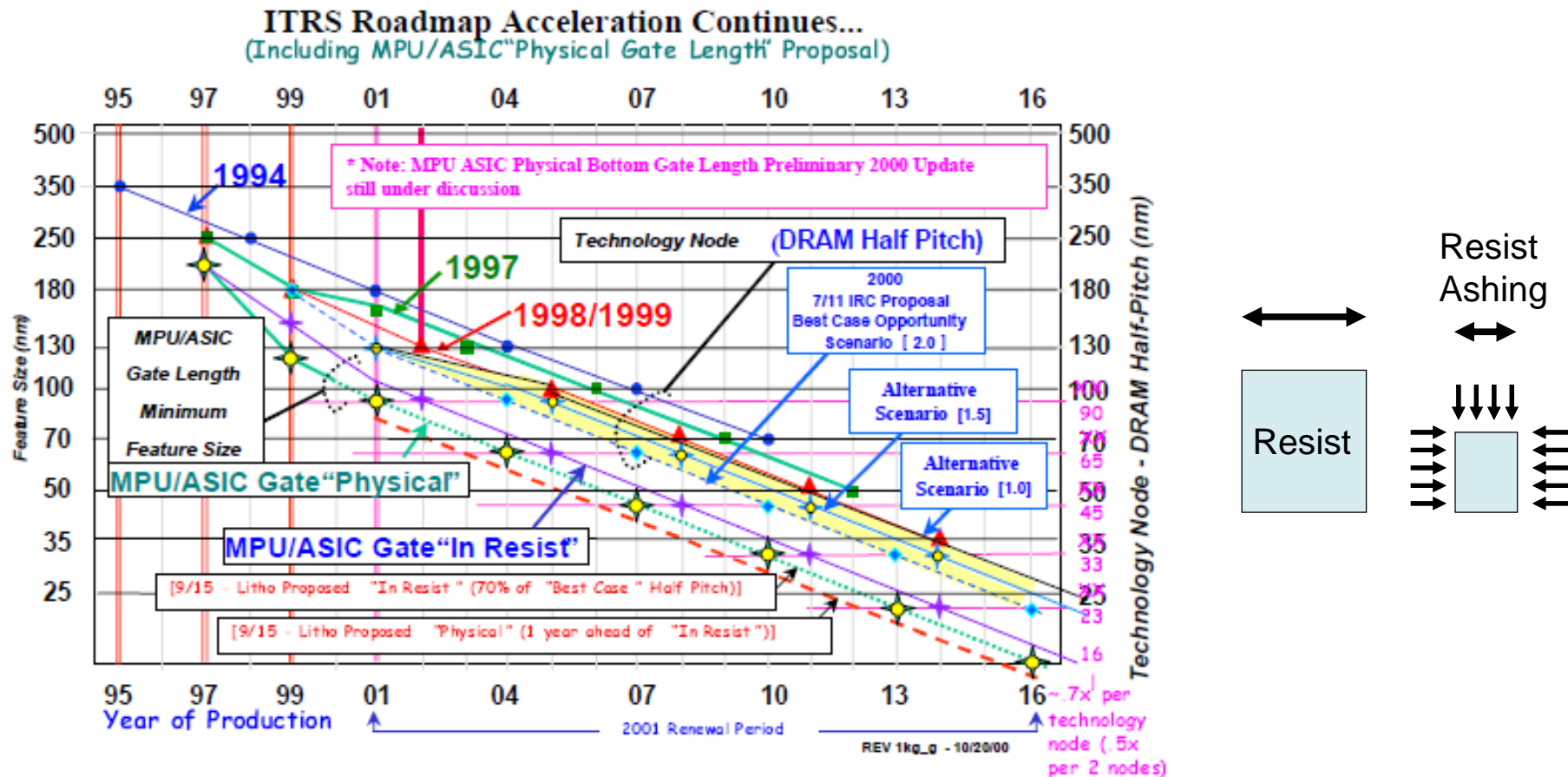
***Cycle: Period or year for which the half pitch becomes X0.71.

- The gate length of logic CMOS became smaller with one or two generations from the half pitch, and 'XX' names ahead of generations have been used for logic CMOS.
- Memory still keeps the half pitch as the value of 'XX'

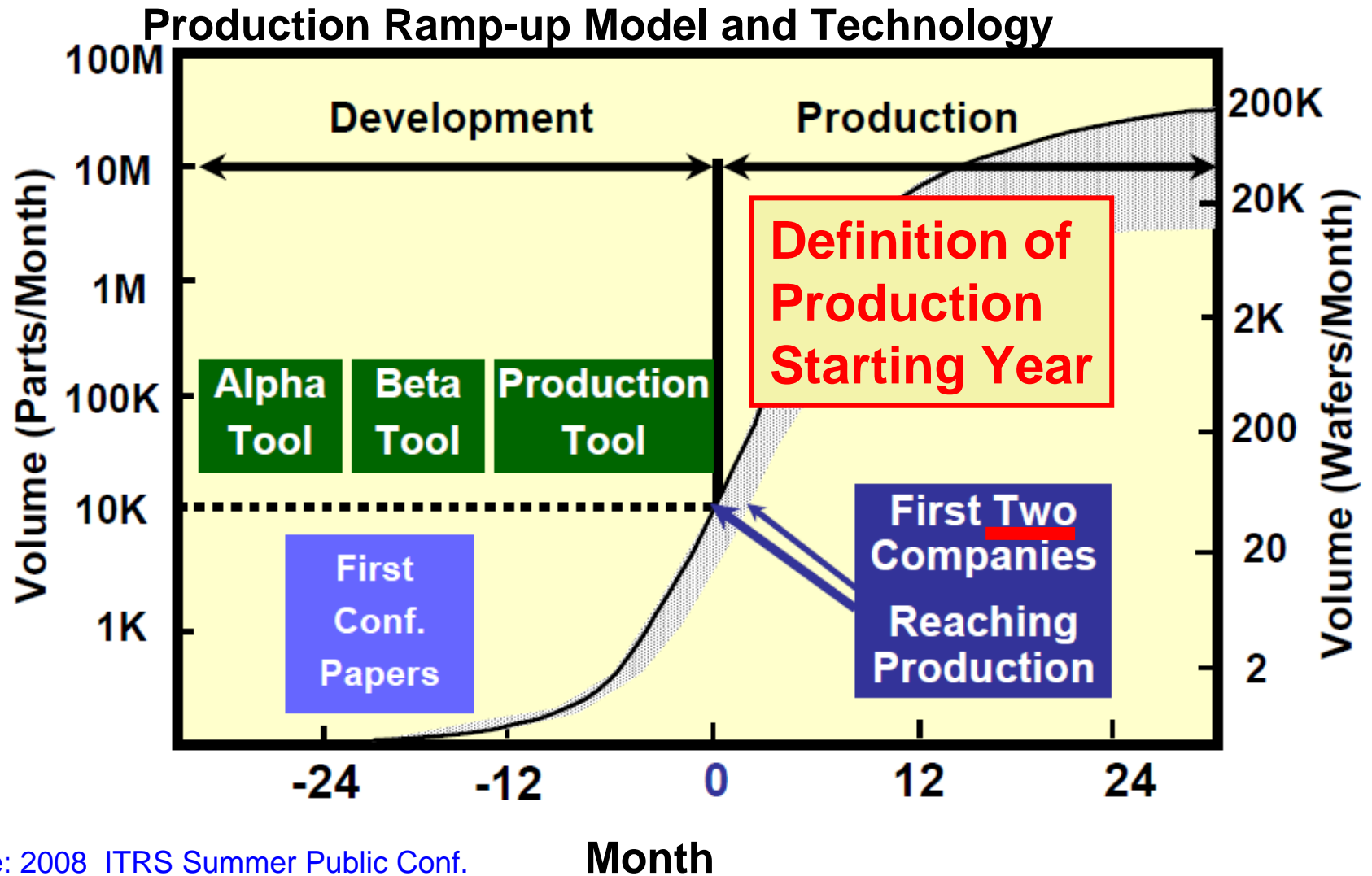
\rightarrow 32nm \rightarrow 22nm \rightarrow 16nm \rightarrow 11nm \rightarrow 8nm?? \rightarrow 5.5nm ??

What does '22 nm' mean in 22 nm CMOS Logic?

Gate length of Logic CMOS became significantly smaller than lithography resolution or half-pitch using special technique such as resist ashing (or trimming) method since 350 nm CMOS.



Source: ITRS 2001 Update

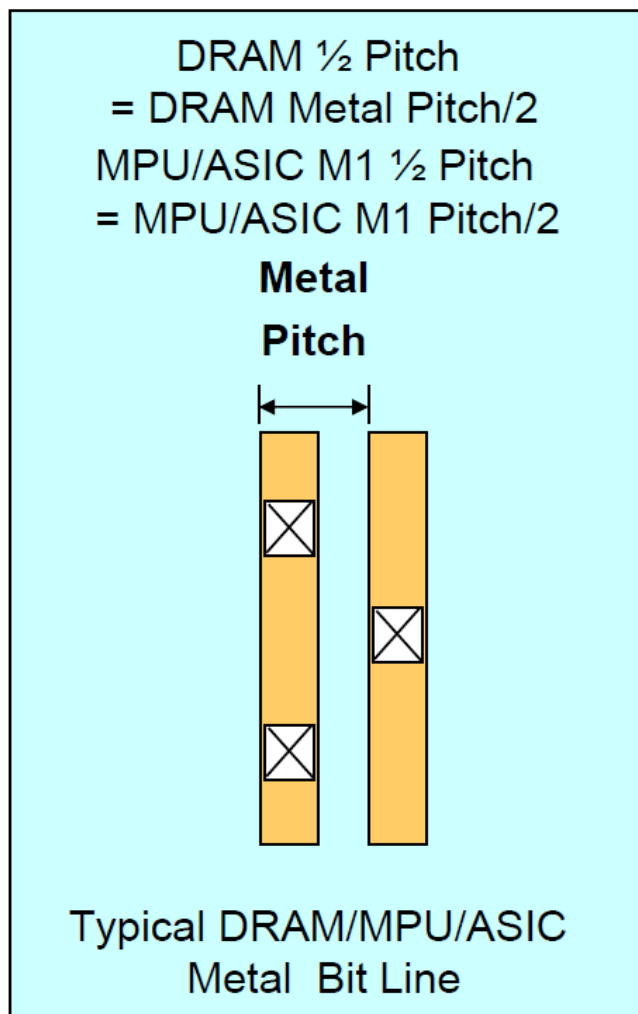


Source: 2008 ITRS Summer Public Conf.

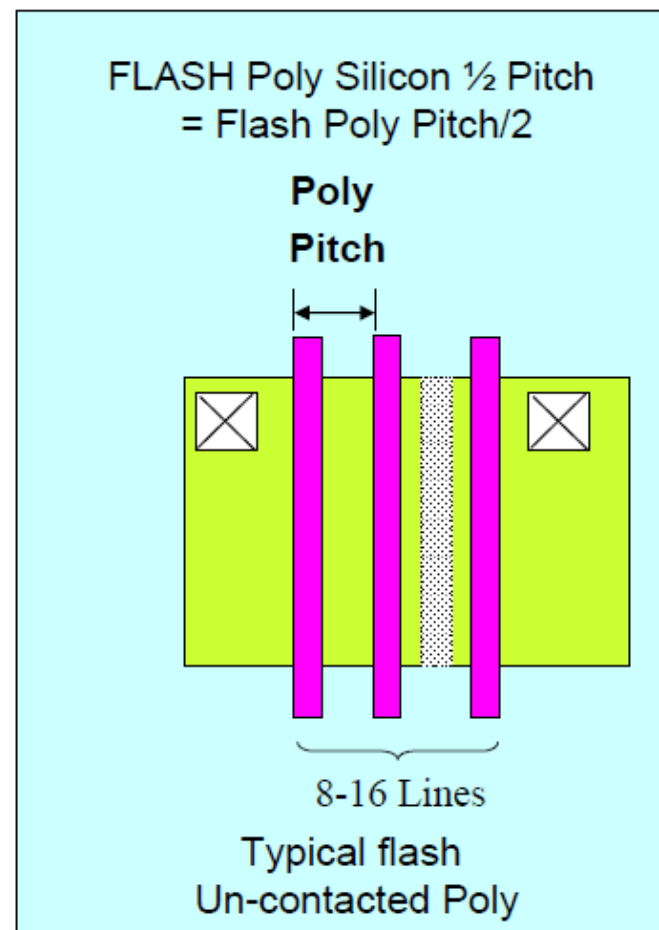
Some Problem: Number of most advanced logic CMOS companies is decreasing in generations.

Definition of the Half Pitch

Logic 1st Metal Half Pitch



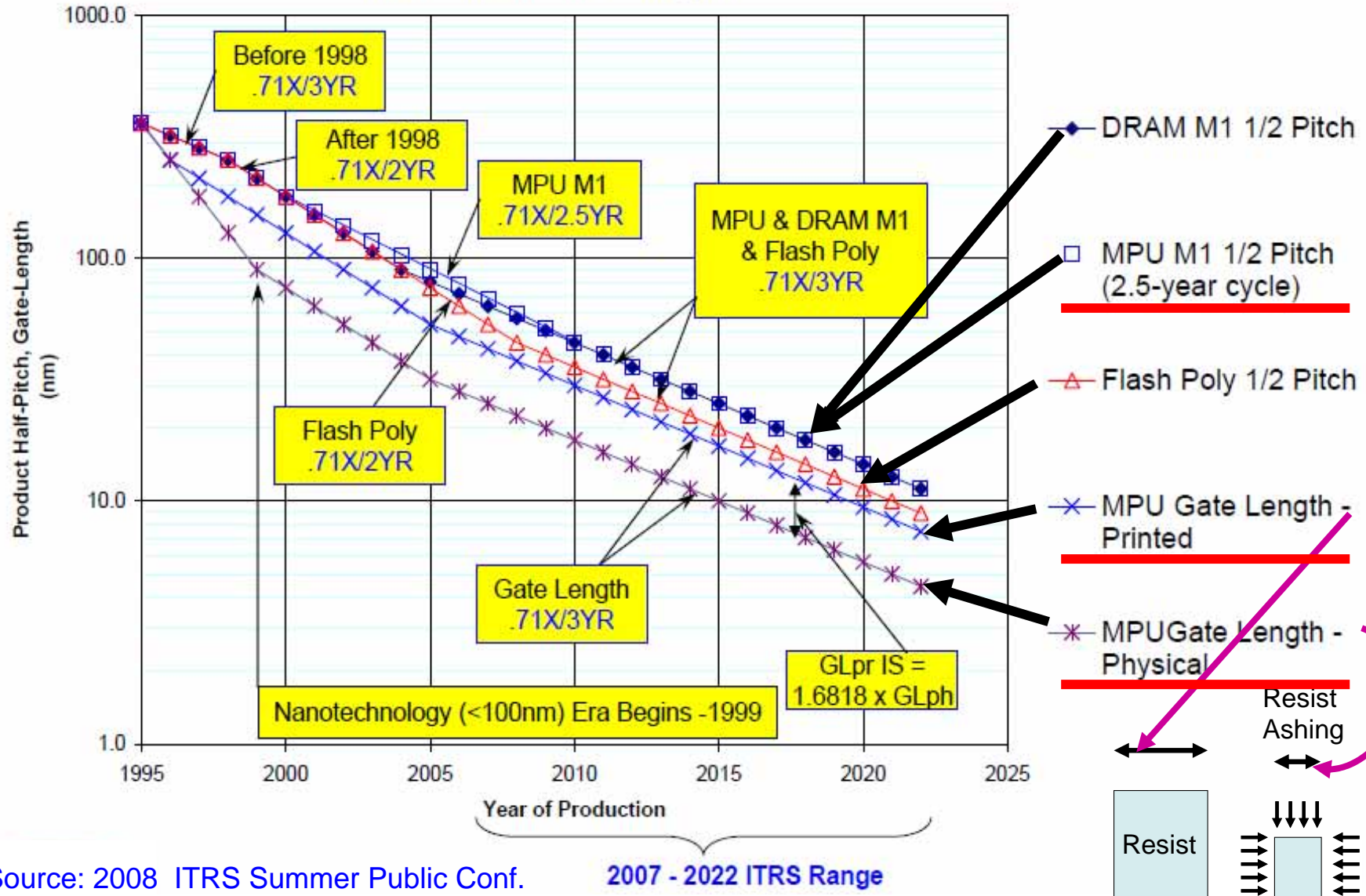
Flash Poly Gate Half Pitch



Source: 2008 ITRS Summer Public Conf.

For example, Typical Half Pitches at ITRS 2007

2007 ITRS Product Technology Trends - Half-Pitch, Gate-Length [WAS]



Source: 2008 ITRS Summer Public Conf.

2007 - 2022 ITRS Range

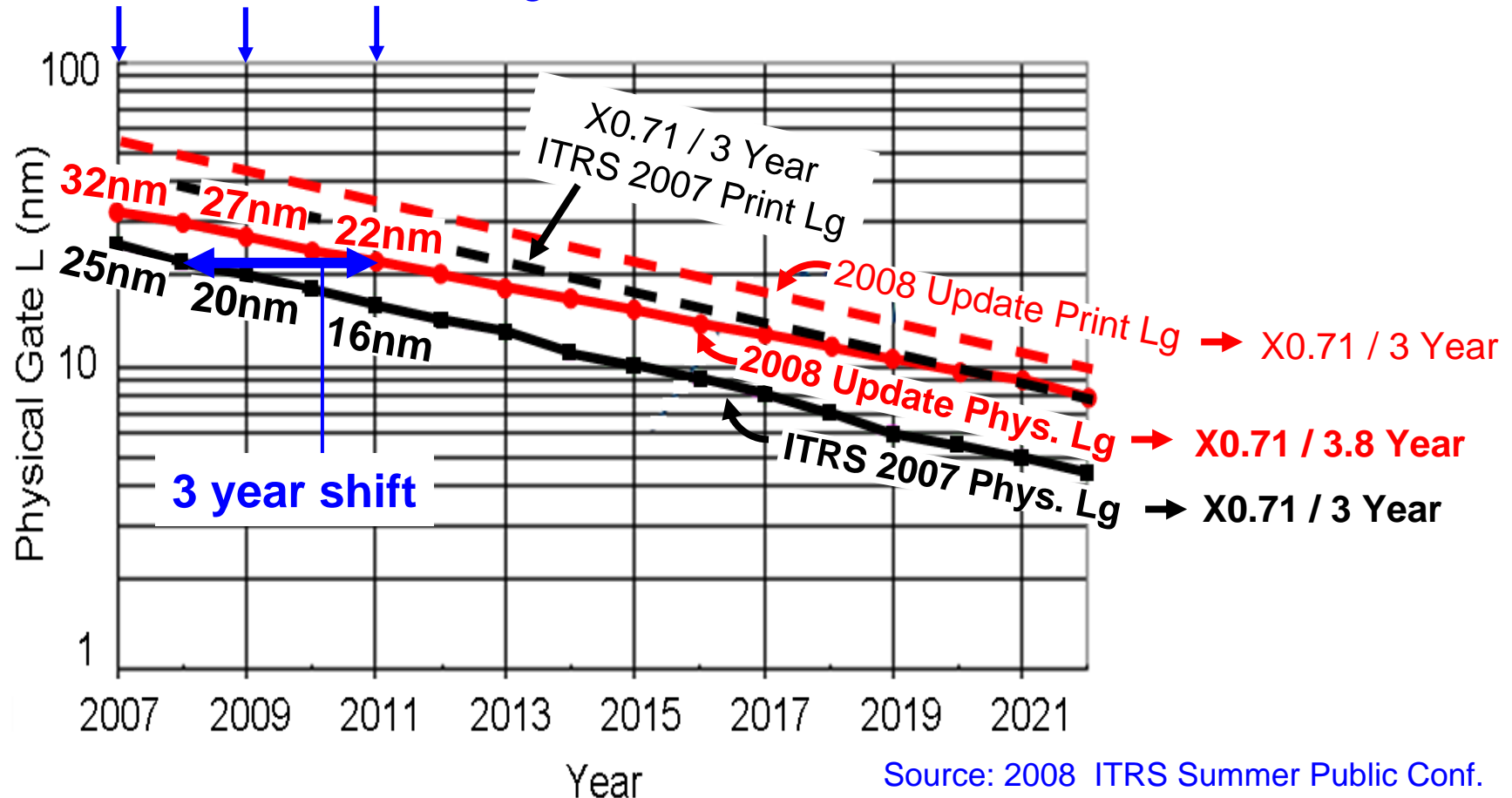
Physical gate length in past ITRS was too aggressive.

The dissociation from commercial product prediction will be adjusted.

Physical gate length of High-Performance logic will shift by 3-5 yrs.

Correspond to

45nm 32nm 22nm Logic CMOS



EOT and Xj shift backward, corresponding to Lg shift

EOT: 0.55 nm → 0.88 nm, Xj: 8 nm → 11 nm @ 22nm CMOS

Likely in 2008 Update Correspond to 22nm Source: 2008/ ITRS Summer Public Conf.

Year of Production	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
2007 MPU/ASIC Lg (nm)	25	23	20	18	16	14	13	11	10	9	8	7	6.3	5.6	5	4.5
2008 MPU/ASIC Lg (nm)	32	29	27	24	22	20	18	17	15	14.0	12.8	11.7	10.7	9.7	8.9	8.1
Shift/Interpolate Formula	2005	intrap	intrap	intrap	intrap	2009	2010	intrap	intrap	2012	intrap	intrap	intrap	intrap	intrap	intrap
EOT w/3E20 poly, bulk MPU (nm)	1.2	0.71	0.54	0.41												
EOT w/3E20 poly, bulk MPU (nm)	1.3	1.2	1.2	1	0.68	0.54	0.41									
EOT w/metal gate, bulk MPU (nm)		0.9	0.75	0.65	0.55	0.50										
EOT w/metal gate, bulk MPU (nm)			1.0	0.95	0.88	0.75	0.65	0.60	0.53	0.5						
Drain Ext. Xj bulk MPU (nm)	12.5	11	10	9	8	7										
Drain Ext. Xj bulk MPU (nm)	11	11	11	11	11		9	8.5	7.7	7						

Likely in 2008 Update

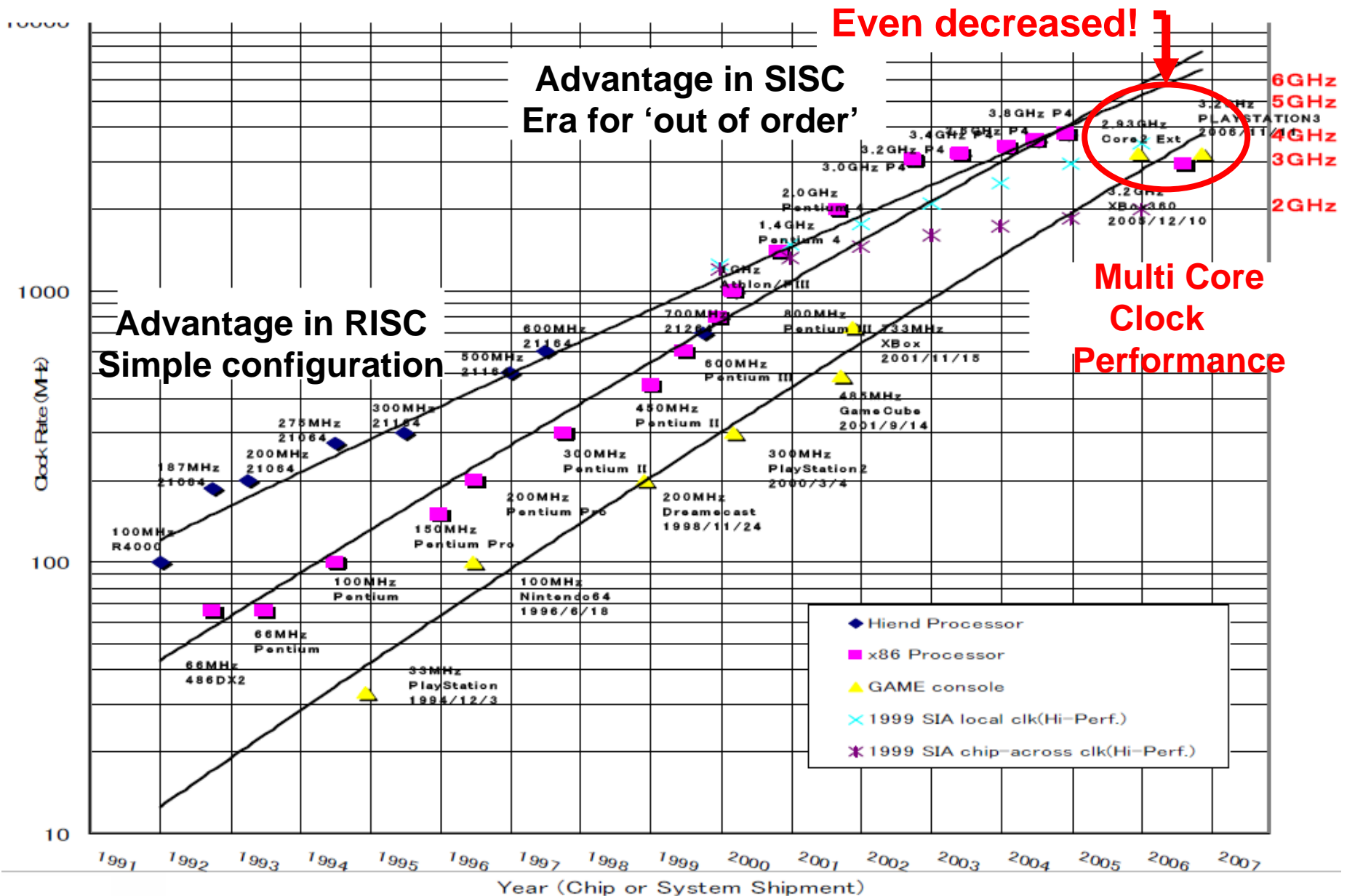
Likely in 2008 Update

Likely in 2008 Update

non-steady trend corrected

filled in for metal gate EOT for 2009/10 based on latest conference presentations

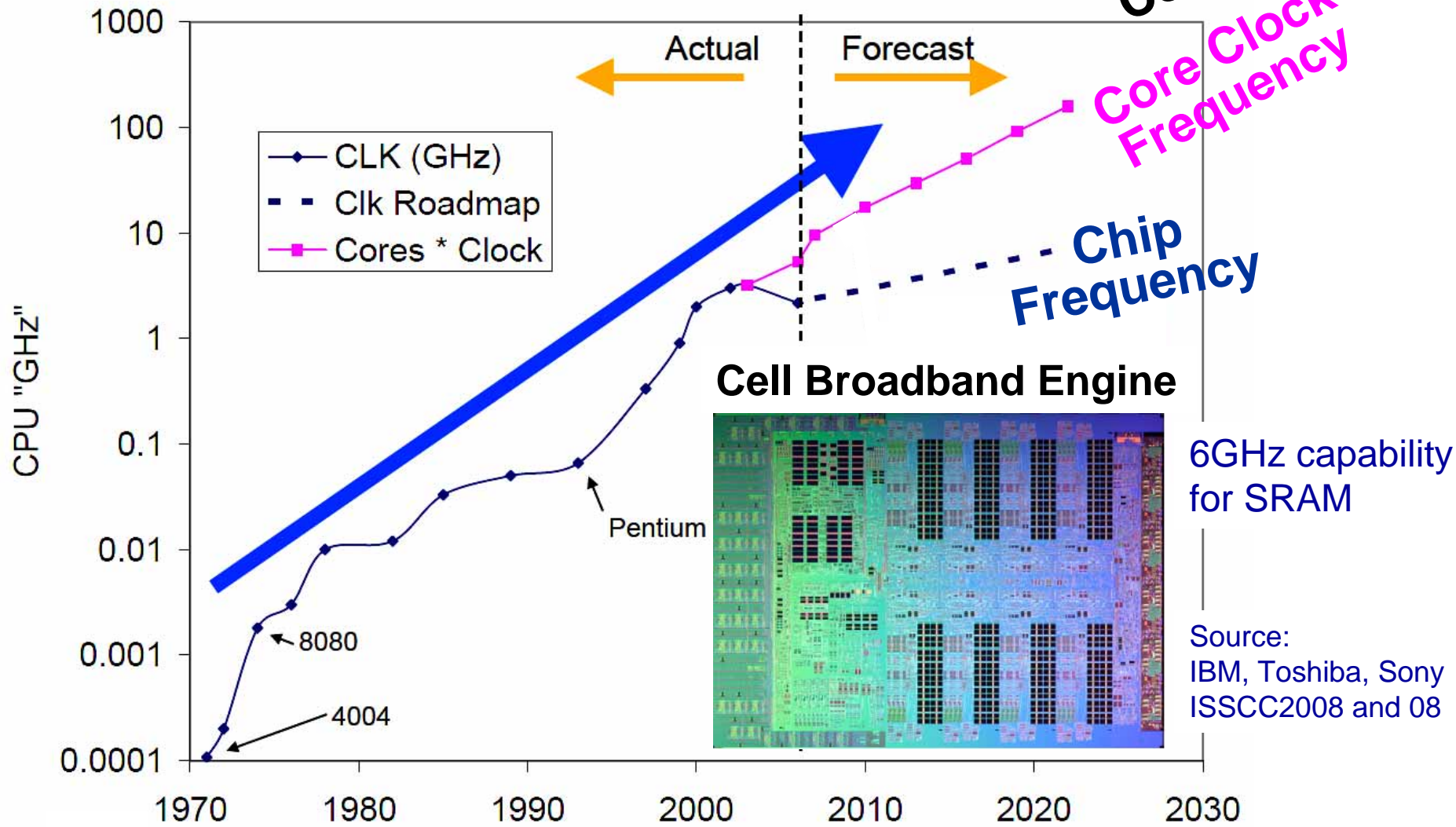
Clock frequency does not increase aggressively anymore.



Source: Mitsuo Saito, Toshiba

MPU "GHz" by "Cores" ITRS2007

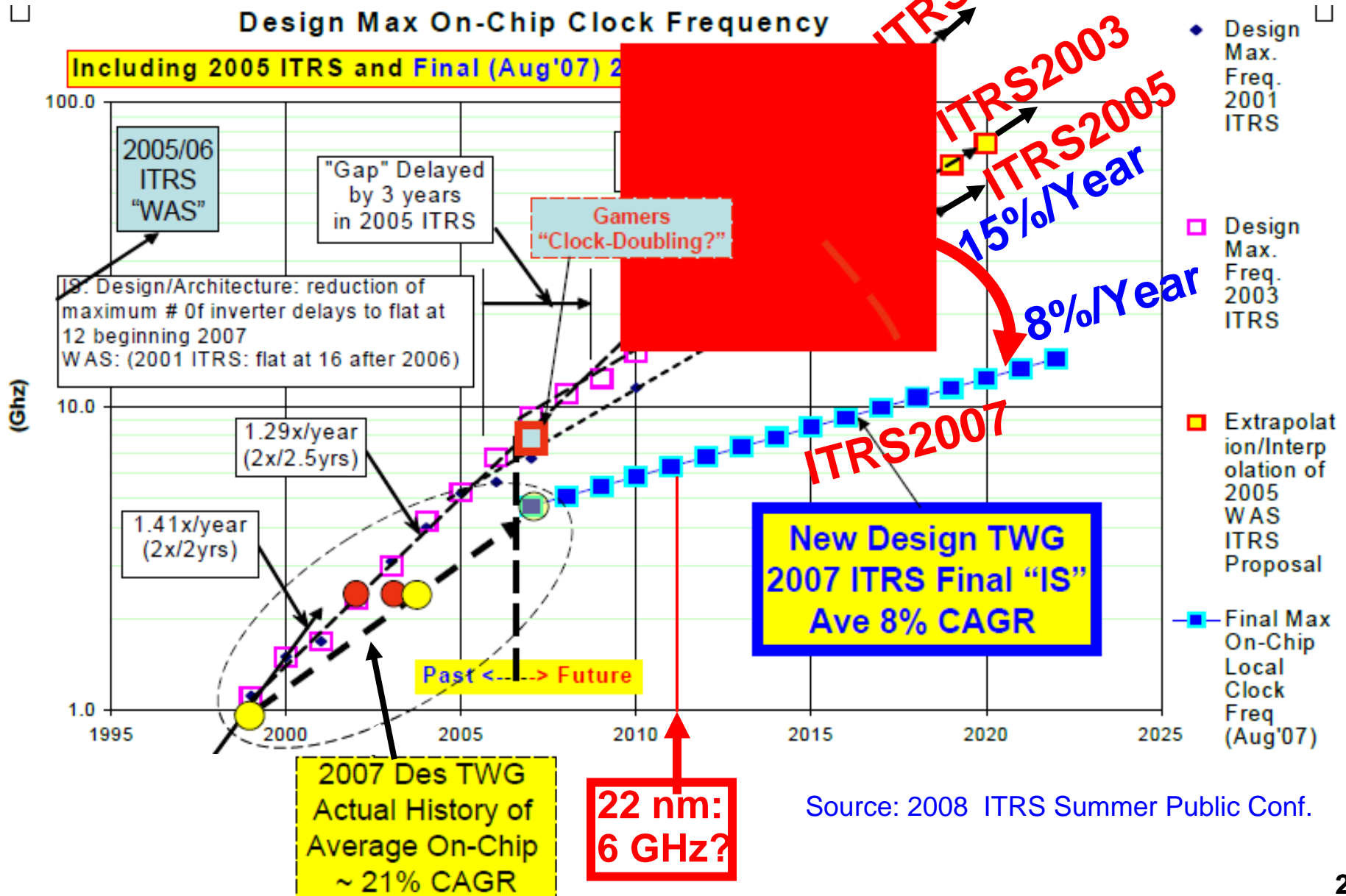
Continued?



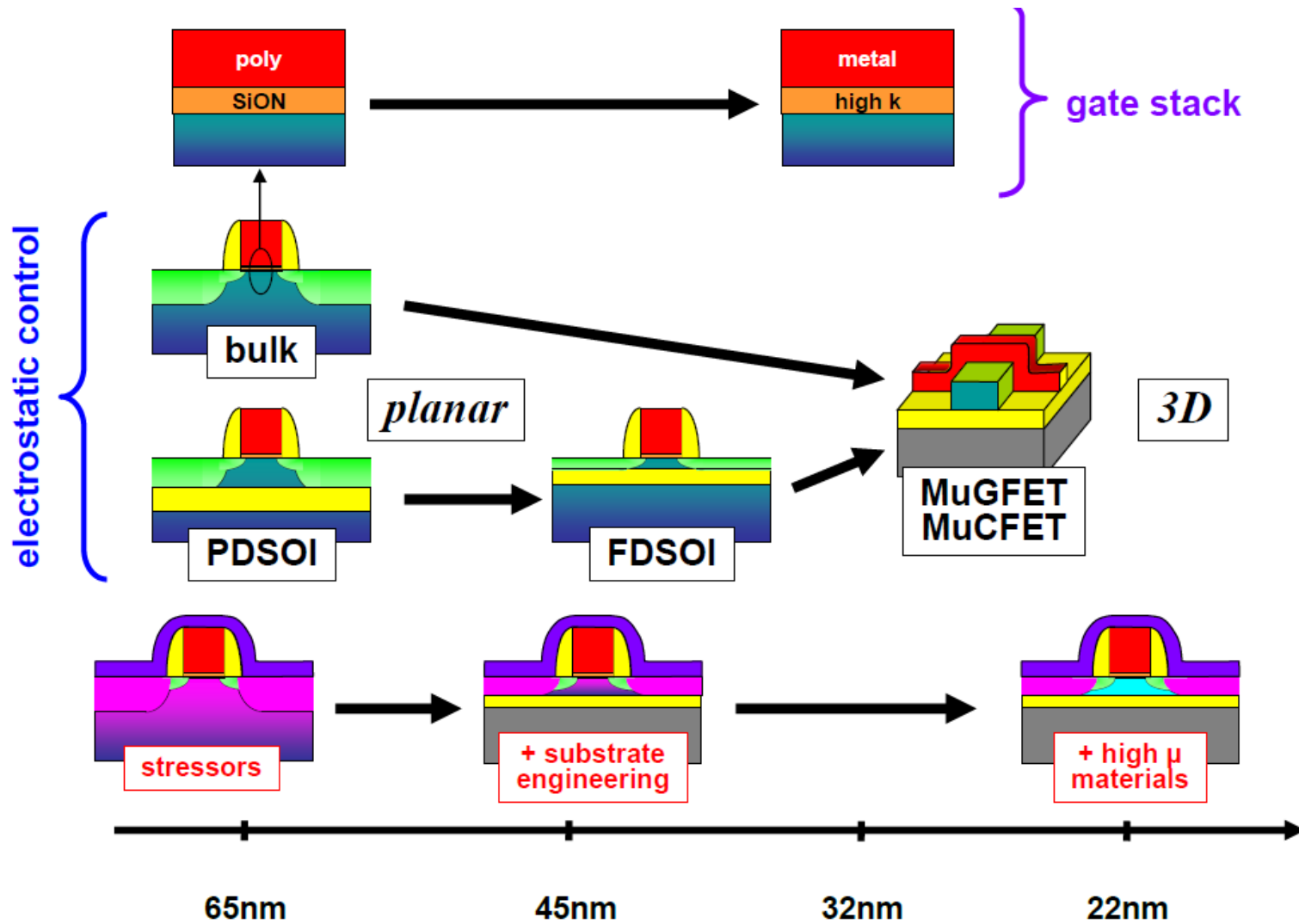
Source: 2007 ITRS Winter Public Conf.

Clock frequency Change in the past ITRS

(Max on chip frequency or 'Core clock')

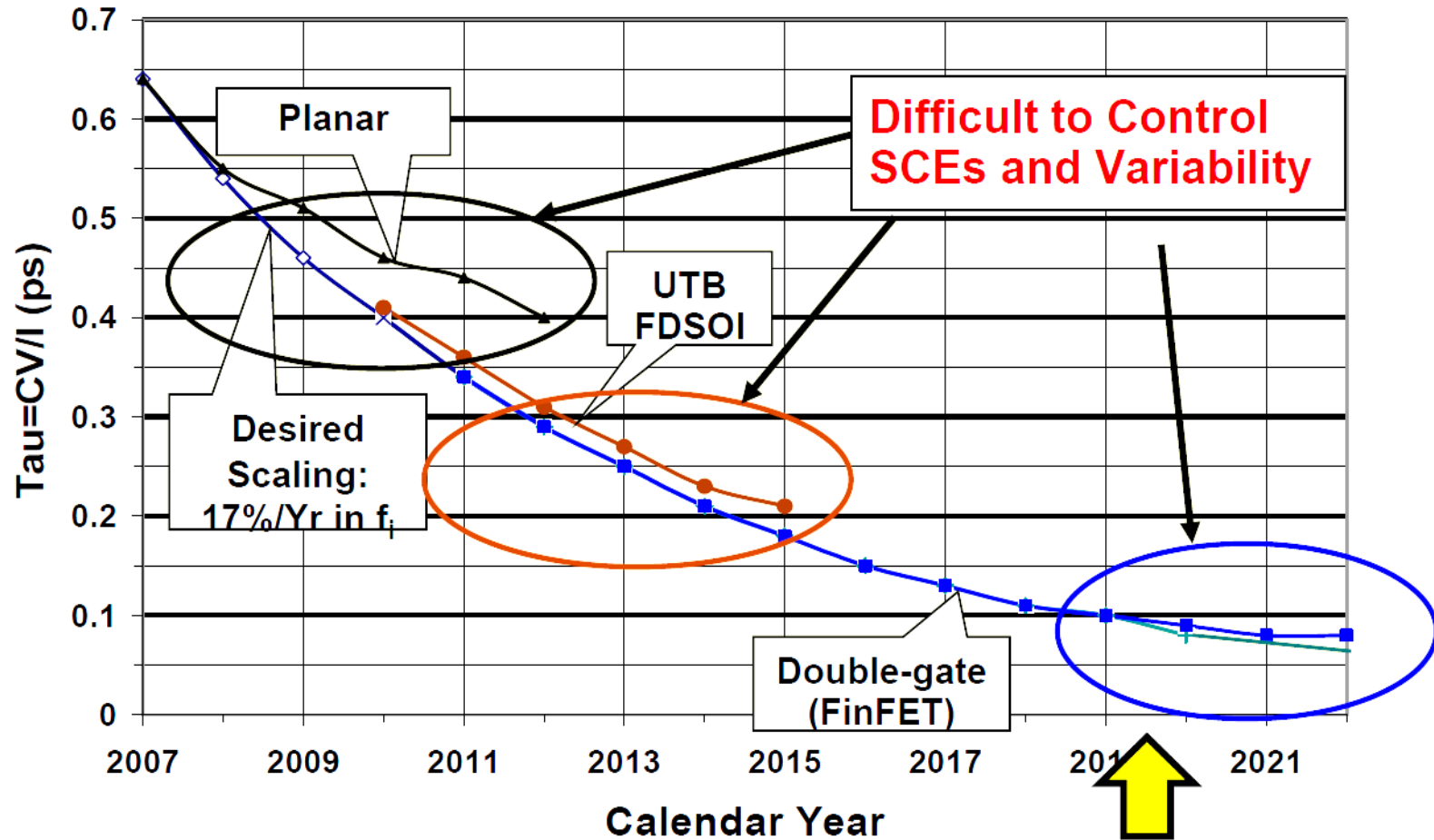


Structure and technology innovation (ITRS 2007)



Source: 2008 ITRS Summer Public Conf.

Technology innovation described in ITRS 2007



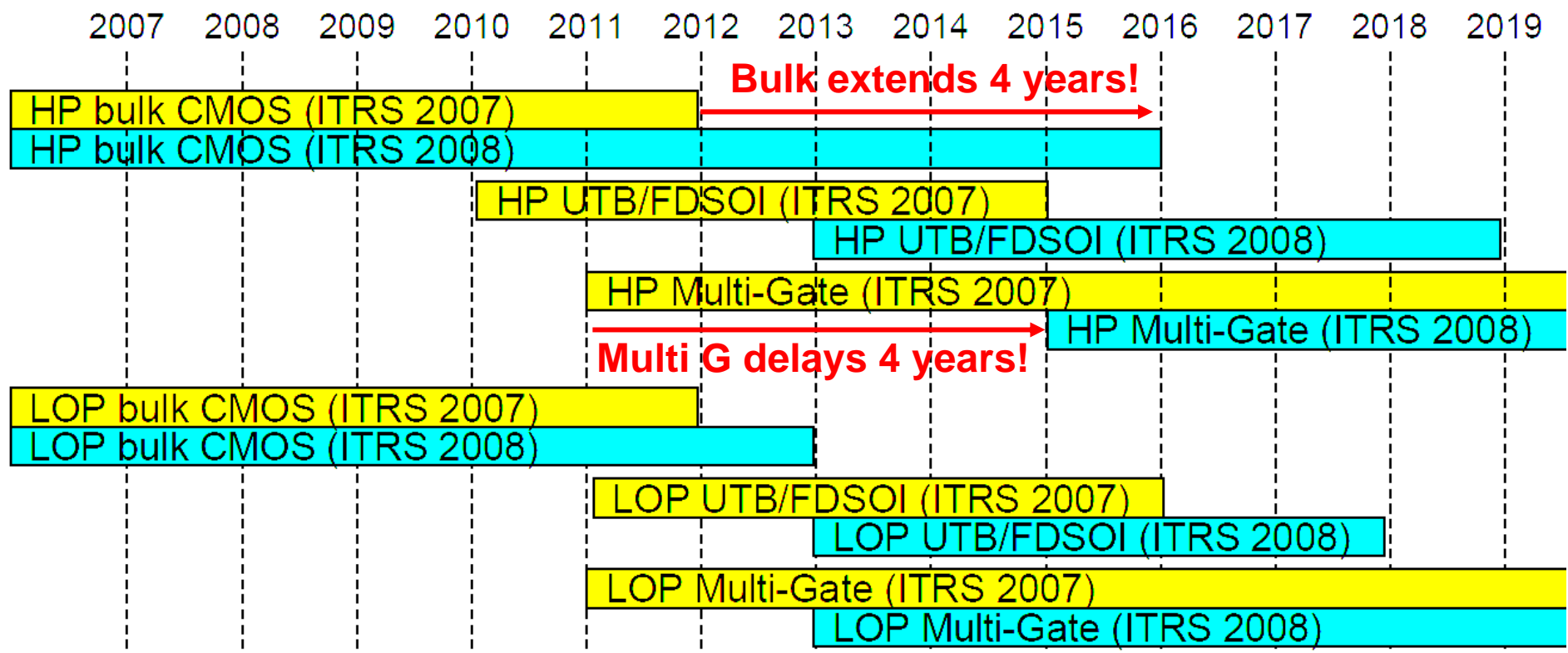
Alternative material (Ge, III-V) and structure (Nanowire) in channel region.

Source: 2007 ITRS Winter Public Conf.

Timing of CMOS innovations shifts backward.

Bulk CMOS has longer life now!

Correspond to 22nm Logic CMOS



Source: 2008 ITRS Summer Public Conf.

Wafer size (ITRS 2007)

Correspond to 22nm

Year of Production	2007	2008	2009	2010	2011	2012	2013	2014	2015
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	68	59	52	45	40	36	32	28	25
MPU Physical Gate Length (nm)	25	23	20	18	16	14	13	11	10
MPU High-Performance Total Chip Area(mm ²)	310	246	195	310	246	195	310	246	195
MPU High-Performance Active Transistor Area(mm ²)	31.7	25.1	20.0	31.7	25.1	20.0	31.7	25.1	20.0
<i>General Characteristics * (99% Chip Yield)</i>									
Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)**	300	300	300	300	300	450	450	450	450

Source: ITRS 2007

→ ??
Maybe delay??

Gate CD (Critical Dimension) Control

ITRS 2007

Correspond to 22nm Logic

Year of Production	2007	2008	2009	2010	2011	2012	2013	2014	2015
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)(contacted)	68	59	52	45	40	36	32	28	25
MPU Physical Gate Length (nm)	25	23	20	18	16	14	13	11	10
Total maximum allowable etch 3σ (nm), including photoresist trim and gate etch [AA]	1.5	1.38	1.2	1.08	0.96	0.84	0.78	0.66	0.6

Source: ITRS 2007

2008 Update

Correspond to 22nm Logic

Year of Production	2007	2008	2009	2010	2011	2012	2013	2014	2015
MPU Physical Gate Length (nm)	32	29	27	24	22	20	18	17	15
L_{gate} 3σ variation (nm) [Z]	3.82	3.49	3.18	2.9	2.65	2.42	2.21	2.02	1.84

Source: 2008 ITRS Summer Public Conf.

Gate CD control color changed to 'white' through 2011 and to 'yellow' for 2012 reflecting the new Lg scaling

ITRS2008 Low-k Roadmap Update

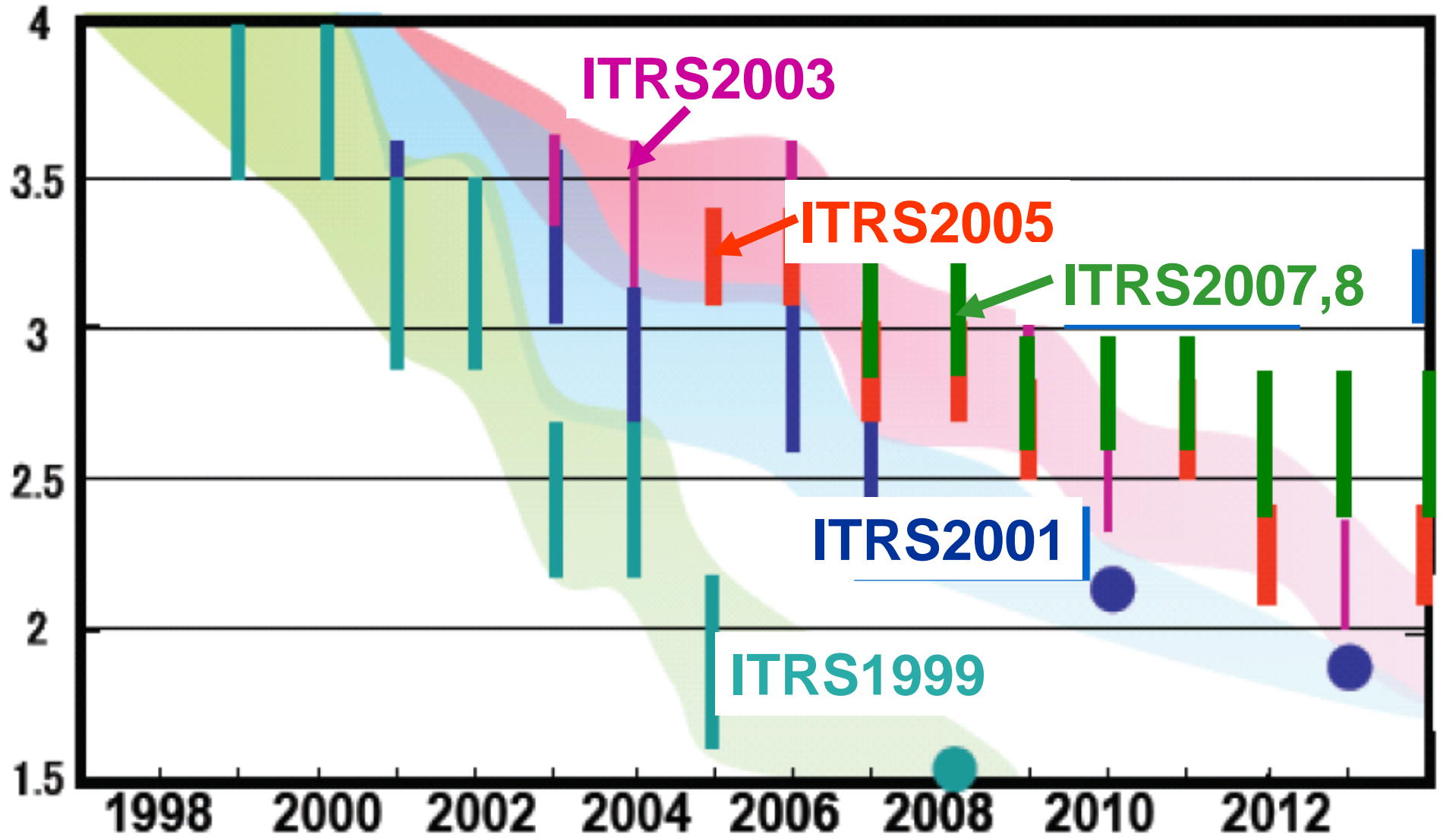
Correspond to 22nm Logic

		Near-term					
	<i>Year of Production</i>	2008	2009	2010	2011	2012	2013
ITRS 2007	Interlevel metal insulator – effective dielectric constant (κ)	2.7-3.0	2.5-2.8	2.5-2.8	2.5-2.8	2.1-2.4	2.1-2.4
Update 2008	Interlevel metal insulator – effective dielectric constant (κ)	2.9-3.3	2.6-2.9	2.6-2.9	2.6-2.9	2.4-2.8	2.4-2.8
ITRS 2007	Interlevel metal insulator – bulk dielectric constant (κ)	2.3-2.7	2.1-2.4	2.1-2.4	2.1-2.4	1.8-2.1	1.8-2.1
Update 2007	Interlevel metal insulator – bulk dielectric constant (κ)	2.5-2.8	2.3-2.6	2.3-2.6	2.3-2.6	2.1-2.4	2.1-2.4

Source: 2008 ITRS Summer Public Conf.

k value increases by 0.1 ~ 0.3

Historical Transition of ITRS Low-k Roadmap



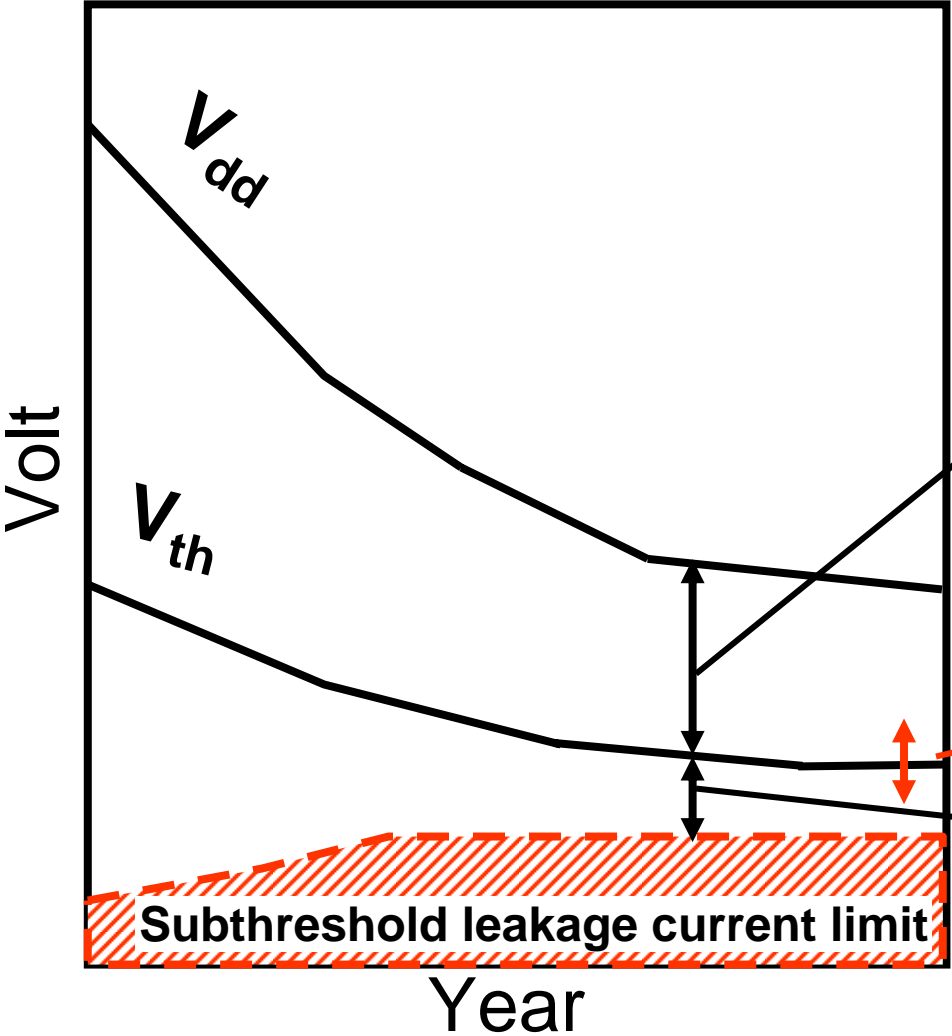
Source: 2008 ITRS Summer Public Conf.

Roadmap towards 22nm technology and beyond

- Physical gate length downsizing rate will be less aggressive.
- Corresponding to the above, performance increase would slow down – Clock frequency, etc.
- Introduction of innovative structures – UTB SOI and DG delayed, and bulk CMOS has longer life than predicted by previous ITRS roadmaps.

3. Voltage Scaling / Low Power and Leakage

Difficulty in Down-scaling of Supply Voltage: V_{dd}



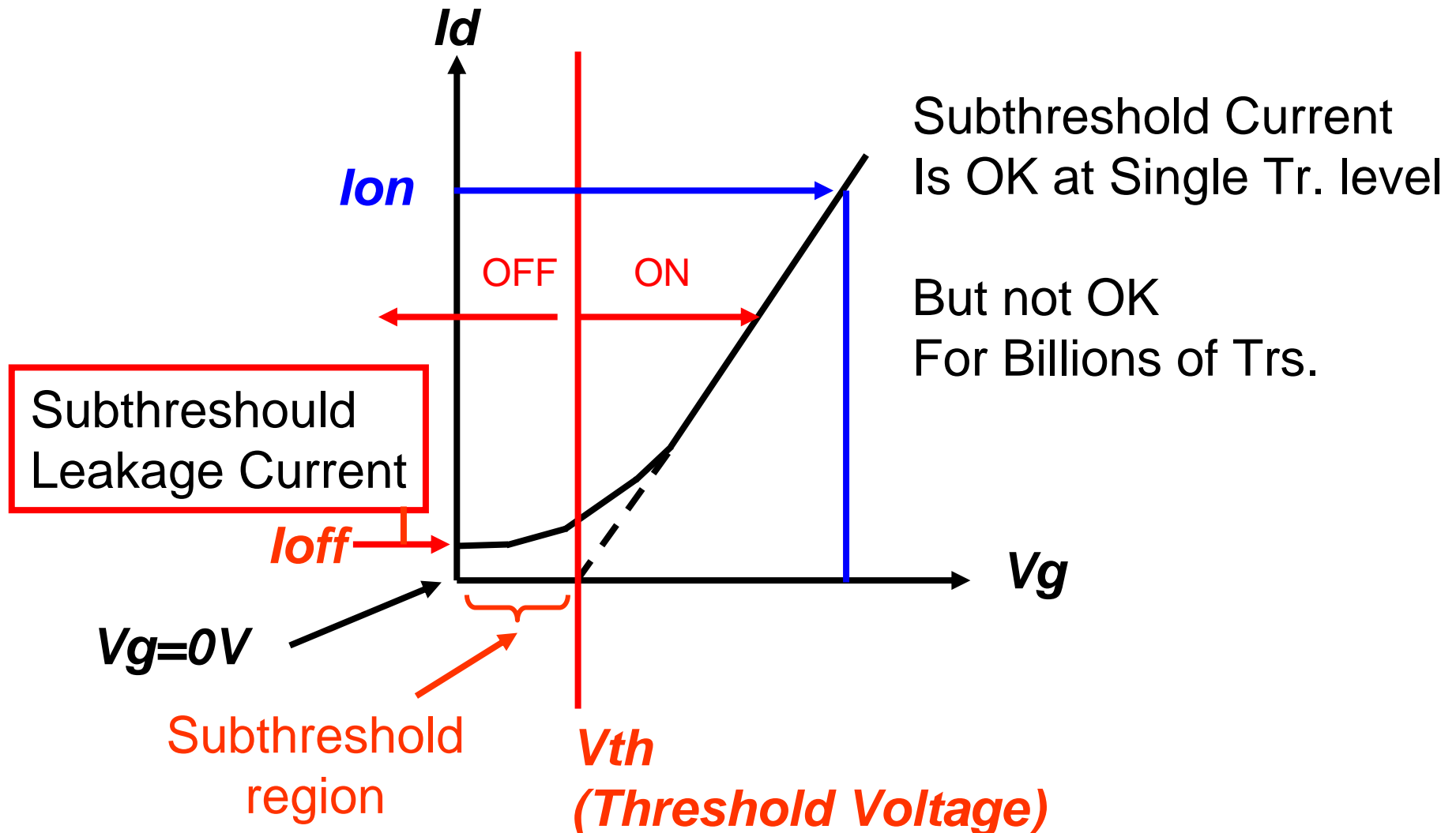
Because, V_{th} cannot be down-scaled anymore, V_{dd} down-scaling is difficult.

$V_{dd} - V_{th}$ determines the performance (High I_d) and cannot be too small.

ΔV_{th} : V_{th} variation

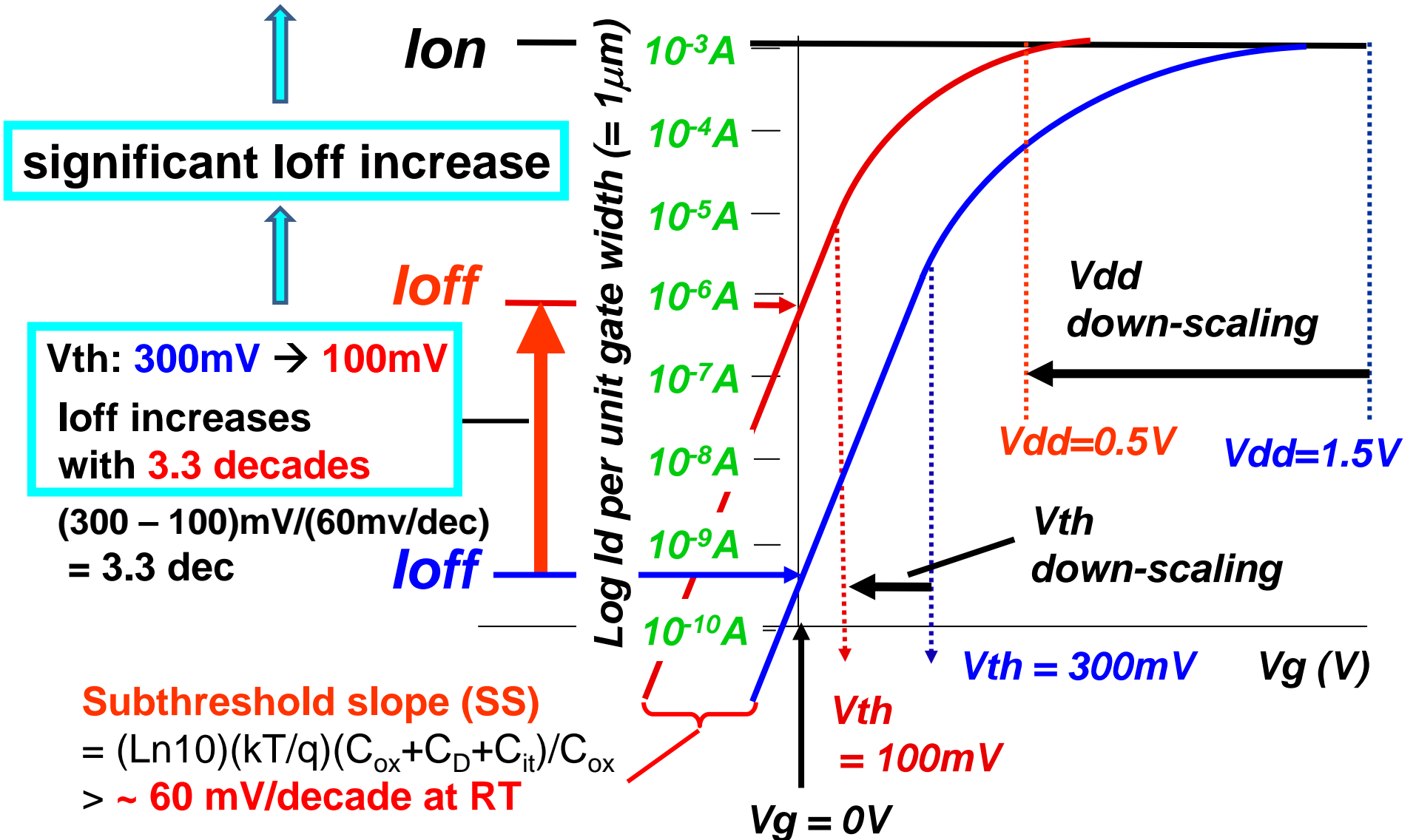
$> \Delta V_{th}$
Margin for V_{th} variation is necessary

Subthreshold leakage current of MOSFET



Vth cannot be decreased anymore

Log scale Id plot



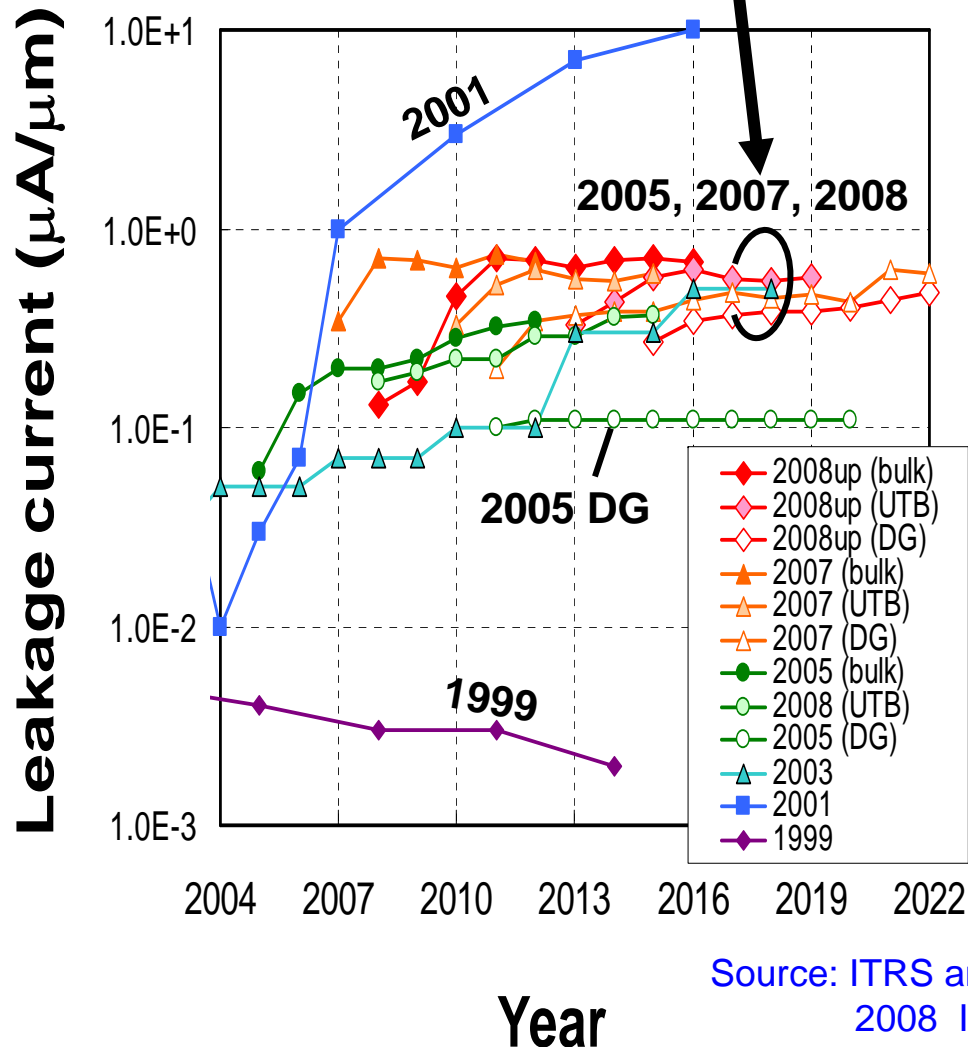
SS value:

Constant and does not become small with down-scaling

ITRS for HP logic

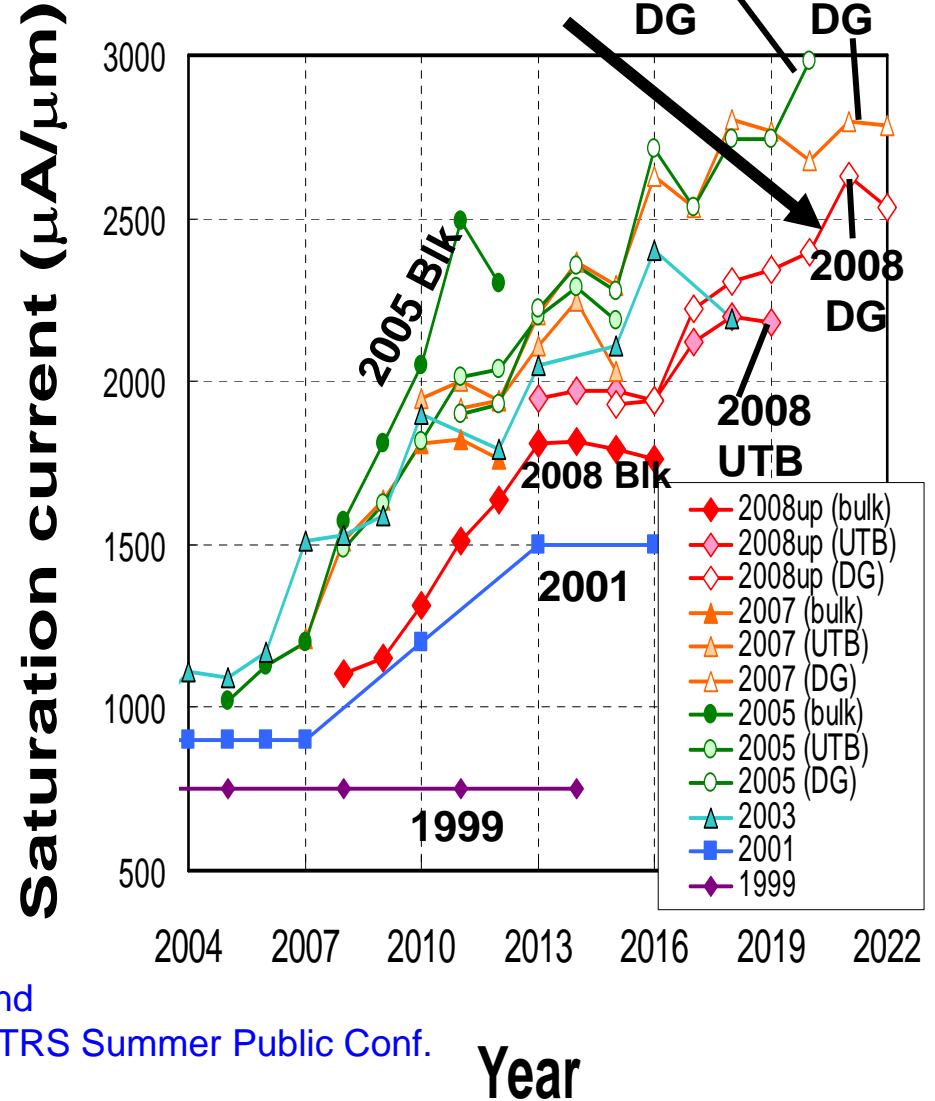
S-D leakage current

Isd-leak has to be stay less than 1 $\mu\text{A}/\mu\text{m}$



Id-sat growth will be modest in 2008 update

Saturated Drain current

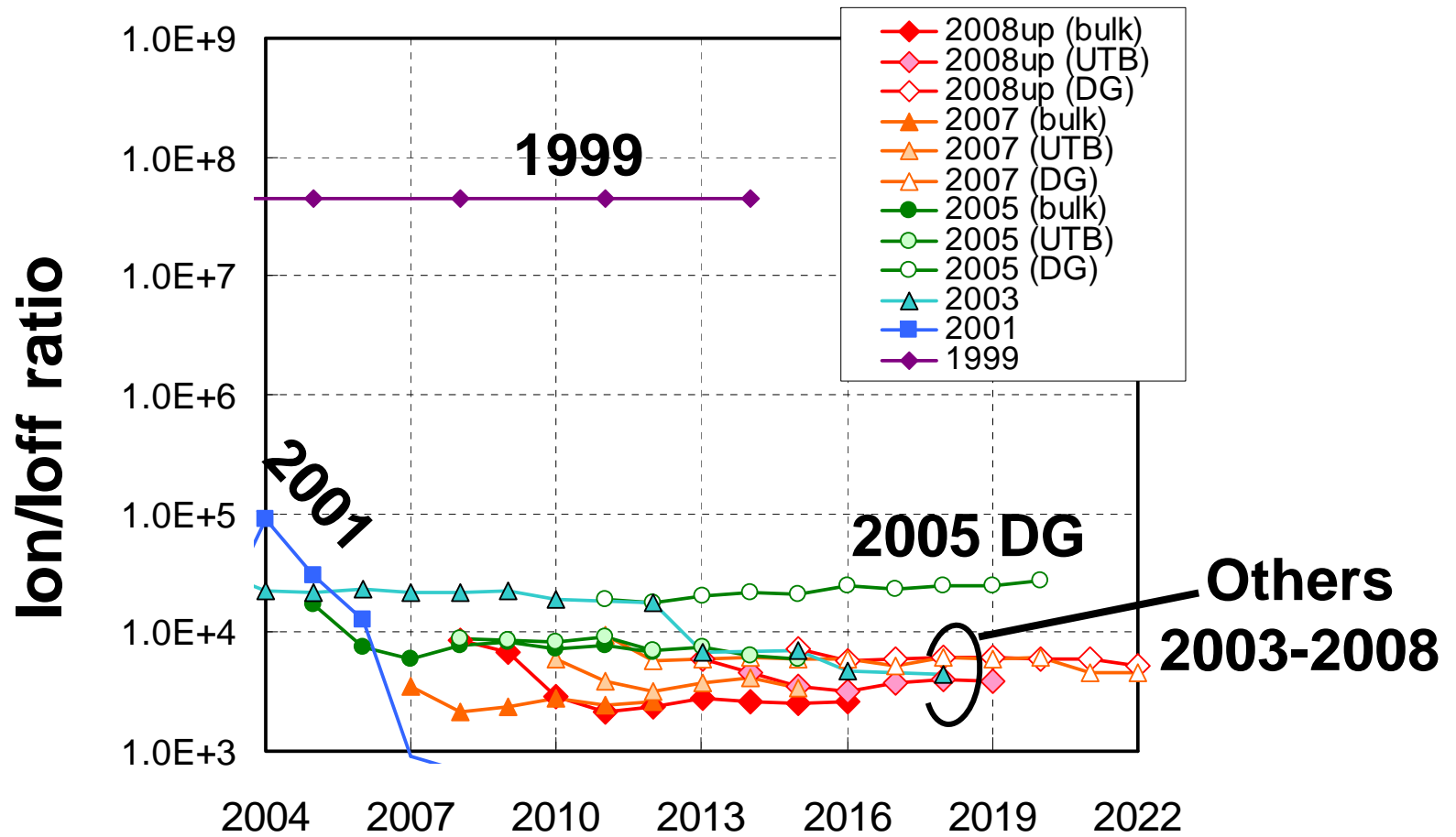


Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion

ITRS for HP logic

Ion/loff ratio

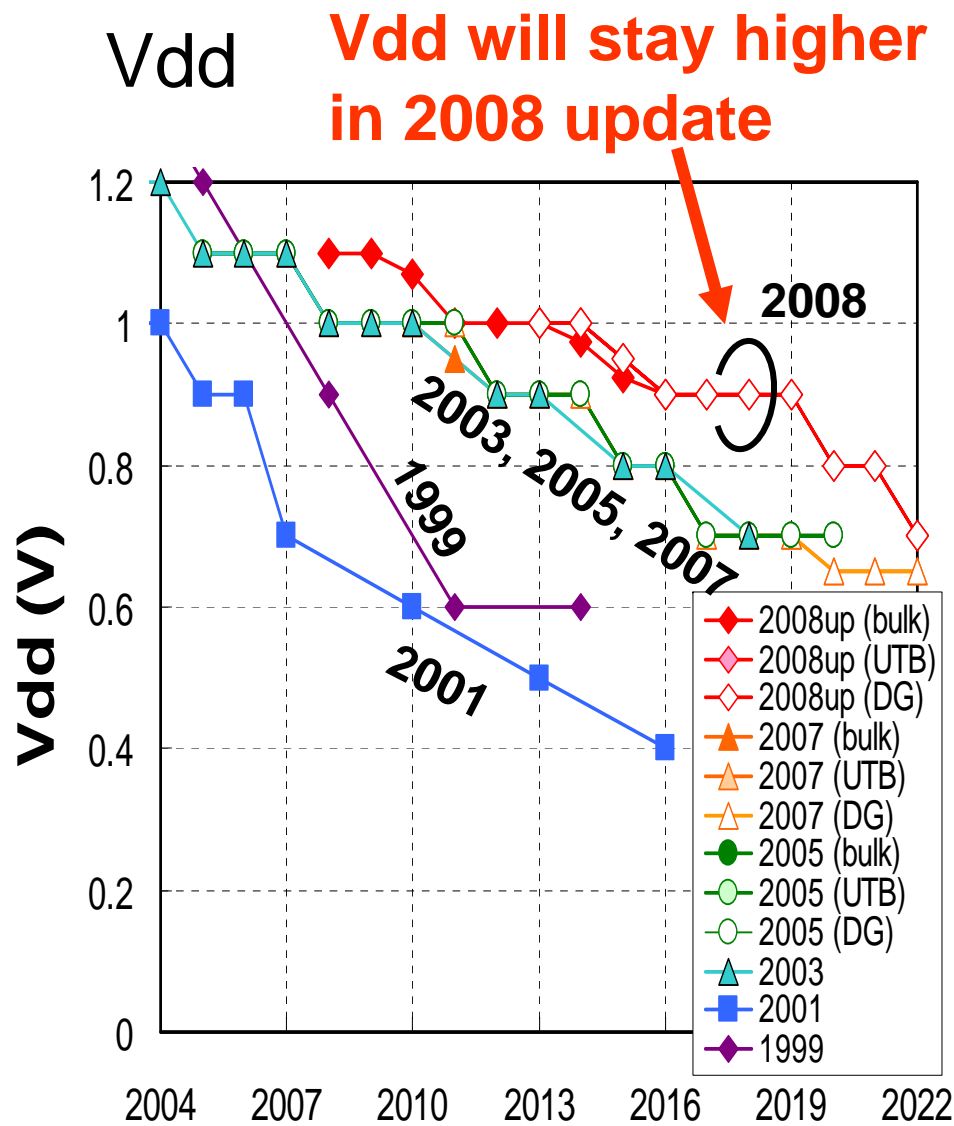


Source: ITRS and
2008 ITRS Summer Public Conf.

Year

2008 Values are from ITRS Public Conf.
and still under discussion

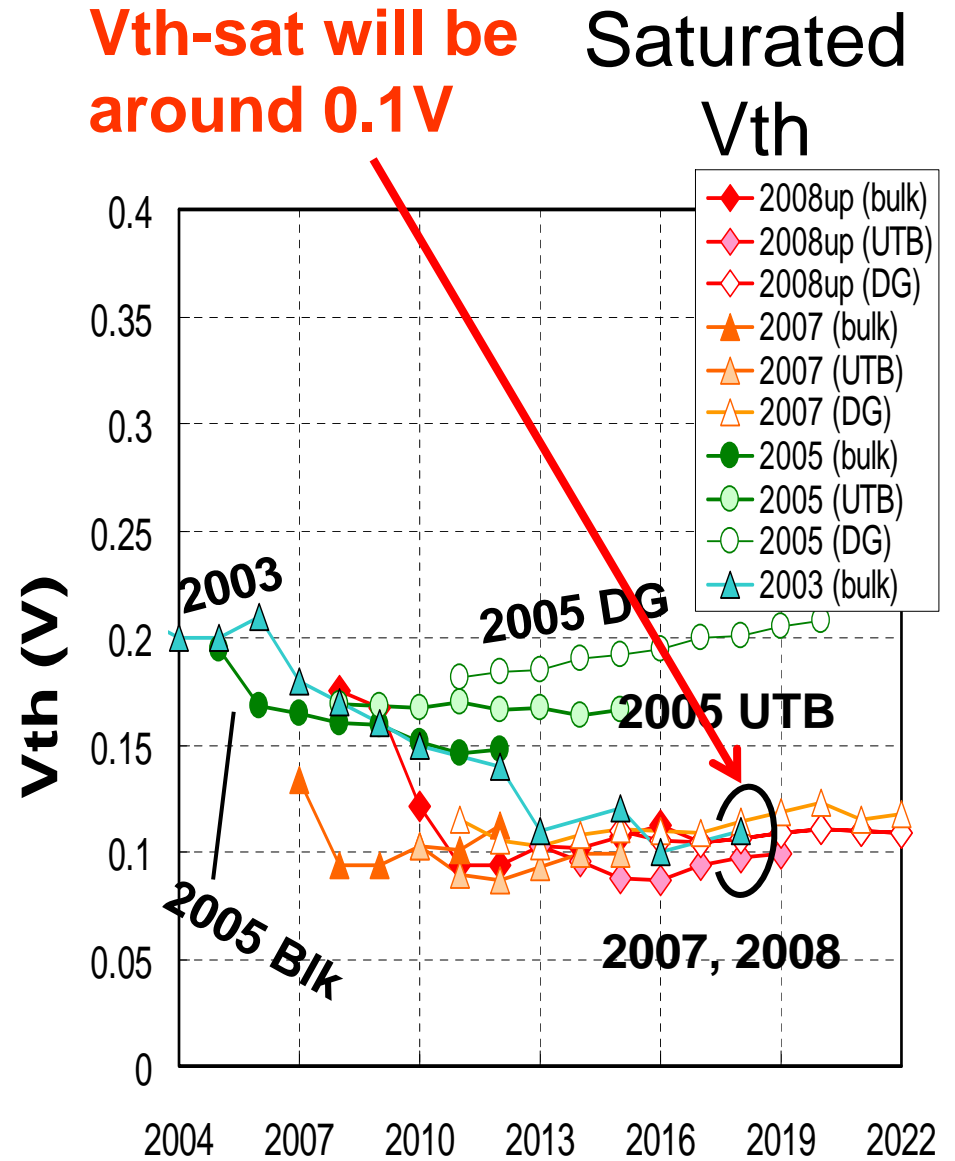
ITRS for HP logic



2008 Values are from ITRS Public Conf. and still under discussion

Year

Source: ITRS and 2008 ITRS Summer Public Conf.

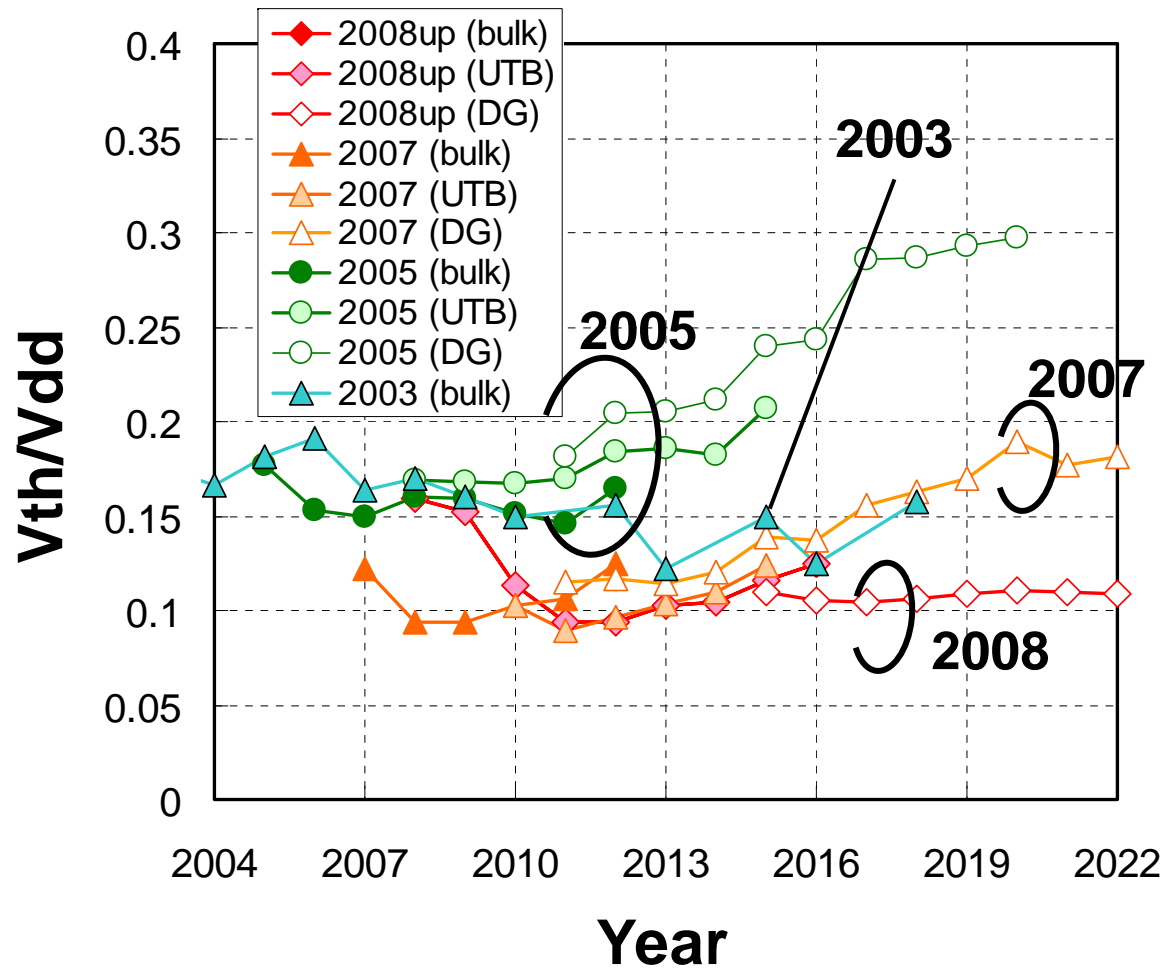


Year

ITRS for HP logic

2008 Values are from ITRS Public Conf.
and still under discussion

V_{th}-sat / V_{dd}

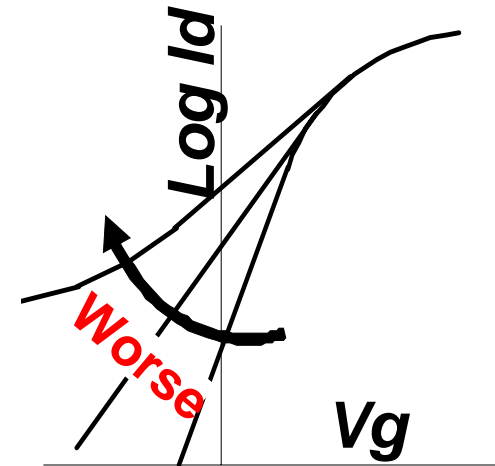


Source: ITRS and
2008 ITRS Summer Public Conf.

SS (Subthreshold Slope) becomes **worse** in the following cases

1. **Improper down-scaling**

Ex. When T_{ox} , W_{dep} , or V_{dd} is not scaled



2. **High impurity doping in channel or substrate**

High impurity Conc.

→ C_D increase

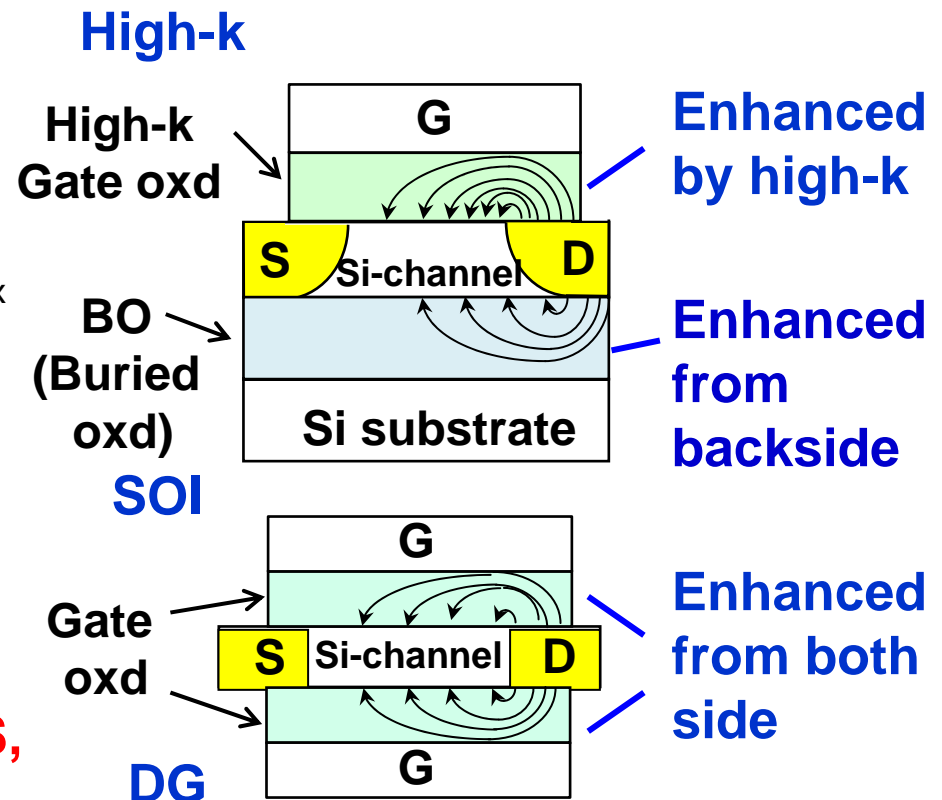
→ SS increase

$$SS = (\ln 10)(kT/q)(C_{ox} + C_D + C_{it})/C_{ox}$$

3. **Enhanced Drain-Electric-field penetration through oxide**

Ex. High-k, SOI, Multi-gate (Double gate: DG)

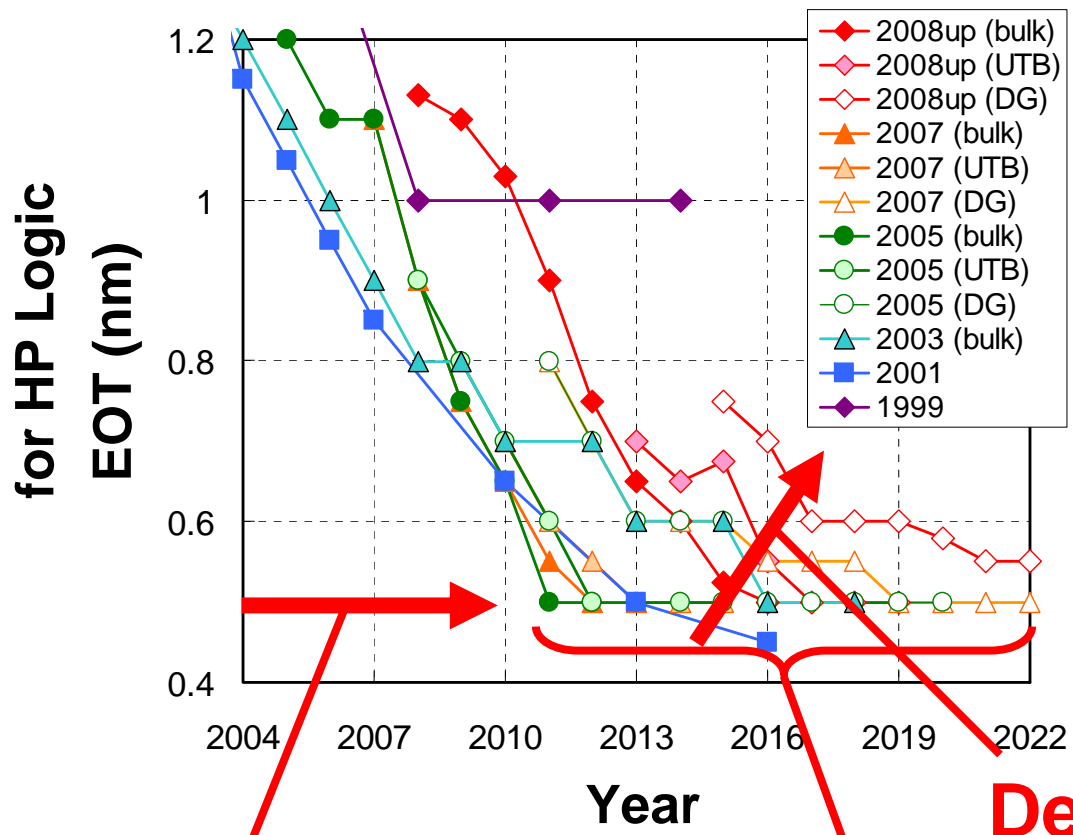
DG and SOI often show better SS, but be careful!



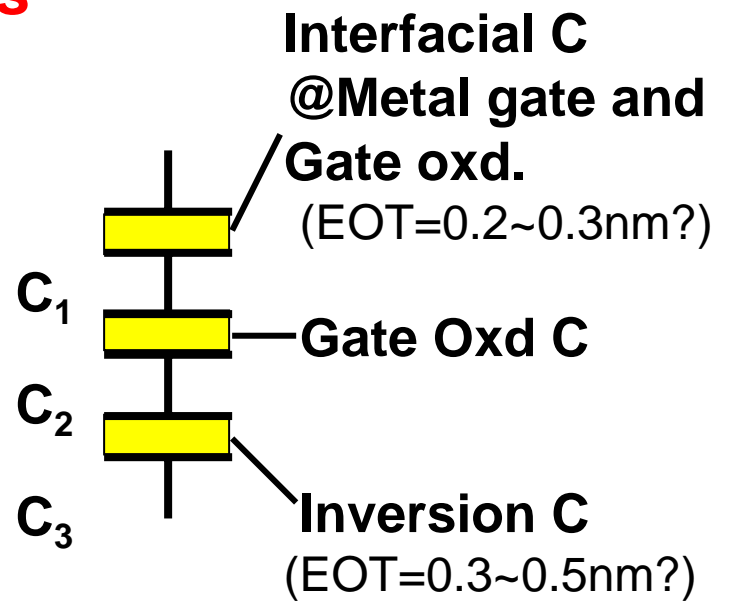
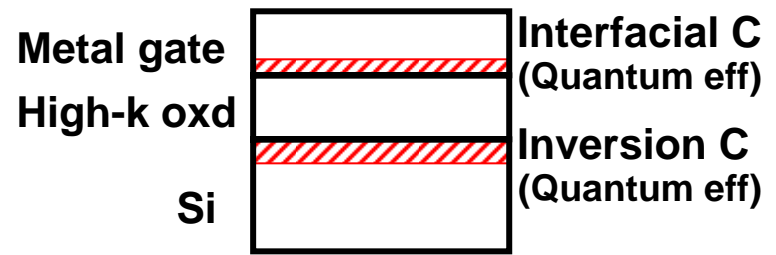
Improper down-scaling

Could we squeeze technologies for ultimate CMOS scaling?

Saturation of EOT thinning is a serious roadblock to proper down-scaling.



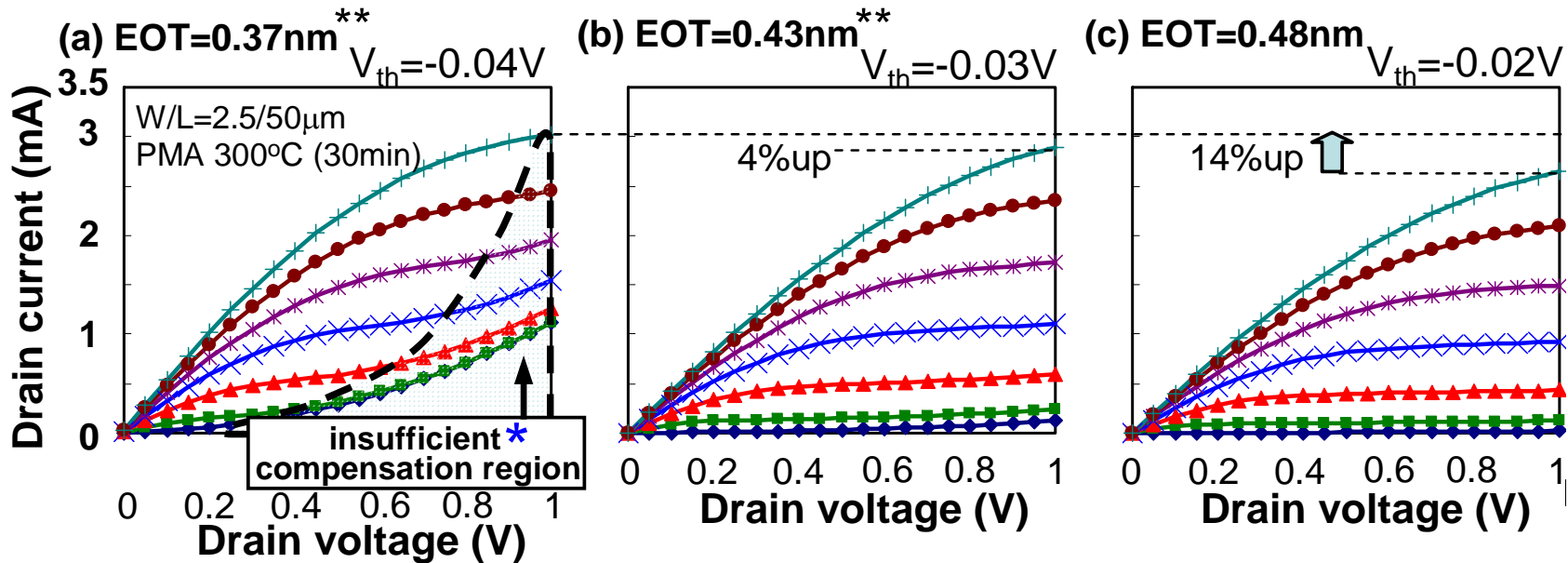
Is 0.5nm real limit? Saturation



$EOT(C_1) + EOT(C_3) > 0.5nm$
 Small effect to decrease $EOT(C_2)$ beyond 0.5nm?

EOT<0.5nm with Gain in Drive Current is Possible

La₂O₃ gate insulator



EOT scaling below 0.5nm

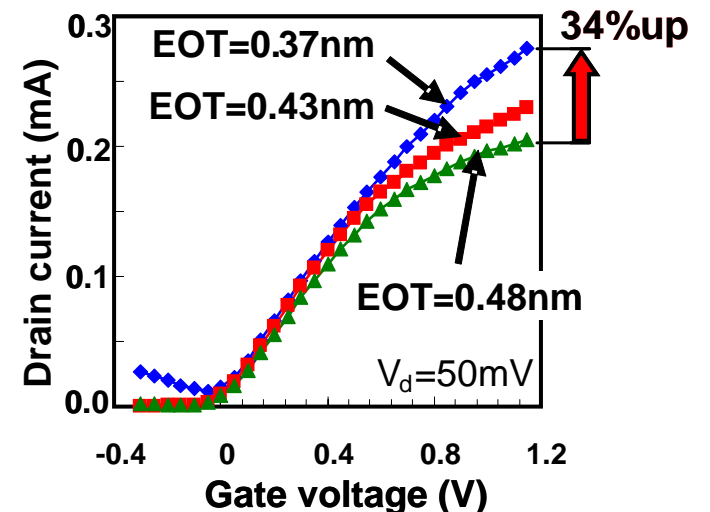


Still useful for larger drain current

Source: K. Kakushima, K. Okamoto, K. Tachi, P. Ahmet, K. Tsutsui, N.i Sugii, T. Hattori, and H. Iwai, IWDTF 2008, Tokyo, November, 2008

* Because L_g is very large (2.5 μ m), gate leakage is large in case (a). The gate leakage component was subtracted from measured data for case (a). However, if we make small gate length, the gate leakage current should become sufficiently small to be ignored compared with I_d as we verified with SiO₂ gate before (Momose et al., IEDM 1994). The gate leakage could be suppressed by modifying material and process in future.

** Estimated by I_d value



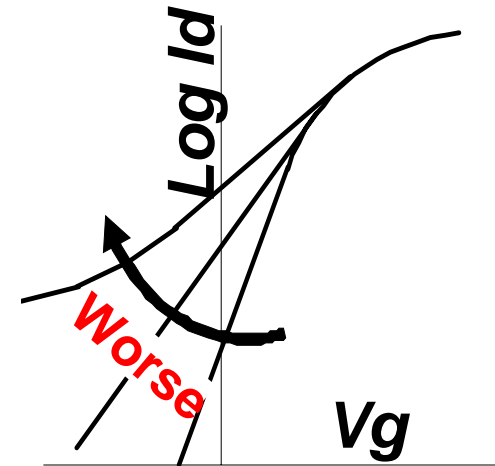
Thus, in future, maybe continuous development of new techniques could make more proper down-scaling possible.

It is difficult to say, but EOT and Vdd may become smaller than expected today.

SS (Subthreshold Slope) becomes **worse** in the following cases

1. **Improper down-scaling**

Ex. When T_{ox} , W_{dep} , or V_{dd} is not scaled



2. **High impurity doping in channel or substrate**

High impurity Conc.

→ C_D increase

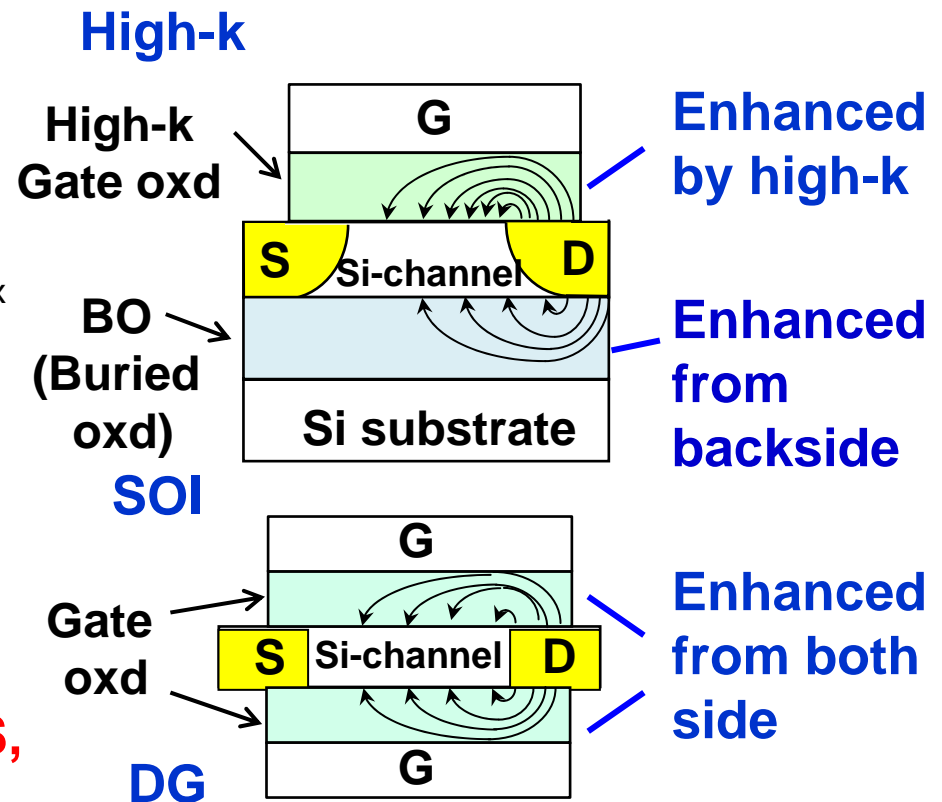
→ SS increase

$$SS = (\ln 10)(kT/q)(C_{ox} + C_D + C_{it})/C_{ox}$$

3. **Enhanced Drain-Electric-field penetration through oxide**

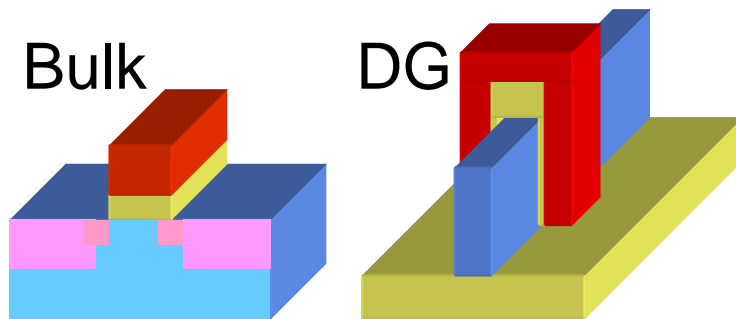
Ex. High-k, SOI, Multi-gate (Double gate: DG)

DG and SOI often show better SS, but be careful!



Enhanced D-Electric-field

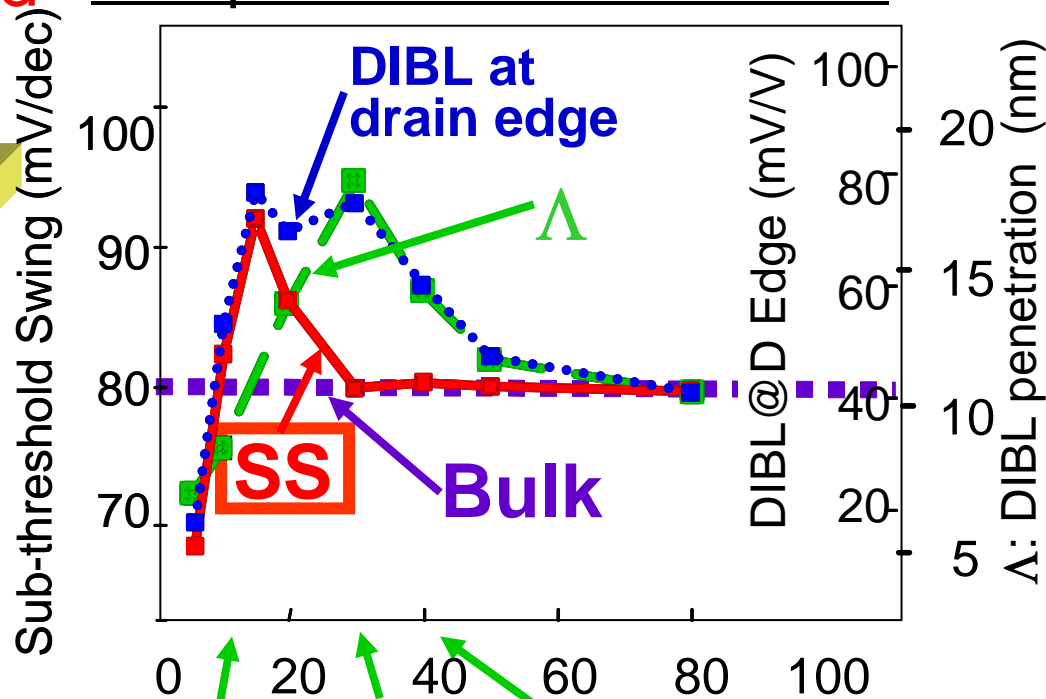
Comparison of Bulk and DG



Same parameter condition for both
(2006 ITRS Bulk parameters are used
for both Bulk and DG)

$L_g = 16\text{nm}$, $t_{ox}(EOT) = 0.5\text{nm}$,
 $Dopant@Channel = 8.1 \times 10^{18}\text{cm}^{-2}$

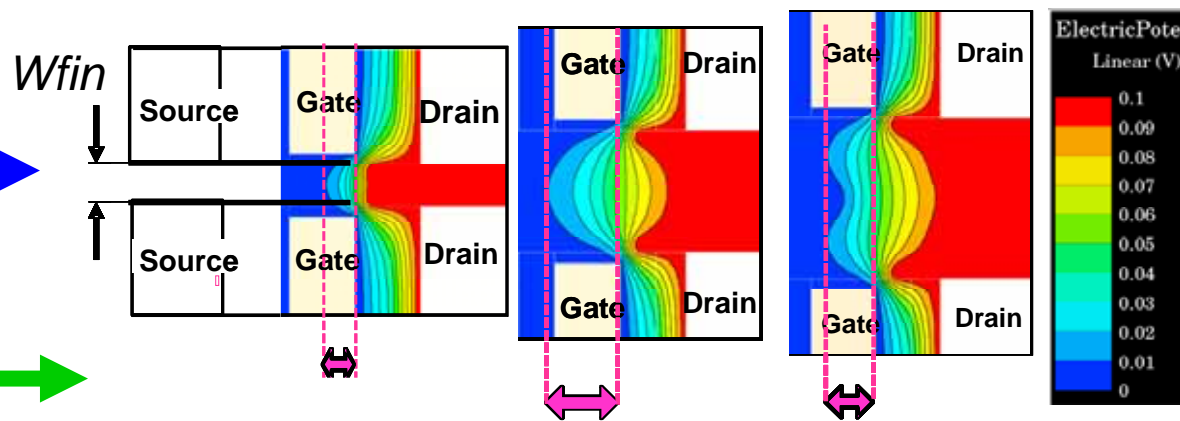
Source: ECS Fall Meeting, Oct 2008, Honolulu,
Y. Kobayashi, A. B. Sachid, K. Tsutsui, K. Kakushima,
P. Ahmet, V. Ramgopal Rao and H. Iwai.



DIBL: Drain Induced Barrier Lowering

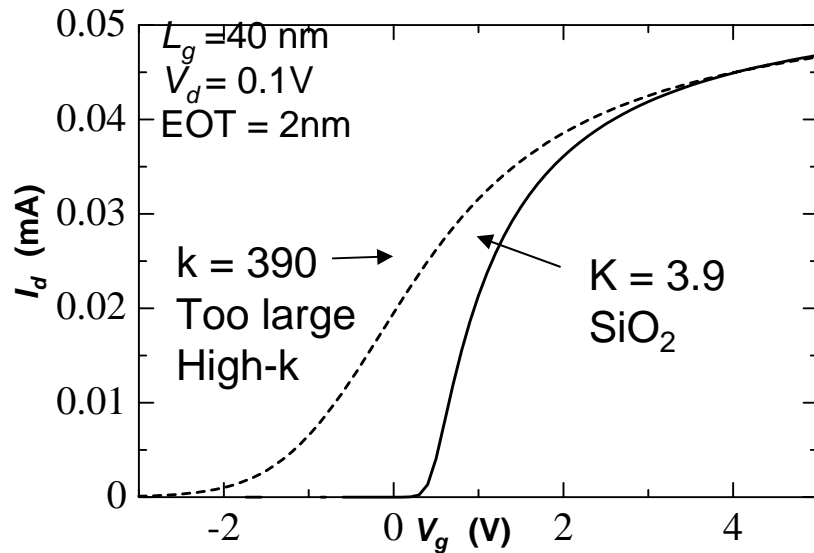
$$\left. \frac{\partial V(x,y)}{\partial V_d} \right|_{V_d=1V}$$

Λ : Penetration Depth of DIBL

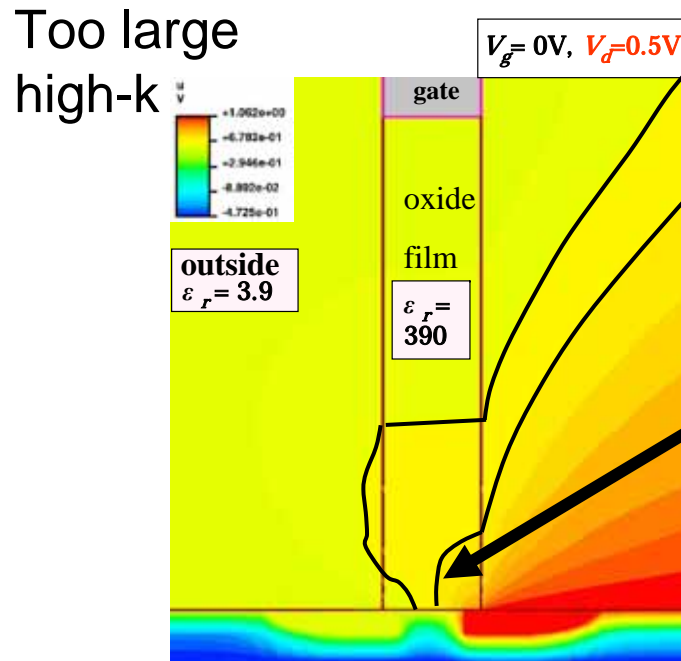
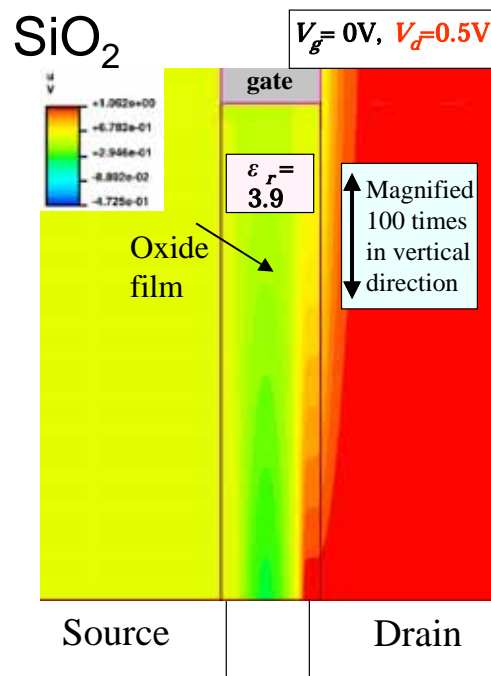
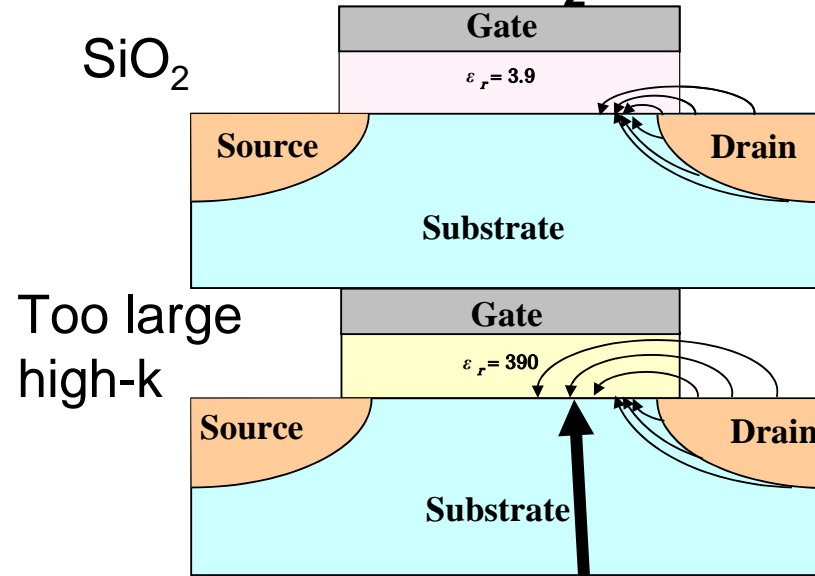


$\Lambda = 7.6\text{nm}$ $\Lambda = 17.1\text{nm}$ $\Lambda = 13.2\text{nm}$ 48

Enhanced D-Electric-field



Comparison of High-k and SiO₂ MOSFETs



Penetration of lateral field from Drain through high-k causes significant short channel effects

R. Fujimura, M. Takeda, K. Sato, S. Ohmi, H. Ishiwara, and H. Iwai, ECS Symp. on ULSI Process Integration II, Volume 2001-2, pp.313-323, 2001,

V_{dd} will stay higher than predicted by previous ITRS roadmaps.

Solution towards Low V_{dd}

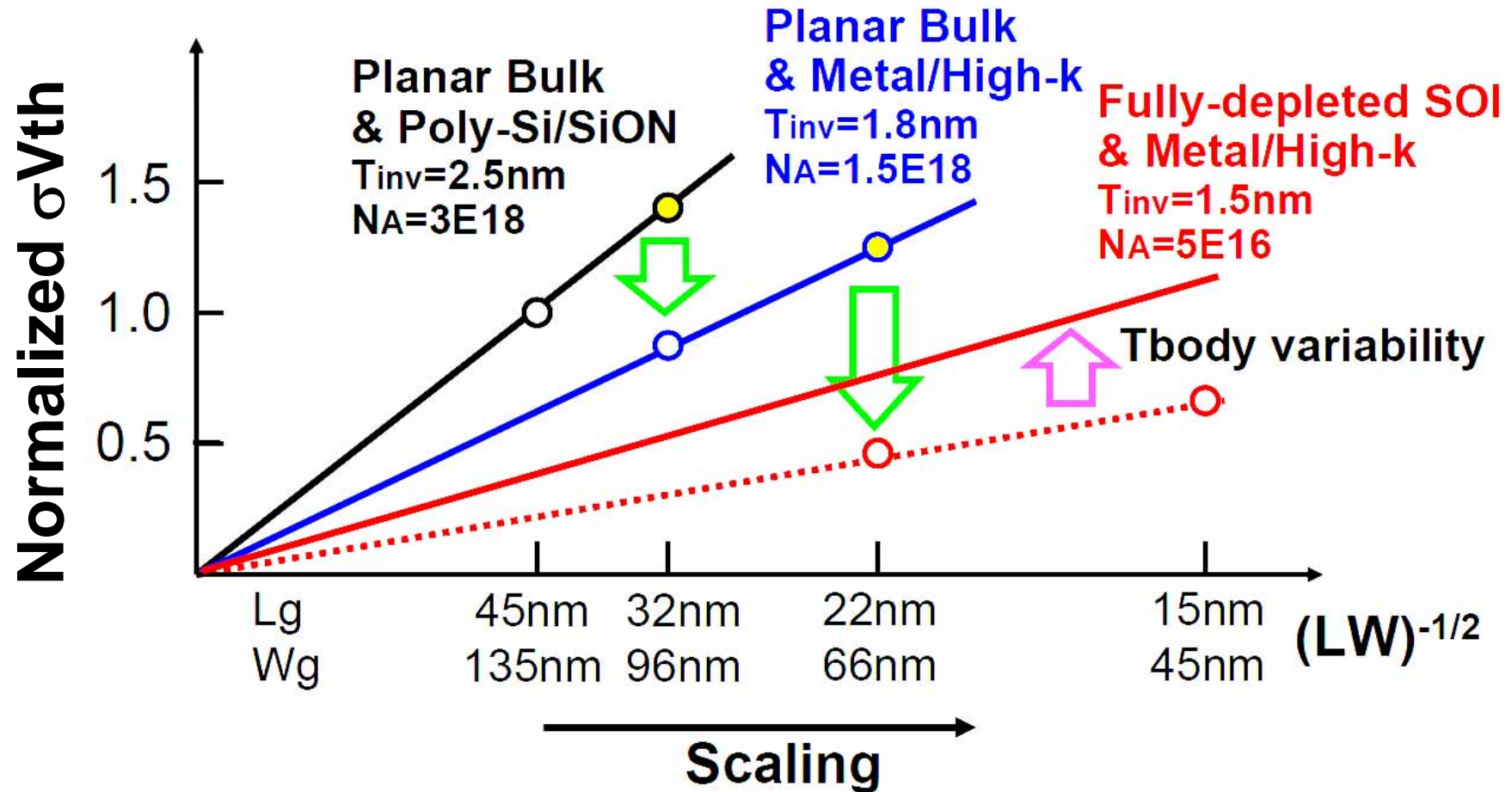
Effort to reduce $I_{sd-leak}$ and increase I_{d-sat} is important

- Scaling: Proper down-scaling
 - Introduction of Next generation high-k, S/D etc.
 - CD* variation control by lithography and etching techniques

* **CD: Critical dimension**

- Structure: Bulk → UTB-SOI → DG → Nanowire
- Variation: Proper scaling by new tech. – High-k, litho. Etc.
 V_{th} adjustment by V_{sub} control
- Circuit techniques: Dynamic and local Multi- V_{dd} , etc.

Random Variability Reduction Scenario in ITRS 2007



Assumption: Random dopant fluctuation is Main source of Random Variability. Line width roughness of L_g and W_g is not considered in this

Source: 2007 ITRS Winter Public Conf.

4. SRAM cell scaling

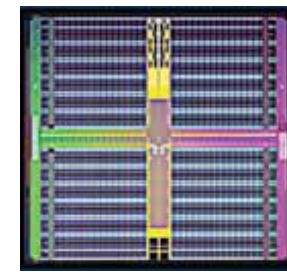
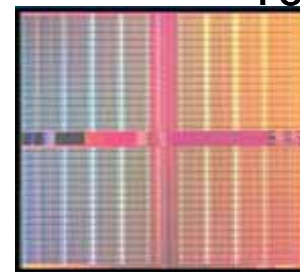
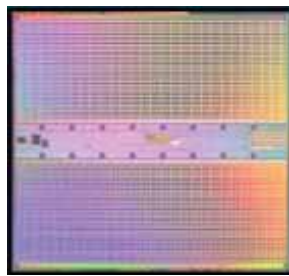
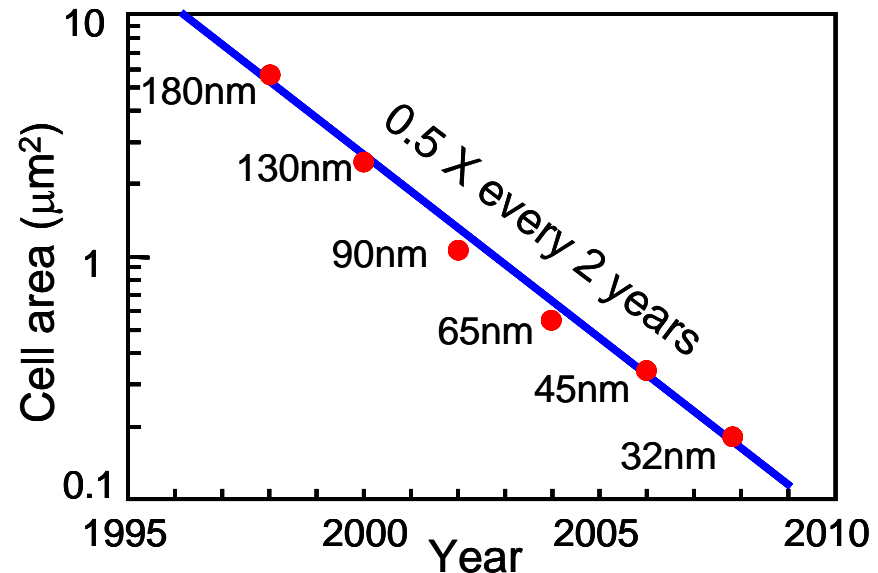
Intel's SRAM test chip trend

Source: B. Krzanich, S. Natrajan, Intel Developer's Forum 2007
http://download.intel.com/pressroom/kits/events/idffall_2007/Briefing_Silicon&TechManufacturing.pdf

SRAM down-scaling trend has been kept until 32nm and probably so to 22nm

Process name	Lithography	1 st production
P1264	65nm	2005
P1266	45nm	2007
P1268	32nm	2009
P1270	22nm	2011

↪ Only schedule has been published



Technology	90 nm Process	65 nm Process	45 nm Process	32 nm Process
Cell size	1.0 µm ² cell	0.57 µm ² cell	0.346 µm ² cell	0.182 µm ² cell
Capacity	50 Mbit	70 Mbit	153 Mbit	291 Mbit
Chip area	109 mm ²	110 mm ²	119 mm ²	118 mm ²
Functional Si	February '02	April '04	January '06	September '07

22 nm technology 6T SRAM Cell: Size = $0.1\mu\text{m}$

Source: <http://www-03.ibm.com/press/us/en/pressrelease/24942.wss>

Announced on Aug 18, 2008

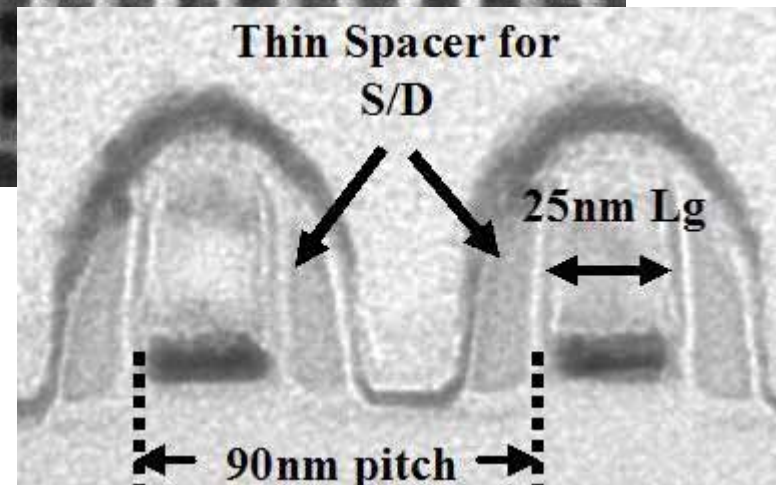
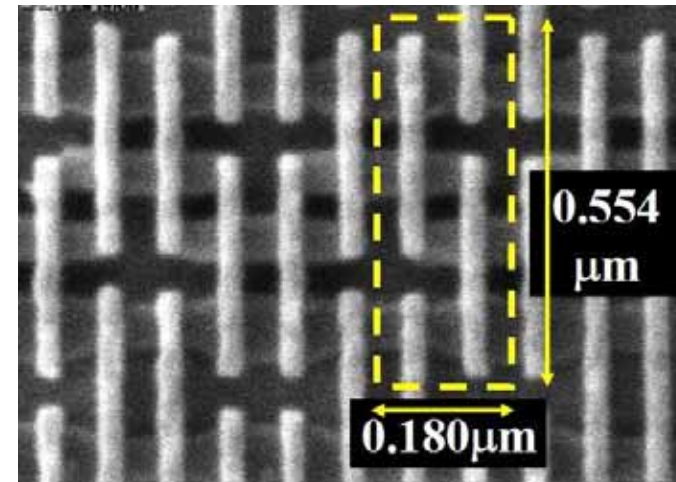
Consortium: IBM (NYSE) , AMD, Freescale, STMicroelectronics, Toshiba and the College of Nanoscale Science and Engineering (CNSE)

$0.1\mu\text{m}$ cell size is almost on the down-scaling trend

New technologies introduced

- High-NA immersion lithography
- High-K metal gate stacks
- **25 nm gate lengths**
- Thin composite oxide-nitride spacers
- Advanced activation techniques
- Extremely thin silicide
- Damascene copper contacts

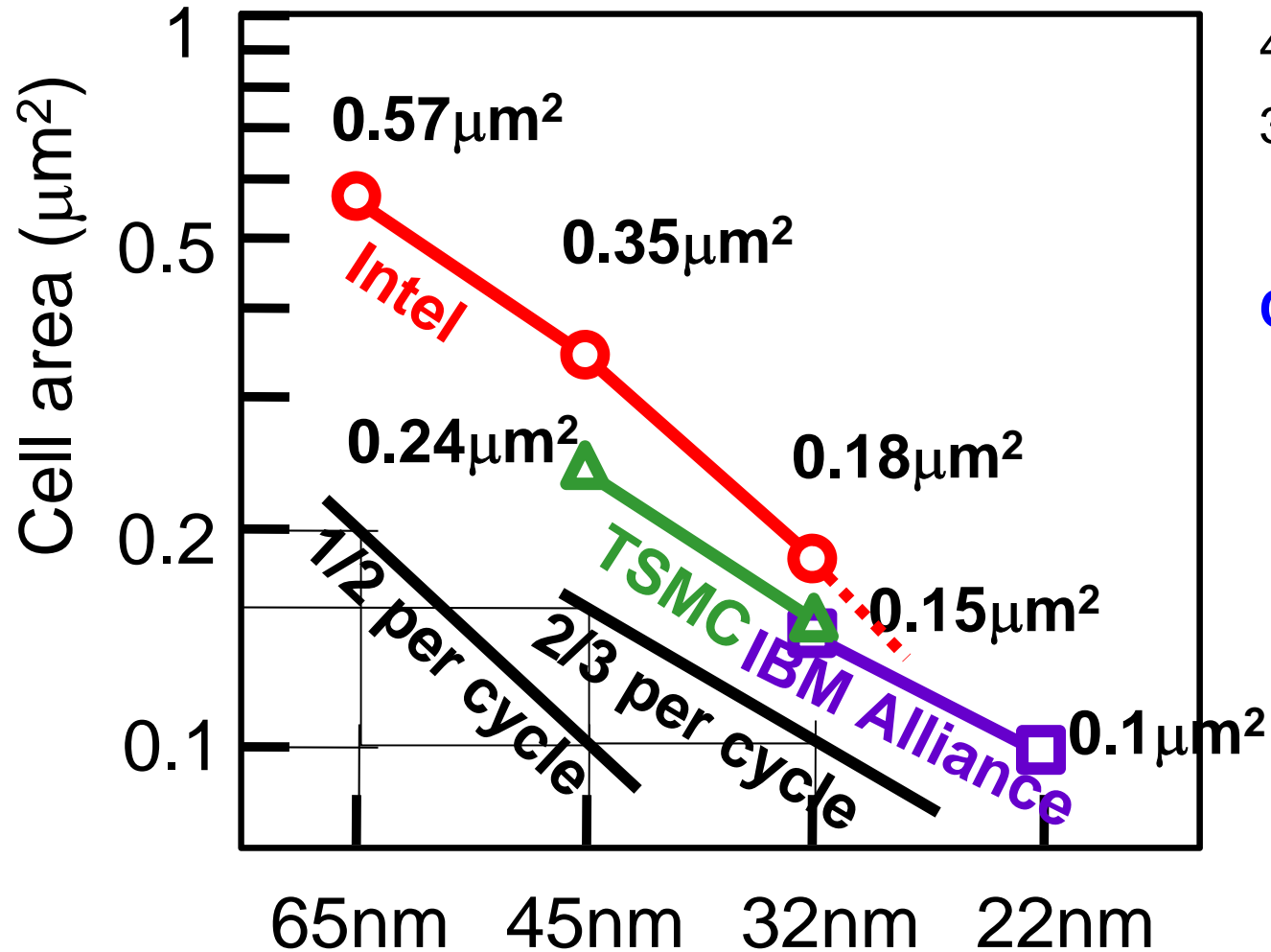
Static noise margin of 220 mV at 0.9 V



Source: IEDM2008 Pre-conference Publicity
<http://www.btbmarketing.com/iedm/>

Cell size reduction trends

1/2 or 2/3 per cycle?



Intel

Functional Si

65nm Apr.2004

45nm Jan.2006

32nm Sep.2007

TSMC

Conference (IEDM)

45nm Dec.2007

32nm Dec.2007

**IBM Alliance
(Consortium)**

Conference (IEDM)

32nm Dec.2007

Press release

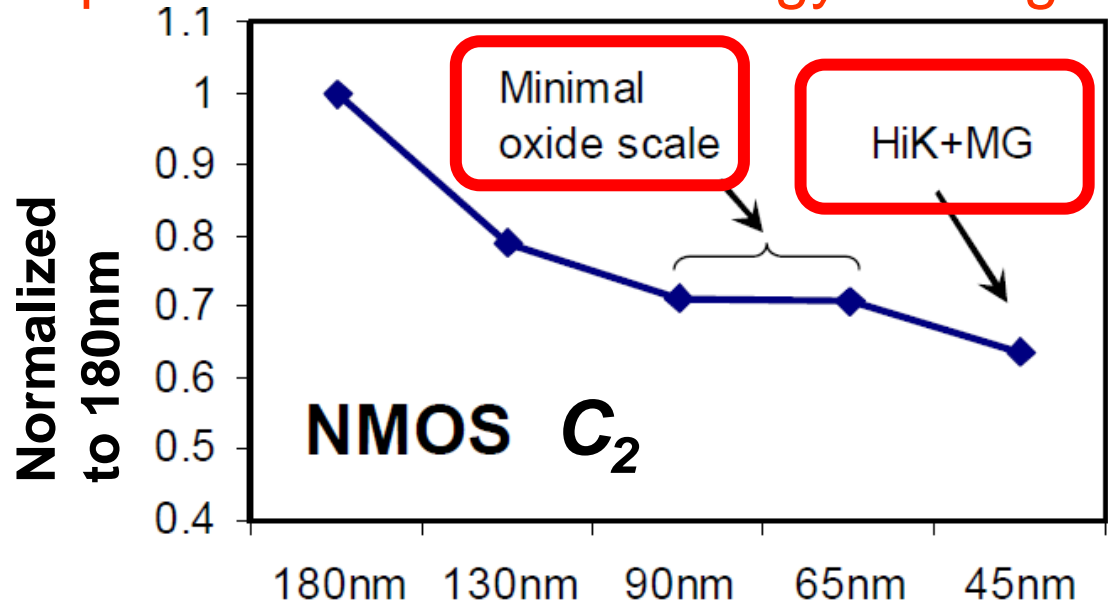
22nm Aug.2008

NMOS Mismatch Coefficient (C_2) improvement with technology scaling

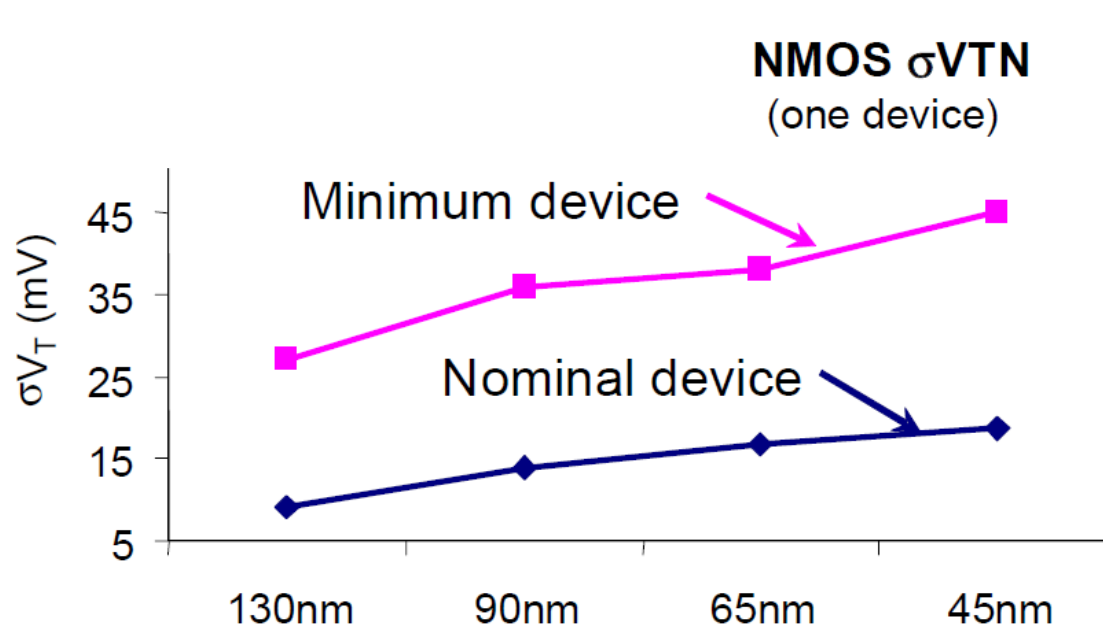
$$\sigma V_{Tran} = \left(\frac{\sqrt[4]{4q^3 \epsilon_{si} \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\epsilon_{ox}} \cdot \left(\frac{\sqrt[4]{N}}{\sqrt{W_{eff} \cdot L_{eff}}} \right)$$

$$= \frac{1}{\sqrt{2}} \left(\frac{C_2}{\sqrt{W_{eff} \cdot L_{eff}}} \right)$$

C_2 →



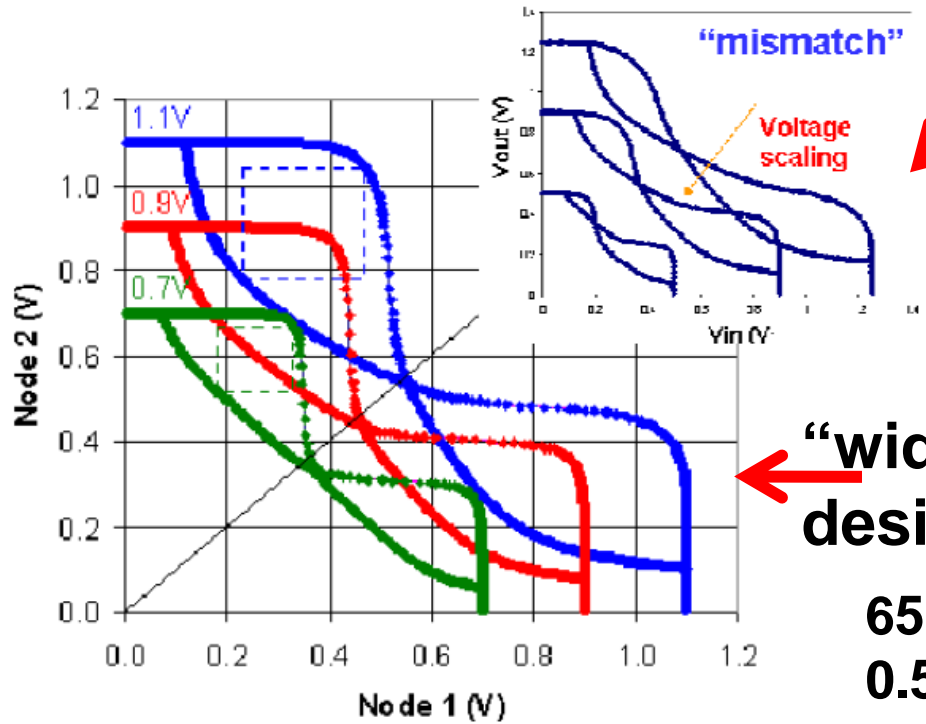
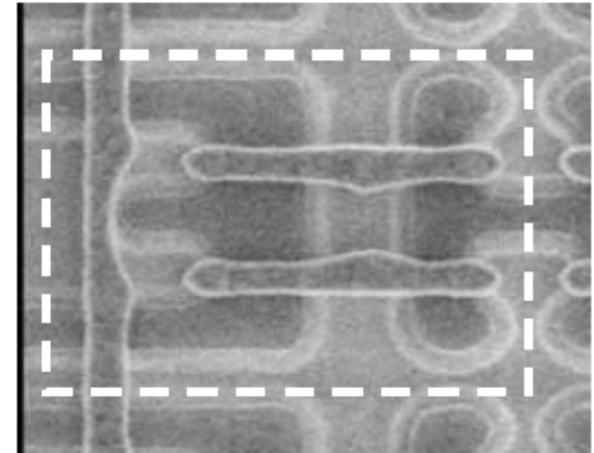
σV_{Tran} →



Source: K.J.Kuhn
IEDM 2007

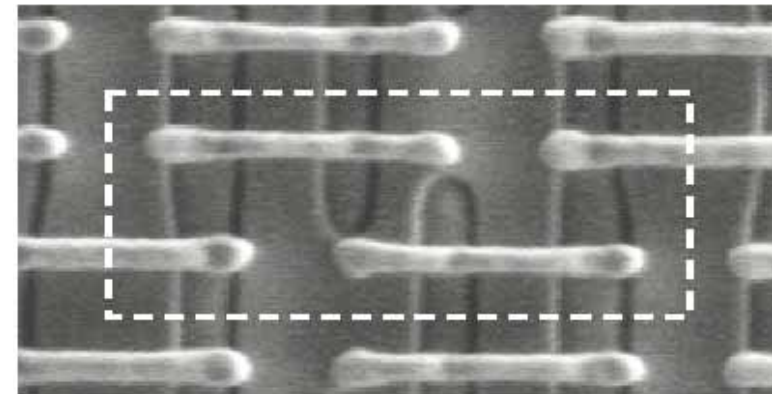
Mismatch improvement by layout (Intel)

“tall” design
90nm : 1.0 μm^2



“wide” design

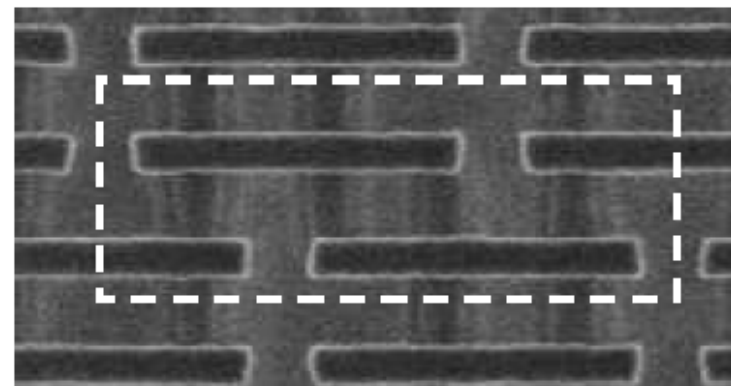
65nm :
0.57 μm^2



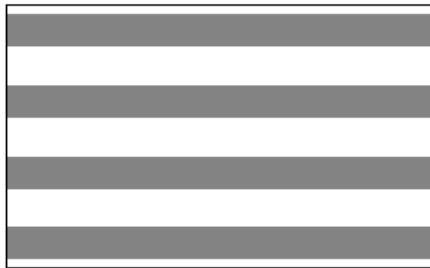
Source: K. J. Kuhn
IEDM2007 Tech. Dig. pp.471

“wide” design
(Square endcaps)

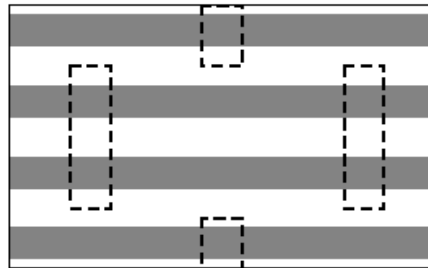
45nm 0.346 μm^2



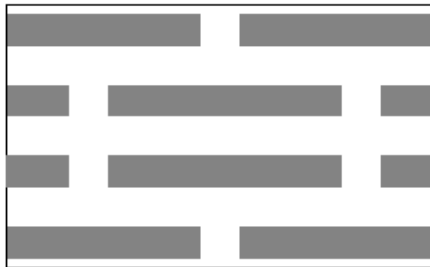
Double patterning for square endcap



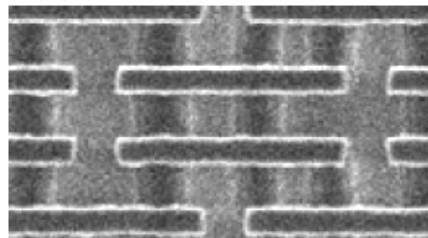
a) Pattern gate lines/spaces



b) Pattern cut mask



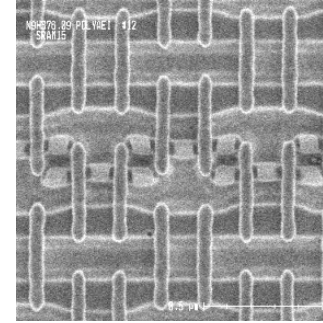
c) Final gate pattern



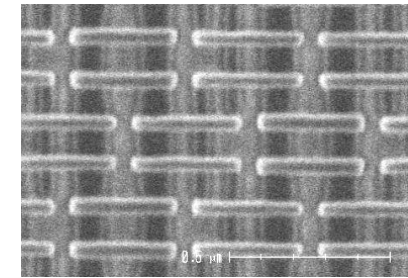
d) Intel 45nm SRAM cell

Source: M. Bohr, ICSICT2008

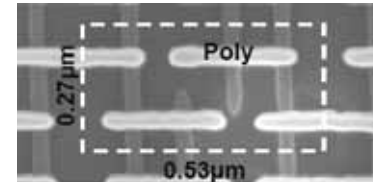
Cell evolution is similar



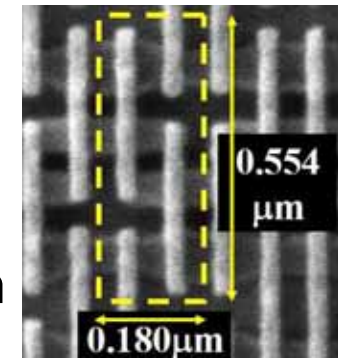
TSMC 45nm
IEDM 2007



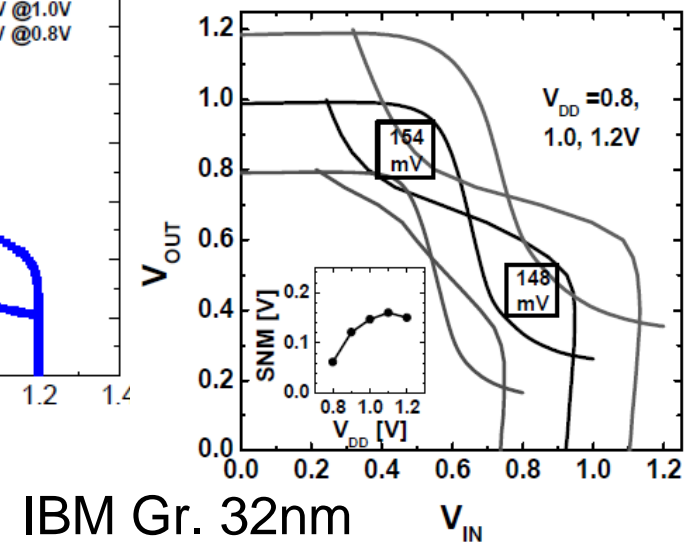
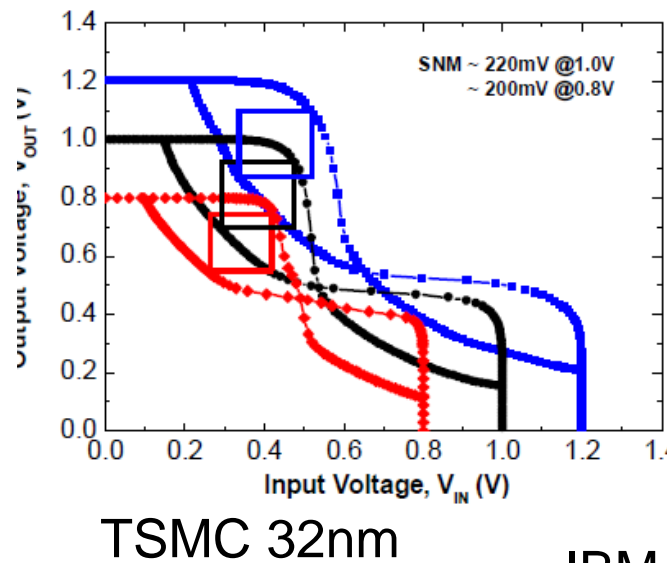
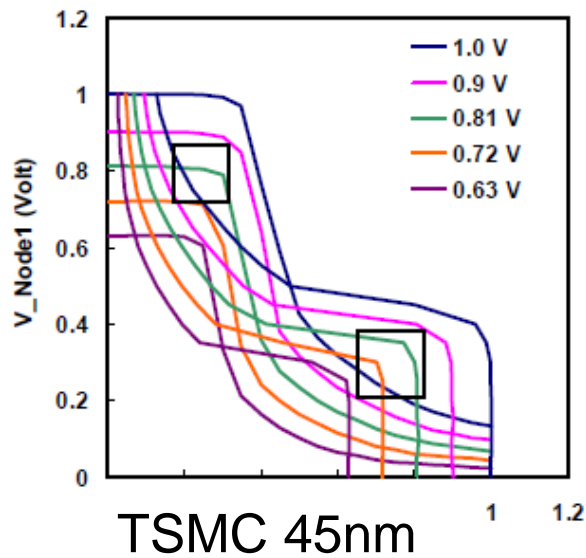
TSMC 32nm
IEDM 2007



IBM Alliance 32nm
IEDM 2004



IBM Alliance 22nm
IEDM 2008



Most Difficult part of SRAM down-scaling is
Vdd down-scaling

Density of on-chip cache SRAM memory is high
and thus, V_{th} cannot be down-scaled too much
because of large I_{sd} -leak

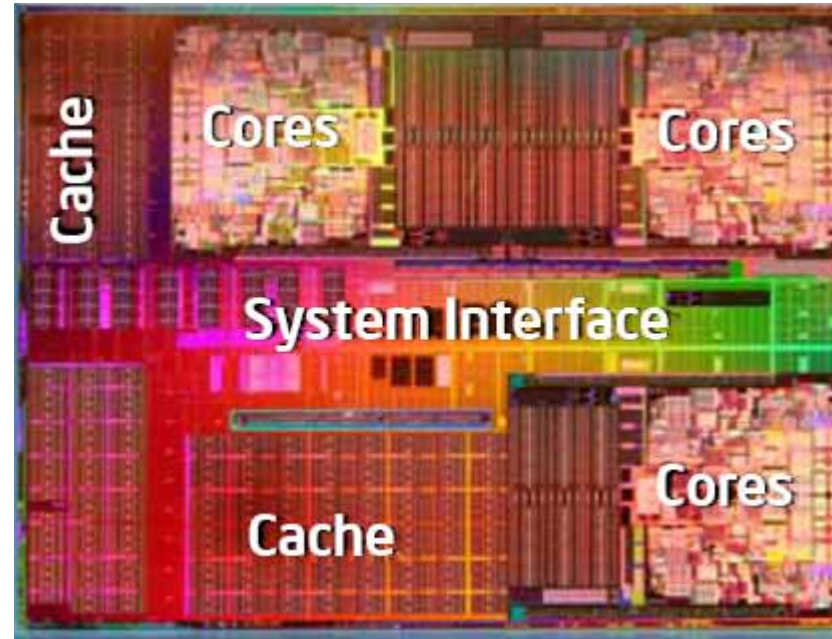
Also, under low Vdd, read- and write margin
degrades, data retention degrade.

Thus, Vdd down-scaling is more severe in SRAM
than logic part of the circuits

Intel® Xeon® 7400 Series (Dunnington)

45 nm high-k6 cores
16MB shared L3 cache

[Source: Intel Developer Forum 2008](#)



Cache occupies huge area

- Cell size of SRAM should be minimized
- Isd-leak should be minimized
 - Vth are often designed to be higher than Min. logic Vth
 - Lg are often designed to be larger than Min. logic Lg

Future Directions For Improving Vmin

- **Application**
 - Improvement in voltage and temperature tolerance
- **Package**
 - Separated array / logic voltage to minimize logic noise effect on SRAM
- **Design**
 - Higher array VDD and improved on-chip supply robustness
 - Increased redundancy
 - Improved timings
 - Cells per BL hierarchical BL structure
 - Write/Read assist and sense-amp design
- **Cell and Process**
 - Improved bit cell optimization
- **NFET/PFET centering and Beta/Gamma control**
- **Minimize device fluctuation by limiting device-geometry scaling**
larger cell
- **Lpoly, Weff, LER**
 - Leakage / defect mechanisms

Source: Harold Pilo IEDM2006 Short Course

Nehalem(Intel) 2,4 or 8 Cores

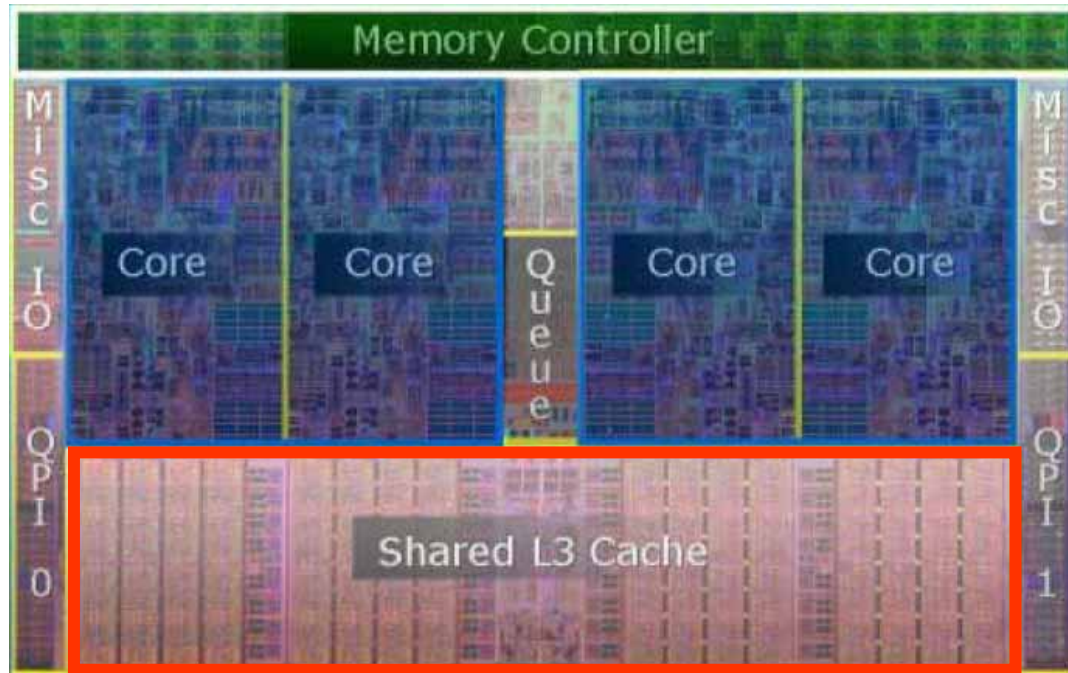
Voltage/Frequency Partitioning

- DDR Vcc
- Core Vcc
- Uncore Vcc

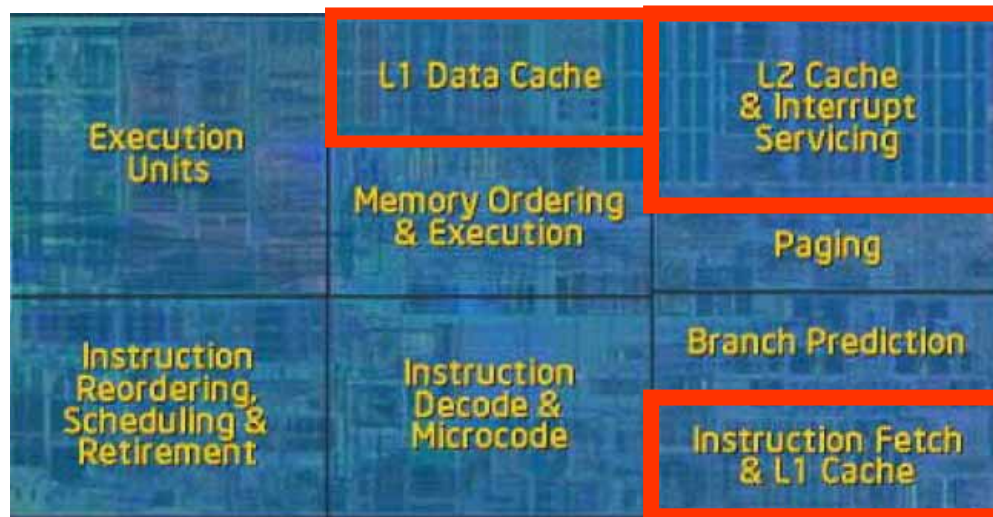
Dynamic Power Management

- 8T SRAMCell**
- 32kB L1 I-cache
- 32kB L1 D-cache
- 256kB L2-cache

- 6T SRAMCell**
- 8 MB L3 cache



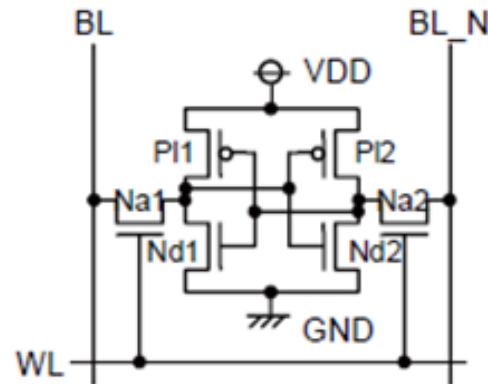
Chip



Core

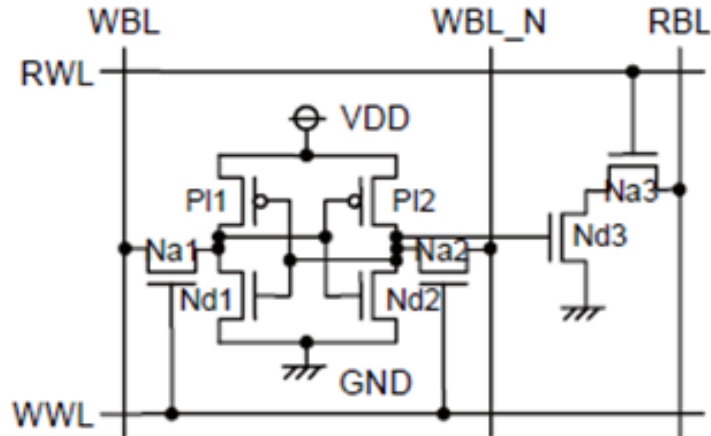
6T and 8T Cell

6T Cell



Cell size is small
For high density use

8T Cell

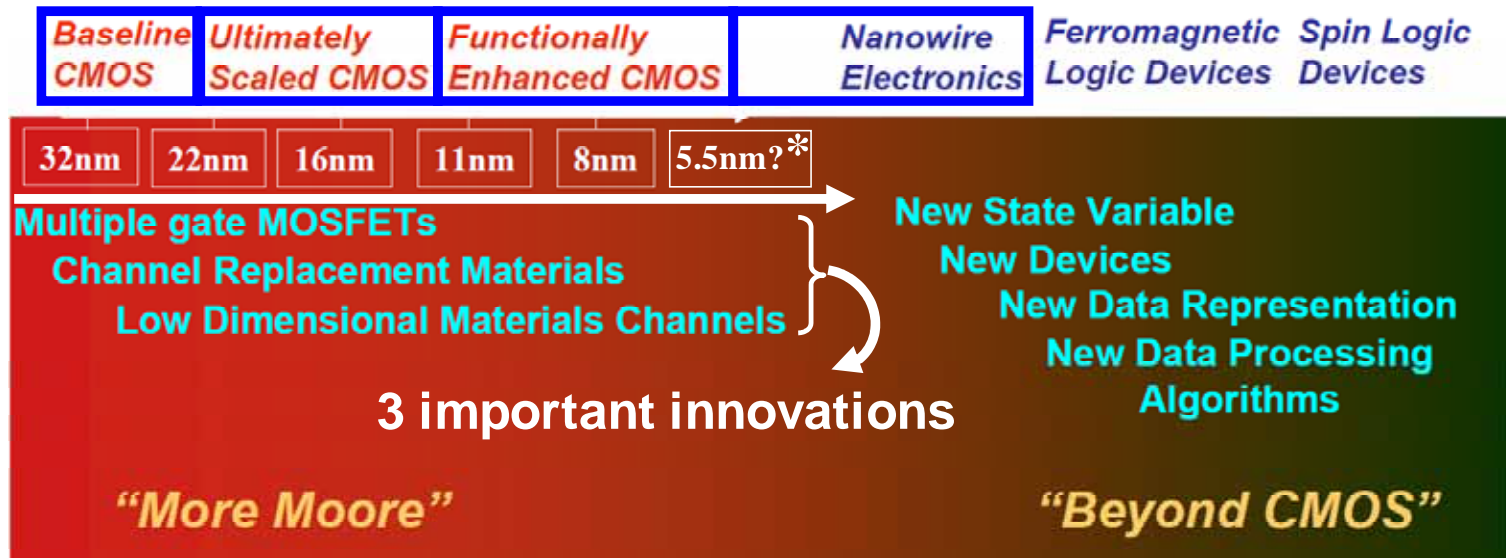


Add separate
read function
Cell size
increase 30%
For low voltage use

Source: Morita et. al, Symp. on VLSI Circ. 2007

5. Roadmap for further future as a Personal View

- There will be still 4~6 cycles (or technology generations) left until we reach 11 ~ 5.5 nm technologies, at which we will reach down-scaling limit, in some year between 2020-30 (H. Iwai, IWJT2008).
- Even After reaching the down-scaling limit, we could still continue R & D, seeking sufficiently higher I_{d-sat} under low V_{dd} .
- Two candidates have emerged for R & D
 1. Nanowire/tube MOSFETs
 2. Alternative channel MOSFETs (III-V, Ge)
- Other Beyond CMOS devices are still in the cloud.



ITRS figure edited by Iwai

Source: 2008 ITRS Summer Public Conf.

* 5.5nm? was added by Iwai

Si nanowire FET with Semi-1D Ballistic Transport

Merit of Si-nanowire

Source: Y. Lee., T. Nagata., K. Kakushima., K. Shiraishi, and H. Iwai, IWDTF 2008, Tokyo, November, 2008

Trade off

Carrier scattering probability

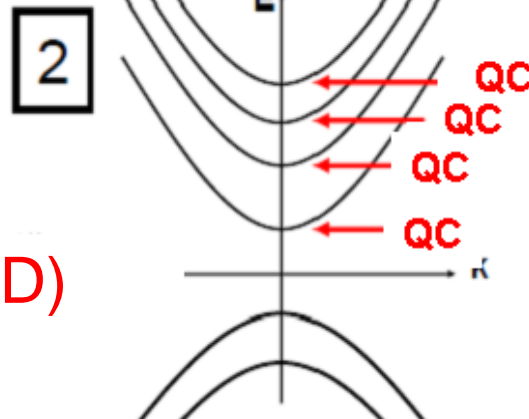
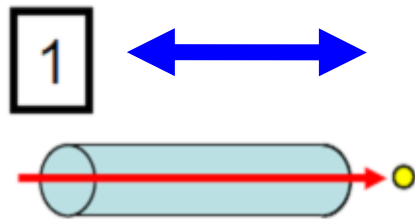
Small

Large

of quantum channel

Small

Large

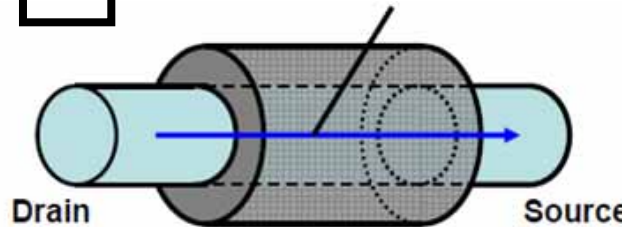


High Conduction (1D)

$G_0 = 77.8 \mu S / \text{wire}$

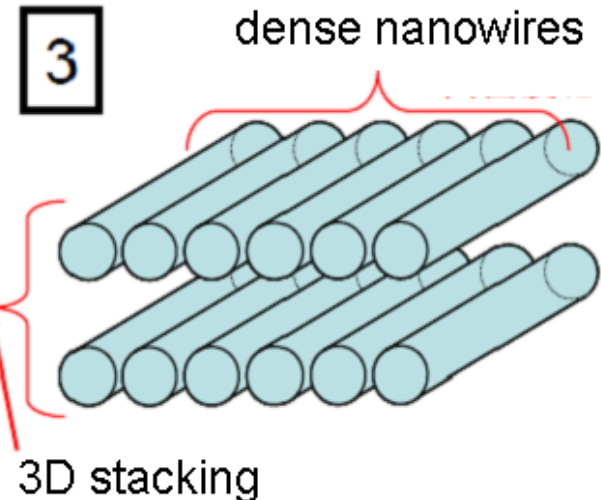
Multiple quantum channel (QC) used for conduction

Reduction in I_{off} (I_{sd} -leak)



Good control of I_{sd} -leak by surrounding gate

Increase in I_{on} (I_{d-sat})



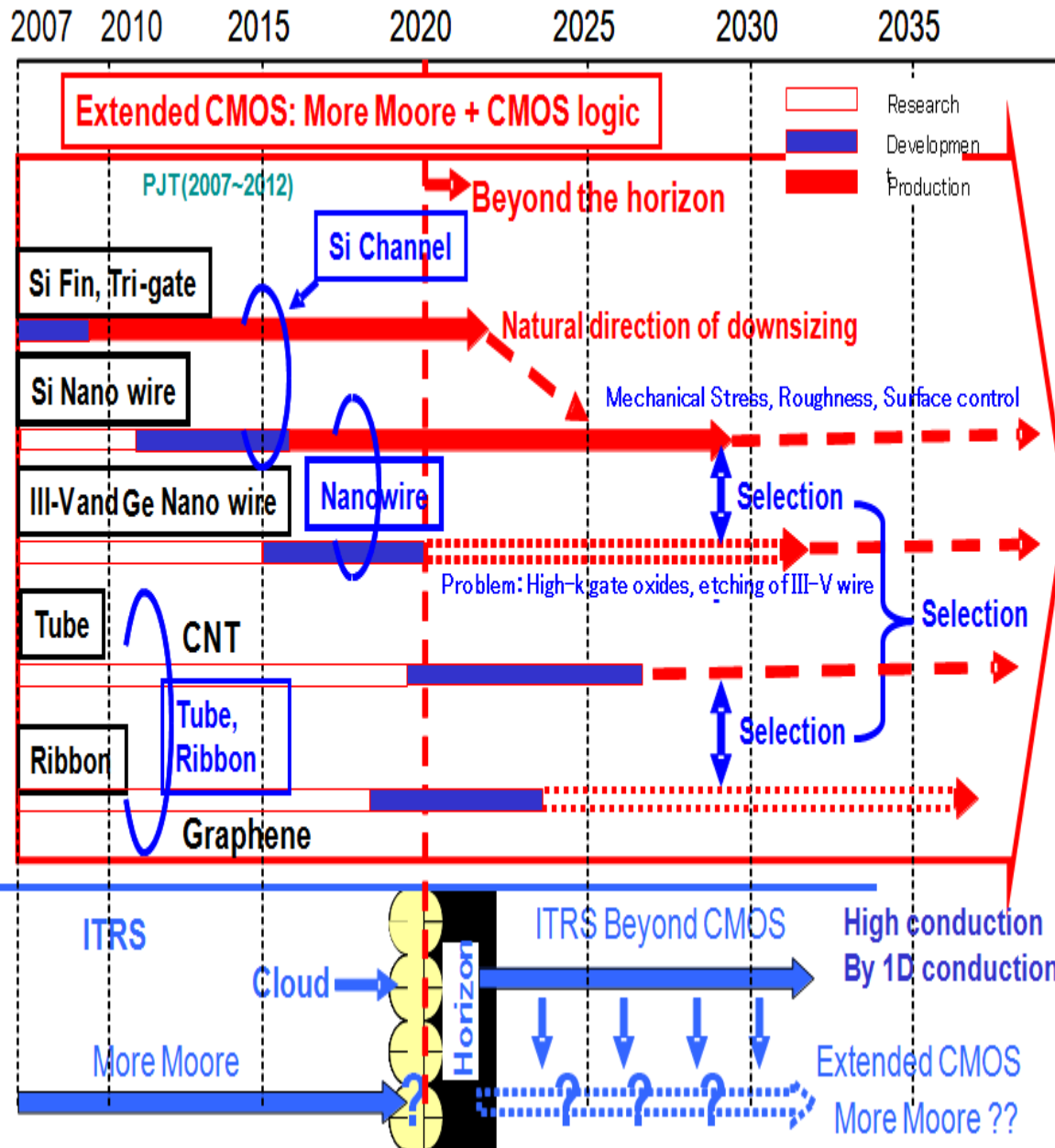
High-density lateral and vertical integration

Source: T. Ohno, K. Shiraishi, and T. Ogawa, Phys. Rev. Lett., 1992

Our roadmap for R & D

Source: H. Iwai, IWJT 2008

Current Issues



Si Nanowire

- Control of wire surface property
- Source Drain contact
- Optimization of wire diameter
- Compact I-V model

III-V & Ge Nanowire

- High-k gate insulator
- Wire formation technique

CNT:

- Growth and integration of CNT
- Width and Chirality control
- Chirality determines conduction types: metal or semiconductor

Graphene:

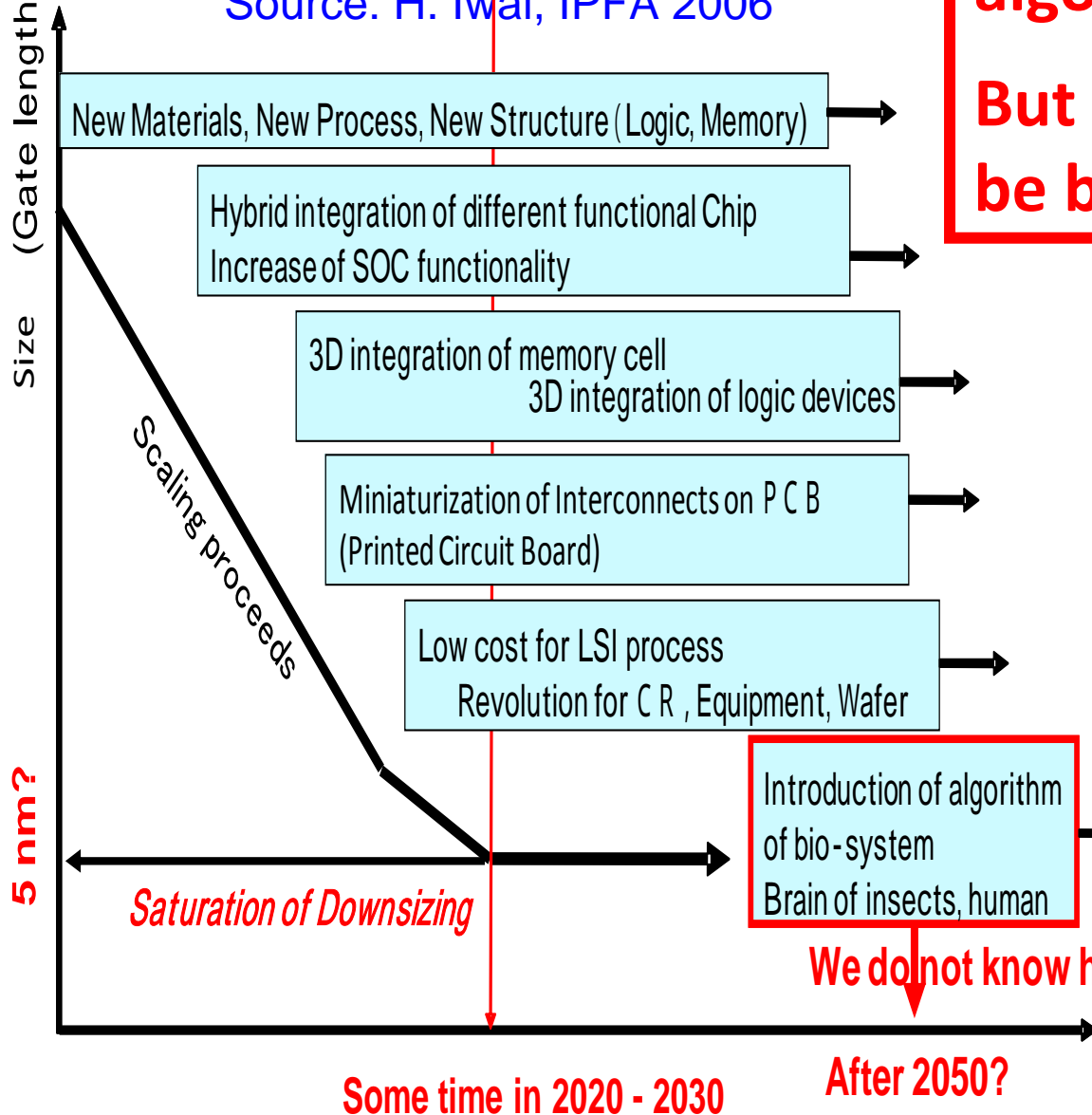
- Graphene formation technique
- Suppression of off-current

Very small bandgap or no bandgap (semi-metal)

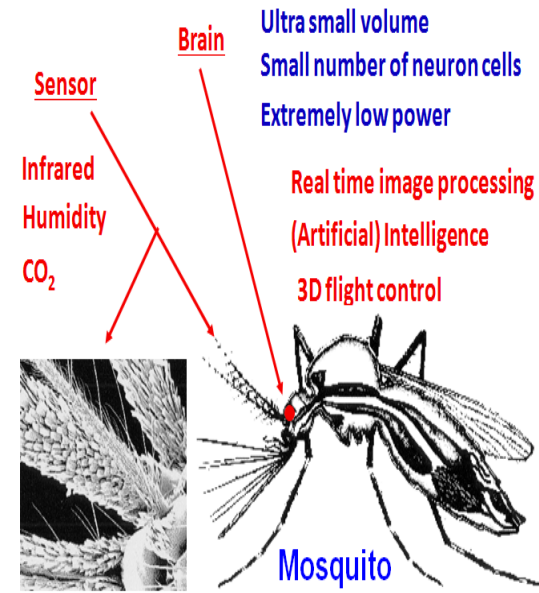
- Control of ribbon edge structure which affects bandgap

Long term roadmap for development

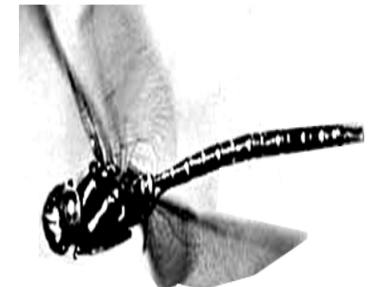
Source: H. Iwai, IPFA 2006



We do know system and algorithms are important!
But do not know how it can be by us for use of bio?



We do not know how? Dragonfly brain has even further higher performance



Acknowledgement

I would like to express deep appreciation to the following people for the useful advice and support for material preparation.

Special thanks to ITRS committee for the permission to refer roadmap and Public conference.

ITRS Committee: Hidemi Ishiuchi (Toshiba), Paolo Gargini (Intel)

Toshiba Corporation: Mitsuo Saito, Yukihiro Urakawa, Tomoaki Yabe

Tsukuba University: Kenji Shiraishi, Kenji Natori

Intel Corporation: Mark Bohr

IBM Alliance : B.S. Haran et al,

Tokyo Institute of Technology: Kuniyuki Kaukshima, Parhat Ahmet,
Takamasa Kawanago, Yeonghun Lee

Thank you
for your attention!