EE 382M VLSI–II

Transistors & Process Scaling:

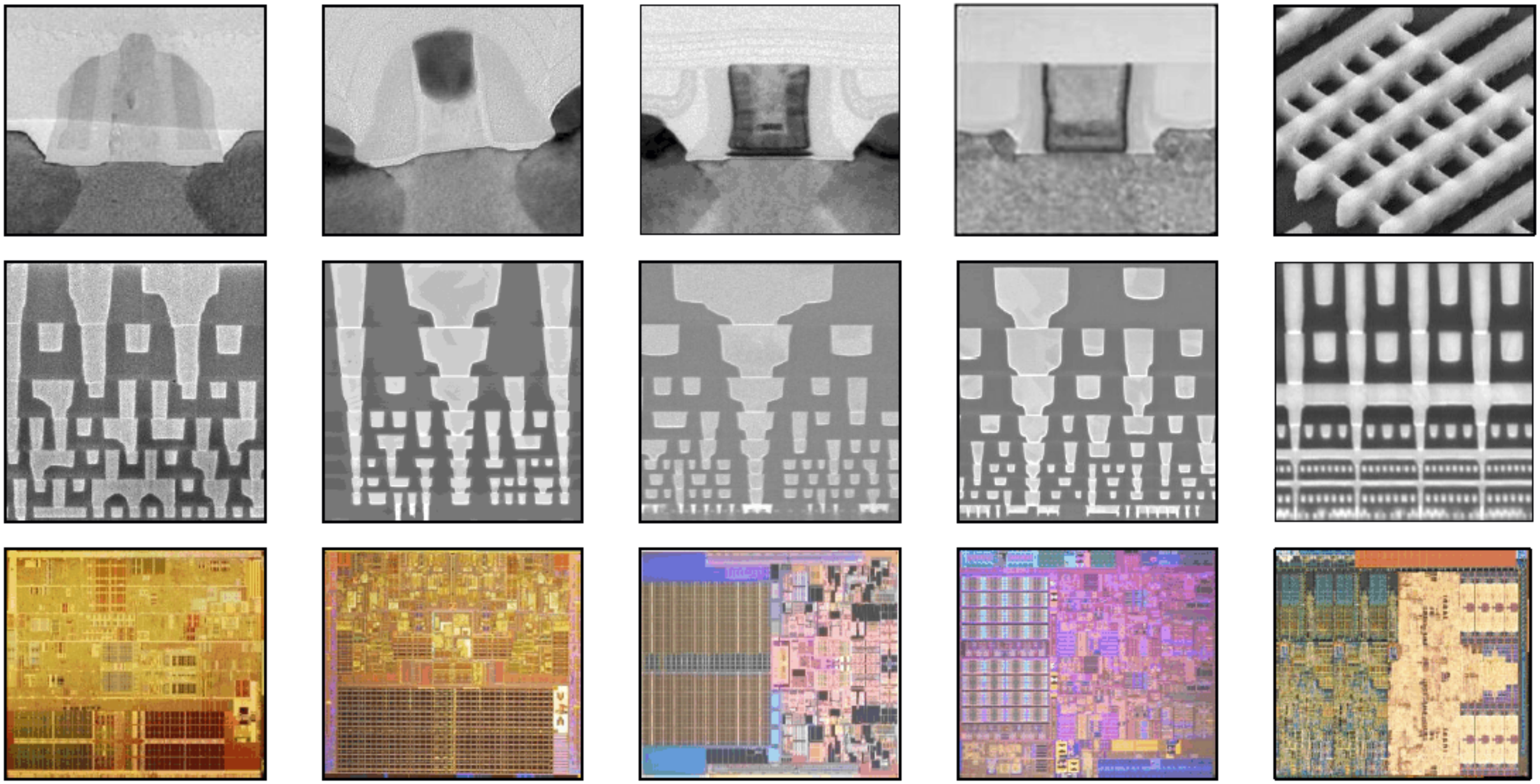
A brief summary of die photos, trends, device performance including 14nm CMOS FIN-FETs

Spring 2017

Gian Gerosa
Mark McDermott

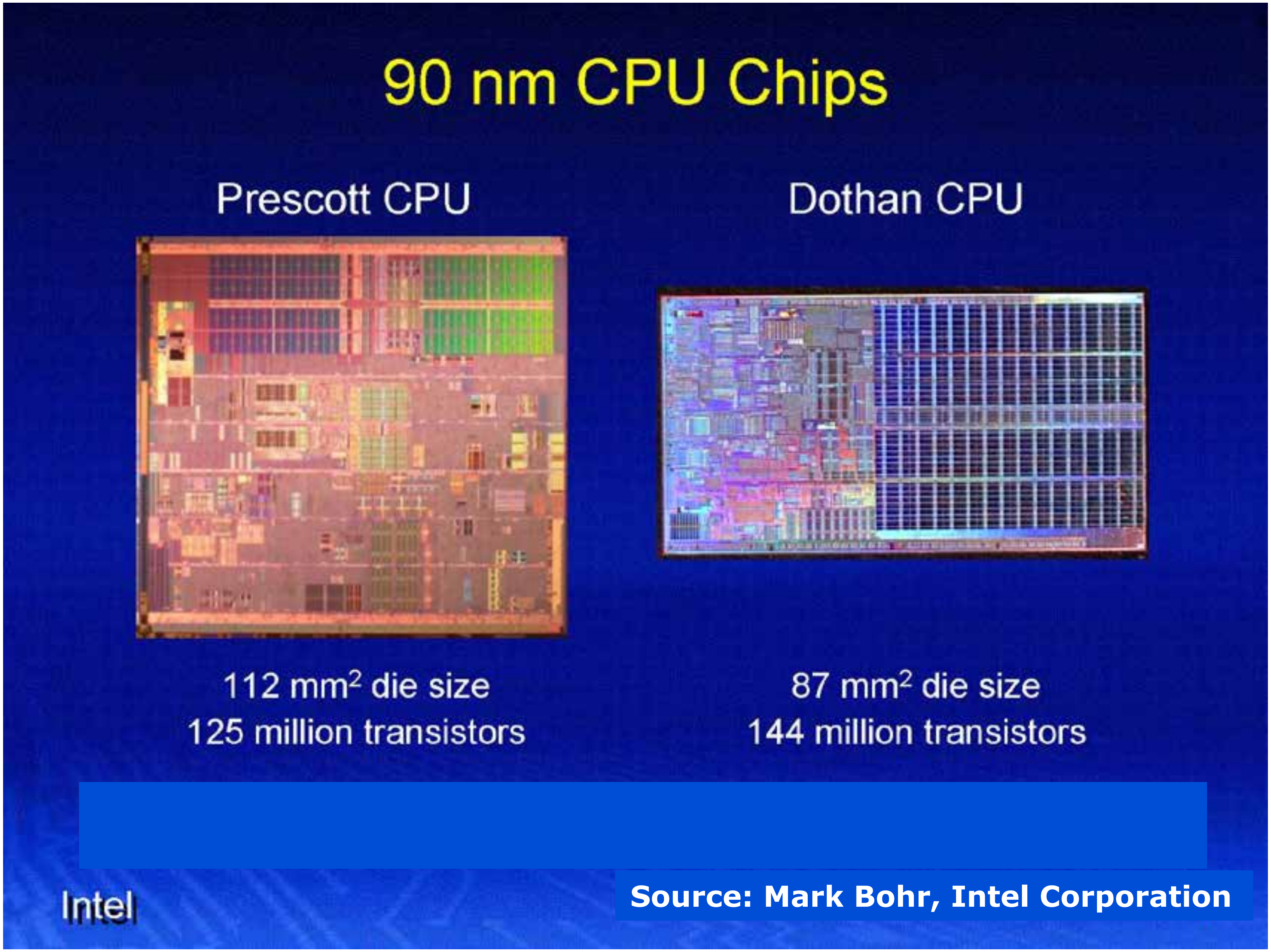
Intel’s 2-Year Technology Cadence

90 nm	65 nm	45 nm	32 nm	22 nm
2003	2005	2007	2009	2011



Source: Mark Bohr, Intel Corporation

P4 Processors in 90nm



90 nm CPU Chips

Prescott CPU

112 mm^2 die size
125 million transistors

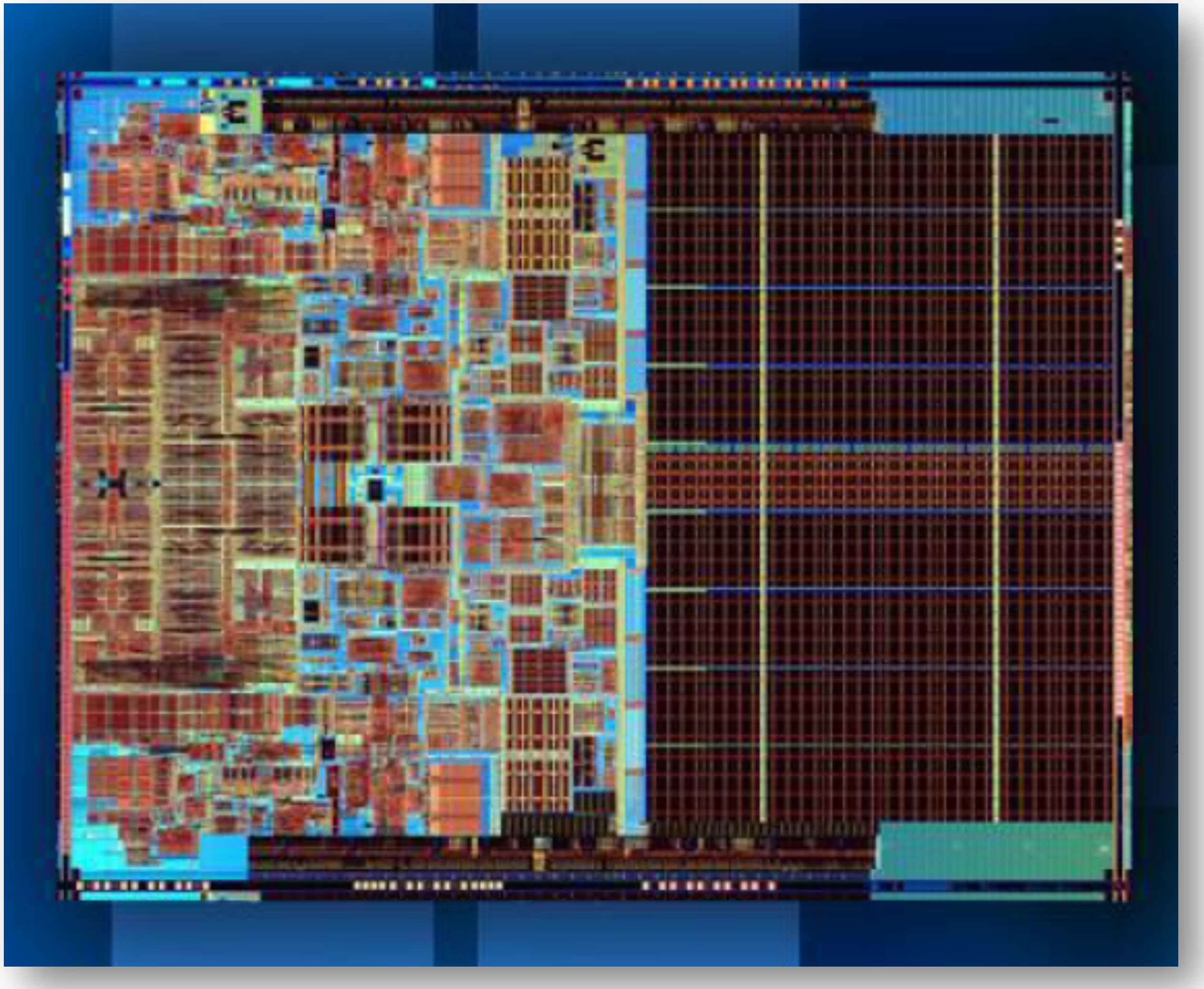
Dothan CPU

87 mm^2 die size
144 million transistors

Intel

Source: Mark Bohr, Intel Corporation

MEROM core2 duo in 65nm

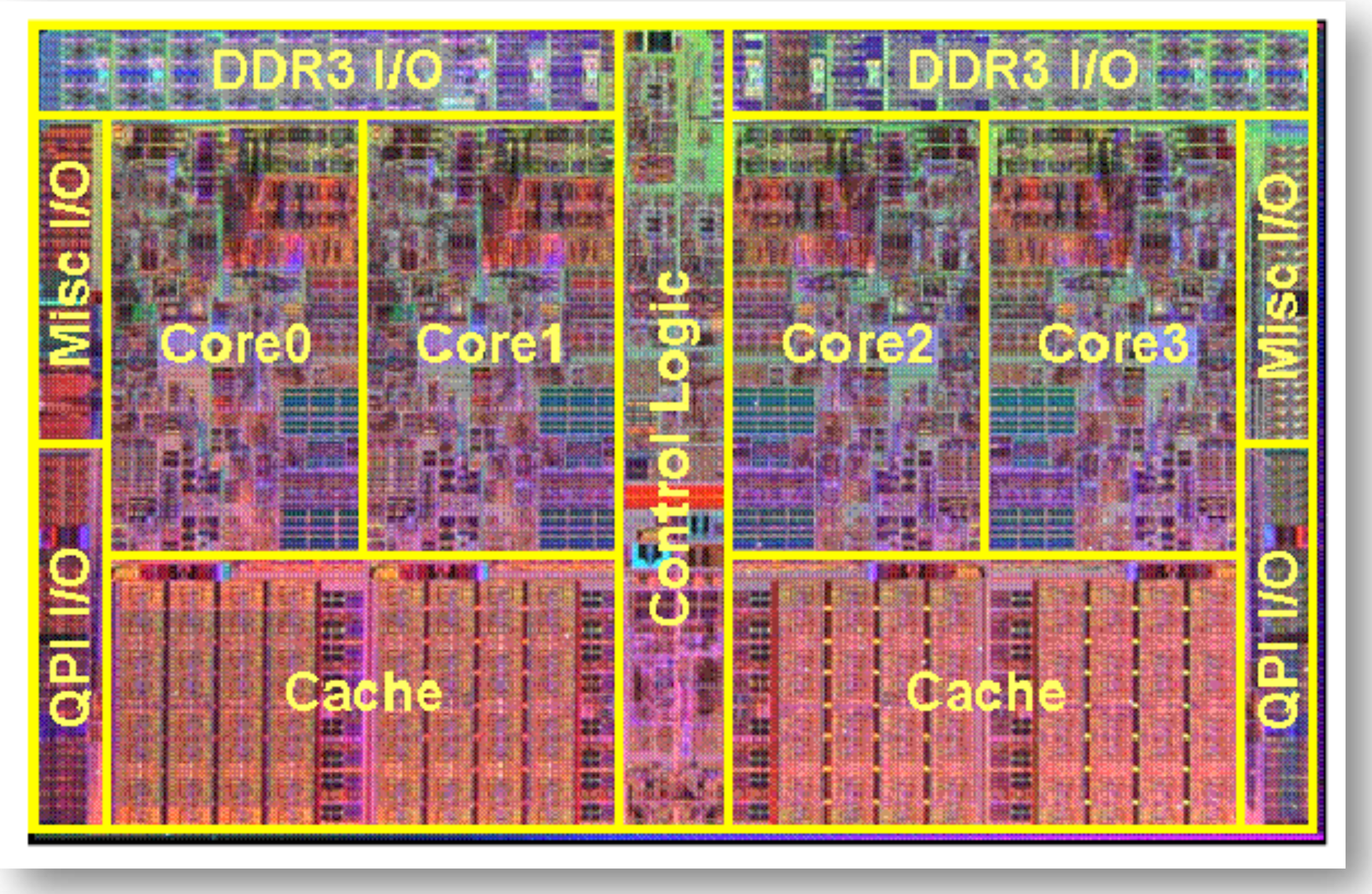


~180 mm2

~ 450 million transistors

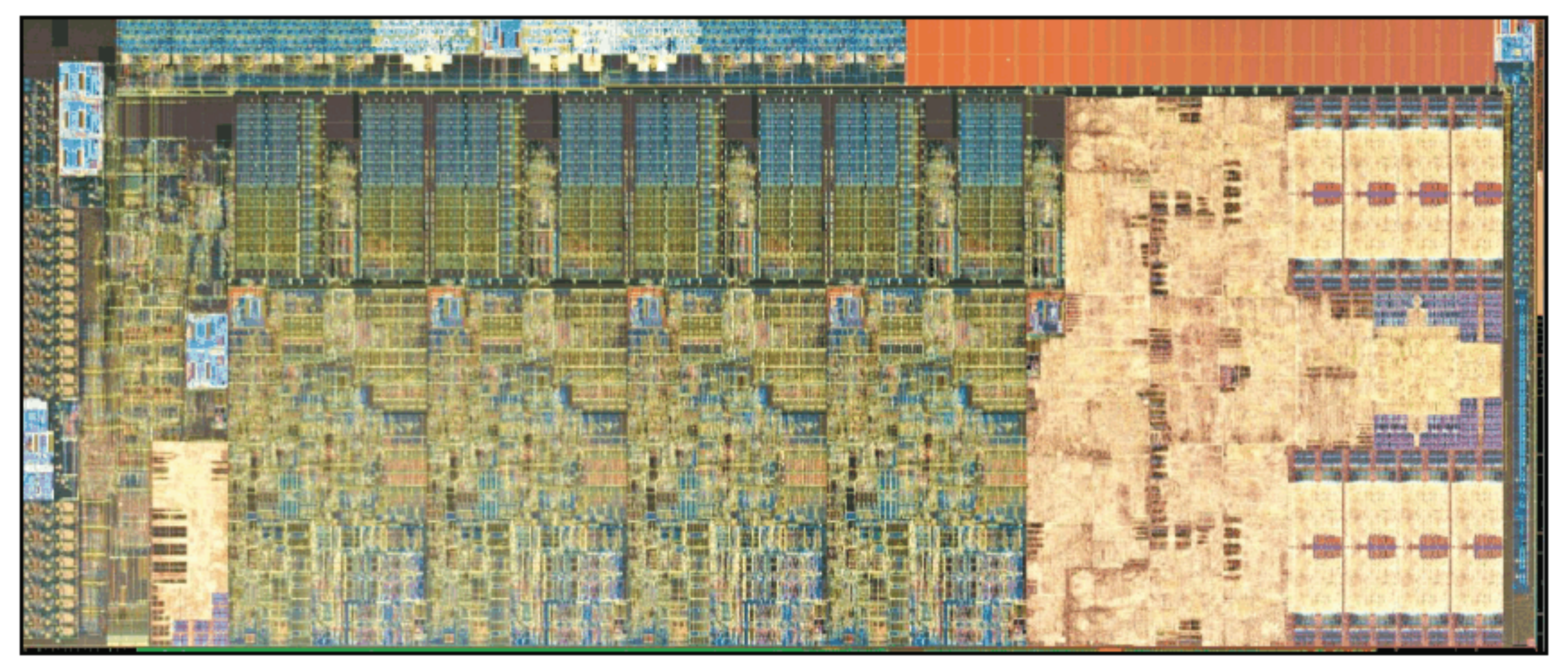
4 MB L2

Four Core Nehalem (core i7) in 45nm



~731 million transistors
8 MB L2

3rd Generation Intel Core Processor in 22nm



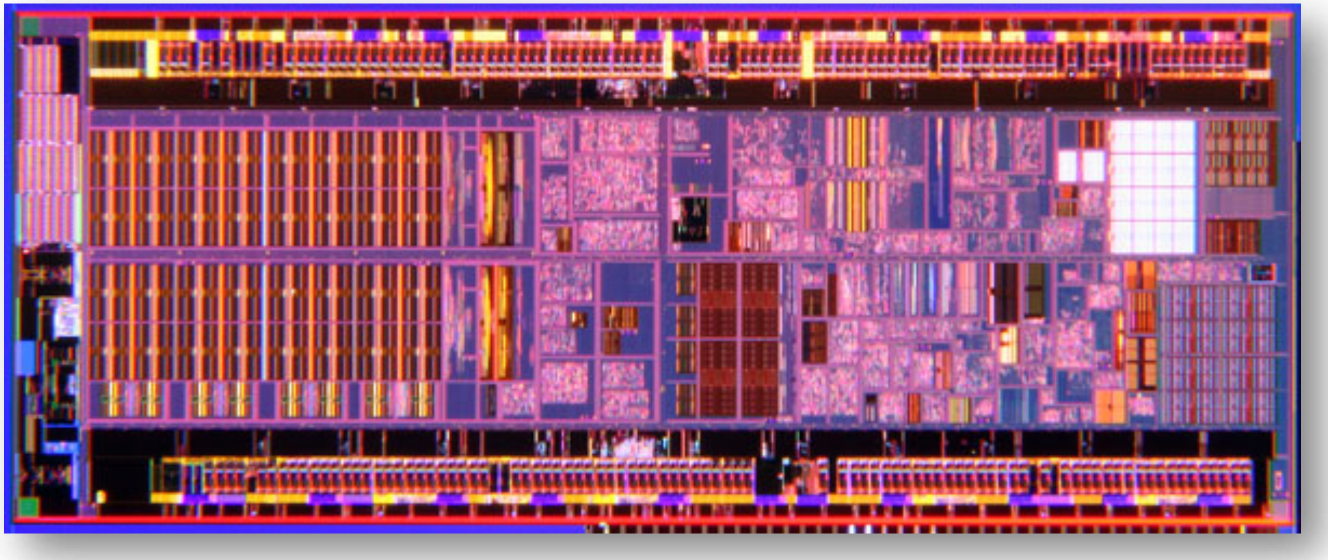
Tri-Gate CMOS

4 Cores + Integrated Graphics

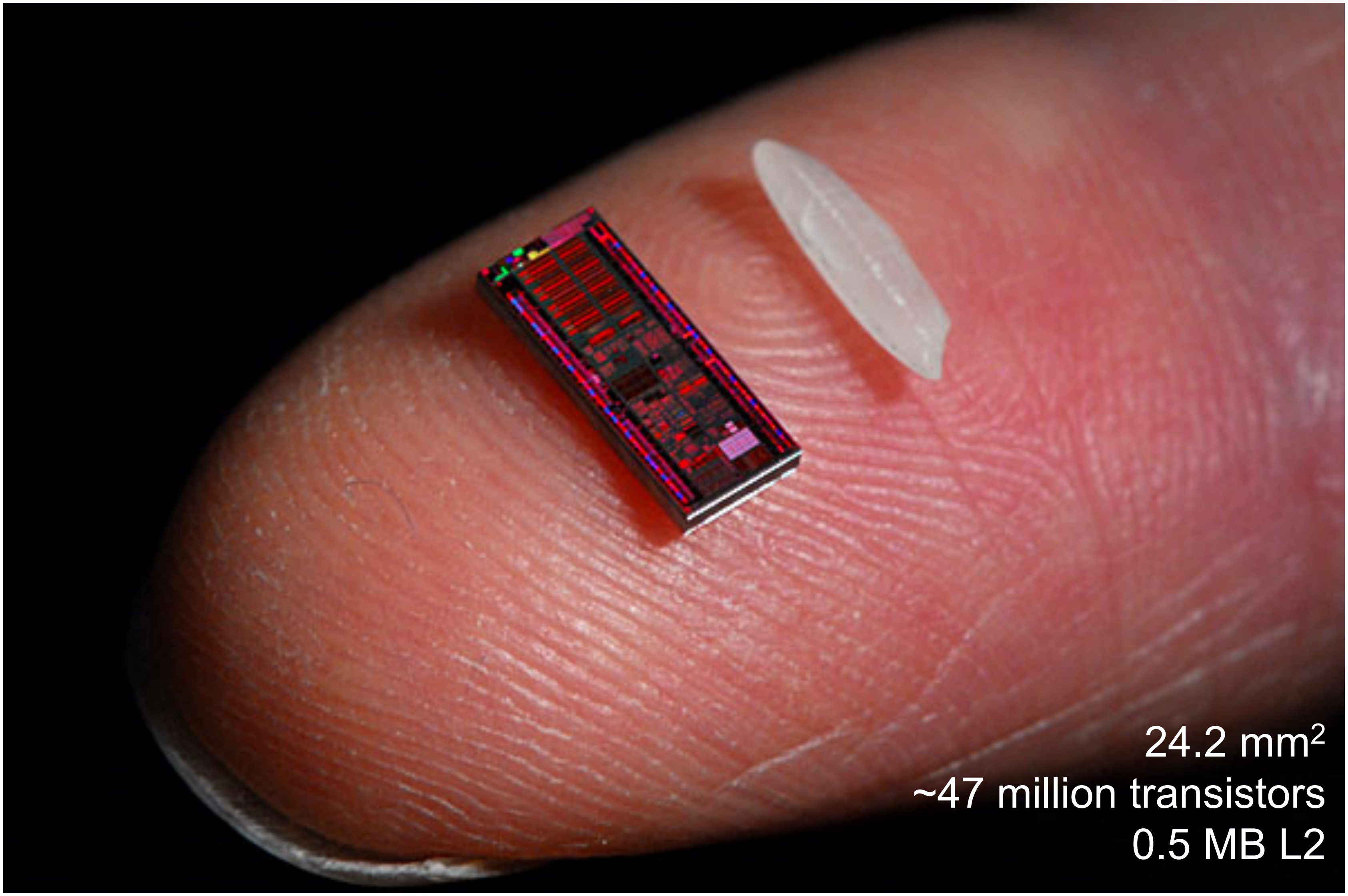
1.4 Billion Transistors, ~160 mm^2

SILVERTHORNE (ATOM Processor) in 45nm

24.2 mm2
~47 million transistors
0.5 MB L2



SILVERTHORNE (ATOM Processor) in 45nm

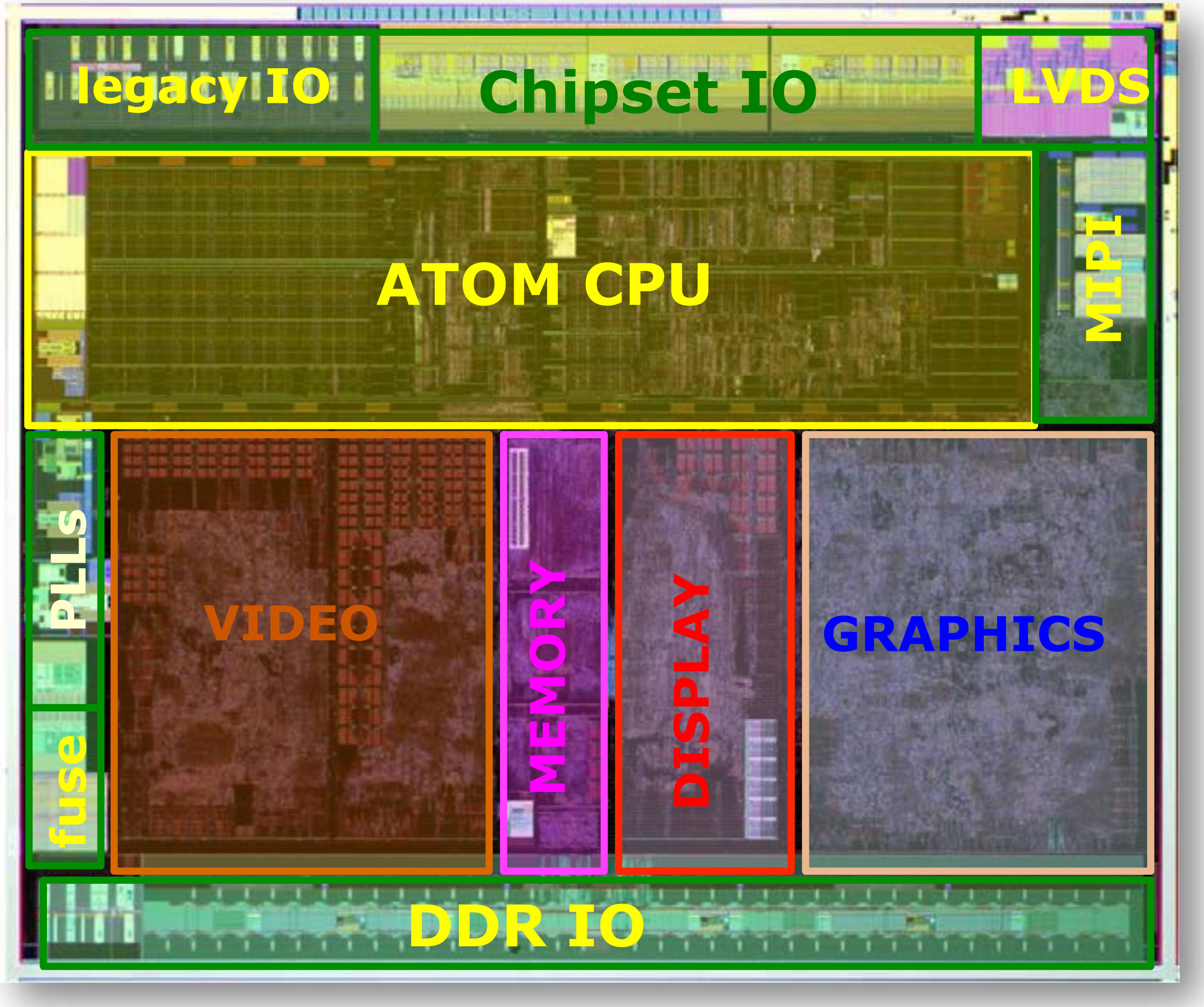


24.2 mm^2

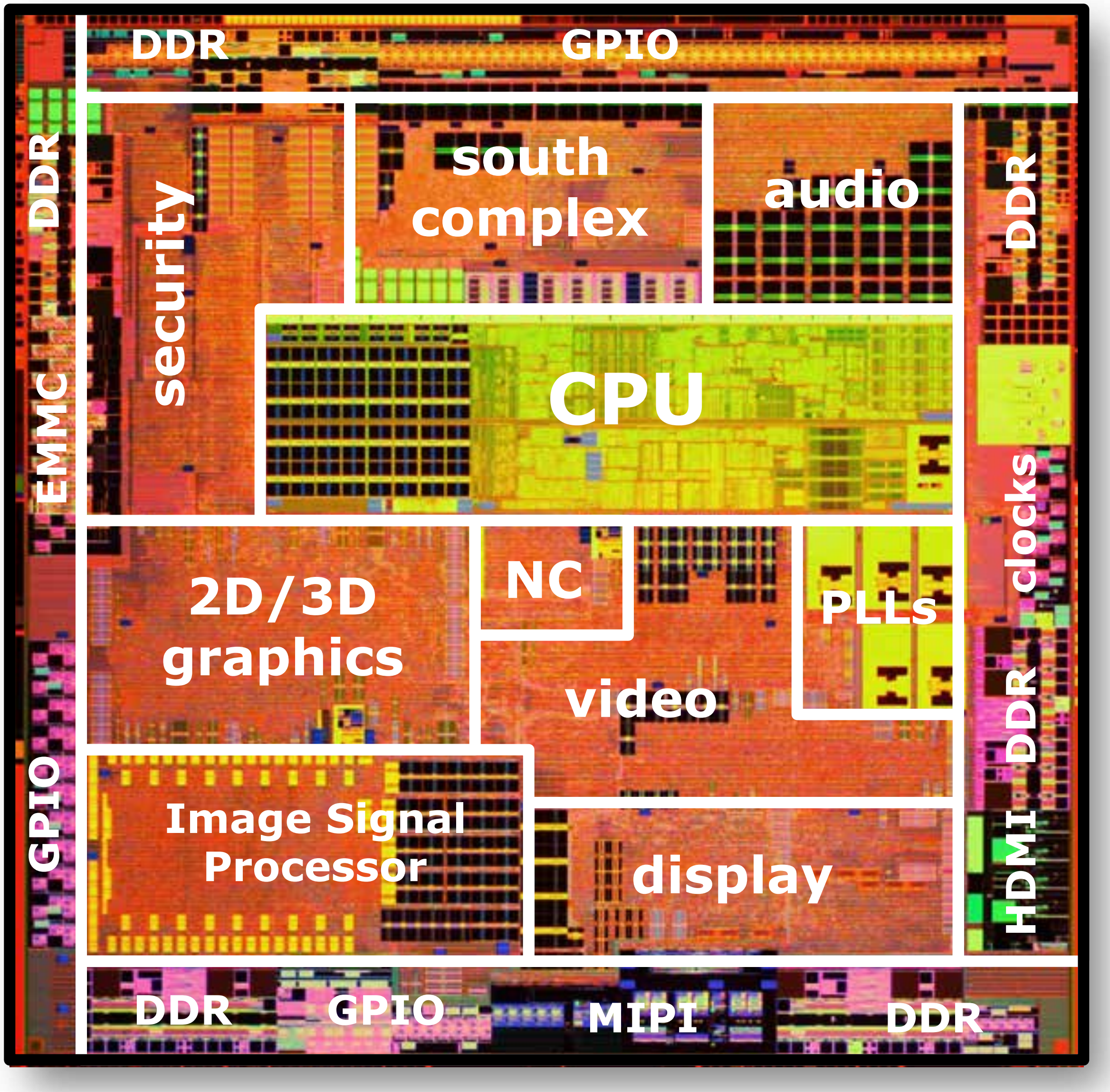
~47 million transistors

0.5 MB L2

ATOM Z6XX SoC in 45nm



ATOM Z2480 SoC in 32nm



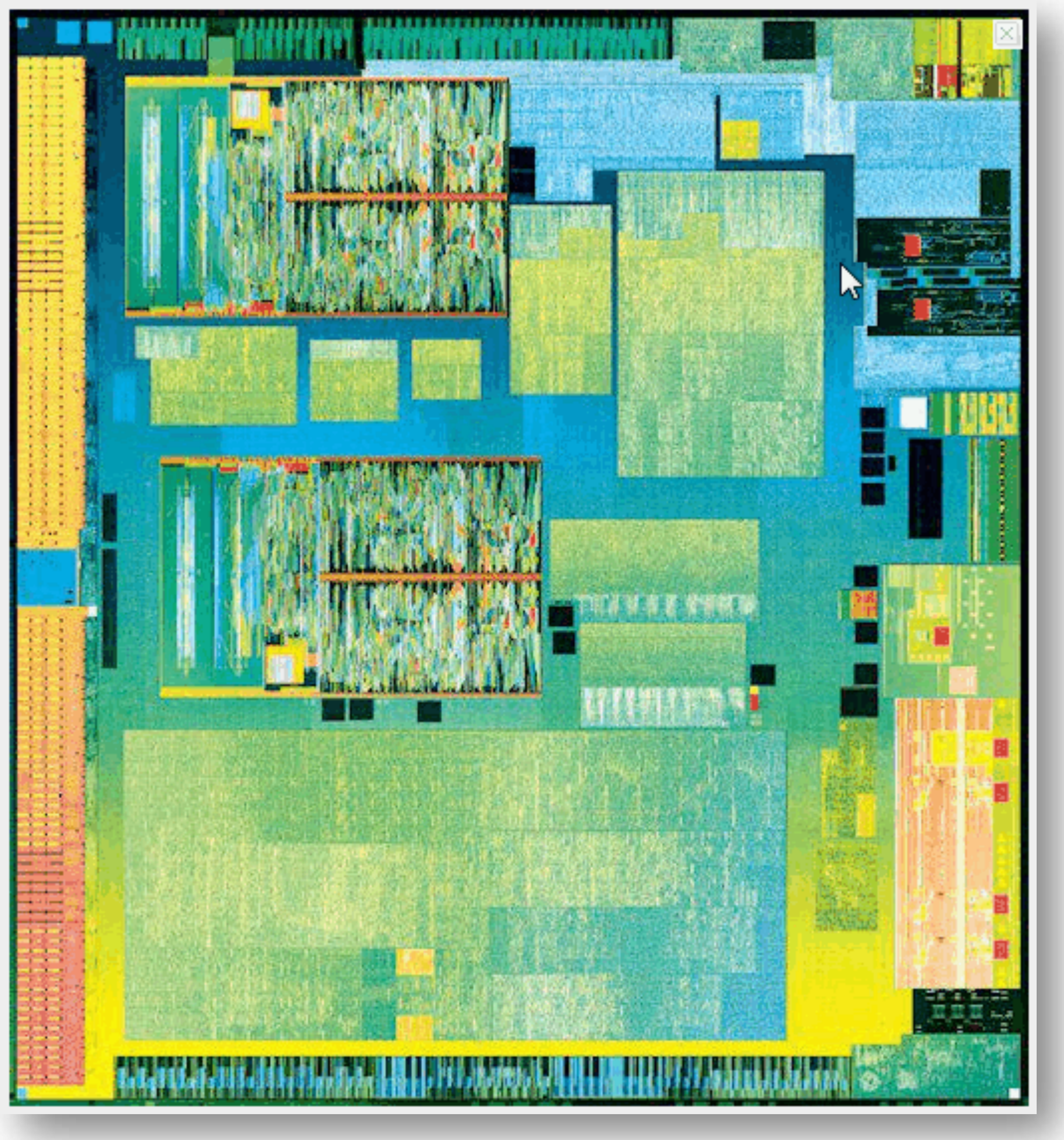
~64 mm2

~ 435 million transistors

Single ATOM CPU Core with 0.5 MB L2

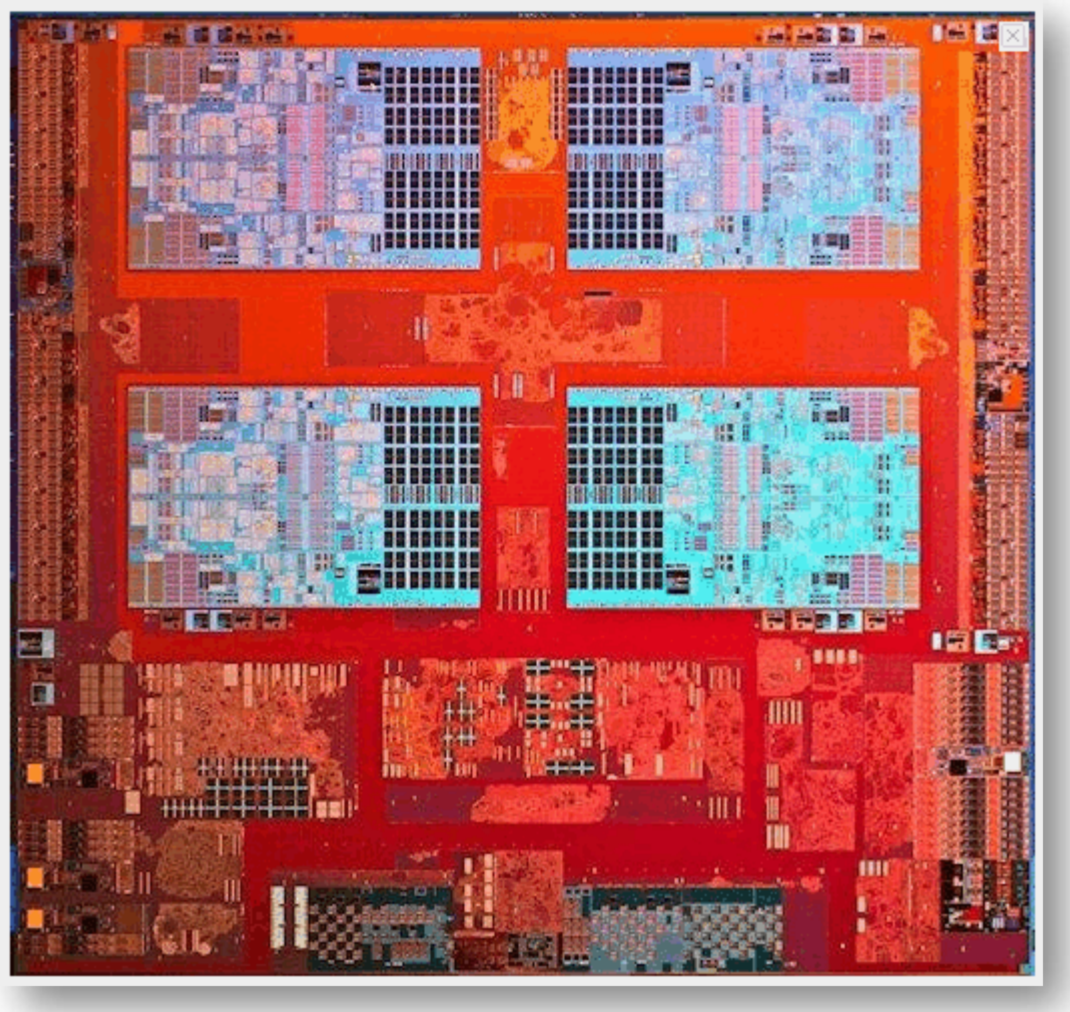
Rumi Zahir, HOT CHIPs #24, Aug. 2012

ATOM Z3xxx SoC in 22nm



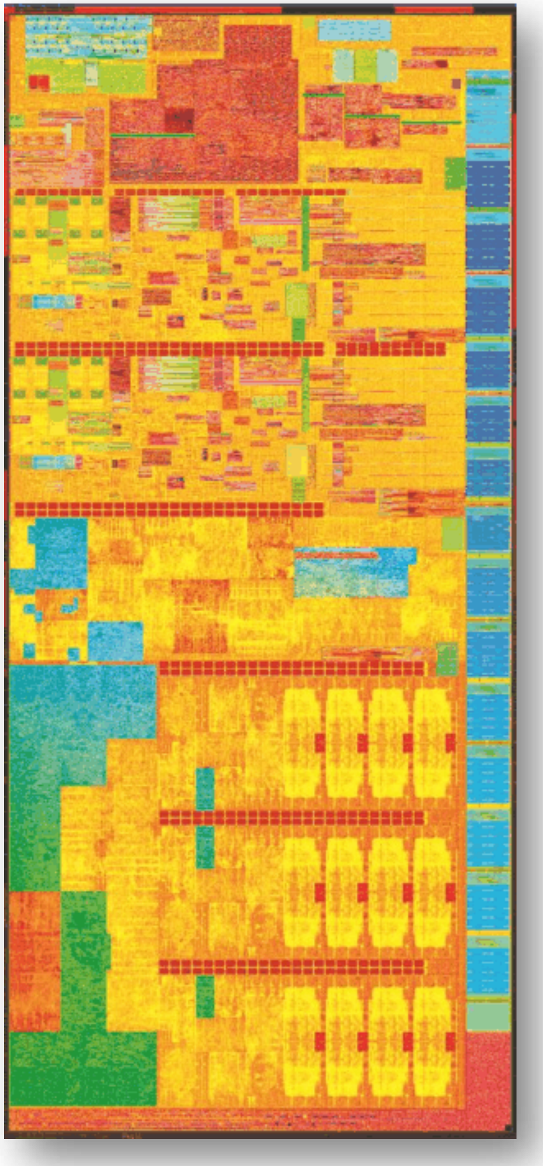
FinFET CMOS
~102 mm2
4 ATOM CPU Cores
2MB L2

ATOM C2000 64b micro-Server SoC in 22nm



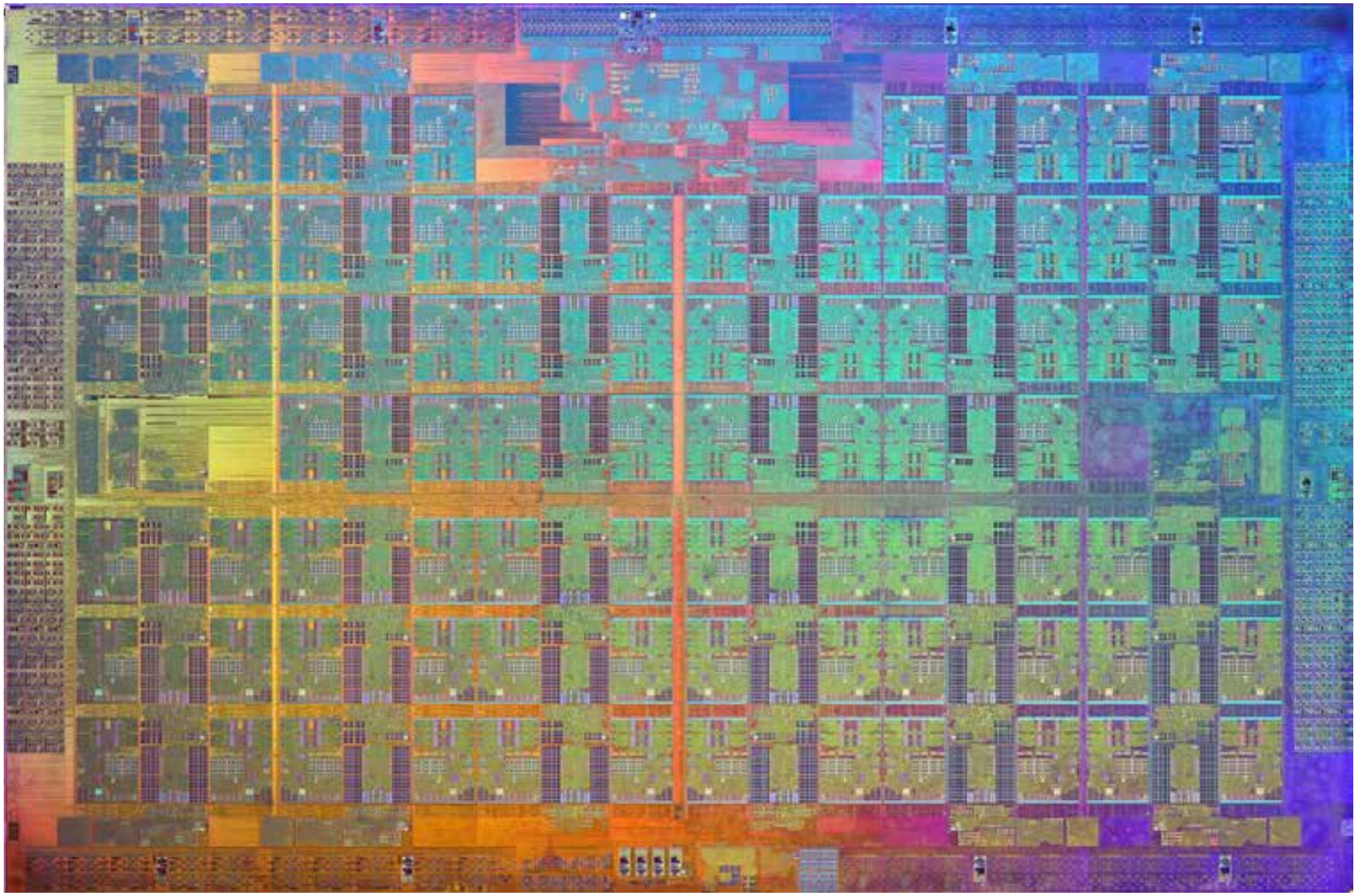
FinFET CMOS
8 ATOM CPU Cores
4MB L2

BROADWELL in 14nm [15]



FinFET CMOS
2 CPU Cores, Intel Graphics
1.3 Billion Transistors, ~82 mm^2

KNIGHTS LANDING (Xeon Phi) in 14nm



FinFET CMOS

72 Silvermont

ATOM Cores

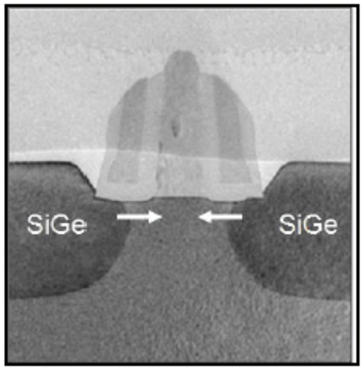
200W TDP

3 TeraFLOPs (dP)

TRANSISTOR HISTORY

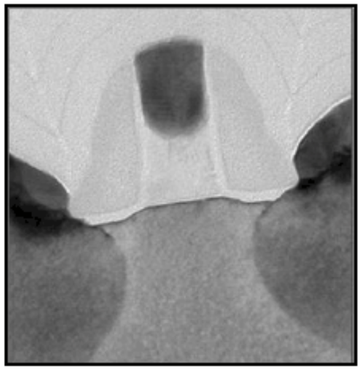
90 nm

2003



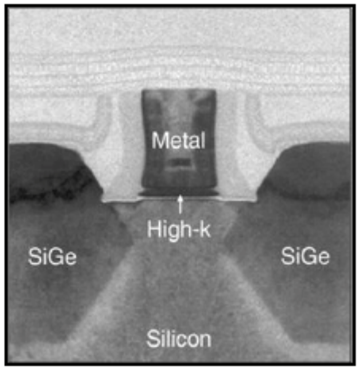
65 nm

2005



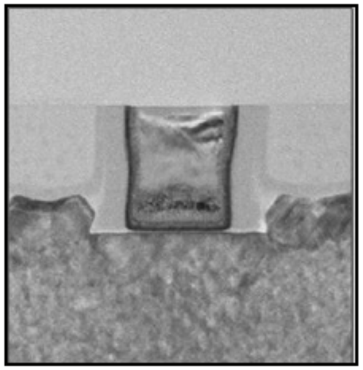
45 nm

2007



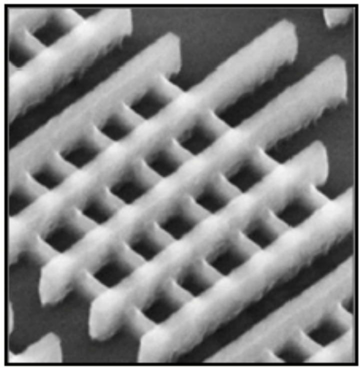
32 nm

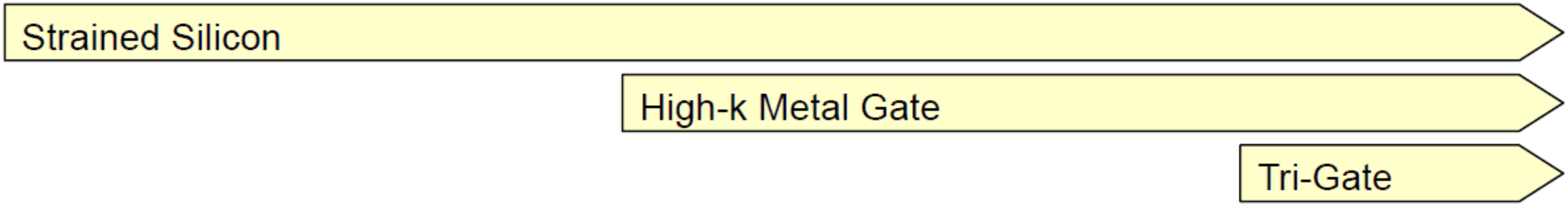
2009



22 nm

2011



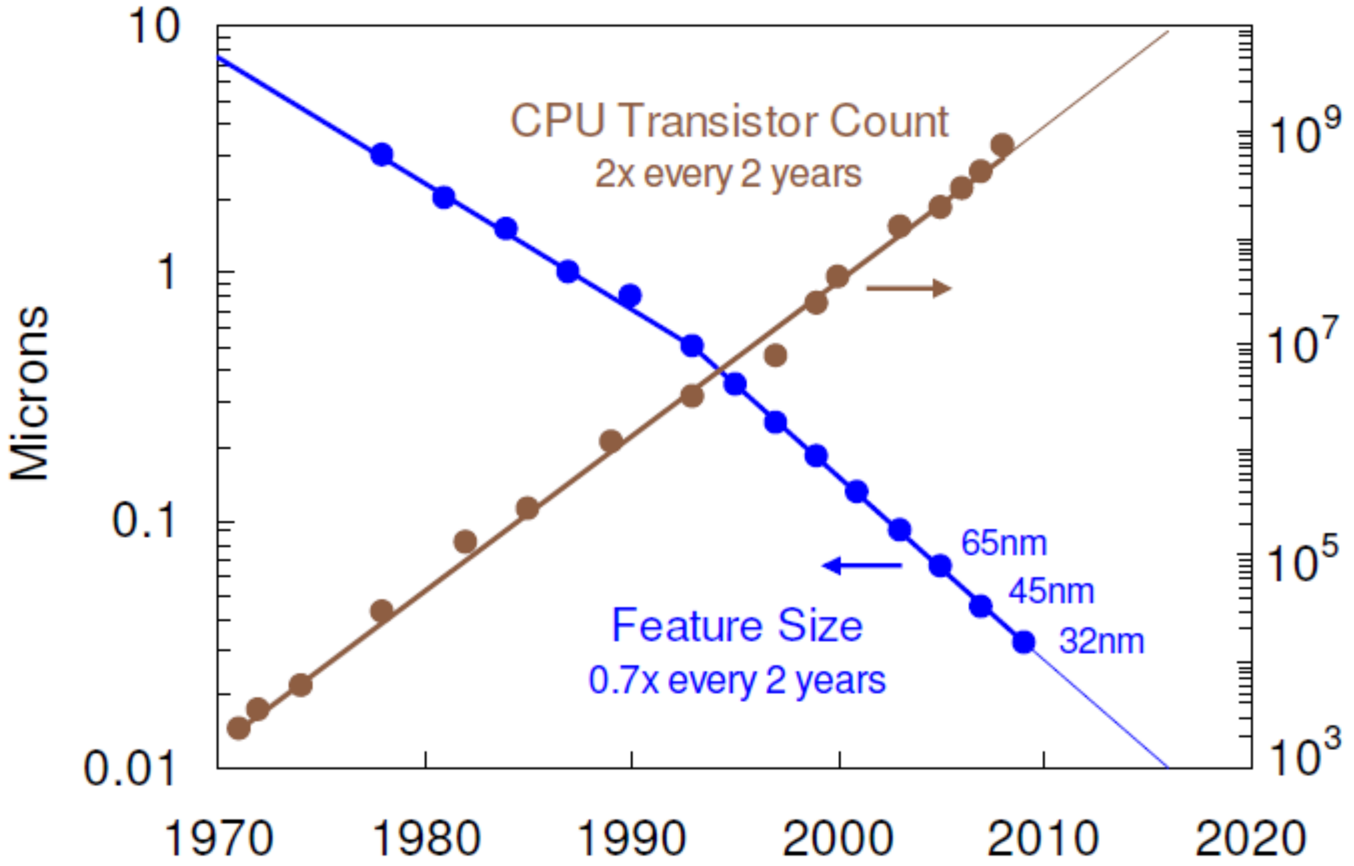


Strained Silicon

High-k Metal Gate

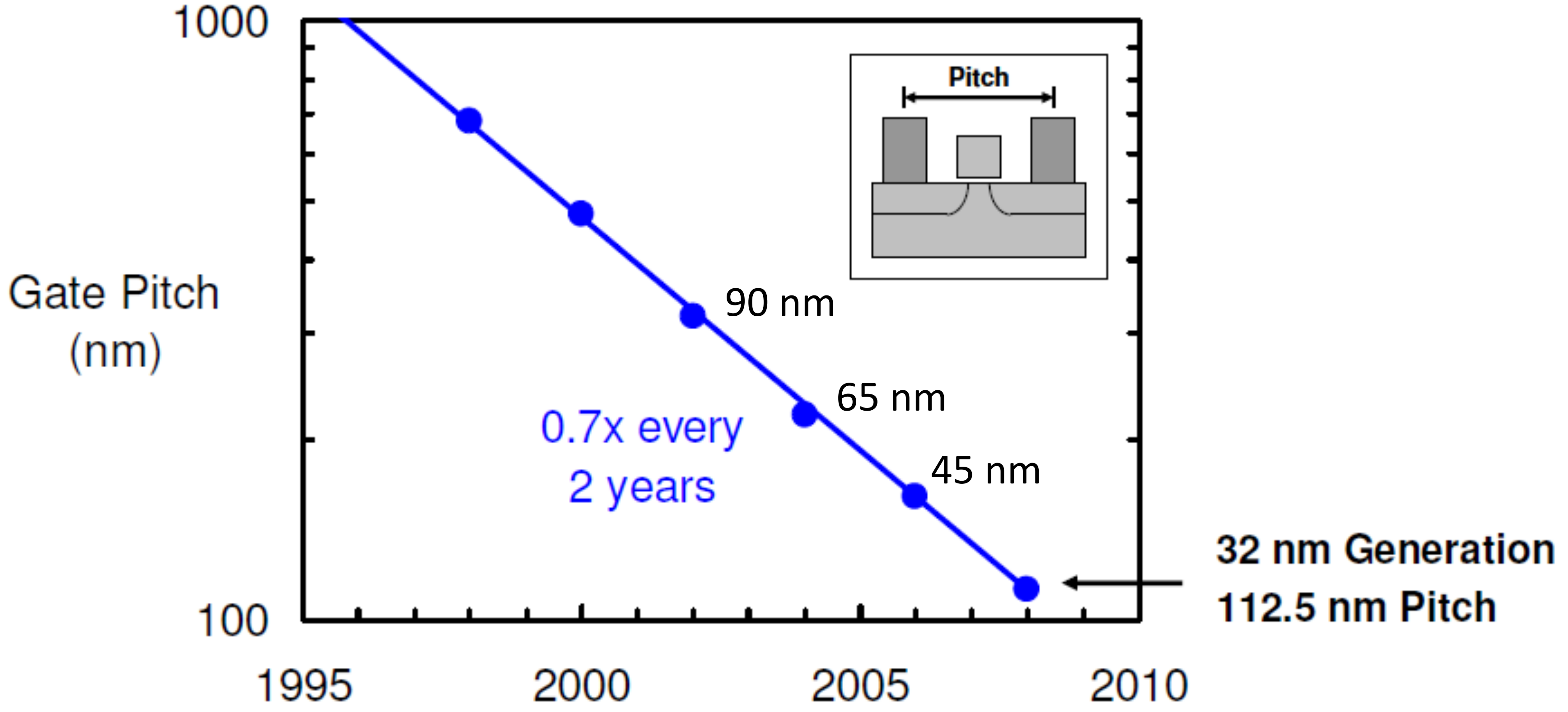
Tri-Gate

Scaling Trends



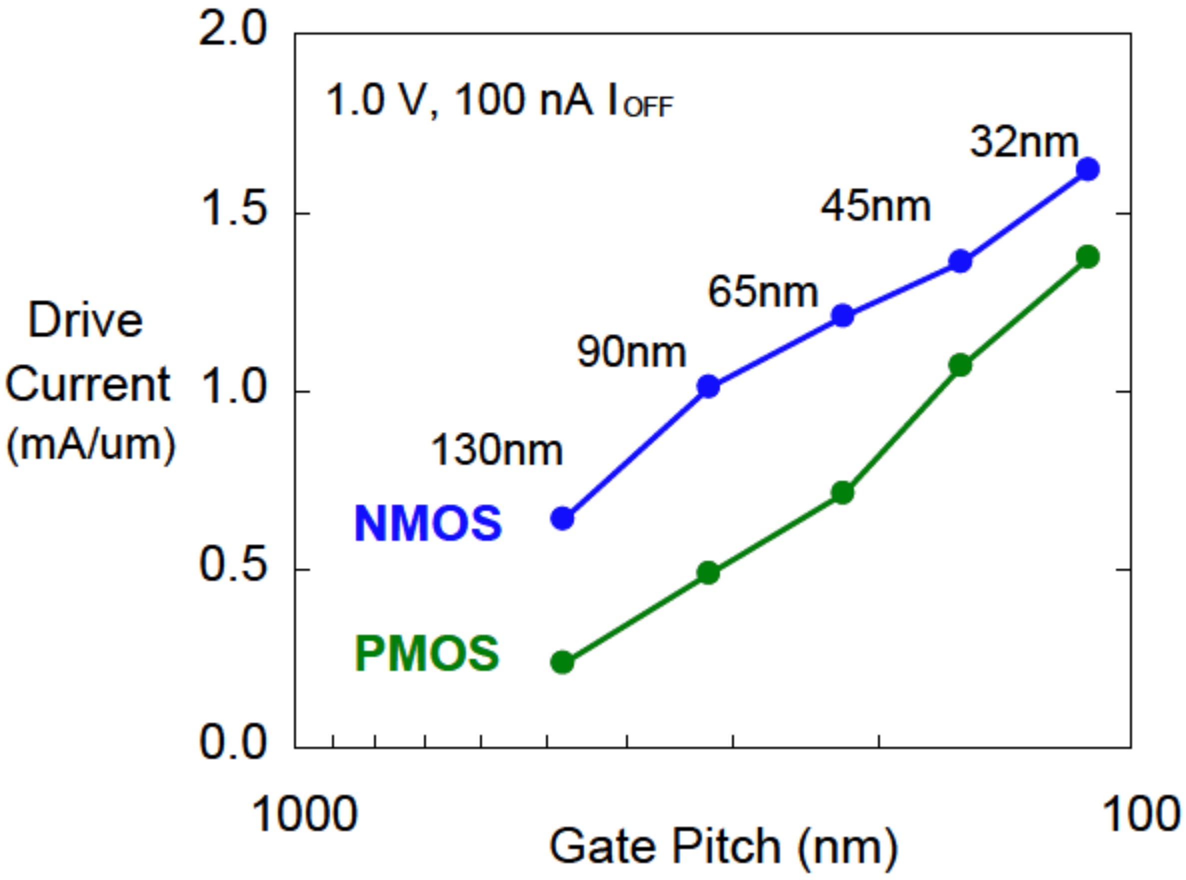
Source: Mark Bohr, Intel Corporation

Contacted Gate Pitch



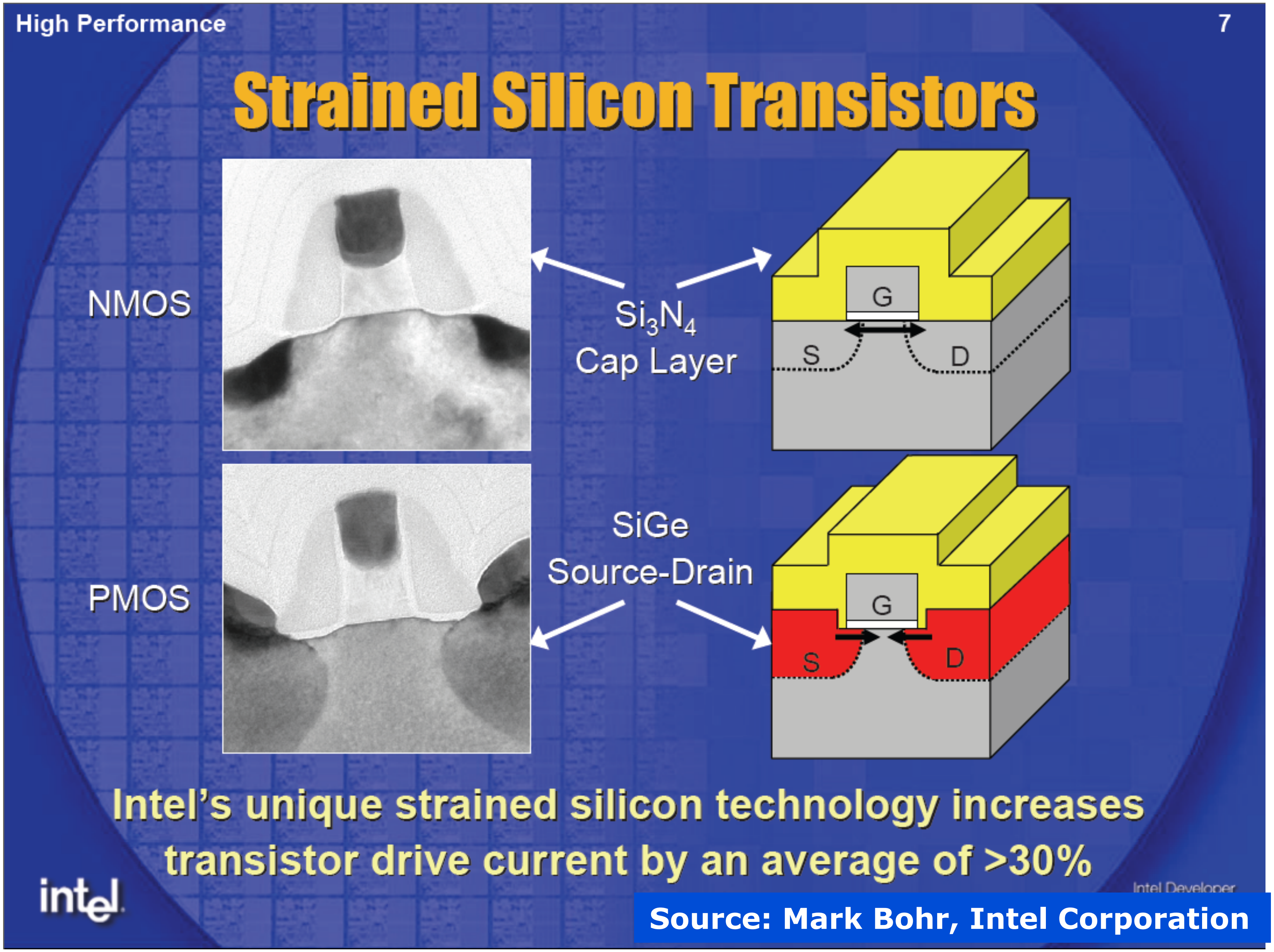
Source: Mark Bohr, Intel Corporation

Normalized Drive Current



P. Packan, et al., “High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors”, Tech. Digest IEDM, Dec 2009.

MOBILITY IMPROVEMENT



High Performance

Strained Silicon Transistors

NMOS

PMOS

Si_3N_4 Cap Layer

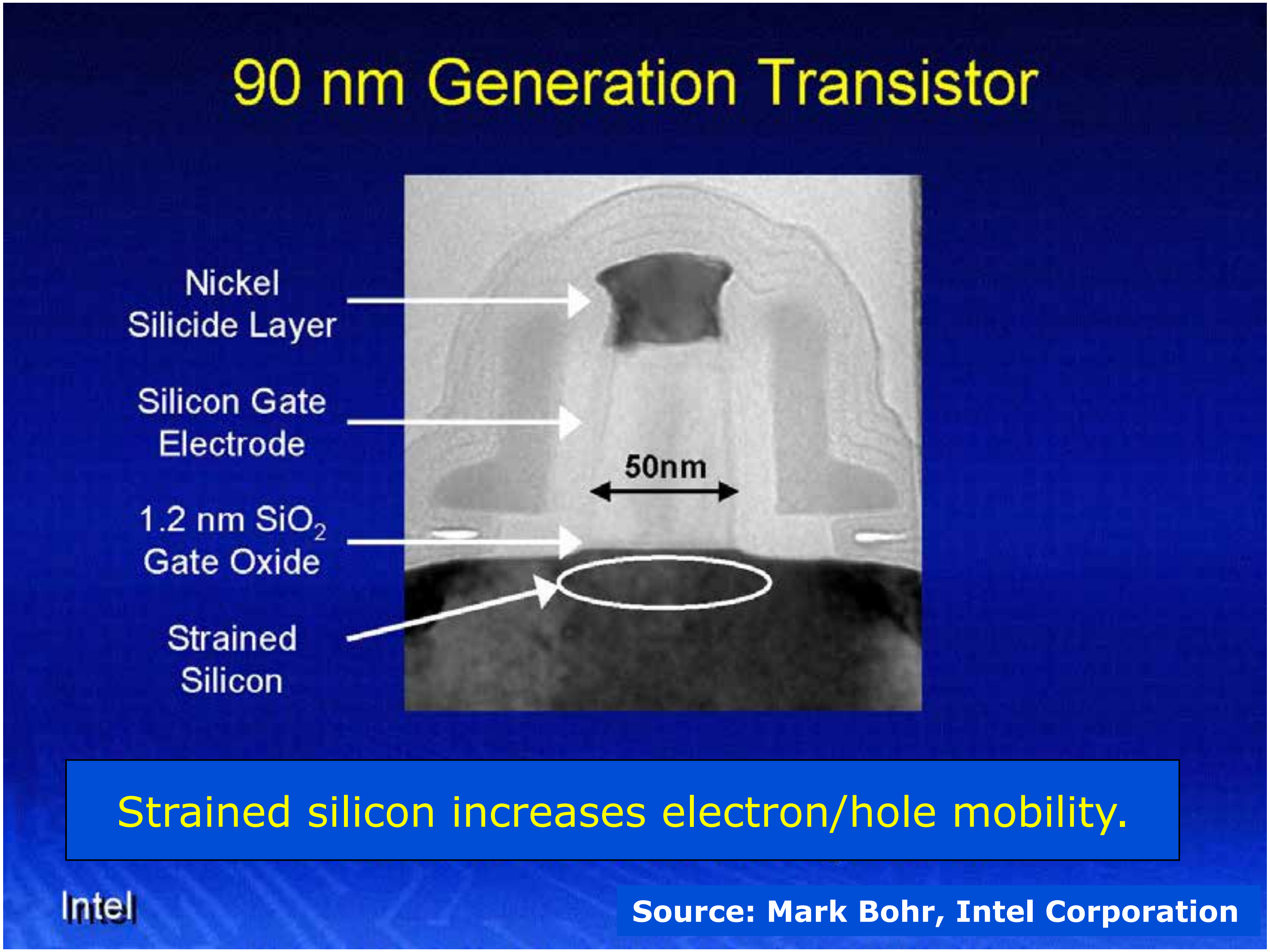
SiGe Source-Drain

Intel's unique strained silicon technology increases transistor drive current by an average of >30%

intel

Intel Developer

Source: Mark Bohr, Intel Corporation



90 nm Generation Transistor

Nickel Silicide Layer

Silicon Gate Electrode

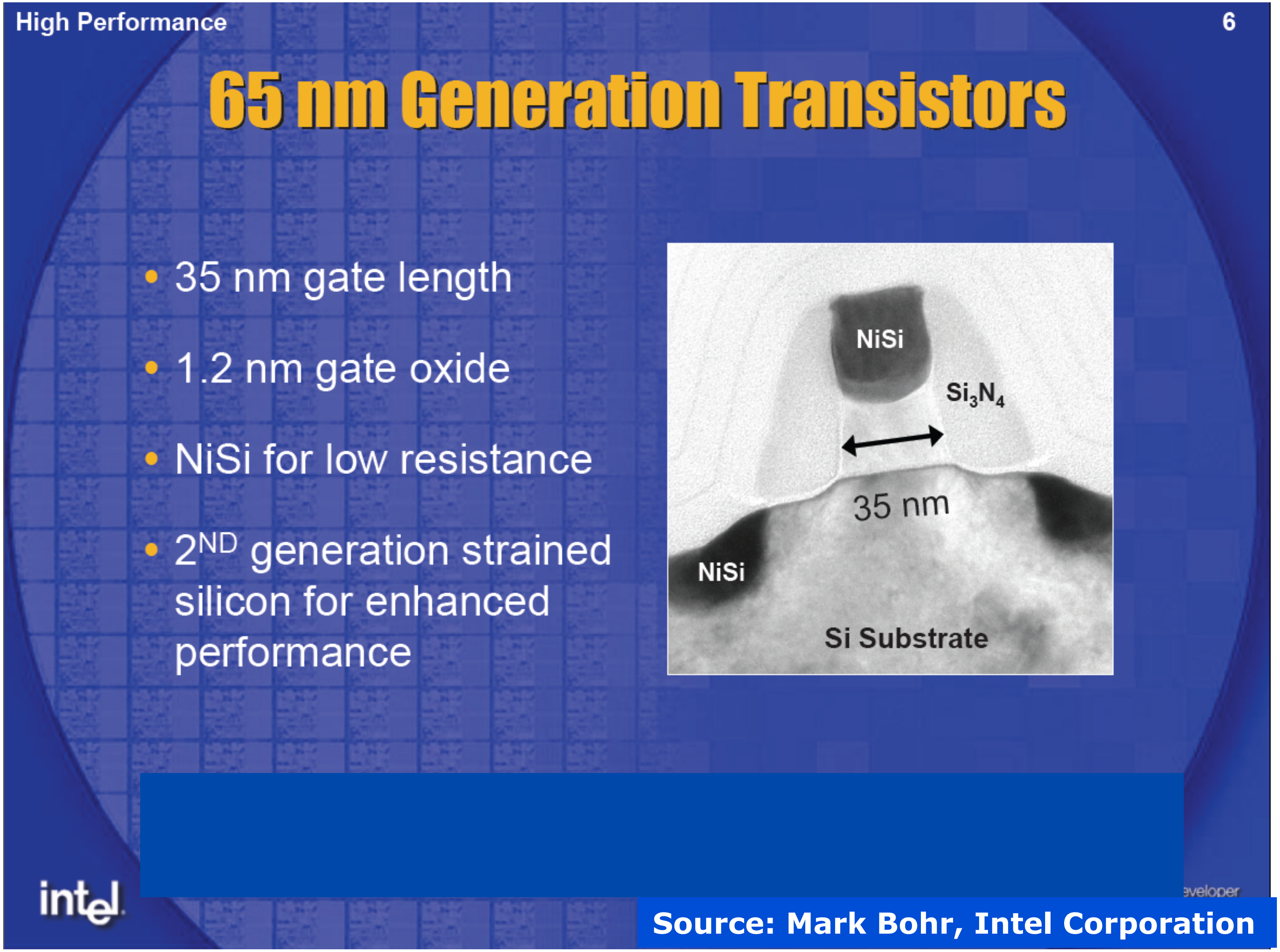
1.2 nm SiO_2 Gate Oxide

Strained Silicon

Strained silicon increases electron/hole mobility.

Intel

Source: Mark Bohr, Intel Corporation



High Performance

6

65 nm Generation Transistors

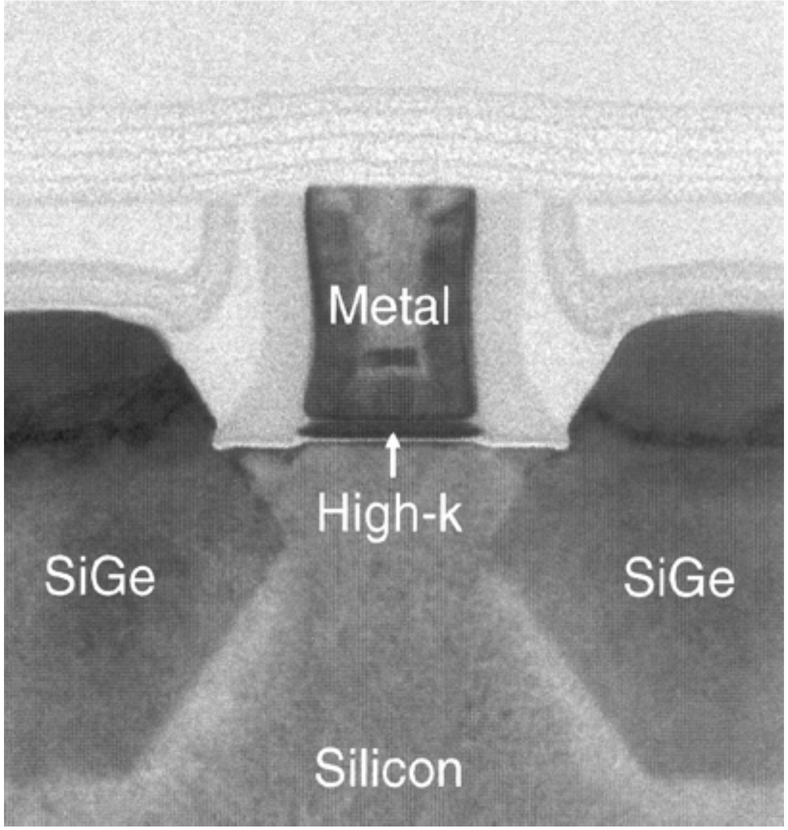
- 35 nm gate length
- 1.2 nm gate oxide
- NiSi for low resistance
- 2^{ND} generation strained silicon for enhanced performance

intel

eveloper

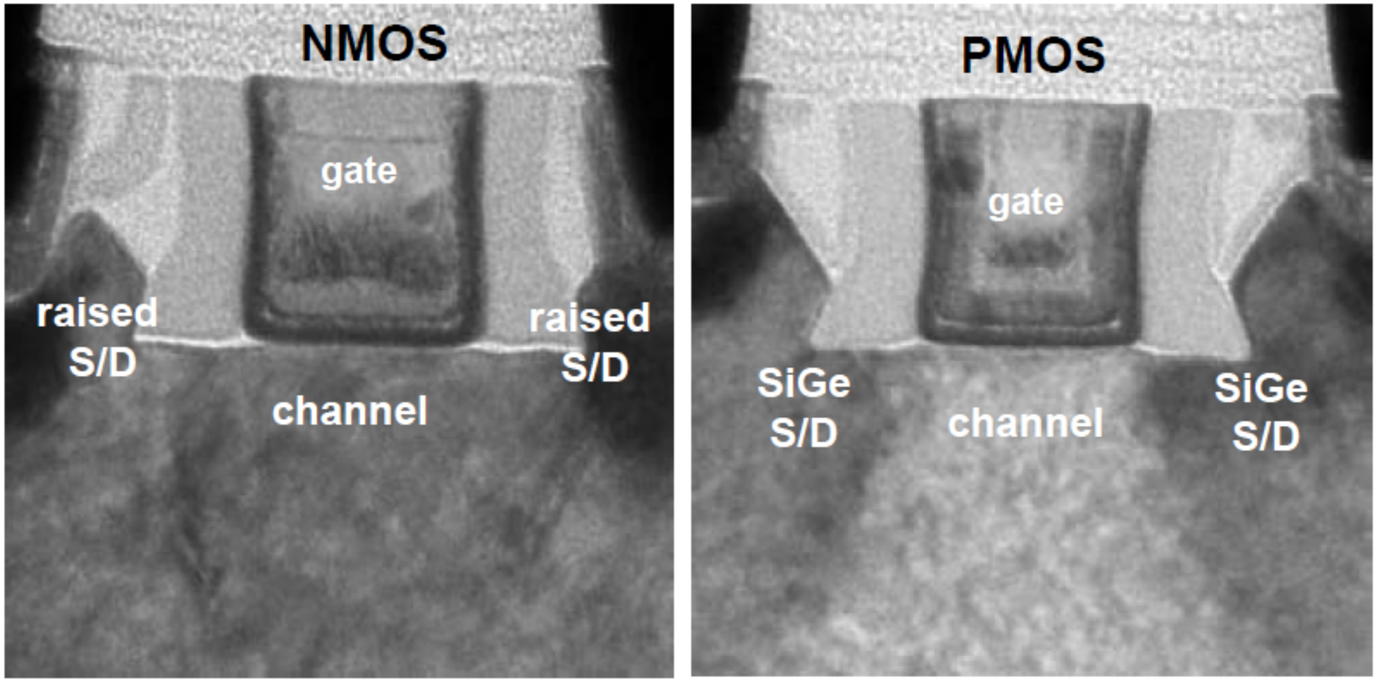
Source: Mark Bohr, Intel Corporation

High-K, Metal Gate 45 nm CMOS (Intel)



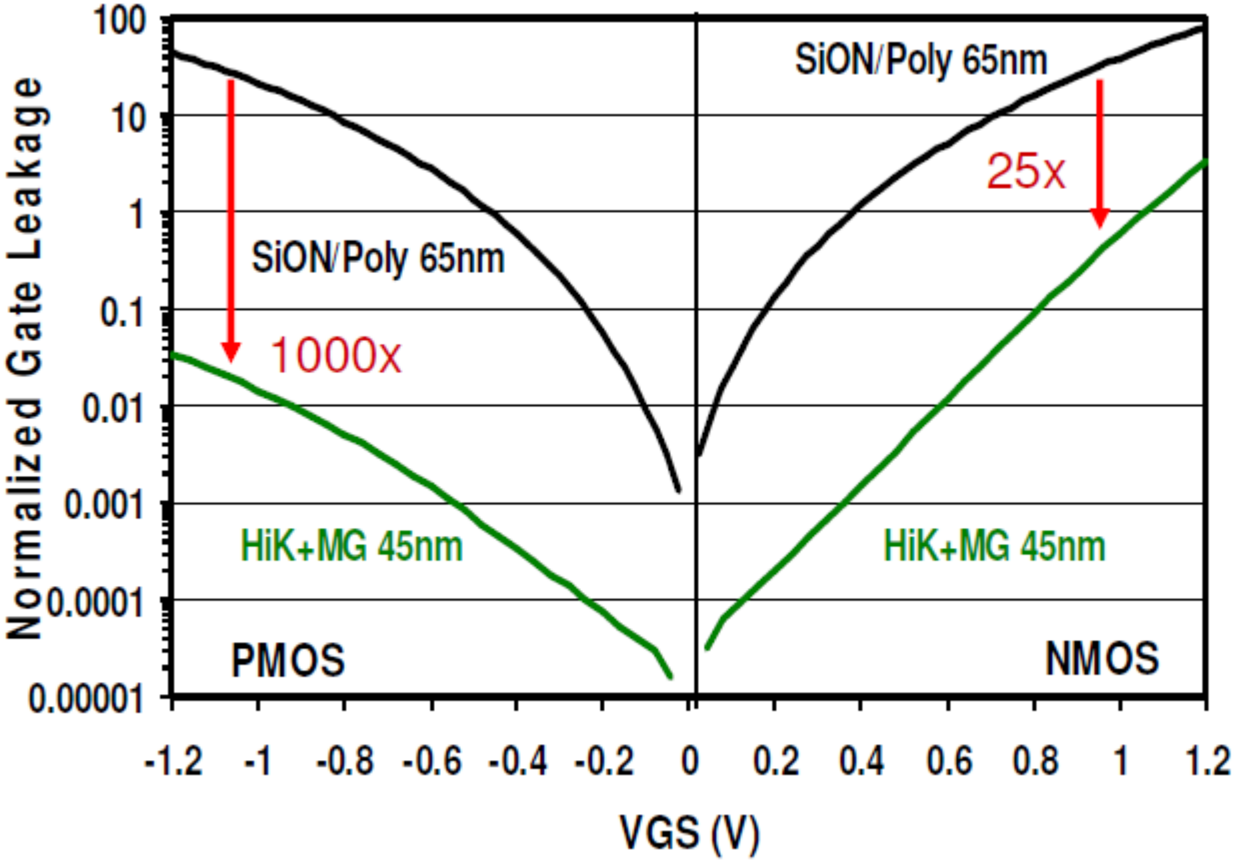
K. Mistry, et al., “A 45nm Logic Technology with High-k+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging”, Tech. Digest IEDM, Dec 2007.

High-K, Metal Gate 32 nm CMOS (Intel)



P. Packan, et al., "High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors", Tech. Digest IEDM, Dec 2009.

Gate Leakage Reduction



High-K and Metal-Gate significantly reduces gate leakage

Performance vs. Leakage Tradeoff

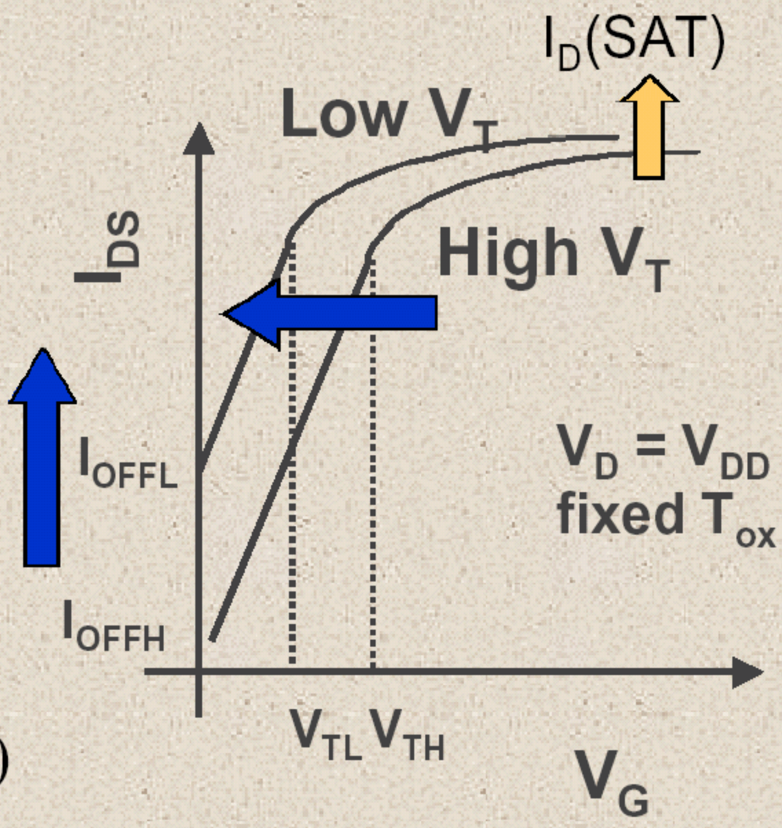
Performance vs Leakage:

$V_T \downarrow I_{OFF} \uparrow I_D(SAT) \uparrow$

$$I_{OFF} \propto I_{subth} \propto \frac{W_{eff}}{L_{eff}} K_1 e^{(V_{GS}-V_T)}$$

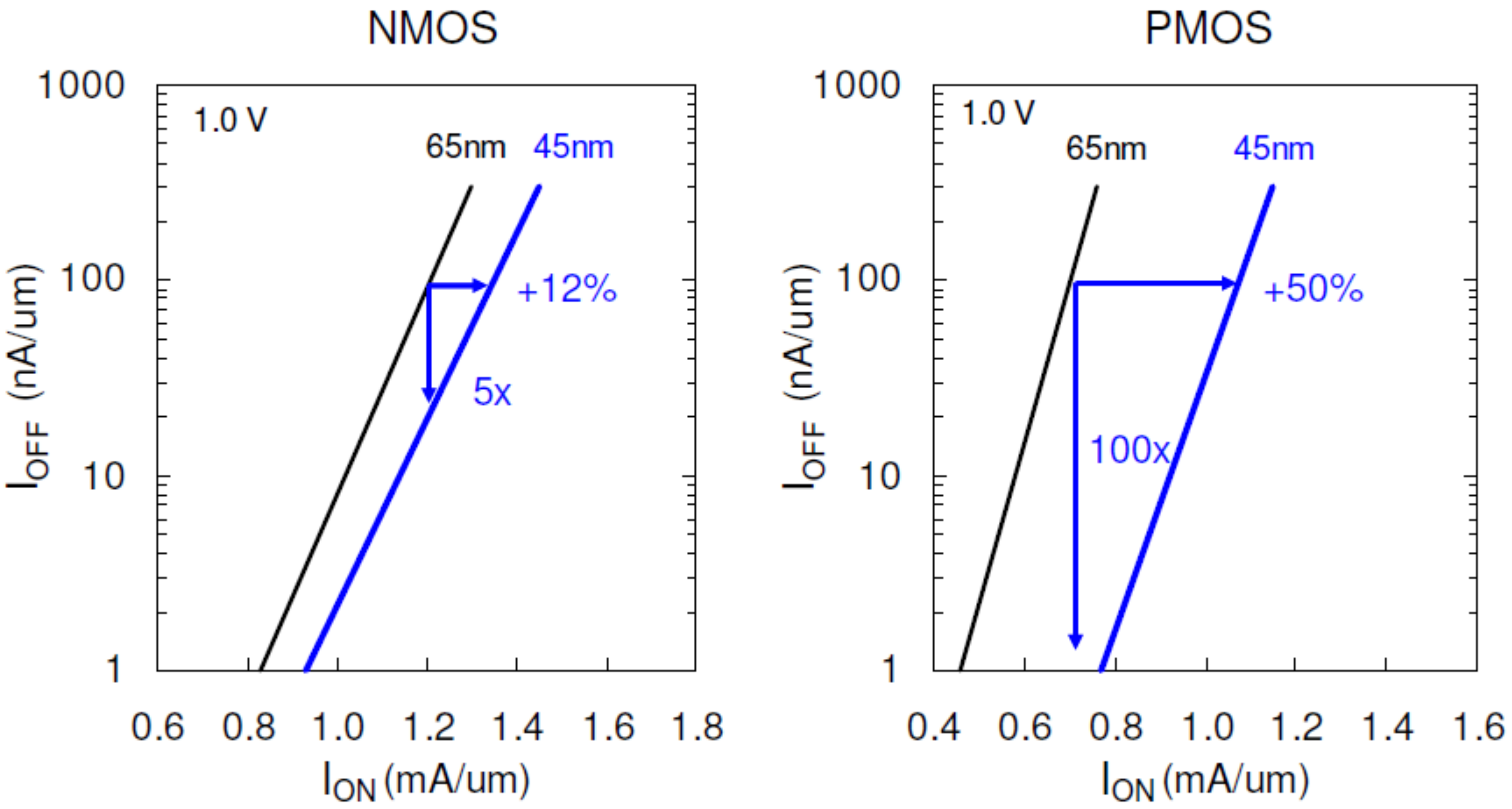
$$I_D(SAT) \propto \frac{W_{eff}}{L_{eff}} K_2 (V_{GS}-V_T)^{\alpha}$$

$$I_D(SAT) \propto K_3 W_{eff} C_{ox} \upsilon_{SAT} (V_{GS}-V_T)$$



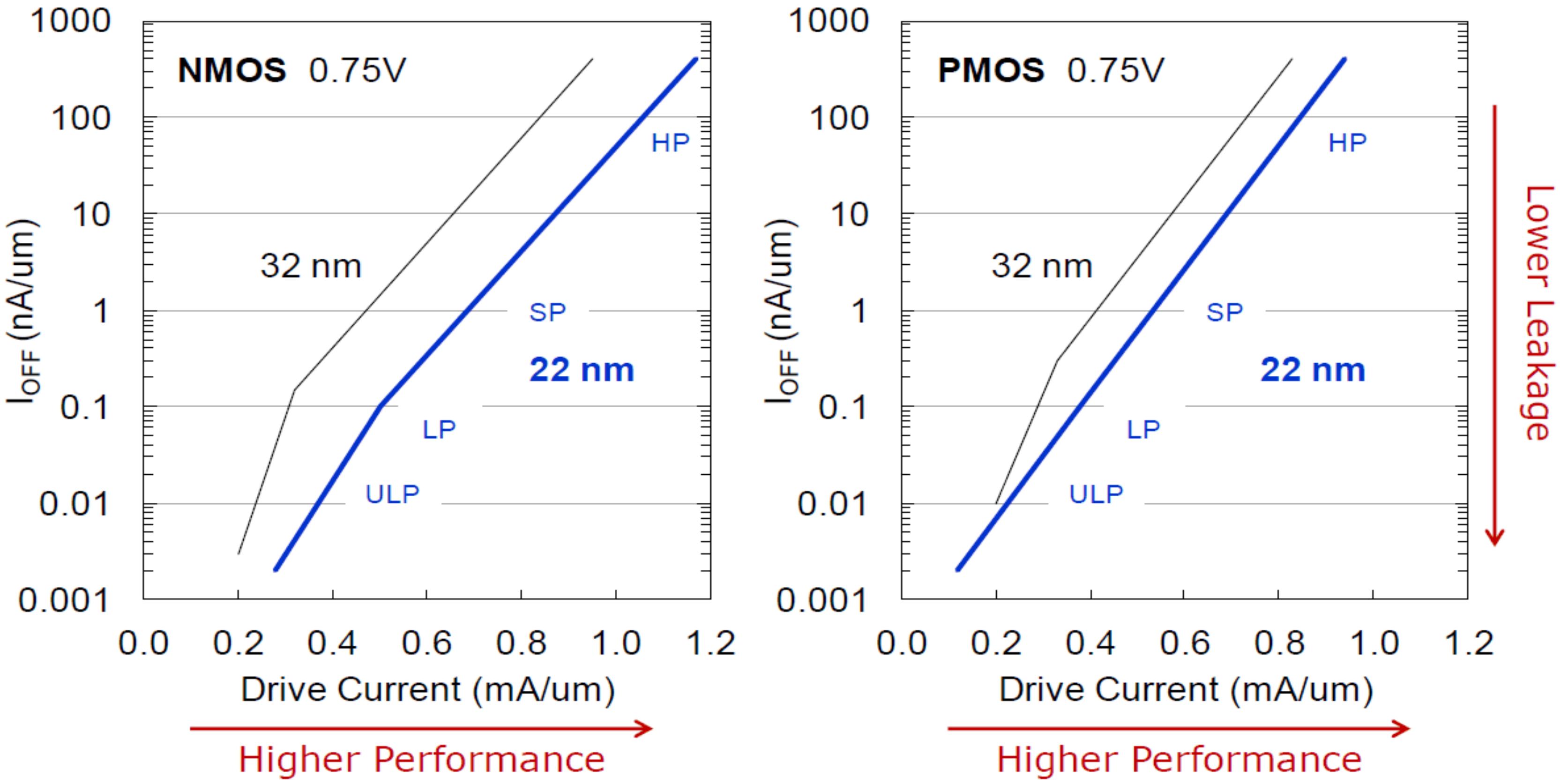
- As V_T decreases, sub-threshold leakage increases
- Leakage is a barrier to voltage scaling

Transistor Performance Increase



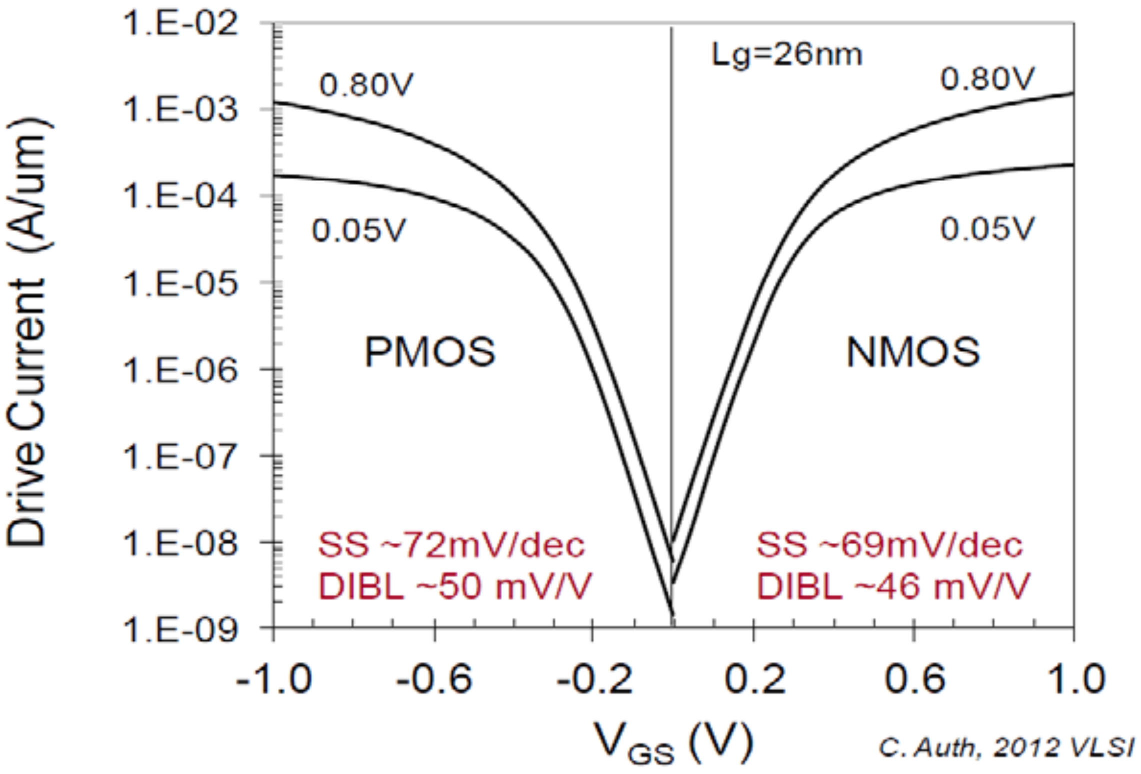
45 nm High-K and Metal-Gate provides average 30% drive current increase or >5x IOFF leakage reduction

Transistor Performance Increase



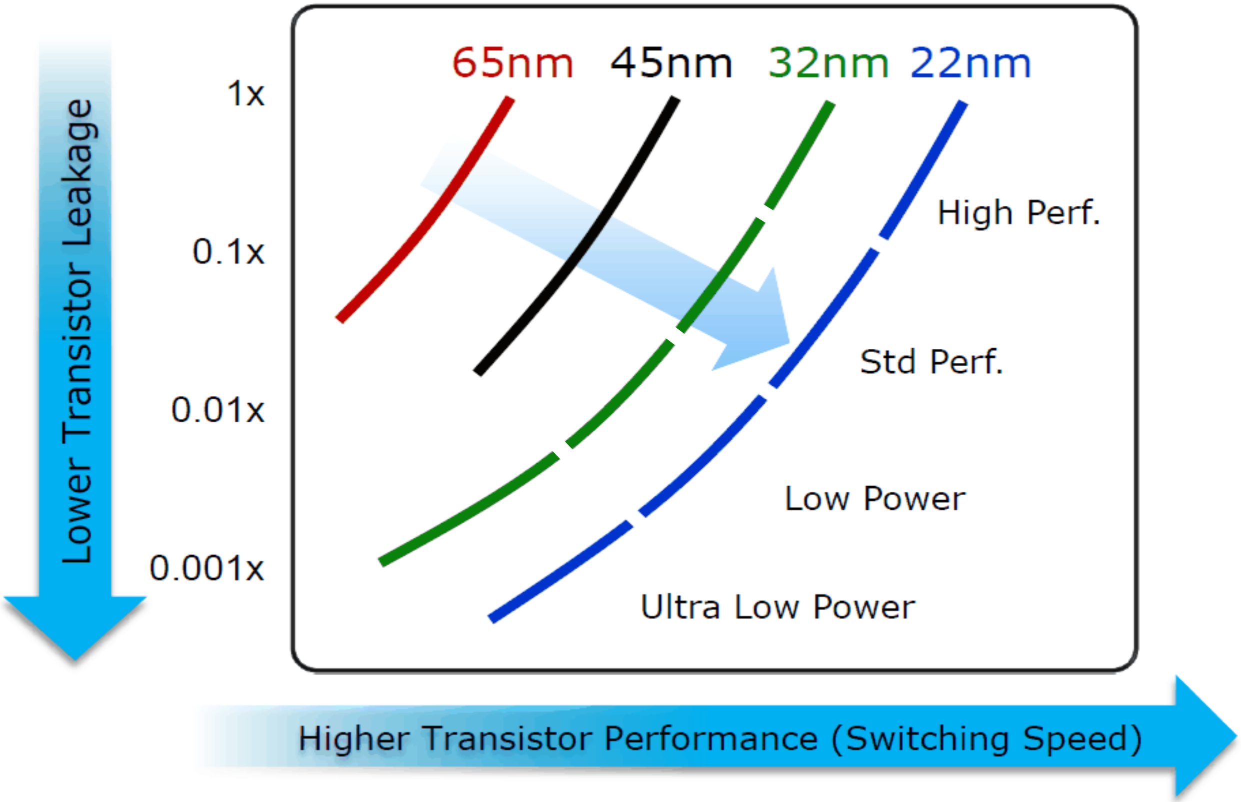
22 nm FinFET Transistors continue to provide a performance improvement at iso-IOFF or ~10X improvement in IOFF at iso-performance.

22nm FinFET I-V Curves



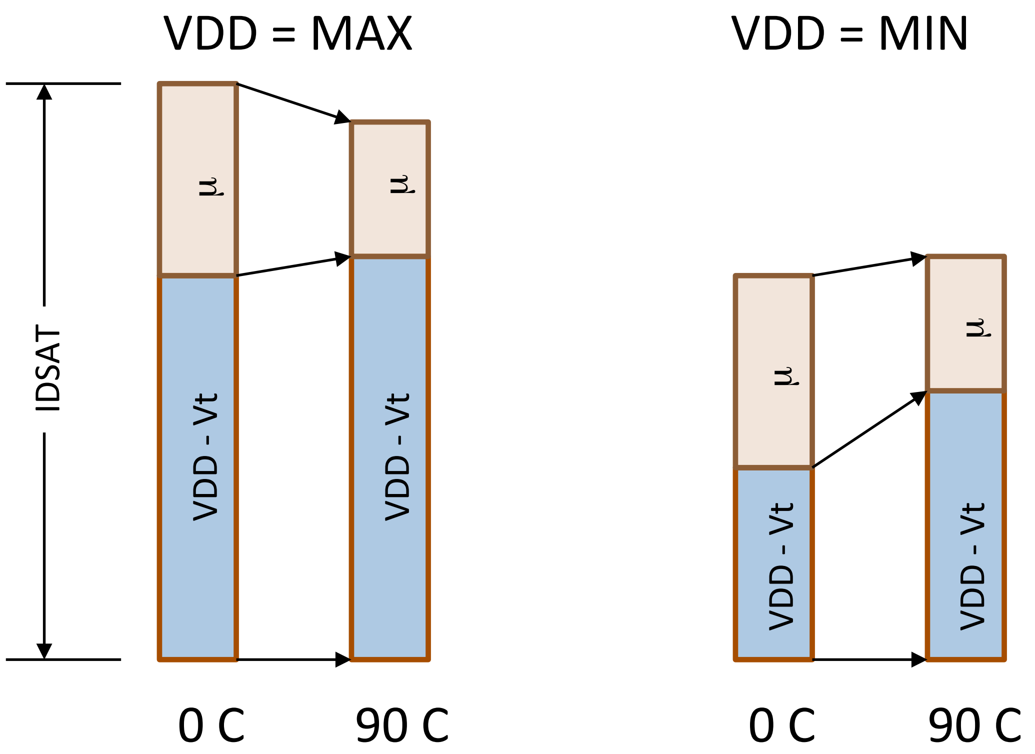
C. Auth, 2012 VLSI

TRANSISTOR PERFORMANCE vs. LEAKAGE



REVERSE TEMPERATURE EFFECT

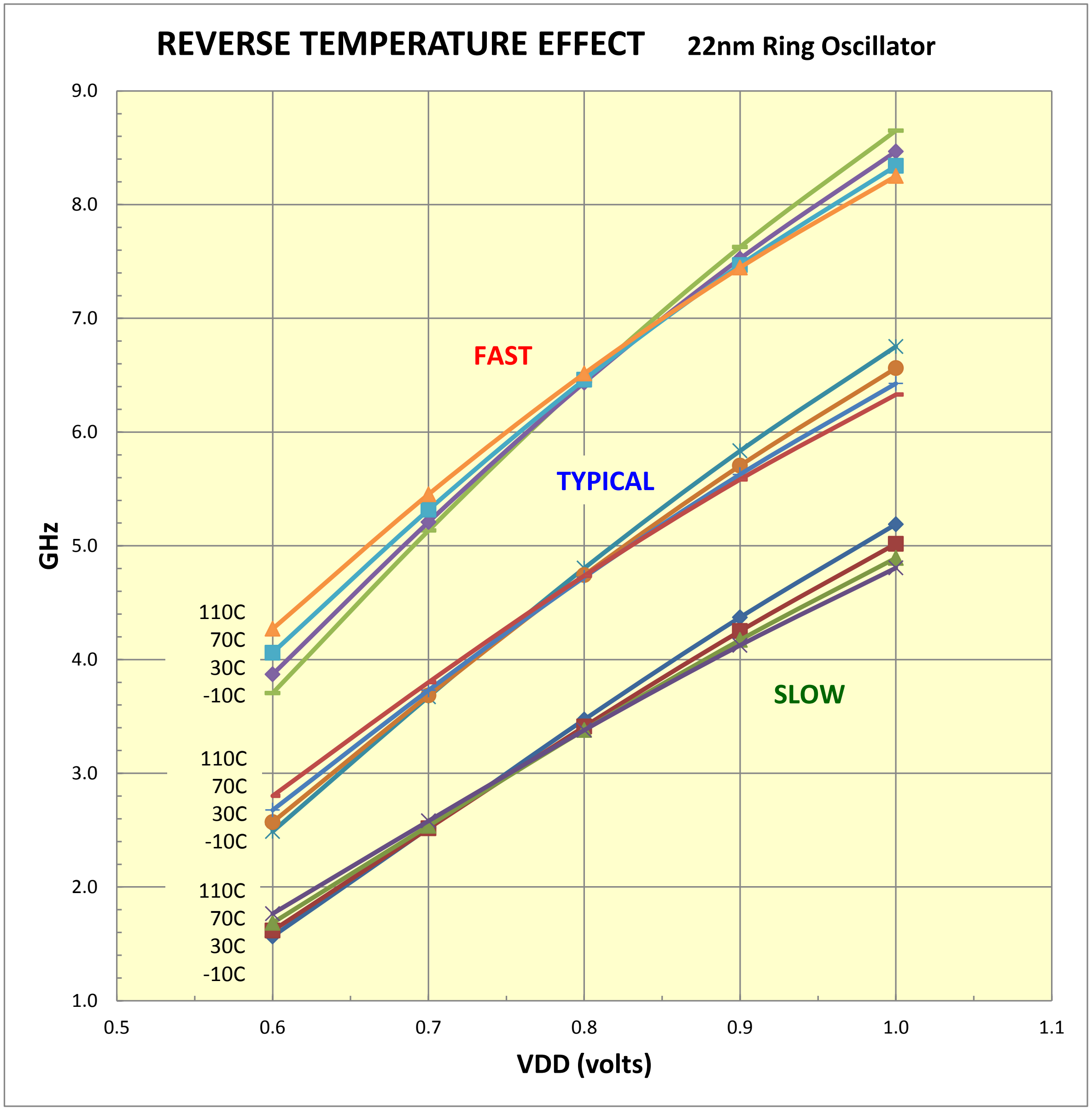
$IDSAT = \mu * W * (VDD - Vt)^{\alpha}$



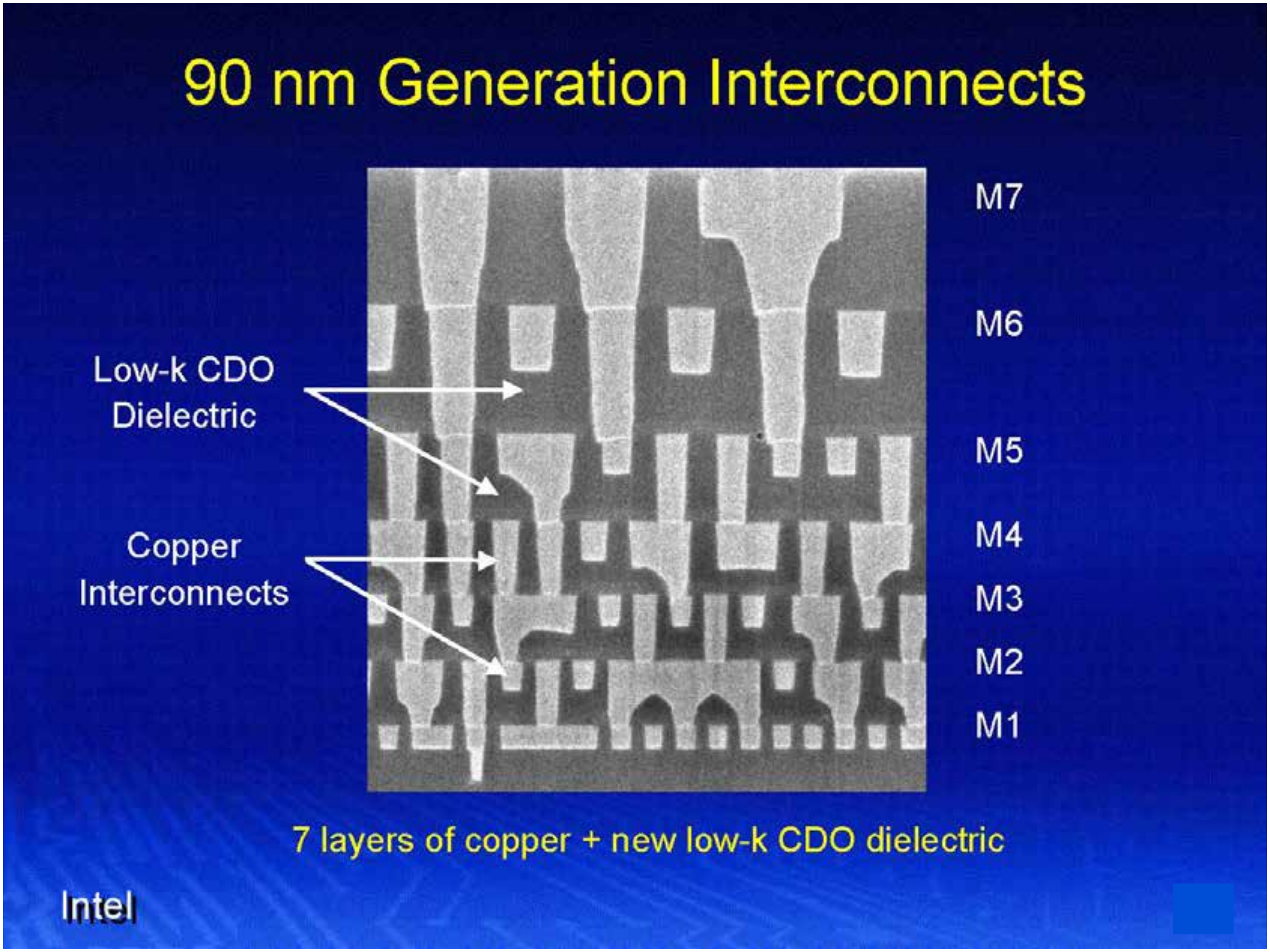
OBSERVATIONS:

- Mobility (μ) and **Vt** both drop at high temperature.
- The (VDD-Vt) quantity increases at high temperature, but it is not enough to compensate for the μ drop; as a result, IDSAT drops for high VDD values.
- The (VDD-Vt) quantity increases at a faster rate for low VDD; it is enough to compensate the μ drop. As a result, IDSAT increases for low VDD and high temperatures; this is called the REVERSE TEMPERATURE EFFECT.

REVERSE TEMPERATURE EFFECT



90nm INTERCONNECT

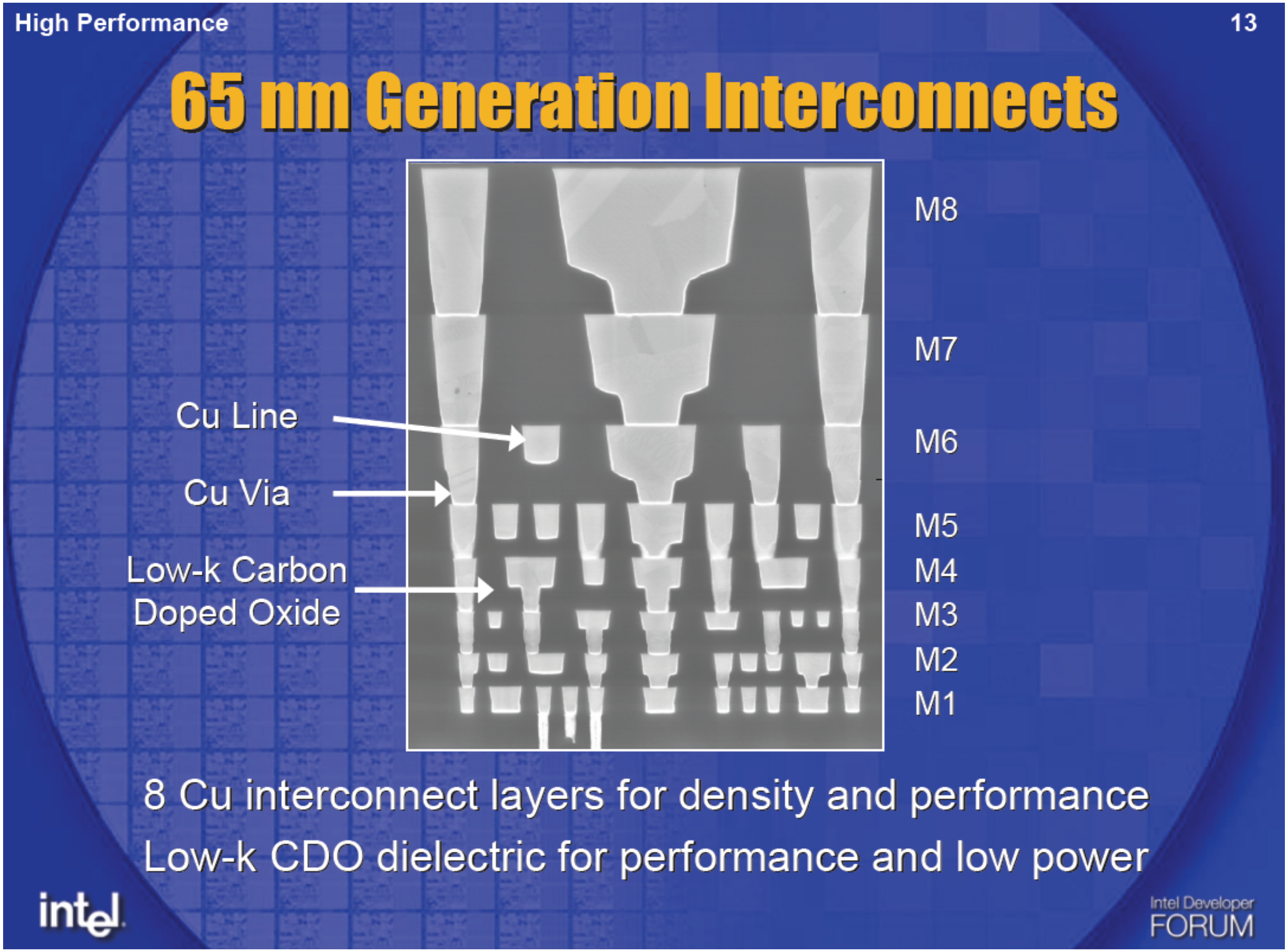


90 nm Generation Interconnects

7 layers of copper + new low-k CDO dielectric

Intel

65nm INTERCONNECT



High Performance

65 nm Generation Interconnects

8 Cu interconnect layers for density and performance

Low-k CDO dielectric for performance and low power

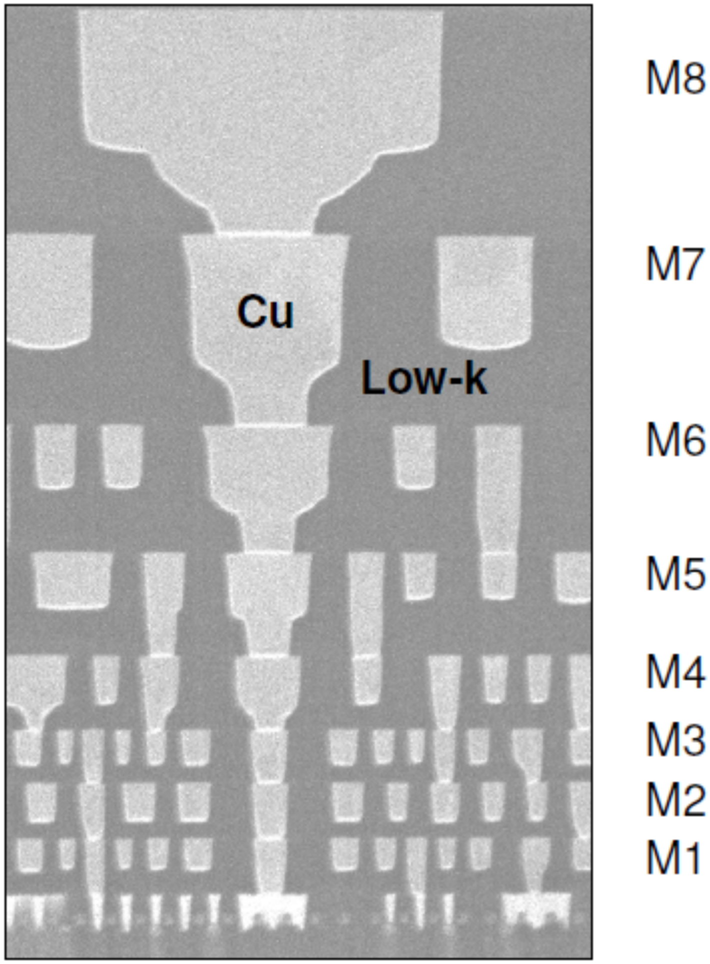
45nm Interconnect

Loose pitch + thick metal on upper layers:

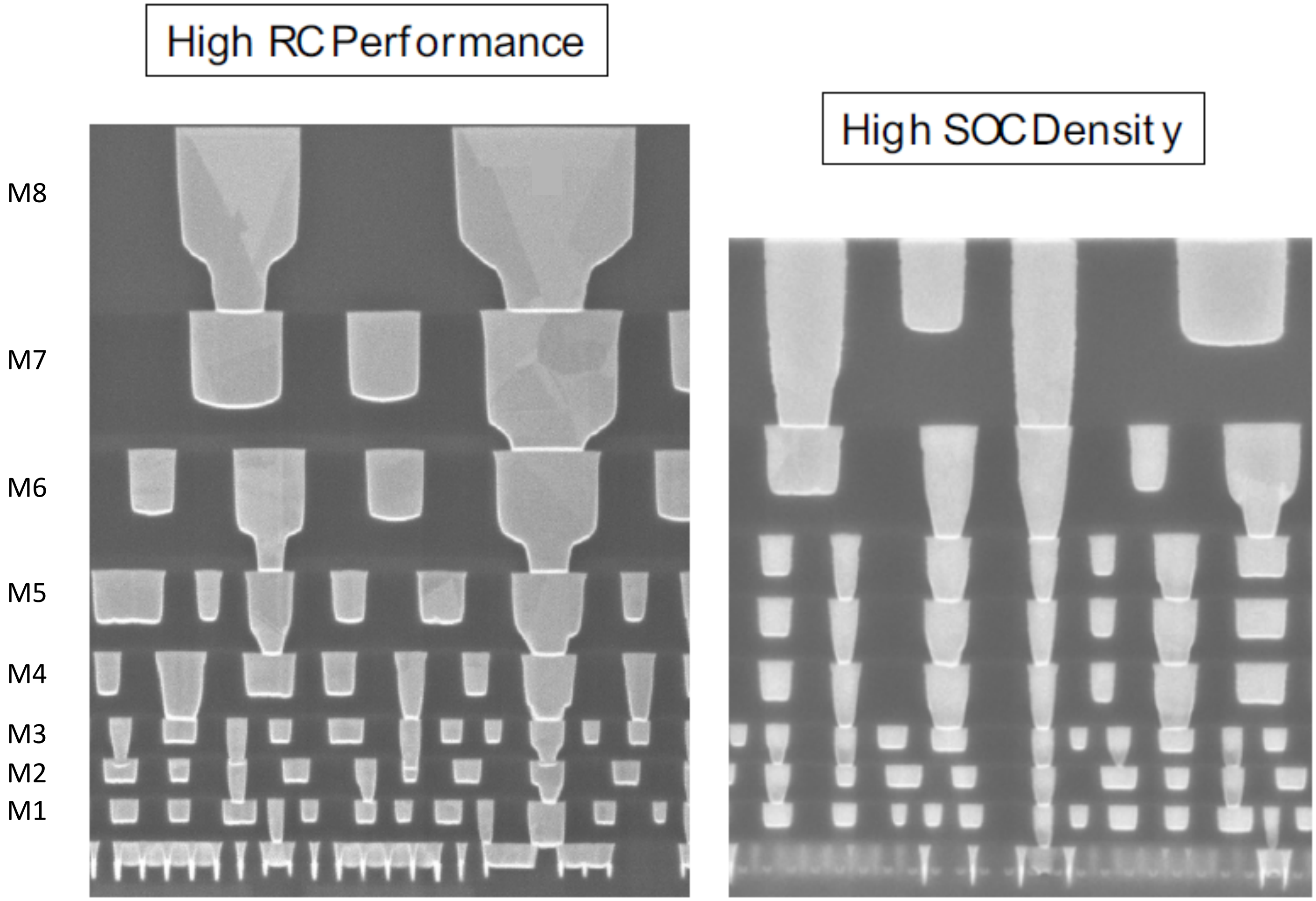
- High speed global wires
- Low resistance power grid

Tight pitch on lower layers:

- Maximum density for local interconnects



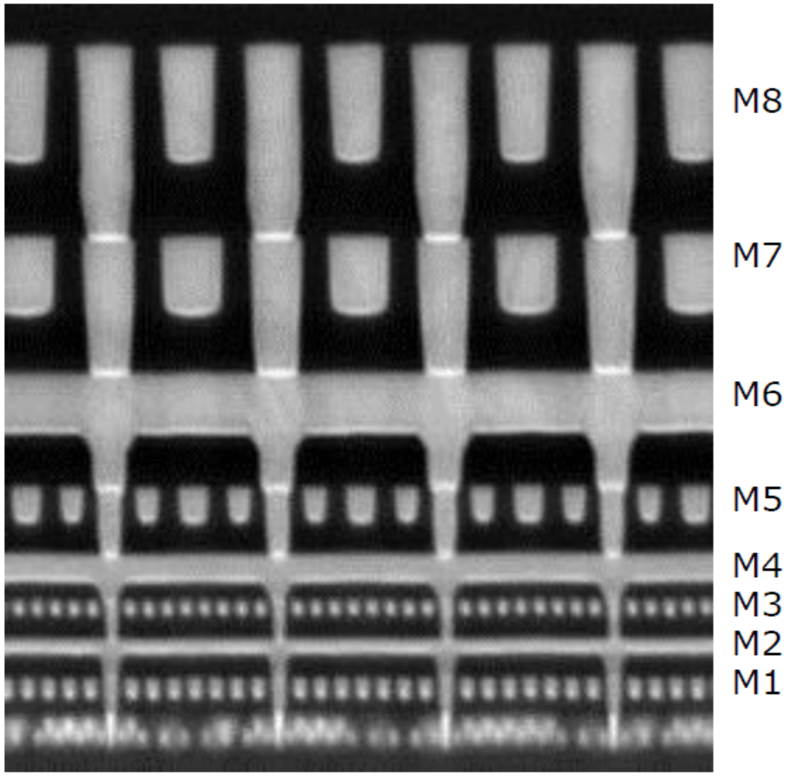
32nm Interconnect



C.-H. Jan, et al., “A 32nm SoC Platform Technology with 2nd Generation High-k/Metal Gate Transistors Optimized for Ultra Low Power, High Performance, and High Density Product Applications”, IEDM, Dec 2009.

22nm Interconnect

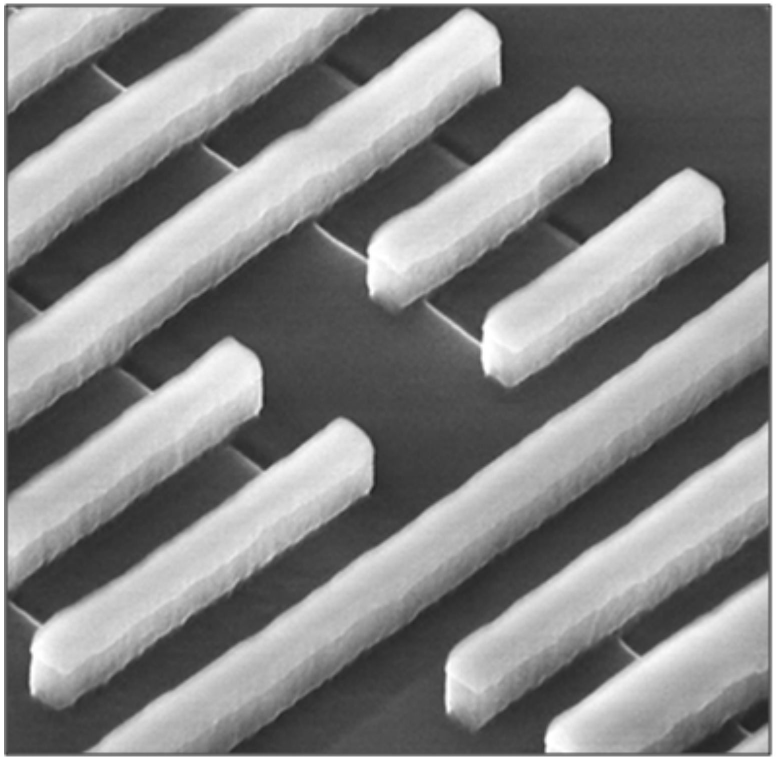
Layer	Pitch
TM	14 um
M8	360 nm
M7	320 nm
M6	240 nm
M5	160 nm
M4	112 nm
M3	80 nm
M2	80 nm
M1	90 nm



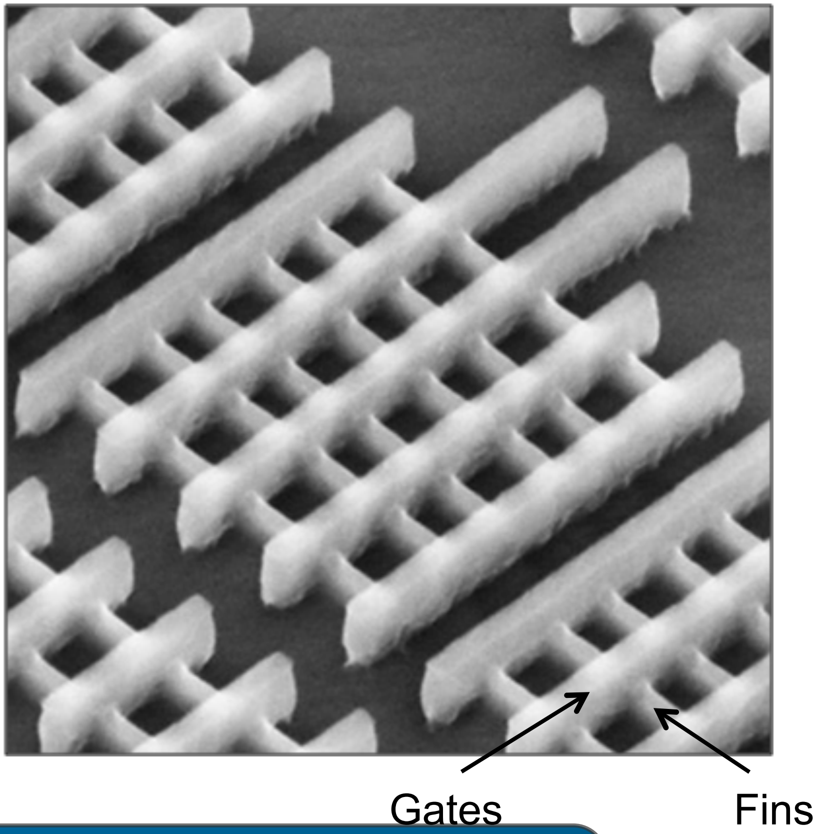
IDF 2012

FIN-FET Transistor Innovation *(courtesy: Mark Bohr, Sr. Intel Fellow)*

32 nm Planar Transistors

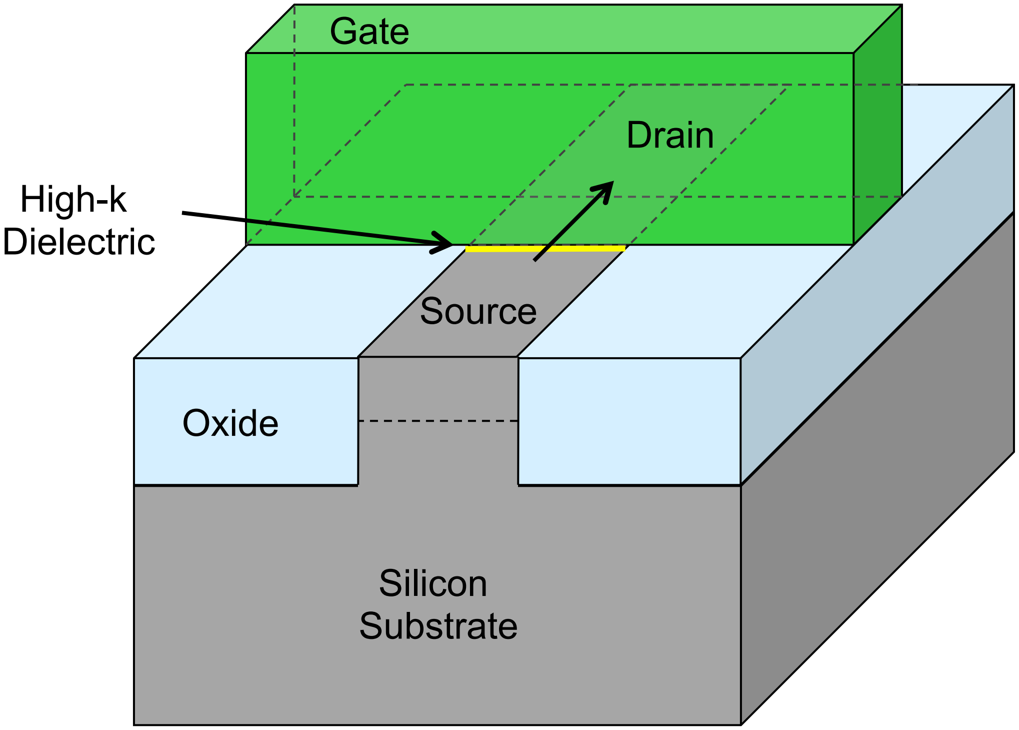


22 nm Tri-Gate Transistors



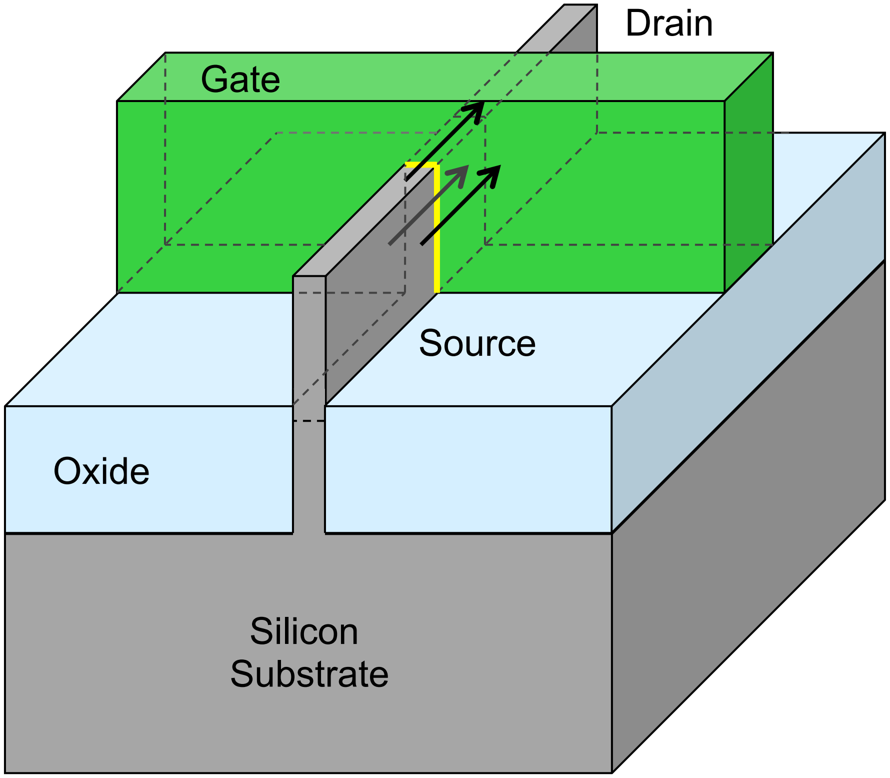
Intel's 22 nm technology introduces revolutionary 3-D Tri-Gate transistors

Traditional Planar Transistor



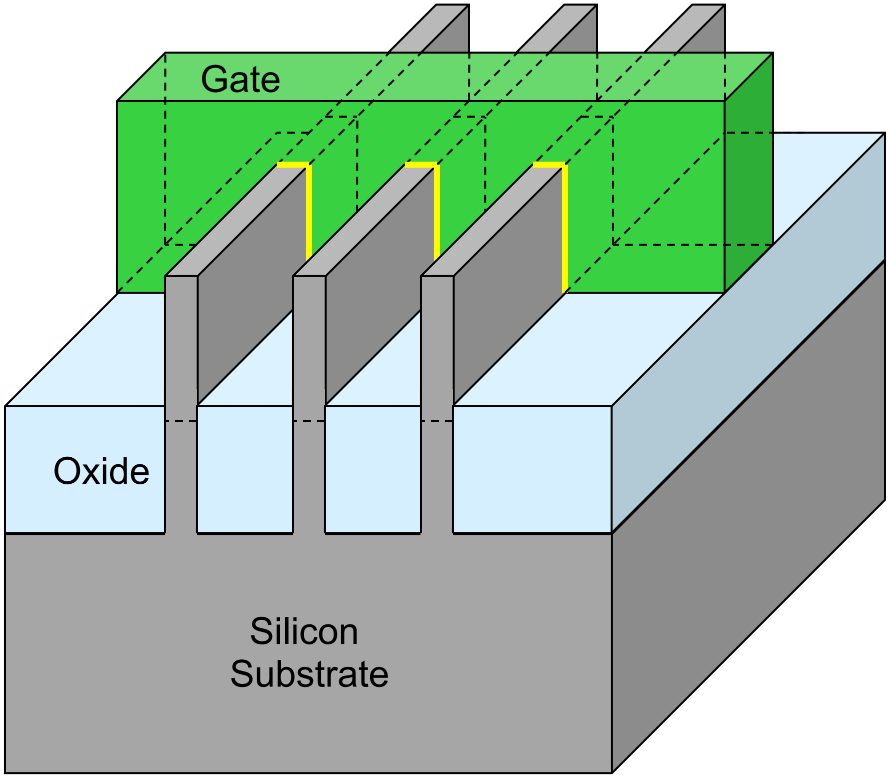
Traditional 2-D planar transistors form a conducting channel on the silicon surface under the gate electrode

22 nm FIN-FET Transistor



3-D Tri-Gate transistors form conducting channels on three sides of a vertical silicon fin

22 nm FIN-FET Transistor

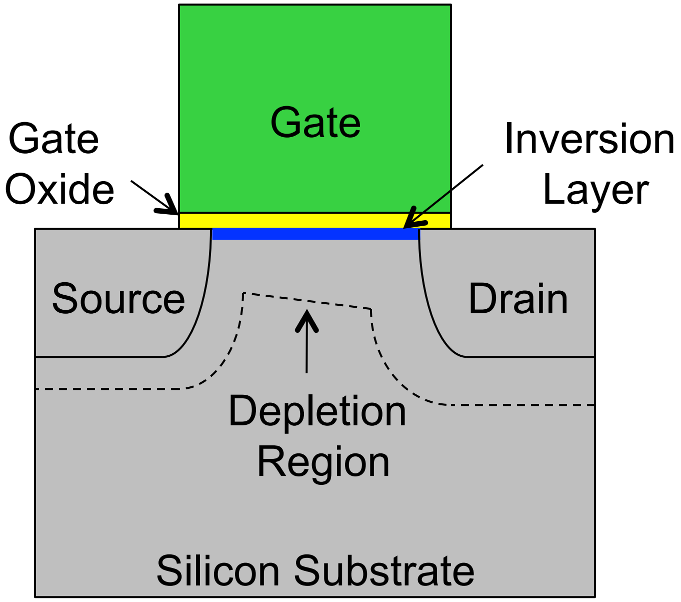


Tri-Gate transistors can connect together multiple fins for higher drive current and higher performance

22 nm FIN-FET Transistors

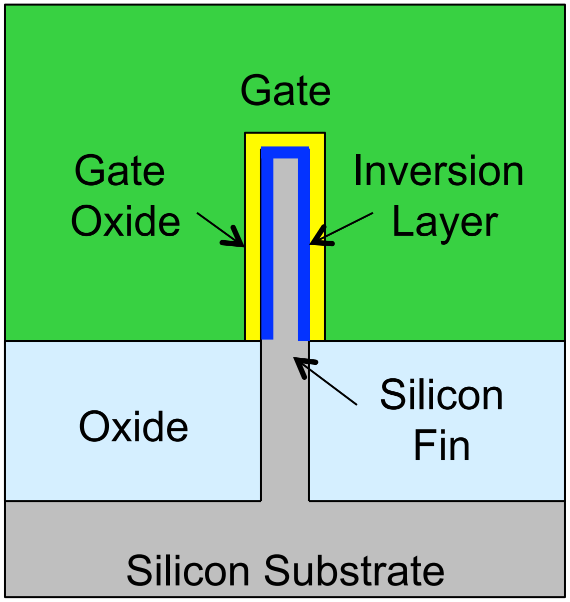
Planar Transistor

Not Fully Depleted



Tri-Gate Transistor

Fully Depleted

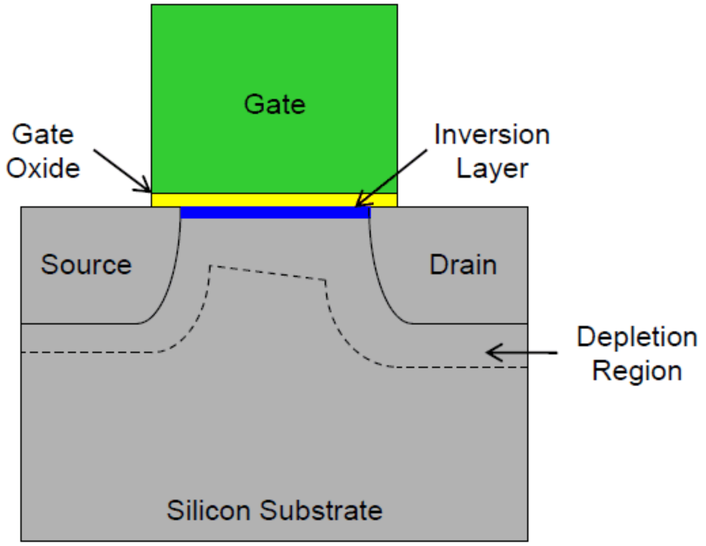


Tri-Gate transistors are "fully depleted" devices that have improved operating characteristics

Std vs. Fully Depleted Transistors

"Transistor 101"

Bulk Transistor



Silicon substrate voltage exerts some electrical influence on the inversion layer (where source-drain current flows)

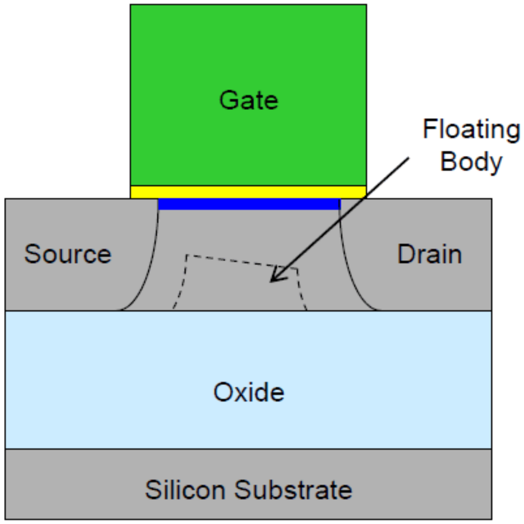
The influence of substrate voltage degrades electrical sub-threshold slope (transistor turn-off characteristics)

NOT fully depleted

Std vs. Fully Depleted Transistors

Partially Depleted SOI (PDSOI)

“Transistor 101”



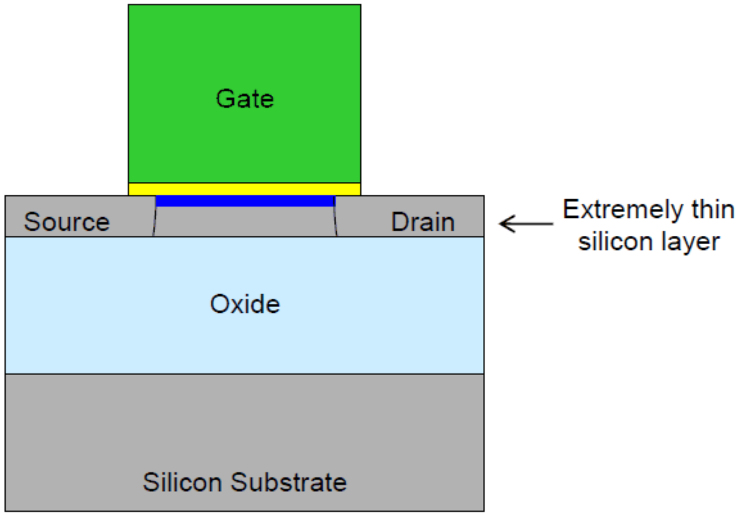
Floating body voltage exerts some electrical influence on the inversion layer, degrading sub-threshold slope

NOT fully depleted

Std vs. Fully Depleted Transistors

“Transistor 101”

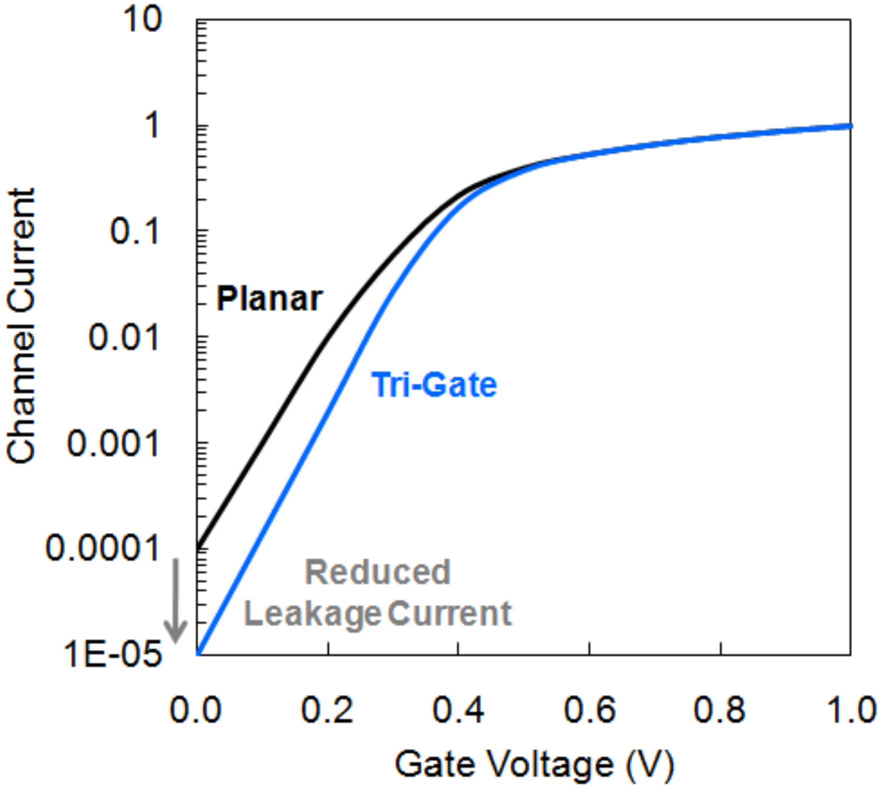
Fully Depleted SOI (FDSOI)

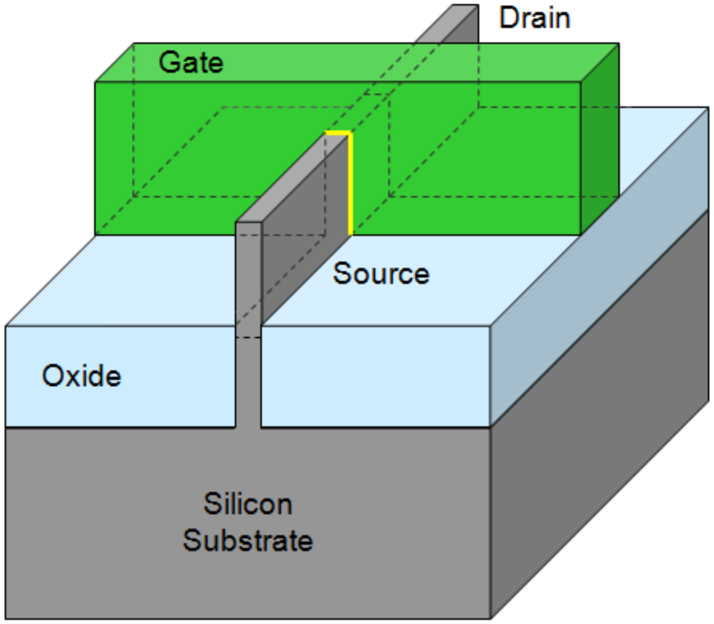


Floating body eliminated and sub-threshold slope improved

Requires expensive extremely-thin SOI wafer,
which adds ~10% to total process cost

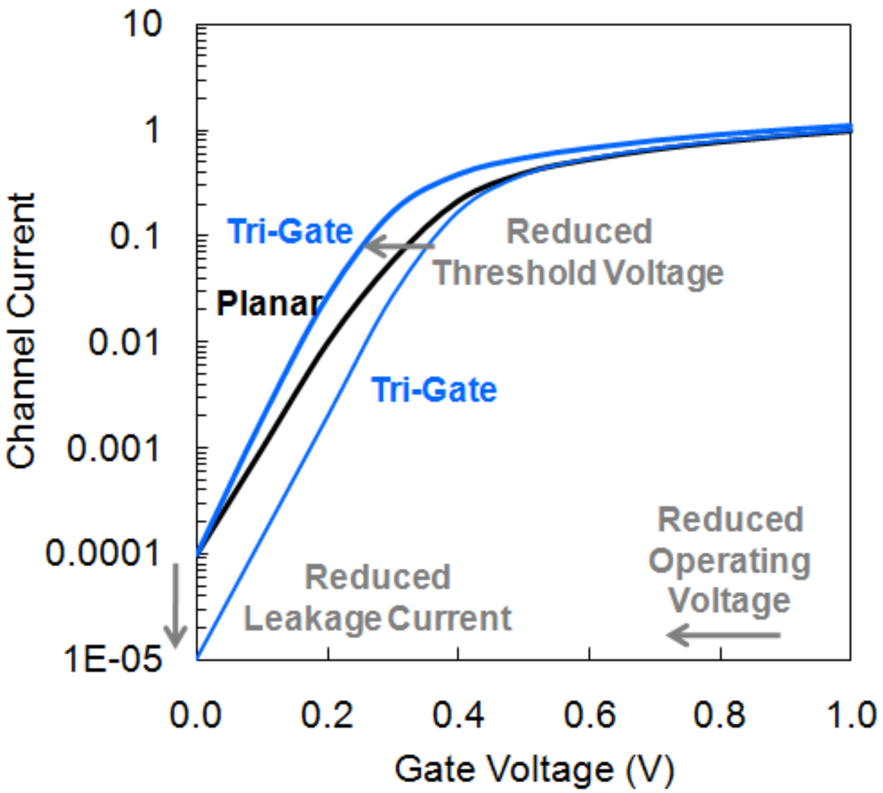
FIN-FET Low Leakage Benefit

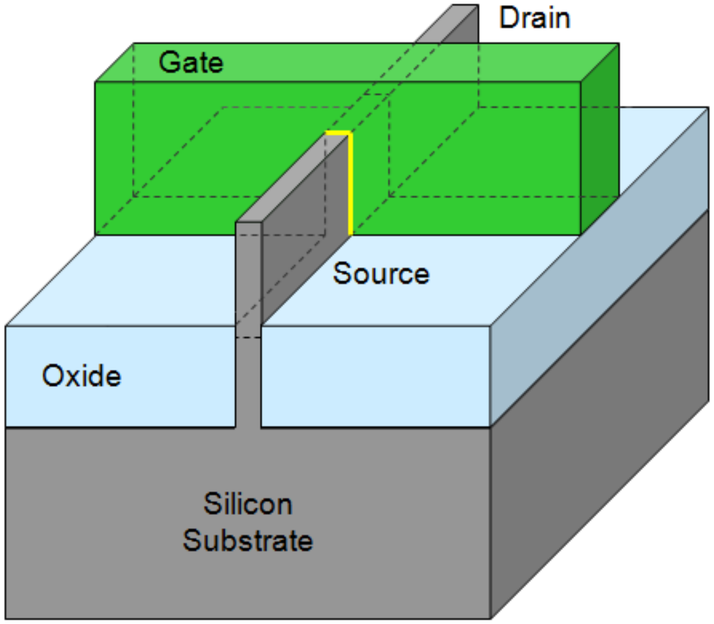




Steeper sub-threshold slope can provide lower leakage, higher performance and lower active power

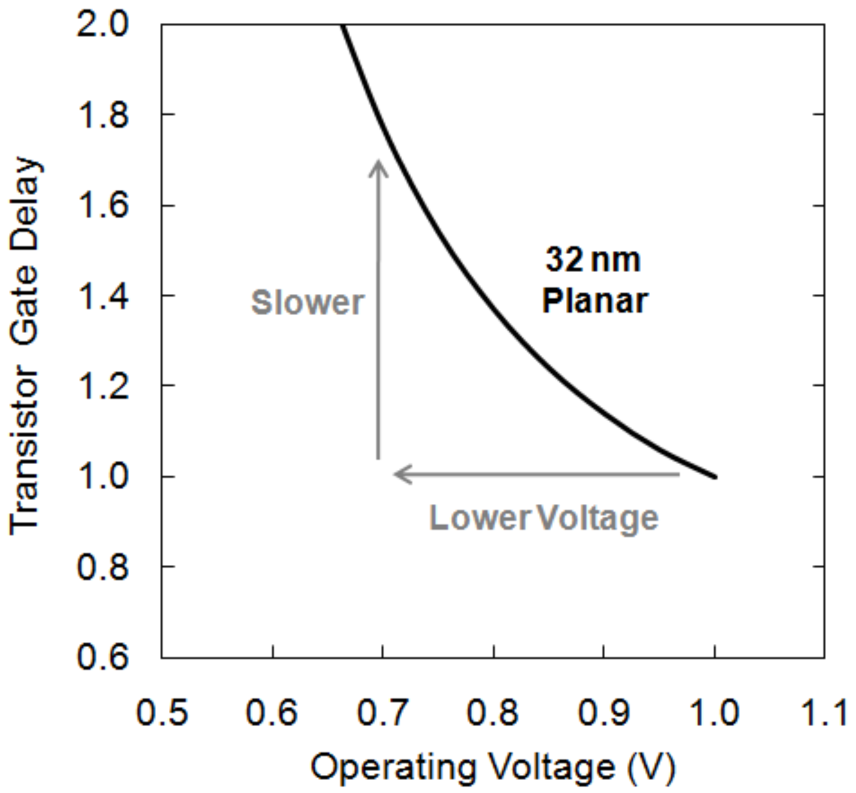
FIN-FET Performance/Power Benefit

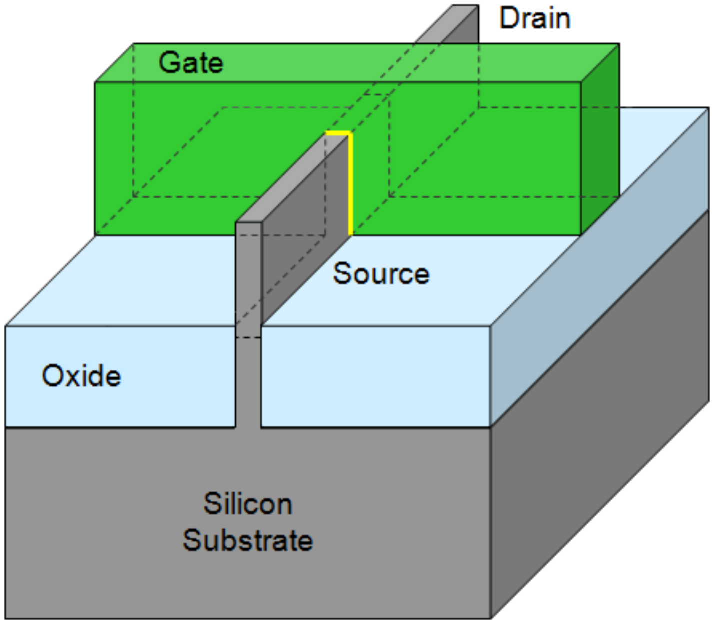




Steeper sub-threshold slope can provide lower leakage, higher performance and lower active power

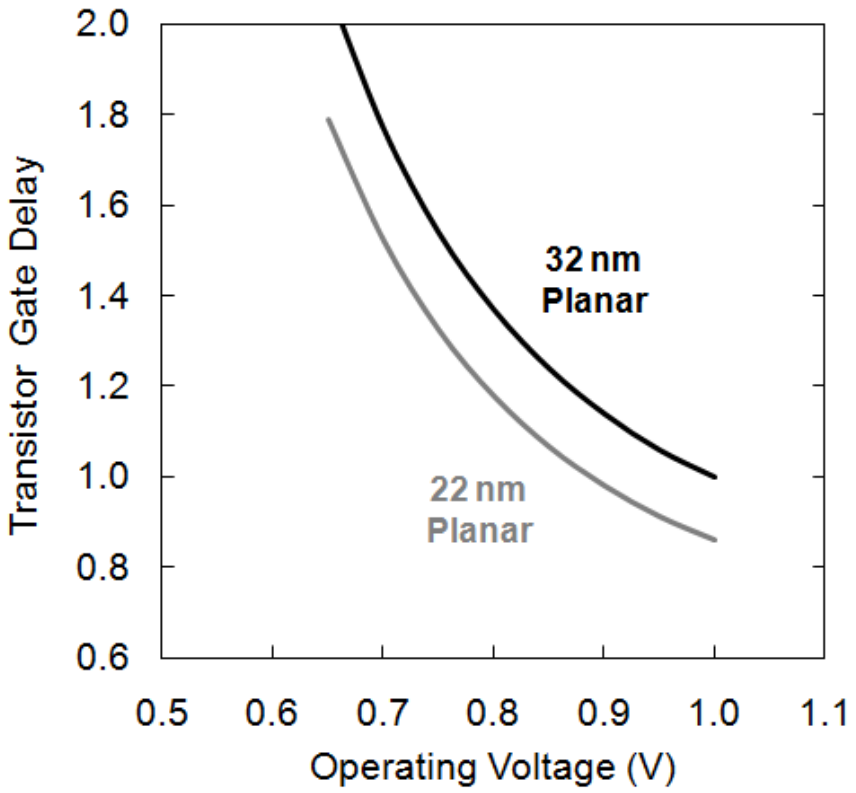
Transistor Delay vs. Voltage

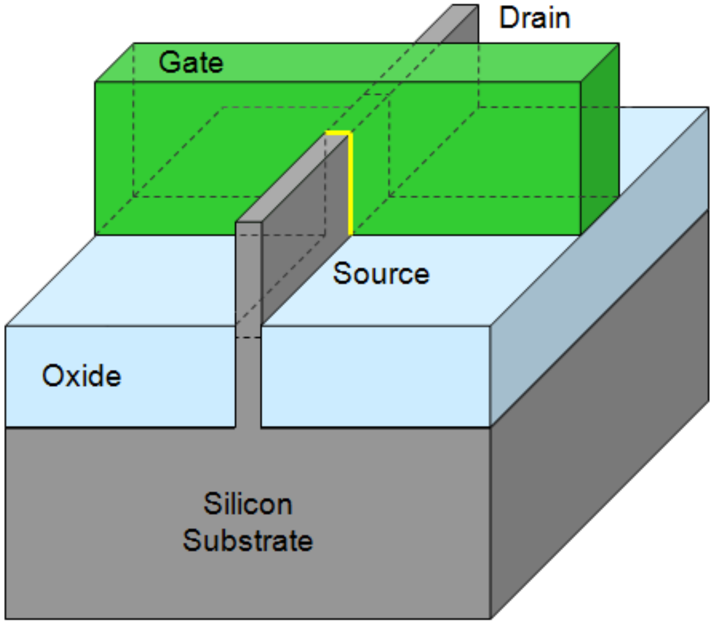




Gate delay is measured on a wide range of microprocessor circuit types, including gate-loaded inverters, interconnect-loaded inverters, NAND circuits, and MUX circuits.

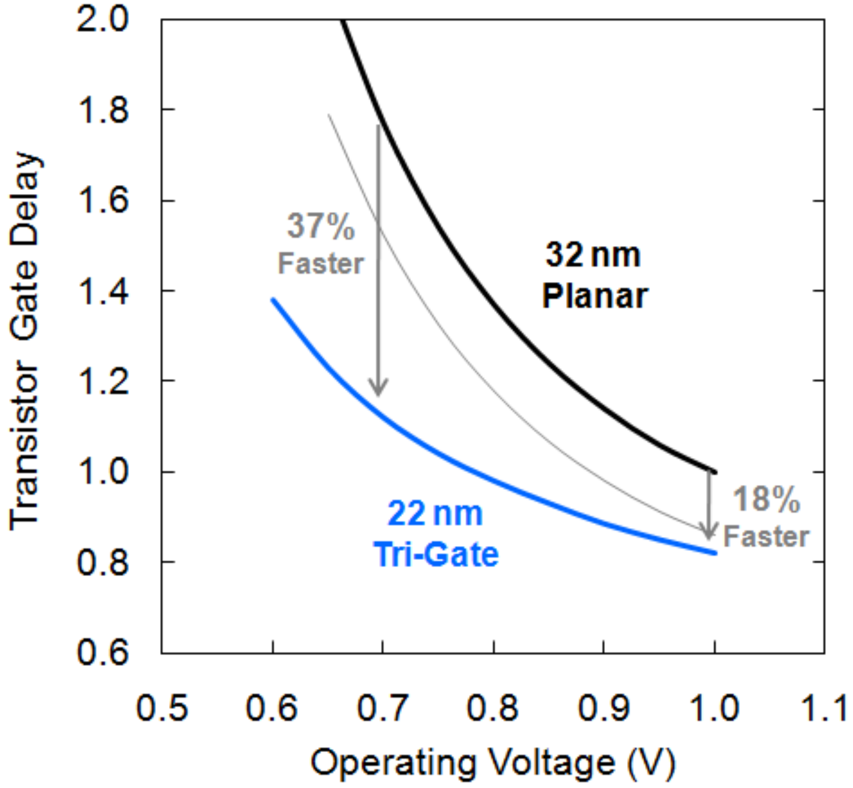
Transistor Delay vs. Voltage

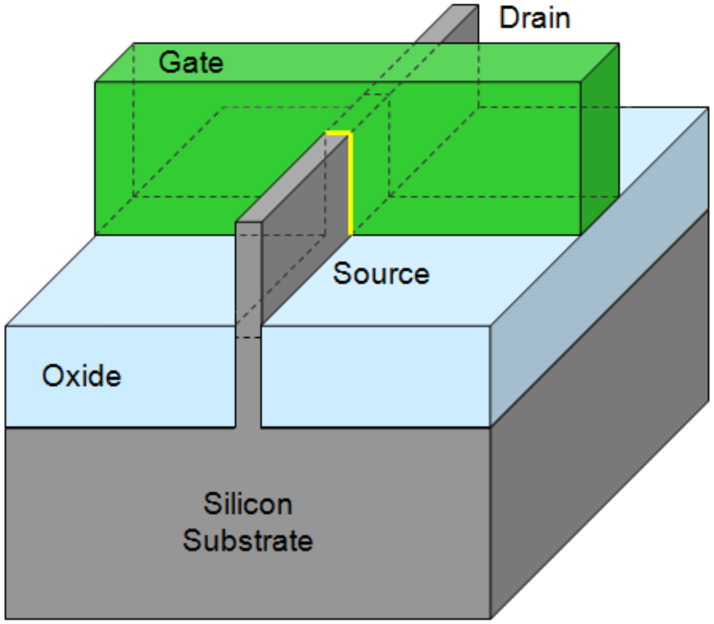




22 nm Planar transistors would provide only a modest improvement in delay vs. voltage

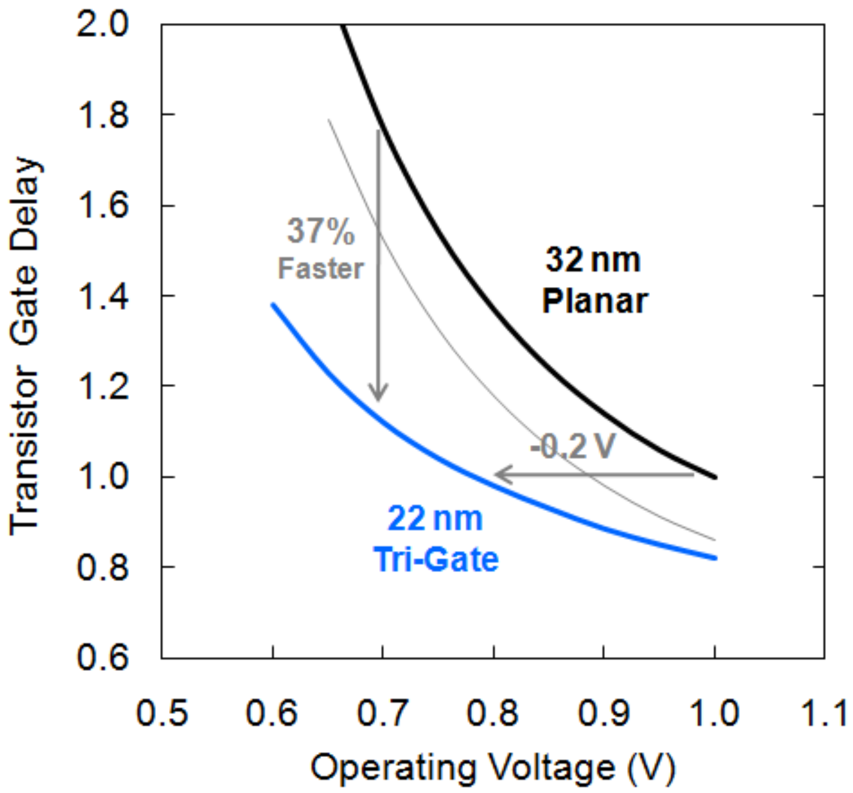
Transistor Delay vs. Voltage

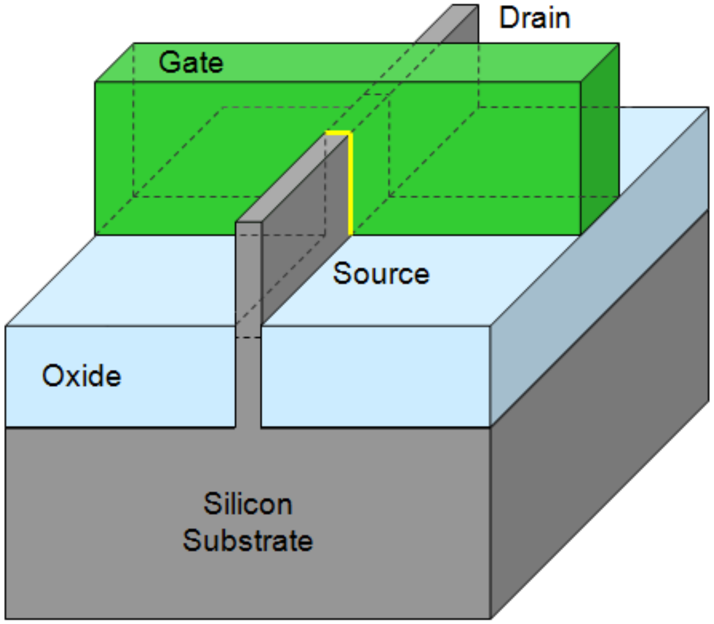




Tri-Gate transistors provide an unprecedented 37% delay improvement at low voltage

Transistor Delay vs. Voltage

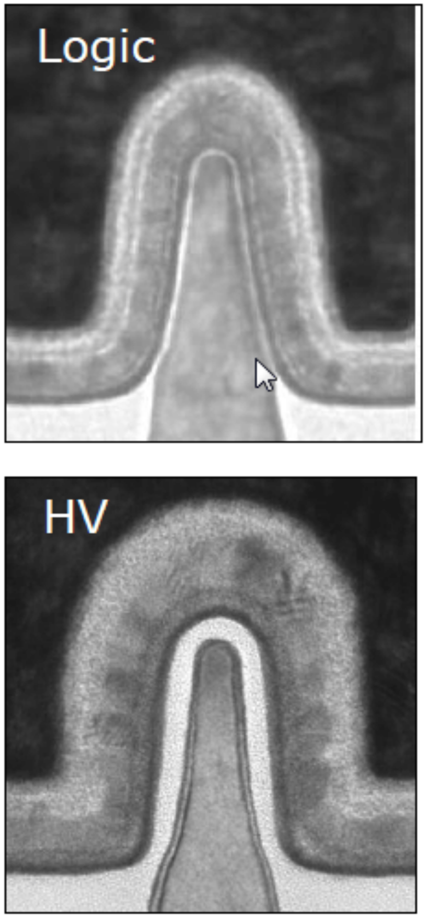




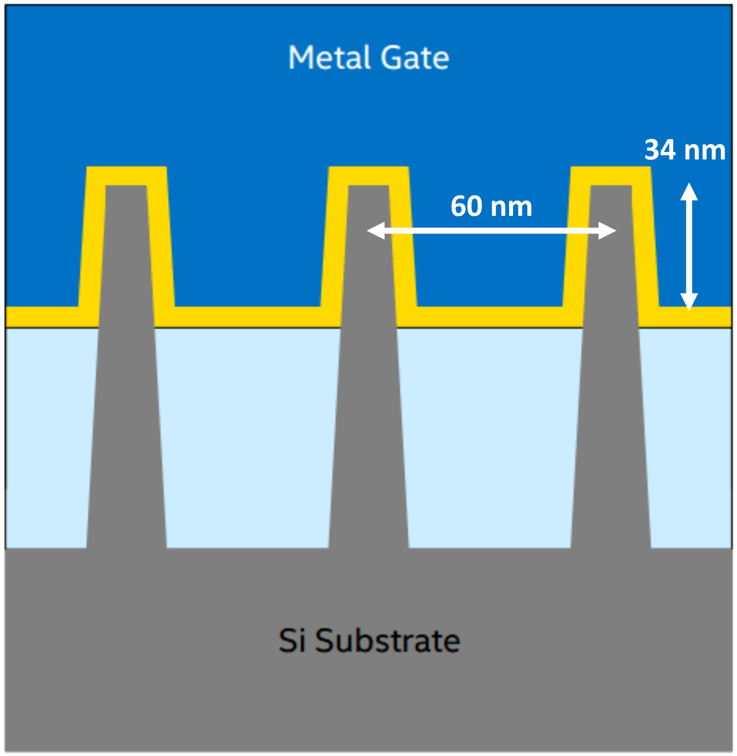
Tri-Gate transistors can operate at lower voltage, providing ~50% active power reduction

1.8V Transistors are also FIN-FET

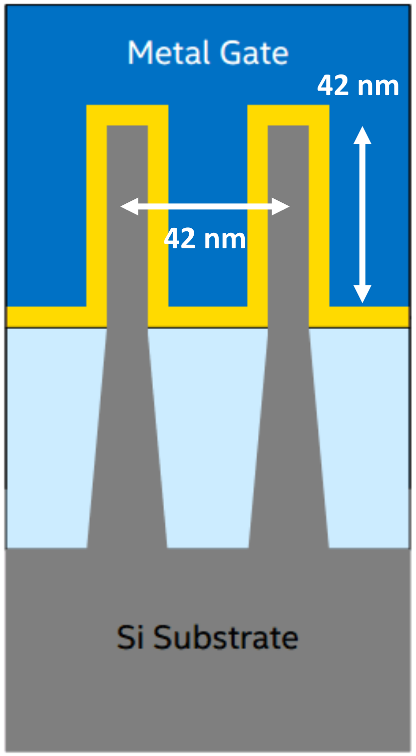
High Voltage
I/O Transistors



14nm Transistor FIN Improvements [15]



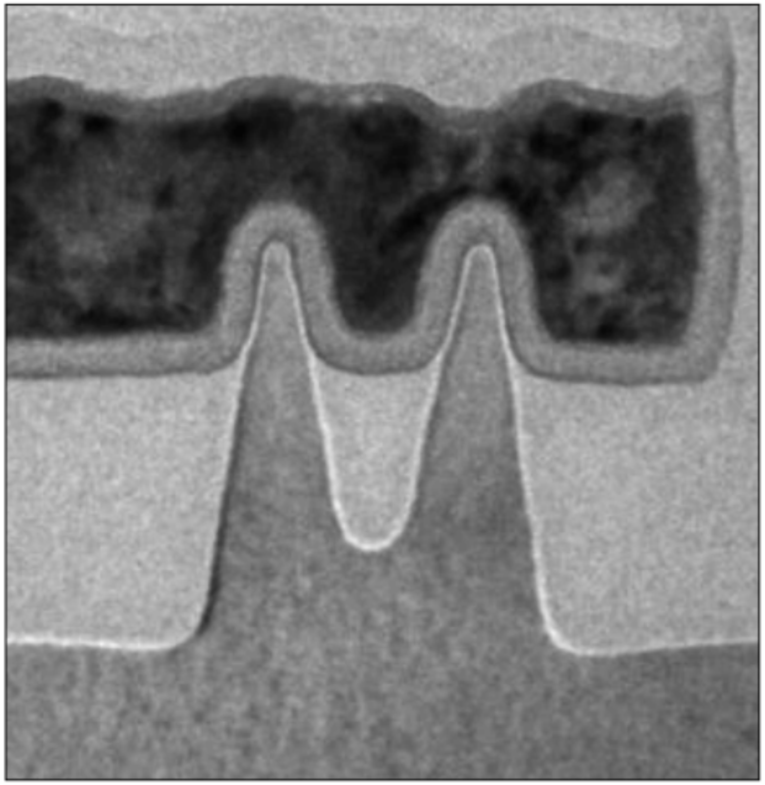
22 nm 1st Generation Tri-gate Transistor



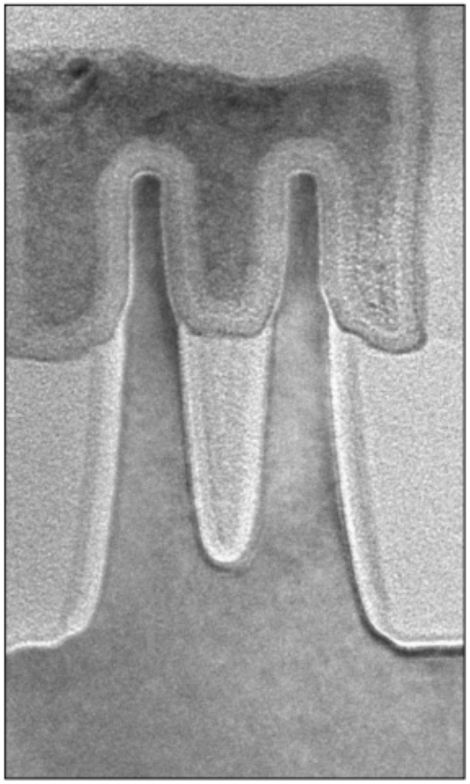
14 nm 2nd Generation Tri-gate Transistor

[15] ***Mark Bohr***, Intel Senior Fellow, Intel's 14nm Technology Announcement, August 11, 2014.

14nm Transistor FIN Improvements [15]



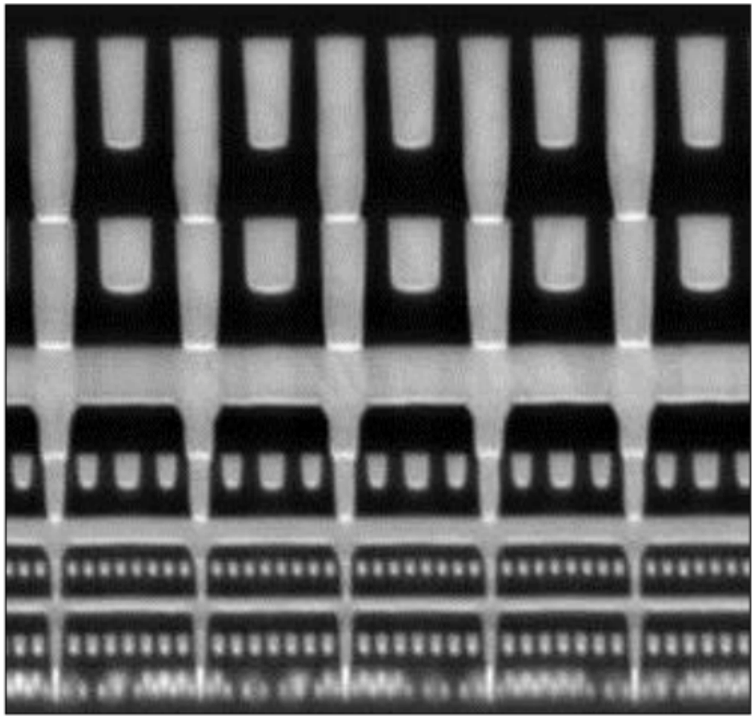
22 nm 1st Generation
Tri-gate Transistor



14 nm 2nd Generation
Tri-gate Transistor

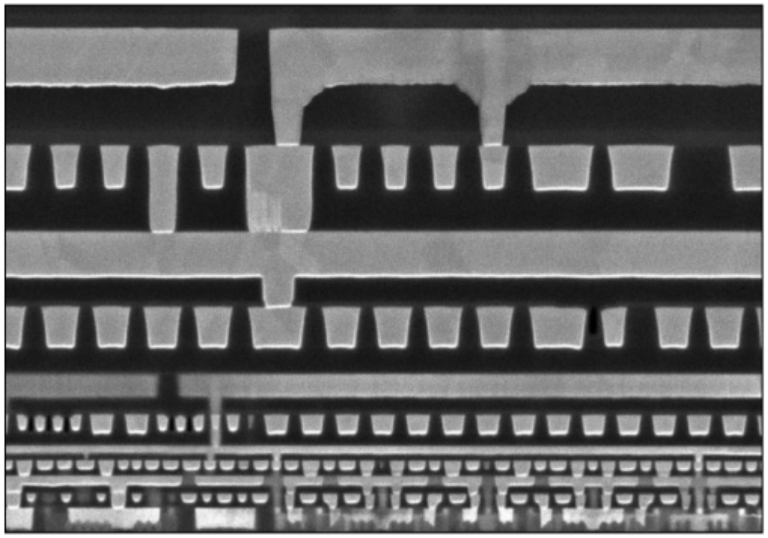
Interconnect Improvements [15]

22 nm Process



80 nm minimum pitch

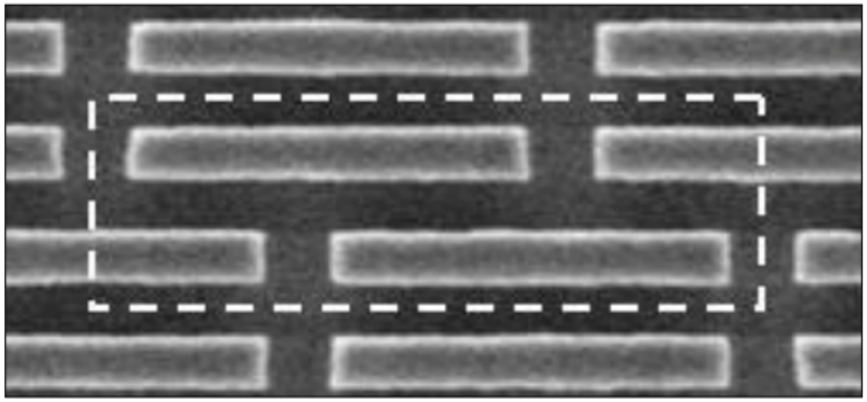
14 nm Process



52 nm (0.65x) minimum pitch

SRAM Memory Cell Improvements [15]

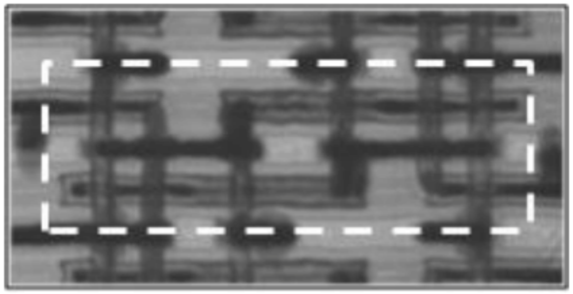
22 nm Process



.108 um^2

(Used on CPU products)

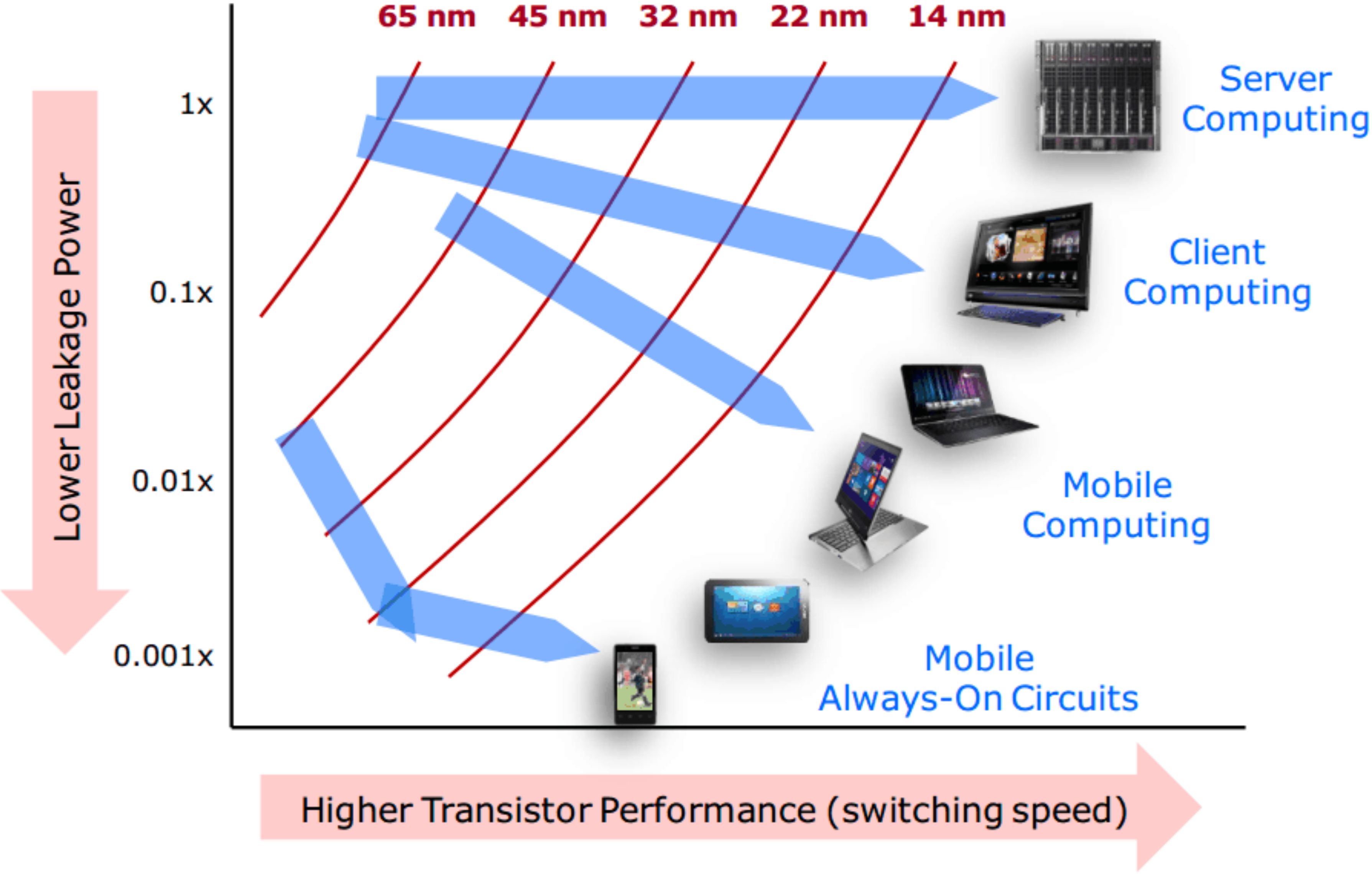
14 nm Process



.0588 um^2

(0.54x area scaling)

Transistor Performance vs. Leakage [15]



References

1. A. Chandrakasan, W.J. Bowhill, F. Fox, Design of High-Performance Microprocessor Circuits, IEEE Press, New York, 2001.

2. Y. Taur, T.H. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, UK, 1998.

3. K. Bernstein et. Al., High Speed CMOS Design Styles, Kluwer Academic Publishers, Boston, 1999.

4. R.J. Baker, H.W. Li, D.E. Boyce, CMOS Circuit Design, Layout, and Simulation, IEEE Press, New York, 1998.

5. J.M. Rabaey, Digital Integrated Circuits, Prentice-Hall, New Jersey, 1996.

6. S. Thompson, et. al., "An Enhanced 130nm Generation Logic Technology Featuring 60nm Transistors Optimized for High Performance and Low Power at 0.7-1.4V", 2001 IEDM Technical Digest, pp. 257-260.

7. S. Thompson, et. al., 90nm Technology, 2002 IEDM Technical Digest, pp. 61-64.

8. P. Bai, et. al., "A 65nm Logic Technology Featuring 35nm Gate Length, Enhanced Channel Strain, 8 Cu Interconnect Layersw, Low-k ILD and 0.57um2 SRAM Cell", 2004 IEDM.

9. M. Bohr, Intel's 90nm Process Starting High Volume Manufacturing: http://www.intel.com/research/silicon

10. K. Mistry, et al., "A 45nm Logic Technology with High-k+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging", Tech. Digest IEDM, Dec 2007.

11. Intel's 45nm CMOS technology: http://www.intel.com/technology/architecture-silicon/45nm-core2/index.htm?iid=tech_45nm+45nm

10. P. Packan, et al., "High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors", Tech. Digest IEDM, Dec 2009.

11. C.-H. Jan, et al., "A 32nm SoC Platform Technology with 2nd Generation High-k/Metal Gate Transistors Optimized for Ultra Low Power, High Performance, and High Density Product Applications", IEDM, Dec 2009.

12. For latest information on Intel's silicon technology, please visit: http://www.intel.com/technology/architecture-silicon/index.htm?iid=tech_home+body_architecturesilicon

13. Intel's 22nm tri-gate transistors: http://www.intel.com/content/www/us/en/silicon-innovations/silicon-technology-leadership-presentation.html

14. Chris Auth, et al., "A 22nm High Performance and Low-Power CMOS Technology Featuring Fully-Depleted Tri-Gate Transistors Self-Aligned Contacts and High Density MIM Capacitors", VLSI Symposium Technical Digest, June 2012.

15. Mark Bohr, Intel Senior Fellow, Intel's 14nm Technology Announcement, August 11, 2014.

Homework #1

High performance Hi-K, metal gate planar 22nm and FinFET 14nm transistors will be characterized in this homework assignment. Models are from the Arizona State University's Predictive Technology Models website: http://ptm.asu.edu/.

1. Measure IDSAT (mA/um) for both NFET and PFETs in 22nm and 14nm CMOS technologies at T=30C, TYPICAL process corner and Vgs=Vds=0.9V. Be aware that each Fin in the 14nm transistors have an effective width of 56nm.
2. Measure Ioff (Amps/um) for both NFET and PFETs in 22nm and 14nm CMOS at 30C, TYPICAL process corner and Vgs=0.0V, Vds=0.9V. Monitor the SOURCE current and not the DRAIN current; why? Plot LOG(ABS(Ids)) .vs. Vgs of the 4 transistors.
3. Measure Cgate (fF/um) for the 4 FETs by integrating the current going into ONE FET in the provided 9-stage FO4 ring oscillator over ONE VDD transition. Using Q=CV, Cgate = (1/VDD)*Integral (Ig(t)dt). Note that the AREA under the Ig(t) curve is the integral over ONE transition.
4. Calculate the Vt's of the 4 FETs using the extrapolated method using the Ids and Gm .vs. Vgs transistor characteristics as follows:
    - Apply 50 milli-volts to Vds; sweep Vgs from 0.0 to 0.9 volts. Gather 91 points.
    - Plot Ids .vs. Vgs; Differentiate Ids with respect to Vgs (this is the Gm of the transistor); plot Gm along the Ids .vs. Vgs curve; Gm will 'peak' at some Vgs value. Draw a tangent on the Ids .vs. Vgs curve at this Vgs value. The Vt of the transistor will be the Vgs value where Ids=0.0 mA 'minus' Vdd/2 (or 25 milli-volts).
    - Use negative values for the PFET characterization; ie. apply -50 milli-volts to Vds, and sweep Vgs from 0.0 to -0.9V. Ids will be negative (Gm will still be positive); add + 25 milli-volts to the final result.
    - How does your 'extrapolated' Vt value compare with the model's Vth0 value?
    - Include all relevant plots.
5. Determine the gate delay of an INVERTER with FO4 (9-stage ring oscillator provided) for both 22nm and 14nm technologies at 30C and TYPICAL process models. Plot stage DELAY (ps) .vs. VDD (volts) for VDD=0.5V to 0.9V in 0.1V increments. Gate delay should be the average of both transitions.
6. What is the performance increase at a constant VDD between the 22nm and 14nm CMOS technologies? What is the power reduction at the same performance level (or iso-GATE DELAY)? Is this performance increase and power reduction constant as a function of VDD?
7. Are these 14nm transistors scaling based on a constant electric field scaling principle?
8. Plot Ids .vs. Vds (0.0V to 0.9V in steps of 0.05V) and Vgs=0.1, 0.3, 0.5, 0.7, and 0.9V for all 4 FETs; use a width=1.0um. Other than the increase in drive current (Ids), what else is different between the planar 22nm transistors and the 14nm FinFET transistors?

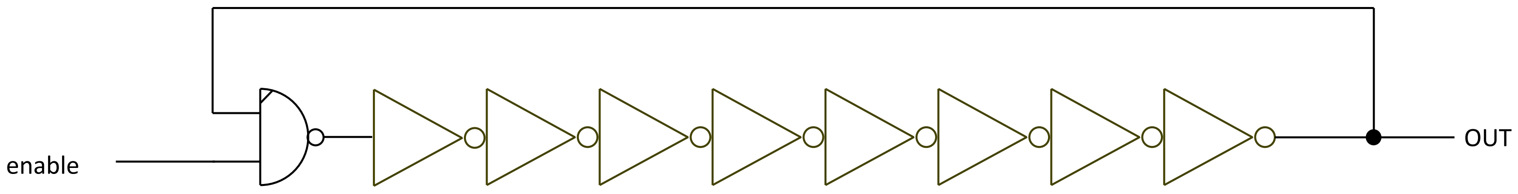
Homework #1

Collaterals you will need to complete Homework Problem 1:

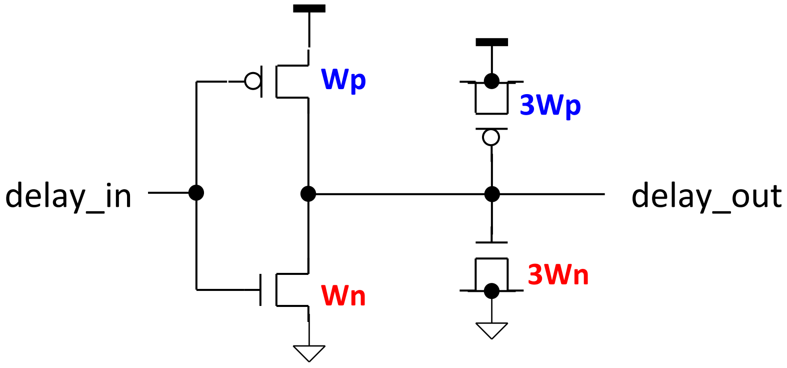
/home/projects/courses/spring_16/ee382m-16745/vlsi/main/hw1

hw1_22nm.sp: top-level HSPICE netlist & simulation deck for a 9-stage FO4 ring oscillator in 22nm
hw1_14nm.sp: top-level HSPICE netlist & simulation deck for a 9-stage FO4 ring oscillator in 14nm
block_22nm.txt: netlist for a 22nm inverter delay element with a built-in FO3
block_14nm.txt: netlist for a 14nm inverter delay element with a built-in FO3
14nm_NMOS_HI_K_HP.txt: NFET MOSFET models for a 14nm semiconductor technology
14nm_PMOS_HI_K_HP.txt: PFET MOSFET models for a 14nm semiconductor technology
22nm_NMOS_HI_K_HP.txt: NFET MOSFET models for a 22nm semiconductor technology
22nm_PMOS_HI_K_HP.txt: PFET MOSFET models for a 22nm semiconductor technology
library-14nm.txt: library of 14nm circuit elements (logic gates, transmission gates, tristate inverters, MSFF, and interconnect elements)
library-22nm.txt: library of 22nm circuit elements (logic gates, transmission gates, tristate inverters, MSFF, and interconnect elements)

9-stage Ring Oscillator

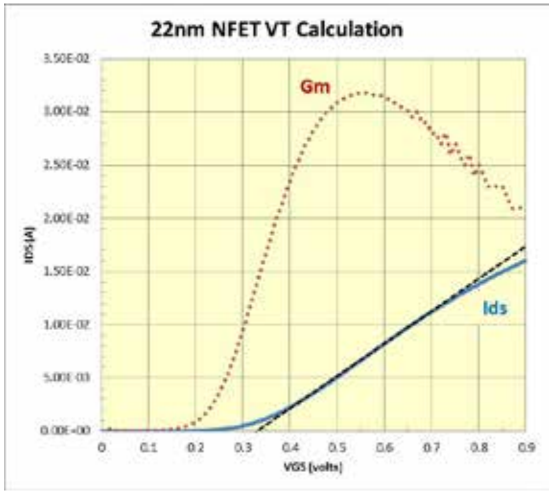


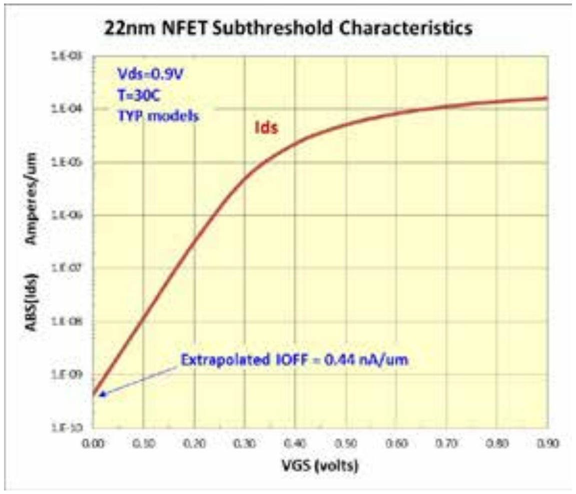
9-stage Ring Oscillator

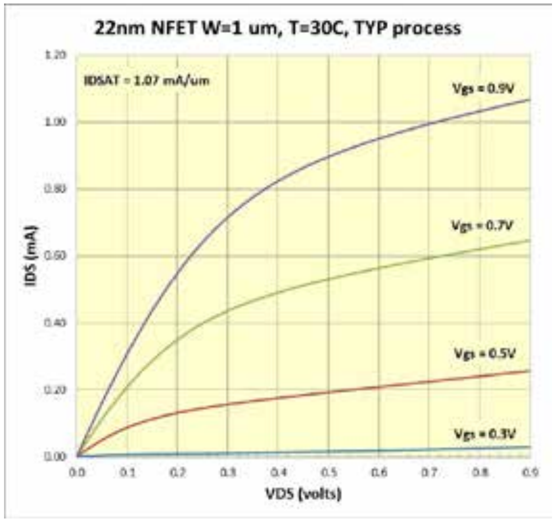


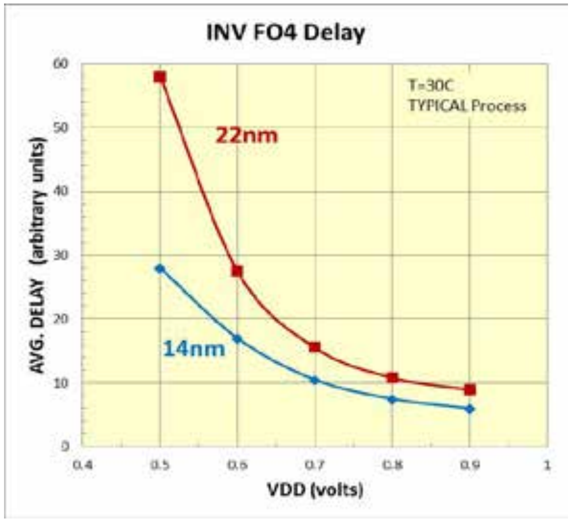
FO4 DELAY ELEMENT

SAMPLE OUTPUTS for HMK#1


22nm NFET VT Calculation

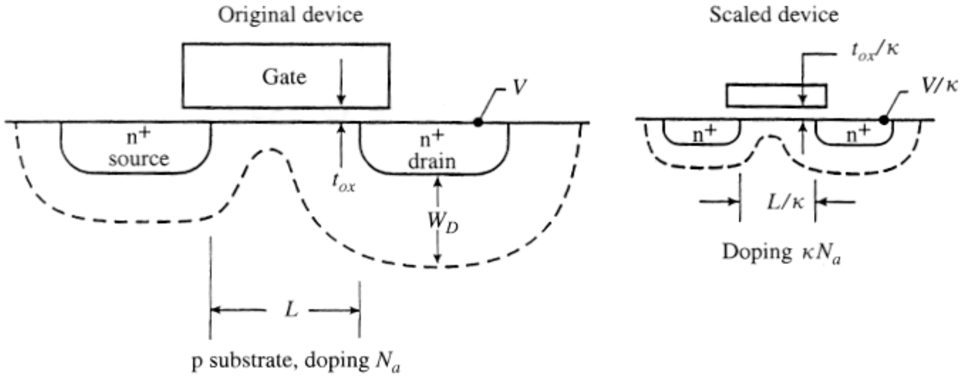

22nm NFET Subthreshold Characteristics


22nm NFET W=1 um, T=30C, TYP process


INV FO4 Delay

Backup

Constant Electric Field Scaling



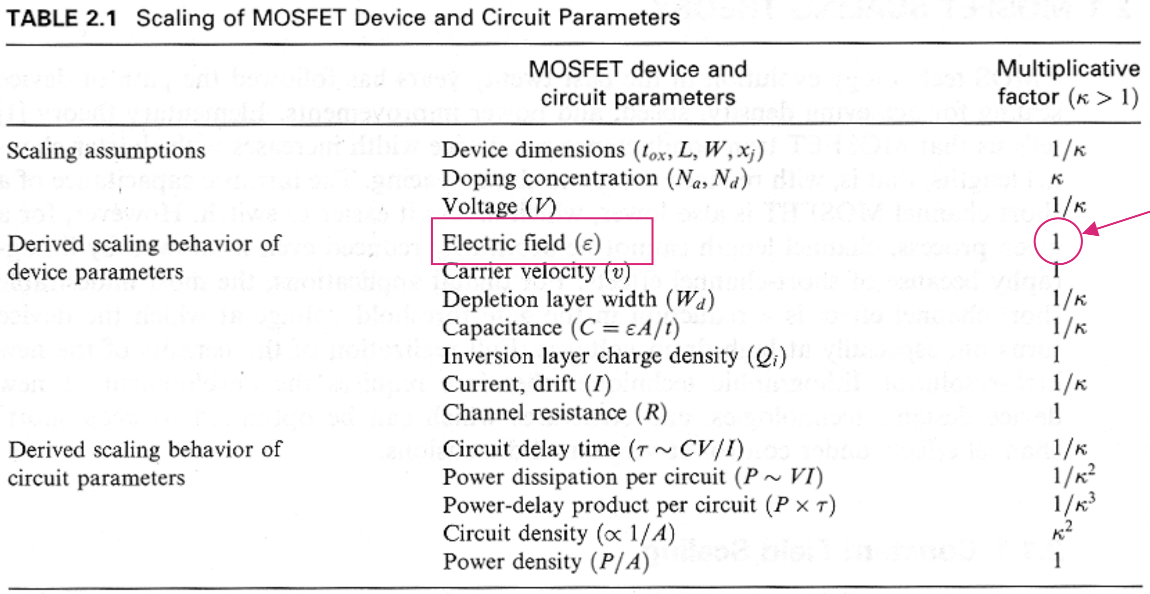


TABLE 2.1 Scaling of MOSFET Device and Circuit Parameters

	MOSFET device and circuit parameters	Multiplicative factor ($\kappa > 1$)
Scaling assumptions	Device dimensions (t_{ox}, L, W, x_j)	$1/\kappa$
	Doping concentration (N_a, N_d)	κ
	Voltage (V)	$1/\kappa$
Derived scaling behavior of device parameters	Electric field (ε)	1
	Carrier velocity (v)	1
	Depletion layer width (W_d)	$1/\kappa$
	Capacitance ($C = \varepsilon A/t$)	$1/\kappa$
	Inversion layer charge density (Q_i)	1
	Current, drift (I)	$1/\kappa$
	Channel resistance (R)	1
Derived scaling behavior of circuit parameters	Circuit delay time ($\tau \sim CV/I$)	$1/\kappa$
	Power dissipation per circuit ($P \sim VI$)	$1/\kappa^2$
	Power-delay product per circuit ($P \times \tau$)	$1/\kappa^3$
	Circuit density ($\propto 1/A$)	κ^2
	Power density (P/A)	1

constant

Design of High-Performance Microprocessor Circuits, IEEE Press, New York, 2001

Some numbers .. Constant Electric Field Scaling

$$Width = W = 0.7,\ Length = L = 0.7,\ t_{ox} = 0.7$$

❶ **Lateral and vertical dimensions reduce 30%**

$$Area\ Cap = C_a = \frac{0.7 \times 0.7}{0.7} = 0.7,$$

~(e*W*L)/Tox

$$Fringing\ Cap = C_f = 0.7,$$

~W

$$Total\ Cap \Rightarrow C = 0.7$$

❷ **Capacitance--area and fringing--reduce 30%**

$$Die\ Area = X \times Y = 0.7 \times 0.7 = 0.7^2$$

❸ **Die area reduces 50%**

Constant Electric Field Scaling –cnt'd-

$$\frac{Cap}{Transistor} = \frac{0.7}{1} = 0.7$$

❹ **Capacitance per transistor reduces 30%**

$$\frac{Cap}{Area} = \frac{0.7}{0.7 \times 0.7} = \frac{1}{0.7}$$

❺ **Capacitance per unit area increases 43%**

$$Vdd = 0.7,\ V_t = 0.7,\ I = \frac{W}{t_{ox}}(Vdd - V_t) = \frac{0.7 \times 0.7}{0.7} = 0.7$$

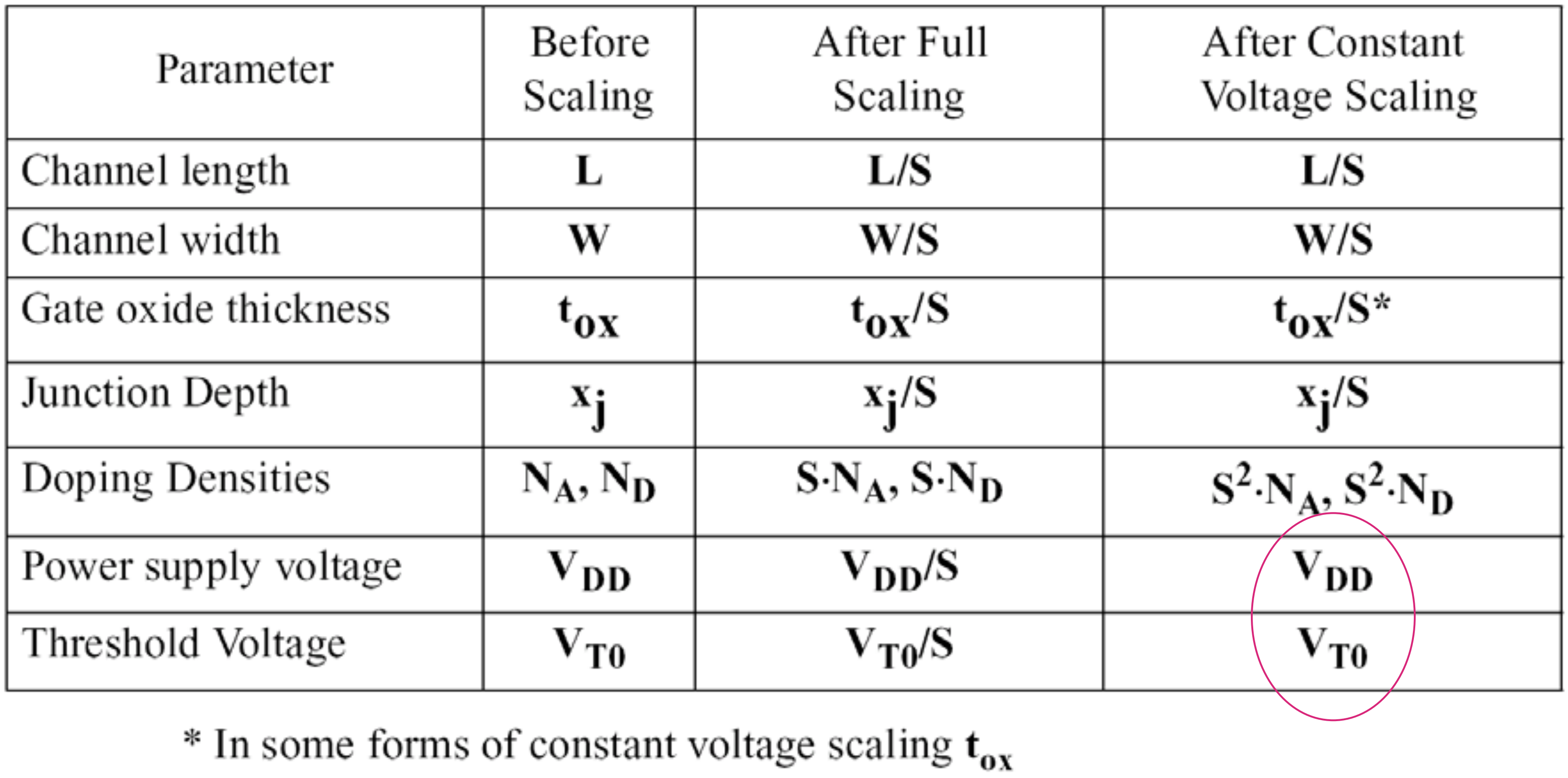
velocity-saturated device

$$T = \frac{C \times Vdd}{I} = \frac{0.7 \times 0.7}{0.7} = 0.7,\ Power = C \times V^2 \times f = \frac{0.7 \times 0.7^2}{0.7} = 0.7^2$$

❻ **Delay reduces 30%, power reduces 50%**

Fmax scales by 1/0.7

What About Constant Voltage Scaling?



Parameter	Before Scaling	After Full Scaling	After Constant Voltage Scaling
Channel length	L	L/S	L/S
Channel width	W	W/S	W/S
Gate oxide thickness	t_{ox}	t_{ox}/S	t_{ox}/S*
Junction Depth	x_j	x_j/S	x_j/S
Doping Densities	N_A, N_D	$S \cdot N_A$, $S \cdot N_D$	$S^2 \cdot N_A$, $S^2 \cdot N_D$
Power supply voltage	V_{DD}	V_{DD}/S	V_{DD}
Threshold Voltage	V_{T0}	V_{T0}/S	V_{T0}

* In some forms of constant voltage scaling t_{ox}

A comparison …….

Constant voltage scaling

$$C = 0.7, V = 1$$

$$I = \frac{W}{t_{ox}}(V - Vt) = \frac{0.7}{0.7} \times 1 = 1$$

$$D = \frac{CV}{I} = \frac{0.7 \times 1}{1} = 0.7$$

$$Power = CV^2F = \frac{0.7 \times 1}{0.7}$$

$$Power = 1$$

$$Power\ Density = 1/0.7^2 = 2$$

Constant electric field scaling

$$C = 0.7, E = \frac{V}{t_{ox}} = \frac{0.7}{0.7} = 1, E = \frac{V}{L} = \frac{0.7}{0.7} = 1$$

$$I = \frac{W}{t_{ox}}(V - Vt) = \frac{0.7}{0.7} \times 0.7 = 0.7$$

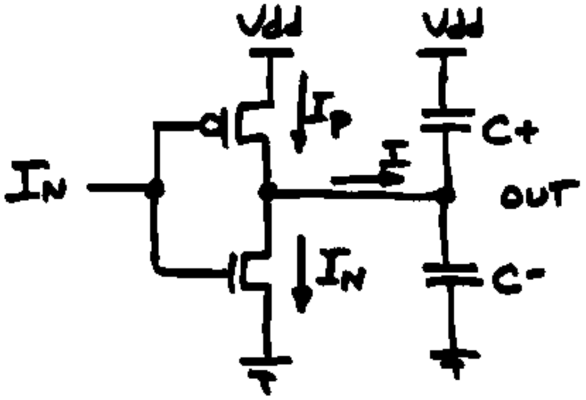
$$D = \frac{CV}{I} = \frac{0.7 \times 0.7}{0.7} = 0.7$$

$$Power = CV^2F = \frac{0.7 \times 0.7^2}{0.7}$$

$$Power = 0.5$$

$$Power\ Density = 0.5/0.7^2 = 1$$

CMOS Inverter Delay



i = dQ/dt
i = D(CV)/dt
i = CdV/dt

Using $C\frac{dV}{dt} = i$

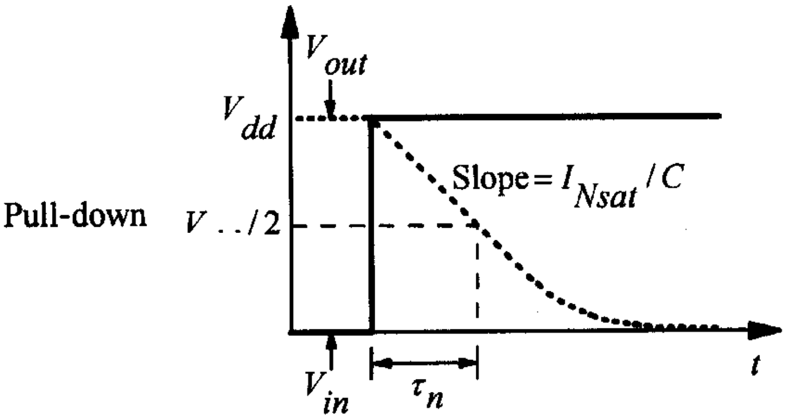
$$C_- \frac{d(V_{out} - 0)}{dt} + C^+ \frac{d(V_{out} - V_{dd})}{dt} = -I_n$$

$$(C_- + C^+)\frac{dV_{out}}{dt} = C\frac{dV_{out}}{dt} = -I_n \quad \text{or} \quad dt = \frac{C\,dV_{OUT}}{-I_n}$$

Figures from: Y. Taur, T.H. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, UK, 1998.

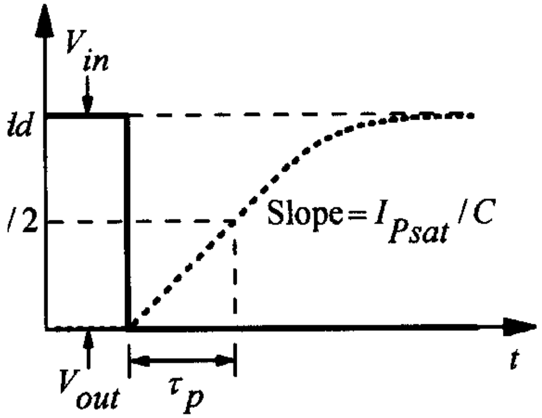
Inverter Delay –cnt'd-

Wn=NFET width, Idsatn = mA/um
Wp=PFET width, Idsatp = mA/um



$$\tau_n = \frac{CV_{dd}}{2I_{NSAT}} = \frac{CV_{dd}}{2W_n I_{DSATN}}$$

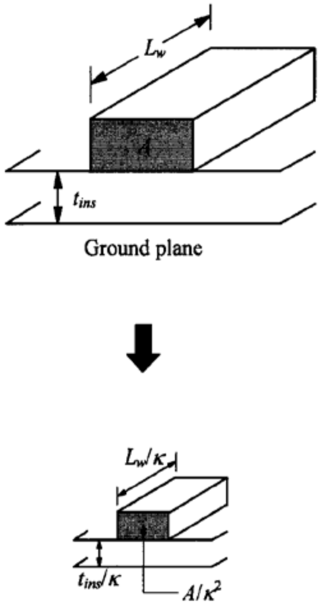
similarly, $$\tau_p = \frac{CV_{dd}}{2W_p I_{DSATP}}$$



$$\tau = (\tau_n + \tau_p)/2$$

$$= \frac{CV_{dd}}{4}\left[\frac{1}{W_n I_{dsatN}} + \frac{1}{W_p I_{dsatP}}\right]$$

Interconnect



thickness

length

width

Inter-dielectric thickness

Metal-to-metal space

	Interconnect Parameters	Scaling Factor ($\kappa \geq 1$)
Scaling assumptions	Interconnect dimensions (t_w, L_w, W_w, t_{ins}, W_{sp})	$1/\kappa$
	Resistivity of conductor (ρ_w)	1
	Insulator permittivity (ε_{ins})	1
Derived wire scaling behavior	Wire capacitance per unit length (C_w)	1
	Wire resistance per unit length (R_w)	κ^2
	Wire *RC* delay (τ_w)	1
	Wire current density ($I/W_w t_w$)	κ

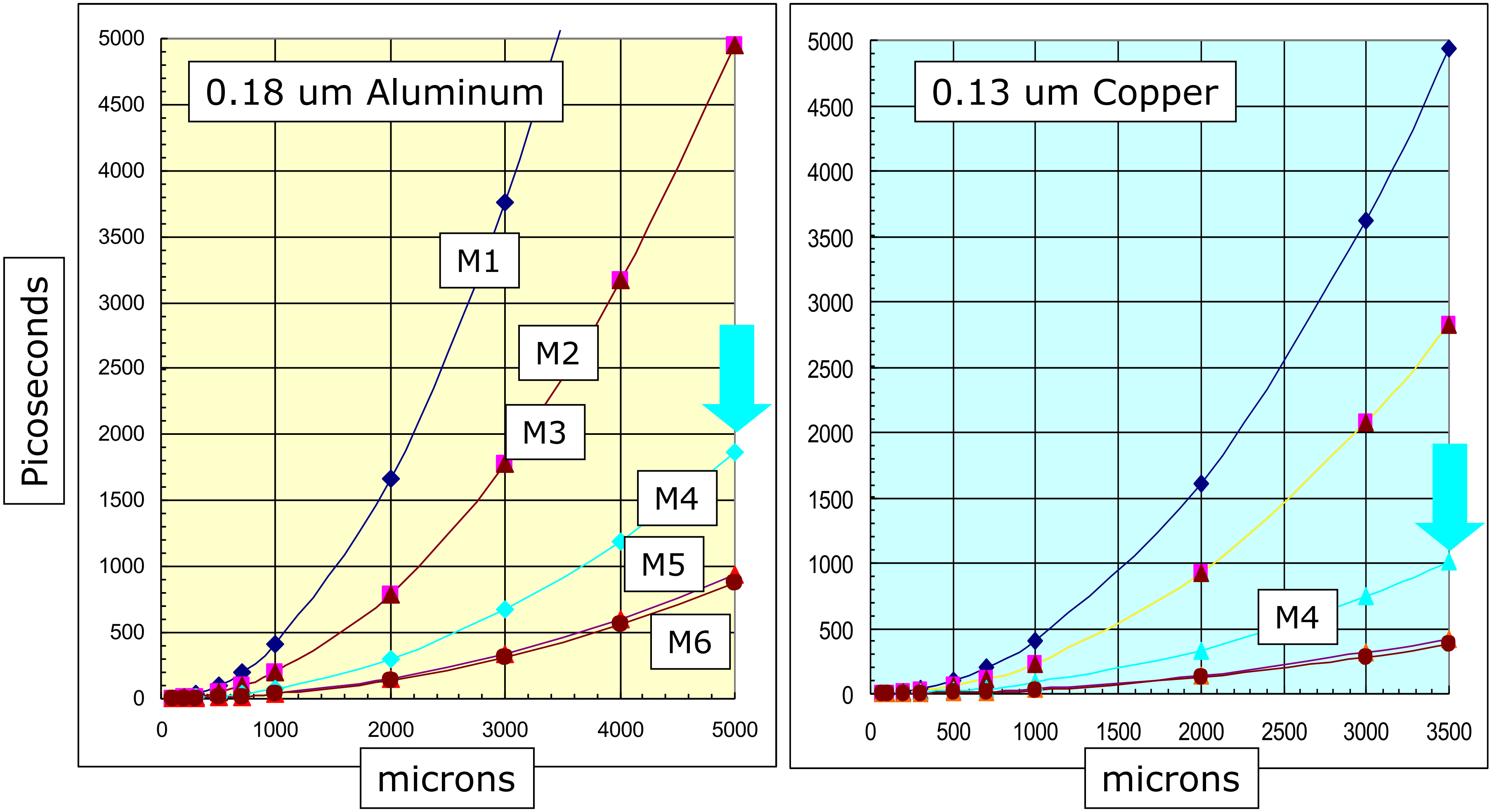
Assuming K = 1/0.7 ~ 1.43

RC delay = [(1/0.7)^2 * Rw * 0.7] * [Cw * 0.7] = RwCw

I / WwTw = [0.7 * I] / [(0.7 * Ww) * (0.7 * Tw)] = I / (0.7 * WwTw)

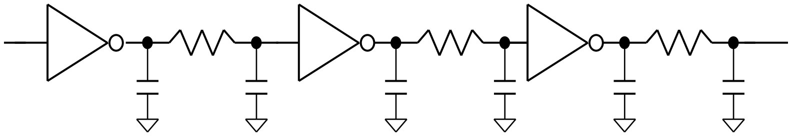
Figures from: Y. Taur, T.H. Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, UK, 1998.

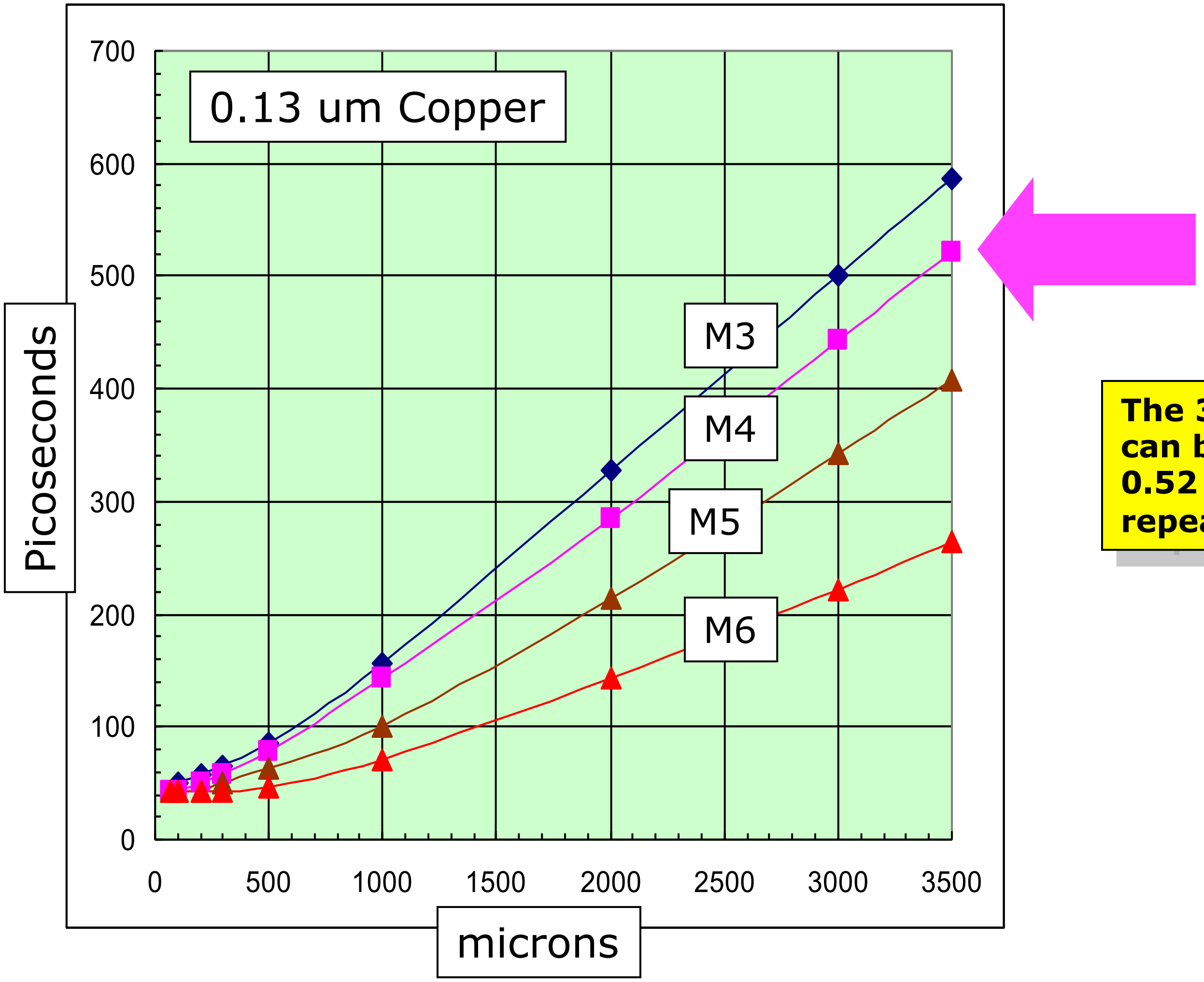
Interconnect Delay Curves



An M4 5mm 0.18um line (1.8ns un-repeated) would scale to 3.5mm in 0.13um; assuming fF/um remains constant, but ohms/um doubles, then the same wire would take 3.6ns. Copper takes this to 1.0ns.

Repeated Interconnect





The 3.5mm M4 line's 1ns can be further reduced to 0.52 ns by adding repeaters.