Digital Integrated Circuits

(83-313)

Lecture 5:

Technology Scaling

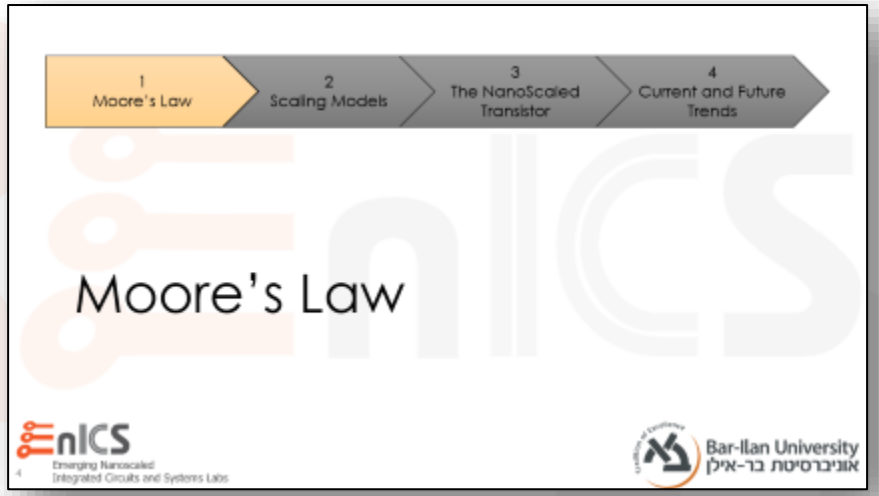
Dr. Adam Teman

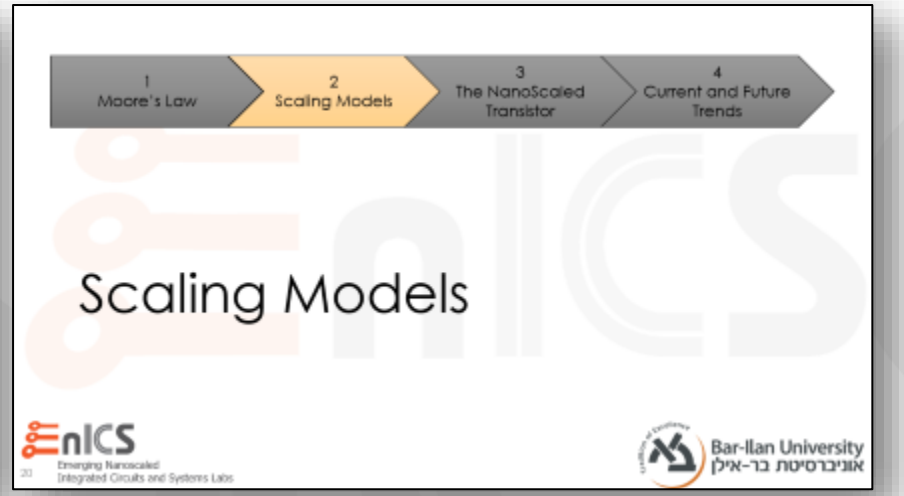
24 May 2020

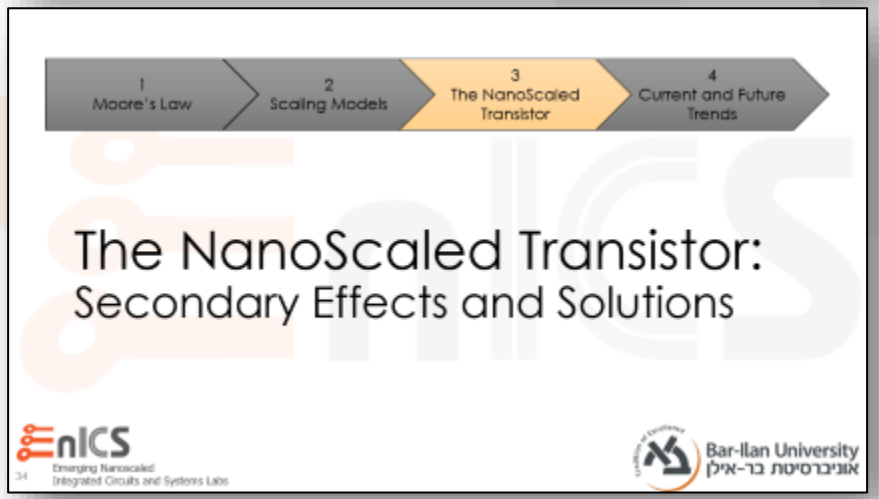
Emerging Nanoscaled Integrated Circuits and Systems Labs

Bar-Ilan University
אוניברסיטת בר-אילן

Lecture Content







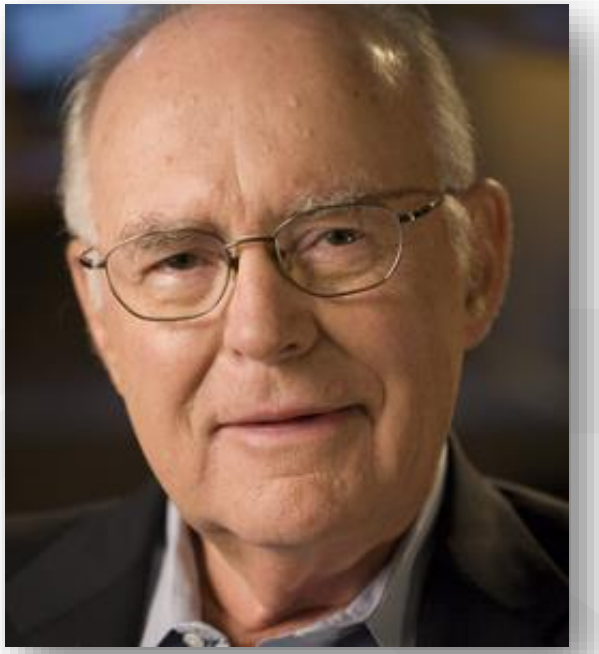
Moore's Law

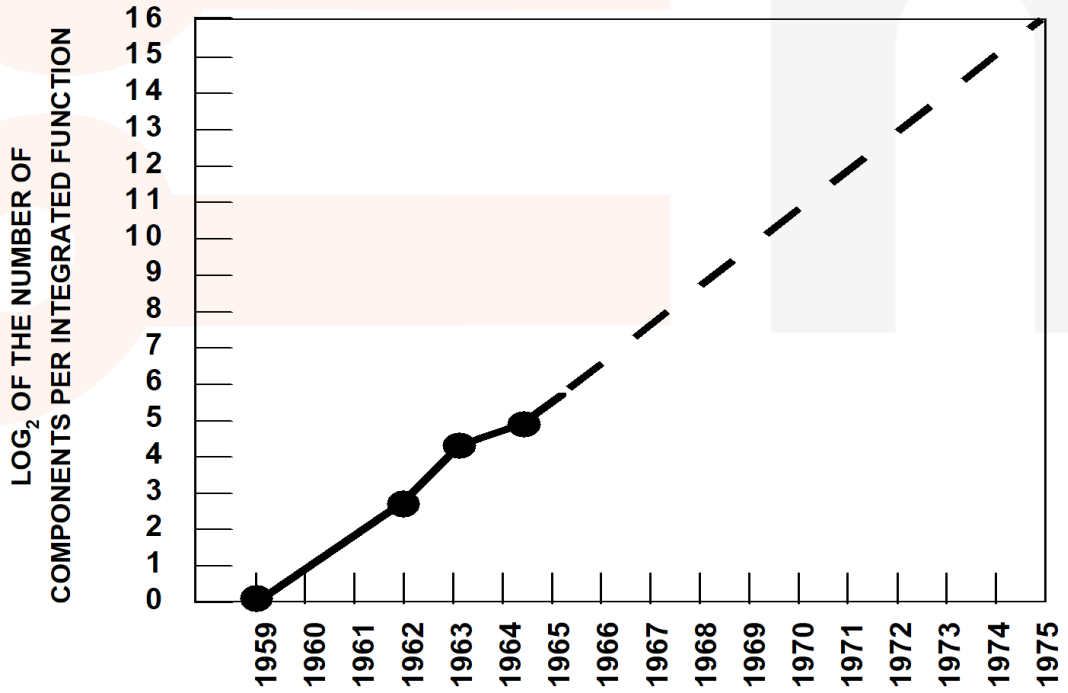
Emerging Nanoscaled Integrated Circuits and Systems Labs

Bar-Ilan University
אוניברסיטת בר-אילן

Moore's Law

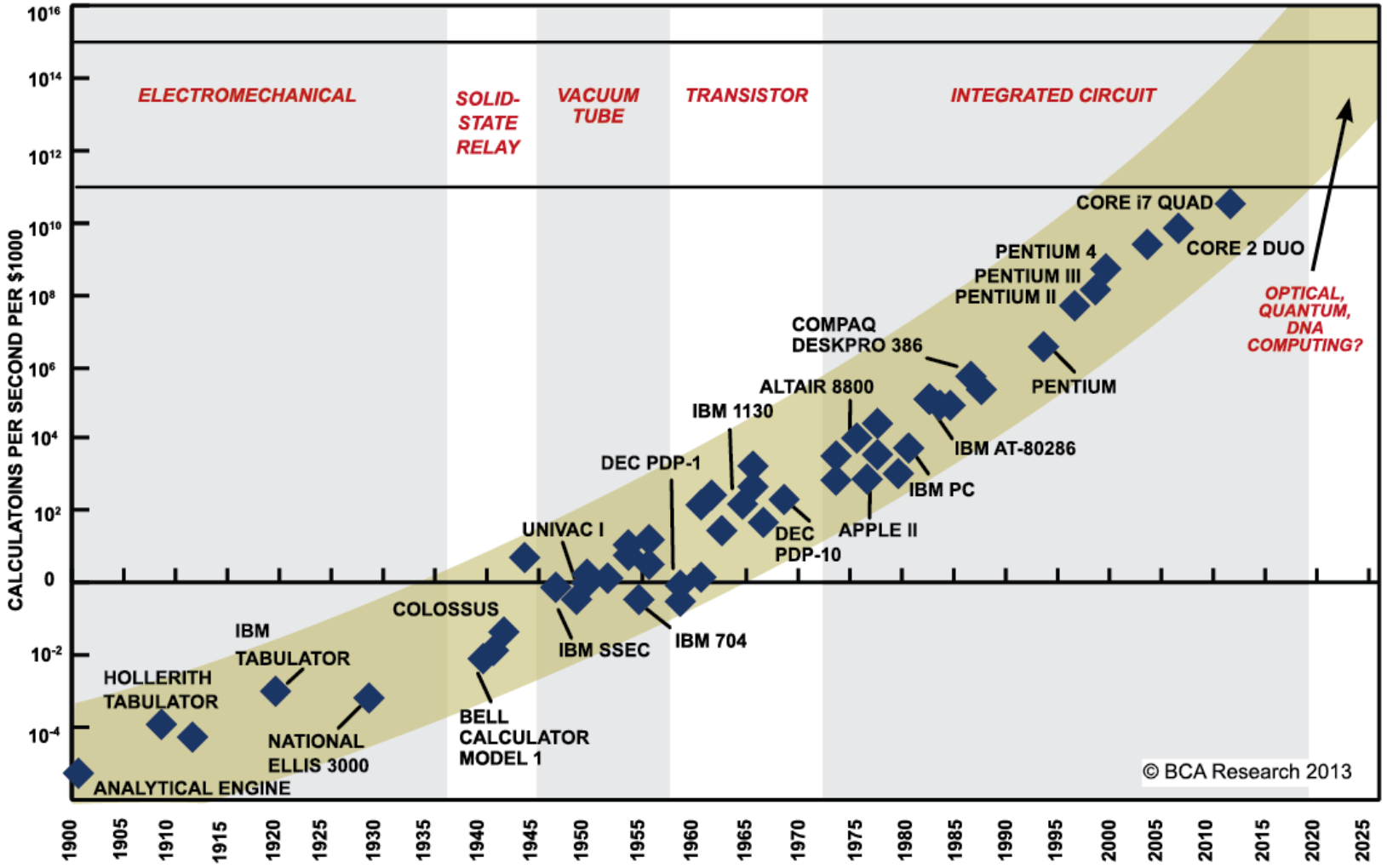
- In 1965, Gordon Moore noted that the number of components on a chip doubled every 18 to 24 months.
- He made a prediction that semiconductor technology will double its effectiveness every 18 months





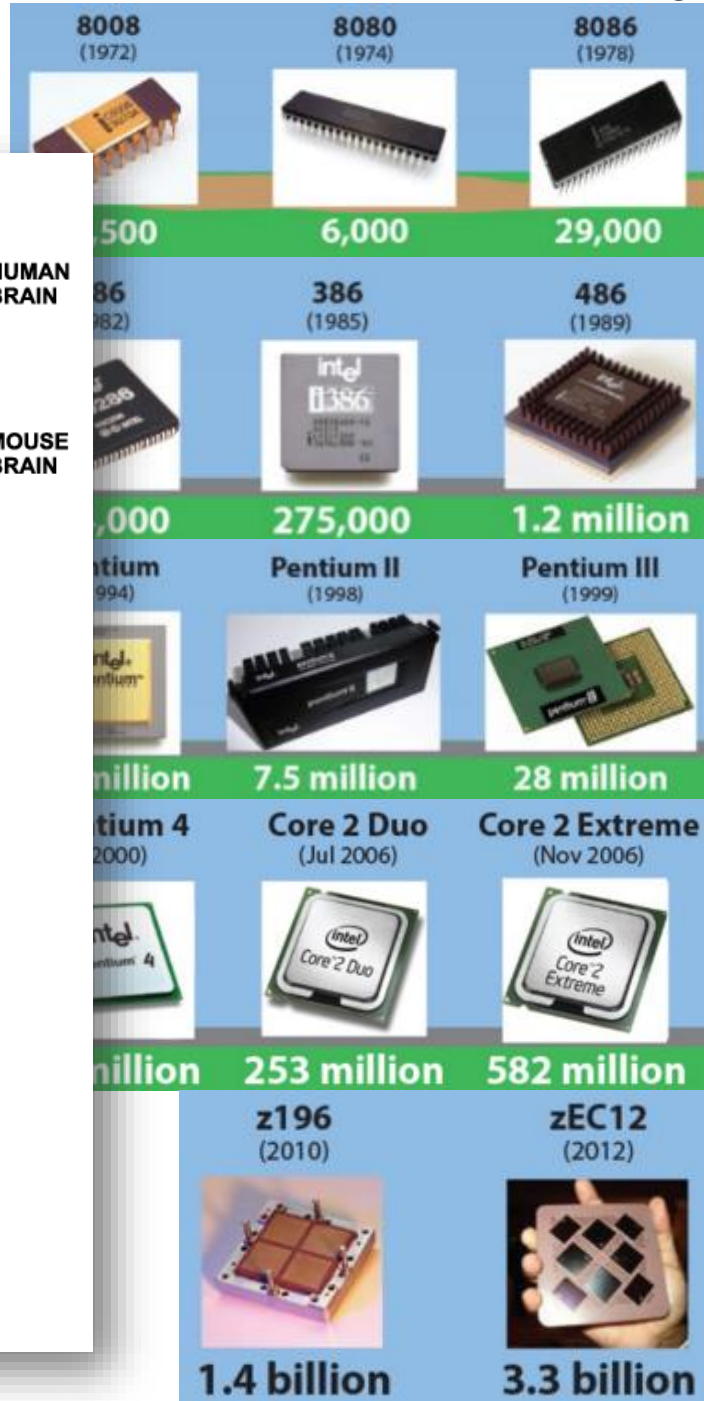
Electronics, April 19, 1965.

Moore's Law



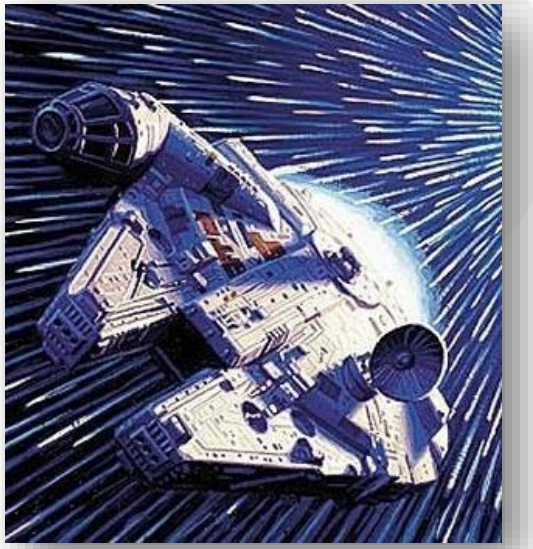
SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

Computersciencezone.org



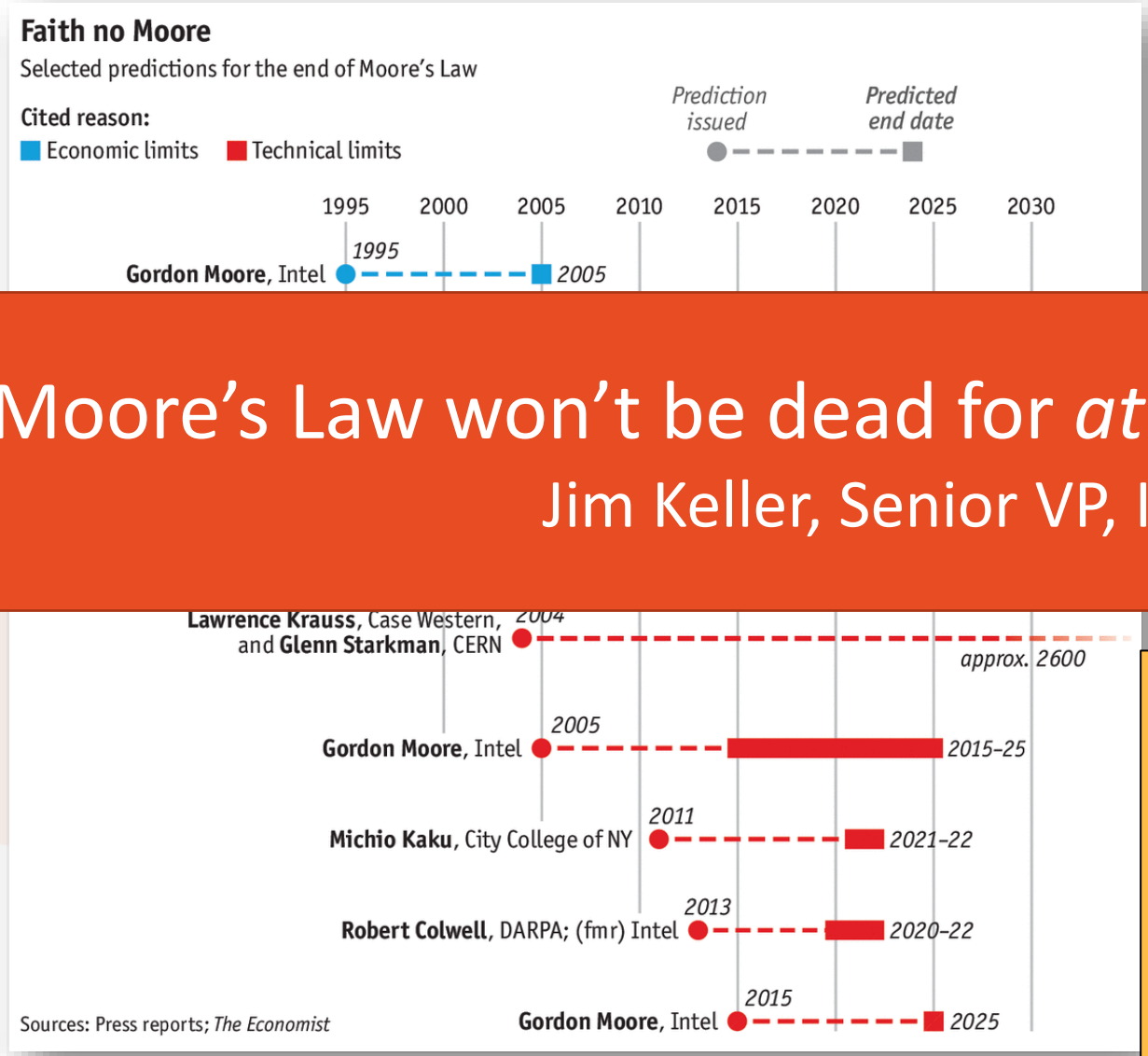
The Scale of Moore's Law

- **Imagine what would happen if other industries experienced innovation at the rate of Moore's Law, i.e., a doubling of capability every two years.**
  - **Car mileage** would be so efficient by now that a car could drive the equivalent distance between the earth and the sun on a single gallon of gas.
  - **Agriculture productivity** would be improved to a level that we could feed the planet on a square kilometer of land.
  - As for **space travel** – by now we could be zooming at 300 times the speed of light.



Stacy Smith, VP Intel, March 2017

Reports of my death were greatly exaggerated



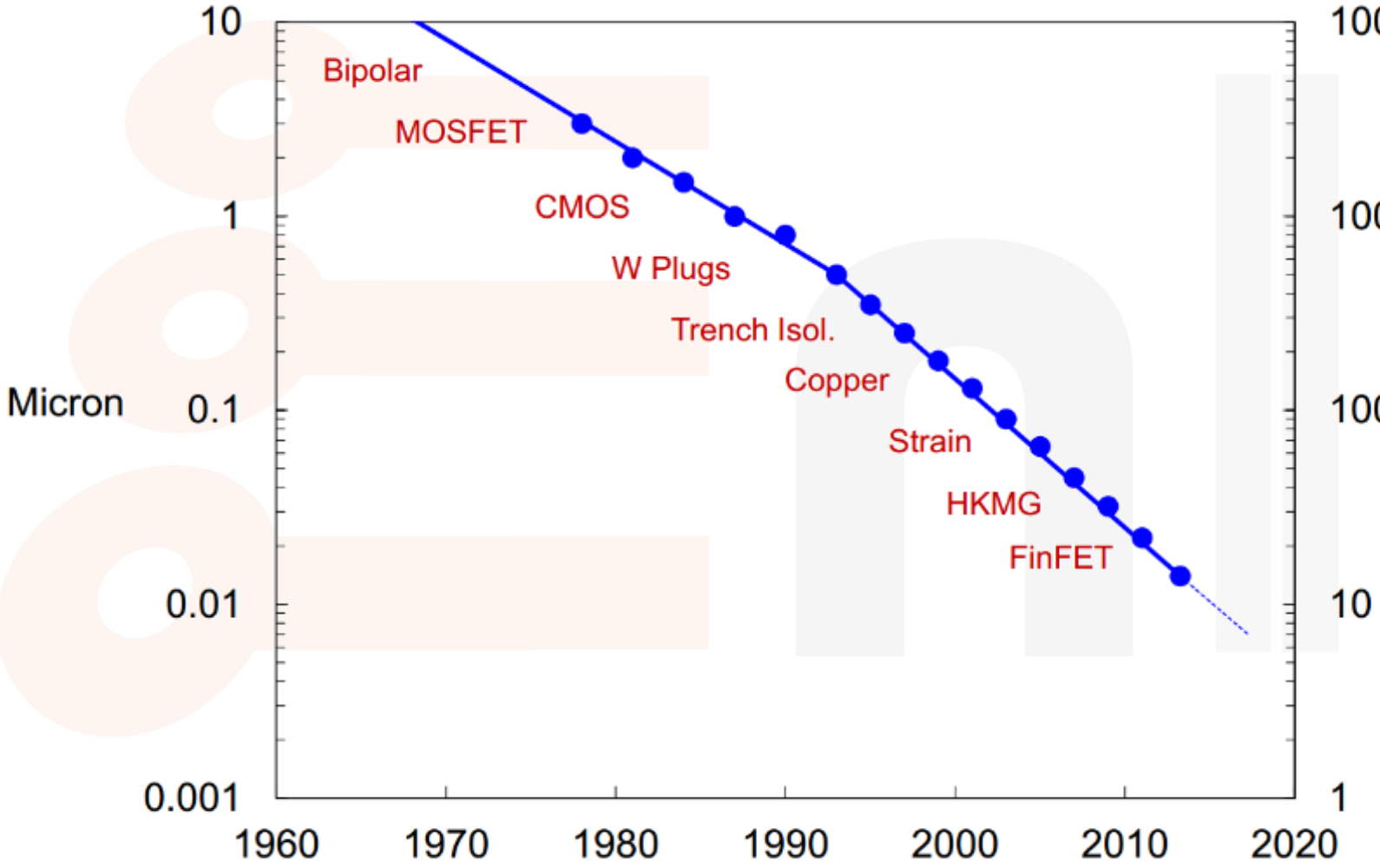
"Moore's Law won't be dead for *at least* another 30 years or so."

Jim Keller, Senior VP, Intel, Dec. 2019

"In my **34 years** in the semiconductor industry, I have witnessed the advertised death of Moore's Law **no less than four times**. As we progress from 14 nanometer technology to 10 nanometer and **plan for 7 nanometer and 5 nanometer** and even beyond, our plans are proof that **Moore's Law is alive and well**"

Bryan Krzanich, CEO Intel, April 2016

Technology supporting Moore's Law



"Yes, someday we may reach a physical limit. But **we don't see that point on our horizon**. I remember in 1990, when the features on the wafer were **the same size as the wavelength of the light** we used to print them: 193 nm. Physics was very clear. **We couldn't go any further**. "

Stacy Smith, VP Intel, March 2017

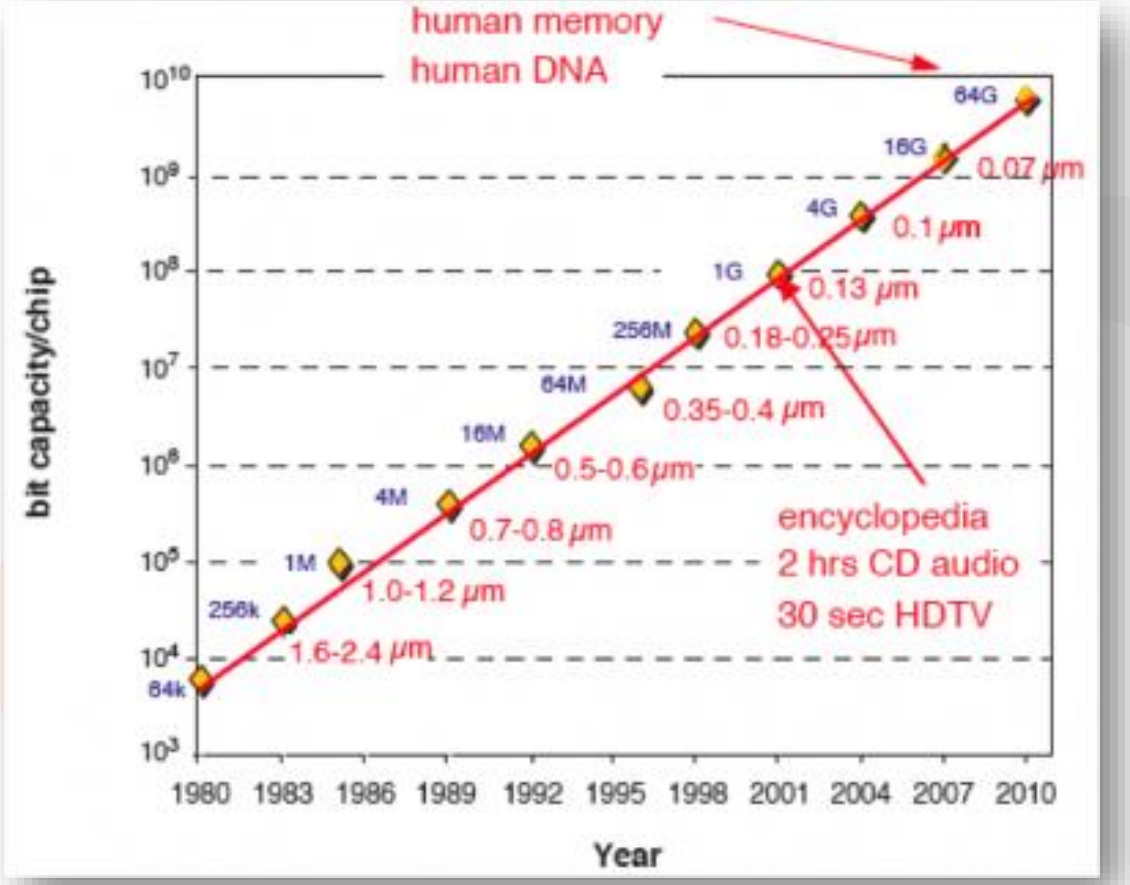
Moore's Law Today (2020)

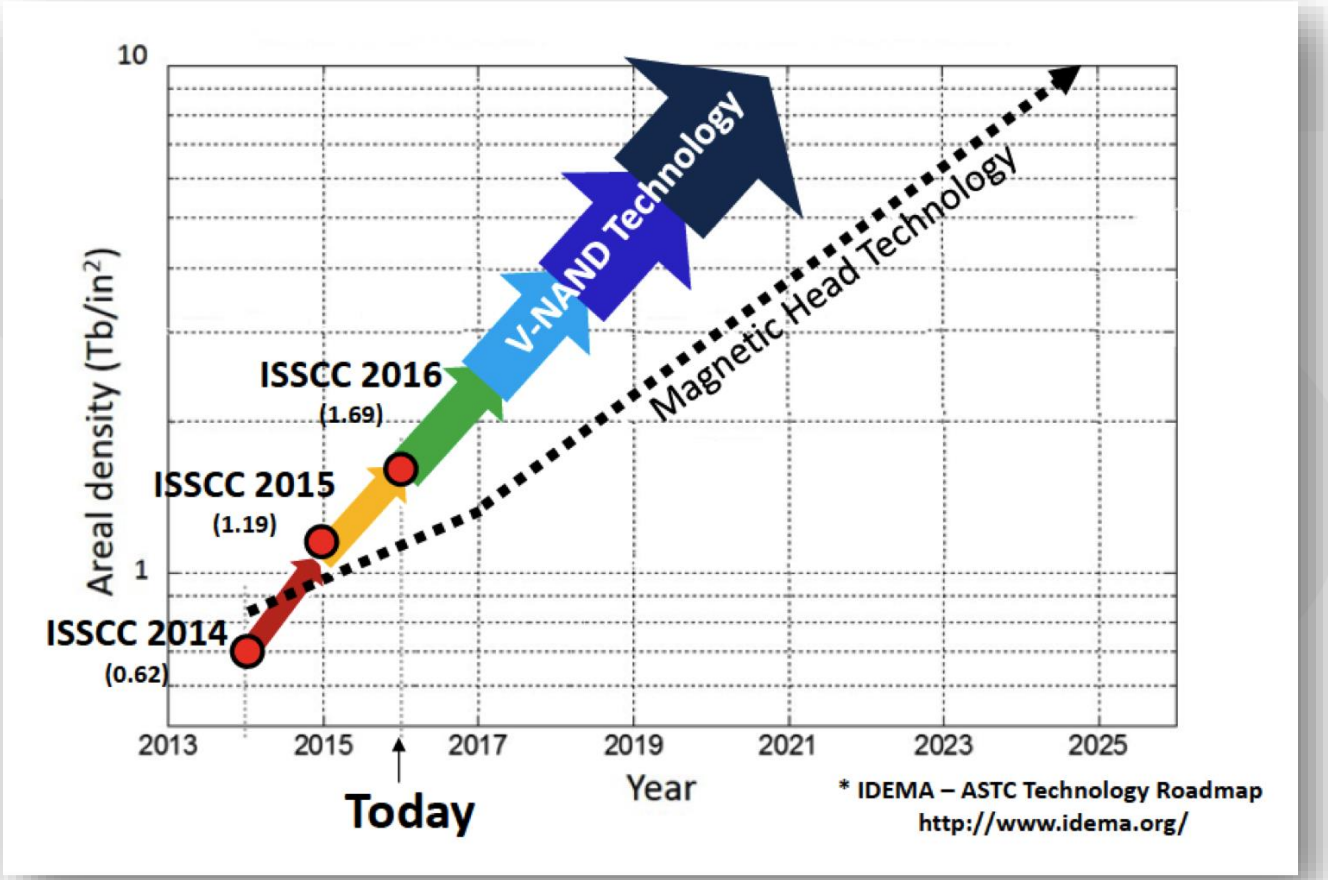
Intel Core i9-10980HK

- 14nm "Comet Lake-H"
- 8 Cores / 16 Threads
- 2.4-5.3 GHz
- 12MB Cache
- Many Billion Transistors
- Introduced April, 2020

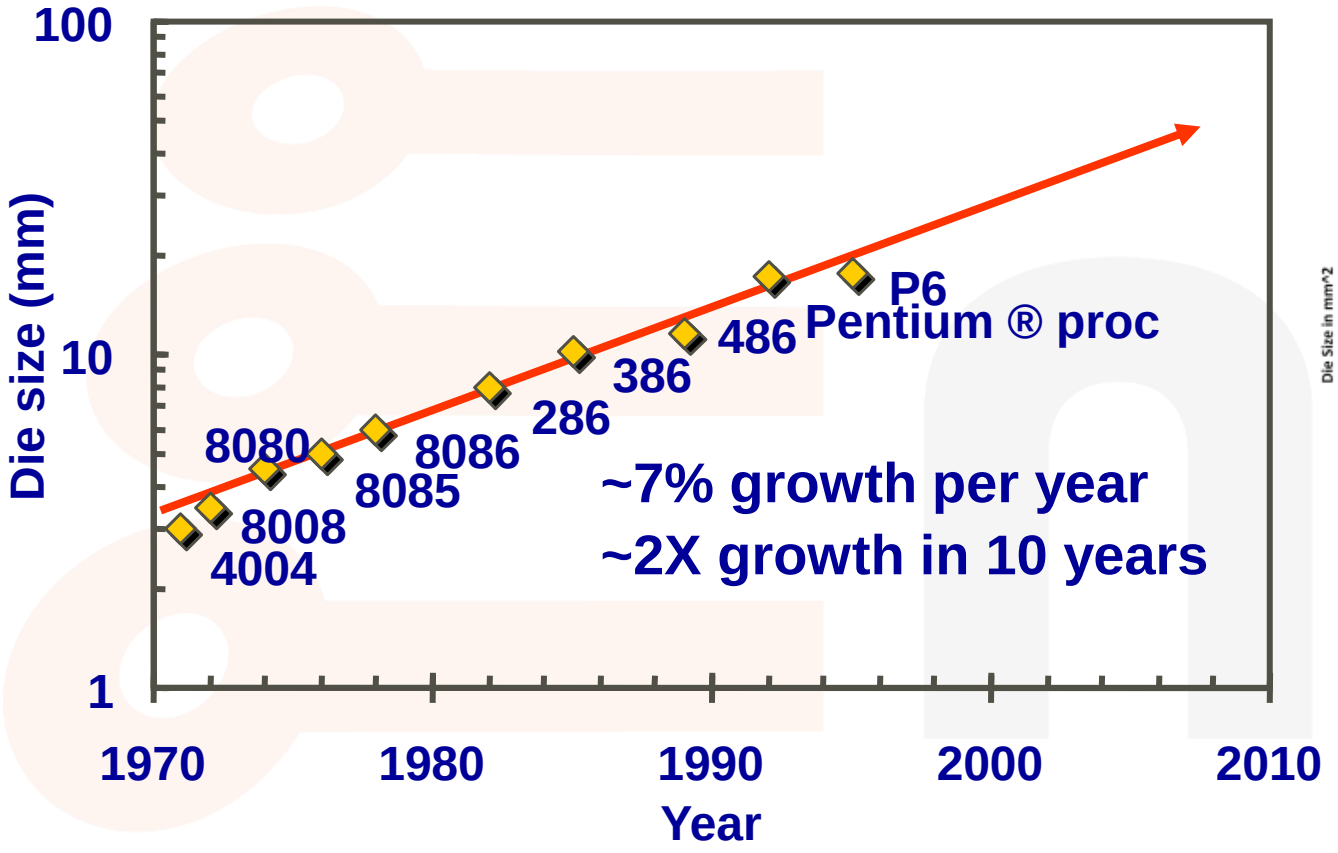
Intel Roadmap – Sep. 2019

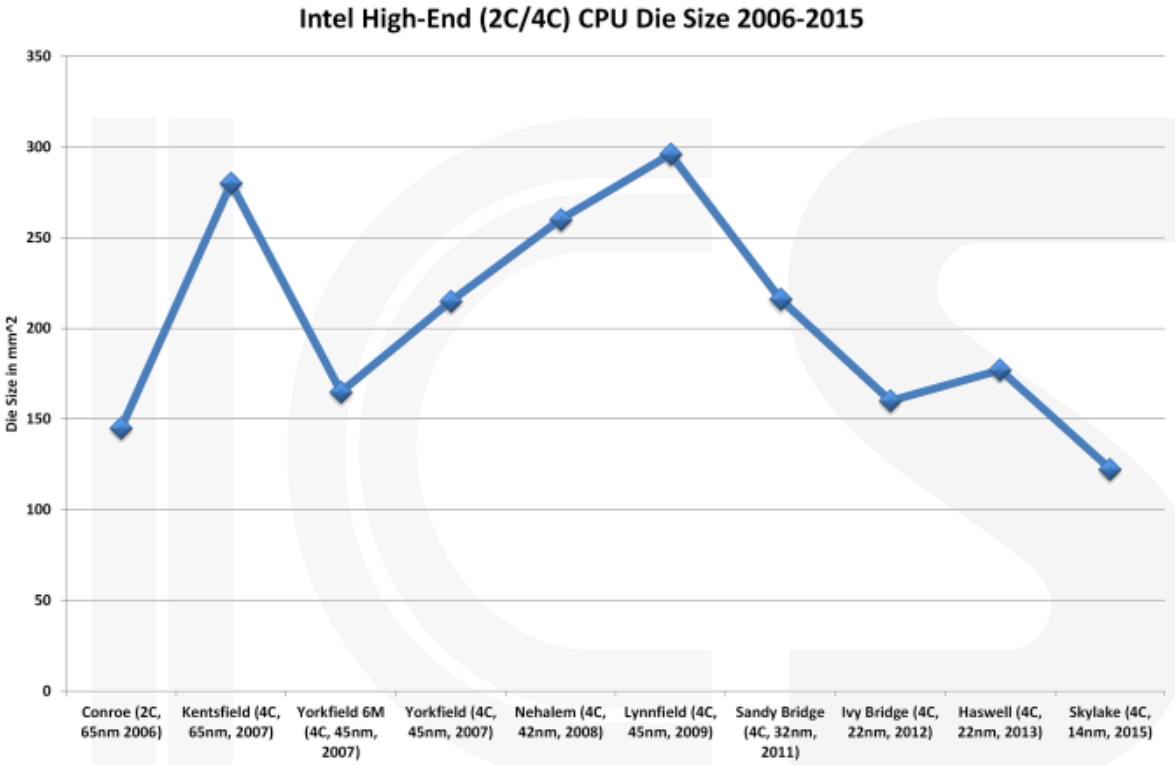
Evolution in Memory Complexity





Die Size Growth



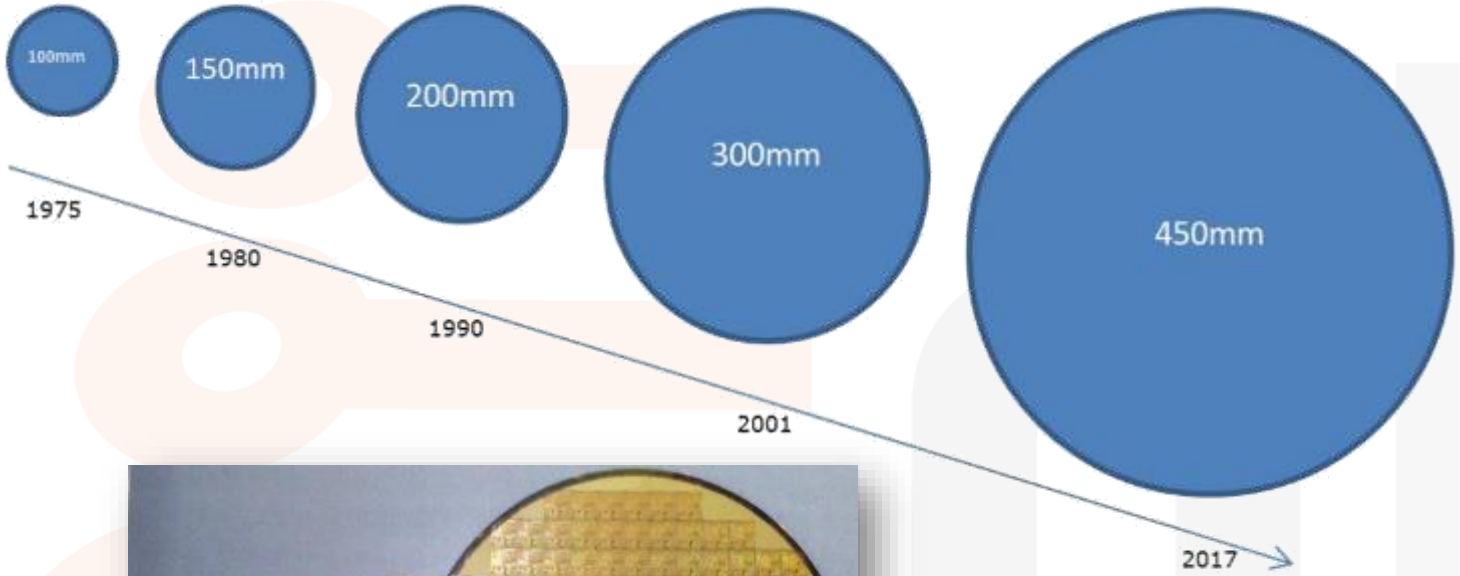


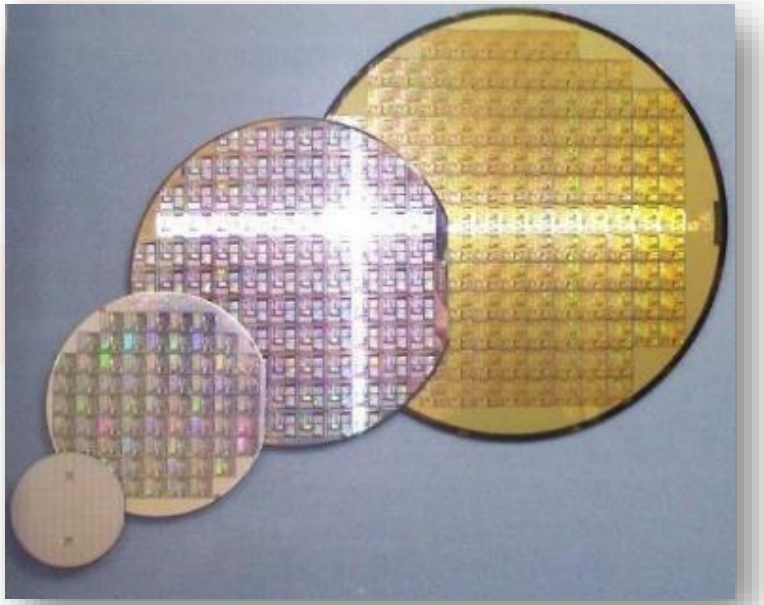
Apparently, that doesn't apply anymore…

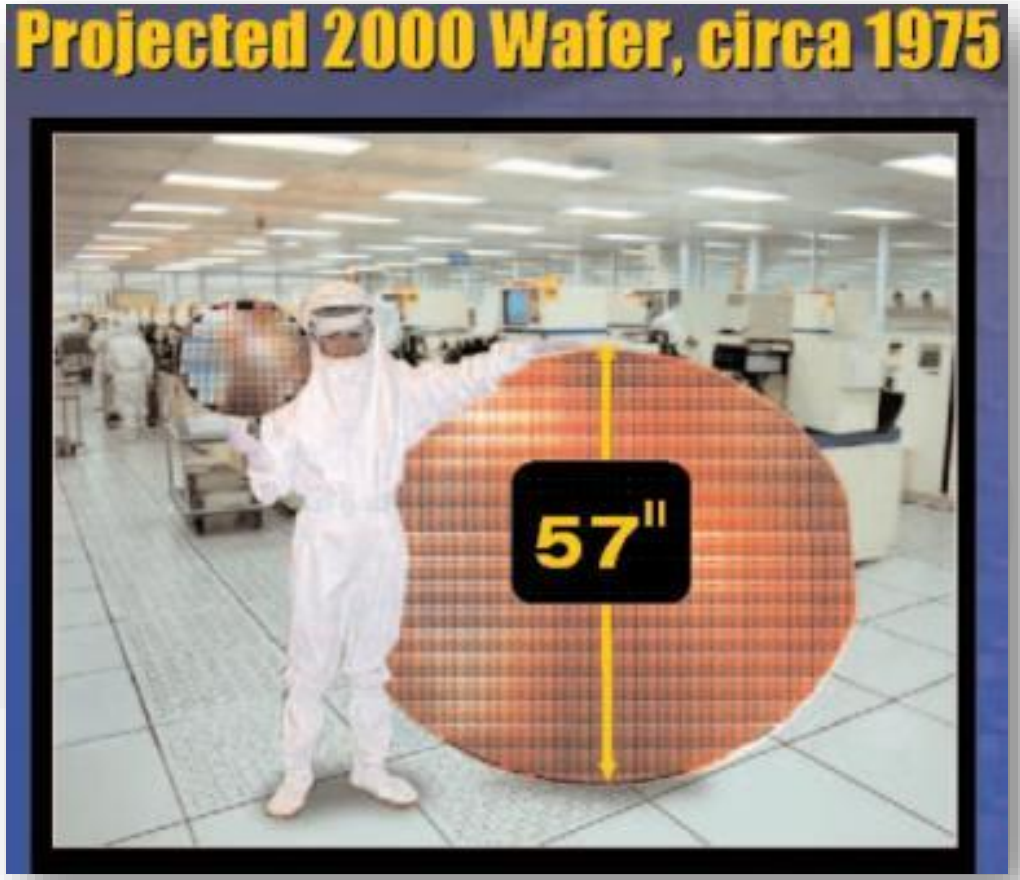
Die size grows by 14% to satisfy Moore's Law

Courtesy, Intel

Moore was not always accurate



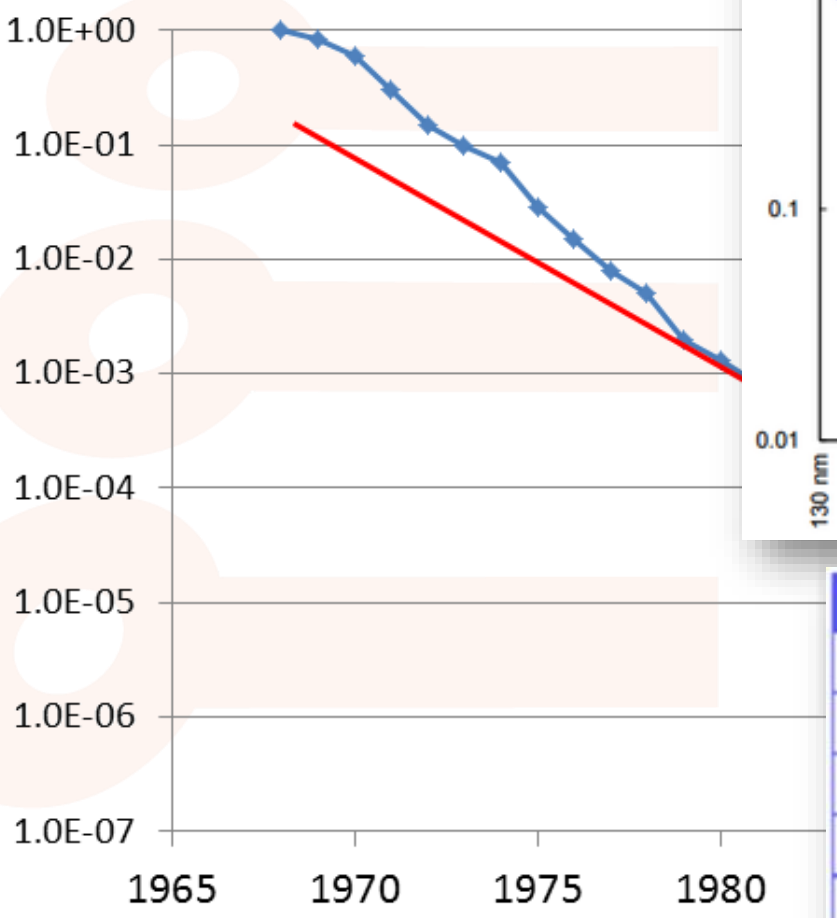




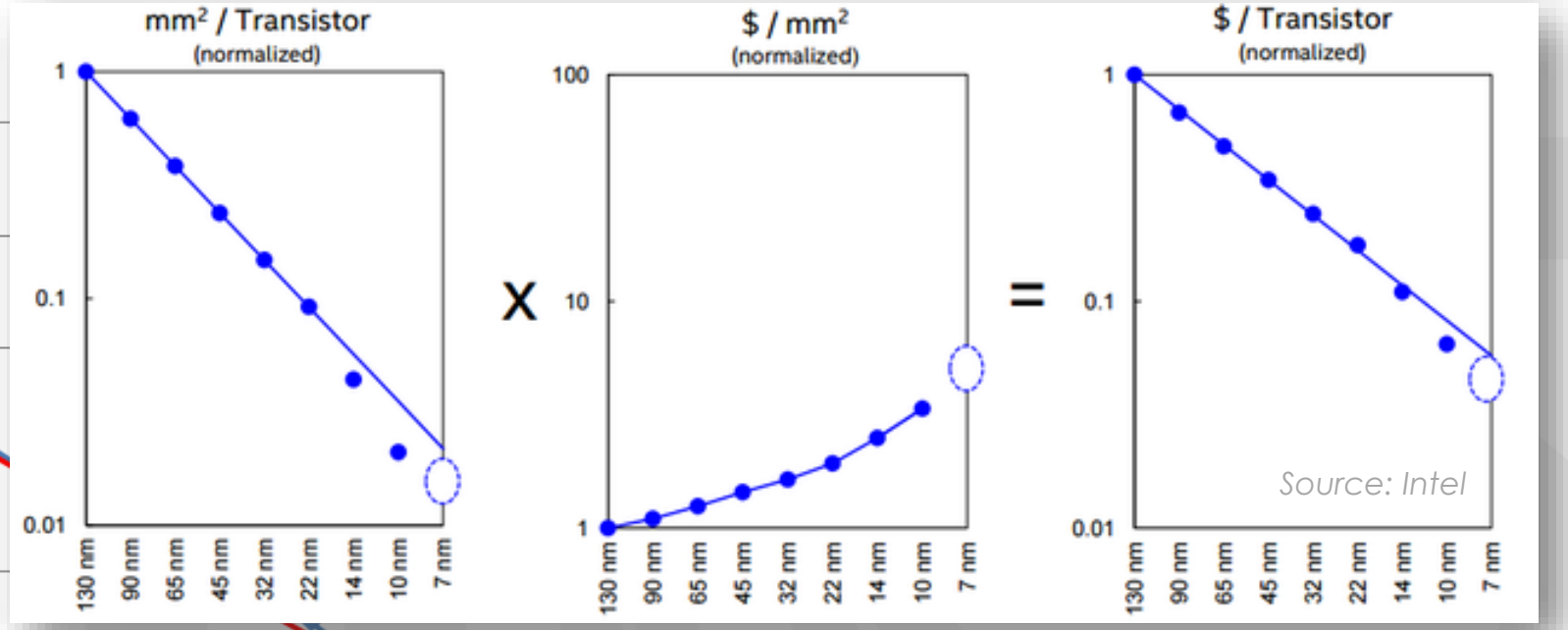
Source: Tech Design Forums

Teman's Law

Cost per Transistor

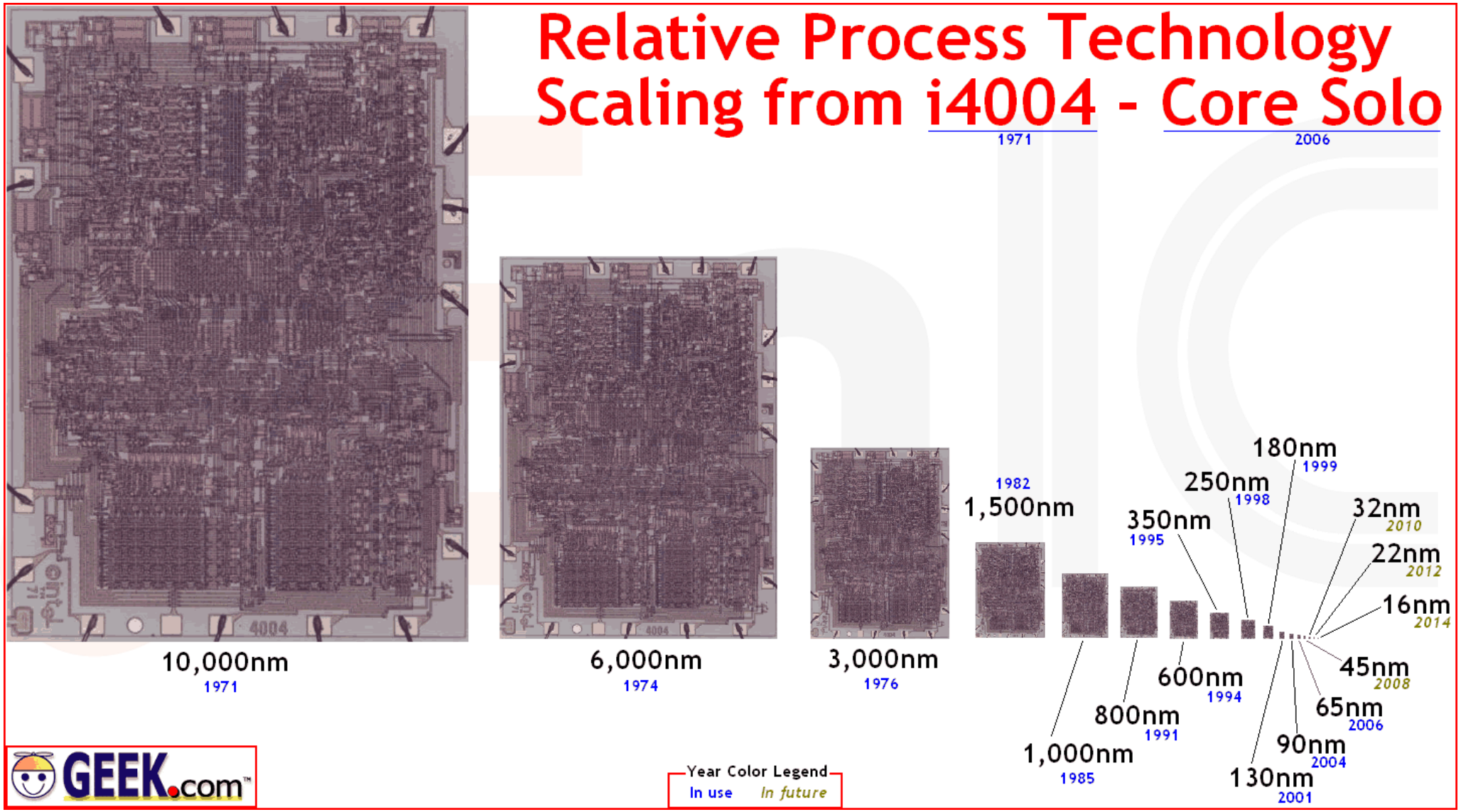


Source: neogaf.com



	16nm	10nm	7nm	5nm	3nm
Chip area (mm²)	125.00	87.66	83.27	85.00	85.00
No. of transistors (BU)	3.3	4.3	6.9	10.5	14.1
Gross die per wafer	478	686	721	707	707
Net die per wafer	359.74	512.44	545.65	530.25	509.04
Wafer price ($)	5,912.00	8,389.00	9,965.00	12,500.00	15,500.00
Die cost ($)	16.43	16.37	18.26	23.57	30.45
Transistor cost per 1B transistors ($)	4.98	3.81	2.65	2.25	2.16

Scaling…



5 MB IBM 350 Disk file (1956)

1 TB USB Flash Drive (2015)

Source: Richard Straub

Goals of Technology Scaling

- **Make things cheaper:**
  - Want to sell more functions (transistors) per chip for the same money
  - Build same products cheaper, sell the same part for less money
  - Price of a transistor **has to be reduced**
- **But also want to be faster, smaller, lower power**

>$900,000 worth of applications in a smart phone today

	Application	$ (2011)	Original Device Name	Year*	MSRP	2011's $
1	Video conferencing	*free*	Compression Labs VC	1982	$250,000	$586,904
2	GPS	*free*	TI NAVSTAR	1982	$119,900	$279,366
3	Digital voice recorder	*free*	SONY PCM	1978	$2,500	$8,687
4	Digital watch	*free*	Seiko 35SQ Astron	1969	$1,250	$7,716
5	5 Mpixel camera	*free*	Canon RC-701	1986	$3,000	$6,201
6	Medical library	*free*	e.g. CONSULTANT	1987	Up to $2,000	$3,988
7	Video player	*free*	Toshiba V-8000	1981	$1,245	$3,103
8	Video camera	*free*	RCA CC010	1981	$1,050	$2,617
9	Music player	*free*	Sony CDP-101 CD player	1982	$900	$2,113
10	Encyclopedia	*free*	Compton's CD Encyclopedia	1989	$750	$1,370
11	Videogame console	*free*	Atari 2600	1977	$199	$744
	Total	**free**				**$902,065**

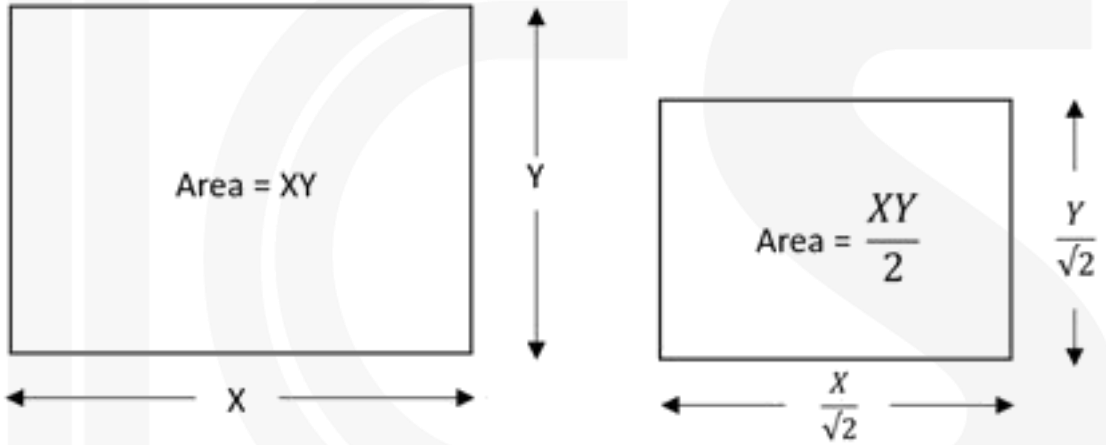
*Year of Launch

Source: Peter Diamandis

Rabaey's Law of Playstations

Technology Scaling – Dennard's Law

- **Benefits of scaling the dimensions by 30% (Dennard):**
  - Double transistor density
  - Reduce gate delay by 30%
    (increase operating frequency by 43%)
  - Reduce energy per transition by 65%
    (50% power savings
    @ 43% increase in frequency



- **Die size used to increase by 14% per generation**

- **Technology generation spans 2-3 years**

The Computer Hall of Fame

- **So yes, we all have a smartphone. But thanks to scaling, 20 years ago, the real techies all had a**

Palm Pilot

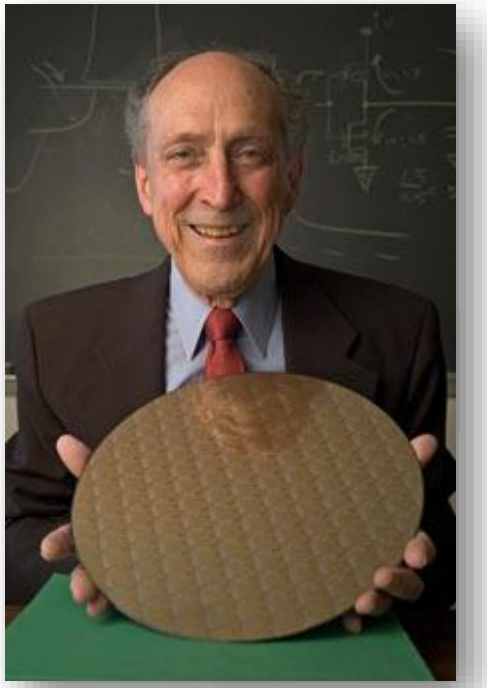
Source: computerhistory.org

- Known as a “PDA” – Personal Digital Assistant
  Introduced in 1996, sold for $299
- Ran on a 16 MHz Motorola DragonBall processor, 128K kB memory,
  160 x 160 pixel monochrome screen
- The Pilot 1000 could store 750 addresses, one year of appointments, 100 to-do items and 100 memos.

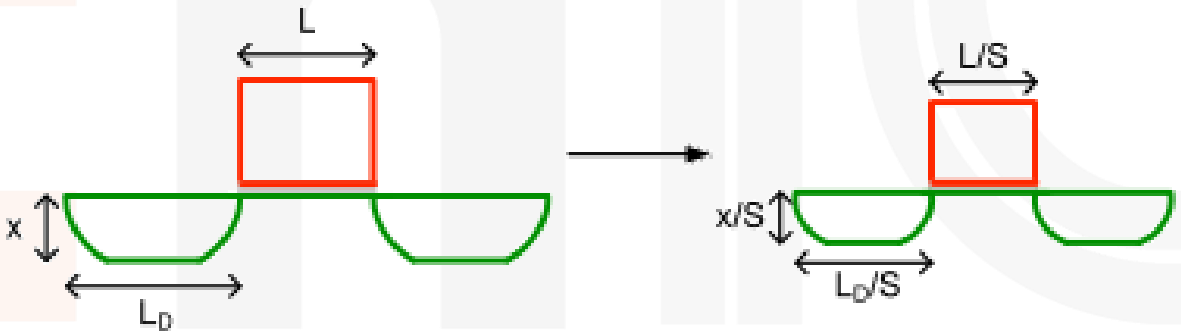
Source: palminfocenter.com

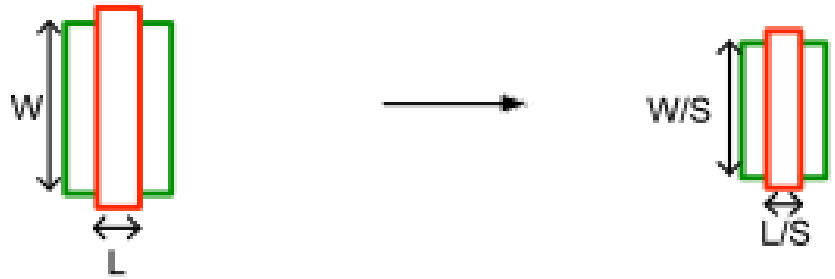
Scaling Models

Dennard (Constant Field) Scaling



- In 1974, Robert Dennard of IBM described the MOS scaling principles that accompanied us for almost forty years.
- As long as we scale all dimensions of a MOSFET by the same amount (*S*), we will arrive at better devices and lower cost:
  - $L - 1/S$
  - $W - 1/S$
  - $t_{ox} - 1/S$
  - $Na - S$
  - $V_{DD} - 1/S$
  - $V_T - 1/S$





Reminder – our simple timing/power models

- **Reminder: the *unified model* for MOS transistor conduction:**

$$I_{DS} = K\left(V_{GT}V_{DSeff} - 0.5V_{DSeff}^2\right)\left(1+\lambda V_{DS}\right)$$

$$V_{DSeff} = \min\left(V_{GT}, V_{DS}, V_{DSAT}\right)$$

$$K = \mu_n C_{ox} W/L$$

$$C_{ox} = \varepsilon_{ox}/t_{ox}$$

$$I_{on} = K_n V_{GT}^2$$

$$R_{on} = V_{DD}/I_{on}$$

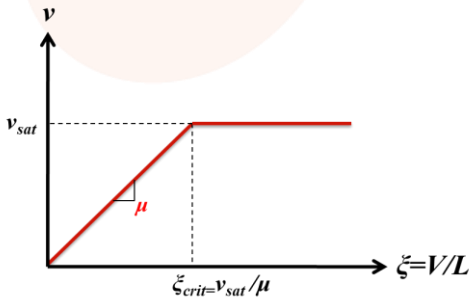
$$t_{pd} = R_{on} C_g$$

$$P_{dyn} = f \cdot C \cdot V_{DD}^2$$

Dennard (Full) Scaling for Long Transistors

$$L \propto S^{-1}$$
$$W \propto S^{-1}$$
$$t_{ox} \propto S^{-1}$$
$$V_{DD} \propto S^{-1}$$
$$V_T \propto S^{-1}$$
$$N_A \propto S$$

$$V_{DSat} = \xi_{crit} L$$



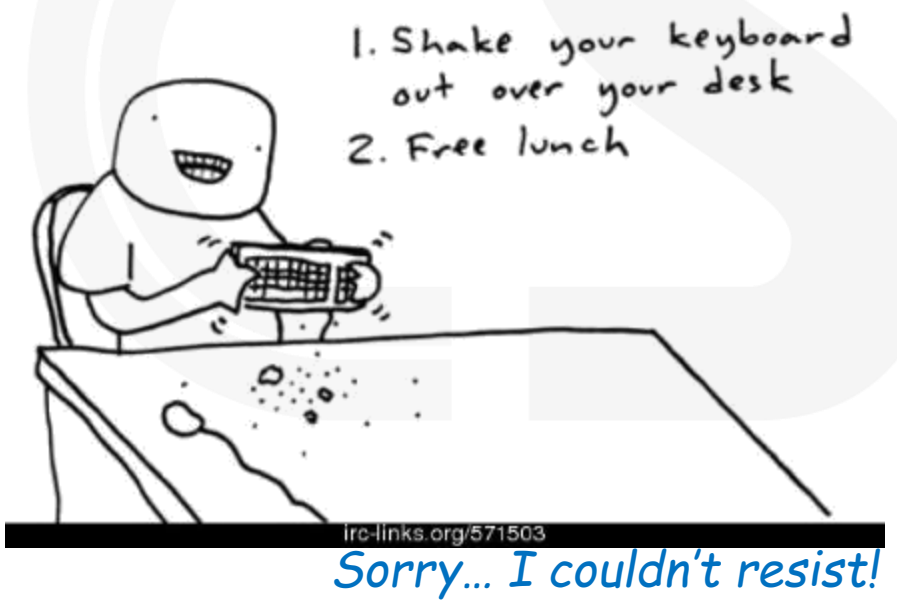
Property	Sym	Equation	Calculation	Scaling	Good?
Oxide Capacitance	C_{ox}	ε_{ox}/t_{ox}	$1/S^{-1}$	S	
Device Area	A	$W \cdot L$	$S^{-1} \cdot S^{-1}$	$1/S^2$	☺
Gate Capacitance	C_g	$C_{ox} \cdot W \cdot L$	$S \cdot S^{-1} \cdot S^{-1}$	$1/S$	☺
Transconductance	K_n	$\mu_n C_{ox} W/L$	$S \cdot S^{-1}/S^{-1}$	S	☺
Saturation Current	I_{on}	$K_n V_{DSat}(K_n W_{GT} - V_{DSat})$	$S \cdot S^{-1} \cdot S^{-1}$	$1/S$	
On Resistance	R_{on}	V_{DD}/I_{on}	S^{-1}/S^{-1}	1	
Intrinsic Delay	t_{pd}	$R_{on} C_g$	$1 \cdot S^{-1}$	$1/S$	☺
Power	P_{av}	$f \cdot C \cdot V_{DD}^2$	$S \cdot S^{-1} \cdot S^{-2}$	$1/S^2$	☺ ☺
Power Density	PD	P_{av}/A	S^{-2}/S^{-2}	1	☺

Dennard Scaling

- **This last slide showed the principal that has led to scaling for the last 50 years.**
  - Assume that we scale our process by **30%** every generation.

$$1/S = 0.7 \longrightarrow S = \sqrt{2}$$

  - Therefore, if the area scales by $1/S^2 = 1/2$, our ***die size*** goes down by **2X** every generation!
  - In addition, our ***speed*** goes up by 30%!
  - And our ***power*** also gets cut in half, without any increase in power density.



- **We have hit one of those rare win-win *free lunch* situations!**

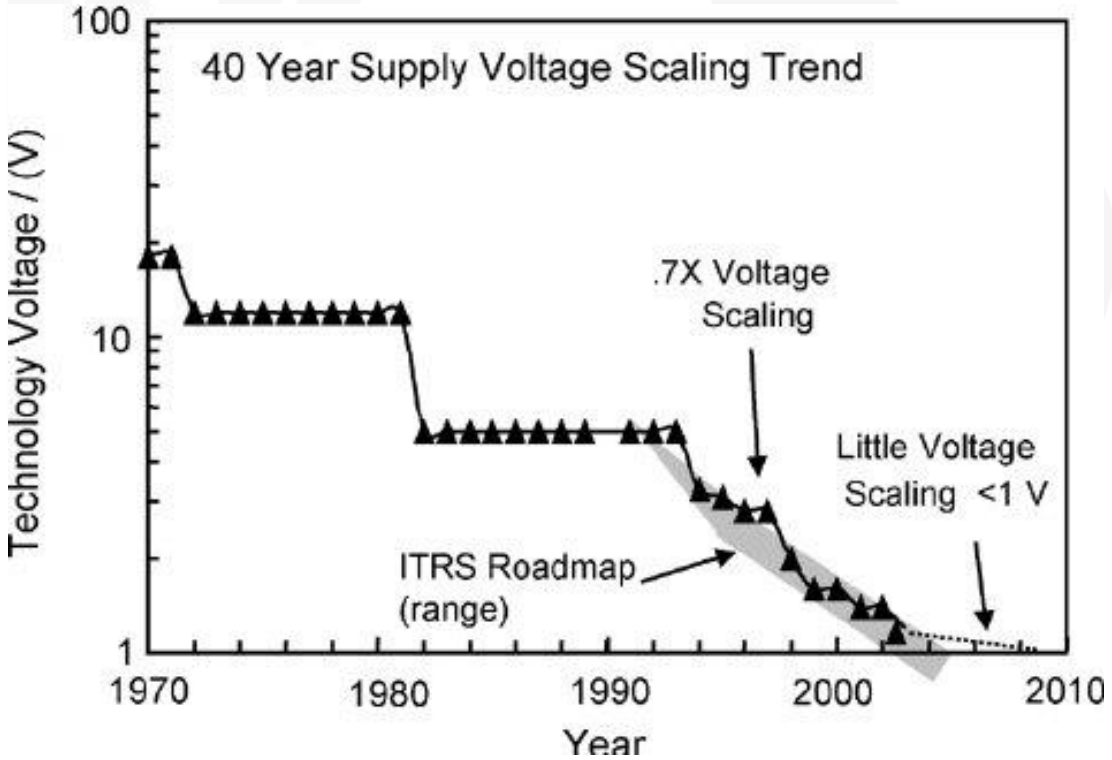
But what if we want more speed?

- We saw that $t_{pd} \propto C_g \cdot V_{DD} / I_{on}$
- We can aggressively increase the speed by keeping the voltage constant.

$$I_{on} \propto K_n V_{GT}^2 \propto S$$

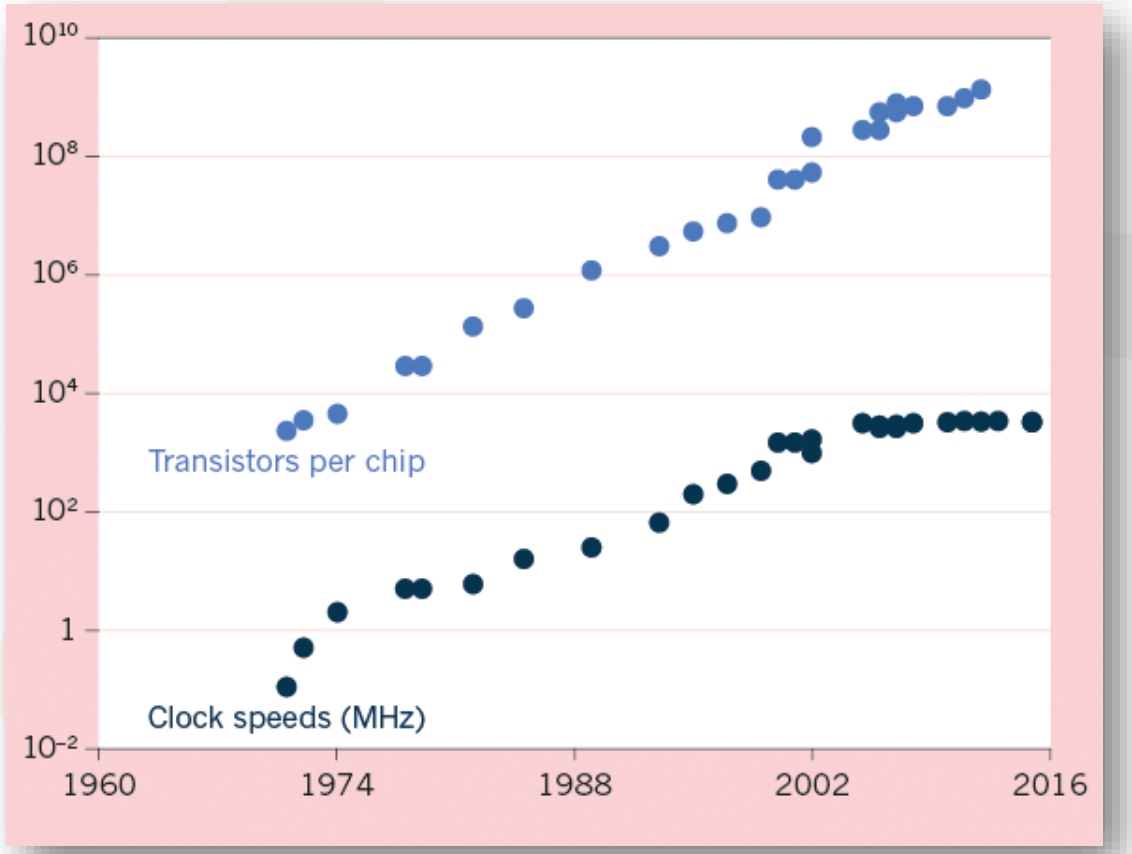
$$\Rightarrow t_{pd} \propto S^{-1} \cdot 1/S = 1/S^2$$

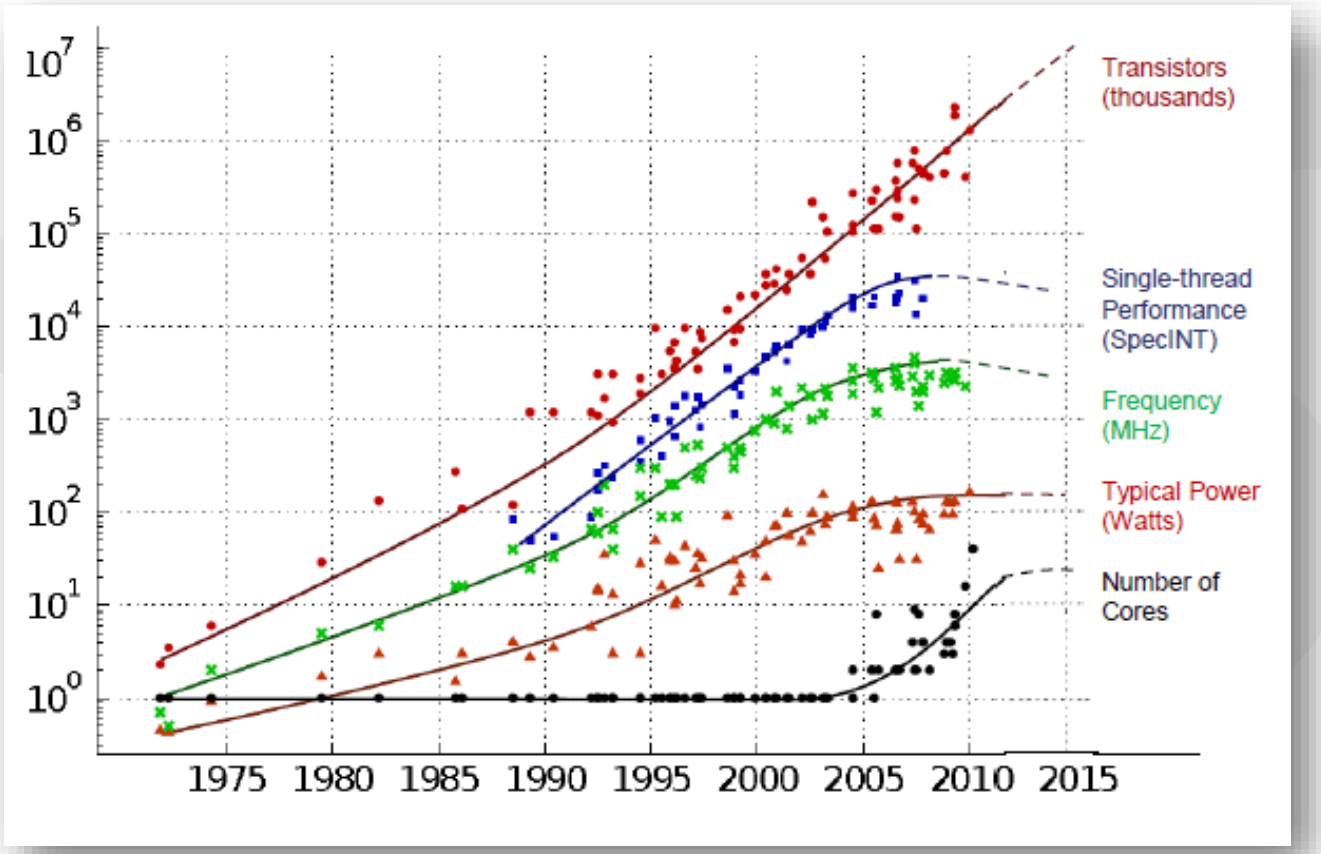
- This led to the *Fixed Voltage Scaling Model*, which was used until the 1990s (V_{DD}=5V)



Source: Thompson, et al., IEEE TSM 2005

Moore's Law in Frequency





Source: Olukotun

Fixed Voltage Scaling

$V_{DD} \propto 1$

$L \propto S^{-1}$

$W \propto S^{-1}$

$t_{ox} \propto S^{-1}$

$V_T \propto S^{-1}$

$N_A \propto S$

Property	Sym	Equation	Calculation	Scaling	Good?
Oxide Capacitance	C_{ox}	ε_{ox}/t_{ox}	$1/S^{-1}$	S	
Device Area	A	$W \cdot L$	$S^{-1} \cdot S^{-1}$	$1/S^2$	☺
Gate Capacitance	C_g	$C_{ox} \cdot W \cdot L$	$S \cdot S^{-1} \cdot S^{-1}$	$1/S$	☺
Transconductance	K_n	$\mu_n C_{ox} W/L$	$S \cdot S^{-1}/S^{-1}$	S	☺
Saturation Current	I_{on}	$K_n V_{GT}^2$	$S \cdot 1$	S	☺
On Resistance	R_{on}	V_{DD}/I_{on}	$1/S$	$1/S$	
Intrinsic Delay	t_{pd}	$R_{on} C_g$	$S^{-1} \cdot S^{-1}$	$1/S^2$	☺ ☺
Power	P_{av}	$f \cdot C \cdot V_{DD}^2$	$S^2 \cdot S^{-1} \cdot 1$	S	✖
Power Density	PD	P_{av}/A	S/S^{-2}	S^3	✖ ✖

Fixed Voltage Scaling – Short Channel

- What happens with velocity saturated devices?

$$I_{on} \propto K_n V_{DSat}\left(V_{GT} - V_{DSat}\right) \propto S \cdot S^{-1} \cdot 1 = 1$$

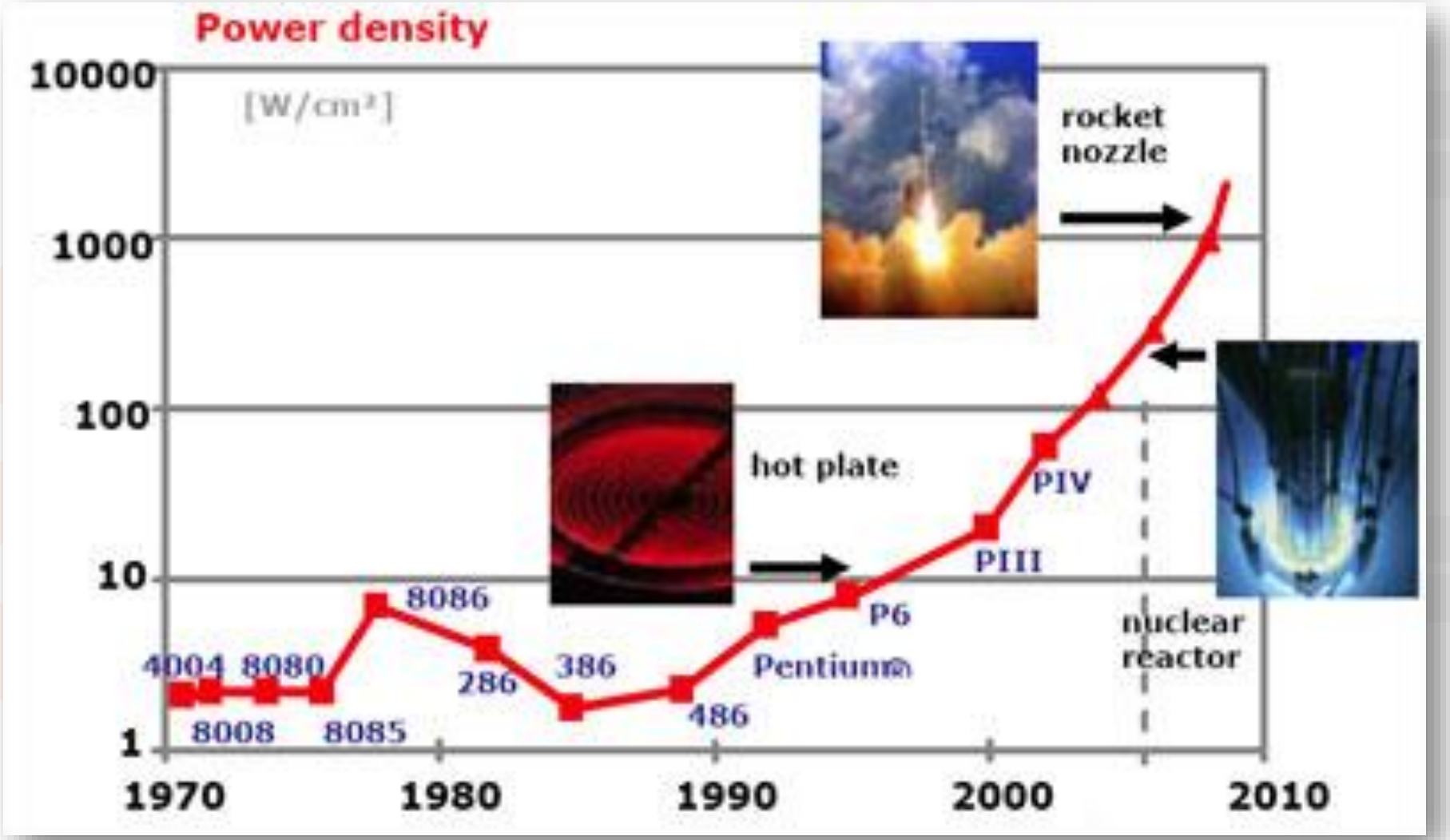
- So the on-current doesn't increase leading to less effective speed increase.

$$t_{pd} \propto R_{on} C_g \propto 1 \cdot S^{-1} = 1/S$$

- The power density still increases quadratically!

$$PD \propto fCV_{DD}^2 / A \propto S \cdot S^{-1} \cdot 1/S^{-2} = S^2$$

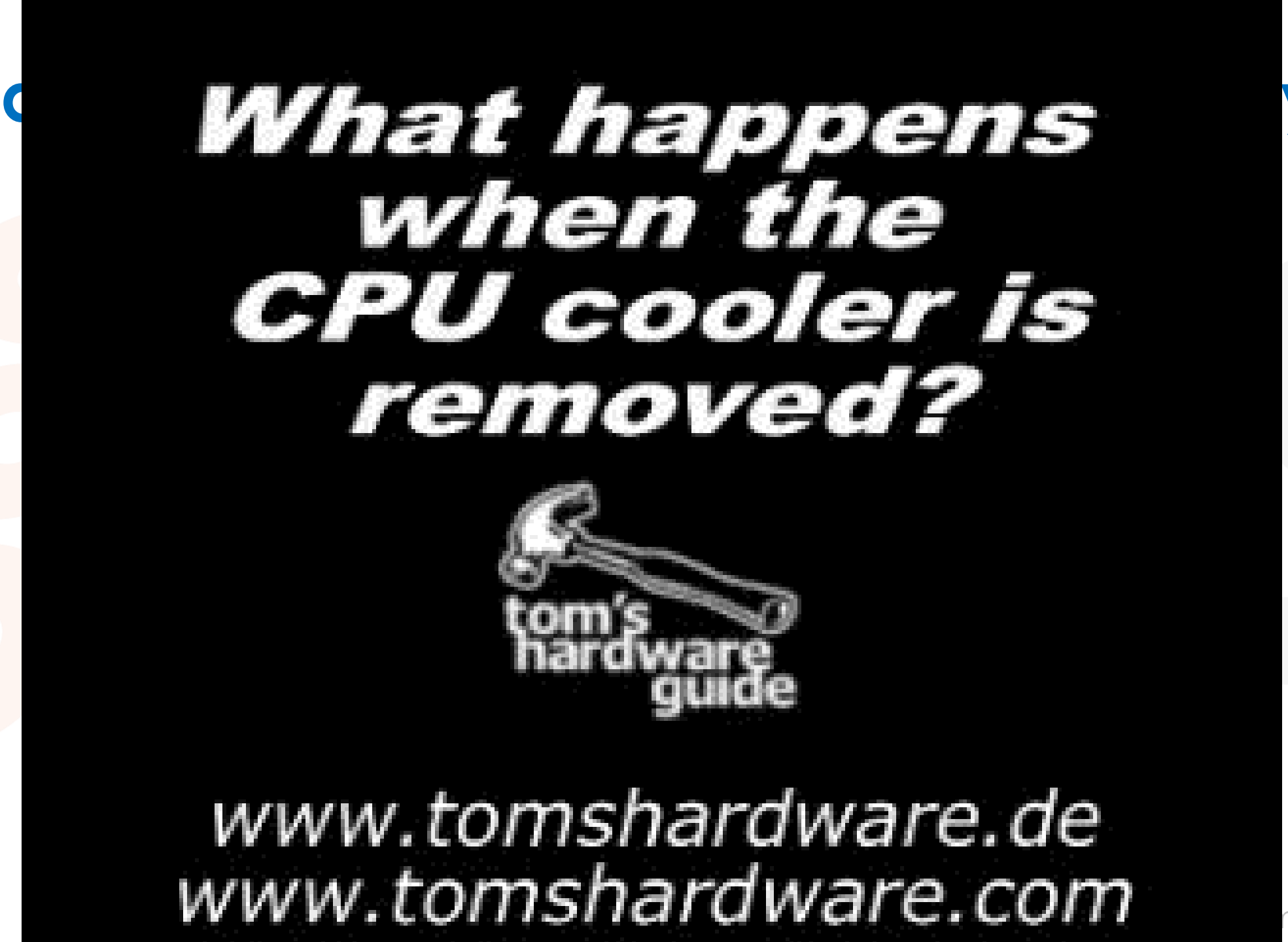
Power density (2004 expectation)



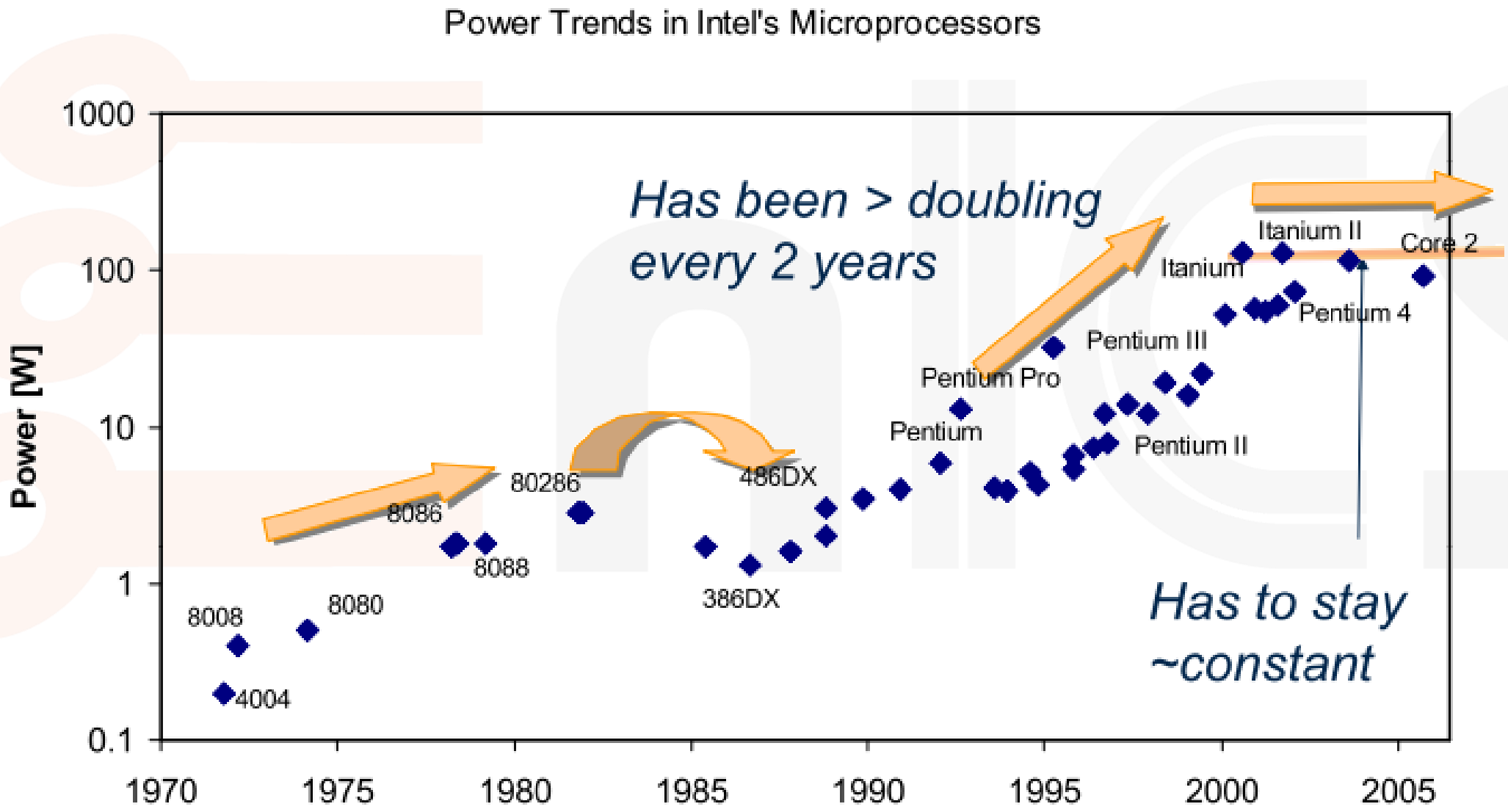
The Power Density Crisis

Patrick Gelsinger, Intel
ISSCC 2001

Who y…?


What happens when the CPU cooler is removed?
tom's hardware guide
www.tomshardware.de
www.tomshardware.com

What actually happened?



Technology Scaling Models

- **Fixed Voltage Scaling**
  - Supply voltages have to be similar for all devices (one battery)
  - Only device dimensions are scaled.
  - 1970s-1990s
- **Full "Dennard" Scaling (Constant Electrical Field)**
  - Scale both device dimensions and voltage by the same factor, *S*.
  - Electrical fields stay constant, eliminates breakdown and many secondary effects.
  - 1990s-2005
- **General Scaling**
  - Scale device dimensions by *S* and voltage by *U*.

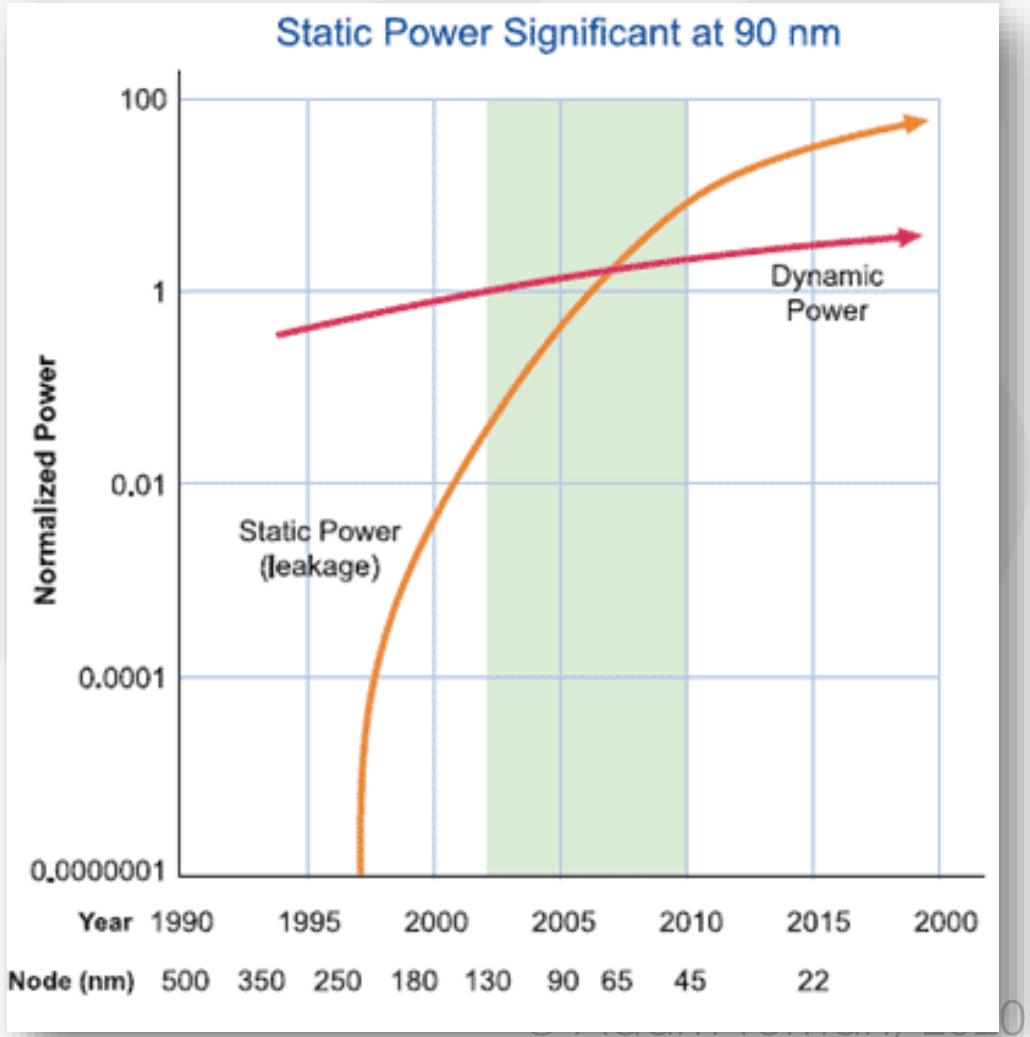
How about Leakage Power?

- We saw that the off current is exponentially dependent on the threshold voltage.

$$I_{off} \propto e^{-V_T / n\phi_T}$$

- In the case of *Full Scaling*, the leakage current *increases exponentially* as V_T is decreased!

- Since the 90nm node, static power is one of the major problems in ICs.



The NanoScaled Transistor:
Secondary Effects and Solutions

Emerging Nanoscaled Integrated Circuits and Systems Labs
Bar-Ilan University
אוניברסיטת בר-אילן

50 Years of Scaling

- = 50 years of headaches…
- How did brilliant process engineers deal with all the problems introduced by pushing physics to the limit?
- Let's remember some phenomena introduced by technology scaling and how they've (temporarily) been solved.

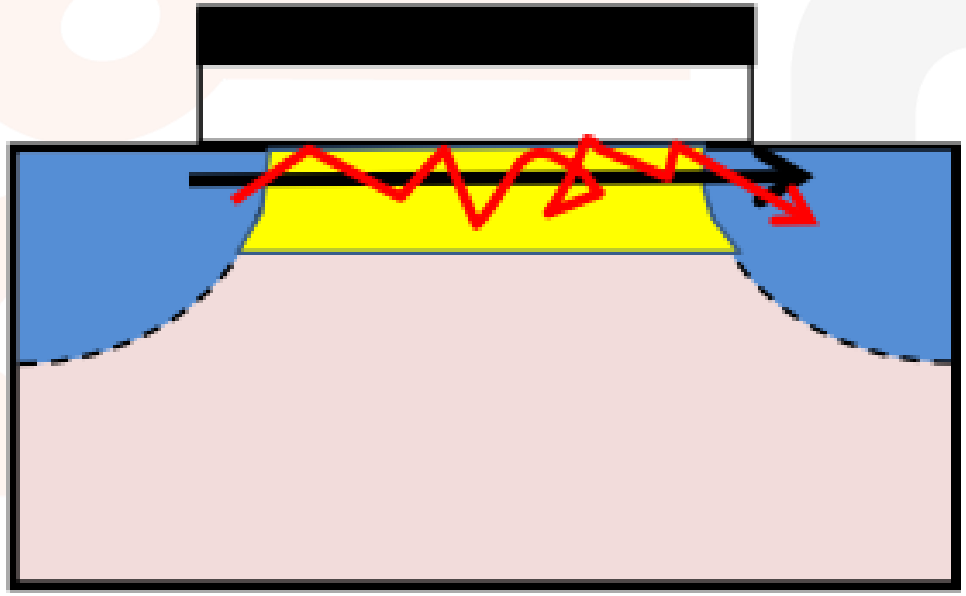
Problem: Mobility Degradation

- **Reminder: we have degraded mobility due to:**
  - Velocity Saturation
  - Surface Scattering
  - …and in general, we want more speed!

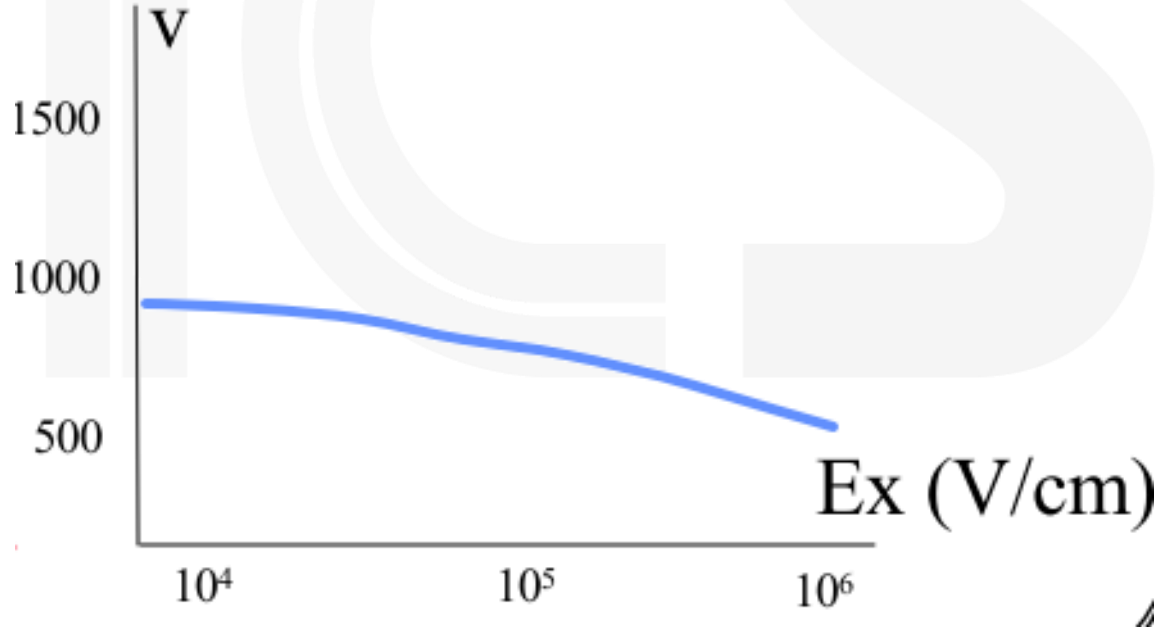
- **There are a number of solutions that are currently used or are being developed:**
  - Strained Silicon
  - Miller Index
  - Different Materials

Problem: Surface Scattering

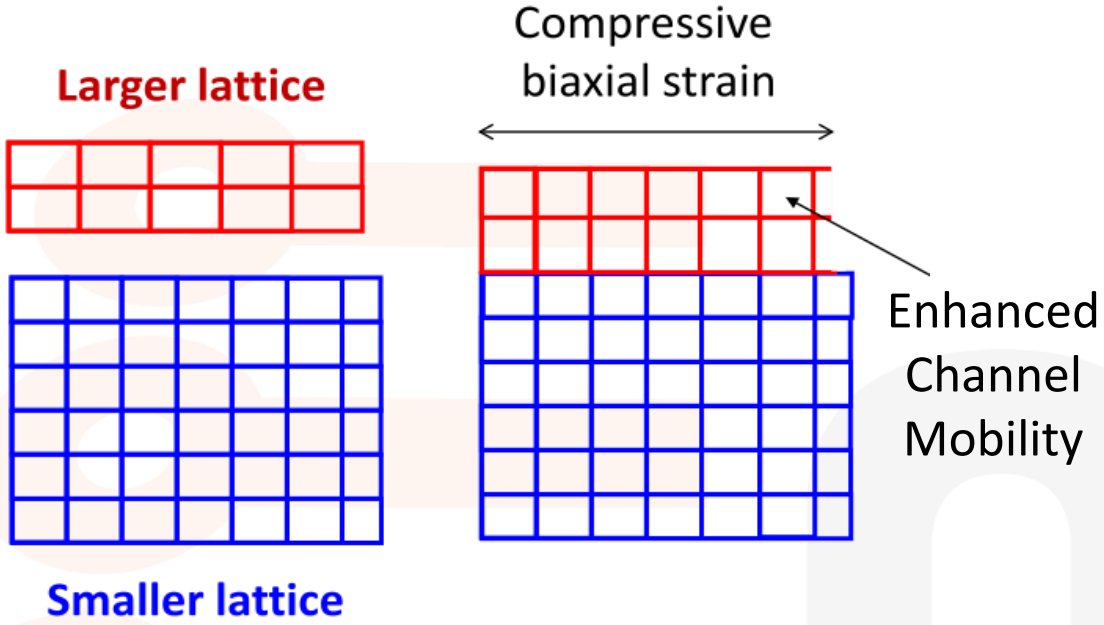
- The mobility at the surface is vertical field (V_G) dependent.
- The stronger the field, the more carriers "hit" the interface and scatter.

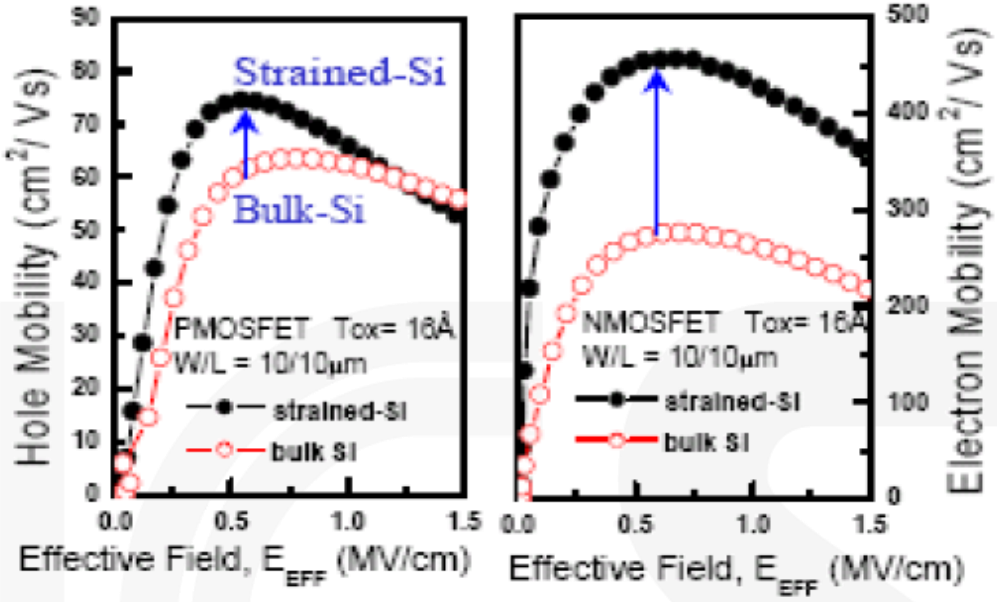


$$\mu \propto \frac{\mu_0}{1+\theta\left(V_{GS}-V_T\right)}$$

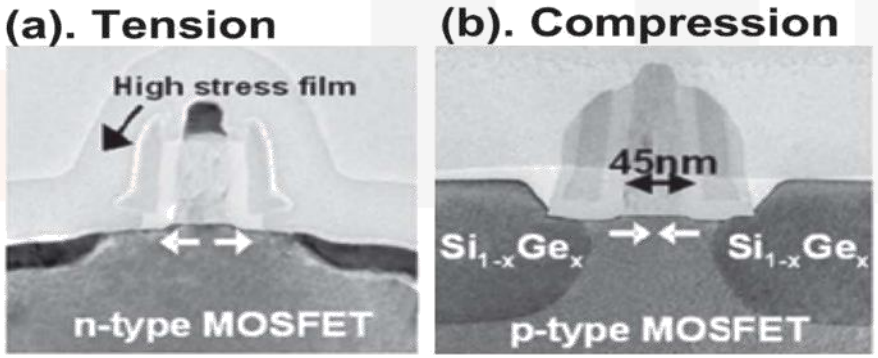


Solution: Strained Silicon

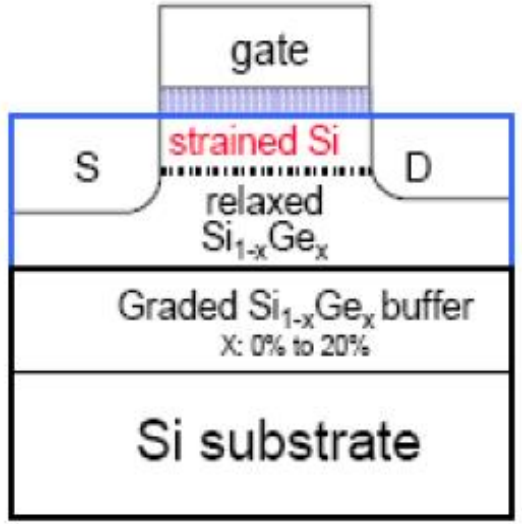




Chang, IEDM 2005



Source: S. E. Thompson et al., "A logic nanotechnology featuring strained-silicon," IEEE TRANSACTIONS ON ELECTRON DEVICES, Vol. 25, pp. 191 - 193, April 2004.

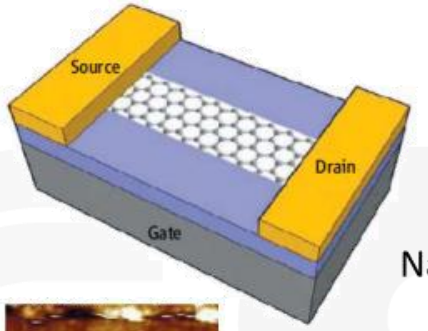


Solution: Improved Channel Materials

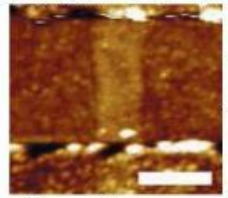
	Speed of Charges in Different Materials ($cm^2/V \cdot s$)				
Charges	Si	GaAs	$In_{0.53}Ga_{0.47}As$	InAs	InSb
Electrons*	300	7000	10,000	15,000	30,000
Holes*	450	400	200	460	1250

*Electron carrier mobilities measured in transistor channels with electron concentration of $1x10^{12}$ cm^{-2}. Hole mobilities in bulk.

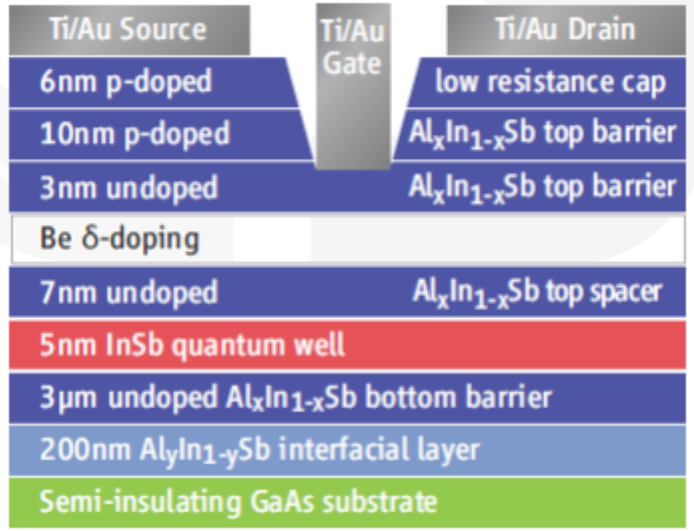
Ge
*
1900



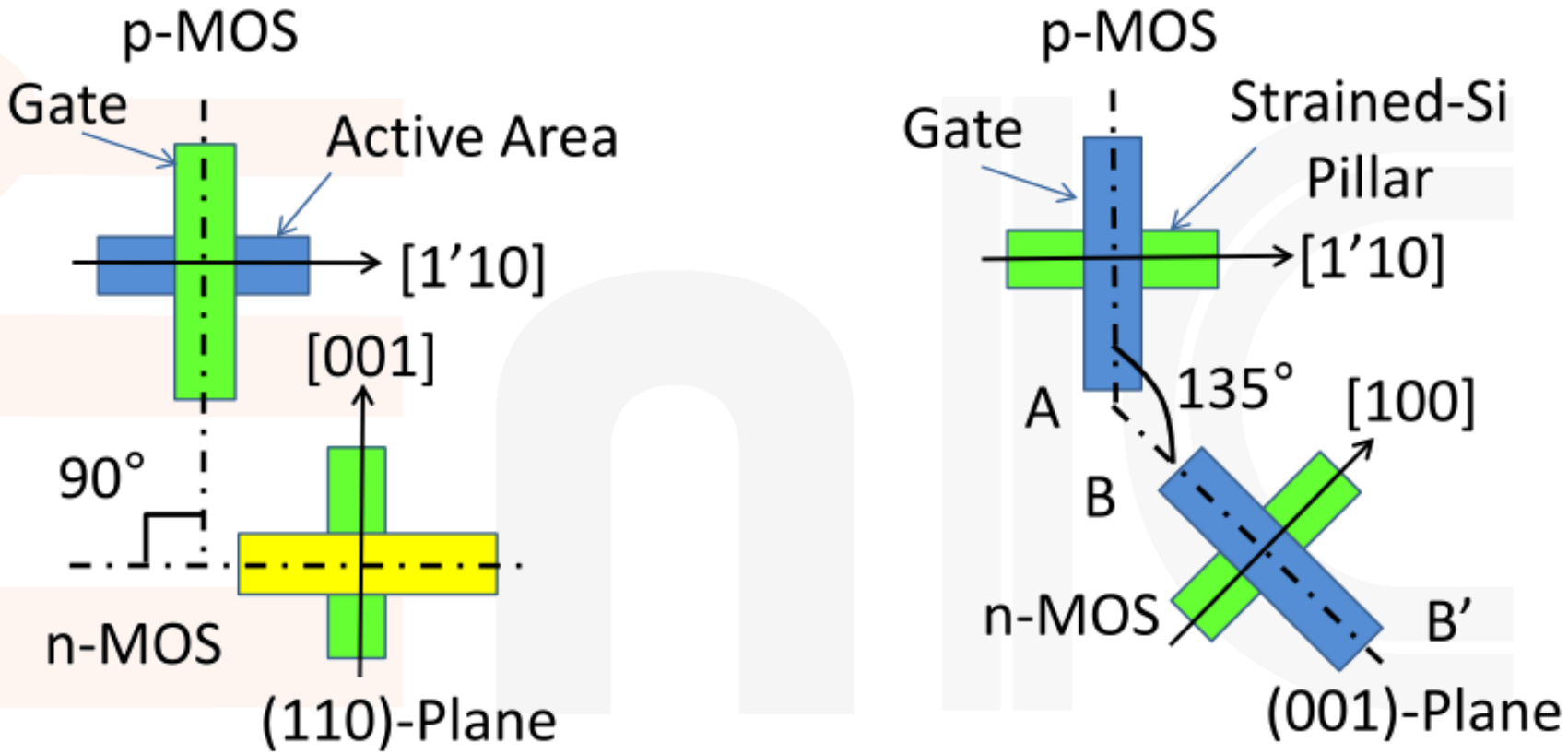
Nature, 2009



New kid. Transistors made from graphene nanoribbons could be blinding fast. But can they perform on an industrial scale?



Solution: Silicon Orientation



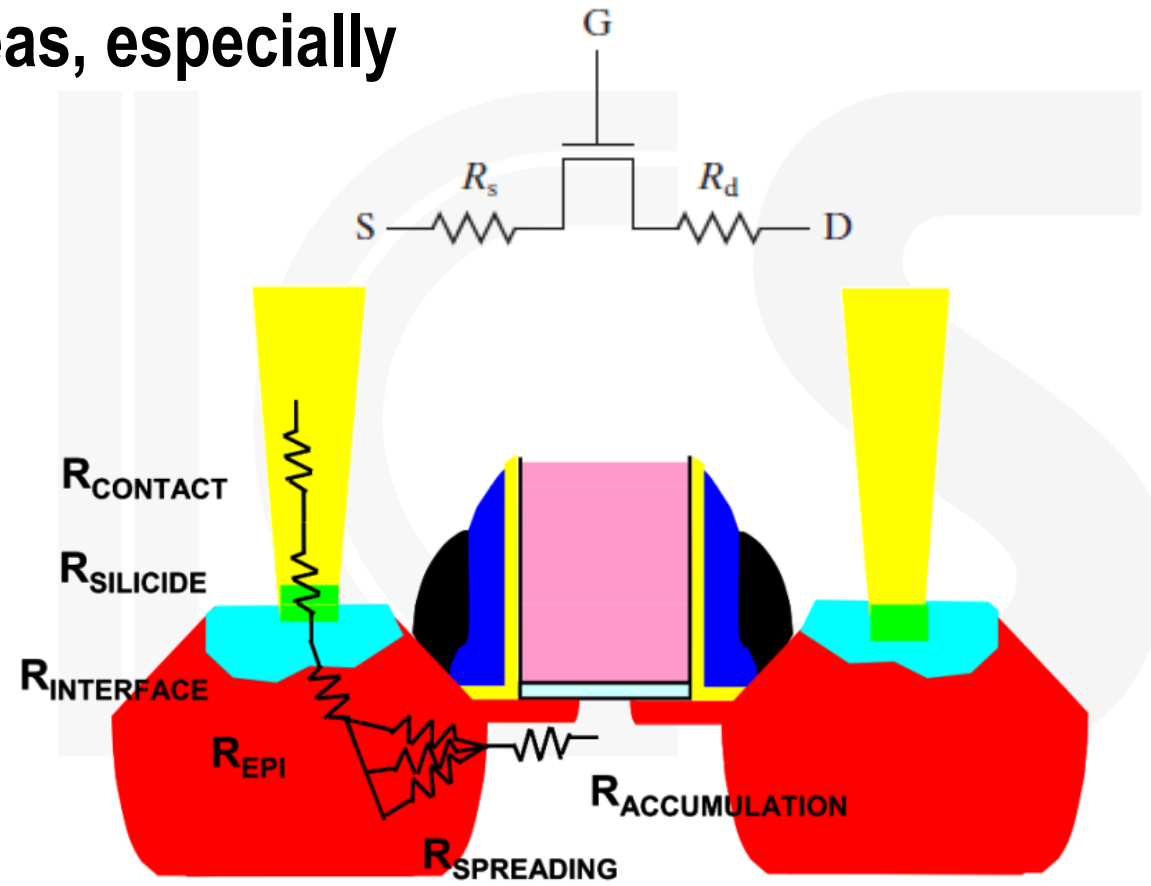
Takagi, TED 52, p.367, 2005

Problem: Serial Resistance

- **The resistance of the Source and Drain areas, especially with *Lightly Doped Drain* (LDD), can have a large impact (>15%) on transistor conductance**

$$I_{\text{Dsat}} = \frac{I_{\text{Dsat}0}}{1 + \frac{I_{\text{Dsat}0} R_S}{(V_{GS} - V_T)}}$$

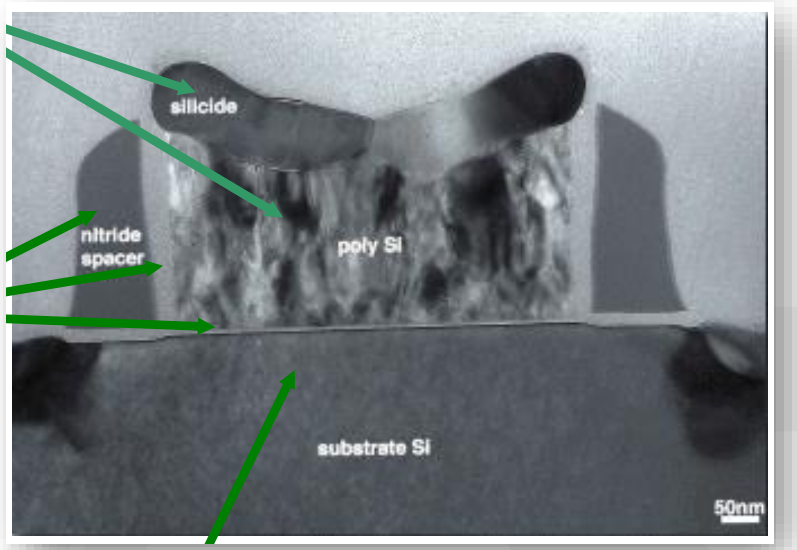
$$V_{\text{Dsat}} = V_{\text{Dsat}0} + I_{\text{Dsat}0}\left(R_S + R_D\right)$$

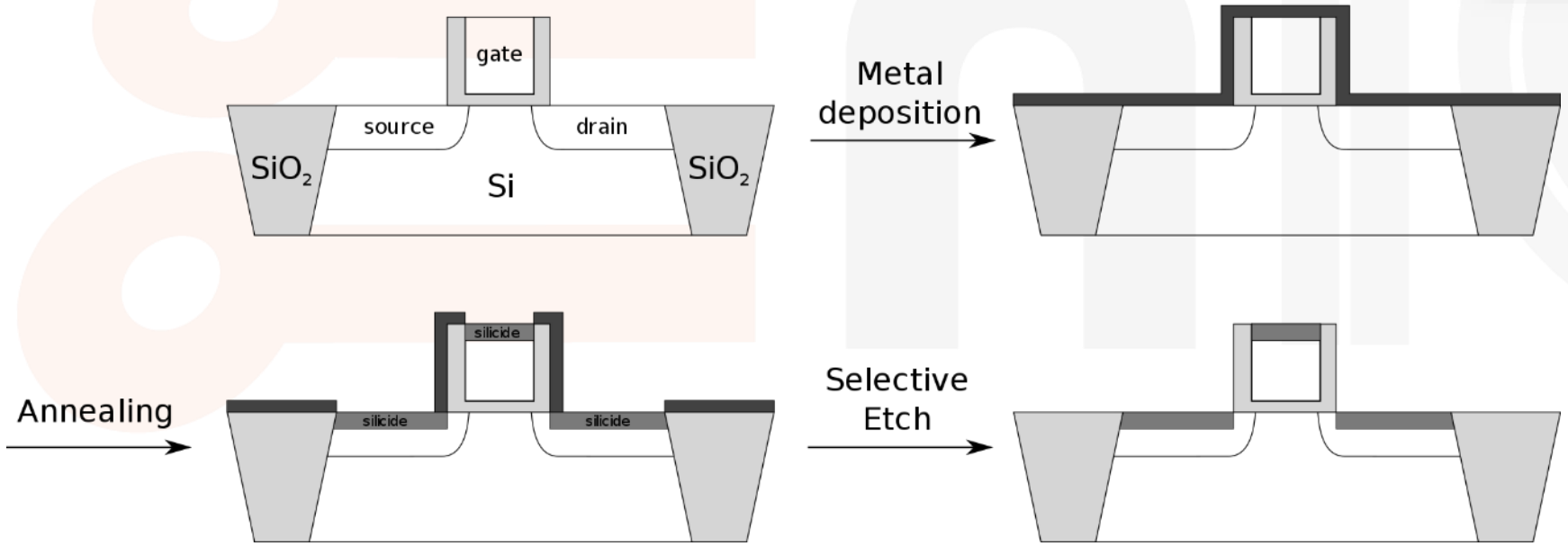


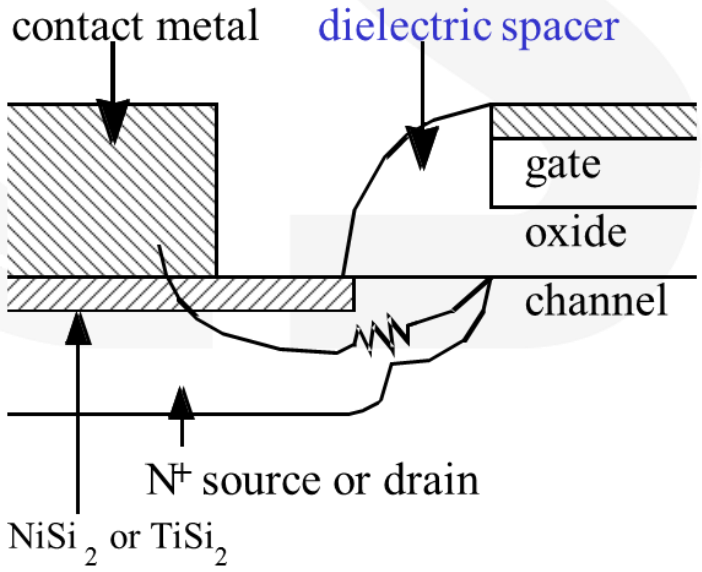
K. Kuhn, IEDM 2008

Solution: Salicides

- **Salicide = Self Aligned Silicide**
  - A low resistance contact is formed on the diffusion and polysilicon surface through annealing.



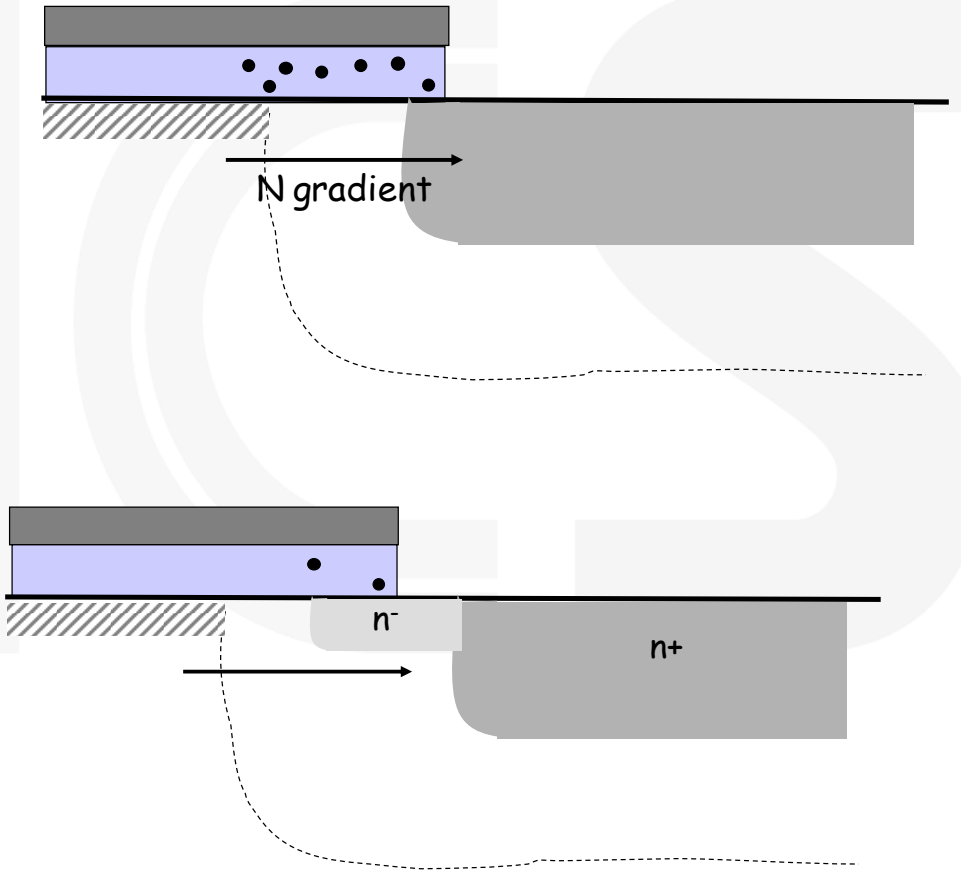




Source: Wikipedia

Problem: Hot Carrier Effects

- **Reminder:**
  - Carriers in strong electric fields jump over the gate energy barrier and get stuck in the oxide.
  - This causes V_T to change over time.
  - Happens mainly close to the drain where strong fields exist.
- **Solution:**
  - To solve this, ***Lightly Doped Drains*** (LDD) are used.
  - The N-area reduces the field gradient near the drain.
  - $N+$ is needed for ohmic contacts.



Problem: Subthreshold Leakage

- **Reminder:**
  - There is a finite number of free carriers in the channel when $V_{GS}<V_T$.
  - The Body-to-Channel Cap limits the gate control over the channel.
  - Roll-Off (SCE) and DIBL cause an exponential current increase.
- **Solutions:**
  - Multi Threshold Devices
  - Body Biasing
  - Shallow Diffusions
  - Thin Oxides
  - Multi-gate Transistors
    - Silicon on Insulator (SOI)
    - FinFET/Tri Gate Transistors

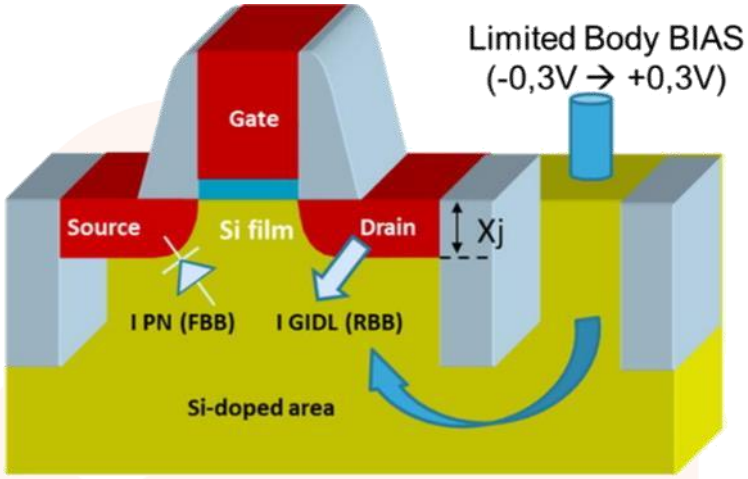
Solution: MTCMOS

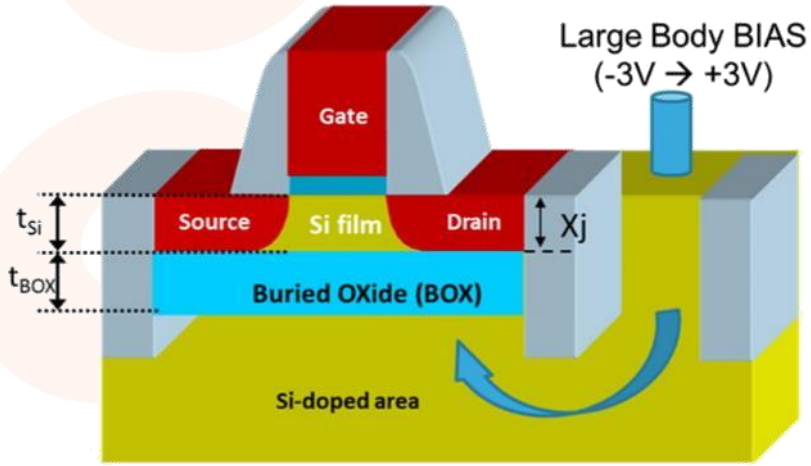
- **A very common solution, already available in most PDKs for many generations is multi-threshold voltage devices.**
  - High-V_T (HVT) devices are slow but have lower leakage
  - Low-V_T (LVT) devices are fast but very leaky
  - Nominal/Standard/Regular-V_T (NVT/SVT/RVT) transistors are best for most operations.
- **This approach provides an easy way to trade off high-performance and low leakage.**
- **Standard cells are often designed with identical footprints in order to exchange various V_T options seamlessly.**

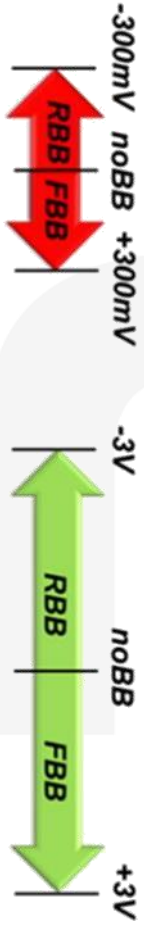
Solution: Body Biasing

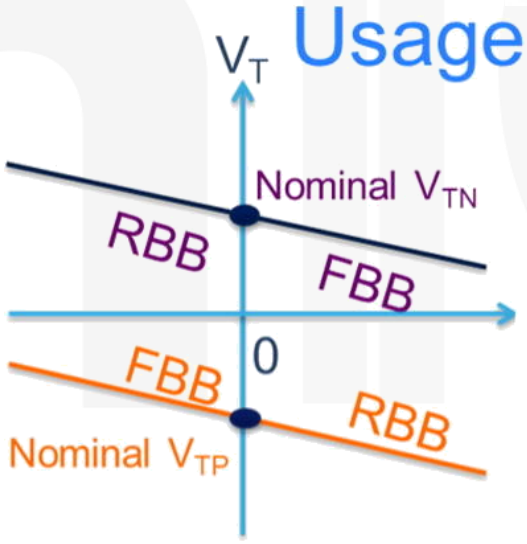
- As we have previously seen, the body voltage of the transistor affects the threshold voltage of the transistor.
- The body voltage of PMOS devices, residing in a shared NWELL can be set without the need for special process steps.
- Changing the body voltage of NMOS devices requires the use of an isolated PWELL (IPW) inside a Deep NWELL (DNW) area.
- The effectiveness of body biasing has thoroughly degraded with process scaling.
- New technologies, such as Fully-depleted Ultra-thin body and buried oxide Silicon-on-Insulator (FD-SOI or UTBB-SOI) have given this technique "new life".

Solution: Body Biasing





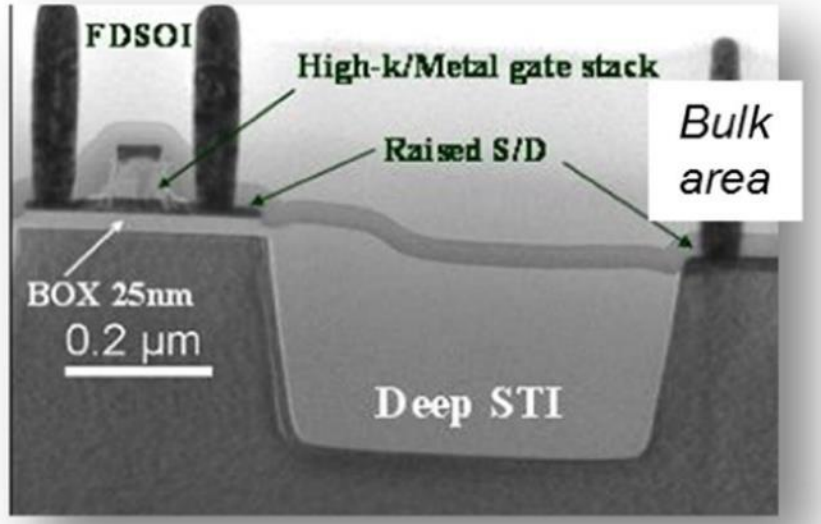




Usage

Vbody bias

- ⇨ *More efficient than on bulk*
- ⇨ *Dynamic regime*
- ⇨ *Not degraded with scaling in FDSOI*
- ⇨ *Inefficient in Finfet*



Source: Monfray, Solid State Electronics, 2016

Solution: Vertical Dimensions

- **A better model of V_T, taking into account both DIBL and Roll-Off shows a strong dependence on the transistor's vertical dimensions:**
  - t_{ox} – gate thickness
  - W_{dep} – depletion width
  - X_j – S/D junction depth

$$V_T = V_{T\text{-long}} - (V_{DS} + 0.4) \cdot e^{-\frac{L}{l_d}}$$

$$l_d \approx \sqrt[3]{t_{ox} W_{dep} X_j}$$

- **Reduction of each of these provides a smaller L_{min} and other advantages.**
  - Smaller t_{ox} means better transconductance.
  - Smaller X_j means lower S/D capacitance
  - Smaller W_{dep} means better drain-channel isolation

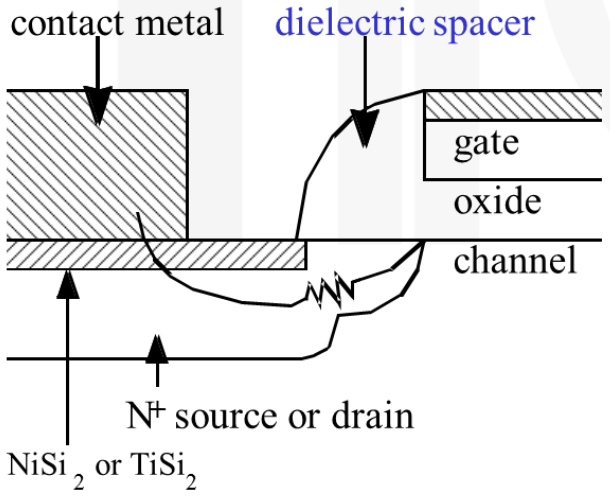
Solution: Vertical Dimensions

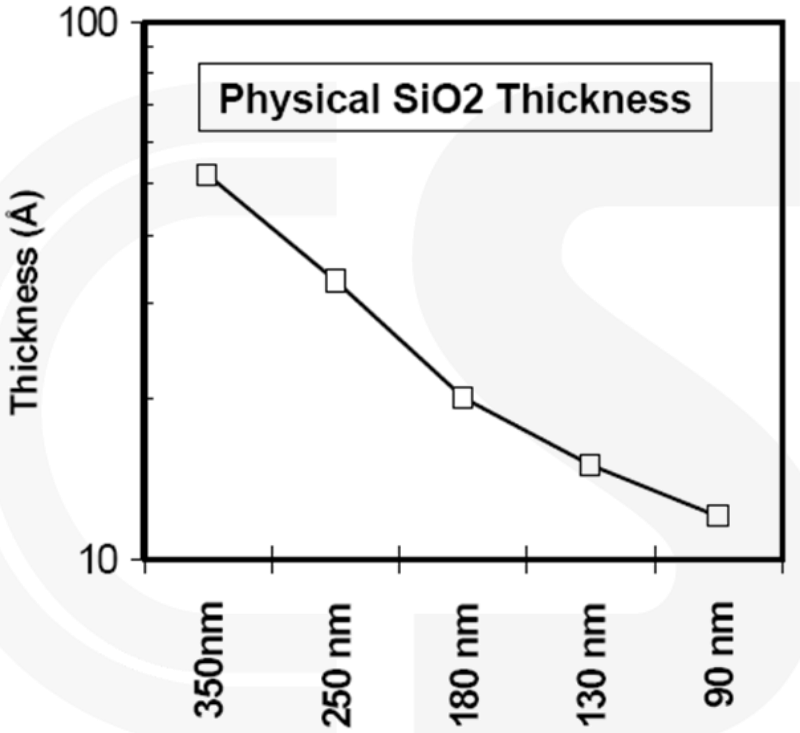
$$l_d \approx \sqrt[3]{t_{ox} W_{dep} X_j}$$

- **Thinner oxide is the most important aspect of scaling, as it allows better gate control.**

- **Shallow junctions are achieved by:**
  - High Substrate Doping (*not wanted*)
  - Lightly Doped Drains (LDD)
  - Metal S/D

- **What about W_{dep}?**

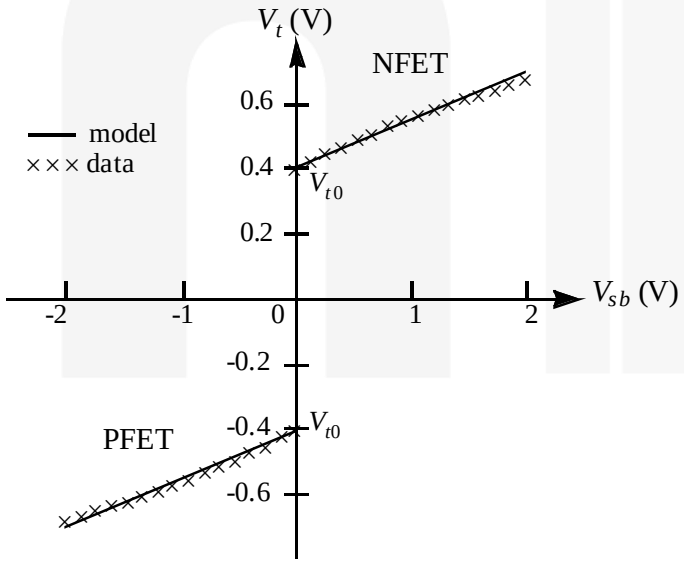


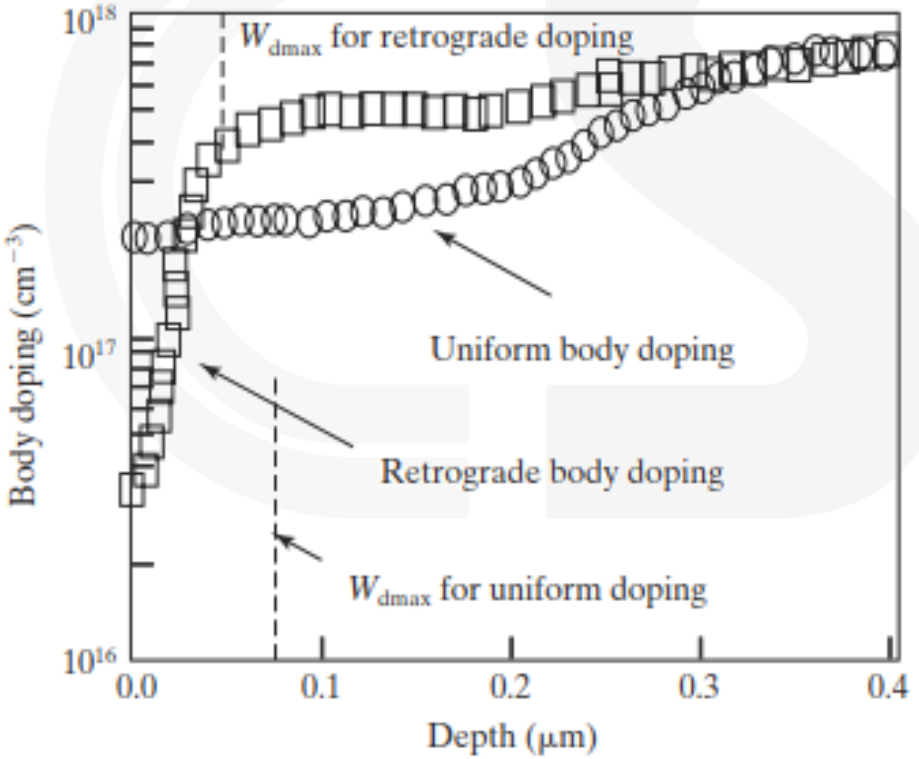


Solution: Vertical Dimensions

$$l_d \approx \sqrt[3]{t_{ox} W_{dep} X_j}$$

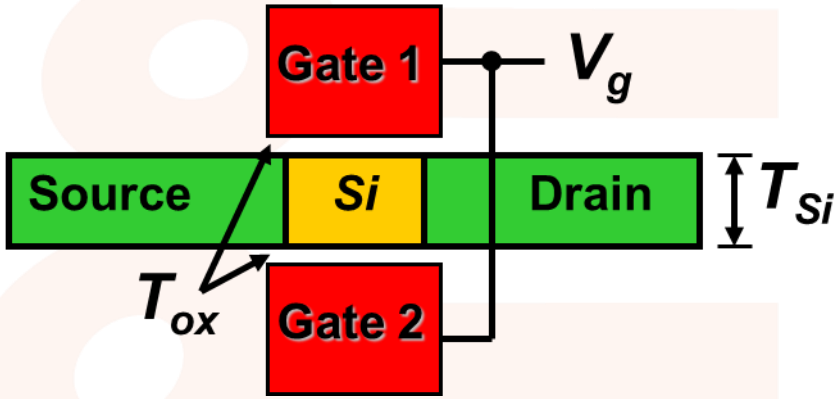
- W_{dep} **can be reduced by "Retrograde Doping"**
  - Reduces impurity scattering (improves mobility).
  - W_{dep} does not vary with V_{SB}.
  - Causes a linear body effect.
  - Reduces punchthrough



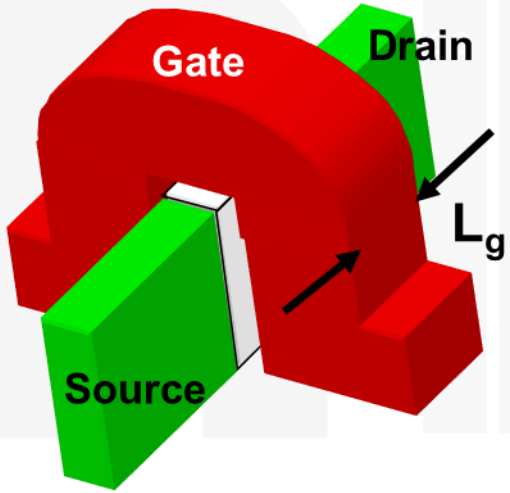


Solution: Multi-gate Transistors

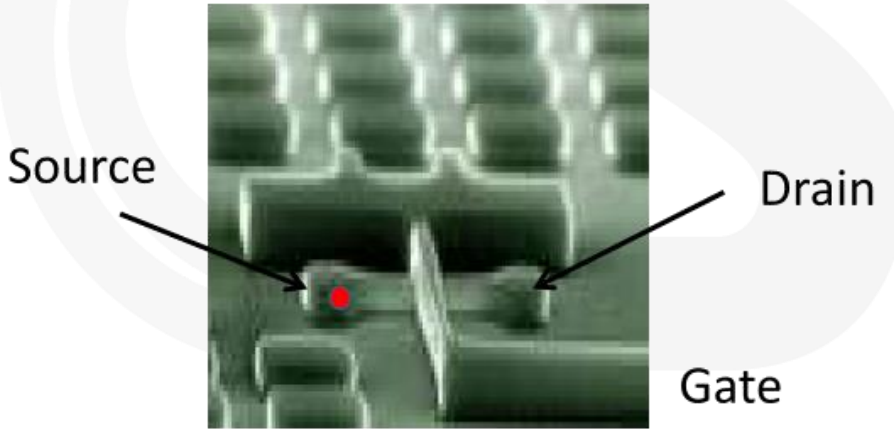
- To gain better control of the channel (better subthreshold slope) use multiple gates:



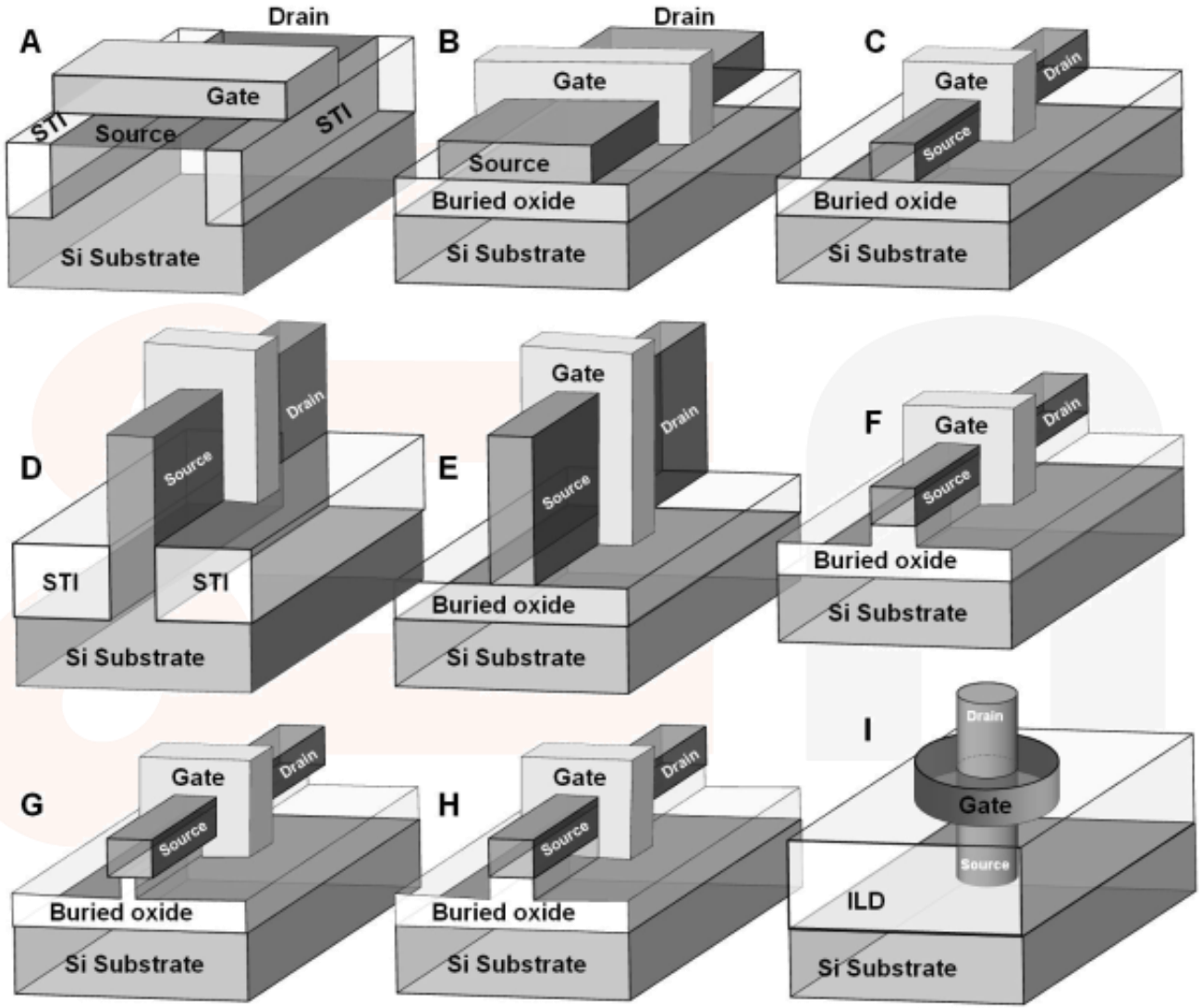
Double-gate MOSFET



FinFET



Solution: Multi-gate Transistors

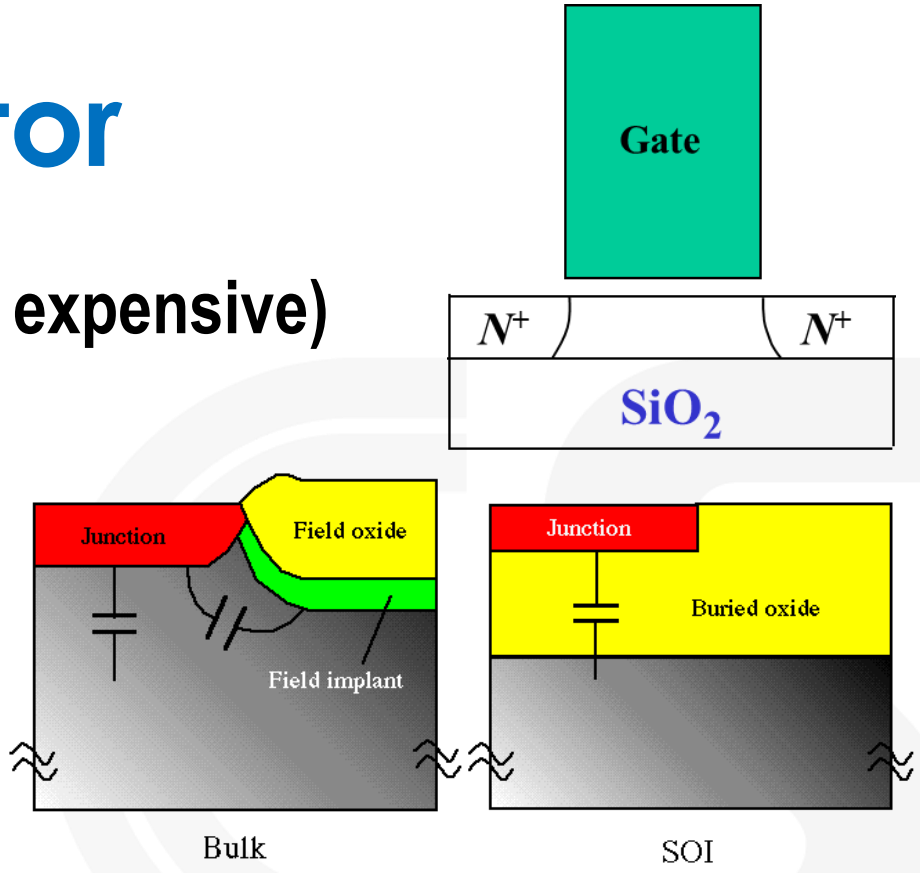


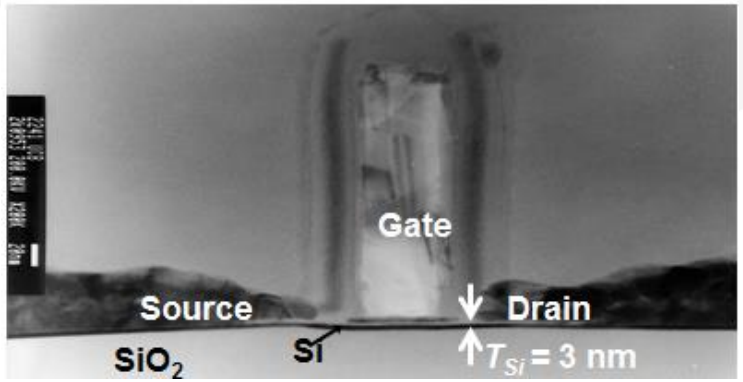
A. Planar bulk MOSFET
B. SOI MOSFET
C. Tri-gate SOI nanowire MOSFET
D. Tri-gate Bulk FinFET
E. Tri-gate SOI FinFET
F. Pi-gate SOI nanowire MOSFET
G. Omega-gate SOI nanowire MOSFET
H. Horizontal Gate-all-around (GAA) nanowire MOSFET
I. Vertical GAA nanowire MOSFET

Source: Ferain, Nature 2011

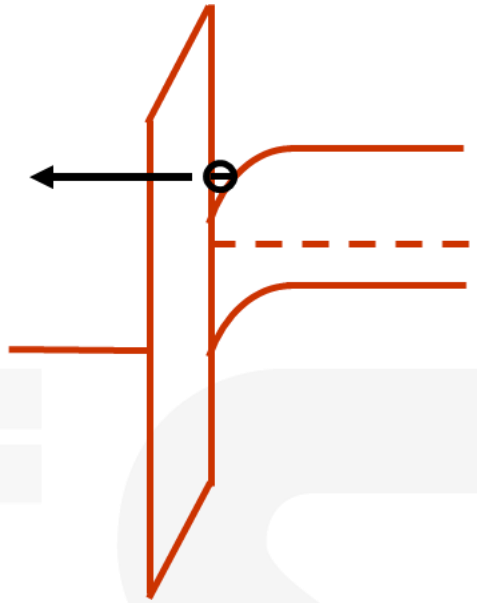
Solution: Silicon-On-Insulator

- **Silicon on Insulator (SOI) is a better (though more expensive) way to make transistors:**
  - Fixed depletion widths (W_{dep}, X_j) – lower Roll-Off/DIBL
  - No Punchthrough.
  - Junction to substrate parasitic capacitance small.
  - No latch-up between NMOS and PMOS (no substrate)
  - Raised Drain/Source for reduced resistivity.





Problem: Gate Leakage



- **Reminder:**
  - Gate leakage is exponentially dependent on t_{ox}.
  - A thin oxide reduces Roll-Off.
  - A thin oxide provides higher transconductance (C_{ox})
  - A thin oxide reduces the subthreshold swing:

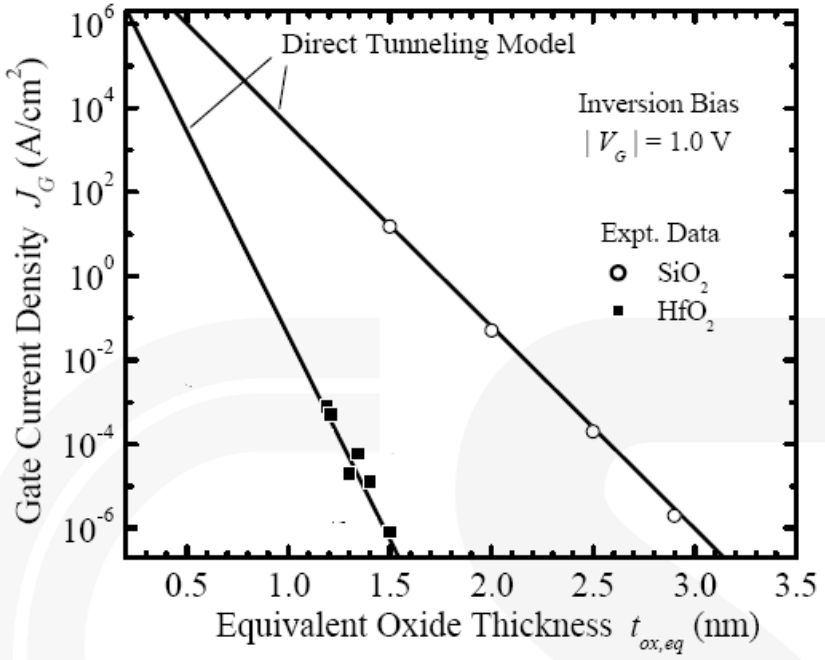
$$n \equiv 1 + \frac{C_{dep}}{C_{oxe}} \qquad S \equiv \frac{1}{n\phi_T \ln 10}$$

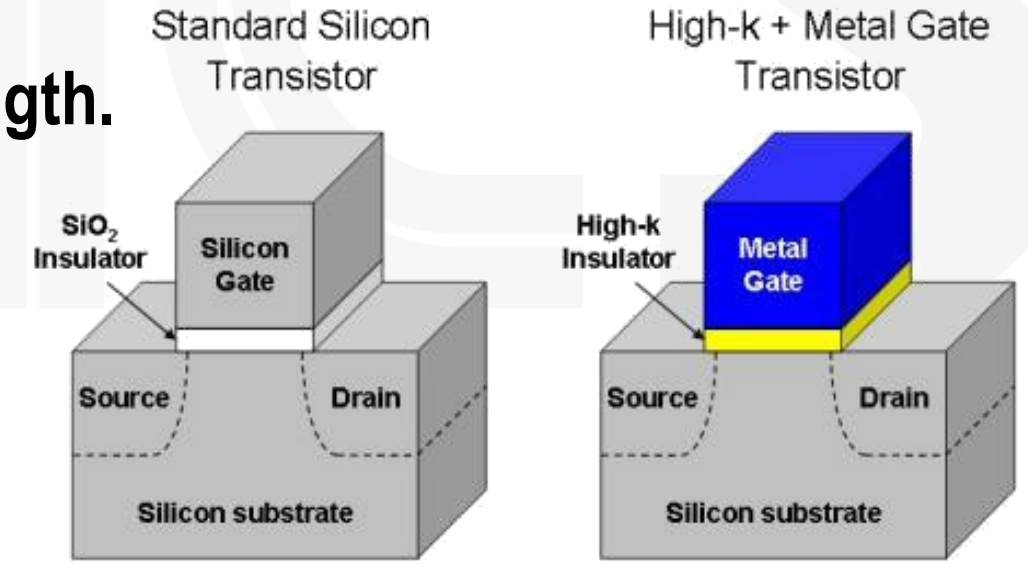
- **Major Problem:**
  - Oxide thickness reached 1.2nm (5 atomic layers) in 65nm.
  - Gate leakage is intolerable under $t_{ox}=1.5\text{nm}$
  - The end of scaling???

Solution: High-K Dielectrics

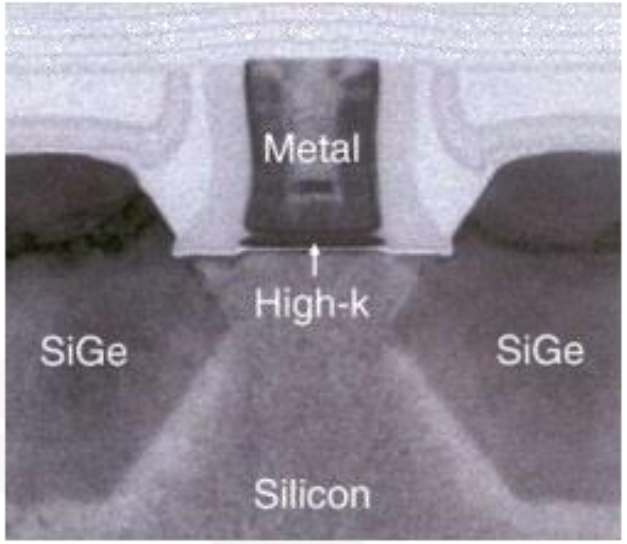
- **Higher K ($\mathcal{E}$) means C_{ox} is increased, therefore:**
  - Transconductance is higher.
  - $n=1+C_{dep}/C_{ox}$ is lower → Better gate control.

- **So we can use <u>thicker gates</u> to eliminate tunneling without losing control or drive strength.**





Solution: High-K Dielectrics

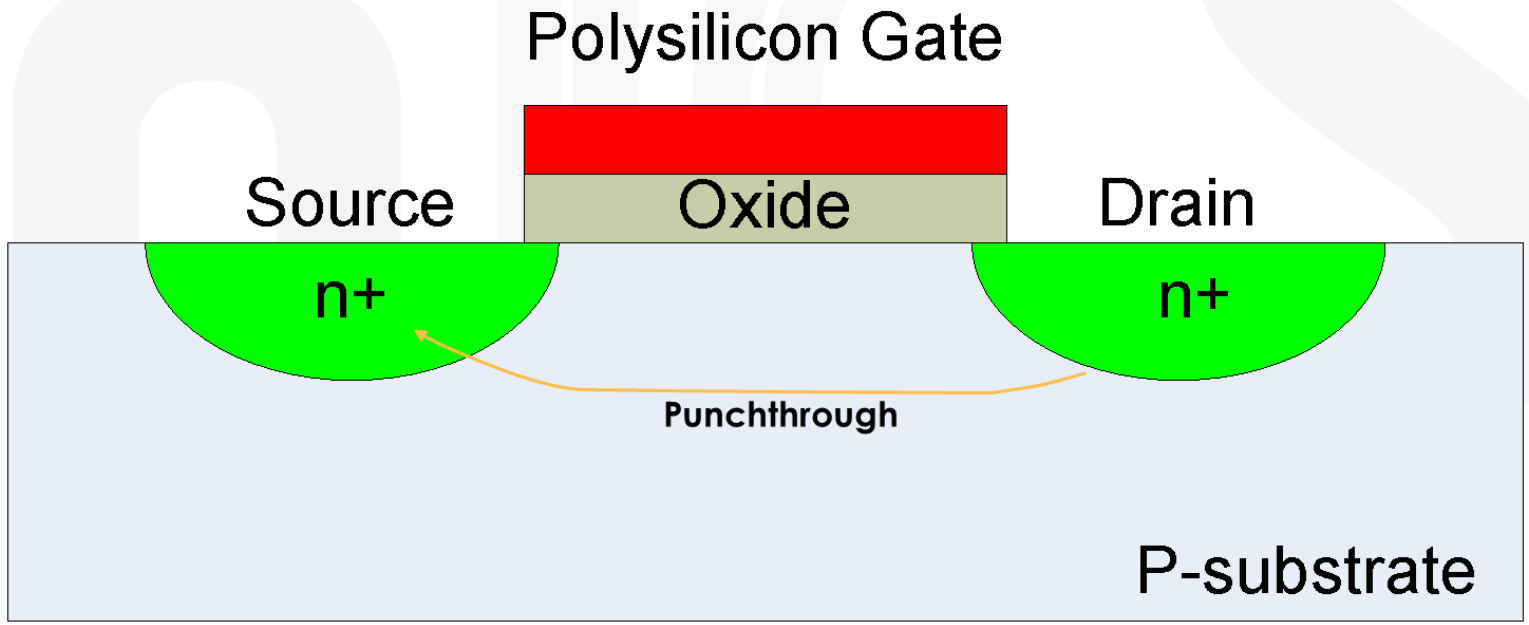


- HfO_2 **has a relative dielectric constant (**k**) of** ~24, **six times larger than that of** SiO_2**.**
  - For the same EOT, the HfO_2 film presents a much thicker (albeit a lower) tunneling barrier to the electrons and holes.
  - $Toxe$ can be further reduced by introducing metal-gate technology since the poly-depletion effect is eliminated.
- **The difficulties of high-k dielectrics:**
  - Chemical reactions between them and the silicon substrate and gate
  - Lower surface mobility than the Si/SiO_2 system
  - Too low a V_T for pMOS (as if there is positive charge in the *high-k* dielectric).
  - Long-term reliability

Problem: Punchthrough Currents

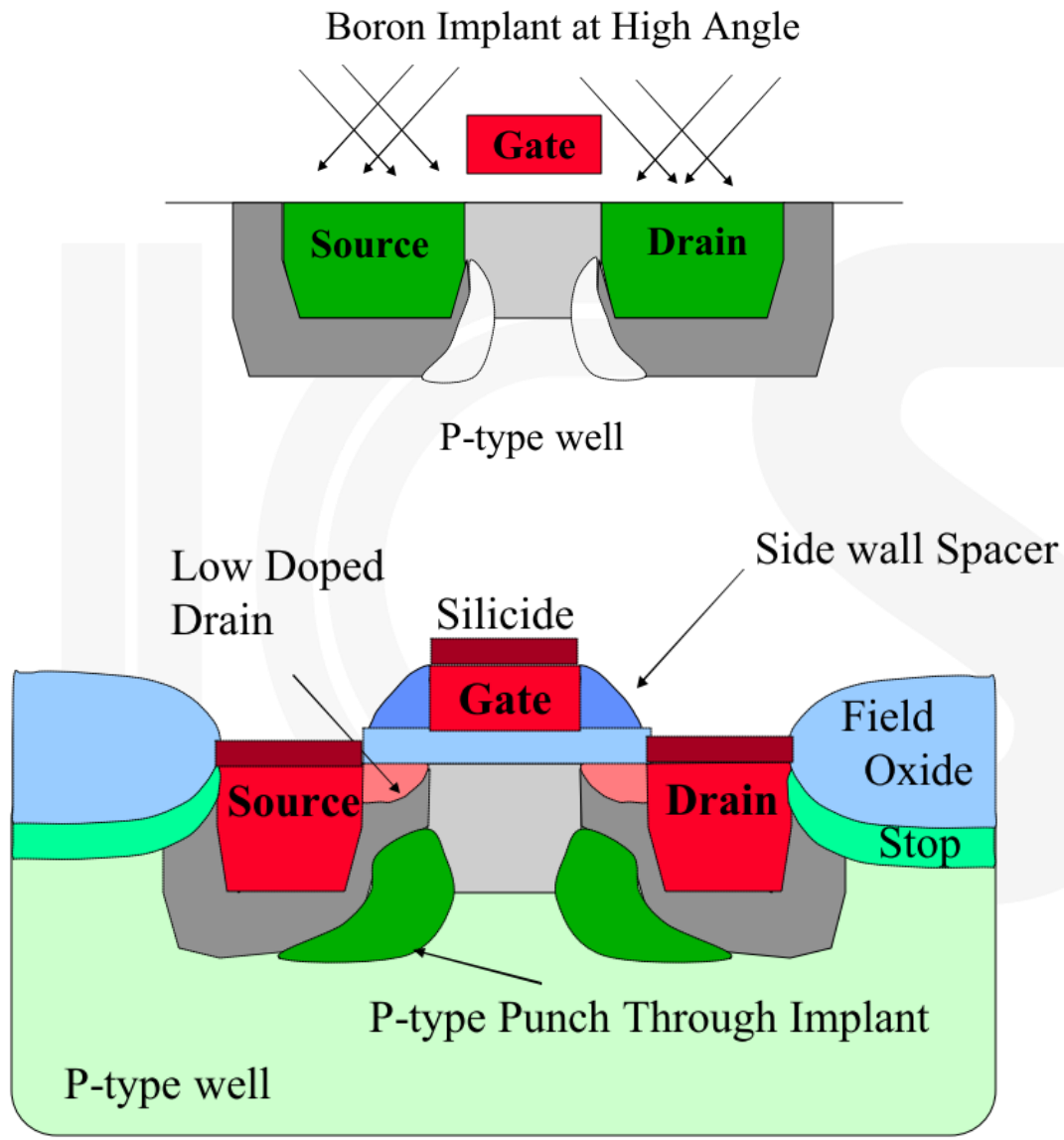
- **Reminder:**
  - Drain and Source depletion regions “connect” to each other deep underneath the channel.

- **Solutions:**
  - Halo Implants
  - LDD
  - Retrograde Doping
  - SOI



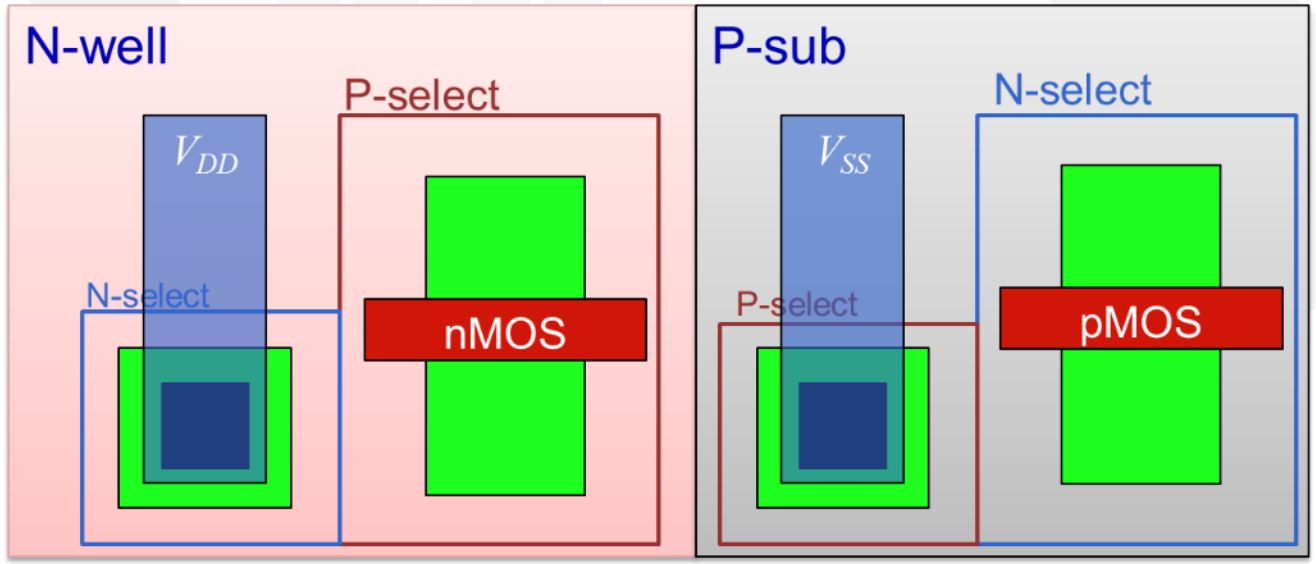
Solution: Halo Implants

- p+ implants suppress the source/drain depletion regions.
- However, they also cause the Reverse Short Channel Effect.



Problem: Latch Up

- **Reminder:**
  - A Thyristor is created because of voltage drop over the substrate (or well).
  - In general, cases of forward biased diodes are often called ***latchup***.



- **Solutions:**
  - Body Taps
  - SOI

Current and Future Trends

Emerging Nanoscaled Integrated Circuits and Systems Labs

Bar-Ilan University
אוניברסיטת בר-אילן

ITRS

- **International Technology Roadmap for Semiconductors**
  - Started in 1998 to predict the future of the semiconductor industry

Year	2009	2012	2015	2018	2021
Feature size (nm)	34	24	17	12	8.4
L_{gate} (nm)	20	14	10	7	5
V_{DD} (V)	1.0	0.9	0.8	0.7	0.65
Billions of transistors/die	1.5	3.1	6.2	12.4	24.7
Wiring levels	12	12	13	14	15
Maximum power (W)	198	198	198	198	198
DRAM capacity (Gb)	2	4	8	16	32
Flash capacity (Gb)	16	32	64	128	256

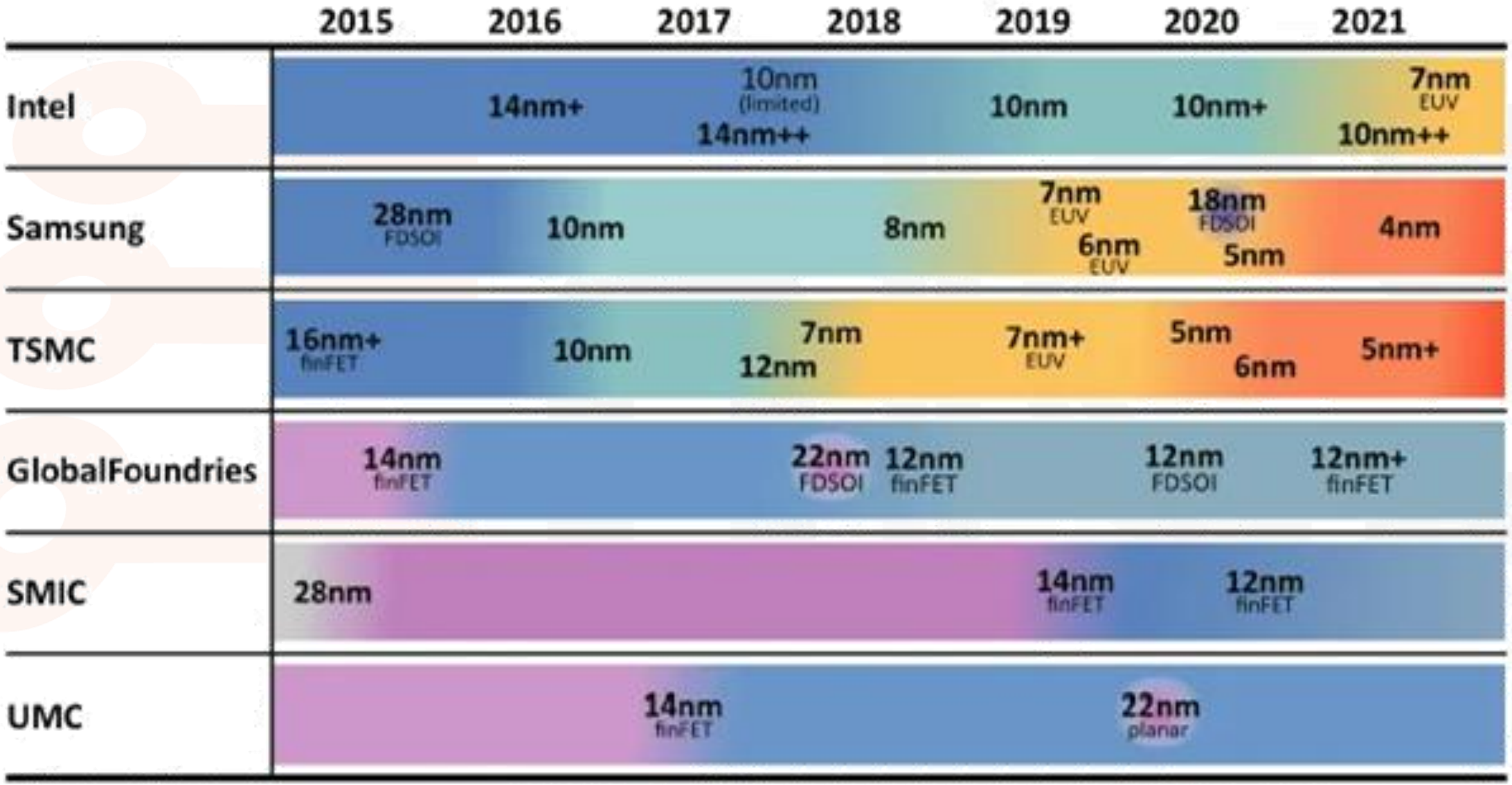
ITRS Retired

- **Moore's Law may not yet be dead, but the ITRS has been retired**
- **Let me introduce the "International Roadmap for Devices and Systems" (IRDS)**
  - Introduced in 2016 to expand the original focus to include systems

YEAR OF PRODUCTION	2018	2020	2022	2025	2028	2031	2034
	G54M36	G48M30	G45M24	G42M21	G40M16	G40M16T2	G40M16T4
Logic industry "Node Range" Labeling (nm)	"7"	"5"	"3"	"2.1"	"1.5"	"1.0 eq"	"0.7 eq"
IDM-Foundry node labeling	i10-f7	i7-f5	i5-f3	i3-f2.1	i2.1-f1.5	i1.5e-f1.0e	i1.0e-f0.7e
Logic device structure options	FinFET	finFET	finFET LGAA	LGAA	LGAA VGAA	LGAA-3D VGAA	LGAA-3D VGAA
Mainstream device for logic	finFET	finFET	finFET	LGAA	LGAA	LGAA-3D	LGAA-3D
Vdd (V)	0.75	0.70	0.70	0.65	0.65	0.60	0.60
Gate length (nm)	20	18	16	14	12	12	12
Number of stacked tiers	1	1	1	1	1	2	4
Number of stacked devices	1	1	1	3	3	4	4
Digital block area scaling - node-to-node	-	0.60	0.75	0.82	0.79	0.57	0.50
Cell height limitation - HD	device	M0	M0	M0	M0	M0	M0
SoC area scalling (stacked) - node-to-node	-	0.70	0.79	0.84	0.83	0.60	0.60
CPU frequency (GHz)	2.90	3.13	3.27	3.64	4.02	3.46	3.30
Frequency scaling - node-to-node	-	0.08	0.04	0.11	0.10	-0.14	-0.05
CPU frequency at constant power density (GHz)	2.90	1.92	1.69	2.14	1.93	1.25	0.72
Power at iso frequency - node-to-node	-	-0.23	-0.14	-0.36	-0.20	-0.12	-0.14
Power density - relative	1.00	1.64	1.94	1.70	2.08	2.78	4.55

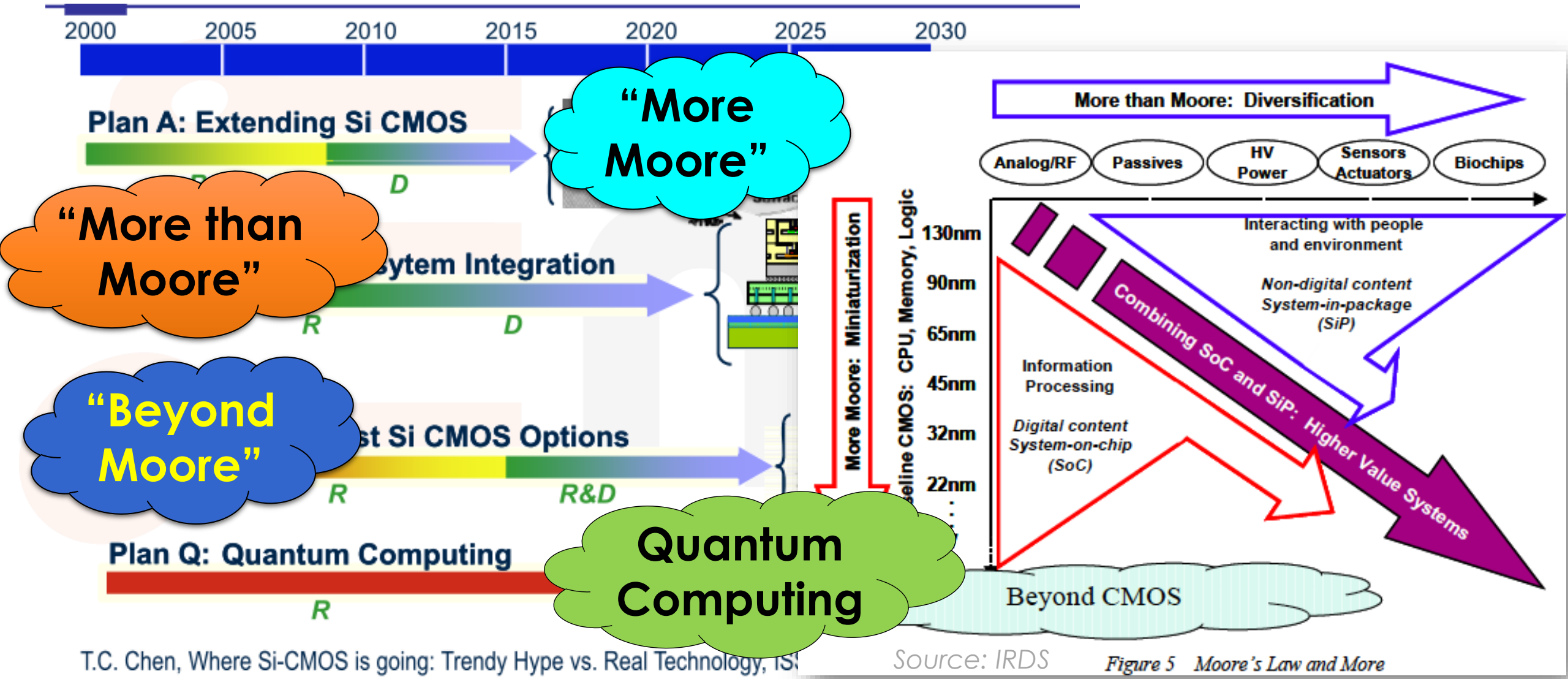
Source: IDRS 2018 Update (published 2019)

Foundry Roadmaps



Source: IC Insights

Technology Strategy Roadmap

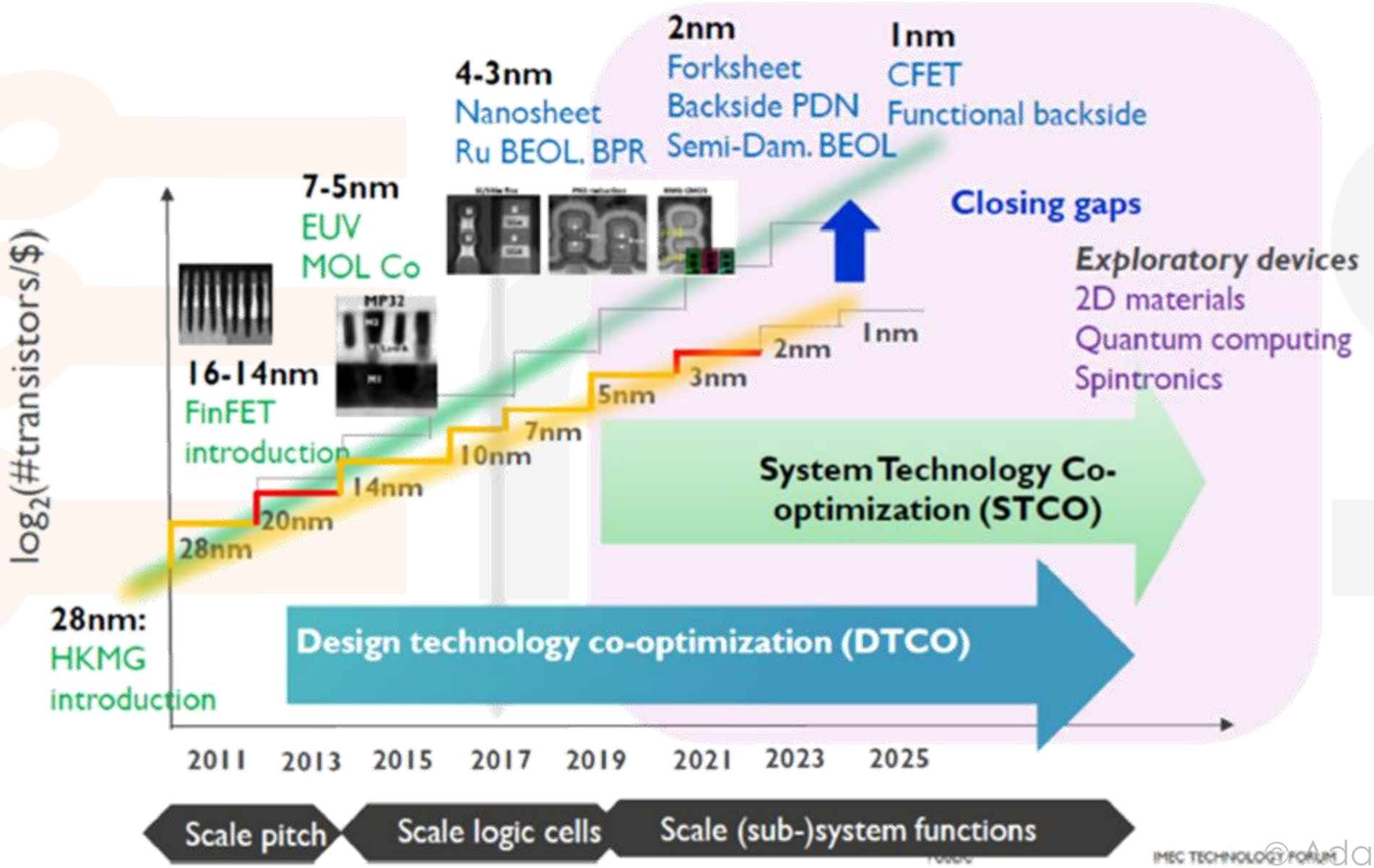


T.C. Chen, Where Si-CMOS is going: Trendy Hype vs. Real Technology, IS

Source: IRDS

Figure 5 Moore's Law and More

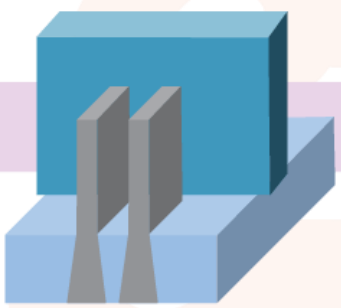
imec View of Technology Scaling



Source: imec

Next stage CMOS Devices

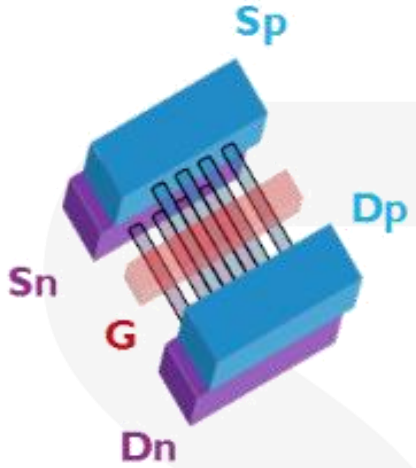
FinFET

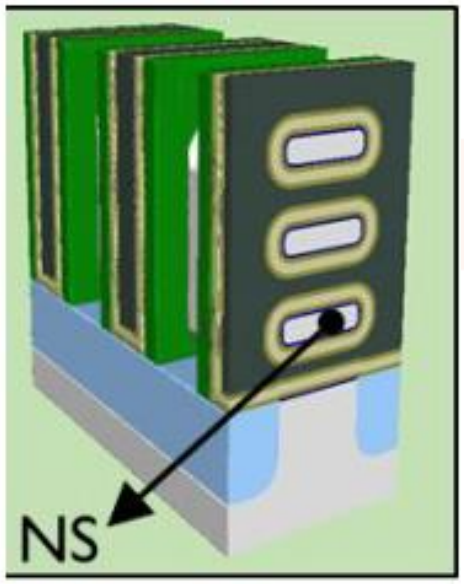


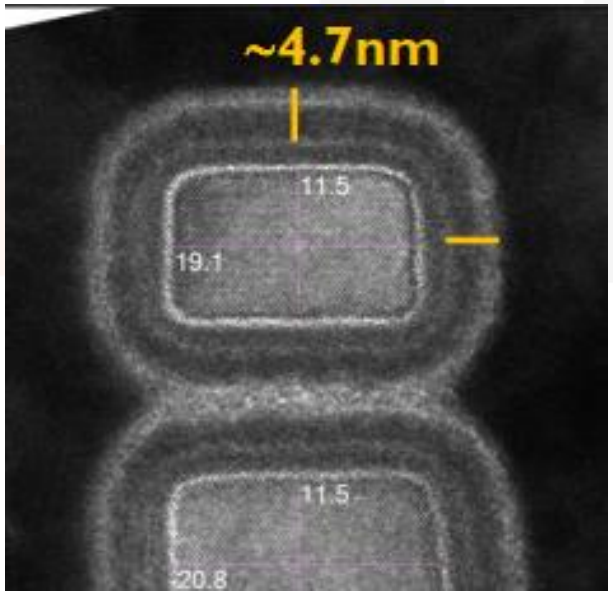
Nanosheet

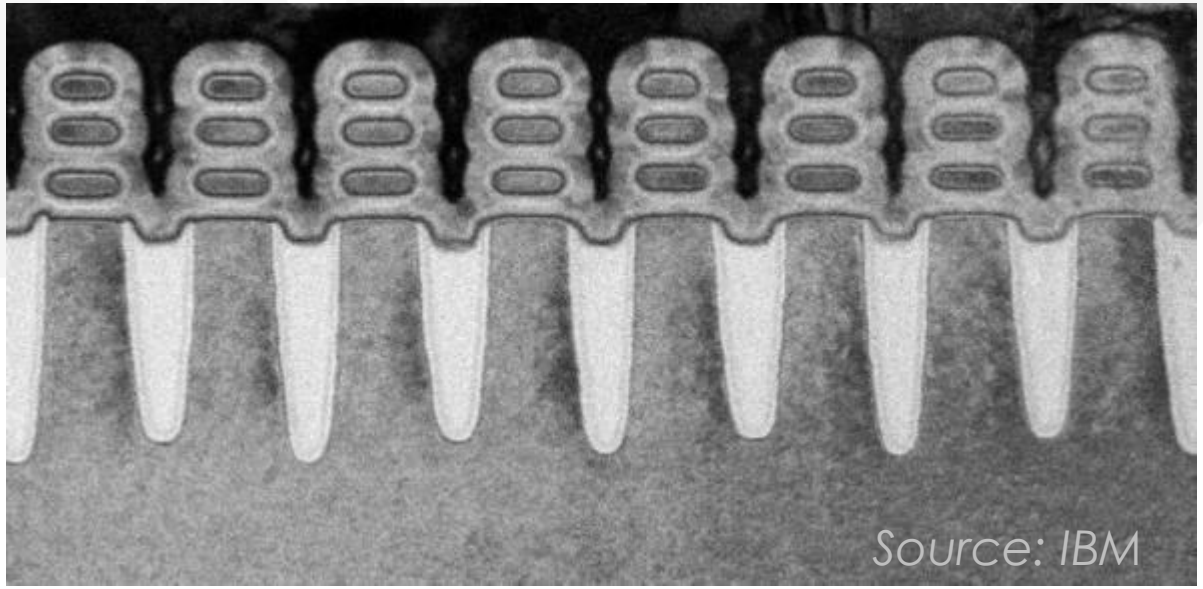
Forksheet

Complementary FET (CFET)





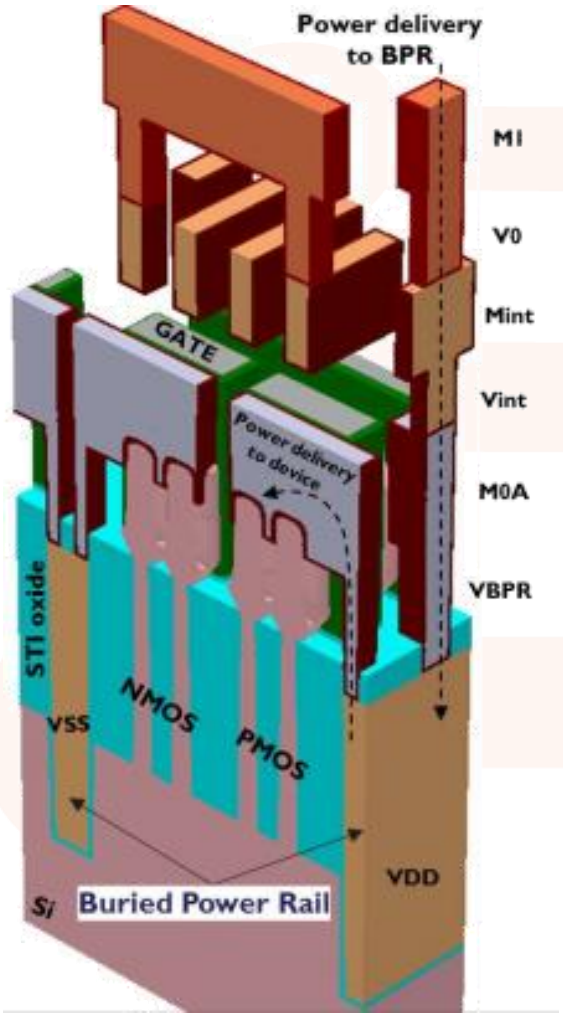




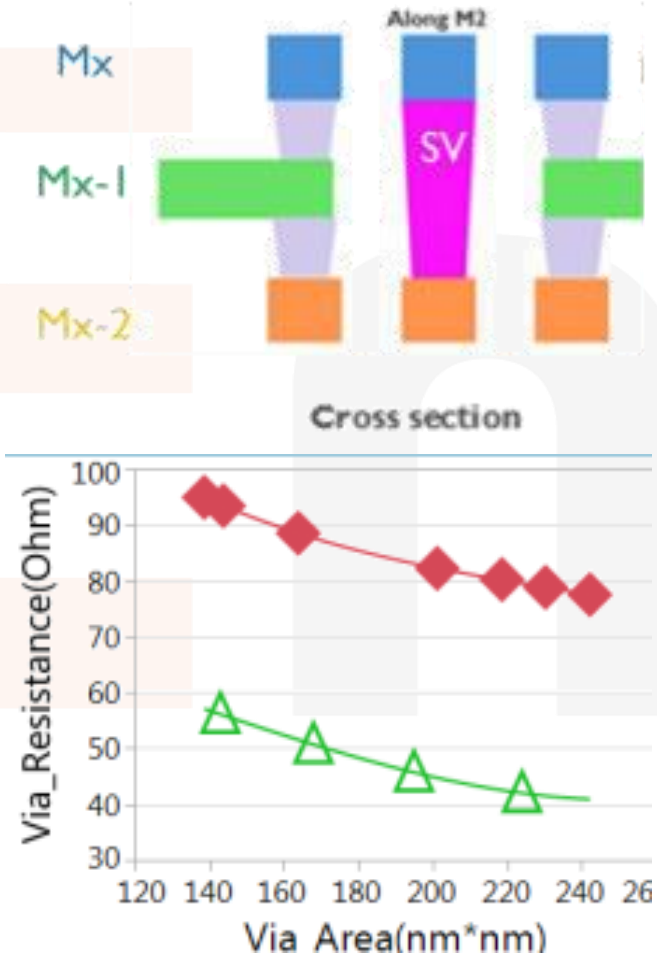
Source: imec

Gate-All-Around NanoSheets

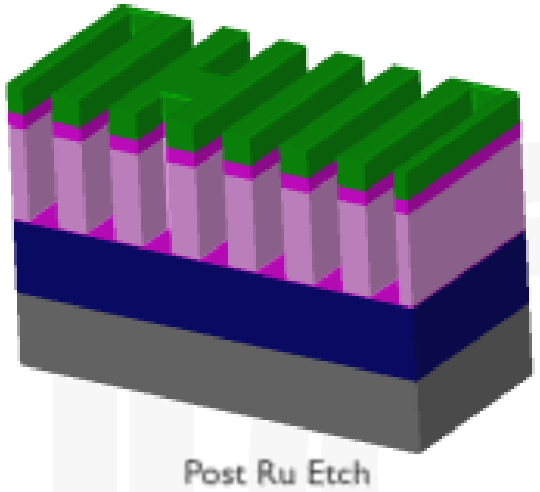
Backend Technology Innovations

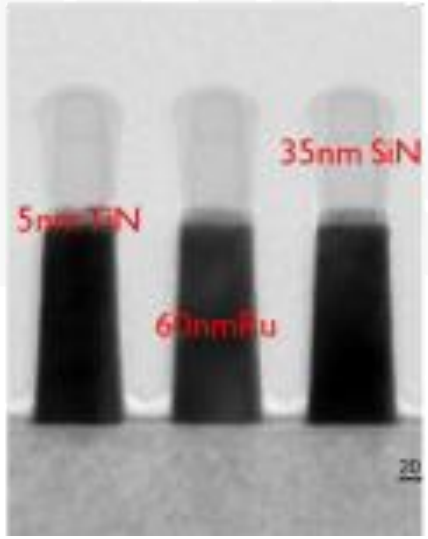


Buried Power Rails (BPR)

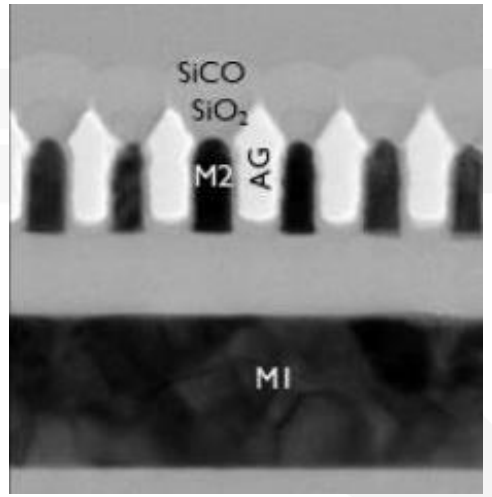


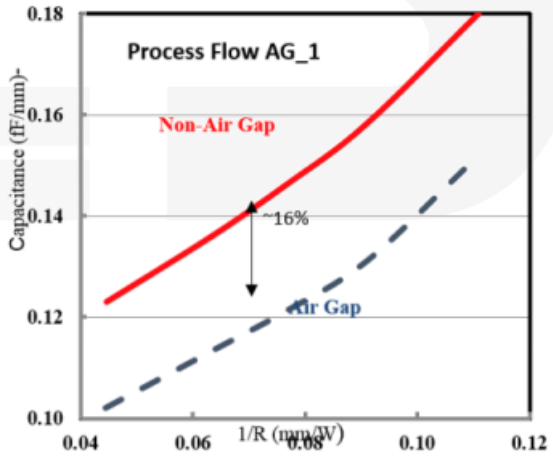
Super Vias





Barrier-less Ru Metalization



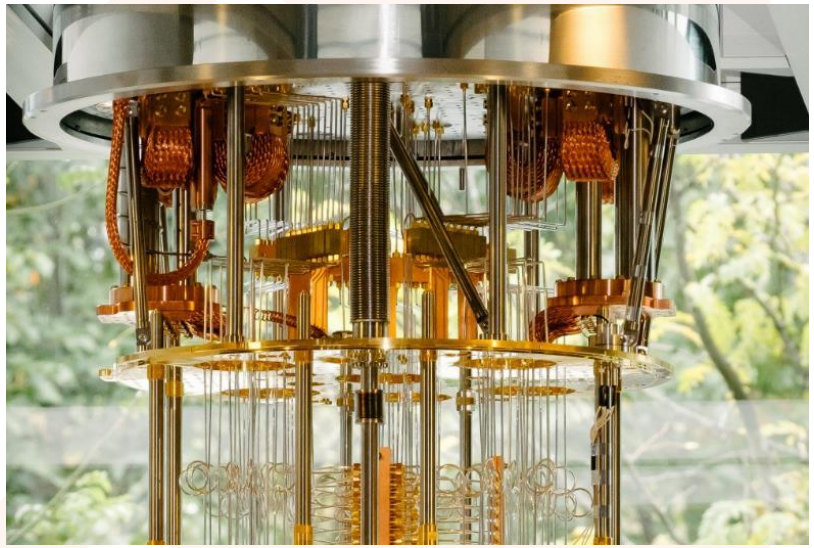


Air Gap ILD

Source: imec

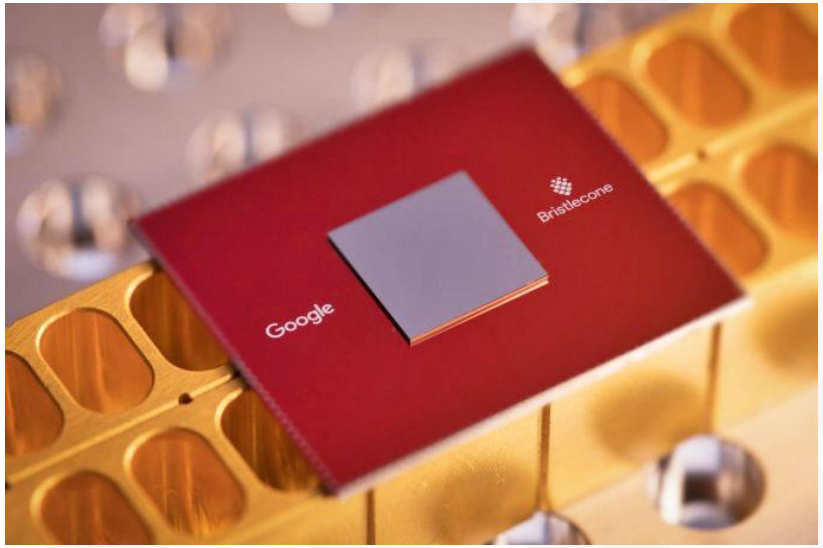
Is Quantum Computing Finally Here?

- **YES!!!!**



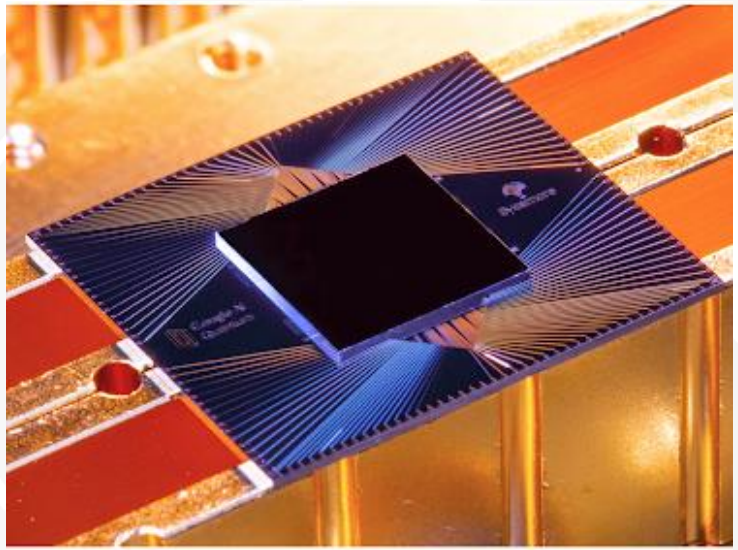
IBM Q System– 50 qubits
November, 2017

MIT Technology Review
IBM.com



Google "Bristlecone" – 72 qubits
March 5, 2018

research.googleblog.com



Google "Sycamore" –
Quantum Supremacy?
October 23, 2019

ai.googleblog.com

Further Reading

- J. Rabaey, “Digital Integrated Circuits” 2003, Chapters 1.3, 2.5, 3.3-3.5
- Weste, Harris “CMOS VLSI Design”, Chapter 7
- C. Hu, “Modern Semiconductor Devices for Integrated Circuits”, 2010, Chapters 4-7 http://www.eecs.berkeley.edu/~hu/Book-Chapters-and-Lecture-Slides-download.html
- Tzividis, et al. “Operation and Modeling of MOS Transistor” New York Press 2011. Chapters 1-5
- E. Alon, Berkeley EE-141, Lecture 2, 9 (Fall 2009) http://bwrc.eecs.berkeley.edu/classes/icdesign/ee141_f09/
- M. Alam, Purdue ECE-606 – lectures 32-38 (2009) http://nanohub.org/resources/5749
- A. B. Bhattacharyya “Compact MOSFET models for VLSI design”, 2009,
- T. Sakurai, “Alpha Power-Law MOS Model” – JSSC Newsletter Oct 2004
- Managing Process Variation in Intel’s 45nm CMOS Technology, Intel Technology Journal, Volume 12, Issue 2, 2008
- Berkeley “BSIM 4.6.4 User’s Manual” http://www-device.eecs.berkeley.edu/~bsim/BSIM4/BSIM464/BSIM464_Manual.pdf