ELEC 402

Introduction to VLSI Systems
Lecture 1

Reza Molavi
Dept. of ECE
University of British Columbia
reza@ece.ubc.ca

Slides Courtesy : Prof. Sudip Shekar, Dr. Res Saleh (UBC)

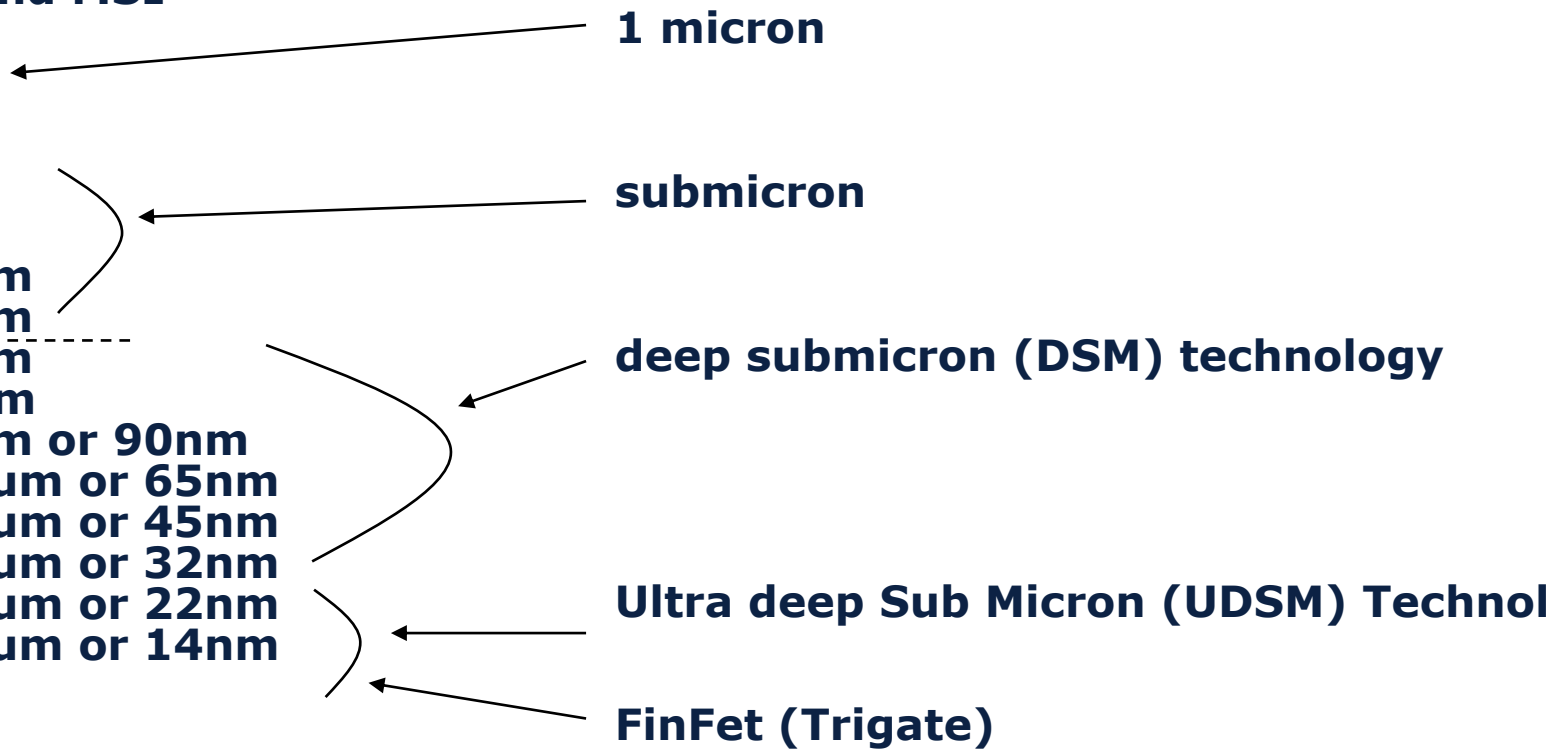
Very Large Scale

- In 2016, ~ 650 billion billion (6.5×10^{20}) transistors were made.
- Every second 20 trillion transistors.
- 60X # of stars in the Milky Way
- 170X # of galaxies in the known universe

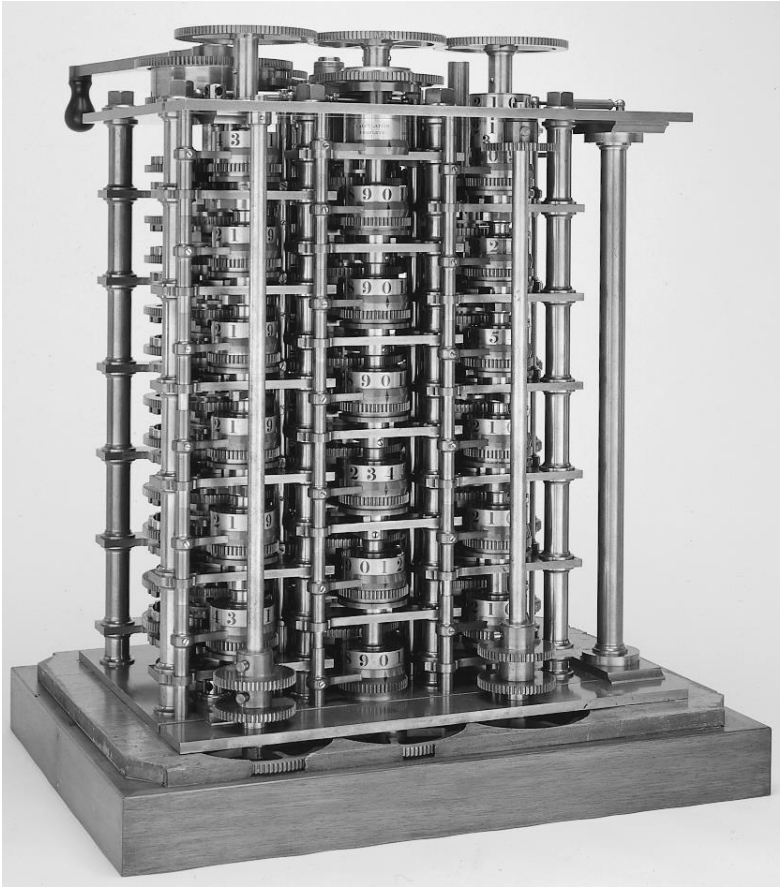
Source: VLSI Research, Apr 2017

Historical Perspective

Transistor Invented - 1940's
Integrated Circuit Invented - 1960's
'60's - SSI and MSI
'70's - LSI
'80's - VLSI
'90's - 1um
0.8um
0.6um
0.5um
0.35um
0.25um
0.18um
'00's - 0.13um
0.09um or 90nm
0.065um or 65nm
0.045um or 45nm
0.032um or 32nm
0.022um or 22nm
0.014um or 14nm

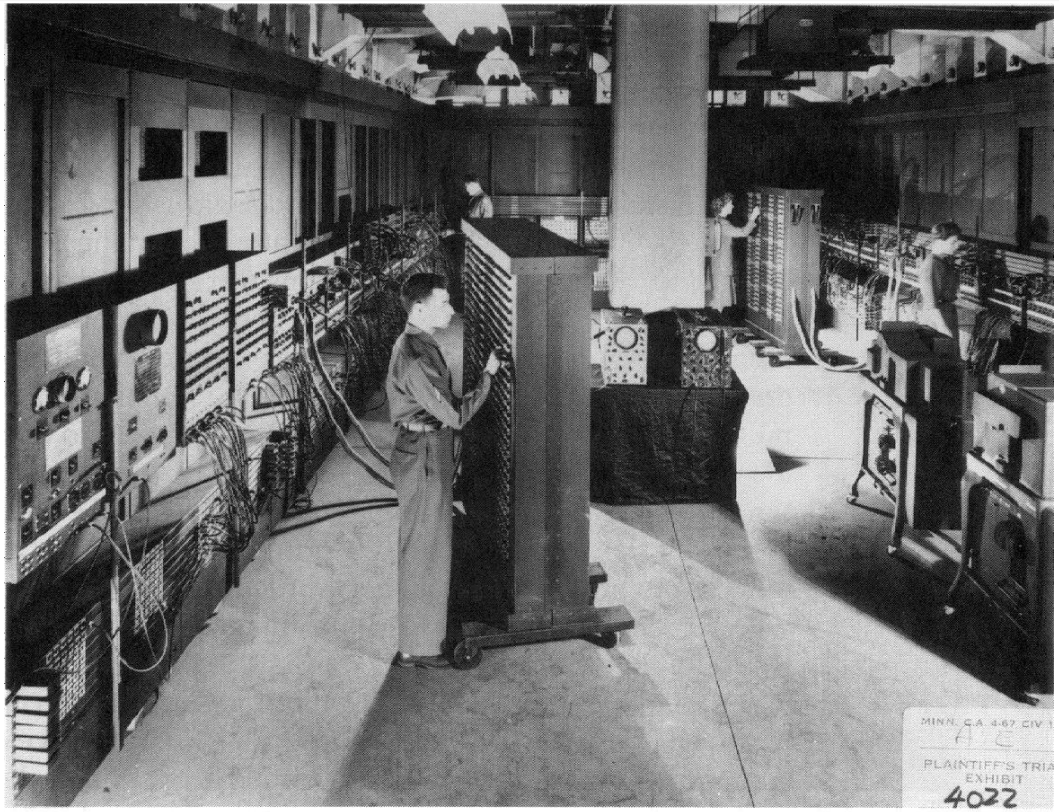


The 1st Computer: Babbage Difference Engine (1832)



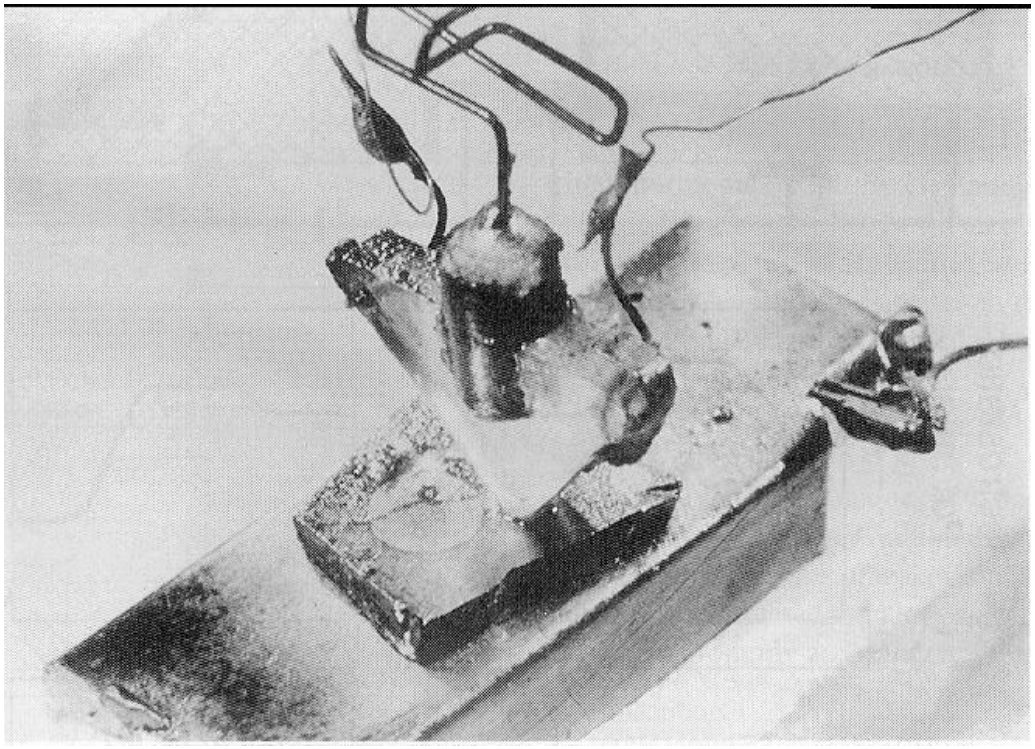
- Arithmetic operations (+-x/)
- 2-cycle sequence operation (store, execute)
- Pipelining to speed up
- 25000 parts, £17470

The 1st Electronic Computer: ENIAC (1946)



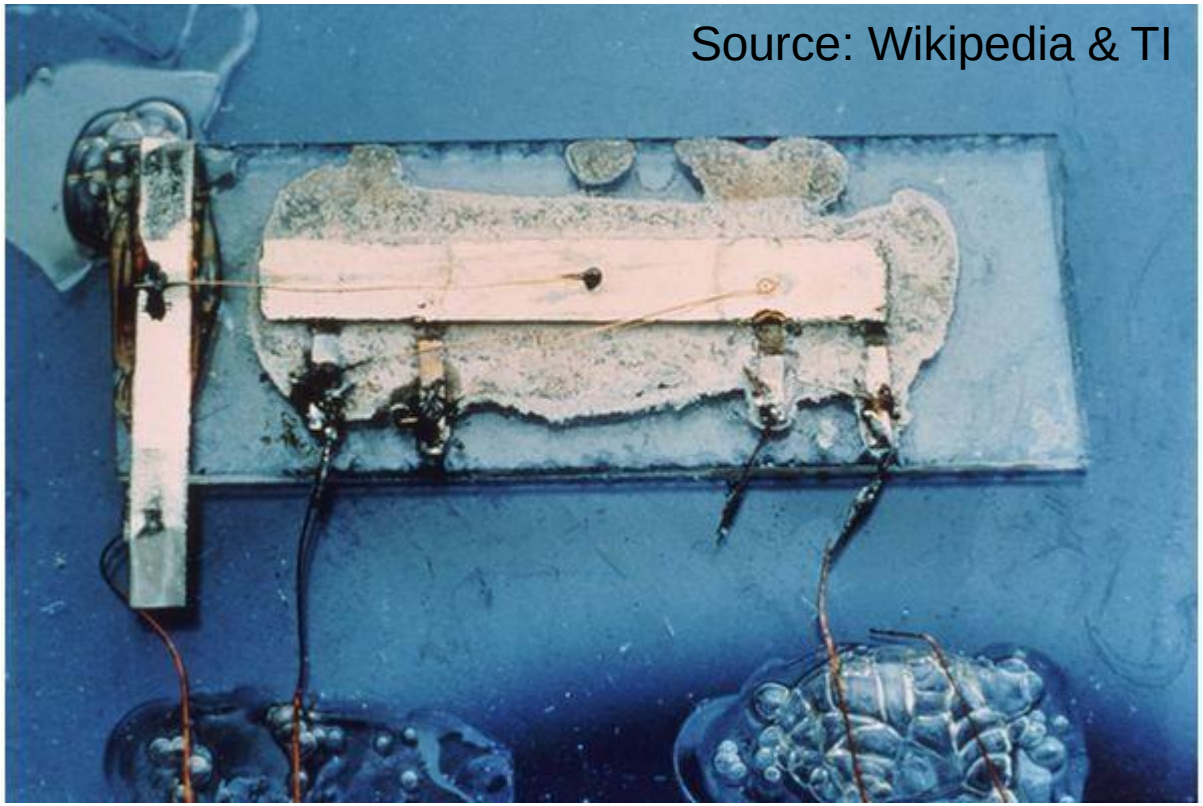
- Electronic Numerical Integrator And Computer
- 17468 Vacuum tubes
- 100’ x 8.5’ x 3’

The 1st Transistor: Bell Labs (1948)



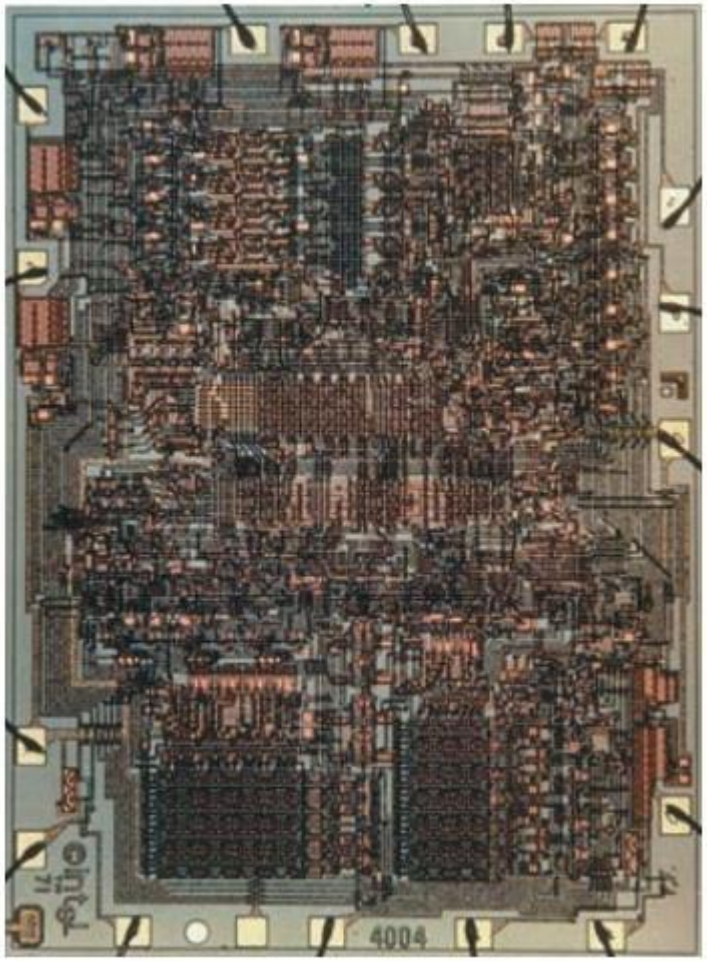
- J. Edgar Lilienfeld filed the first patent on Field-Effect Transistor (FET) in 1925
- Bardeen, Brattain and Shockley made a working prototype of Ge transistor in 1948.

The 1st Integrated Circuit (IC): Jack Kilby (1958)

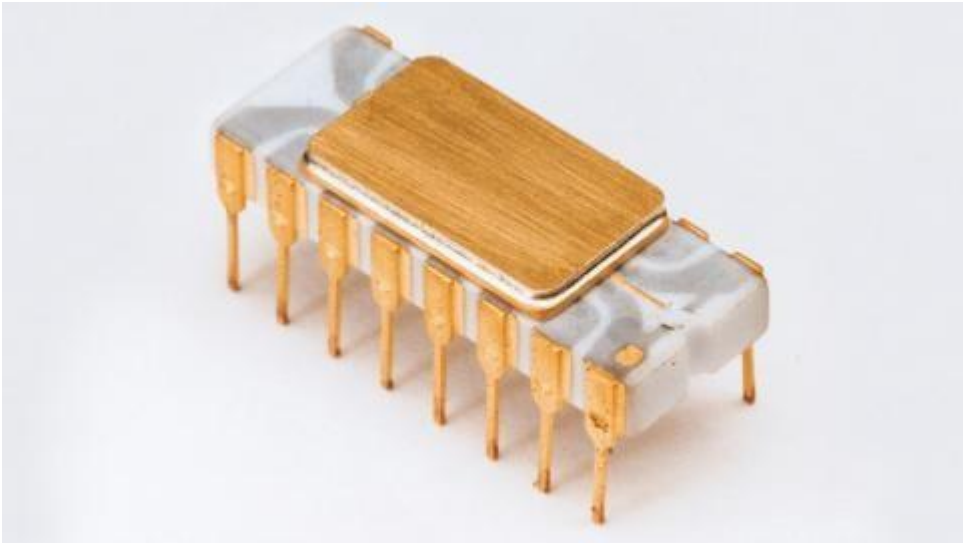


- **Jack Kilby at TI showed the first IC in 1958**
- **Robert Noyce developed another *better* IC within 6 months. Co-founds Fairchild Semiconductor and Intel – gave Silicon Valley its name**

The 1st μProcessor: Intel 4004 (1971)



Source: Intel



1971
Intel® 4004 processor

Initial clock speed:
108KHz

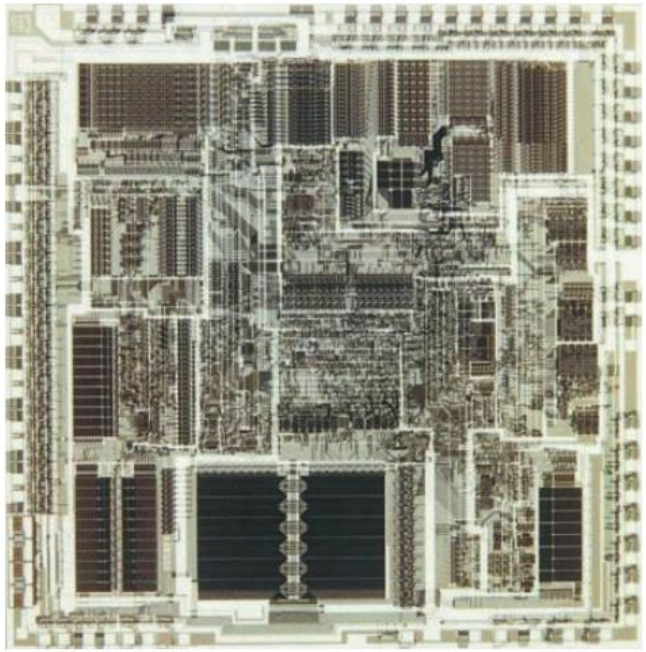
Transistors:
2,300

Manufacturing technology:
10 micron

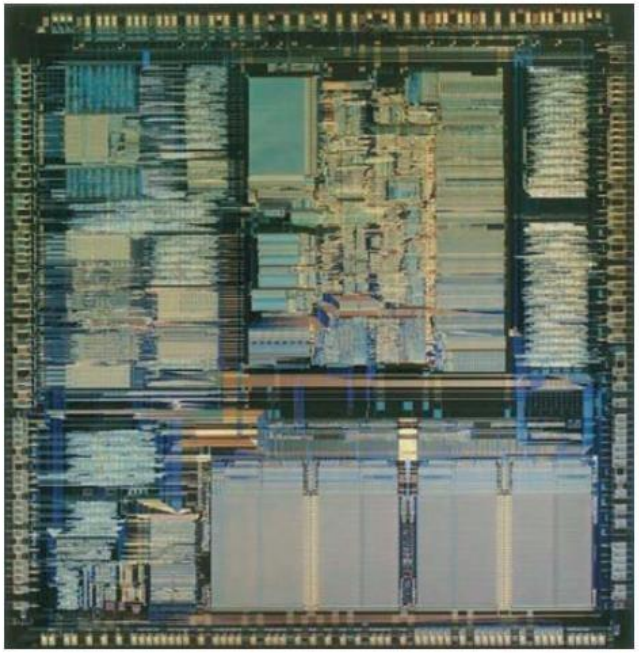
Source: Intel

Exponential Growth: 1982-1989

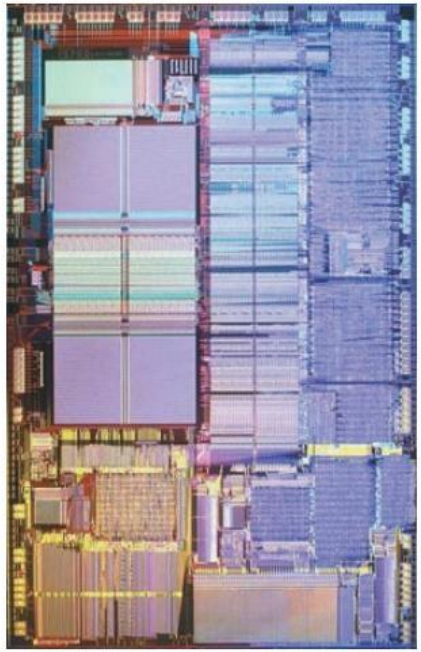
Source: Intel



5



6



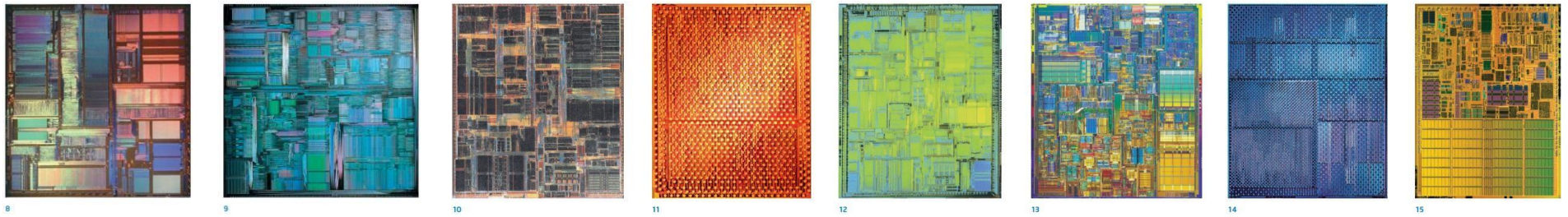
7

5	6	7
1982 Intel® 286™ processor	**1985 Intel386™ processor**	**1989 Intel486™ processor**
Initial clock speed: 6MHz	Initial clock speed: 16MHz	Initial clock speed: 25MHz
Transistors: 134,000	Transistors: 275,000	Transistors: 1.2 million
Manufacturing technology: 1.5 micron	Manufacturing technology: 1.5 micron	Manufacturing technology: 1 micron

Source: Intel

Exponential Growth: 1993-2003

Source: Intel



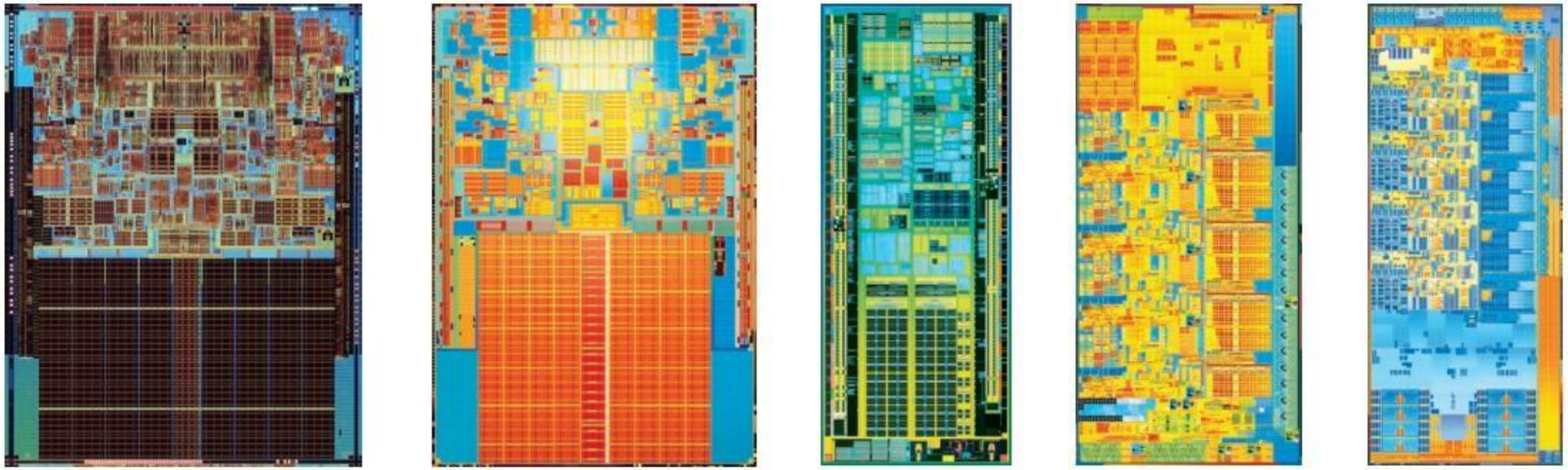
8	9	10	11	12	13	14	15
1993 **Intel® Pentium® processor**	**1995** **Intel® Pentium® Pro processor**	**1997** **Intel® Pentium® II processor**	**1998** **Intel® Celeron® processor**	**1999** **Intel® Pentium® III processor**	**2000** **Intel® Pentium® 4 processor**	**2001** **Intel® Xeon® processor**	**2003** **Intel® Pentium® M processor**
Initial clock speed: 66MHz	Initial clock speed: 200MHz	Initial clock speed: 300MHz	Initial clock speed: 266MHz	Initial clock speed: 600MHz	Initial clock speed: 1.5GHz	Initial clock speed: 1.7GHz	Initial clock speed: 1.7GHz
Transistors: 3.1 million	Transistors: 5.5 million	Transistors: 7.5 million	Transistors: 7.5 million	Transistors: 9.5 million	Transistors: 42 million	Transistors: 42 million	Transistors: 55 million
Manufacturing technology: 0.8 micron	Manufacturing technology: 0.35 micron	Manufacturing technology: 0.25 micron	Manufacturing technology: 0.25 micron	Manufacturing technology: 0.25 micron	Manufacturing technology: 0.18 micron	Manufacturing technology: 0.18 micron	Manufacturing technology: 90nm

Source: Intel

Exponential Growth: 2006-2014

Source: Intel

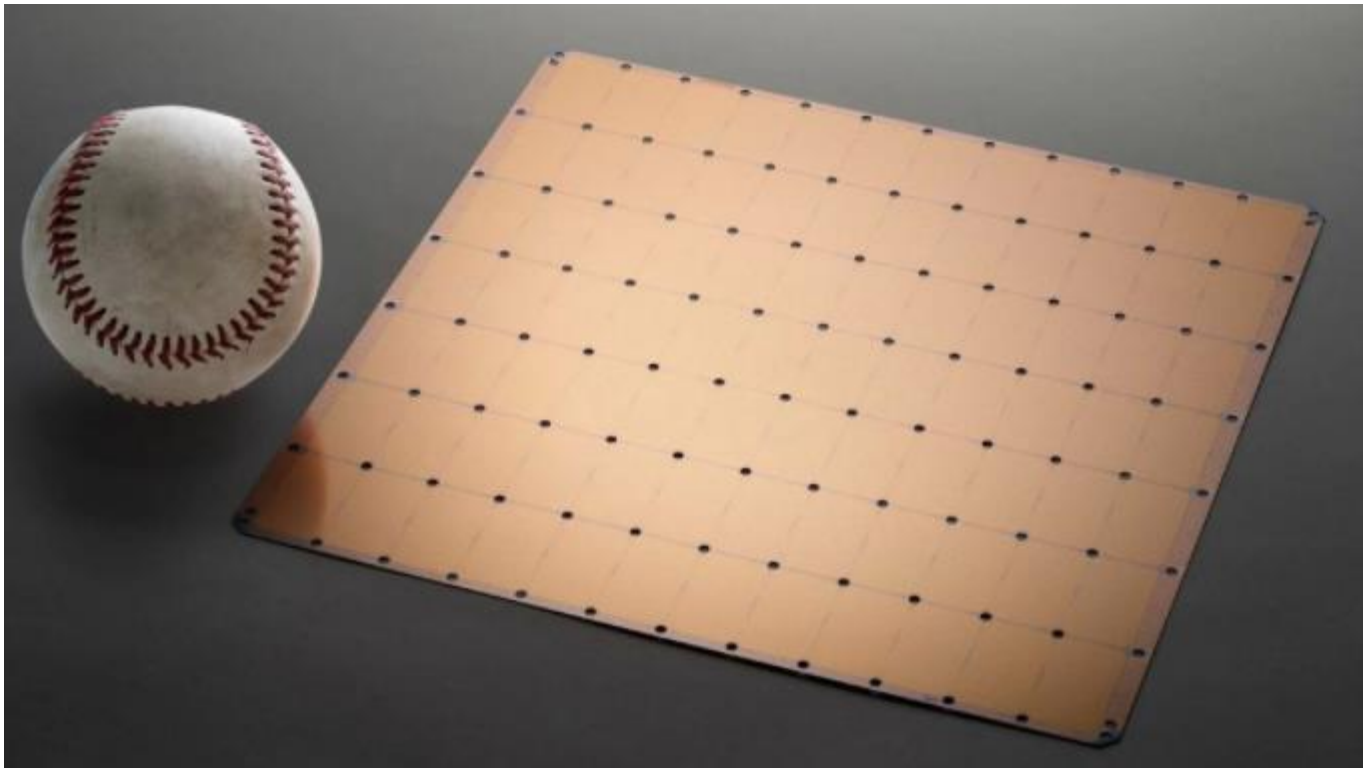
High-k Metal Gate **3D Tri-Gate**



16	17	18	19	20
2006 Intel® Core™2 Duo processor	2008 Intel® Core™2 Duo processor	2008 Intel® Atom™ processor	2010 2nd generation Intel® Core™ processor	2012 3rd generation Intel® Core™ processor
Initial clock speed: 2.66GHz	Initial clock speed: 2.4GHz	Initial clock speed: 1.86GHz	Initial clock speed: 3.8GHz	Initial clock speed: 2.9GHz
Transistors: 291 million	Transistors: 410 million	Transistors: 47 million	Transistors: 1.16 billion	Transistors: 1.4 billion
Manufacturing technology: 65nm	Manufacturing technology: 45nm	Manufacturing technology: 45nm	Manufacturing technology: 32nm	Manufacturing technology: 22nm

Source: Intel

First Trillion Transistor processor – Aug 2019

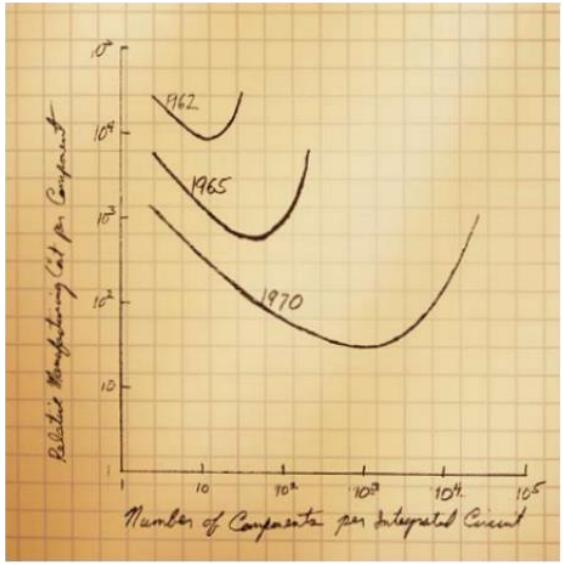


- **1.2 Trillion Transistors, 400,000 AI optimized Cores, 46000 mm^2,**
- **ASIC design for ML/AI Applications**
- **Regularity allows dense implementation**

Source: Intel

Moore's Law

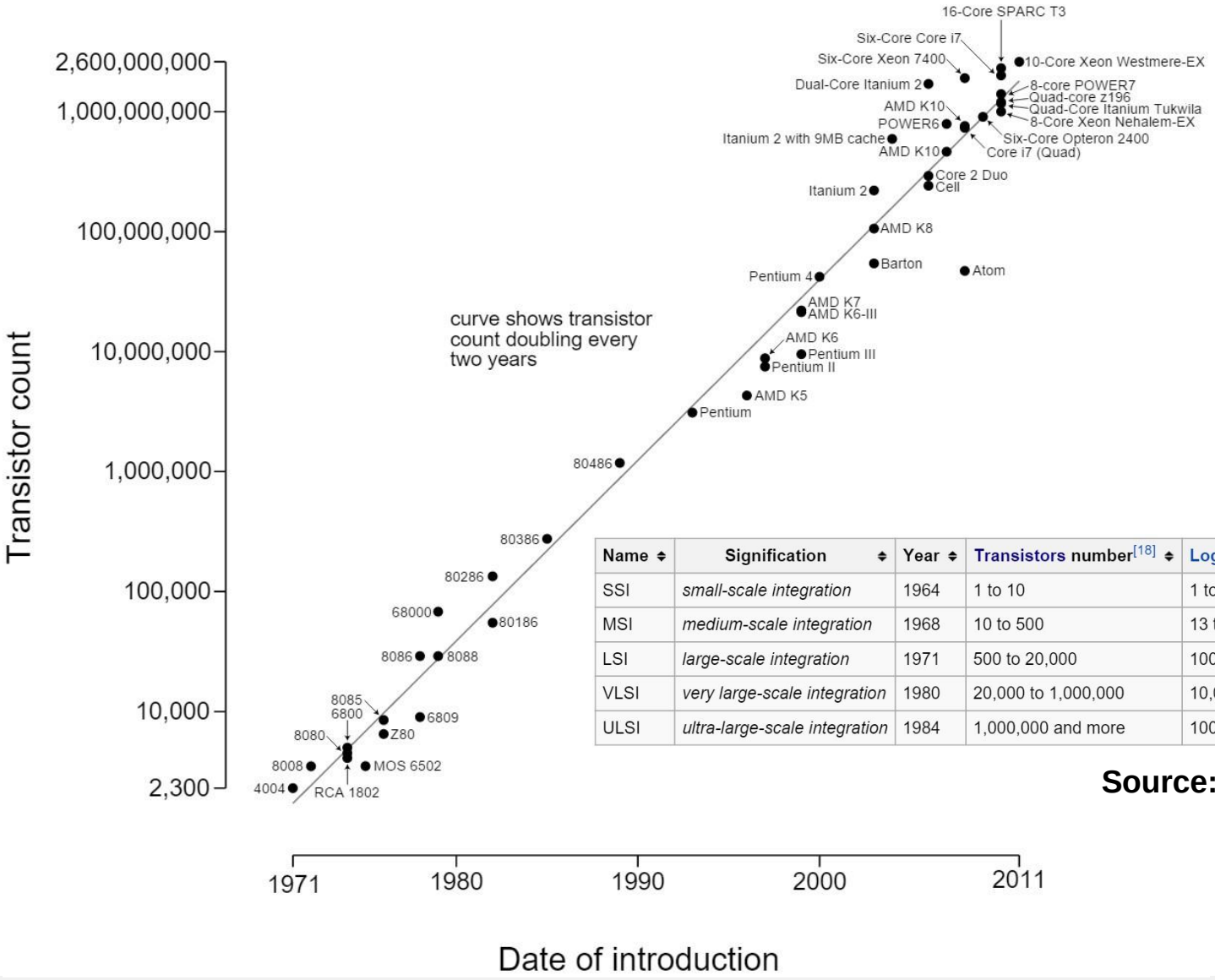
In 1965, Gordon Moore predicted that the number of transistors on a piece of silicon would double every year—an insight later dubbed "Moore's Law." Intended as a rule of thumb, it has become the guiding principle for the industry to deliver ever more powerful semiconductor chips at proportionate decreases in cost. In 1975, Moore updated his prediction that the number of transistors that the industry would be able to place on a computer chip would double every couple of years. The original Moore's Law graph is shown here.



Moore's Law has been amazingly accurate over time. In 1971, the Intel 4004 processor held 2,300 transistors. In 2008, the Intel® Core™2 Duo processor holds 410 million transistors. A dramatic reduction in cost has also occurred. In 1965, a single transistor cost more than one dollar. By 1975, the cost was reduced to one cent, and today Intel can manufacture transistors that sell for less than 1/10,000 of a cent each.
ensures th
Moore's La

http://spectrum.ieee.org/geek-life/history/moores-law-milestones **Source: Intel**

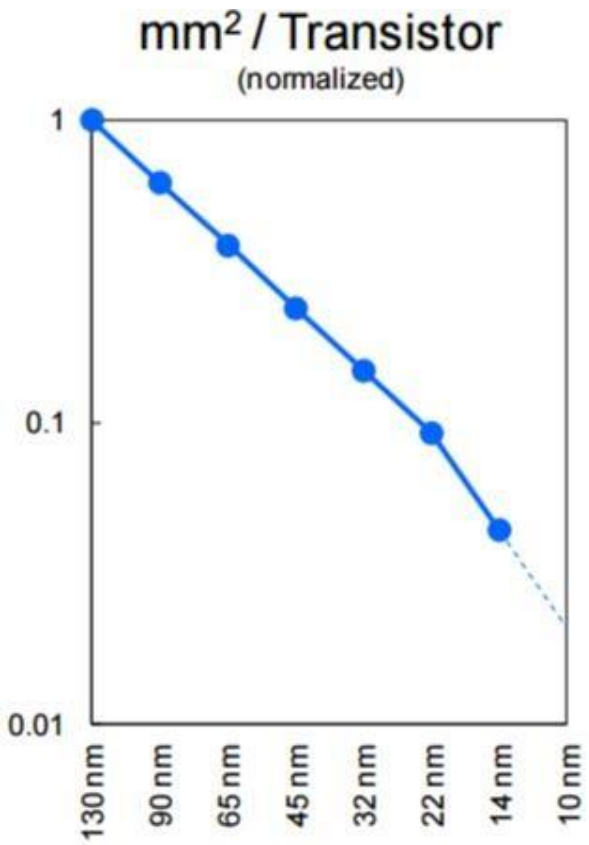
Moore's Law & μProcessor Transistor Counts

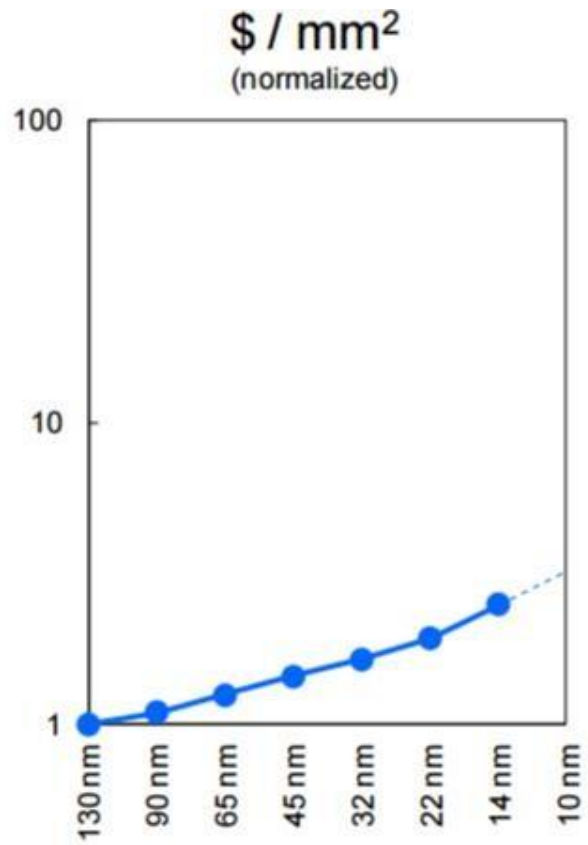


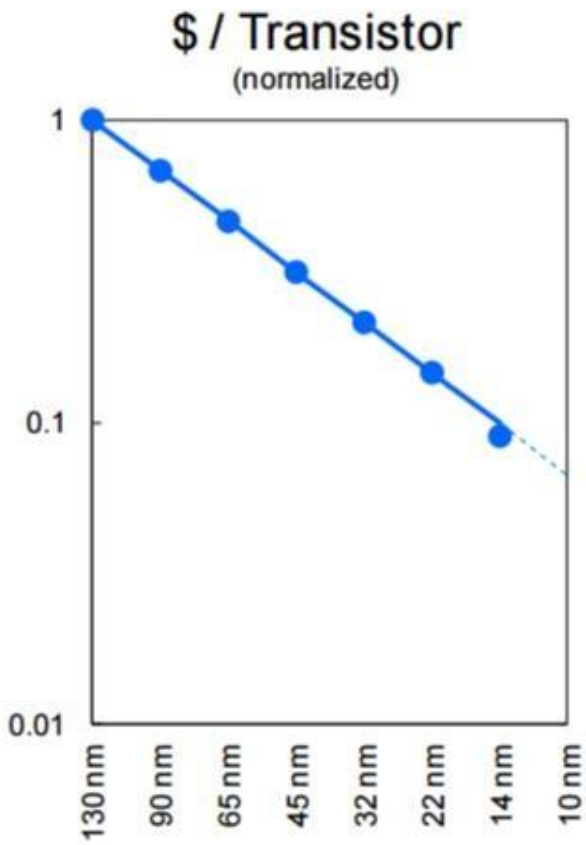
Name	Signification	Year	Transistors number[18]	Logic gates number[19]
SSI	*small-scale integration*	1964	1 to 10	1 to 12
MSI	*medium-scale integration*	1968	10 to 500	13 to 99
LSI	*large-scale integration*	1971	500 to 20,000	100 to 9,999
VLSI	*very large-scale integration*	1980	20,000 to 1,000,000	10,000 to 99,999
ULSI	*ultra-large-scale integration*	1984	1,000,000 and more	100,000 and more

Source: Wikipedia

Moore's Law & Cost/Transistor Scaling



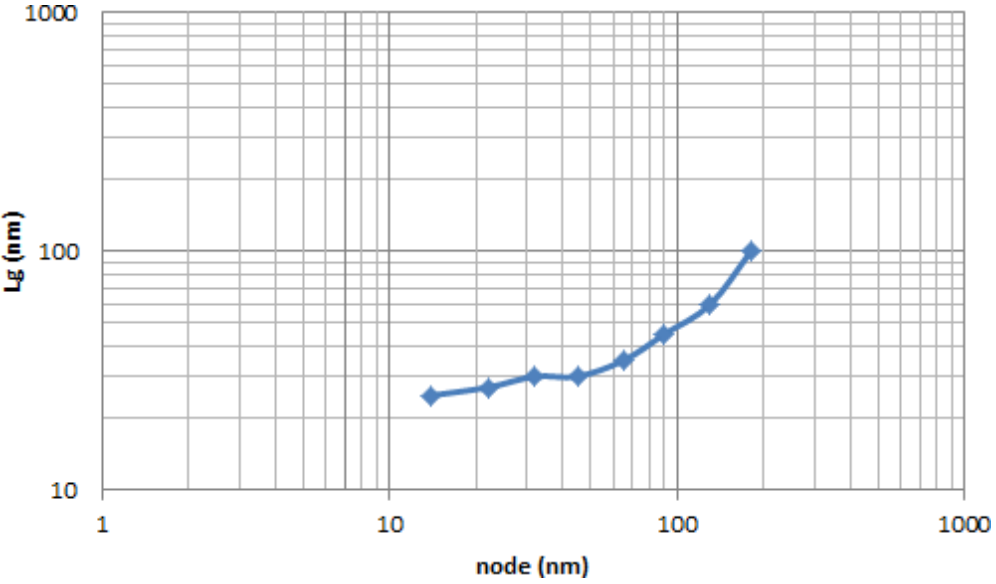




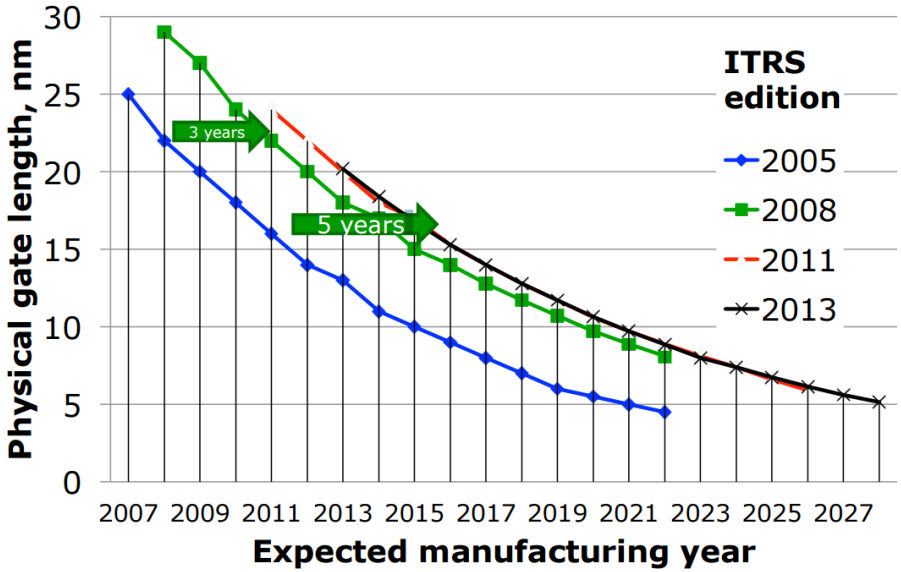
- **Area/transistor is scaling down, with even better-than-normal at 14 nm using advanced double patterning techniques**
- **However, wafer cost increasing due to added masking steps**
- **For now, lower cost/transistor continues**

Source: Intel

Moore's Law & Transistor Gate Length Scaling?



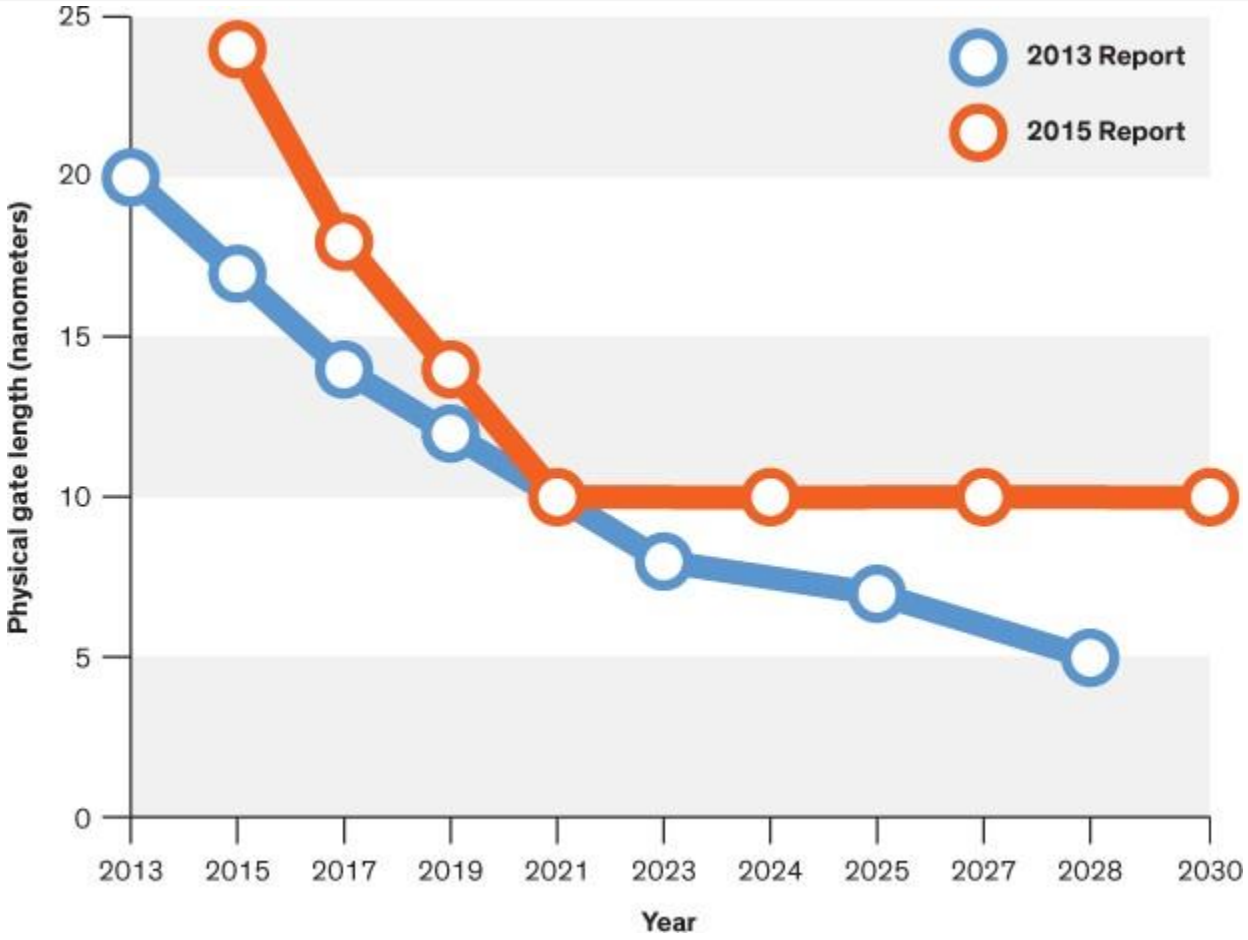
Source: Wikipedia



Source: W. Gropp

- Transistor gate length trend points to a slow-down
- ITRS projections adjusted by 3 and then 5 years.
- Approaching the size of atoms which is a fundamental barrier

End of Moore's Law?



- **Final ITRS projection – Moore's law to end near 2021**
- **Some further scaling may be possible with 3D transistors**

http://spectrum.ieee.org/semiconductors/devices/transistors-could-stop-shrinking-in-2021

Dennard's Scaling

Scaling Results for Circuit Performance

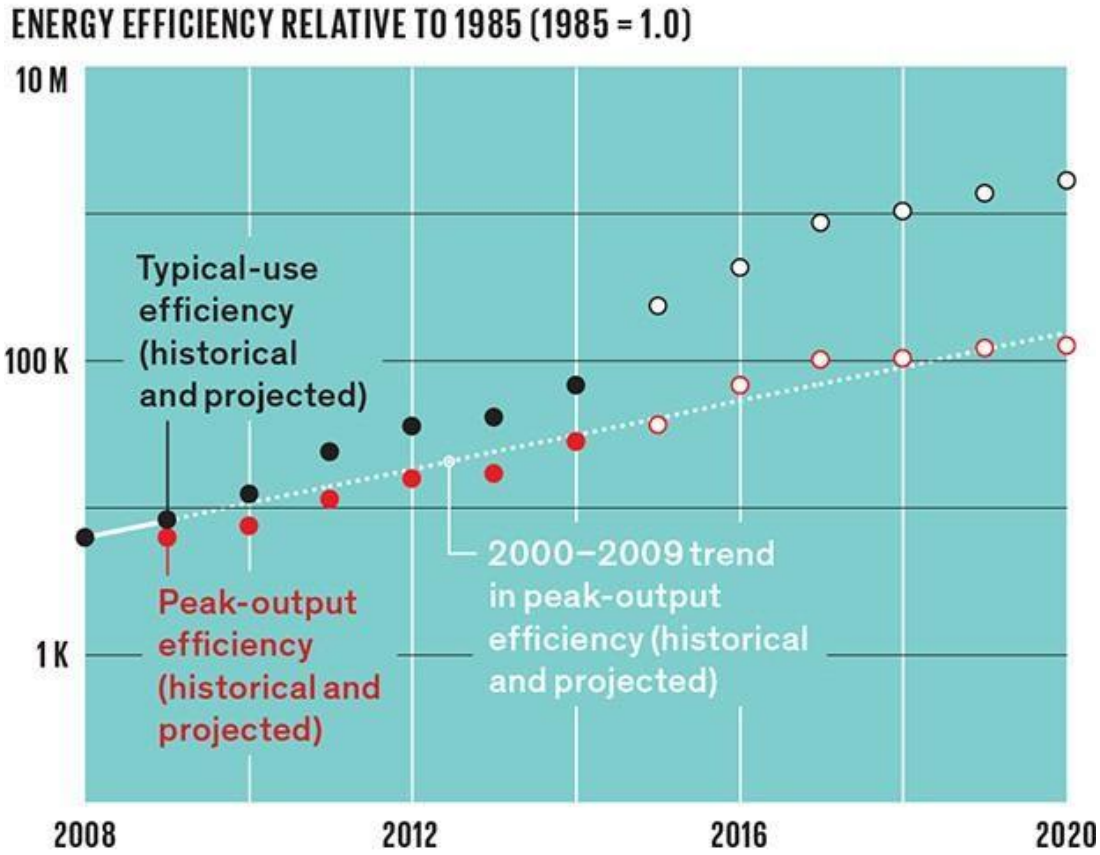
Device or Circuit Parameter	Scaling Factor
Device dimension t_{ox}, L, W	$1/\kappa$
Doping concentration N_a	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\epsilon A/t$	$1/\kappa$
Delay time/circuit VC/I	$1/\kappa$
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1

- **As transistors shrink in dimension, both voltage and current scale down as well**
- **Power density stays constant → Power dissipation proportional to area**
- **Dennard Scaling & Moore's law → Performance/Watt improves**

Read section IX of

R Dennard et al, "Design of ion-implanted MOSFET's with very small physical dimensions", *IEEE J. Solid-State Circuits*, Oct. 1974

Moore's Law & Energy Efficiency

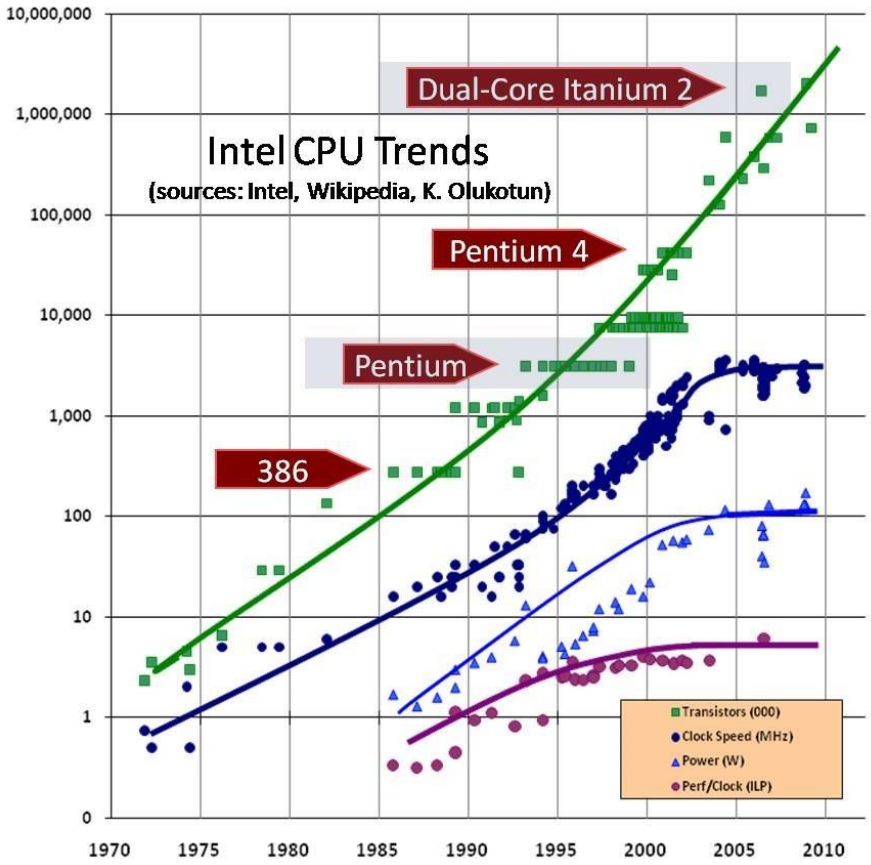


Peak-output efficiency: # computations that can an be performed/ kW-hr of electricity consumed.

Typical-use efficiency: # computations over a year/total electricity consumed—a weighted sum of the energy a processor and its supporting circuitry use in different modes over that same period.

Source: IEEE Spectrum, Mar 2015

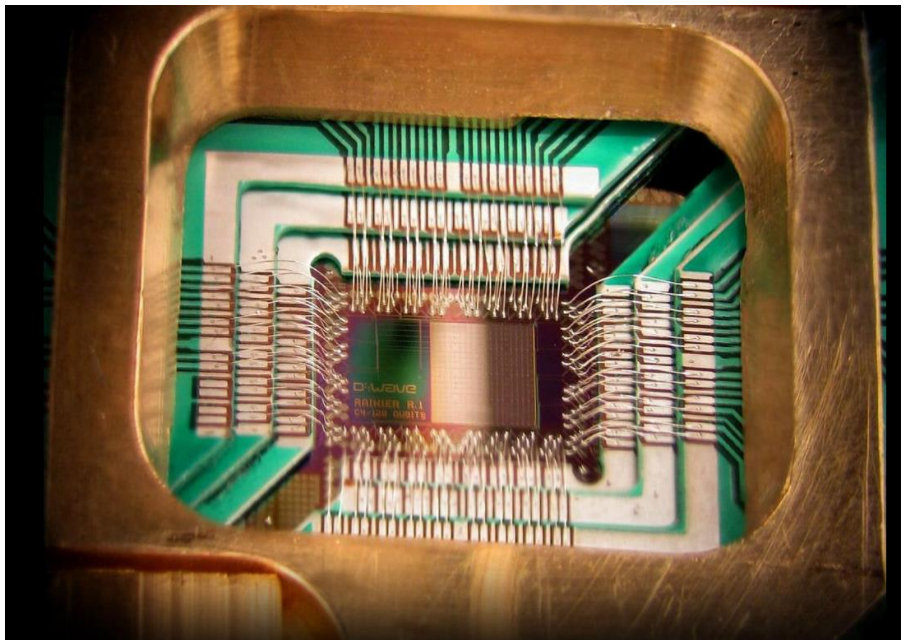
Moore's Law & Dennard Scaling



Source: W. Gropp

- Clock-speed has slowed down.
- Current leakage and VTH/VDD bigger issue at smaller nodes, increasing power consumption and creating a power wall for processors. (...later)
- Industry moved towards multicore processors.

New Computing Paradigm – Quantum Computing



- The limitation of our current technology is that it assesses one combination of bits at a time resulting in excess delays for complex calculations

- Quantum computing has emerged as an alternative to assess all combinations at the same time using quantum physics properties such as superposition and tunneling

- Superconducting integrated circuits can generate qubits, D-Wave has 2000 qubits processors now, Google, IBM, and Microsoft efforts on the way

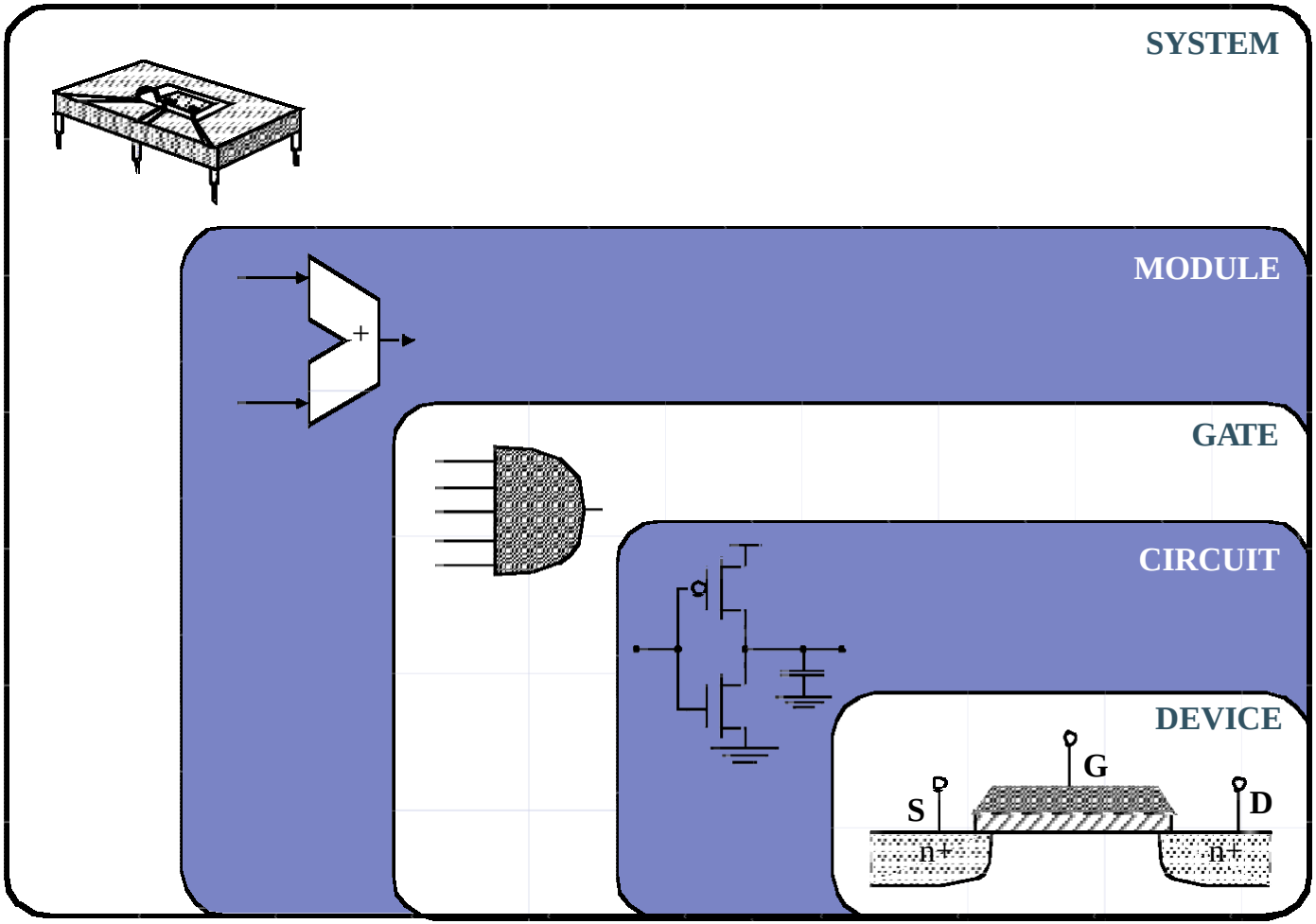
Deep Submicron Characteristics

What's so special about deep submicron (DSM)?

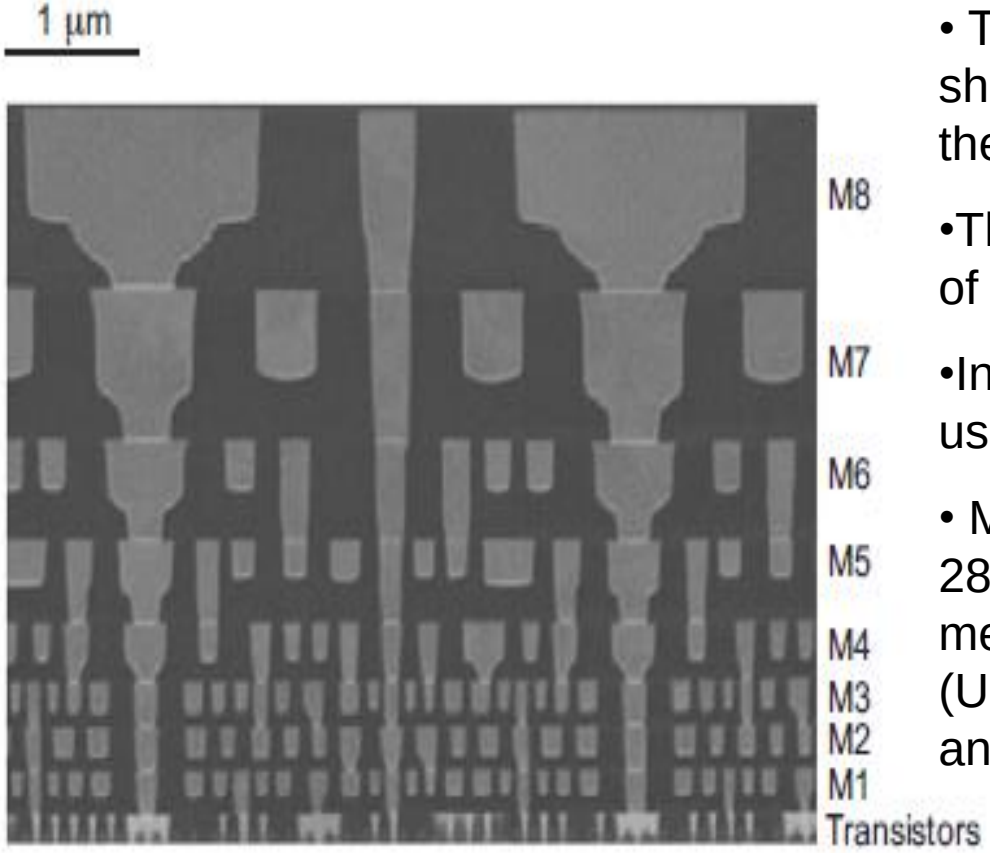
- MOS device behavior is much more complex (velocity saturation)
- Wires become as important as devices (in some ways even more important).

DSM Devices	DSM Wires
short-channel effects on V_T	interconnect RC delays
velocity saturation	IR drop + Ldi/dt
thin-oxide (tunneling/breakdown)	capacitive coupling
subthreshold current	inductive coupling
DIBL	electromigration
hot-carrier effects	antenna effects

Digital VLSI Design Abstraction Levels



DSM Process

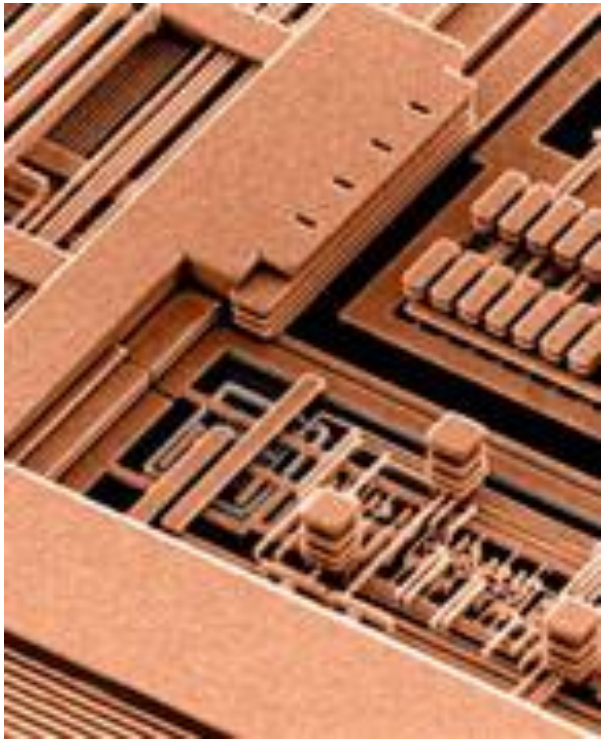


- This is an SEM of Intel's 45 nm process showing 8 layers of metal (copper) and the corresponding vias.
- The figure demonstrates the importance of interconnect in deep submicron
- In this process, the transistors are usually dwarfed by the interconnect
- More recent technology nodes, e.g. 28nm CMOS, possess more than 10 metal layers with an option for Thick (Ultrathick) Metal layers and MiM caps and metal/vias are made of copper

What is this course about?

This course is about at the details of

System Level Models
MOS Models
Combination/Sequential Circuit Design
Interconnect Models
Area/Timing/Power tradeoffs
Memory Design
Performance Optimization



IBM Copper interconnect

Fundamentally, I want you to:

"Think like an IC designer"

Need to develop models for design and analysis; models are approximations of the real world; we need to understand the type of approximations being made and where they breakdown.

Models

Models are an approximation of the real world

Must leave many details out
Must (to be useful) retain the important details
Appropriate level depends on questions you want to answer

CAUTION:

Simulation and analysis do not tell you what the circuit does
It tells you what your MODEL of the circuit does
So remember:
Garbage in, garbage out

Some of the hardest work is figuring out the right model for a problem

HSPICE/Cadence/

SPICE is the most widely use circuit simulator for detailed analysis of transistor level designs

It uses very accurate models so we can verify our hand analysis against the "correct" answer

The problem sets will use a version of SPICE called HSPICE. This version has a number of features (like parameter sweeps and optimization) that will make your life easier.
There is a library provided for the class

15 nm FinFET / 45 nm GPDK CMOS technology

Tutorial sessions will explain how to use the key features in HSPICE/Cadence.

What Needs to be Modeled?

Transistors

nMOS, pMOS

Wires

They are not ideal connectors

How complex should the model be?

- Resistance effects, IR drops in lines?
- Coupling, Inductance?

Circuit Environment

Temperature, Power Supply, Substrate Voltage, Chip Gnd vs. Board Gnd. We won't be spending too much time on these aspects but they are all important at the chip level!

Circuit Design Priorities

- **Functionality**
- **Timing**
- **Power**
- **Circuit Noise Tolerance**
- **Area**
- **Cost**
- **Time-to-market**
- **Supply variations**

- Testability
- Packaging
- Process Variations
- Yield
- Temperature variations
- Short/Long-term Reliability

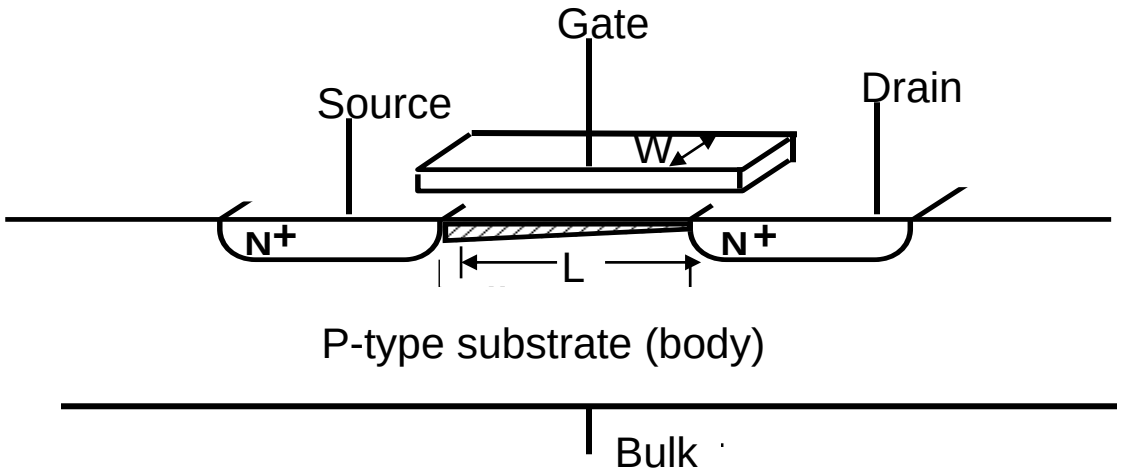
MOS Transistor Basics

Older terms: FET, IGFET, MOST, MOSFET

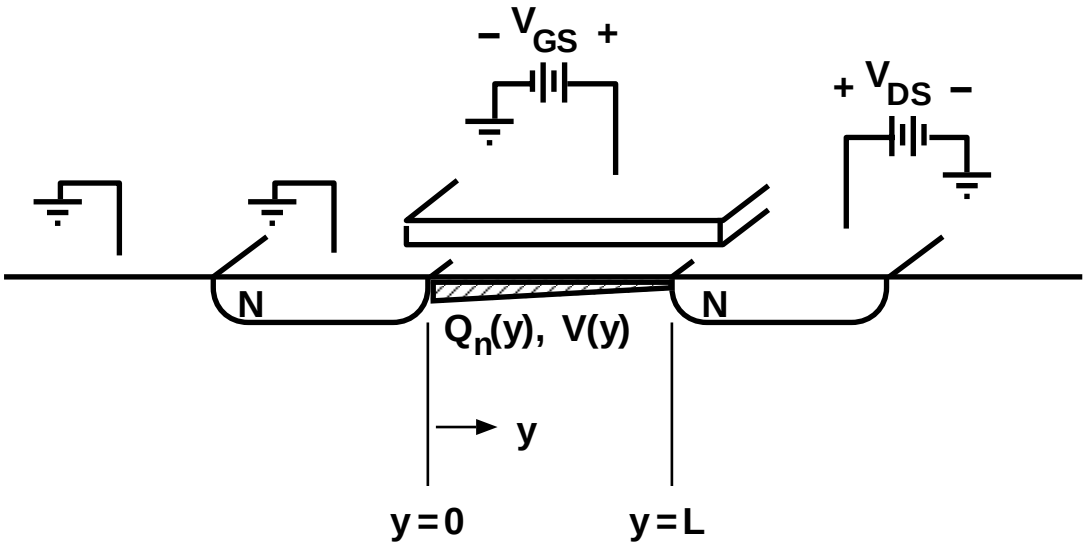
Recent terms: MOS transistor, MOS device

New terms: 3D transistors, FinFET, MG - MOS, GAA -FET

Transistor = a four terminal semiconductor device wherein current flow between two of the terminals is controlled from the third terminal.



MOS I-V Characteristics



Velocity Saturation controls device operation:

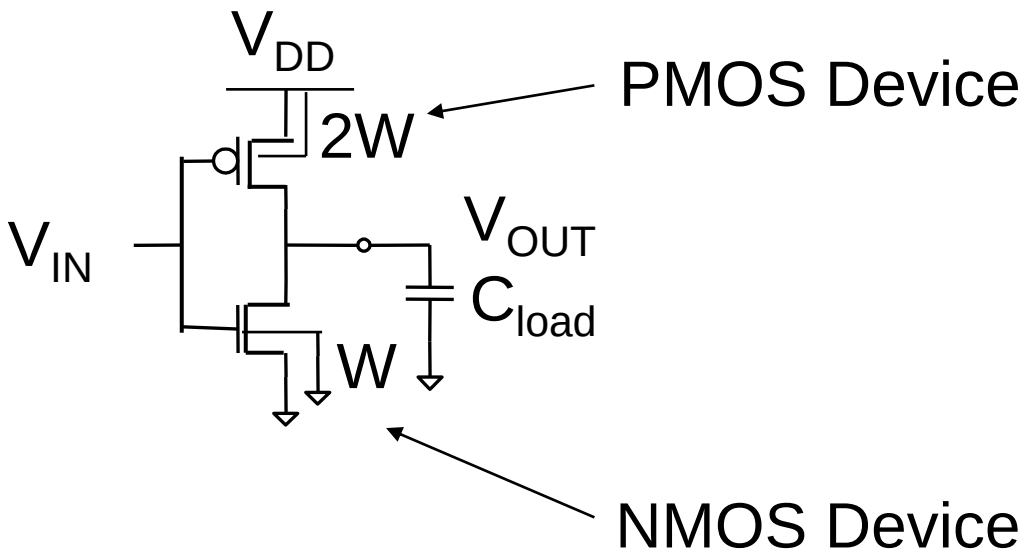
- carriers reach maximum velocity before they reach end of channel region

- sets limit on current level

$$Ids \approx W\nu_{sat} C_{ox} (V_{GS} - V_{th})$$

CMOS Inverter

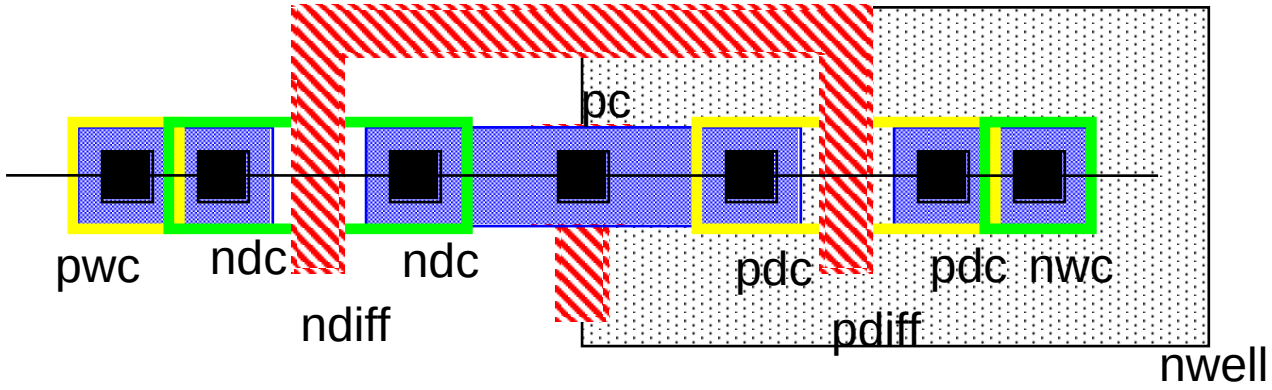
Need NMOS and PMOS device to form an inverter

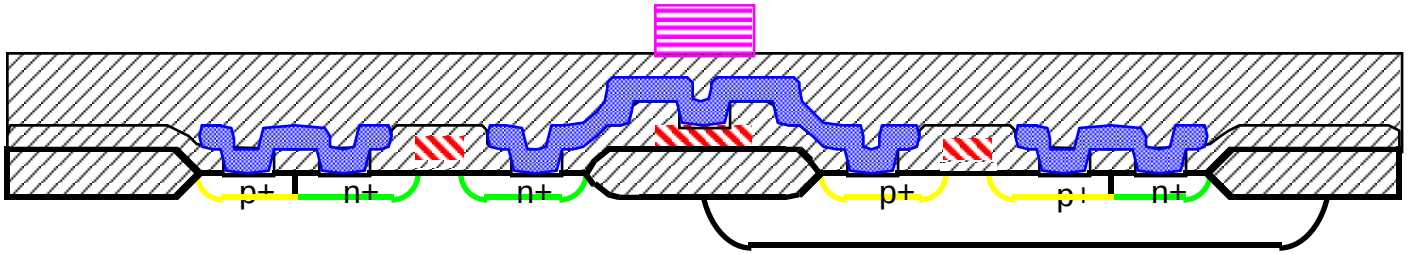


When Vin is low, the NMOS device is off and the PMOS device pulls the output to Vdd When Vin is high, the NMOS device is on and it pulls the output to Gnd

Fabrication and Layout

Layout of cell and final cross-section of inverter cell

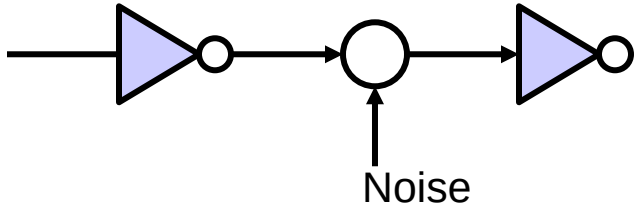




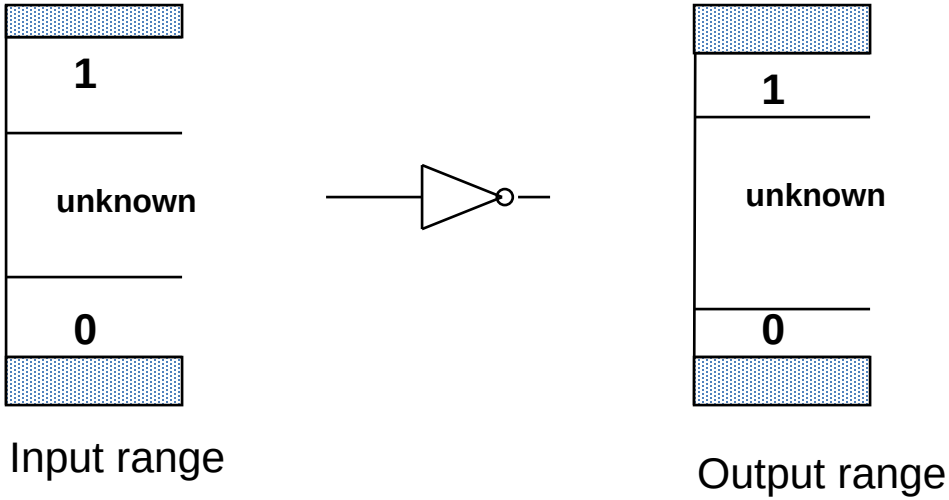
The Digital Abstraction

Signals are represented by voltages

Voltages are not fundamentally quantized
Signals will have noise



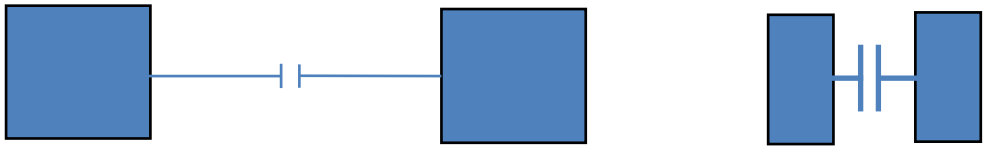
In robust systems, noise should not affect output
Divide voltage range into regions: 0, X, 1



Interconnect Scaling Effects

Dense multilayer metal increases coupling capacitance

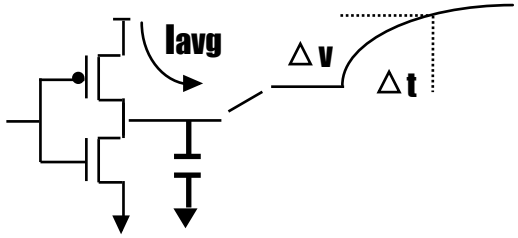
Old Assumption **DSM**



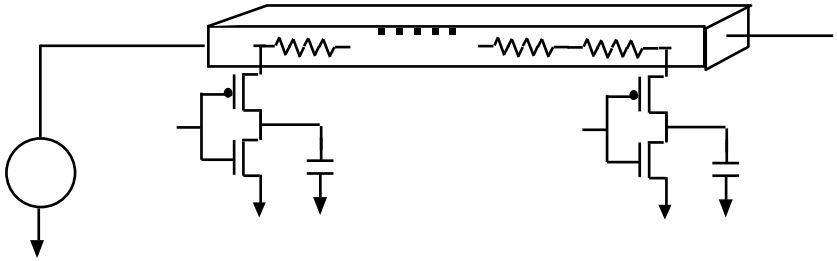
Long/narrow line widths further increases resistance of interconnect

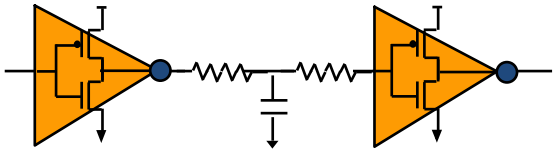
Technology Scaling Effects

At 0.5um and above: Simple capacitance



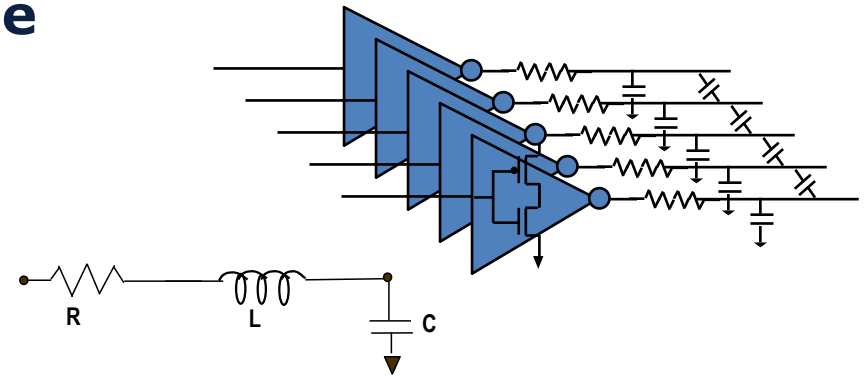
At 0.35um and below: Resistance





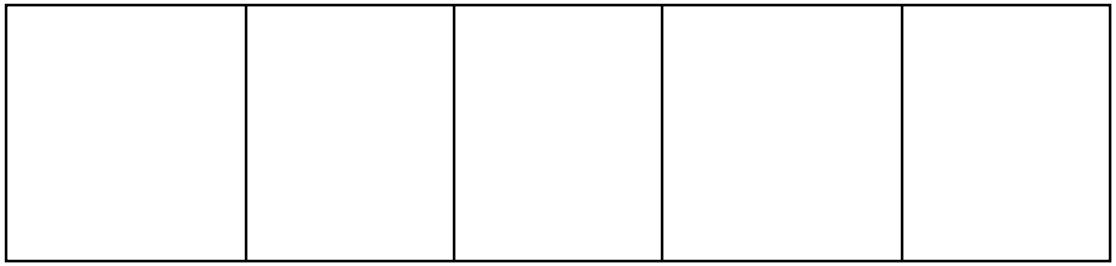
At 0.18um and below : Coupling Capacitance

At 90 nm and below: Inductance



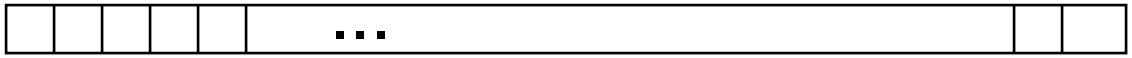
Example: IR drop

Consider a metal line used for power bus routing:



Older technology:

5 squares of resistance

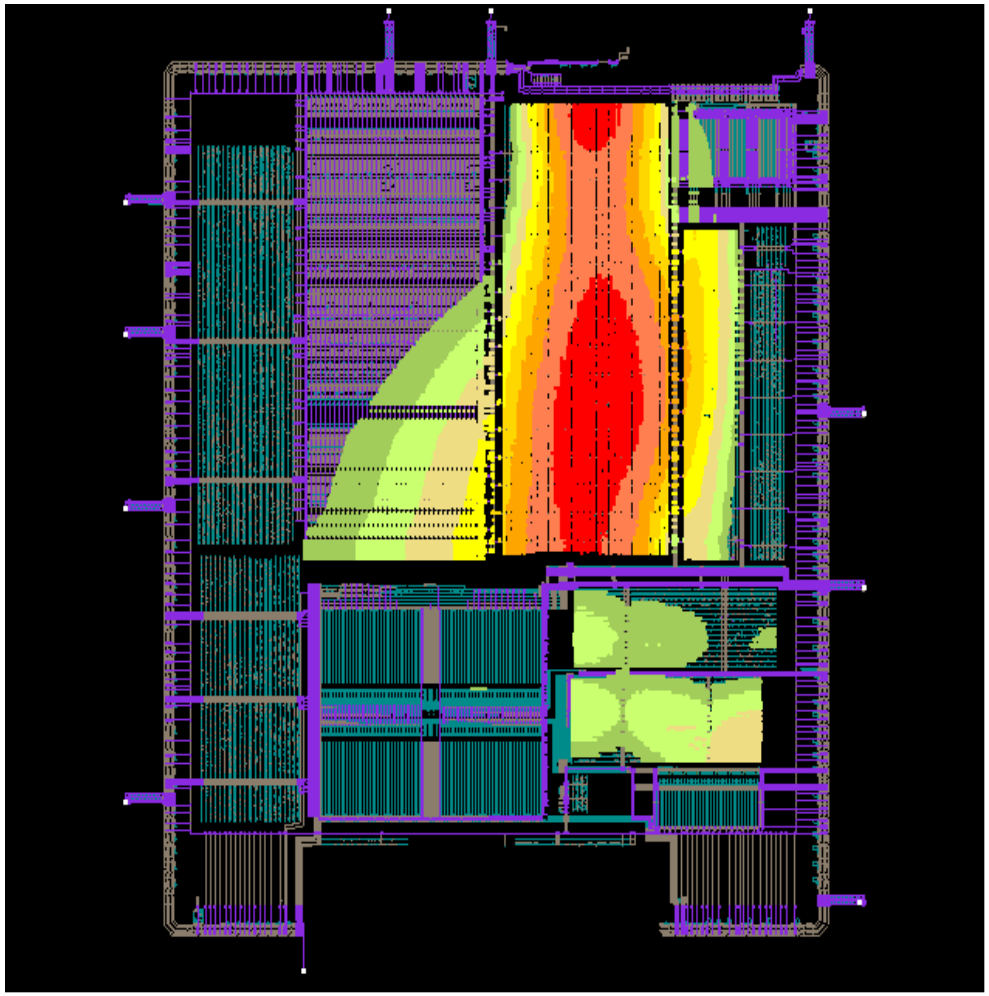


Newer technology:

35 squares of resistance

Scaled technologies increase the resistance going from the Vdd pad to a gate. For example, Vdd may be set to 1.2V at the pad but may be around 0.9V by the time it gets to a gate. This is called IR drop and it will effect the timing of the gate by amount that is dependent on the level of IR drop.

IR Drop Problems: Global Routing Inadequate



- Arrangement of blocks and global power routing determines IR drop
- IR drop impacts gate timing since it reduces the drive capability of devices
- 3.3V --> 0.5ns
- 2.9V --> 0.7ns

Summary

Deep submicron has introduced new issues in MOS integrated circuit design for both devices and interconnect

In order to understand these issues, and design in the presence of the new problems, we need to understand the device and interconnect models in detail
This course will give you the background you need to design in various technology nodes from 180 nm down to below 15 nm

Also, I think it is a lot of fun, and I hope you will find it interesting

Class Notes (live lecture)