



고려대학교
KOREA UNIVERSITY

COM515 Advanced Computer Architecture

Lecture 1. Technology Trend

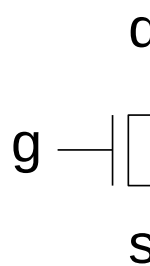
Prof. Taeweon Suh

Computer Science Education

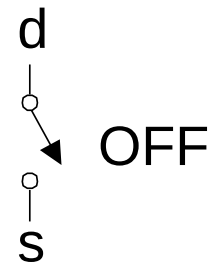
Korea University

Transistor Basics

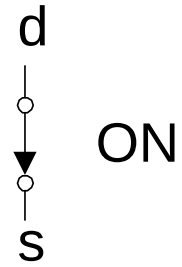
- All semiconductor chips are collections and integrations of transistors
- Transistor is a three-ported voltage-controlled switch
 - Two of the ports are connected depending on the voltage on the third port
 - For example, in the switch below the two terminals (d and s) are connected (ON) only when the third terminal (g) is 1



$g = 0$

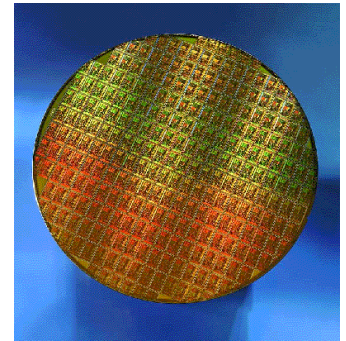


$g = 1$

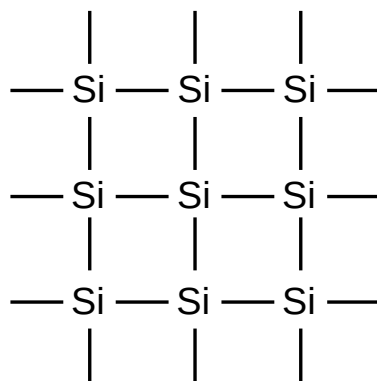


Silicon

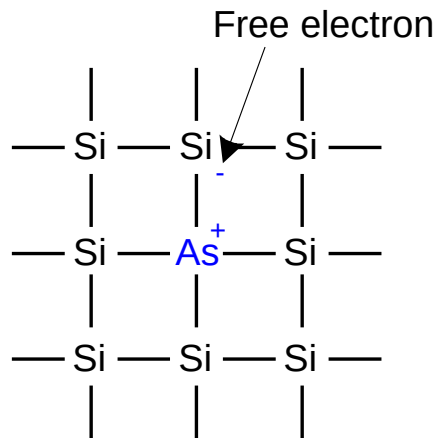
- Transistors are built out of silicon, a semiconductor
- Silicon is not a conductor
- Doped silicon is a conductor
 - n-type (free **n**egative charges, electrons)
 - p-type (free **p**ositive charges, holes)



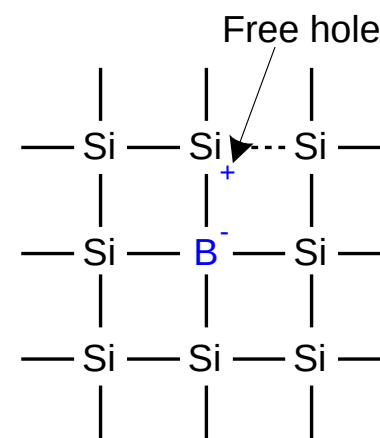
wafer



Silicon Lattice



n-Type



p-Type

Periodic Table of the Elements

PERIODIC TABLE
Atomic Properties of the Elements

Frequently used fundamental physical constants
For the most accurate values of these and other constants, visit physics.nist.gov/constants
1 second = 9 192 631 770 periods of radiation corresponding to the transition between the two hyperfine levels of the ground state of ^{133}Cs

speed of light in vacuum	c	$299\,792\,458\text{ m s}^{-1}$	(exact)
Planck constant	h	$6.626\,069\,57 \times 10^{-34}\text{ J s}$	($h = h/2\pi$)
elementary charge	e	$1.602\,176\,634 \times 10^{-19}\text{ C}$	
electron mass	m_e	$9.109\,381\,86 \times 10^{-31}\text{ kg}$	
proton mass	m_p	$1.672\,621\,63 \times 10^{-27}\text{ kg}$	
fine-structure constant	α	$1/137.035\,999\,074$	
Rydberg constant	R_∞	$10\,973\,731.762\,178\text{ m}^{-1}$	
Boltzmann constant	k	$1.380\,658 \times 10^{-23}\text{ J K}^{-1}$	

Legend:
■ Solids
■ Liquids
■ Gases
■ Artificially Prepared

Group
1 IA, 2 IIA, 3 IIIB, 4 IVB, 5 VB, 6 VIB, 7 VIIB, 8, 9, 10, 11 IB, 12 IIB, 13 IIIB, 14 IVA, 15 VA, 16 VIA, 17 VIIA, 18 VIIIA

Period
1, 2, 3, 4, 5, 6, 7

Atomic Number
Symbol
Name
Atomic Weight
Ground-state Configuration
Ionization Energy (eV)

Example: Cerium (Ce)
 Atomic Number: 58
 Symbol: Ce
 Name: Cerium
 Atomic Weight: 140.116
 Ground-state Configuration: $[\text{Xe}]4f15d16s2$
 Ionization Energy (eV): 5.5387

Periodic Table Data (Selected Elements):

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H (1.00794)																	He (4.002602)
2	Li (6.941)	Be (9.012182)																B (10.811)
3	Na (22.989769)	Mg (24.3050)																C (12.011)
4	K (39.0983)	Ca (40.078)	Sc (44.955910)	Ti (47.88)	V (50.9415)	Cr (51.9961)	Mn (54.93804)	Fe (55.845)	Co (58.933200)	Ni (58.6934)	Cu (63.546)	Zn (65.409)	Ga (69.723)	Ge (72.64)	As (74.92160)	Se (78.96)	Br (79.904)	Kr (83.796)
5	Rb (85.4678)	Sr (87.62)	Y (88.90585)	Zr (91.224)	Nb (92.90638)	Mo (95.94)	Tc (98.90625)	Ru (101.07)	Rh (102.90550)	Pd (106.42)	Ag (107.8682)	Cd (112.411)	In (114.818)	Sn (118.710)	Sb (121.760)	Te (127.60)	I (126.90447)	Xe (131.29)
6	Cs (132.90545)	Ba (137.327)		Hf (178.49)	Ta (180.9479)	W (183.84)	Re (186.207)	Os (190.23)	Ir (192.222)	Pt (195.084)	Au (196.96657)	Hg (200.59)	Tl (204.3833)	Pb (207.2)	Bi (208.98039)	Po (209)	At (210)	Rn (222)
7	Fr (223)	Ra (226)		Rf (261)	Db (262)	Sg (266)	Bh (264)	Hs (277)	Mt (268)	Uun (281)	Uuu (272)	Uub (285)	Uut (289)	Uuq (293)	Uuh (297)			

Lanthanides: 57 La (138.90547), 58 Ce (140.116), 59 Pr (140.90765), 60 Nd (144.24), 61 Pm (144.9127), 62 Sm (150.36), 63 Eu (151.964), 64 Gd (157.25), 65 Tb (158.92534), 66 Dy (162.500), 67 Ho (164.93032), 68 Er (167.259), 69 Tm (168.93421), 70 Yb (173.04), 71 Lu (174.967)

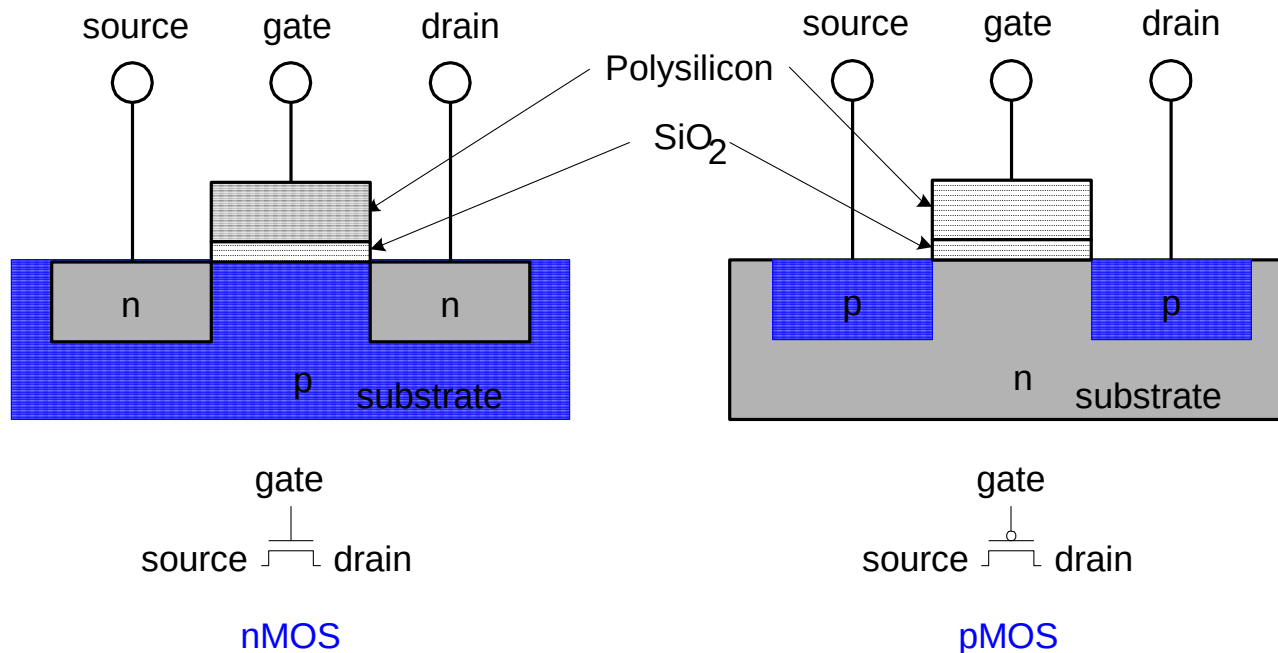
Actinides: 89 Ac (227), 90 Th (232.0381), 91 Pa (231.03688), 92 U (238.02891), 93 Np (237), 94 Pu (244), 95 Am (243), 96 Cm (247), 97 Bk (247), 98 Cf (251), 99 Es (252), 100 Fm (257), 101 Md (258), 102 No (259), 103 Lr (262)

Footnote: ¹Based upon ^{12}C . () indicates the mass number of the most stable isotope.

Source: For a description of the data, visit physics.nist.gov/data NIST SP 966 (September 2003)

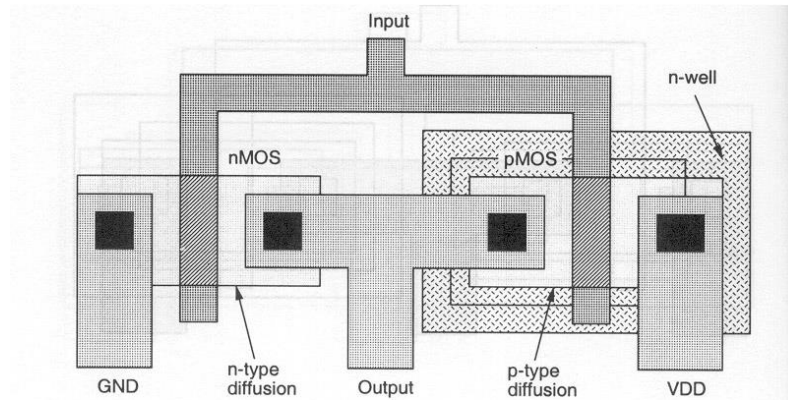
MOS Transistors

- Metal oxide silicon (MOS) transistors:
 - Polysilicon (used to be **Metal**) gate
 - **Oxide** (silicon dioxide) insulator
 - Doped **Silicon** substrate and wells



MOS Transistors

- Top view



- Cross-section

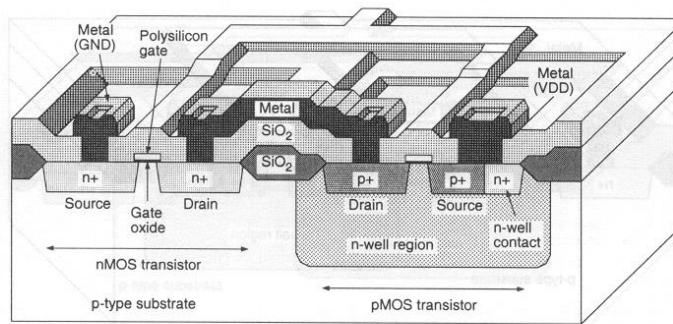
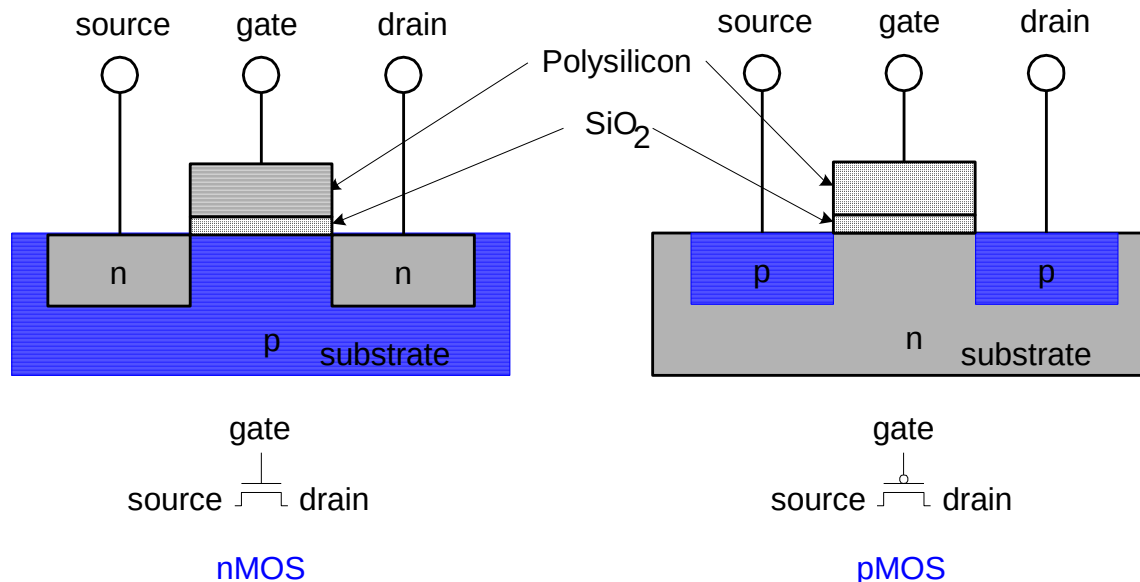


Figure 2.11. The composite layout and the resulting cross-sectional view of the chip, showing one nMOS and one pMOS transistor (in the n-well), and the polysilicon and metal interconnections. The final step is to deposit the passivation layer (for protection) over the chip, except over wire-bonding pad areas. After *Atlas of IC Technologies*, by W. Maly [1].

MOS Transistors

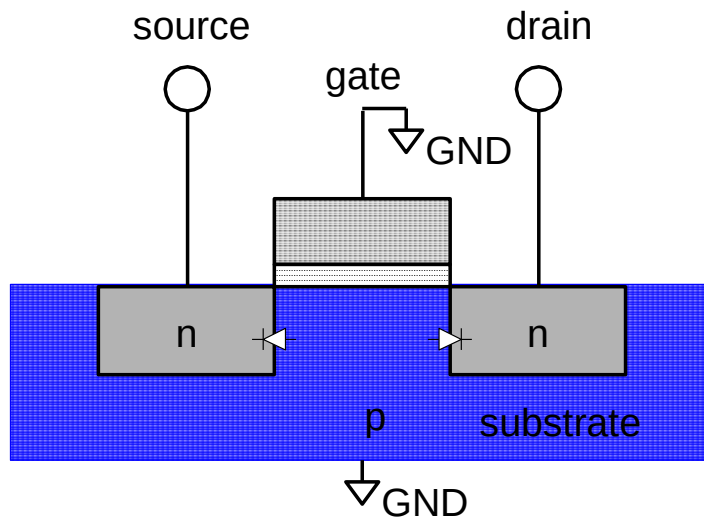
- The MOS sandwich acts as a capacitor (two conductors with insulator between them)
- When voltage is applied to the gate, the opposite charge is attracted to the semiconductor on the other side of the insulator, which could form a *channel* of charge



Transistors: nMOS

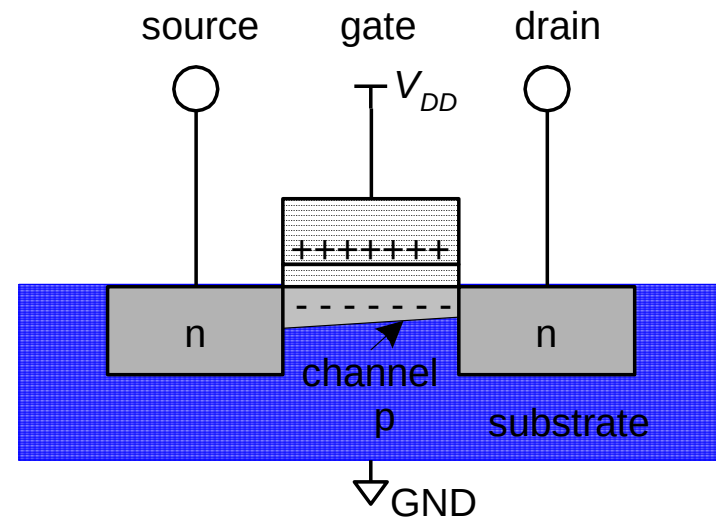
Gate = 0, so it is OFF

(no connection between source and drain)

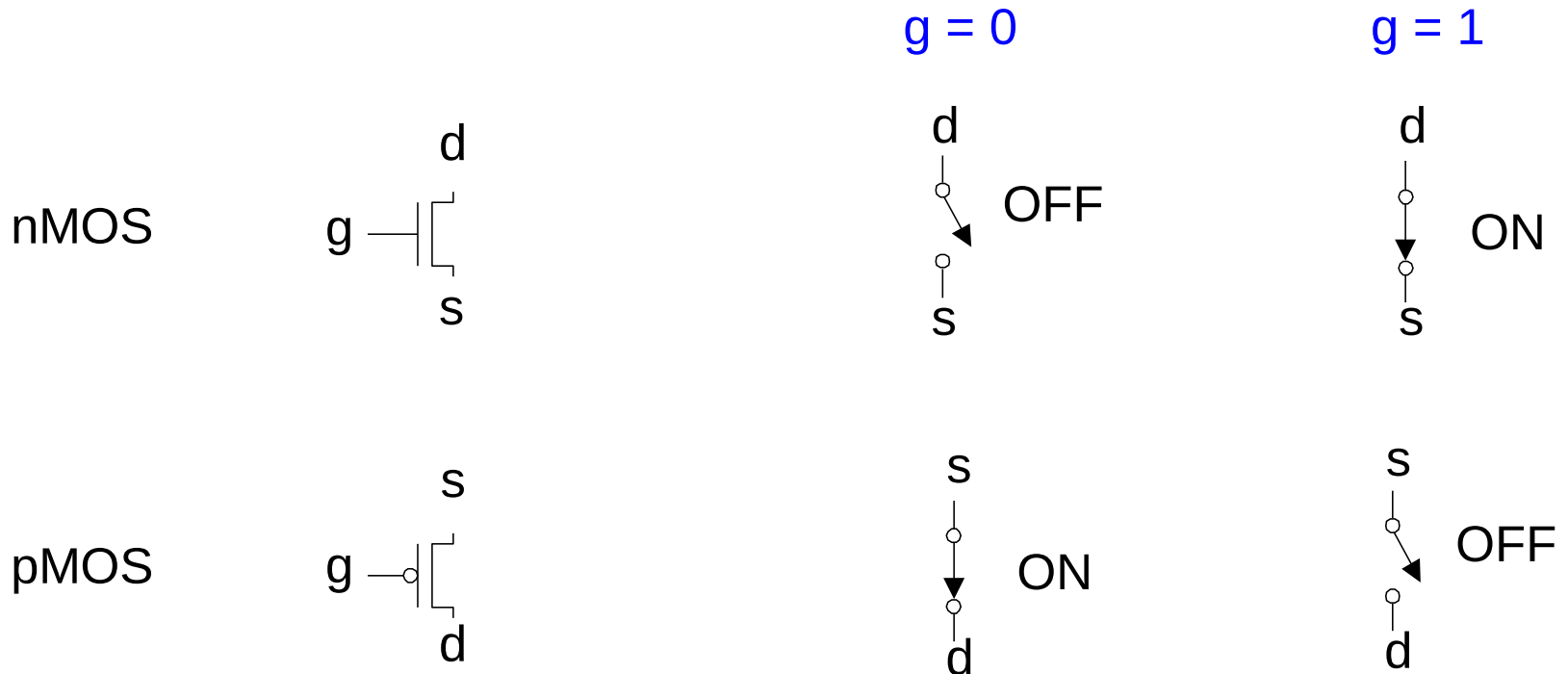


Gate = 1, so it is ON

(connection between source and drain)

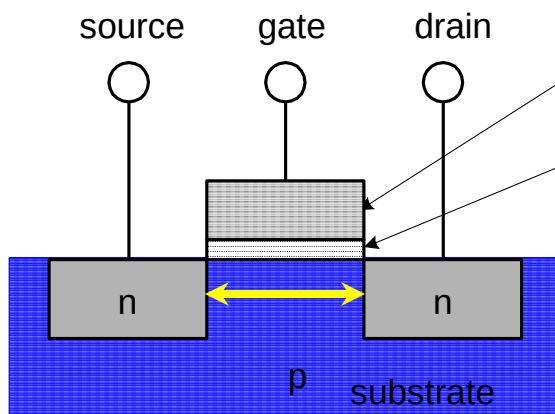


Transistor Function



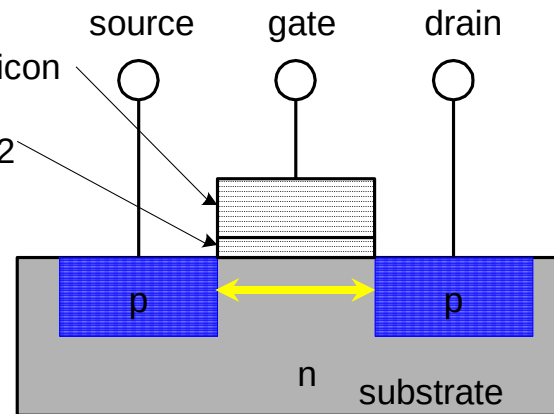
(Semiconductor) Technology

- **Transistor** is simply an on/off switch controlled by electricity
- **IC (Integrated Circuit)** combined dozens to hundreds of transistors into a single chip
- **VLSI (Very Large Scale Integration)** is used to describe the tremendous increase in the number of transistors in a chip
- **(Semiconductor) Technology**: How small can you make a transistor
 - **0.1 μm (100nm), 90nm, 65nm, 45nm, 32nm technologies**



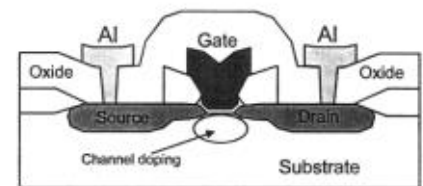
gate
source — drain

nMOS

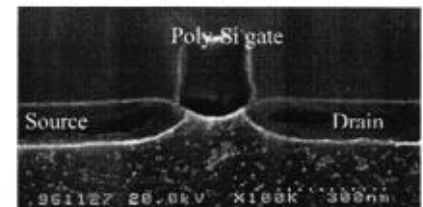


gate
source — drain

pMOS



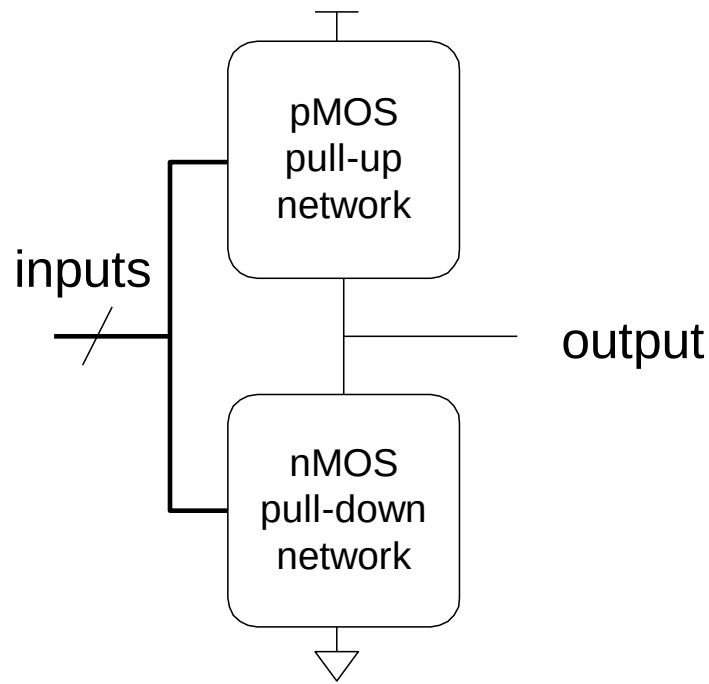
(a)



(b)

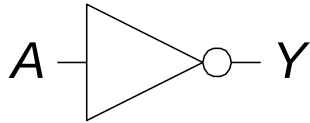
CMOS (Complementary MOS)

- nMOS transistors pass good 0's, so connect source to GND
- pMOS transistors pass good 1's, so connect source to V_{DD}



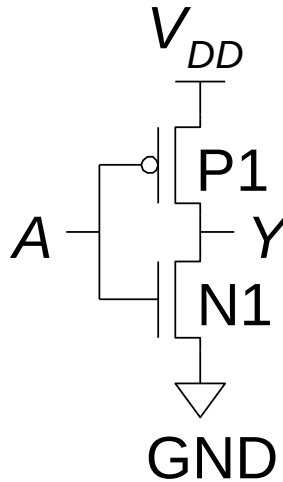
CMOS Gates: NOT Gate

NOT

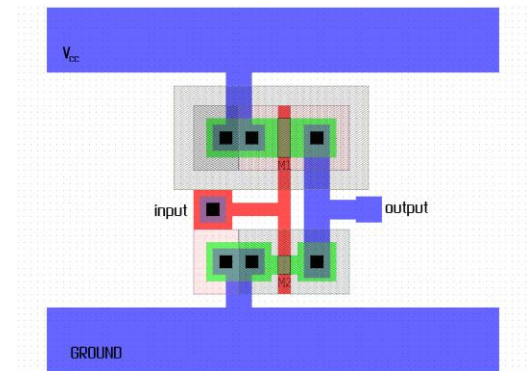


$$Y = \overline{A}$$

A	Y
0	1
1	0



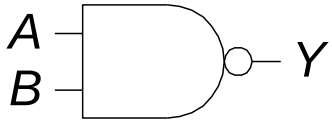
Layout (top view)



A	P1	N1	Y
0	ON	OFF	1
1	OFF	ON	0

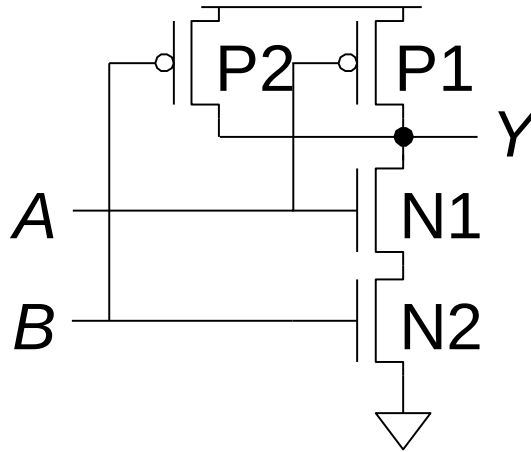
CMOS Gates: NAND Gate

NAND

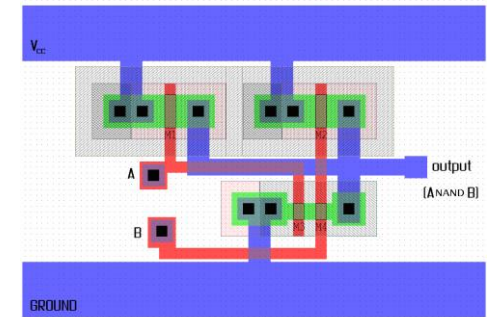


$$Y = \overline{AB}$$

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0



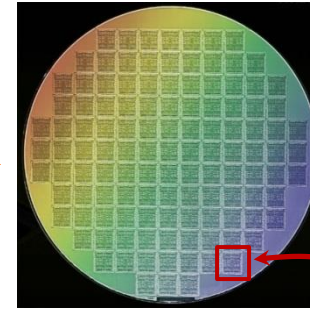
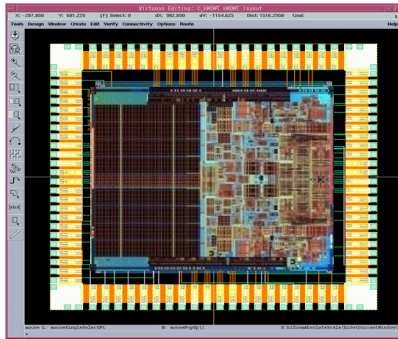
Layout



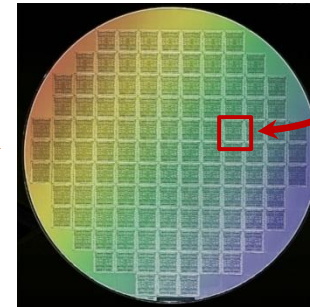
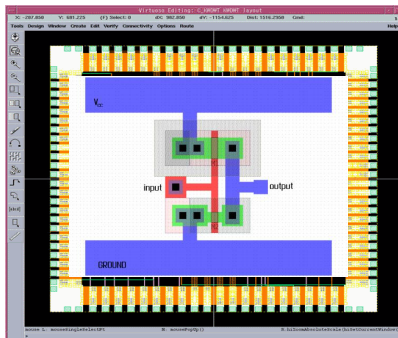
A	B	P1	P2	N1	N2	Y
0	0	ON	ON	OFF	OFF	1
0	1	ON	OFF	OFF	ON	1
1	0	OFF	ON	ON	OFF	1
1	1	OFF	OFF	ON	ON	0

Now, Let's Make an Inverter Chip

Core 2 Duo



Your Inverter chip



die

- **Yield** means how many dies are working correctly after fabrication

x86?

- What is x86?
 - Generic term referring to processors from Intel, AMD and VIA
 - Derived from the model numbers of the first few generations of processors:
 - 80**86**, 802**86**, 803**86**, 804**86** → **x86**
 - Now it generally refers to processors from Intel, AMD, and VIA
 - x86-16: 16-bit processor
 - x86-32 (aka IA32): 32-bit processor * IA: Intel Architecture
 - x86-64: 64-bit processor
- Intel takes about 80% of the PC market and AMD takes about 20%
 - Apple also have been introducing Intel-based Mac from Nov. 2006



Power under the hood.

Inside the MacBook is a powerful 2.13GHz Intel Core 2 Duo processor based on advanced Core microarchitecture. MacBook provides a fast 1066MHz frontside bus and 3MB of shared L2 cache, so you'll have more than enough horsepower to get the job done.



x86 History (as of 2008)

40 years of changing the world

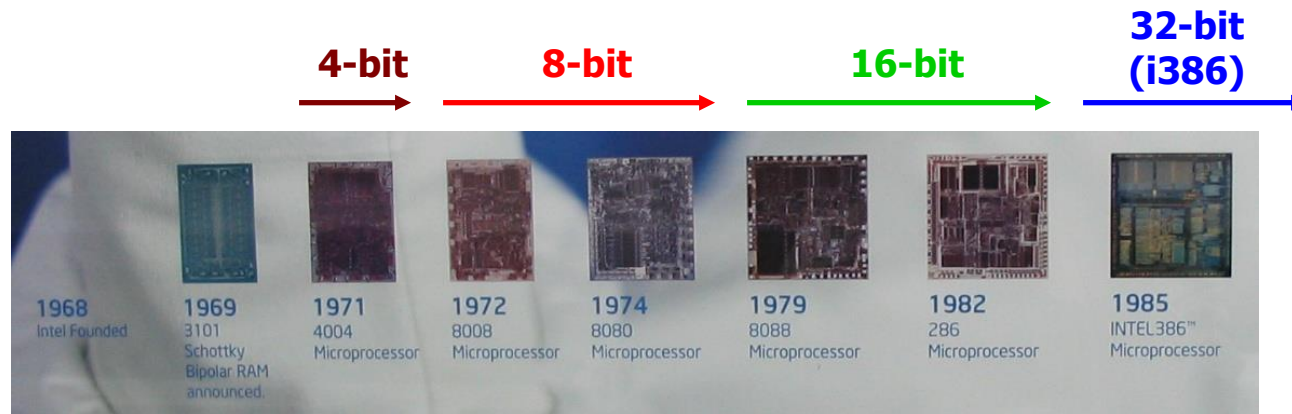
Since Intel's earliest days, we've harnessed the power of the transistor to transform what's possible for people around the world. In our first 40 years we've redefined how people live, work and play. Imagine what we can do together in the next 40.

Learn more at next40.intel.com

intel Leap ahead™

Year	Processor
1968	Intel 4004
1969	Intel 8008
1971	Intel 8080
1972	Intel 8085
1974	Intel 8088
1979	Intel 80286
1982	Intel 286
1985	Intel 386
1989	Intel 486
1993	Intel Pentium
1997	Intel Pentium Pro
1999	Intel Pentium 4
2000	Intel Pentium 4
2006	Intel Core 2 Duo
2008	Intel Core 2 Duo

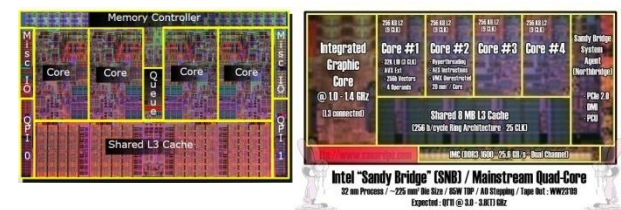
x86 History (Cont.)



32-bit (i586)

32-bit (i686)

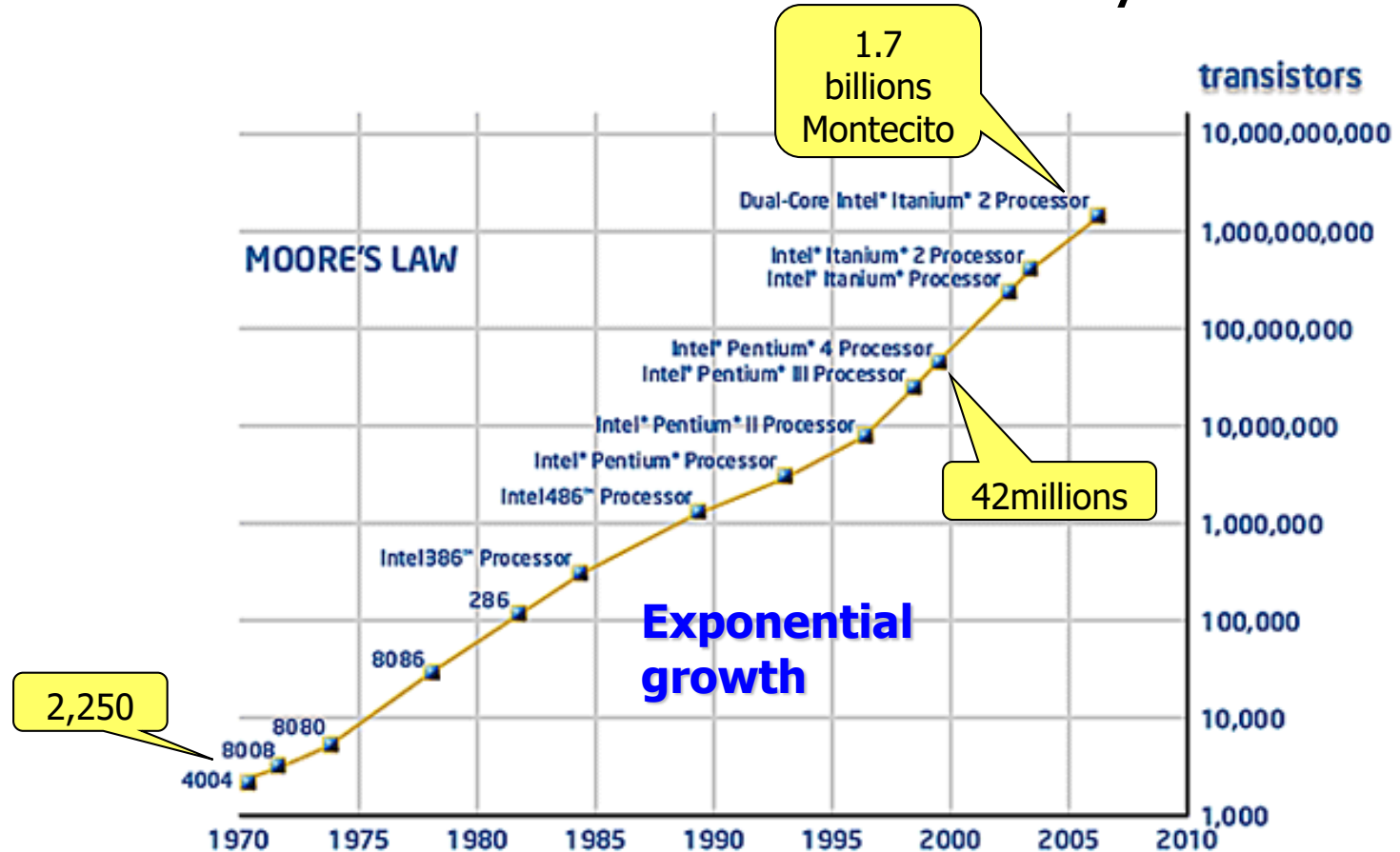
64-bit (x86_64)



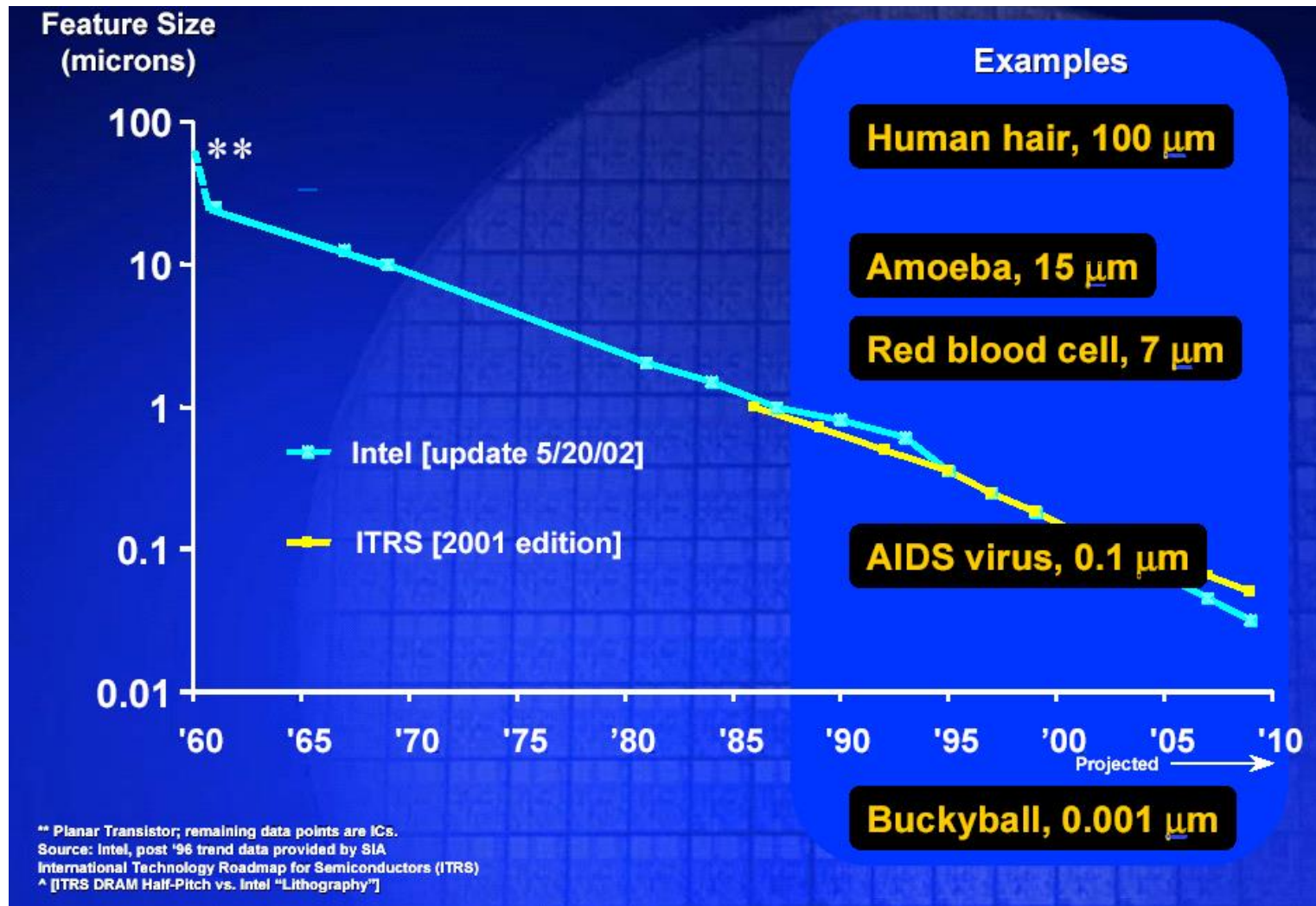


Moore's Law

- Transistor count will be doubled every 18 months



Feature Size (Technology) Trend



Power Dissipation

- By early 2000, Intel and AMD made every effort to increase clock frequency to enhance the performance of their CPUs
- But, the power consumption is the problem

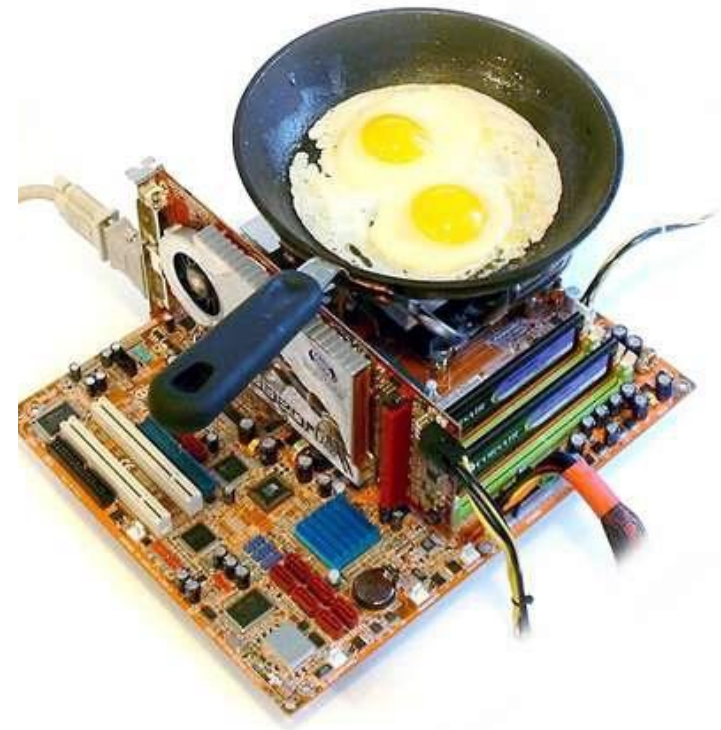
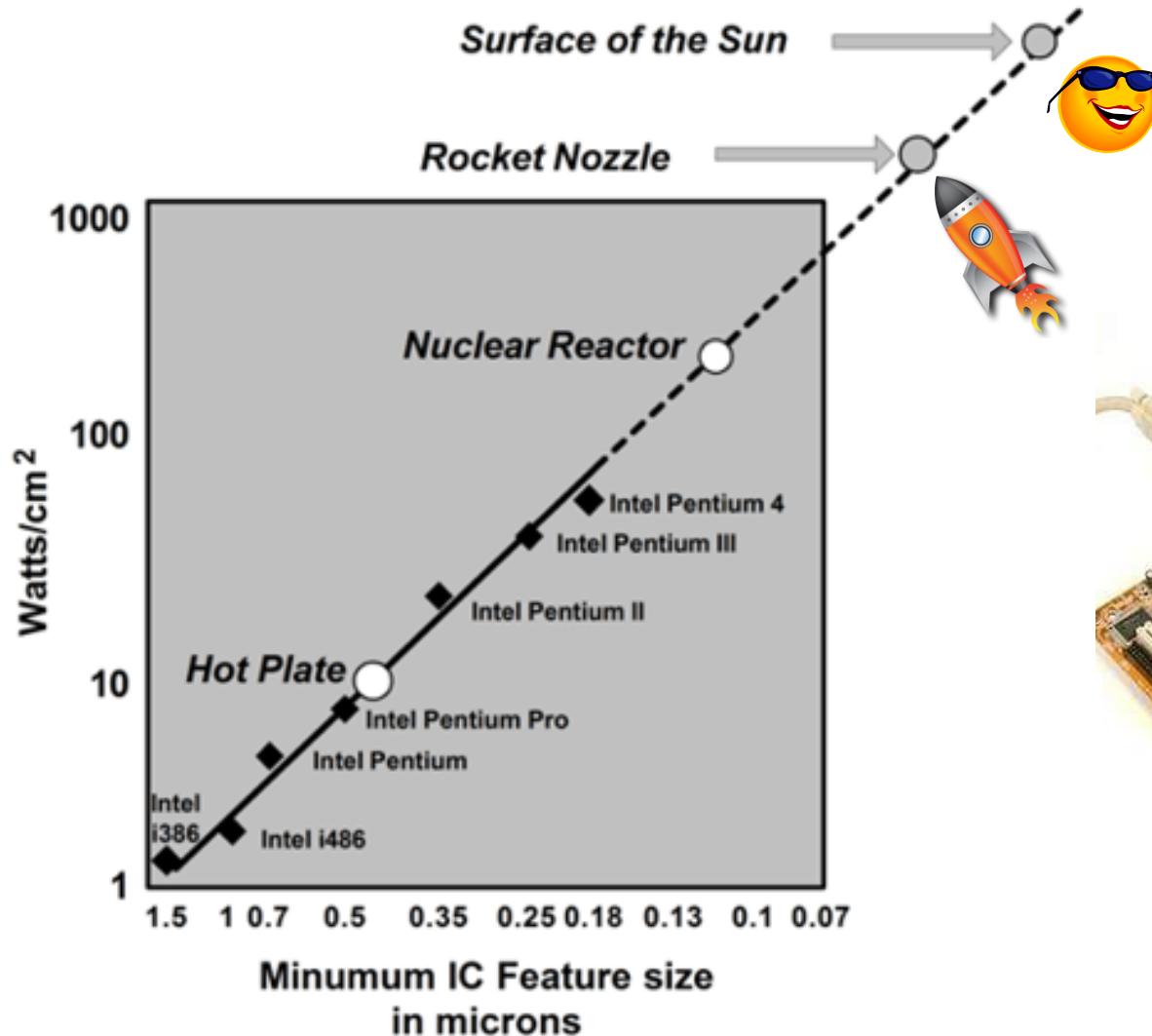
$$P \approx CV_{DD}^2 f$$

C: Capacitance

V_{DD}: Voltage

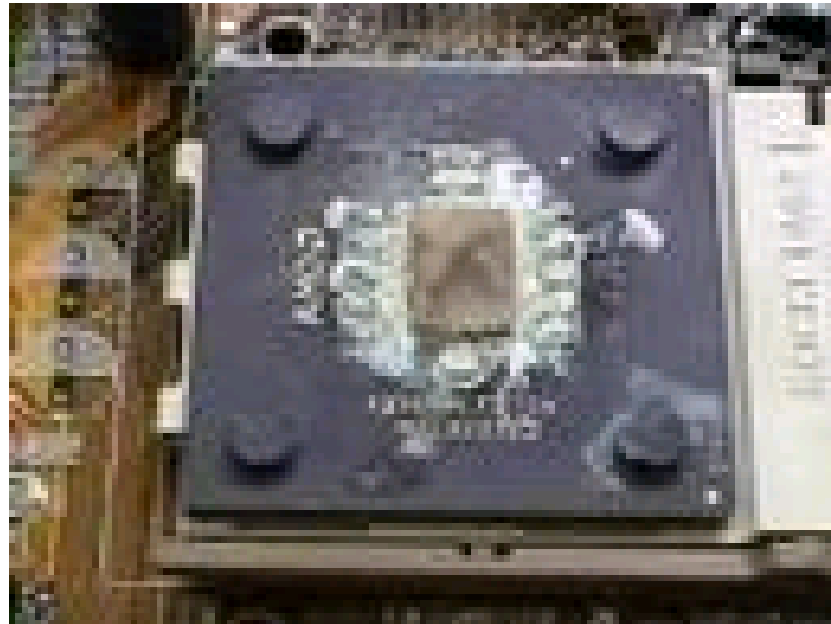
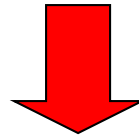
f: Frequency

Power Density Trend



Watch this!

Click the chip



How to Reduce Power Consumption?

- Reduce supply voltage with new technologies
 - i.e., reducing transistor size
- Keep the clock frequency in modest range
 - No longer increase the clock frequency
- Then... what would be the problem?

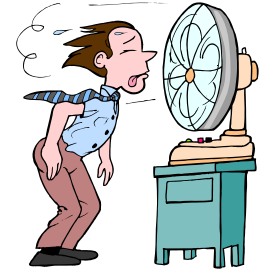
Performance

- So, the strategy is to integrate simple many CPUs in a chip

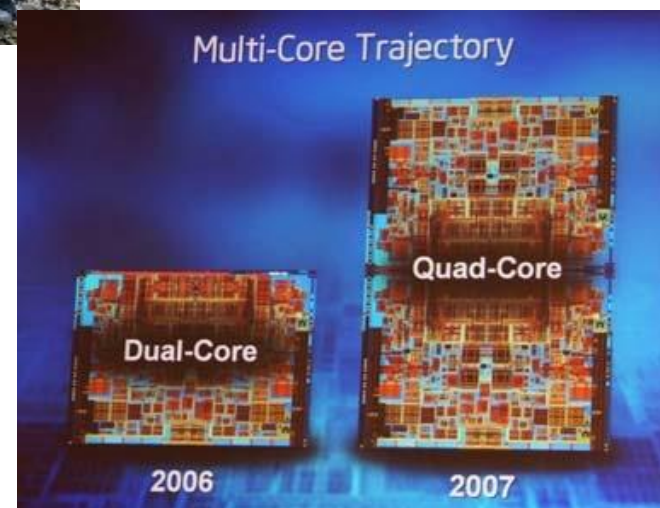
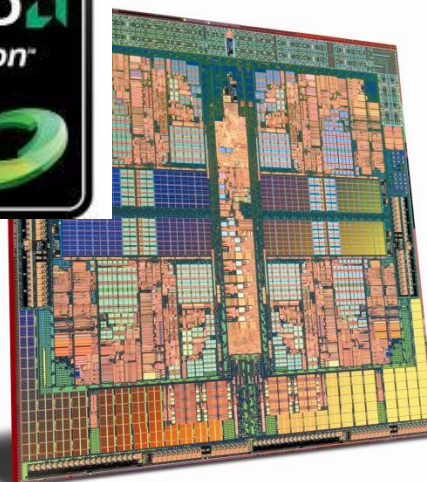
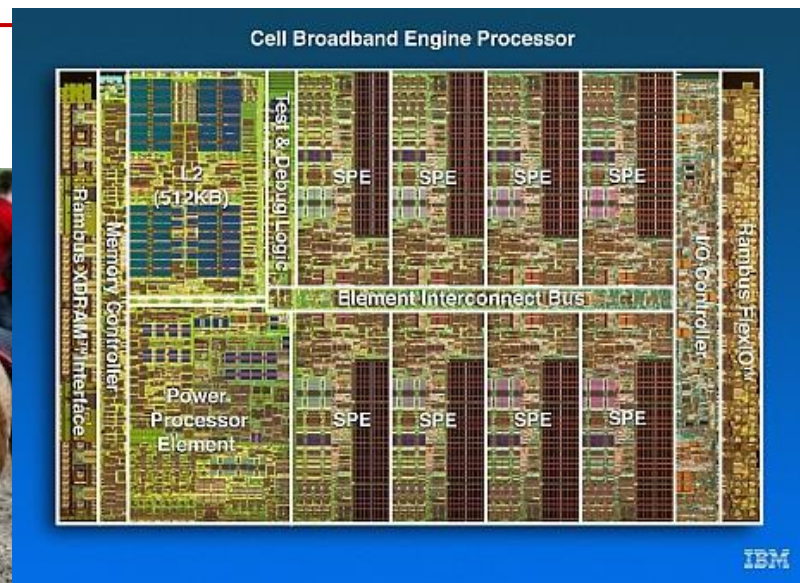
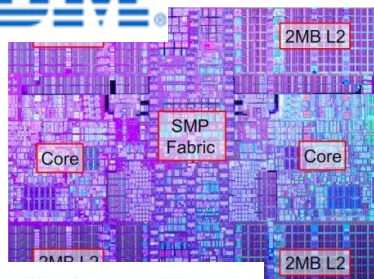
Dual Core, Quad Core....

Reality Check, circa 200x

- Conventional processor designs run out of steam
 - Power wall (**thermal**)
 - Complexity (verification)
 - Physics (CMOS scaling)
- Unanimous direction → **Multi-core**
 - Simple cores (massive number)
 - Keep
 - Wire communication on leash
 - Gordon Moore happy (Moore's Law)
 - Architects' menace: kick the ball to the other side of the court?

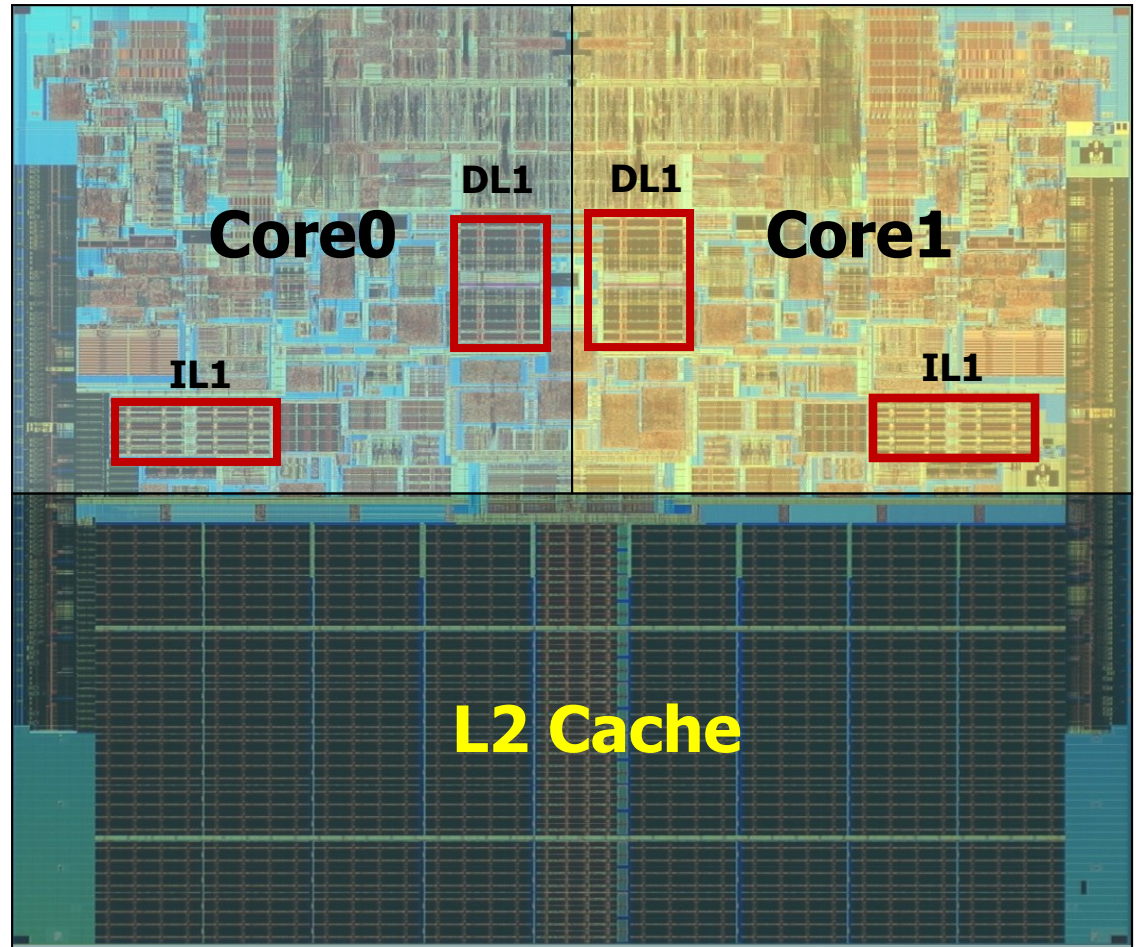


Multi-core Processor Gala



Intel's Core 2 Duo

- 2 cores on one chip
- Two levels of caches (L1, L2) on chip
- 291 million transistors in 143 mm² with 65nm technology



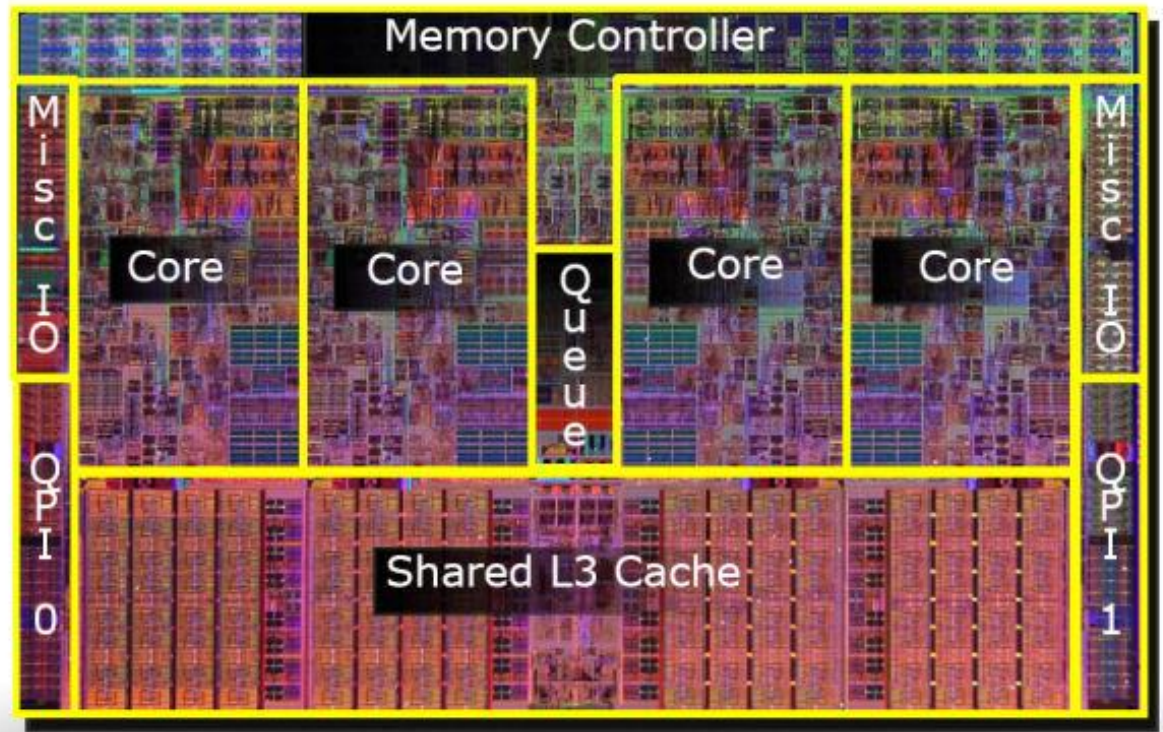
Source: <http://www.sandpile.org>

Intel's Core i7

- 4 cores on one chip
- Three levels of caches (L1, L2, L3) on chip
- 731 million transistors in 263 mm² with 45nm technology



The First Nehalem Processor

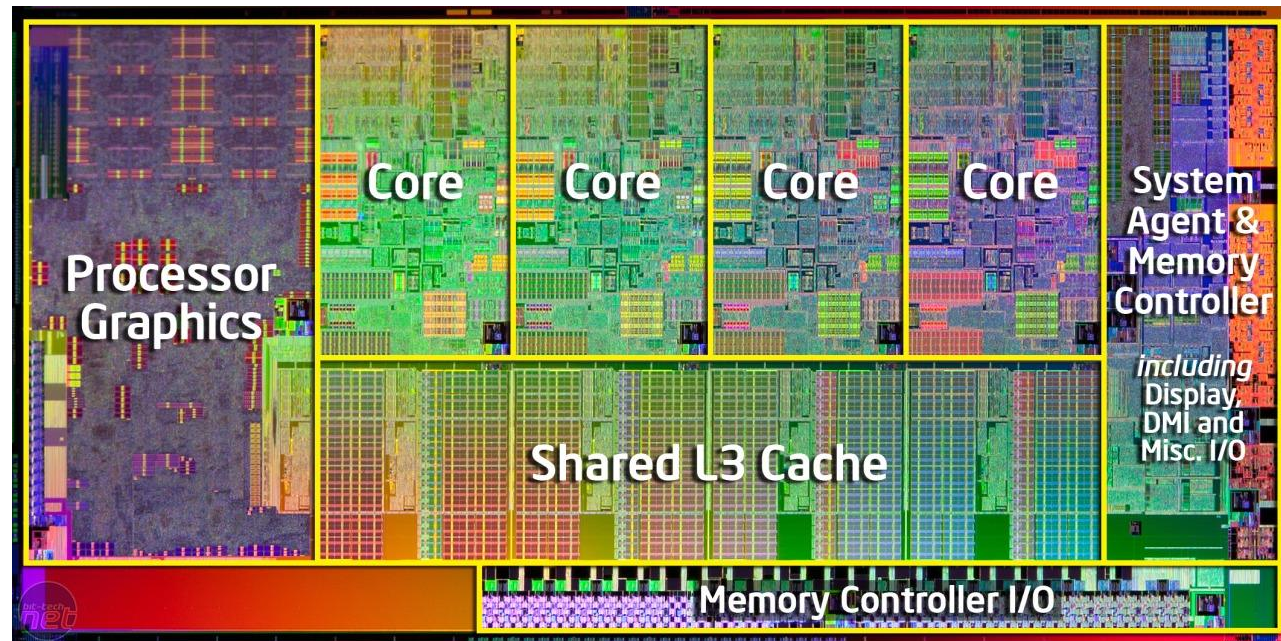


Intel's Core i7 (2nd Gen.)

2nd Generation Core i7

Sandy Bridge

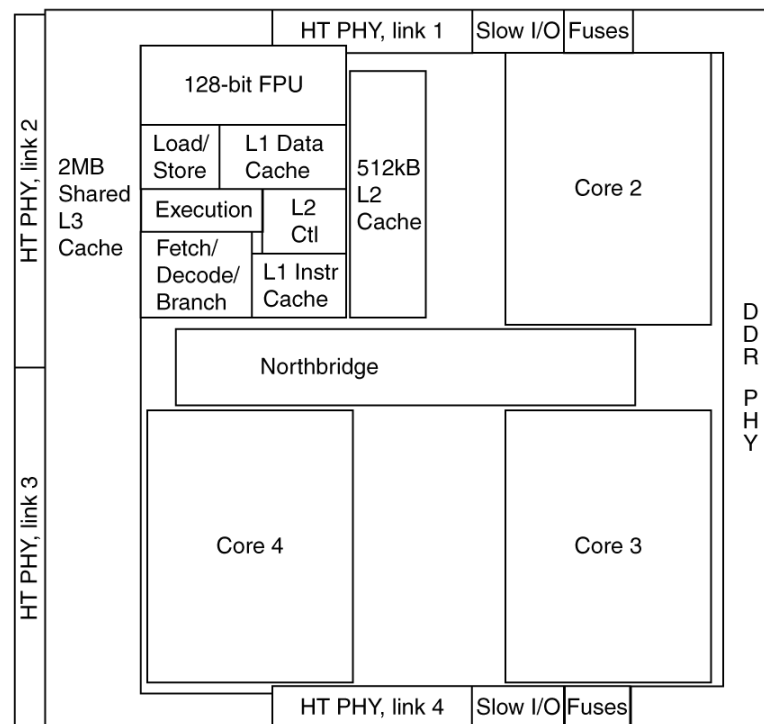
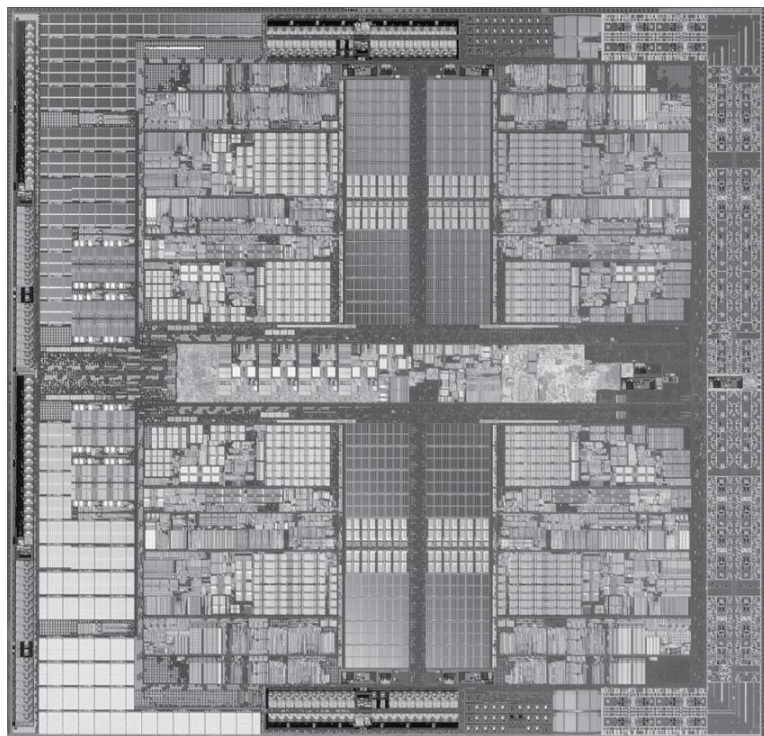
L1	32 KB
L2	256 KB
L3	8MB



**995 million transistors
in 216 mm² with 32nm
technology**

AMD's Opteron – Barcelona (2007)

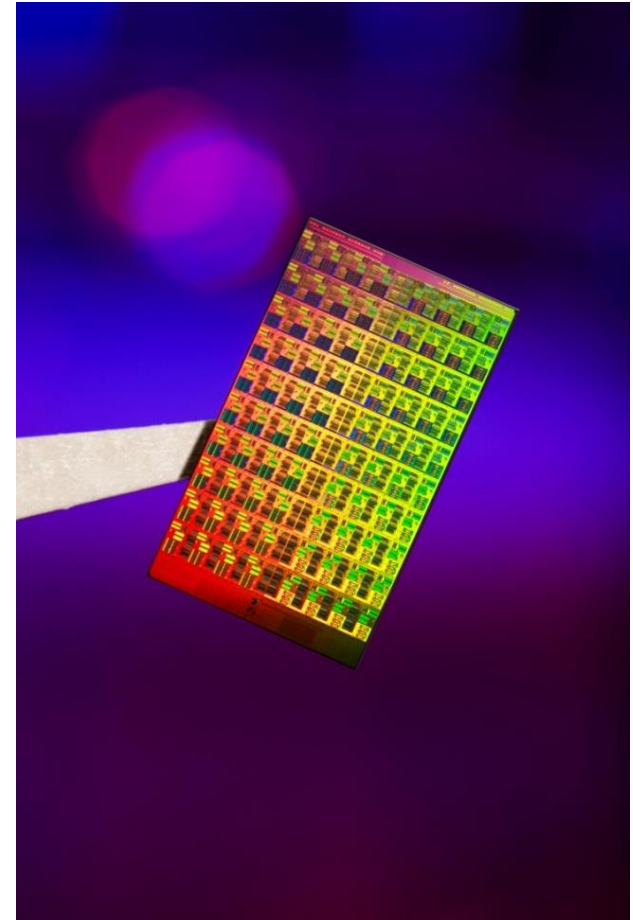
- 4 cores on one chip
- 1.9GHz clock
- 65nm technology
- Three levels of caches (L1, L2, L3) on chip
- Integrated North Bridge



Intel Teraflops Research Chip

- 80 CPU cores
- Deliver more than 1 trillion floating-point operations per second (**1 Teraflops**) of performance

Computer Performance	
Name	FLOPS
yottaFLOPS	10^{24}
zettaFLOPS	10^{21}
exaFLOPS	10^{18}
petaFLOPS	10^{15}
teraFLOPS	10^{12}
gigaFLOPS	10^9
megaFLOPS	10^6
kiloFLOPS	10^3



Introduced in September 2006

Intel's 48 Core Processor

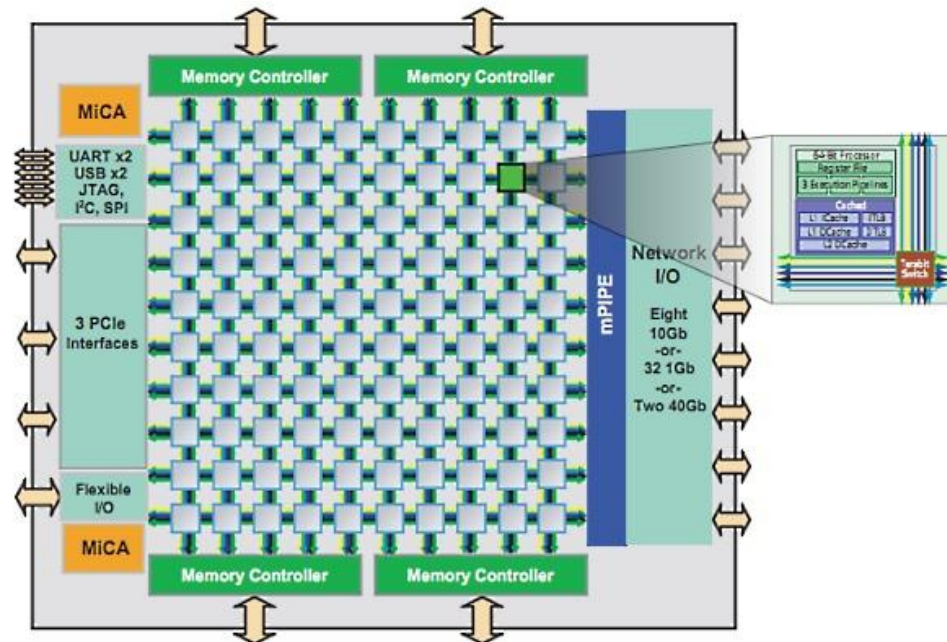
- 48 x86 cores manufactured with 45nm technology
- Nicknamed "single-chip cloud computer"



Debuted in December 2009

Tilera's 100 cores (June 2011)

- Tilera has introduced a range of processors (64-bit Gx family: 36 cores, 64 cores and 100 cores), aiming to take on Intel in servers that handle high-throughput web applications
 - 64-bit cores running up to 1.5GHz
 - Manufactured in 40nm technology



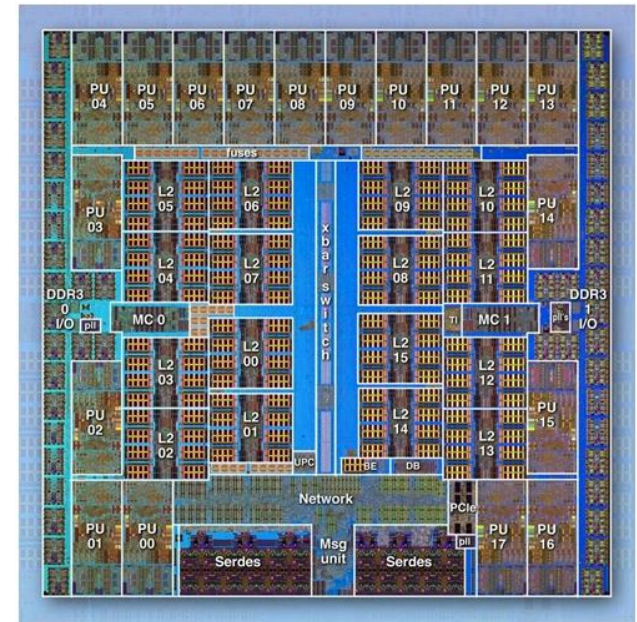
[TILE Gx 3000 Series Overview](#)

IBM Bluegene/Q Processor

- The Bluegene/Q processors will power the 20 petaflops Sequoia supercomputer being built by IBM for Lawrence Livermore National Labs.
- Bluegene/Q has 18 cores
 - First processor supporting hardware transactional memory
 - Each core is a 64-bit 4-way multithreaded PowerPC A2
 - 16 cores are used for running actual computations; one will be used for running the operating system; the other is used to improve chip reliability
 - 1.47 billion transistors
 - 1.6 GHz



Bluegene/P Supercomputer in Argonne National Lab.



IBM's Bluegene/Q Processor (2011)

Performance

- If you edit your ms-word document on dual core, would it be running twice faster?

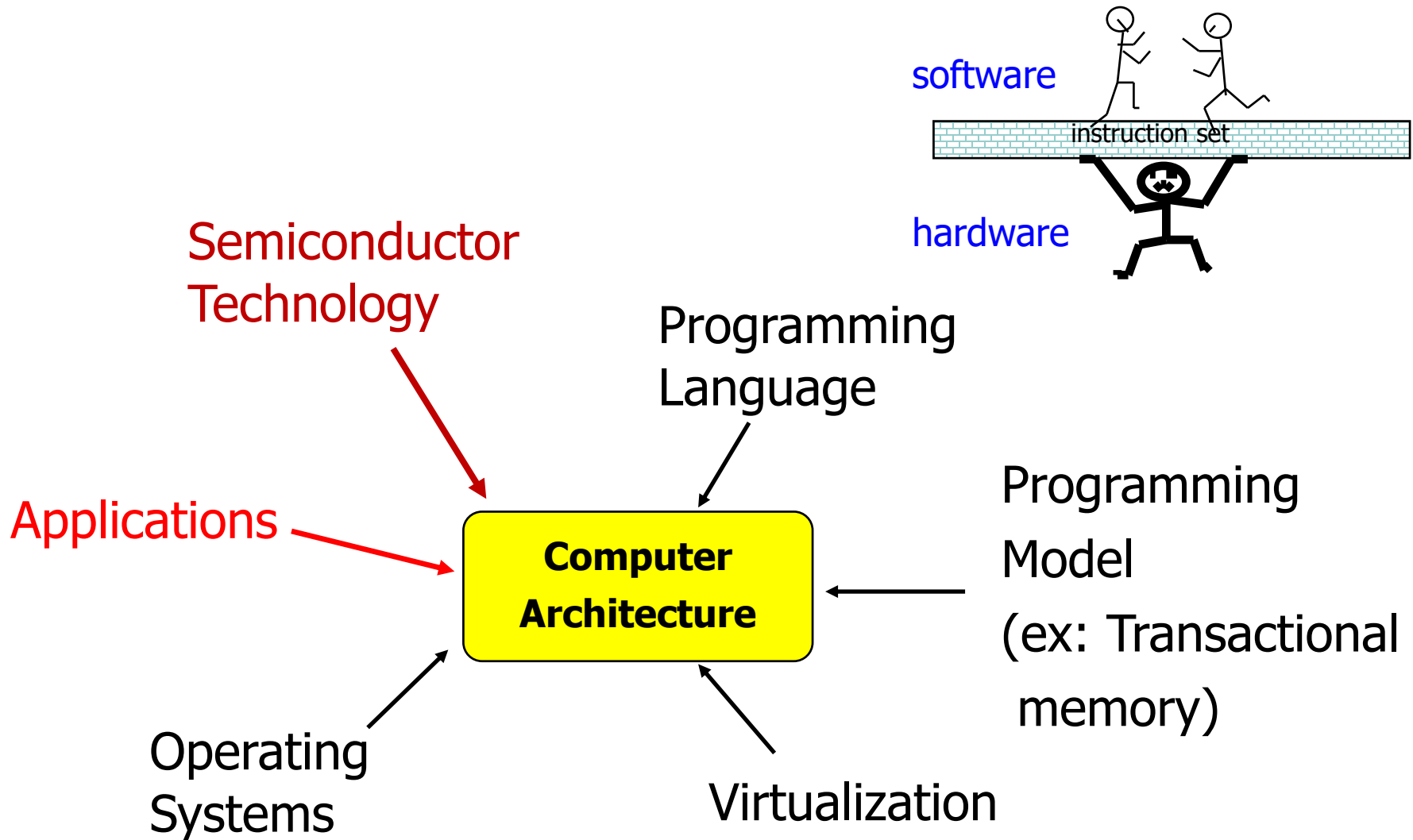
No!

- The problem now is how to parallelize applications and efficiently use hardware resources (available cores)...
- If you were plowing a field, which would you rather use: Two strong oxen or 1024 chickens?
 - Seymour Cray (the father of supercomputing)

Well, it is hard to say in Computing World



Focus on Computer Architecture



Changing Definition

- 50s to 60s: Computer Architecture ~ Computer Arithmetic
- 70s to mid 80s: Instruction Set Design, especially ISA appropriate for compilers
- 90s: Speculation: Predict this, predict that; memory system; I/O system; Multiprocessors; Networks
- 2000s: Power efficiency, Communication, On-die Interconnection Network, Multi-this, Multi-that.
- 2010s and beyond: Thousand-core processors, Self adapting systems? Self organizing structures? DNA Systems/Quantum Computing?