Introduction

brief history - trends

Contents

- Early development of computers.
- Computer Architecture
- Trends in performance
- Trends in Technology
- Trends in Power and Energy
- Measuring performance
- Amdahl’s Law
- Processor Performance Equation

Bibliography

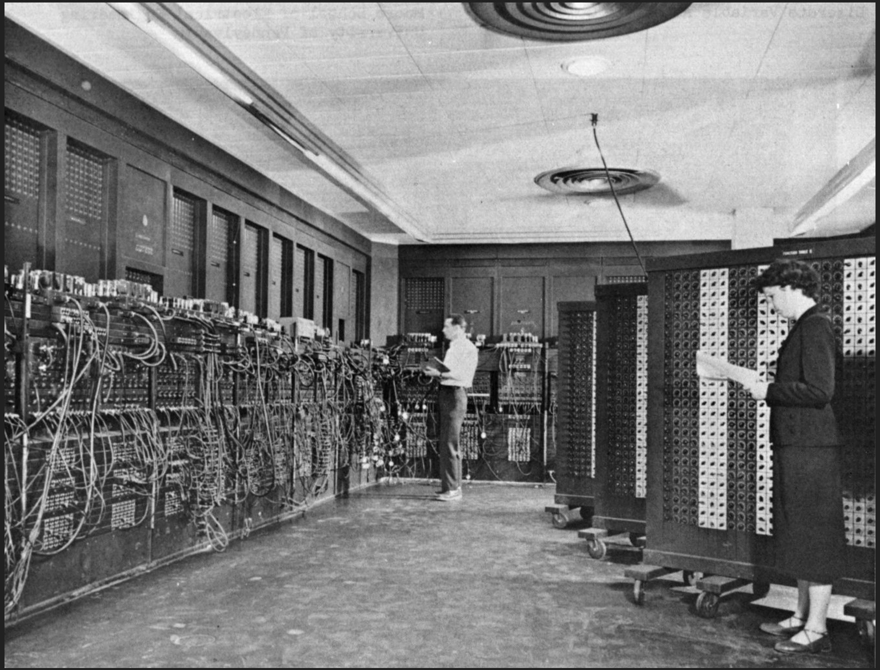
- Computer Architecture, Sixth Edition_ A Quantitative Approach. John L. Hennessy, David A. Patterson.
- Computer Organization and Design - Risc-V Edition
- The essentials of computer organization and architecture / Linda Null, Julia Lobur.
- Organización y arquitectura de computadores Séptima Edición WILLIAM STALLINGS

Early development of computers.

- Generation Zero: Mechanical Calculating Machines
- 1st Generation: Vacuum tubes
- 2nd Generation: Transistor
- 3rd Generation: IC
- Last Generation: VLIC

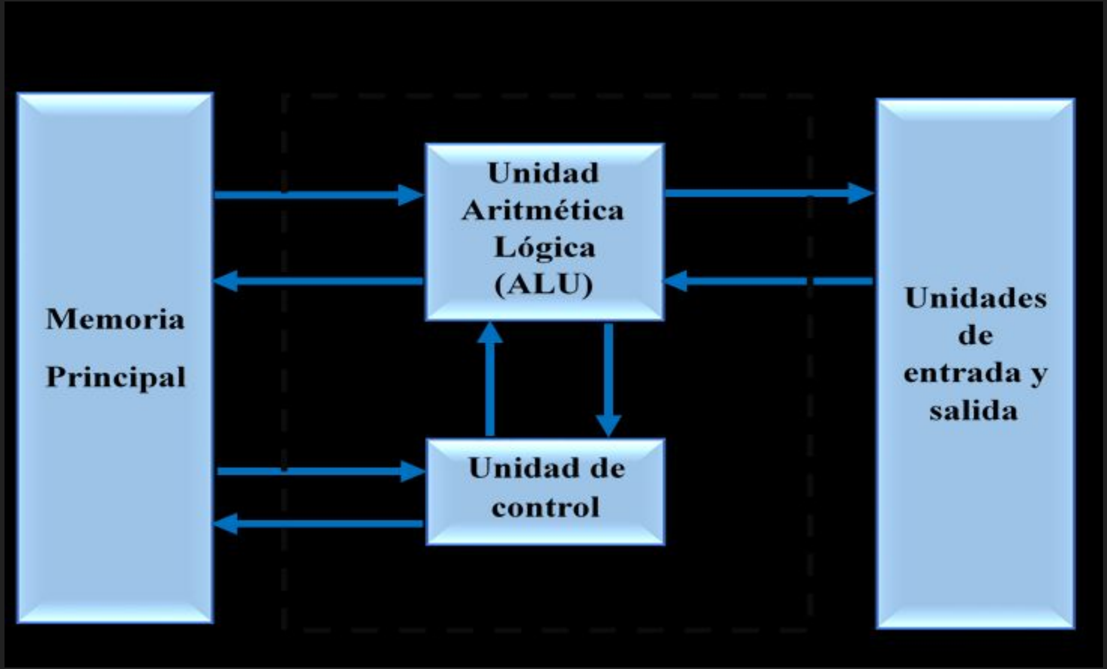
1st Generation: Vacuum tubes

- 1946 - ENIAC (Electronic numerical integrator and calculator)
- Ballistics Research Laboratory
- 18.000 tubes - 30 tons - 140 KW
- 20 10-digit registers
- 5000 adds per sec (200 us)



1st Generation - von Neumann

- stored-program computers
- IAS - 1952

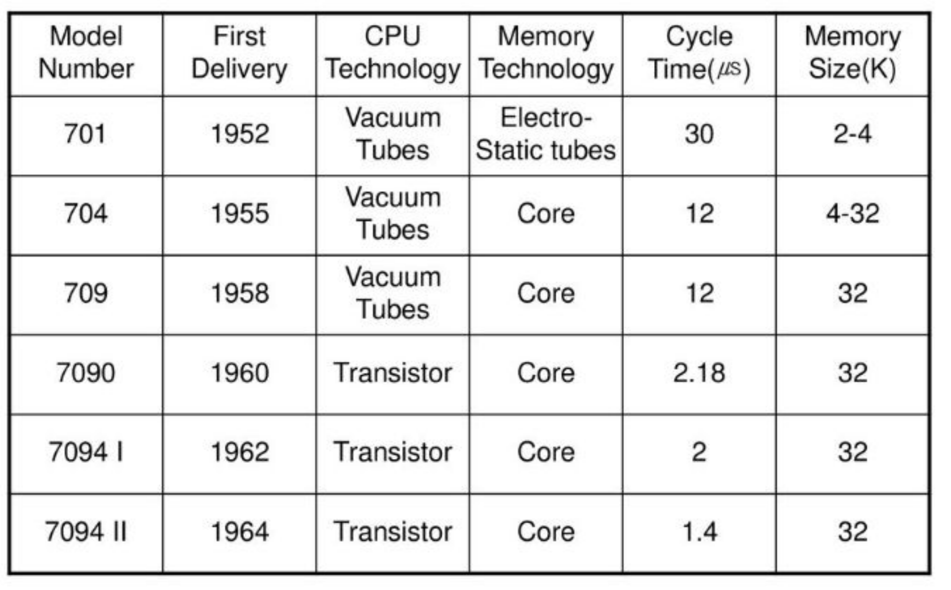


1st Generation - Commercial developments

- Eckert-Mauchly - Sperry-Rand. UNIVAC 1 - 1951
- IBM 1953 - 701, 702, 704, 705

2nd Generation: Transistor

- IBM 7000
- DEC PDP-1



Model Number	First Delivery	CPU Technology	Memory Technology	Cycle Time(μs)	Memory Size(K)
701	1952	Vacuum Tubes	Electro-Static tubes	30	2-4
704	1955	Vacuum Tubes	Core	12	4-32
709	1958	Vacuum Tubes	Core	12	32
7090	1960	Transistor	Core	2.18	32
7094 I	1962	Transistor	Core	2	32
7094 II	1964	Transistor	Core	1.4	32

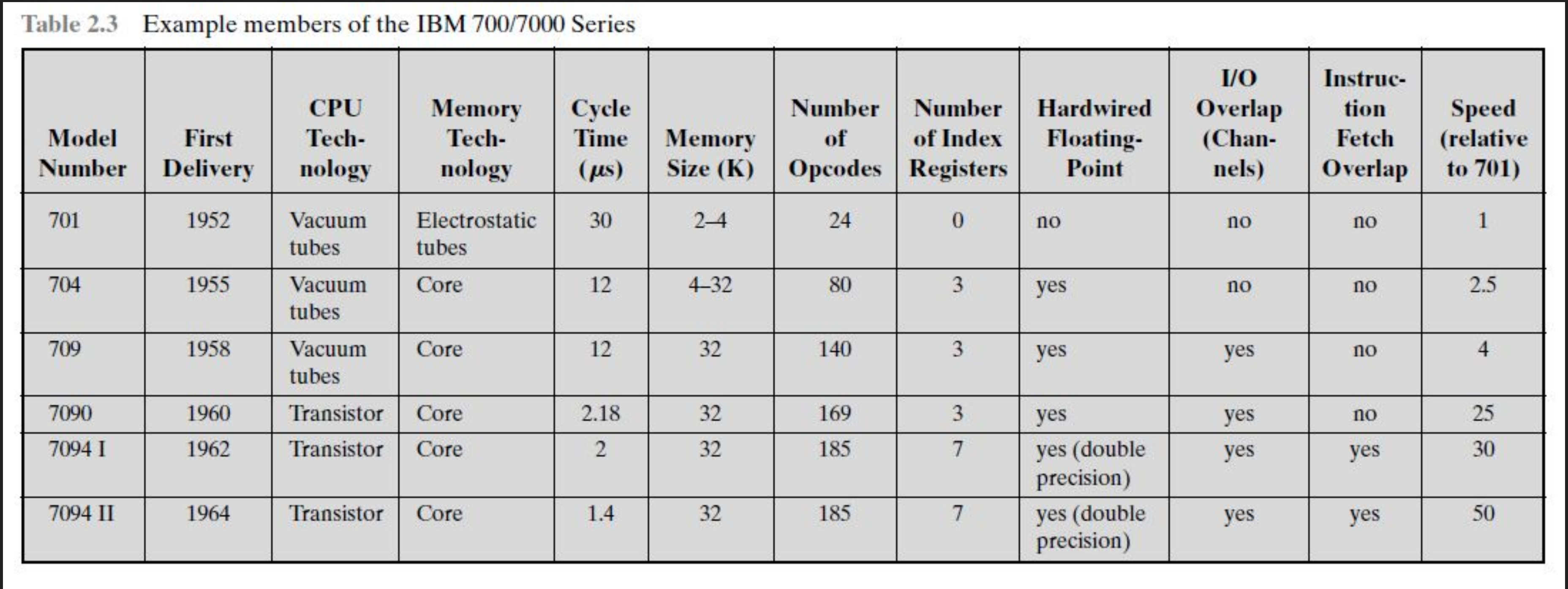
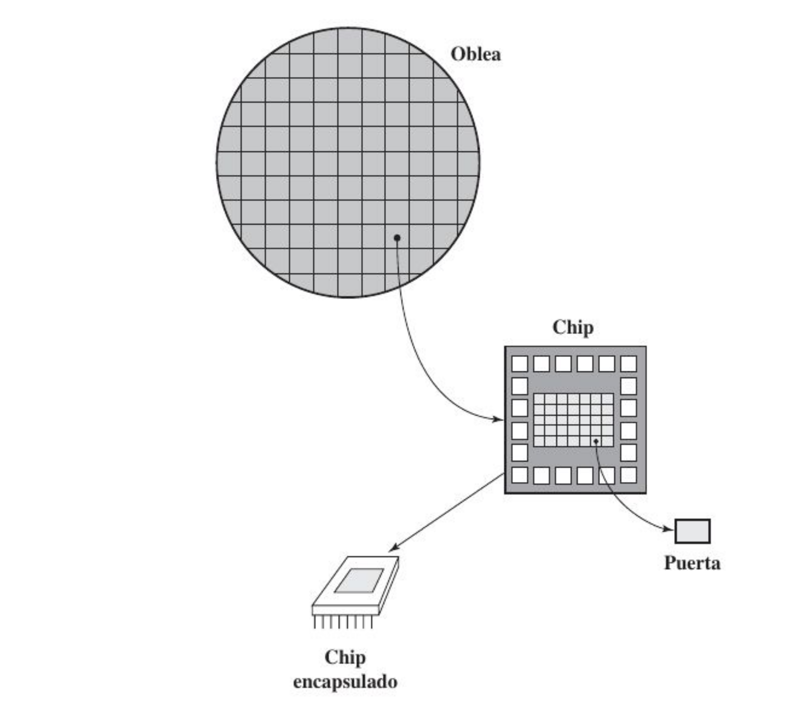


Table 2.3 Example members of the IBM 700/7000 Series

Model Number	First Delivery	CPU Technology	Memory Technology	Cycle Time (μs)	Memory Size (K)	Number of Opcodes	Number of Index Registers	Hardwired Floating-Point	I/O Overlap (Channels)	Instruction Fetch Overlap	Speed (relative to 701)
701	1952	Vacuum tubes	Electrostatic tubes	30	2–4	24	0	no	no	no	1
704	1955	Vacuum tubes	Core	12	4–32	80	3	yes	no	no	2.5
709	1958	Vacuum tubes	Core	12	32	140	3	yes	yes	no	4
7090	1960	Transistor	Core	2.18	32	169	3	yes	yes	no	25
7094 I	1962	Transistor	Core	2	32	185	7	yes (double precision)	yes	yes	30
7094 II	1964	Transistor	Core	1.4	32	185	7	yes (double precision)	yes	yes	50

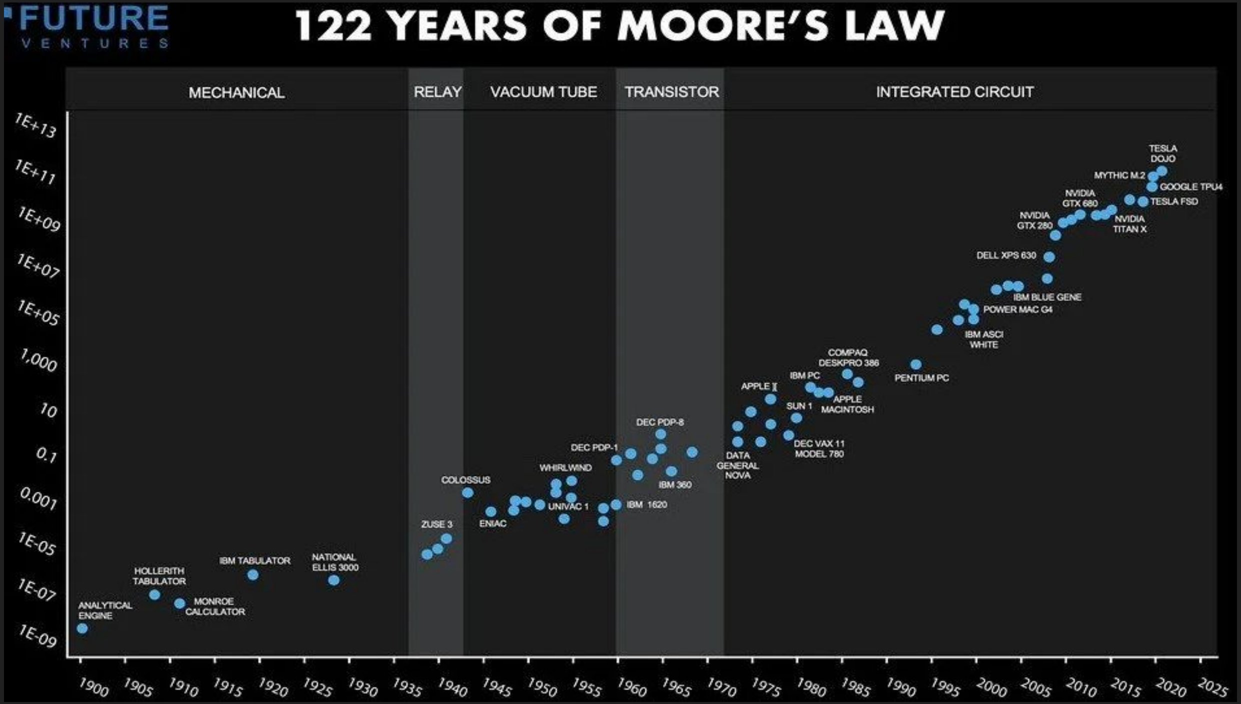
3rd Generation - Integrated Circuits

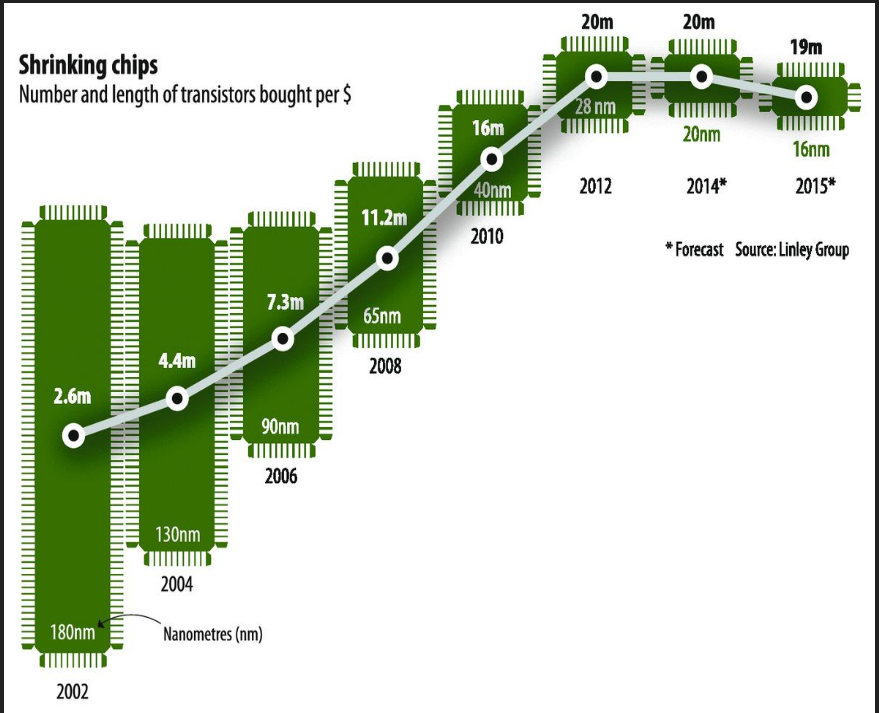
- IBM Sys/360
- DEC PDP-8



3rd Generation - Moore’s Law

- Processing speed
- Price

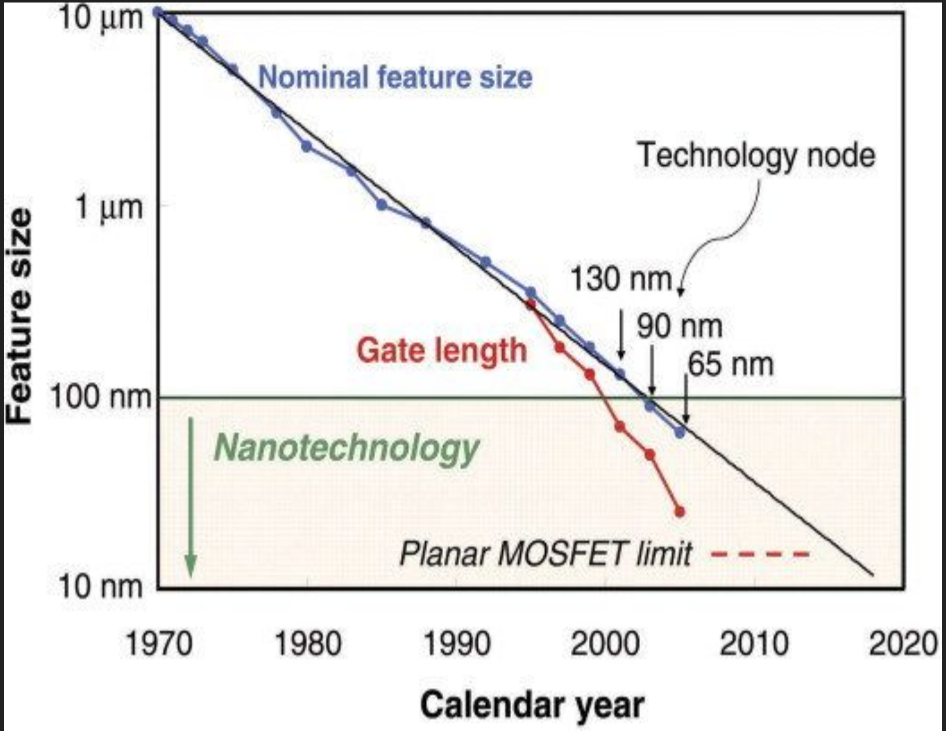




Shrinking chips

Number and length of transistors bought per $

* Forecast Source: Linley Group

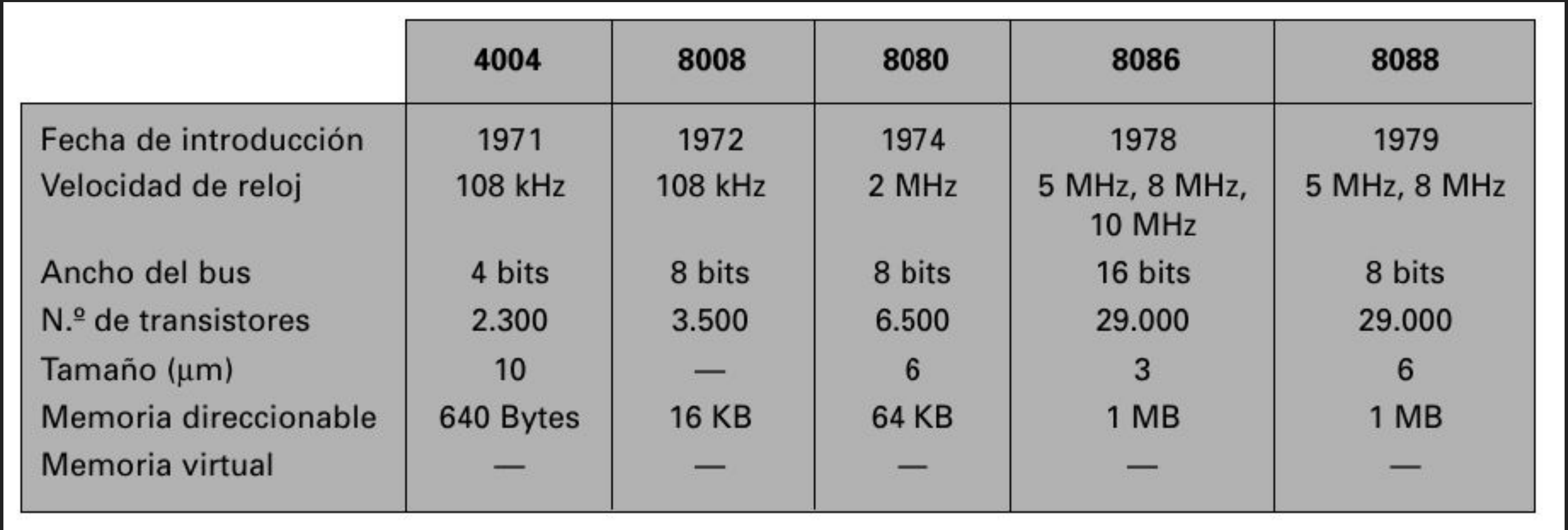


Last Generation.

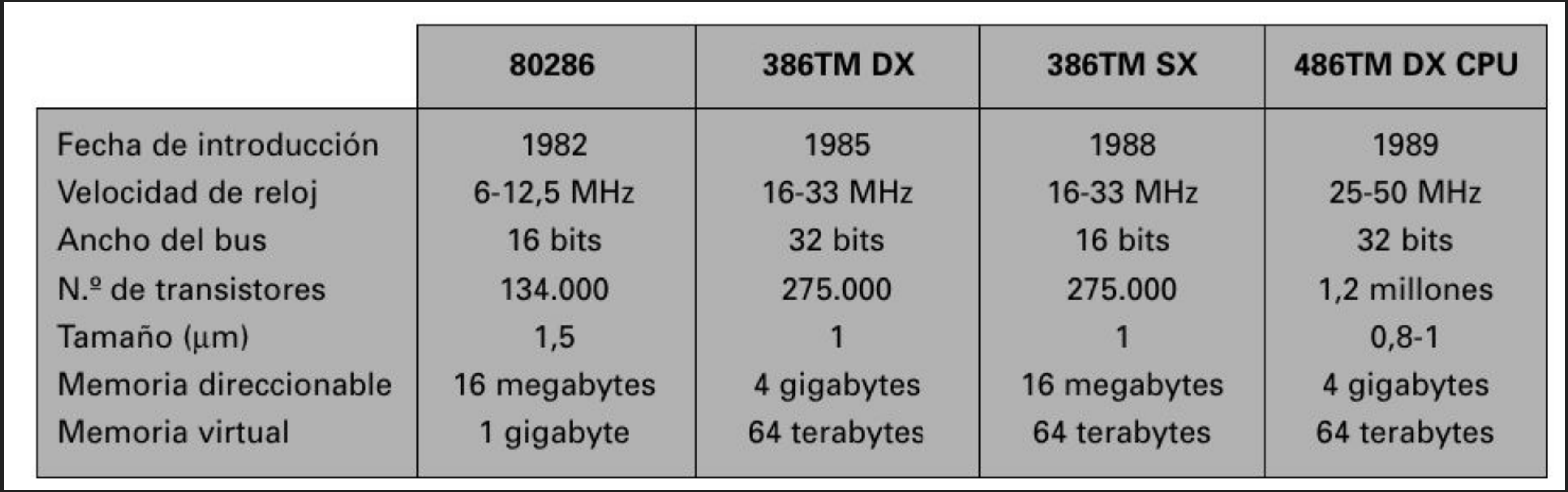
- SSI (small scale integration) - 10 to 100 components per chip
- MSI (medium scale integration) - 100 to 1,000 components per chip
- LSI (large scale integration) - 1,000 to 10,000 components per chip
- VLSI (very large scale integration) - > 10,000 components per chip. → Intel 4004

microprocessor

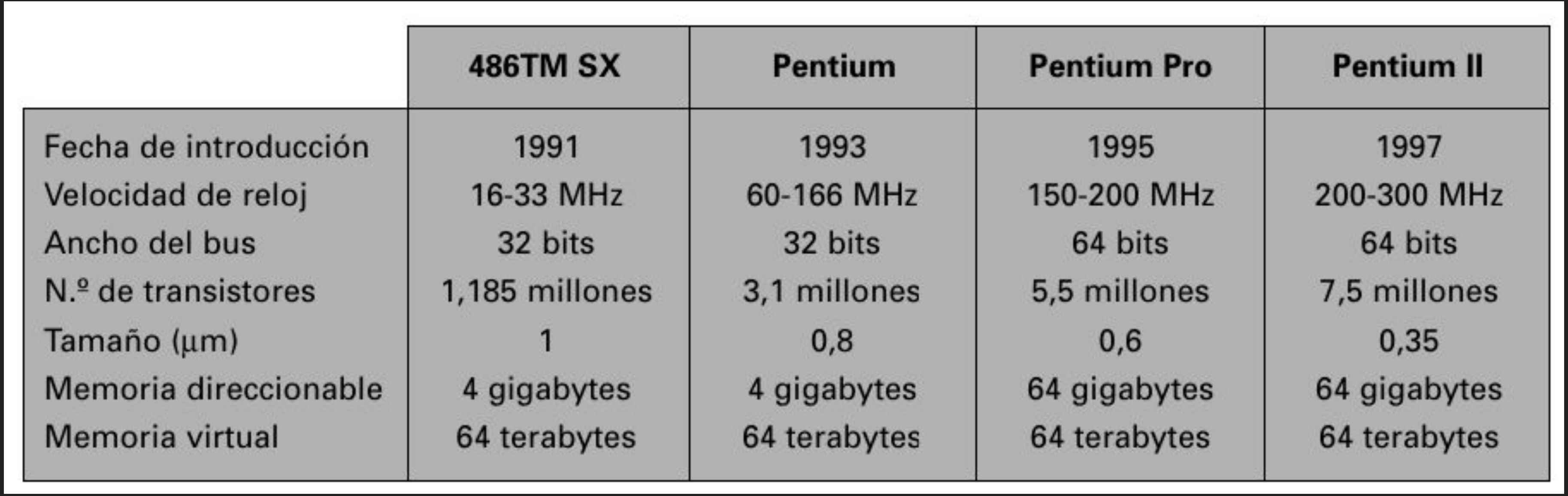
- 1971 - Intel 4004 - 4 bits
- 1972 - Intel 8008 - 8 bits
- 1974 - Intel 8080 - 8 bits
- 1978 - Intel 8086 - 16 bits
- 1982 - Intel 80286 - 16 bits
- 1985 - Intel 386 - 32 bits
- ….



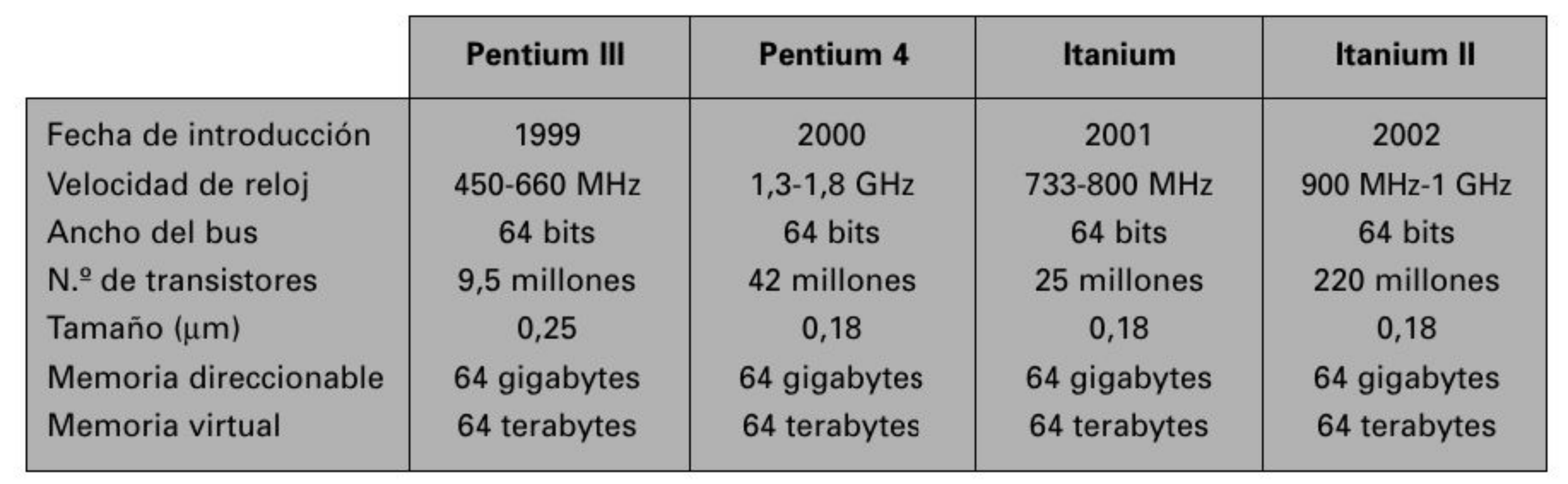
	4004	8008	8080	8086	8088
Fecha de introducción	1971	1972	1974	1978	1979
Velocidad de reloj	108 kHz	108 kHz	2 MHz	5 MHz, 8 MHz, 10 MHz	5 MHz, 8 MHz
Ancho del bus	4 bits	8 bits	8 bits	16 bits	8 bits
N.º de transistores	2.300	3.500	6.500	29.000	29.000
Tamaño (µm)	10	—	6	3	6
Memoria direccionable	640 Bytes	16 KB	64 KB	1 MB	1 MB
Memoria virtual	—	—	—	—	—



	80286	386TM DX	386TM SX	486TM DX CPU
Fecha de introducción	1982	1985	1988	1989
Velocidad de reloj	6-12,5 MHz	16-33 MHz	16-33 MHz	25-50 MHz
Ancho del bus	16 bits	32 bits	16 bits	32 bits
N.º de transistores	134.000	275.000	275.000	1,2 millones
Tamaño (µm)	1,5	1	1	0,8-1
Memoria direccionable	16 megabytes	4 gigabytes	16 megabytes	4 gigabytes
Memoria virtual	1 gigabyte	64 terabytes	64 terabytes	64 terabytes



	486TM SX	Pentium	Pentium Pro	Pentium II
Fecha de introducción	1991	1993	1995	1997
Velocidad de reloj	16-33 MHz	60-166 MHz	150-200 MHz	200-300 MHz
Ancho del bus	32 bits	32 bits	64 bits	64 bits
N.º de transistores	1,185 millones	3,1 millones	5,5 millones	7,5 millones
Tamaño (μm)	1	0,8	0,6	0,35
Memoria direccionable	4 gigabytes	4 gigabytes	64 gigabytes	64 gigabytes
Memoria virtual	64 terabytes	64 terabytes	64 terabytes	64 terabytes



	Pentium III	Pentium 4	Itanium	Itanium II
Fecha de introducción	1999	2000	2001	2002
Velocidad de reloj	450-660 MHz	1,3-1,8 GHz	733-800 MHz	900 MHz-1 GHz
Ancho del bus	64 bits	64 bits	64 bits	64 bits
N.º de transistores	9,5 millones	42 millones	25 millones	220 millones
Tamaño (µm)	0,25	0,18	0,18	0,18
Memoria direccionable	64 gigabytes	64 gigabytes	64 gigabytes	64 gigabytes
Memoria virtual	64 terabytes	64 terabytes	64 terabytes	64 terabytes

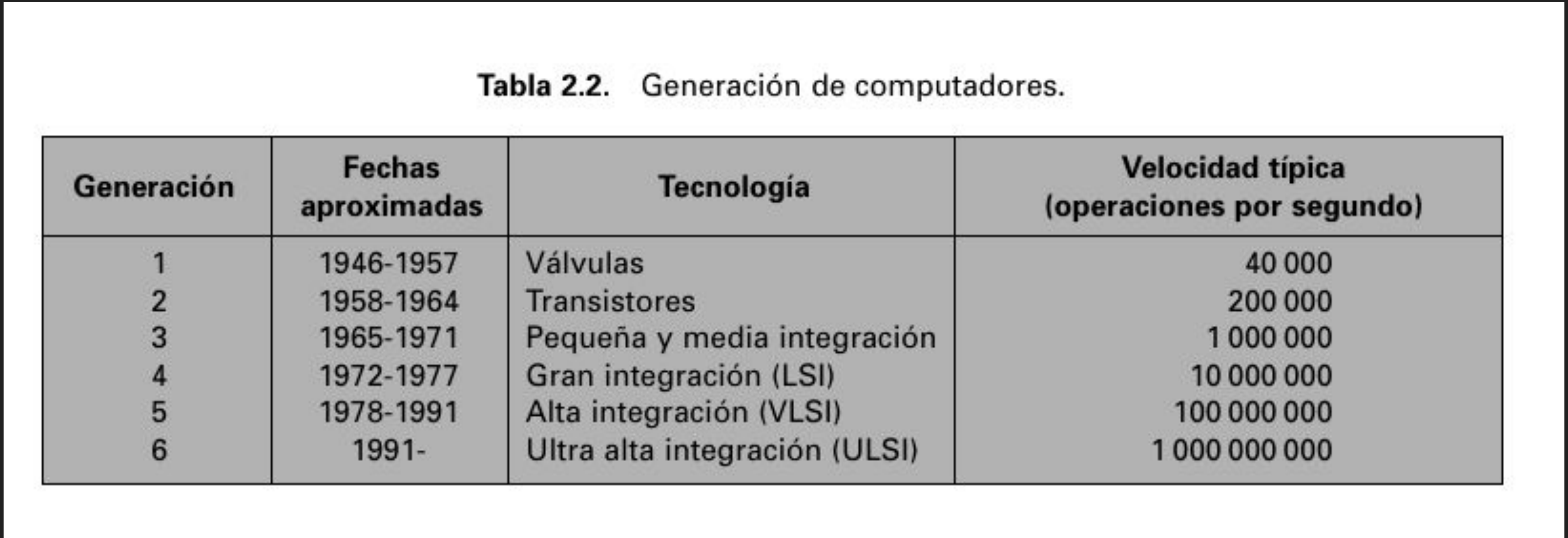


Tabla 2.2. Generación de computadores.

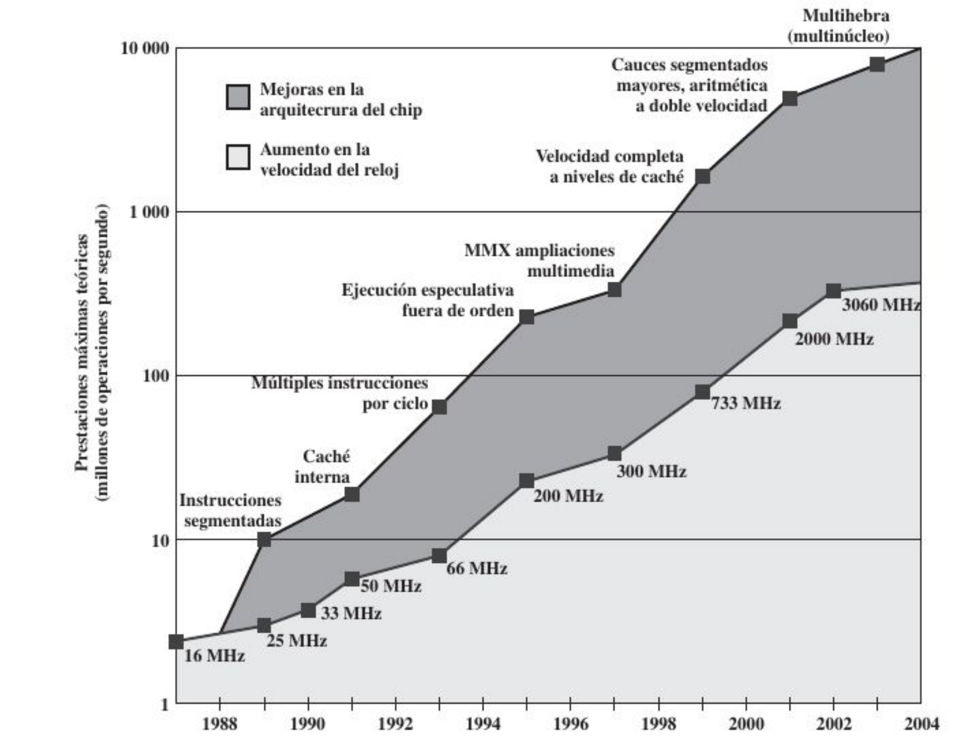
Generación	Fechas aproximadas	Tecnología	Velocidad típica (operaciones por segundo)
1	1946-1957	Válvulas	40 000
2	1958-1964	Transistores	200 000
3	1965-1971	Pequeña y media integración	1 000 000
4	1972-1977	Gran integración (LSI)	10 000 000
5	1978-1991	Alta integración (VLSI)	100 000 000
6	1991-	Ultra alta integración (ULSI)	1 000 000 000

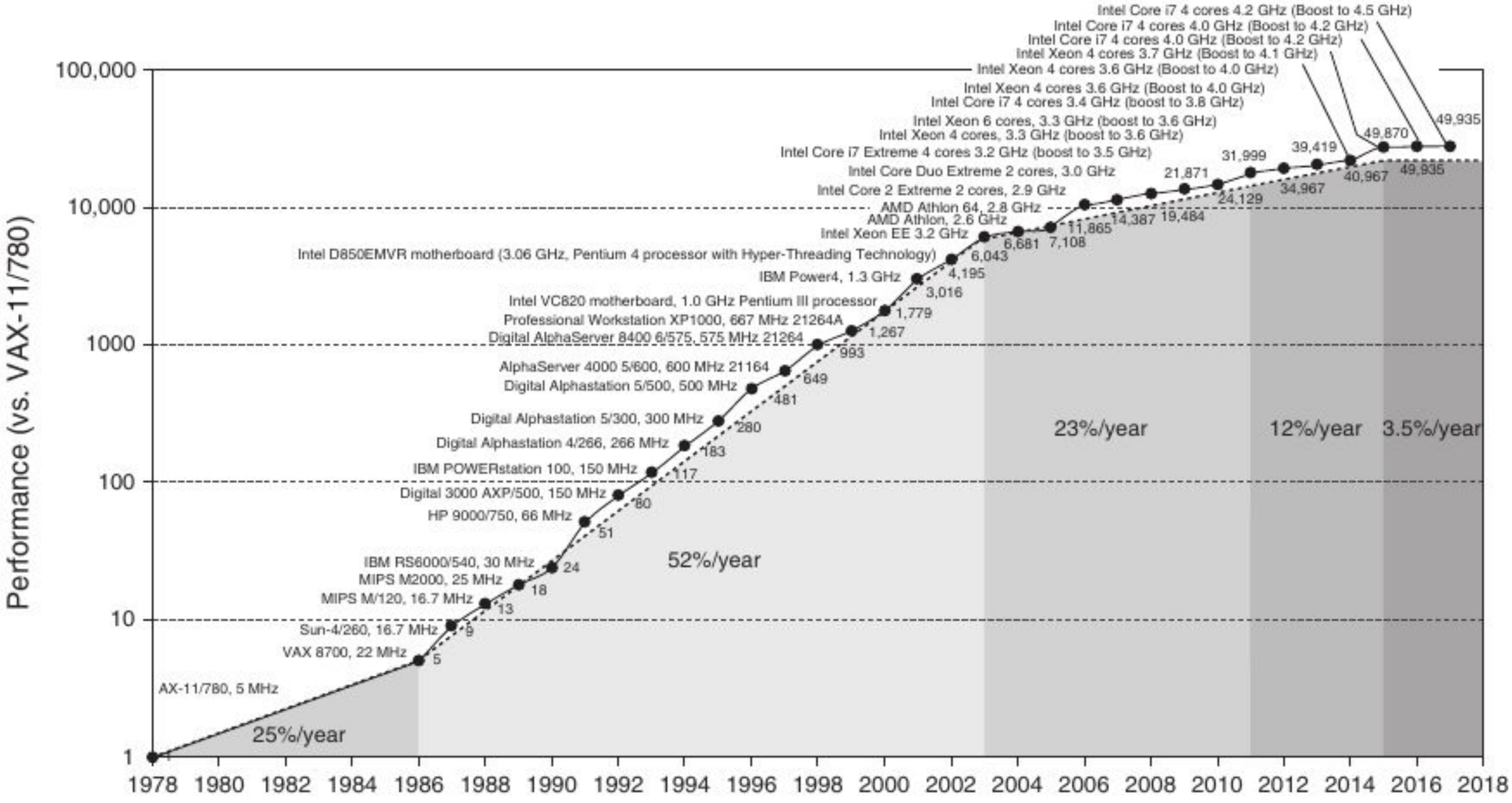
Computer Architecture

- ISA (Instruction set architecture)
    - programmer's interface
    - classes: register-memory, load-store
    - memory addressing
    - addressing modes
    - type and size operands
    - operations
    - control flow instructions
    - encoding
- HW components
- Organization (micro-architecture)

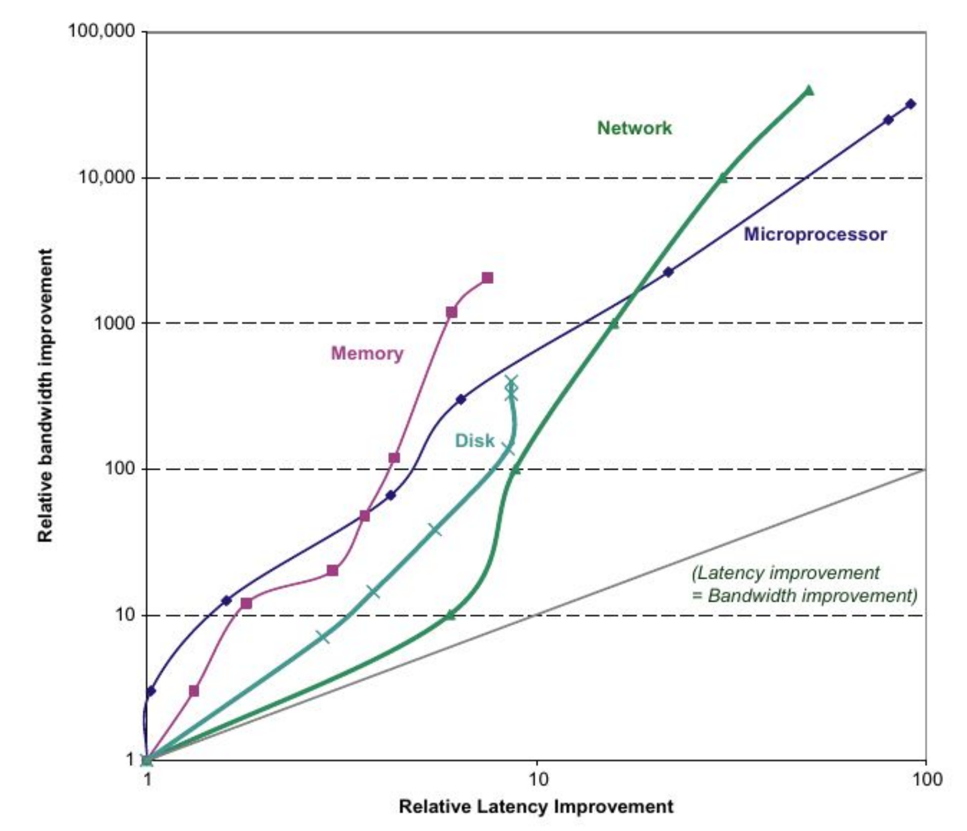
Trends in performance

- processors performance:
  - Technology size / clock frequency
    - power density
    - Delay
  - micro-architecture techniques

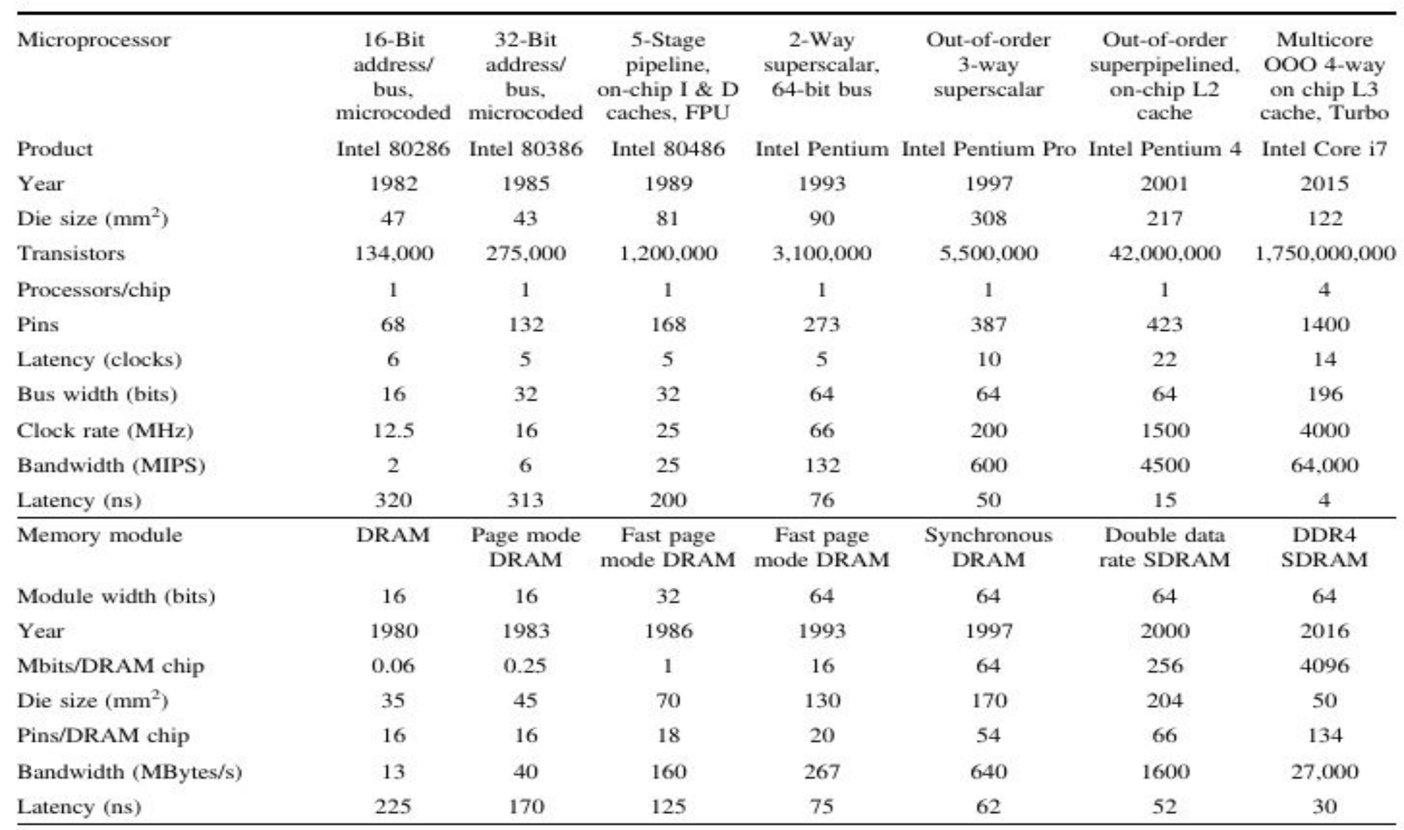



Intel Core i7 4 cores 4.2 GHz (Boost to 4.5 GHz)
Intel Core i7 4 cores 4.0 GHz (Boost to 4.2 GHz)
Intel Core i7 4 cores 4.0 GHz (Boost to 4.2 GHz)
Intel Xeon 4 cores 3.7 GHz (Boost to 4.1 GHz)
Intel Xeon 4 cores 3.6 GHz (Boost to 4.0 GHz)
Intel Xeon 4 cores 3.6 GHz (Boost to 4.0 GHz)
Intel Core i7 4 cores 3.4 GHz (boost to 3.8 GHz)
Intel Xeon 6 cores, 3.3 GHz (boost to 3.6 GHz)
Intel Xeon 4 cores, 3.3 GHz (boost to 3.6 GHz)
Intel Core i7 Extreme 4 cores 3.2 GHz (boost to 3.5 GHz)
Intel Core Duo Extreme 2 cores, 3.0 GHz
Intel Core 2 Extreme 2 cores, 2.9 GHz
AMD Athlon 64, 2.8 GHz
AMD Athlon, 2.6 GHz
Intel Xeon EE 3.2 GHz
Intel D850EMVR motherboard (3.06 GHz, Pentium 4 processor with Hyper-Threading Technology)
IBM Power4, 1.3 GHz
Intel VC820 motherboard, 1.0 GHz Pentium III processor
Professional Workstation XP1000, 667 MHz 21264A
Digital AlphaServer 8400 6/575, 575 MHz 21264
AlphaServer 4000 5/600, 600 MHz 21164
Digital Alphastation 5/500, 500 MHz
Digital Alphastation 5/300, 300 MHz
Digital Alphastation 4/266, 266 MHz
IBM POWERstation 100, 150 MHz
Digital 3000 AXP/500, 150 MHz
HP 9000/750, 66 MHz
IBM RS6000/540, 30 MHz
MIPS M2000, 25 MHz
MIPS M/120, 16.7 MHz
Sun-4/260, 16.7 MHz
VAX 8700, 22 MHz
AX-11/780, 5 MHz
49,935
49,870
39,419
31,999
21,871
40,967
49,935
34,967
24,129
14,387
19,484
11,865
6,681
7,108
6,043
4,195
3,016
1,779
1,267
993
649
481
280
183
117
80
51
24
18
13
9
5
25%/year
52%/year
23%/year
12%/year
3.5%/year
Performance (vs. VAX-11/780)
100,000
10,000
1000
100
10
1
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018

Trends in Technology: bandwidth vs latency

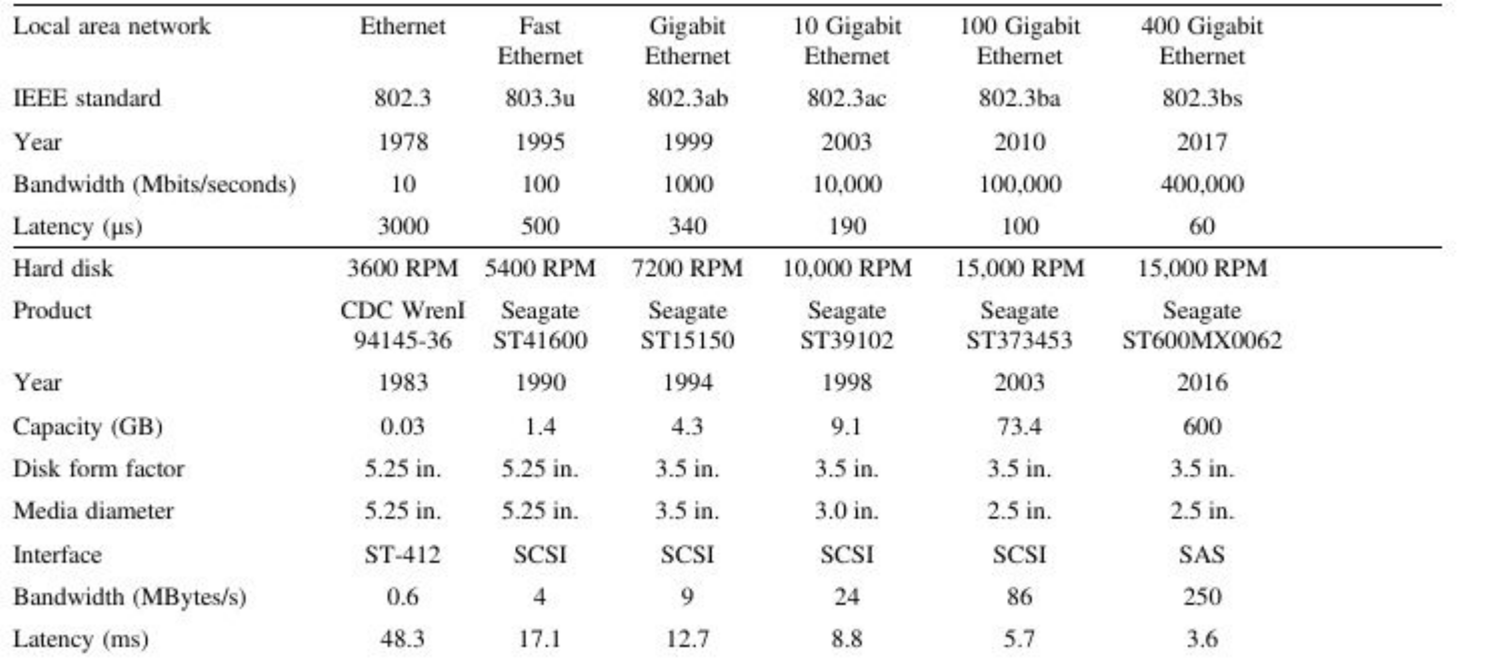


Trends in technology: scaling of transistors



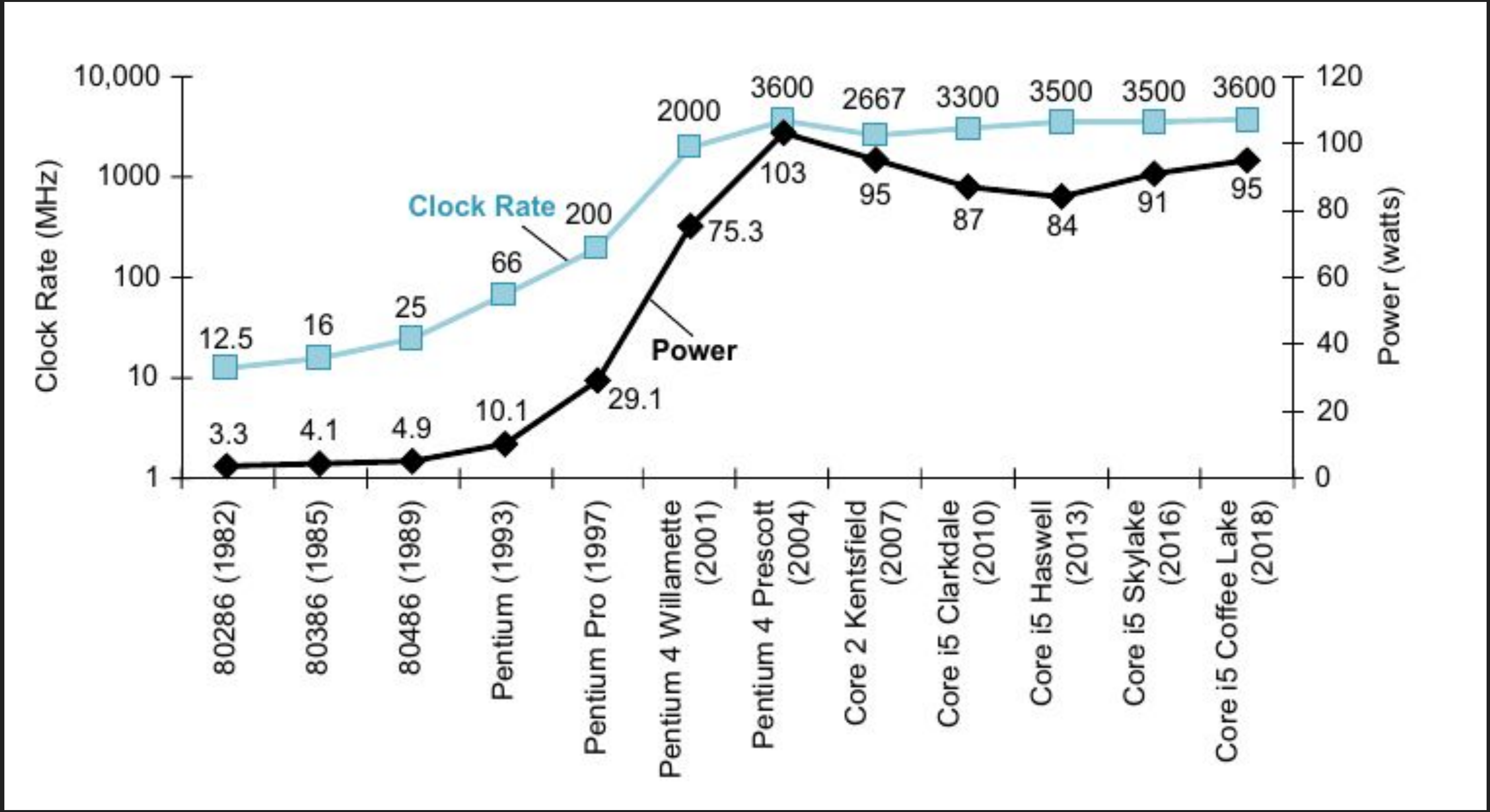
Microprocessor	16-Bit address/ bus, microcoded	32-Bit address/ bus, microcoded	5-Stage pipeline, on-chip I & D caches, FPU	2-Way superscalar, 64-bit bus	Out-of-order 3-way superscalar	Out-of-order superpipelined, on-chip L2 cache	Multicore OOO 4-way on chip L3 cache, Turbo
Product	Intel 80286	Intel 80386	Intel 80486	Intel Pentium	Intel Pentium Pro	Intel Pentium 4	Intel Core i7
Year	1982	1985	1989	1993	1997	2001	2015
Die size (mm^2)	47	43	81	90	308	217	122
Transistors	134,000	275,000	1,200,000	3,100,000	5,500,000	42,000,000	1,750,000,000
Processors/chip	1	1	1	1	1	1	4
Pins	68	132	168	273	387	423	1400
Latency (clocks)	6	5	5	5	10	22	14
Bus width (bits)	16	32	32	64	64	64	196
Clock rate (MHz)	12.5	16	25	66	200	1500	4000
Bandwidth (MIPS)	2	6	25	132	600	4500	64,000
Latency (ns)	320	313	200	76	50	15	4
Memory module	DRAM	Page mode DRAM	Fast page mode DRAM	Fast page mode DRAM	Synchronous DRAM	Double data rate SDRAM	DDR4 SDRAM
Module width (bits)	16	16	32	64	64	64	64
Year	1980	1983	1986	1993	1997	2000	2016
Mbits/DRAM chip	0.06	0.25	1	16	64	256	4096
Die size (mm^2)	35	45	70	130	170	204	50
Pins/DRAM chip	16	16	18	20	54	66	134
Bandwidth (MBytes/s)	13	40	160	267	640	1600	27,000
Latency (ns)	225	170	125	75	62	52	30

Trends in technology: scaling of transistors



Local area network	Ethernet	Fast Ethernet	Gigabit Ethernet	10 Gigabit Ethernet	100 Gigabit Ethernet	400 Gigabit Ethernet
IEEE standard	802.3	803.3u	802.3ab	802.3ac	802.3ba	802.3bs
Year	1978	1995	1999	2003	2010	2017
Bandwidth (Mbits/seconds)	10	100	1000	10,000	100,000	400,000
Latency (µs)	3000	500	340	190	100	60
Hard disk	3600 RPM	5400 RPM	7200 RPM	10,000 RPM	15,000 RPM	15,000 RPM
Product	CDC WrenI 94145-36	Seagate ST41600	Seagate ST15150	Seagate ST39102	Seagate ST373453	Seagate ST600MX0062
Year	1983	1990	1994	1998	2003	2016
Capacity (GB)	0.03	1.4	4.3	9.1	73.4	600
Disk form factor	5.25 in.	5.25 in.	3.5 in.	3.5 in.	3.5 in.	3.5 in.
Media diameter	5.25 in.	5.25 in.	3.5 in.	3.0 in.	2.5 in.	2.5 in.
Interface	ST-412	SCSI	SCSI	SCSI	SCSI	SAS
Bandwidth (MBytes/s)	0.6	4	9	24	86	250
Latency (ms)	48.3	17.1	12.7	8.8	5.7	3.6

Trends in Power and Energy

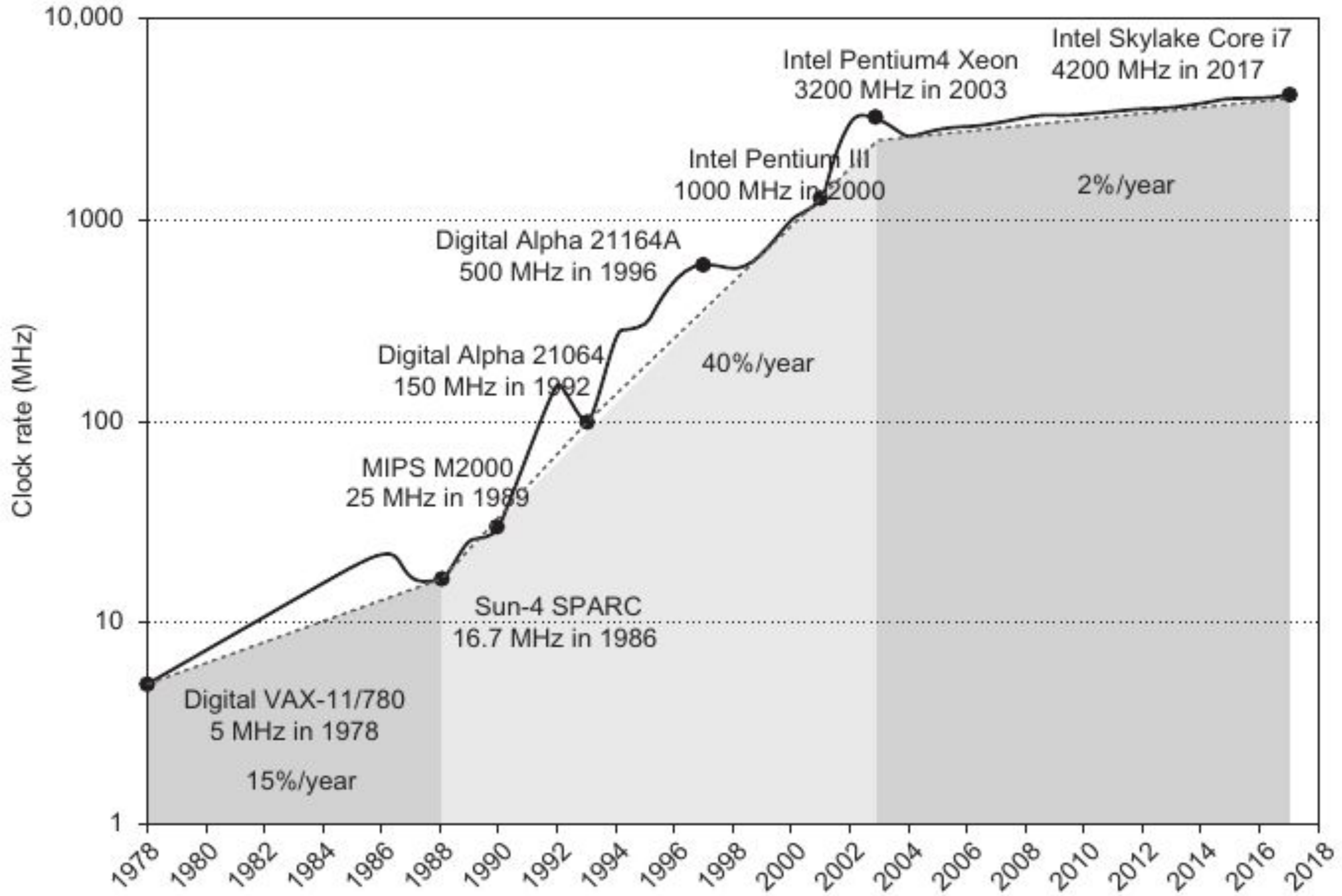


Energy d = C x V2   (0->1->0; 1->0->1)

Energy d = ½ x C x V2 (0->1; 1->0)

Power d = ½ x C x V2 x Freq

- Intel 80386 used about 2 W, whereas a 4.0 GHz Intel Core i7-6700K consumes 95 W.
- Given that this heat must be dissipated from a chip that is about 1.5 cm on a side, we are near the limit of what can be cooled by air, and this is where we have been stuck for nearly a decade.


10,000
1000
100
10
1
Clock rate (MHz)
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Intel Skylake Core i7
4200 MHz in 2017
Intel Pentium4 Xeon
3200 MHz in 2003
Intel Pentium III
1000 MHz in 2000
2%/year
Digital Alpha 21164A
500 MHz in 1996
Digital Alpha 21064
150 MHz in 1992
40%/year
MIPS M2000
25 MHz in 1989
Sun-4 SPARC
16.7 MHz in 1986
Digital VAX-11/780
5 MHz in 1978
15%/year

techniques to improve energy efficiency

- Do nothing well
- Dynamic voltage-frequency scaling
- Design for the typical case
- Overclocking

Static power

Power s = Current static x Voltage

- leakage current flows even when a transistor is off
- static power is proportional to the number of devices.

→power gating

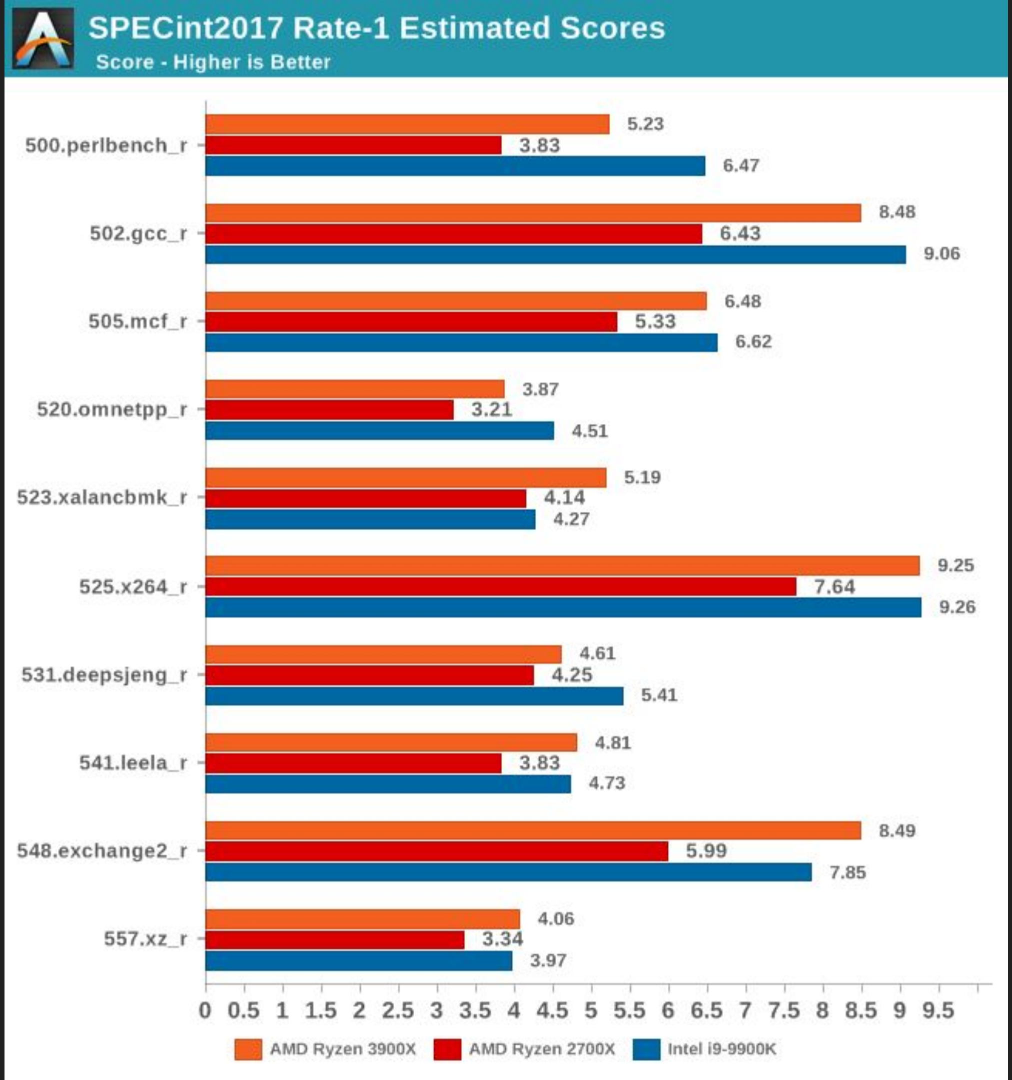
Measuring performance

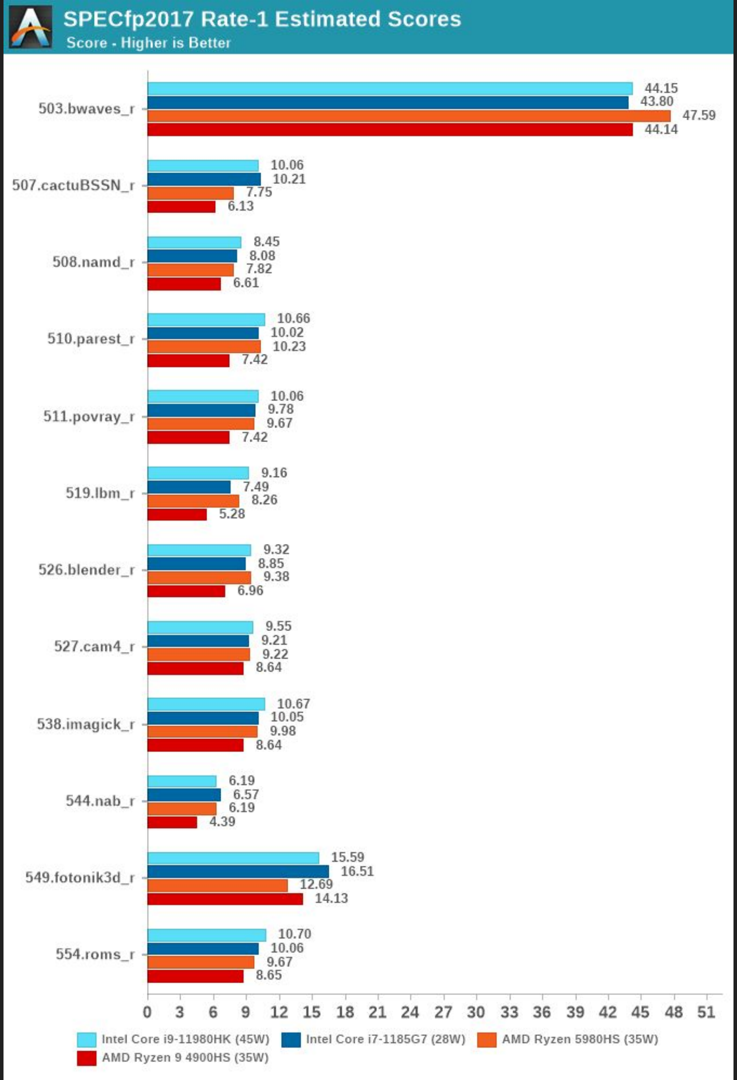
When we say one computer is faster than another one is, what do we mean?

execution time? throughput?

Benchmarks

- Programs, which are programs that many companies use to establish the relative performance of their computers.
- SPEC (Standard Performance Evaluation Corporation) http://www.spec.org.
- SPEC CPU2017 consists of a set of 10 integer benchmarks (CINT2017) and 17 floating-point benchmarks (CFP2017).
- real programs modified to be portable and to minimize the effect of I/O on performance. The integer benchmarks vary from part of a C compiler to a go program to a video compression. The floating-point benchmarks include molecular dynamics, ray tracing, and weather forecasting


SPECint2017 Rate-1 Estimated Scores
Score - Higher is Better
500.perlbench_r
5.23
3.83
6.47
502.gcc_r
8.48
6.43
9.06
505.mcf_r
6.48
5.33
6.62
520.omnetpp_r
3.87
3.21
4.51
523.xalancbmk_r
5.19
4.14
4.27
525.x264_r
9.25
7.64
9.26
531.deepsjeng_r
4.61
4.25
5.41
541.leela_r
4.81
3.83
4.73
548.exchange2_r
8.49
5.99
7.85
557.xz_r
4.06
3.34
3.97
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5
AMD Ryzen 3900X
AMD Ryzen 2700X
Intel i9-9900K

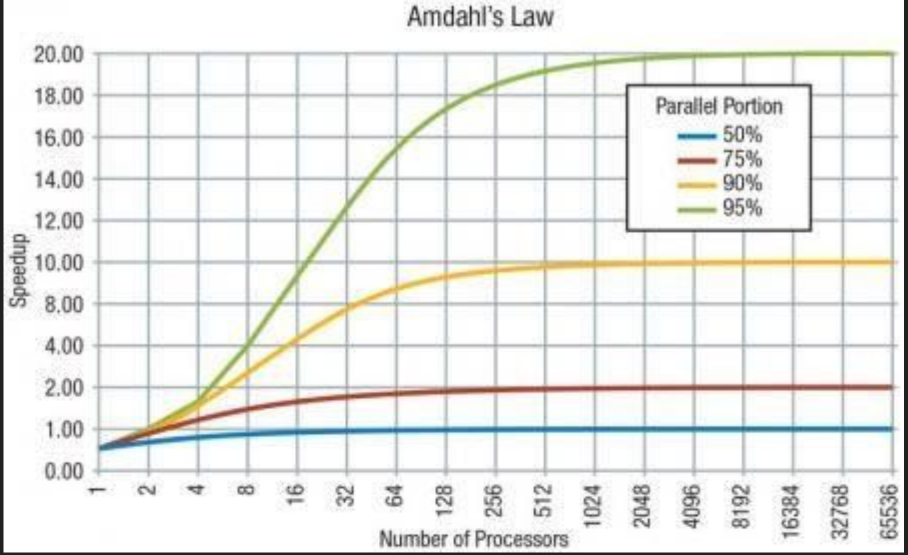

SPECfp2017 Rate-1 Estimated Scores
Score - Higher is Better
503.bwaves_r
44.15
43.80
47.59
44.14
507.cactuBSSN_r
10.06
10.21
7.75
6.13
508.namd_r
8.45
8.08
7.82
6.61
510.parest_r
10.66
10.02
10.23
7.42
511.povray_r
10.06
9.78
9.67
7.42
519.lbm_r
9.16
7.49
8.26
5.28
526.blender_r
9.32
8.85
9.38
6.96
527.cam4_r
9.55
9.21
9.22
8.64
538.imagick_r
10.67
10.05
9.98
8.64
544.nab_r
6.19
6.57
6.19
4.39
549.fotonik3d_r
15.59
16.51
12.69
14.13
554.roms_r
10.70
10.06
9.67
8.65
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Intel Core i9-11980HK (45W)
Intel Core i7-1185G7 (28W)
AMD Ryzen 5980HS (35W)
AMD Ryzen 9 4900HS (35W)

Amdahl’s Law

- The performance gain that can be obtained by improving some portion of a computer can be calculated using Amdahl’s Law.
- The performance improvement to be gained from using some faster mode of execution is limited by the fraction of the time the faster mode can be used.

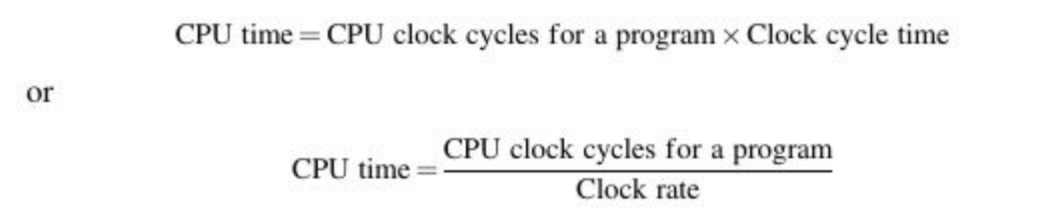
$$\text{Speedup}_{\text{overall}} = \frac{\text{Execution time}_{\text{old}}}{\text{Execution time}_{\text{new}}} = \frac{1}{\left(1 - \text{Fraction}_{\text{enhanced}}\right) + \dfrac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

- if an enhancement is usable only for a fraction of a task, then we can’t speed up the task by more than the reciprocal of 1 minus that fraction.


Amdahl's Law
20.00
18.00
16.00
14.00
12.00
10.00
8.00
4.00
2.00
1.00
0.00
Speedup
1
2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
Number of Processors
Parallel Portion
50%
75%
90%
95%

Processor Performance Equation

- CPU time for a program can be expressed in two ways:

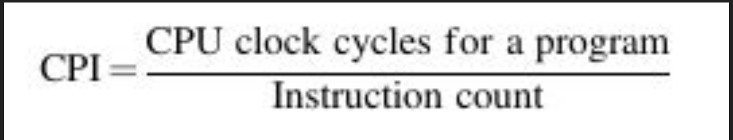


$$\text{CPU time} = \text{CPU clock cycles for a program} \times \text{Clock cycle time}$$

or

$$\text{CPU time} = \frac{\text{CPU clock cycles for a program}}{\text{Clock rate}}$$

- If we know the number of clock cycles and the instruction count, we can calculate the average number of clock cycles per instruction (CPI).



$$\text{CPI} = \frac{\text{CPU clock cycles for a program}}{\text{Instruction count}}$$

- transposing the instruction count in the preceding formula, clock cycles can be defined as IC CPI.

$$\text{CPU time} = \text{Instruction count} \times \text{Cycles per instruction} \times \text{Clock cycle time}$$

- Expanding the first formula into the units of measurement shows how the pieces fit together:

$$\frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Clock cycles}}{\text{Instruction}} \times \frac{\text{Seconds}}{\text{Clock cycle}} = \frac{\text{Seconds}}{\text{Program}} = \text{CPU time}$$

- processor performance is dependent upon three characteristics: clock cycle (or rate), clock cycles per instruction, and instruction