MATA48
Arquitetura de Computadores

AULA 01
Visão geral do projeto de arquiteturas de computadores

Prof. Marcos E. Barreto
UFBA / DCC - [2019.2]

Roteiro

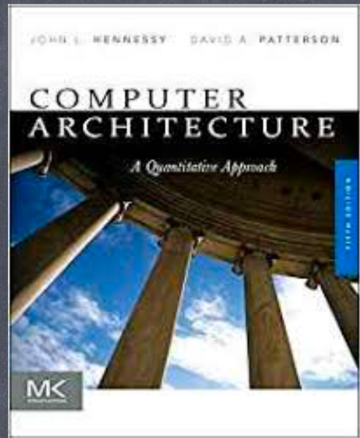
- Conceitos iniciais
- Aspectos históricos e evolutivos
- Muito breve introdução ao Conjunto de Instruções (ISA)
- Aspectos de desempenho
- Questões de projeto

Referências

- PATTERSON, David; HENNESSY, John. Organização e projeto de computadores: a interface hardware/software. Rio de Janeiro: Elsevier, 2017.

- HENNESSY, J. L.; PATTERSON, D. A. Computer architecture - a quantitative approach. 5 ed. Waltham: Morgan Kaufmann, 2012.

- WEBER, Raul Fernando. Fundamentos de arquitetura de computadores. 4 ed. Porto Alegre: Bookman, 2012.



Categorias (classes) de computadores

- Quais características são importantes para um novo sistema computacional?

- Depende do foco do projeto: i) Alto desempenho; ii) Alta confiabilidade (disponibilidade); iii) Baixo consumo.

- 3 classes: PCs (desktops), servidores e sistemas embarcados.

Feature	Personal mobile device (PMD)	Desktop	Server	Clusters/warehouse-scale computer	Embedded
Price of system	\$100–\$1000	\$300–\$2500	\$5000–\$10,000,000	\$100,000–\$200,000,000	\$10–\$100,000
Price of micro-processor	\$10–\$100	\$50–\$500	\$200–\$2000	\$50–\$250	\$0.01–\$100
Critical system design issues	Cost, energy, media performance, responsiveness	Price-performance, energy, graphics performance	Throughput, availability, scalability, energy	Price-performance, throughput, energy proportionality	Price, energy, application-specific performance

Figure 1.2 A summary of the five mainstream computing classes and their system characteristics. Sales in 2010

List of computer size categories

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(October 2013) (Learn how and when to remove this template message)*

This **list of computer size categories** attempts to list commonly used categories of computer by the physical size of the device and its chassis or case, in descending order of size. One generation's "supercomputer" is the next generation's "mainframe", and a "PDA" does not have the same set of functions as a "laptop", but the list still has value, as it provides a ranked categorization of devices. It also ranks some more obscure computer sizes.

This illustration shows the development of computers, from huge room-sized computers in the 1940s and 1950s (left-most photo) to smaller personal computers over the 1980s to the 2010s.

Contents [hide]

Computer form factors

Listed by PCB size (mm)

WTX (356×425) · AT (350×305) · Baby-AT (330×216) · BTX (325×267) · ATX (305×244) · SSI CEB (305×267) · EATX (Extended) (305×330) · LPX (330×229) · microBTX (264×267) · NLX (254×228) · Ultra ATX (367×244) · microATX (244×244) · DTX (203×244) · FlexATX (229×191) · Mini-DTX (203×170) · EBX (203×146) · Mini-ITX (170×170) · EPIC (Express) (165×115) · Mini ATX (150×150) · ESM (149×71) · Mini-STX (140×147) · Nano-ITX (120×120) · COM Express (125×95) · ESMexpress (125×95) · ETX (114×95) · XTX (114×95) · NUC (102×102) · Pico-ITX (100×72) · PC/104 *(-Plus)* (96×90) · ESMini (95×55) · SMARC (82×80) · Qseven (70×70) · mobile-ITX (60×60) · CoreExpress (58×65)

V · T · E

Supercomputers [edit]

- Minisupercomputer

Mainframe [edit]

Mainframe computers are large and expensive but powerful, so they can handle hundreds and thousands of connected users at the same time.

Minicomputers [edit]

- Superminicomputer

Definindo arquitetura...

- Determinar quais atributos são importantes para um novo sistema computacional.

- Projetar um sistema que maximize o desempenho e a eficiência energética levando em consideração restrições relativas a custo, consumo (energia) e confiabilidade / disponibilidade.

Elementos básicos

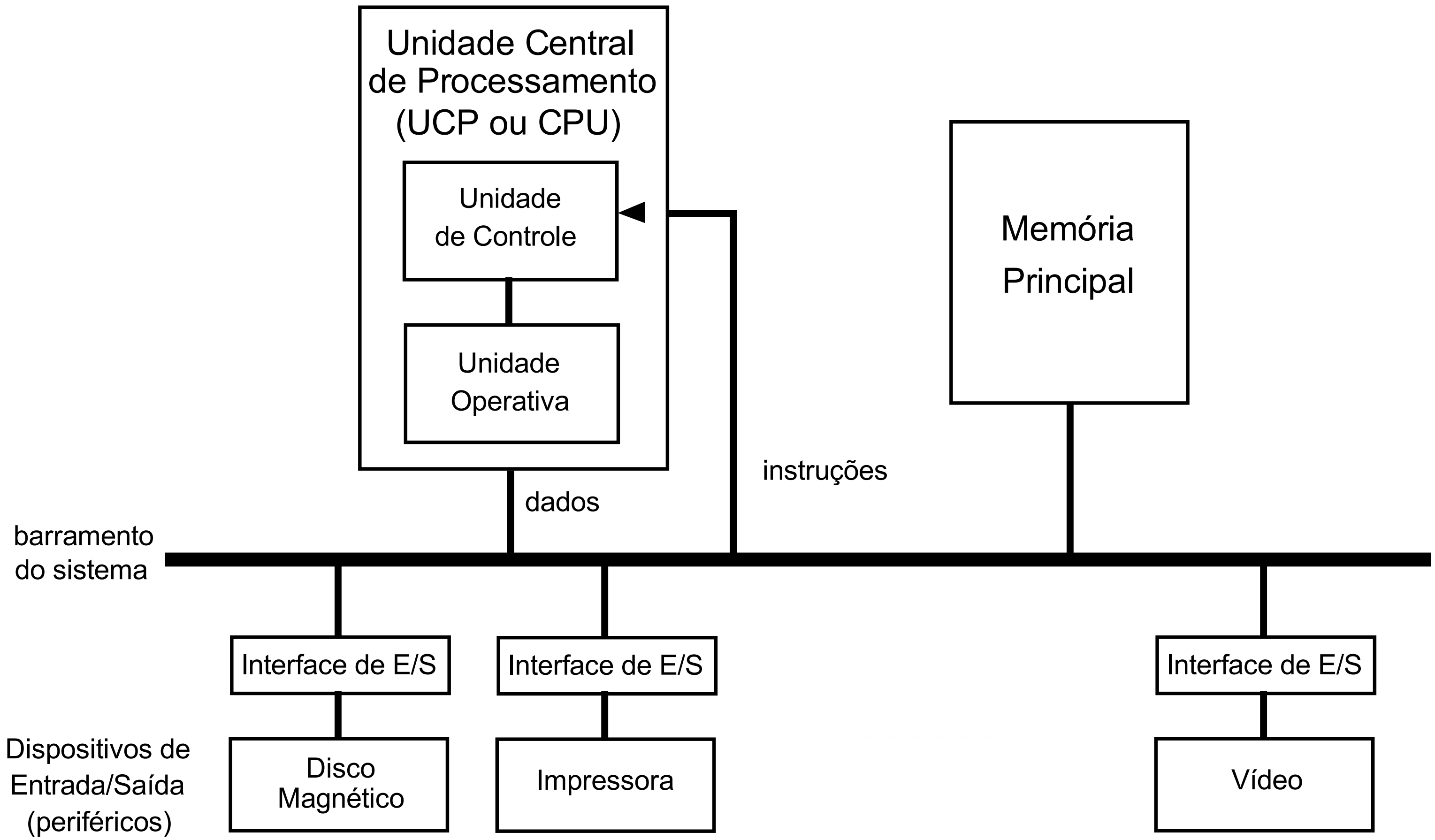
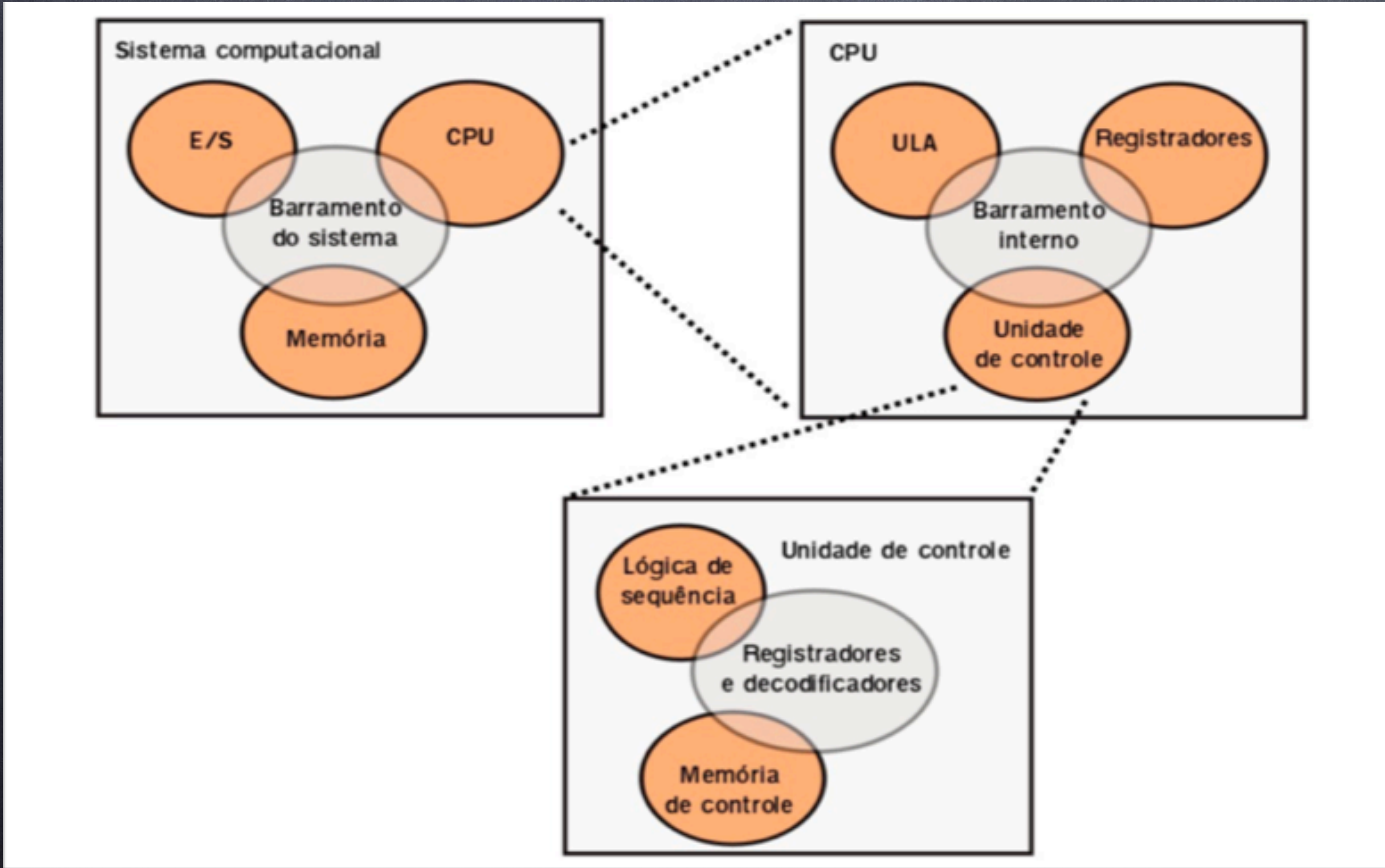


Figura 1 - Diagrama de blocos de um computador com arquitetura convencional.

Elementos básicos

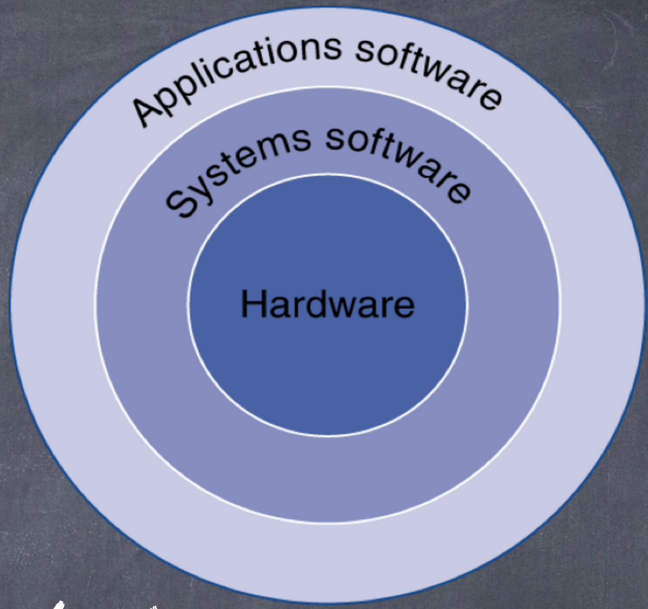
- Componentes e suas inter-relações



Visão hierárquica do sistema computacional

- Arquitetura X Organização
  - Aspectos visíveis ao programador X detalhes de implementação
- FUNÇÃO X ESTRUTURA
  - Processamento de dados. Processador
  - Armazenamento de dados. Memória
  - Movimentação de dados. Controladores
  - Controle

Níveis de abstração



- Software de aplicação
  - Escritos em linguagem de alto nível.
- Software de sistema
  - Compiladores, sistemas operacionais, bibliotecas específicas.
- Hardware
  - processador, memória, controladores etc.

Níveis de abstração

High-level language program (in C)

```
swap(int v[], int k)
{int temp;
   temp = v[k];
   v[k] = v[k+1];
   v[k+1] = temp;
}
```

Compiler

Assembly language program (for MIPS)

```
swap:
      multi $2, $5,4
      add   $2, $4,$2
      lw    $15, 0($2)
      lw    $16, 4($2)
      sw    $16, 0($2)
      sw    $15, 4($2)
      jr    $31
```

Assembler

Interface software / hardware

Binary machine language program (for MIPS)

```
00000000101000100000000100011000
0000000010000010000100000100001
10001101111000100000000000000000
10001110000100100000000000000100
10101110000100100000000000000000
10101101111000100000000000000100
00000011111000000000000000001000
```

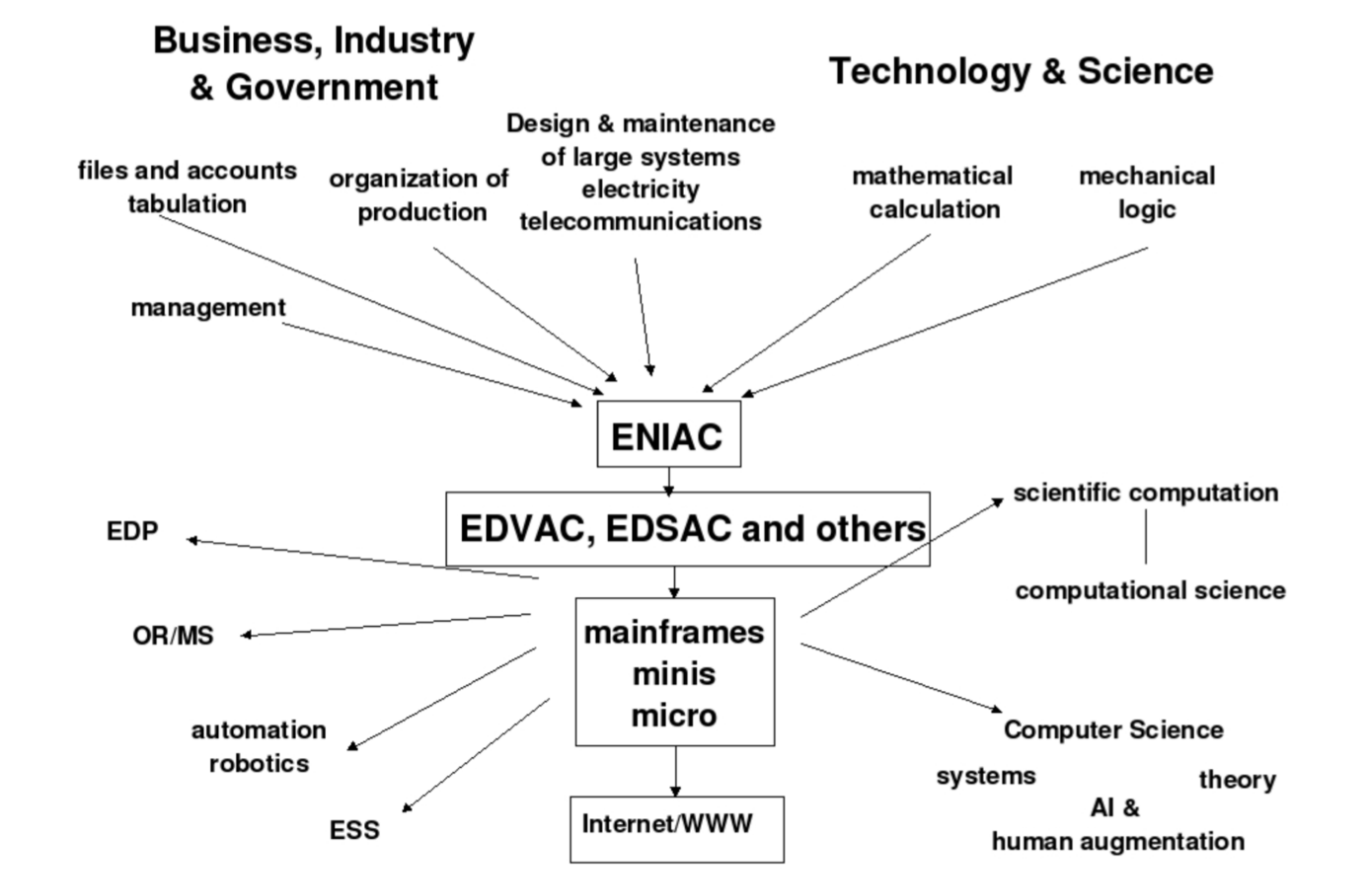
Aspectos históricos

- Gerações

Year	Technology used in computers	Relative performance/unit cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit	900
1995	Very large-scale integrated circuit	2,400,000
2013	Ultra large-scale integrated circuit	6,200,000,000

FIGURE 1.10 Relative performance per unit cost of technologies used in computers over time. Source: Computer Museum, Boston, with 2013 extrapolated by the authors. See **Section 1.12**.

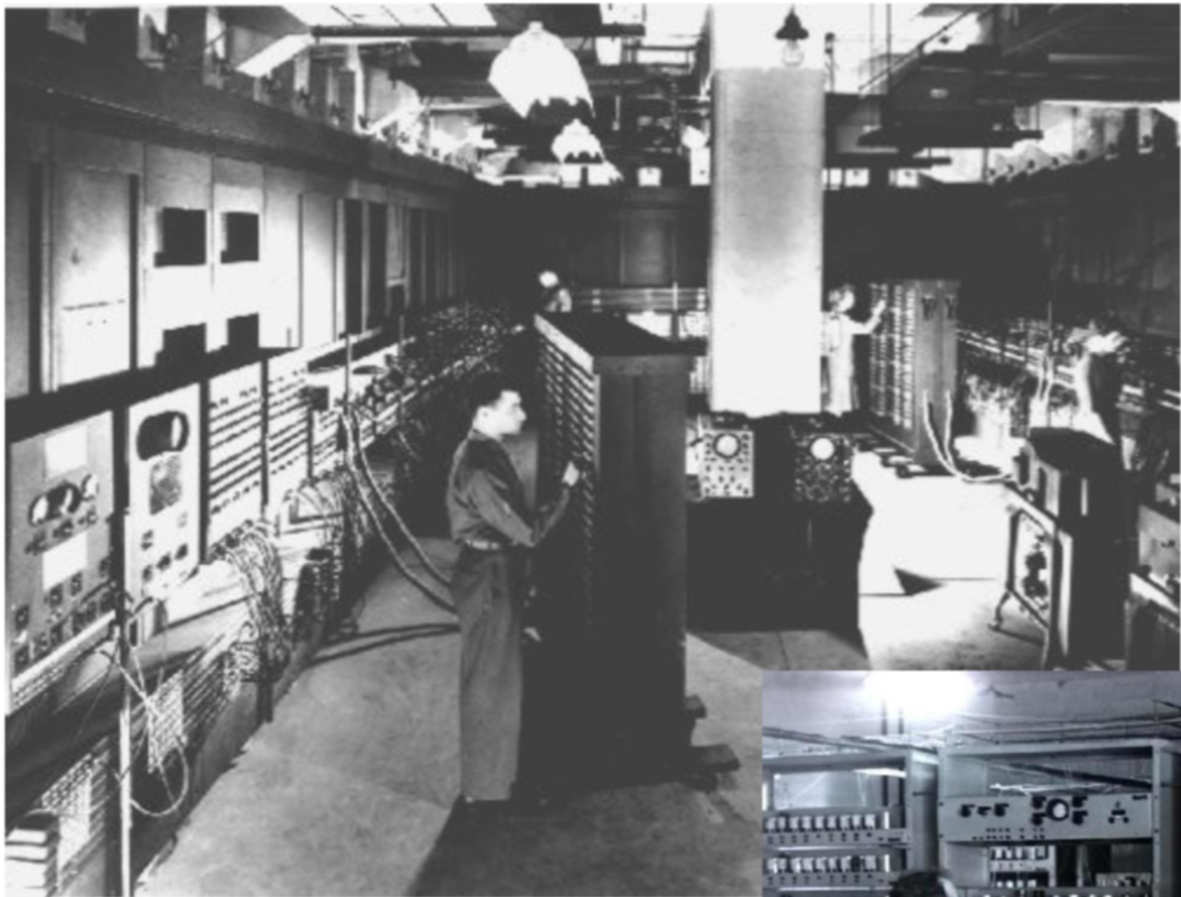
Aspectos históricos



ENIAC (*Electronic Numerical Integrator and Computer*)
Projeto: 1943 – 1946
Operação: até 1955

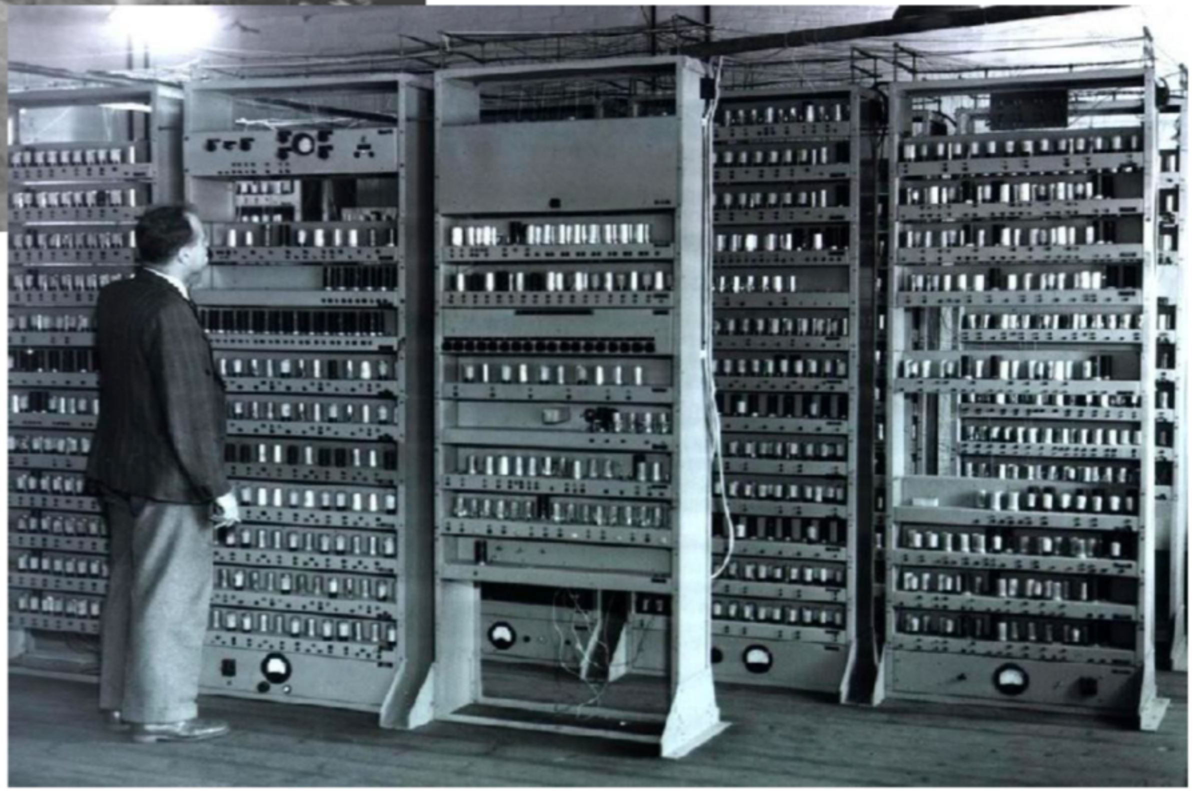
Sua memória podia registrar até 20 números de 10 dígitos cada um.

Primeiro computador digital eletrônico



ENIAC

Projeto: 1943 – 1946
Operação: até 1955



EDSAC

Projeto: 1949
Operação: até 1958

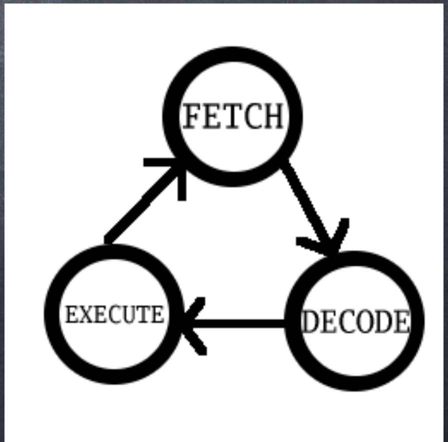
Aspectos históricos

- Leituras complementares

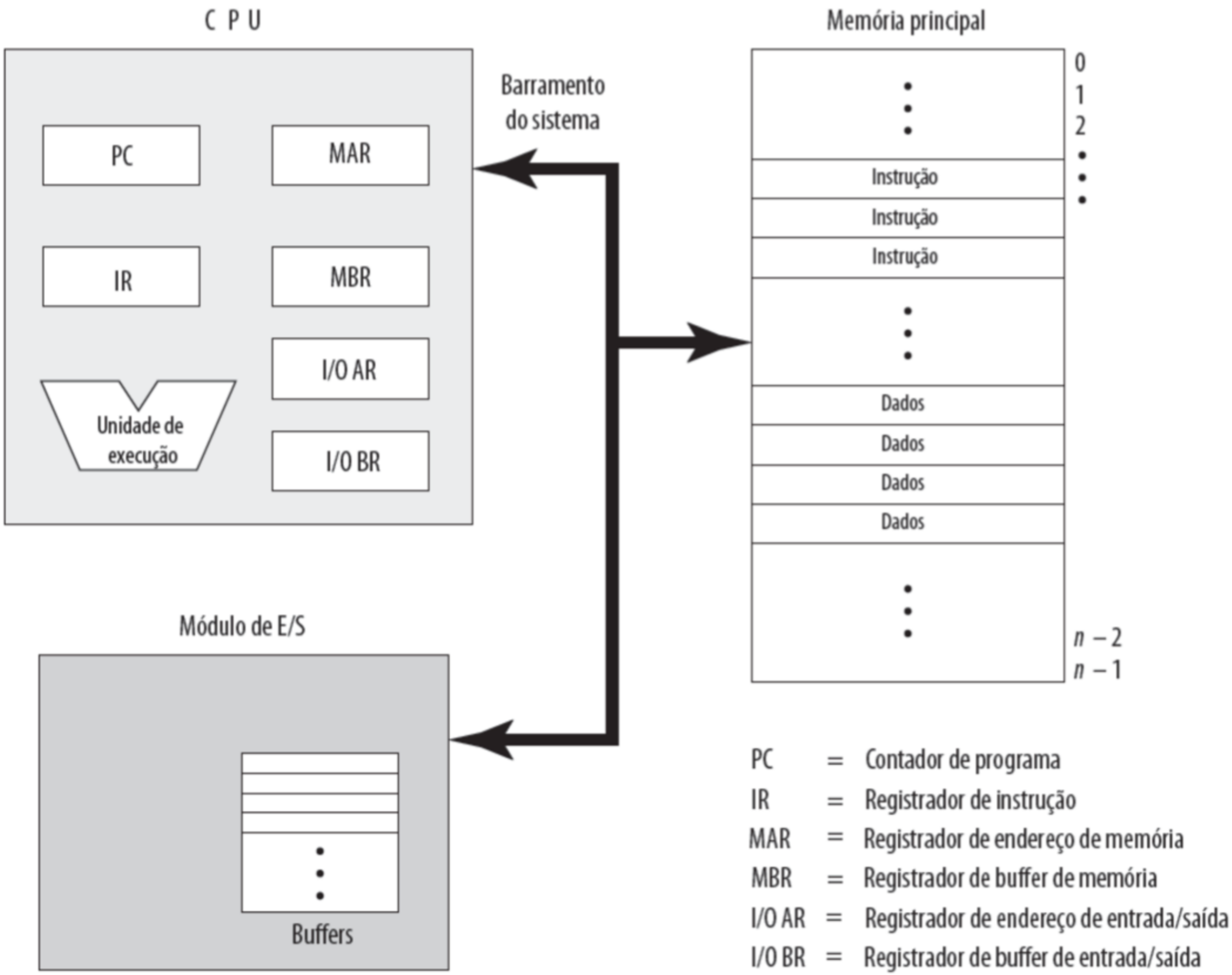
  - Seção 1.12 – Historical perspective.
  - Artigo: The histories of computing(s).
    - Michael Mahoney
  - Histórico da Computação
    - Profa. Magali Longhi (CESUP/UFRGS)

Aspectos históricos

- Arquitetura de von Neumann
  - Modelo de computador digital de programa armazenado => dados + instruções na memória.
  - Elementos: memória, UCP (unidade central de processamento), ULA (unidade lógica e aritmética) e unidade de controle.
  - Primeira implementação: IAS (Princeton Institute for Advances Studies), 1946.
  - Modelo de funcionamento: ciclo de instrução.
  - Baseada na abordagem CISC.
  - Problema associado: gargalo de von Neumann



- Máquina de von Neumann – blocos básicos



Blocos básicos da arquitetura

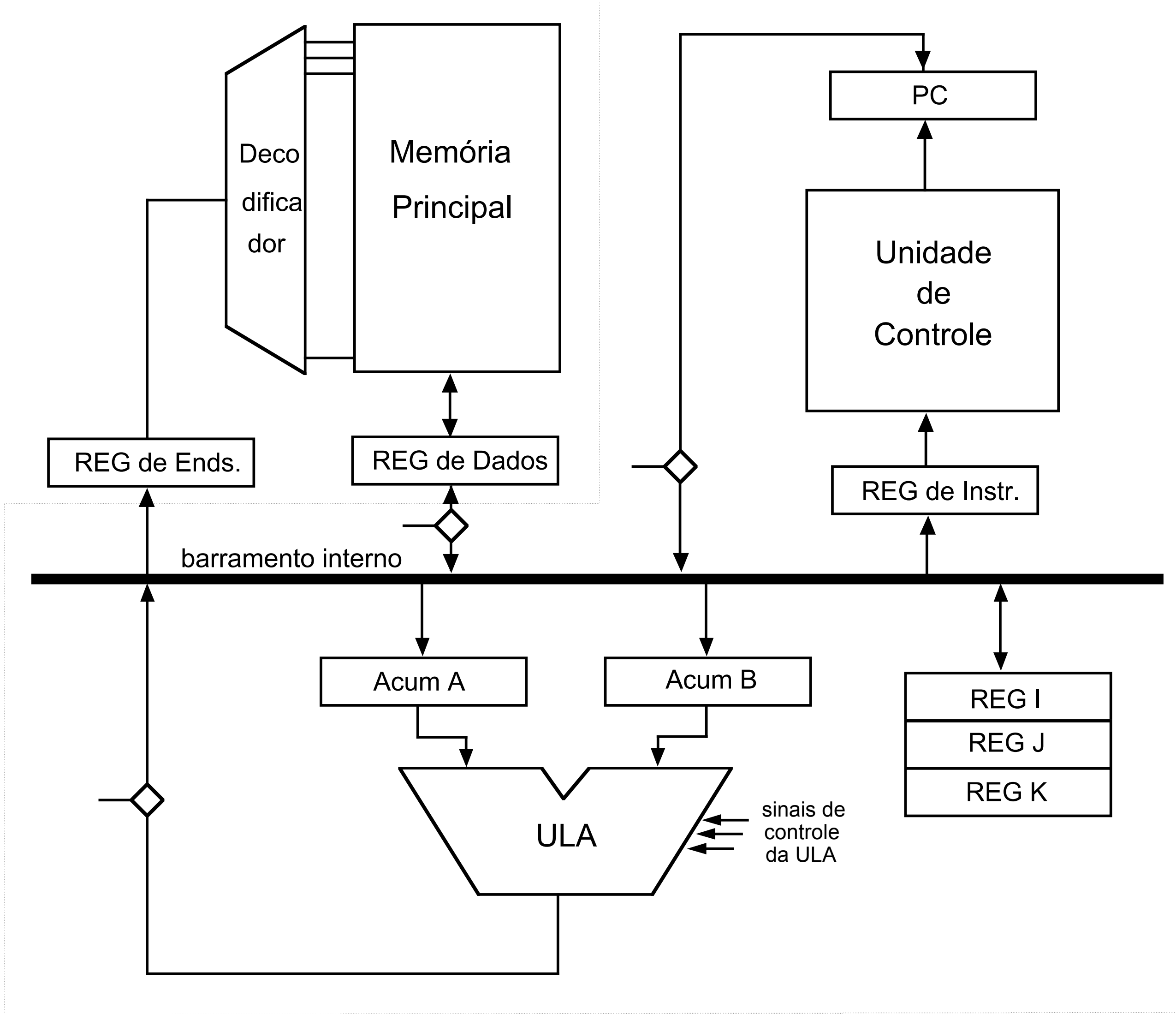
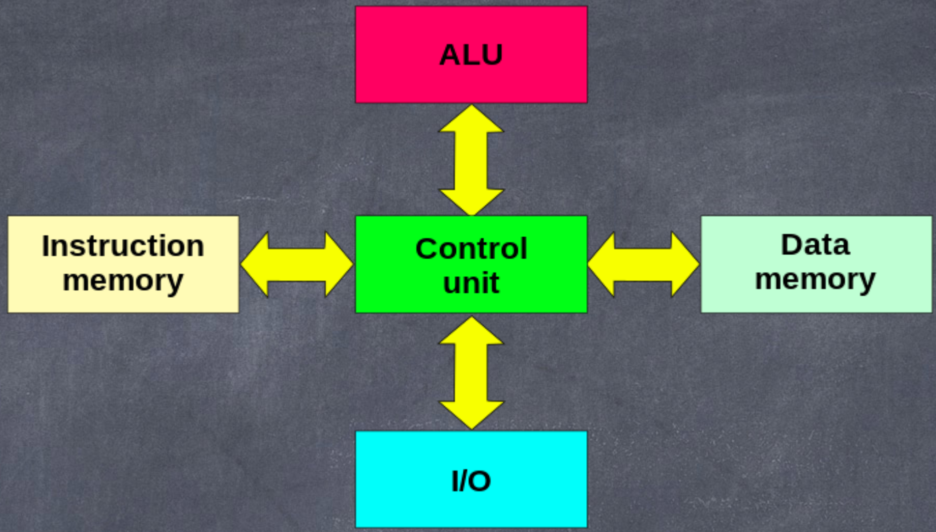


Figura 2 - Diagrama de blocos parcial para o computador da figura 1, detalhando a CPU (processador) e a memória principal.

Aspectos históricos

- Arquitetura Harvard



  - Modelo cujo propósito era fazer o microcontrolador trabalhar mais rápido => criado em 1937 (computador Harvard Mark I).
  - Estratégia: barramentos e memórias separadas para dados e instruções.
  - Vantagem: leitura de instruções e de alguns operandos ocorre em paralelo à execução de outras instruções => princípio da técnica de pipelining.
  - Implementações de referência: microcontroladores PIC, DSP, ARM => abordagem de projeto RISC.

Aspectos históricos

- Fatores associados à evolução dos computadores
  - Diminuição no tamanho dos componentes.
  - Aumento na capacidade de memória.
  - Aumento na capacidade e na velocidade dos mecanismos de E/S.
  - Aumento na velocidade dos processadores.

Conjunto de instruções

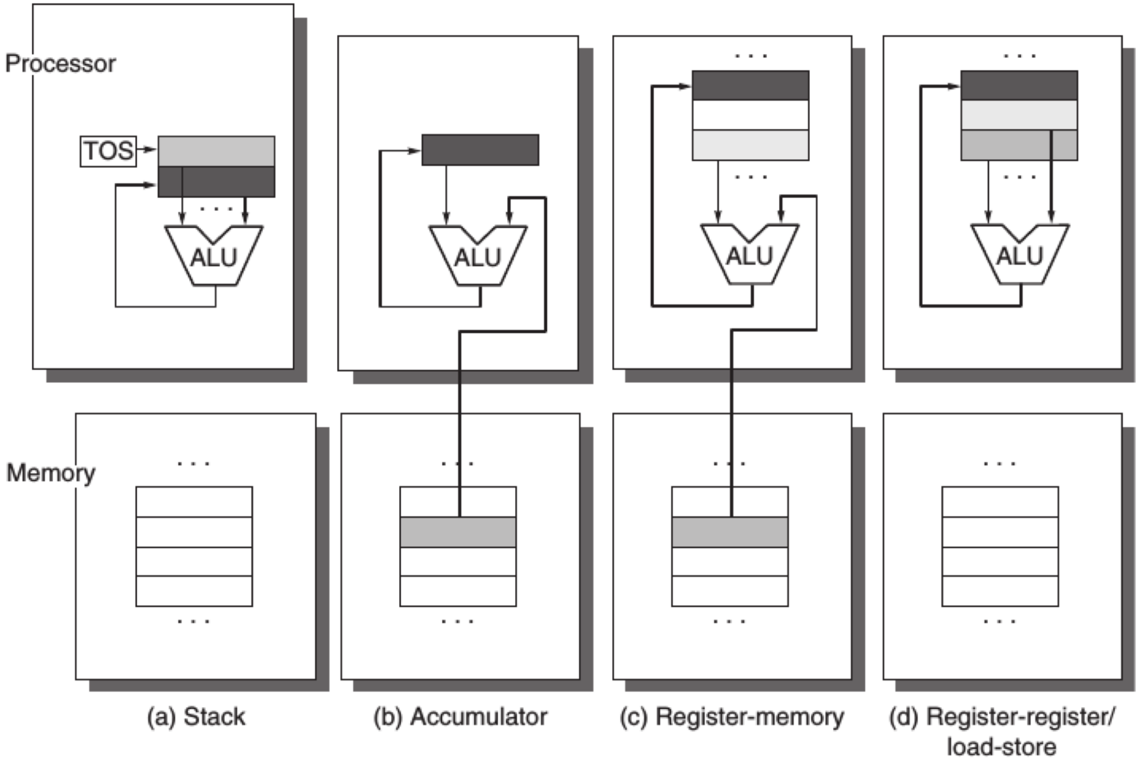
- Instruction Set Architecture (ISA)
  - Parte da arquitetura que é visível ao programador.
  - Interface entre o software e o hardware.
  - Vários exemplos: 80x86, ARM, MIPS, PowerPC etc.

Conjunto de instruções (ISA)

- 7 dimensões (antiga visão de projeto):
  - Classe da ISA
  - Endereçamento de memória
  - Modos de endereçamento de operandos
  - Tipos e tamanhos dos operandos
  - Operações (instruções implementadas)
  - Instruções de controle de fluxo
  - Formato das instruções (codificação da ISA)

Conjunto de instruções (ISA)

- Classe da ISA



C = A + B

PUSH A	LOAD A	LOAD R1, A	LOAD R1, A
PUSH B	ADD B	ADD R3, R1, B	LOAD R2, B
ADD	STORE C	STORE R3, C	ADD R3, R1, R2
POP C			STORE R3, C

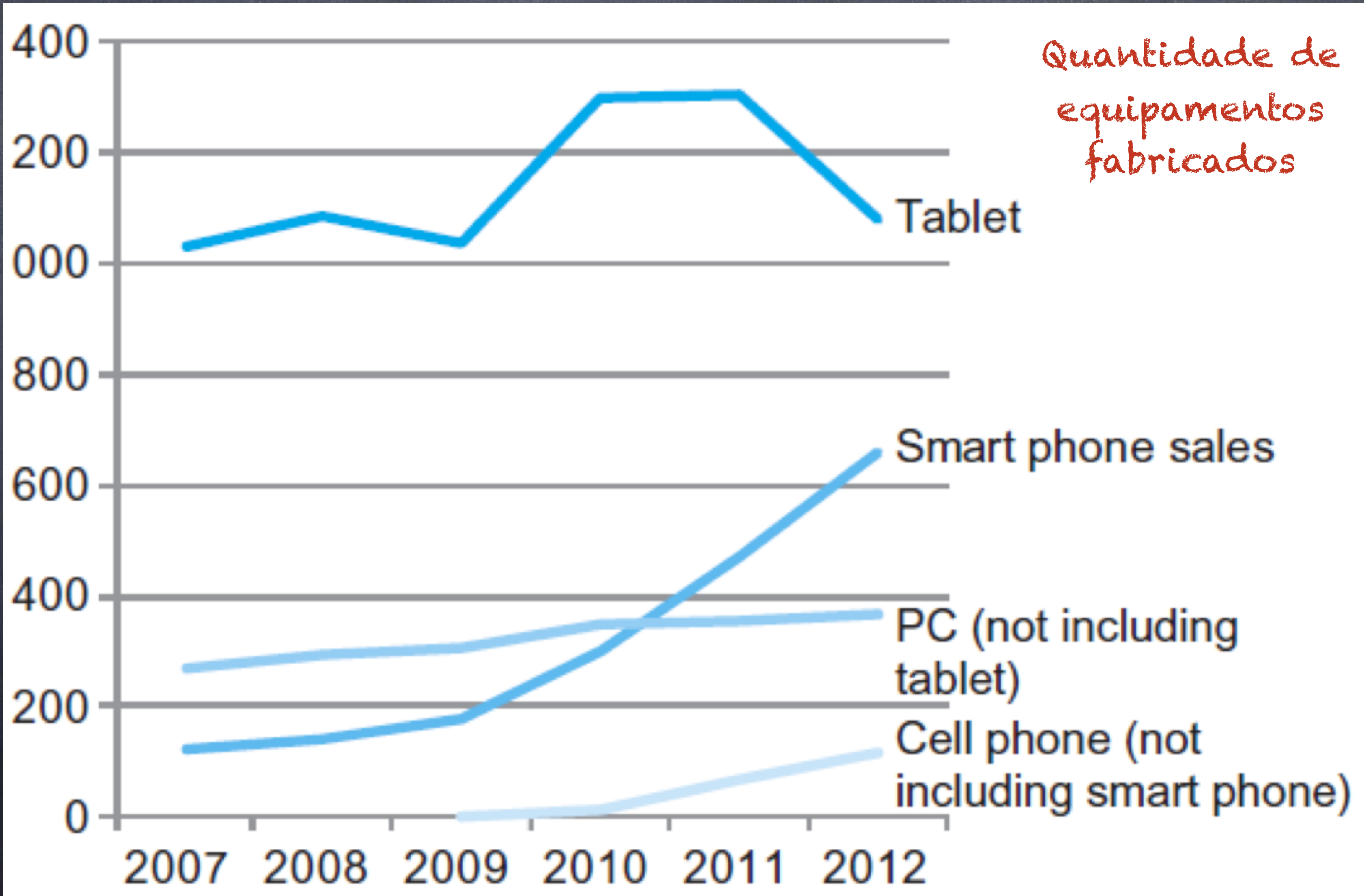
Visão real do projeto de sistemas computacionais

- Requisitos específicos da arquitetura-alvo.
- Foco em desempenho e eficiência energética, mas considerando restrições de custo, consumo e confiabilidade.
- 3 partes:
  - Conjunto de instruções (ISA)
  - Microarquitetura (ou organização)
  - Hardware

Era pós-PC

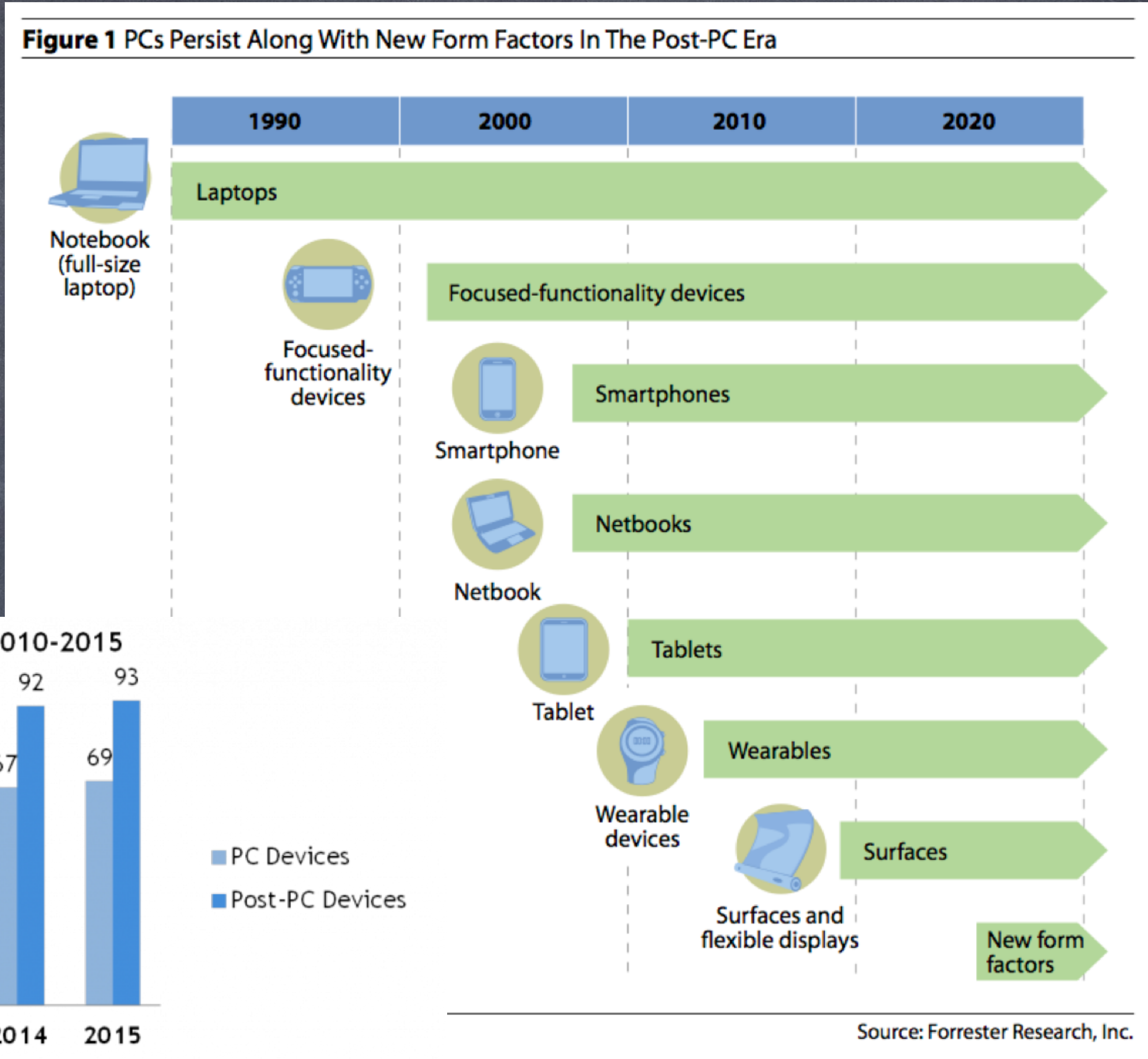
- Alimentação por bateria, conexão com internet, usuários instalam software ("apps"), entrada via fala ou tela sensível ao toque, custo de centenas de reais.

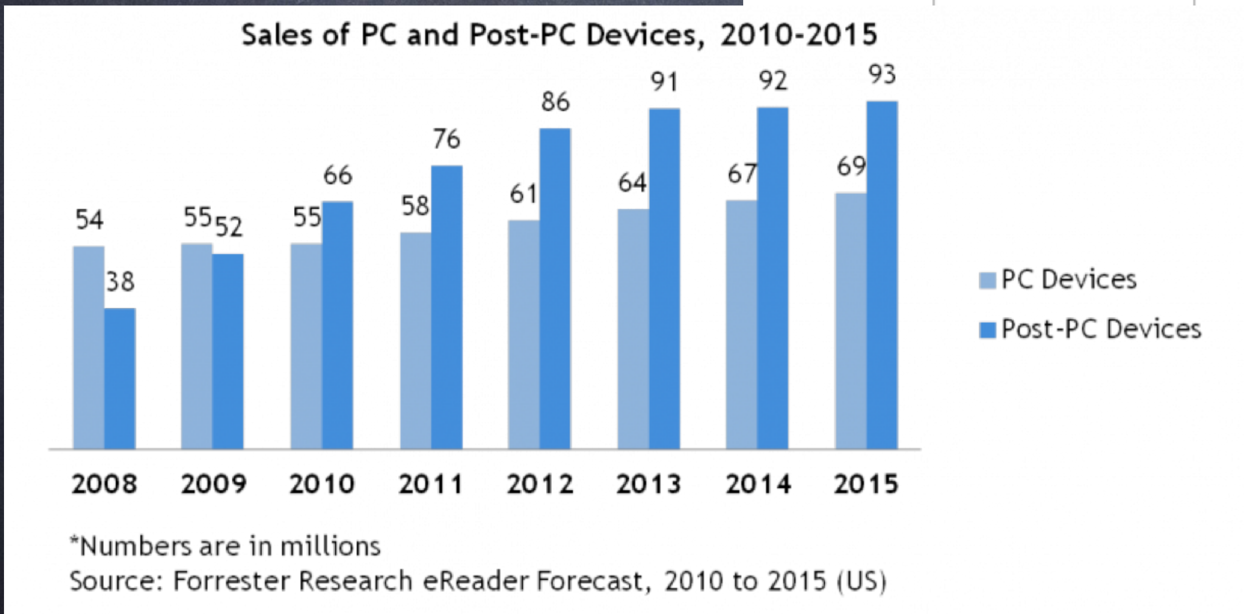
Milhões



Quantidade de equipamentos fabricados

Era pós-PC

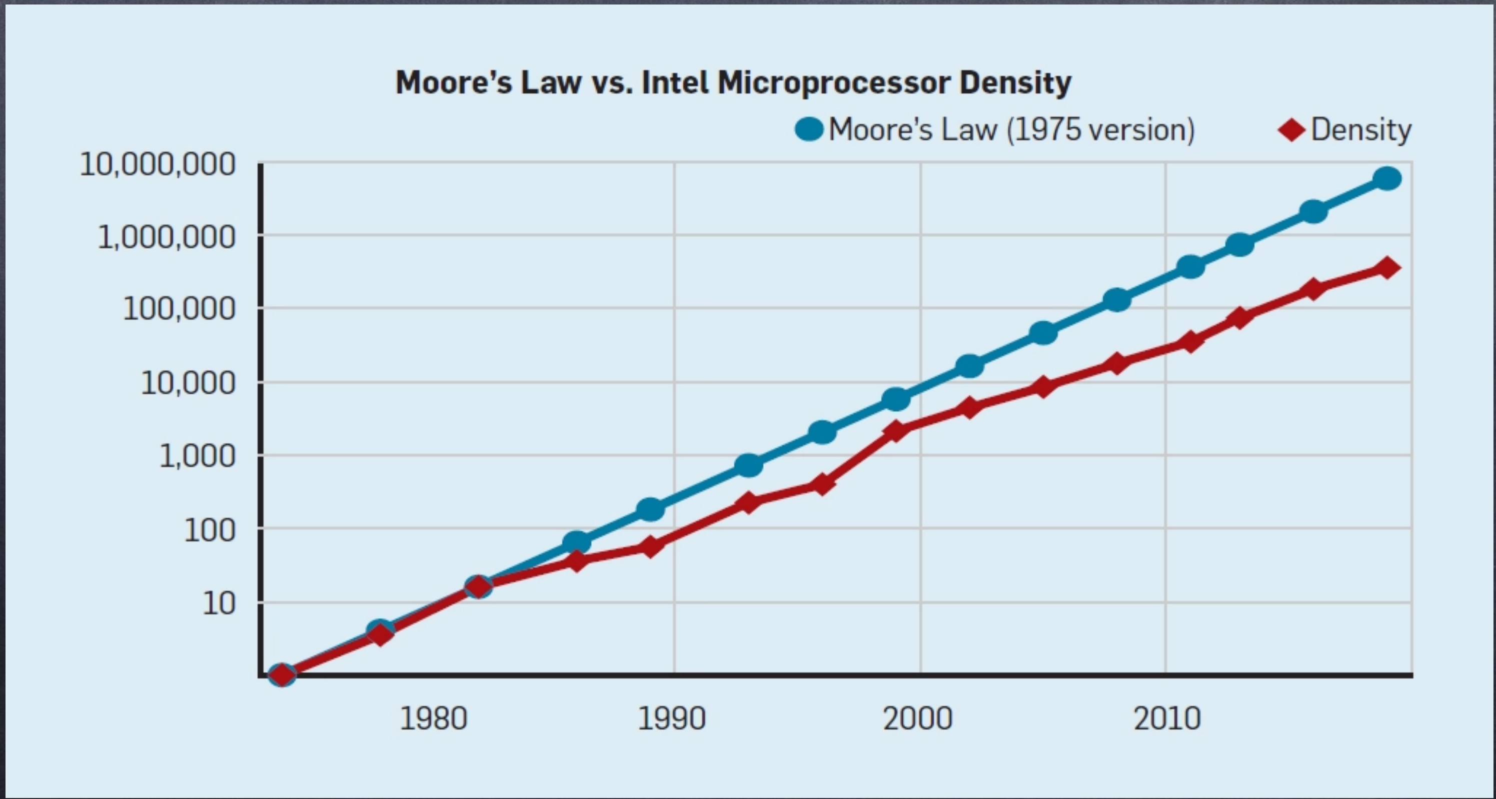




https://ipcarrier.blogspot.com.br/2012/11/post-pc-affects-device-usage-design.html

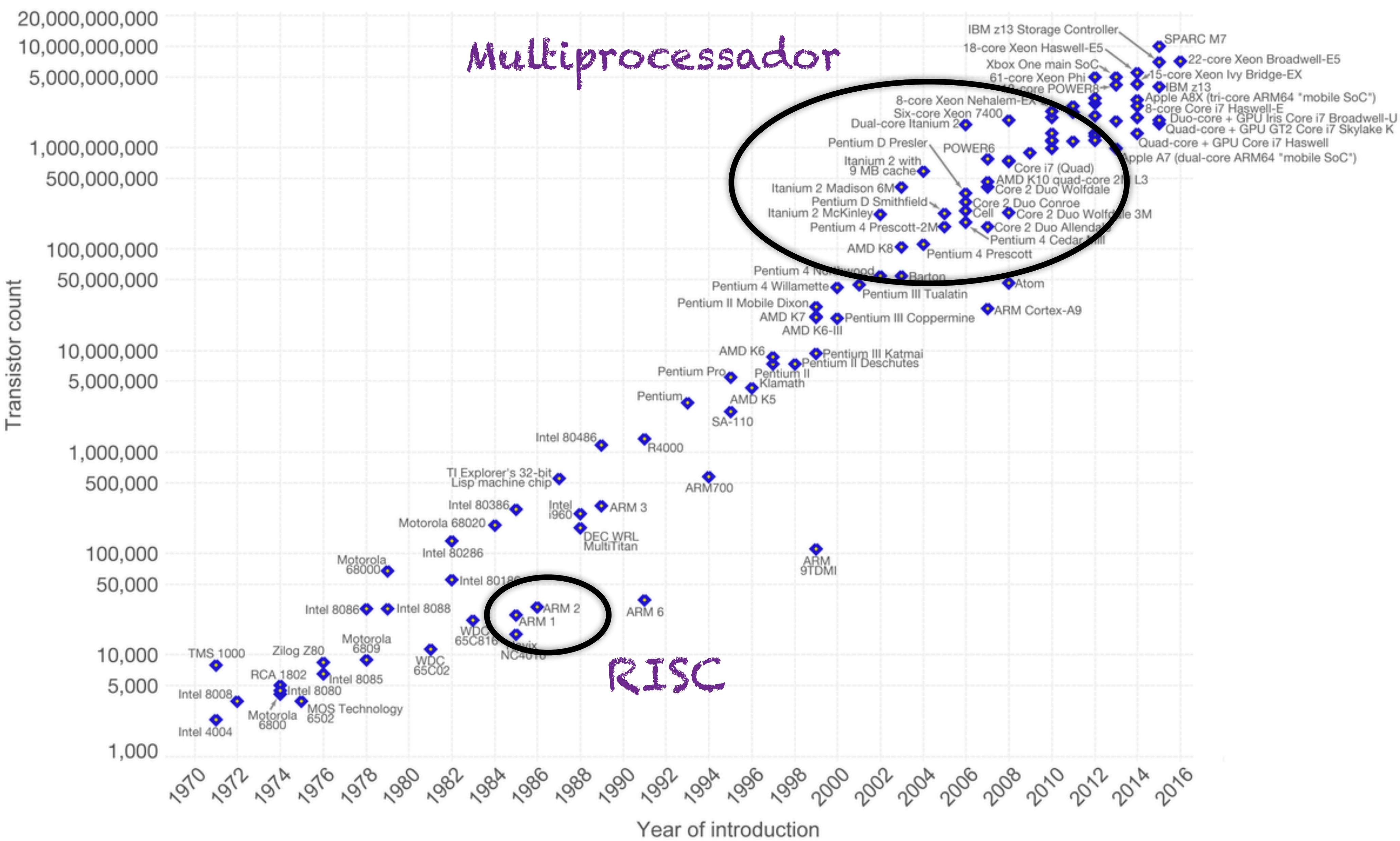
- Aspectos de desempenho
- Lei de Moore
- Gordon Moore (cofundador Intel):
- 1965: capacidade dos processadores vai dobrar a cada ano.
- 1975: revisada - a cada 2 anos.

Densidade = quantidade de elementos

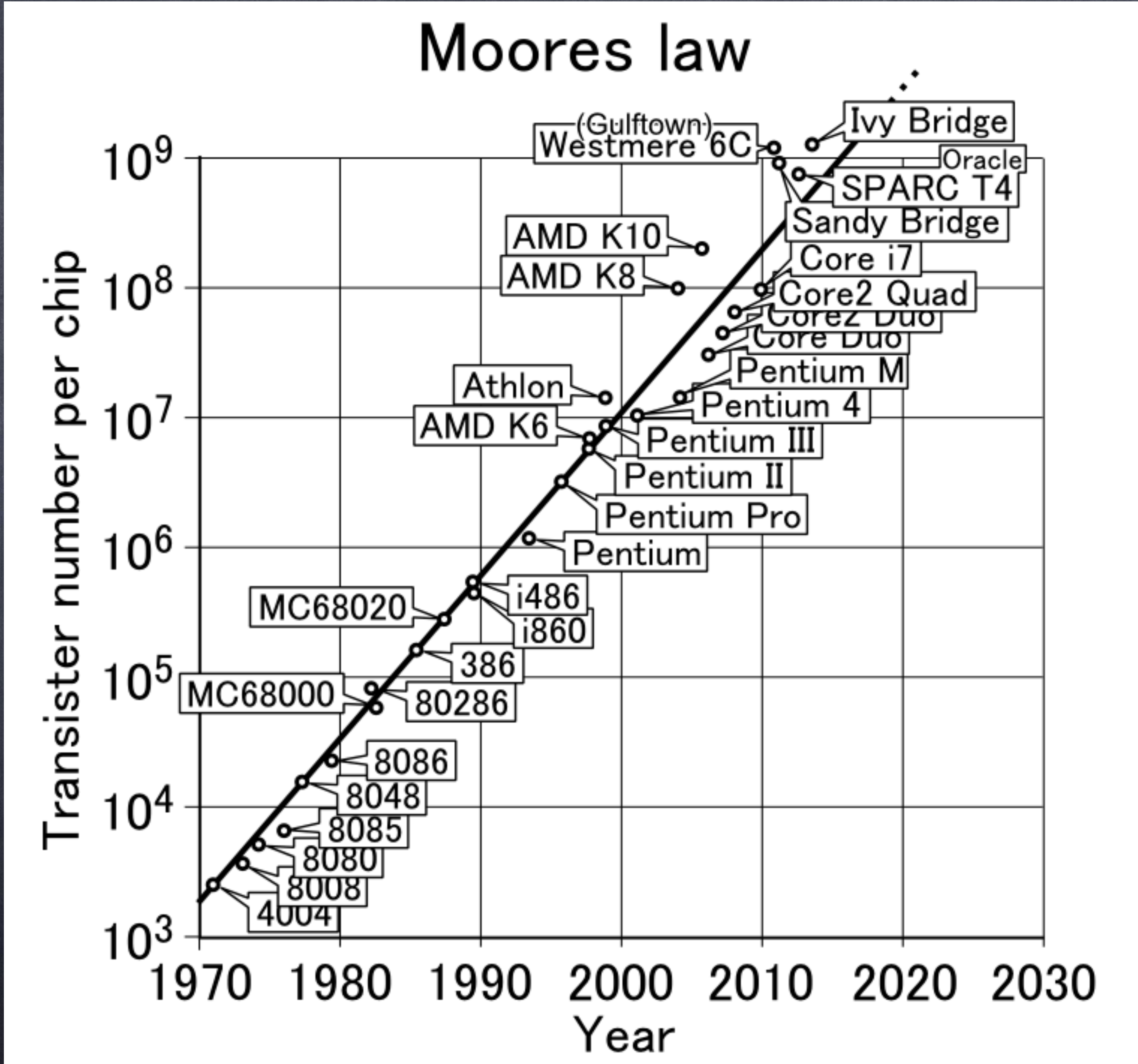


Moore's Law – The number of transistors on integrated circuit chips (1971-2016)

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.



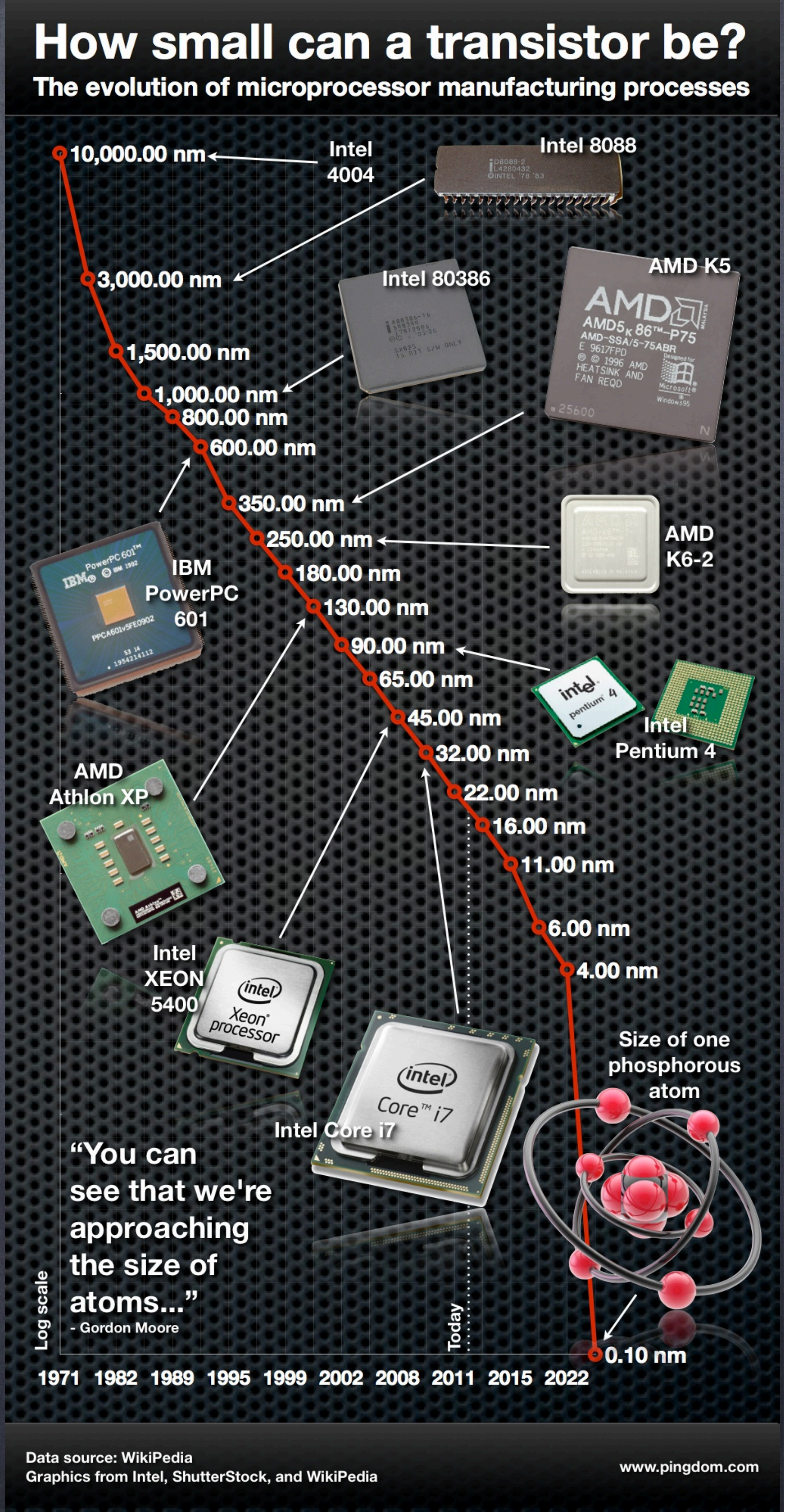
Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.



Moores law
Transister number per chip
Year
10^3
10^4
10^5
10^6
10^7
10^8
10^9
1970
1980
1990
2000
2010
2020
2030
4004
8008
8080
8085
8048
8086
MC68000
80286
386
MC68020
i486
i860
Pentium
Pentium Pro
Pentium II
Pentium III
AMD K6
Athlon
Pentium 4
Pentium M
Core Duo
Core2 Duo
Core2 Quad
Core i7
AMD K8
AMD K10
Sandy Bridge
SPARC T4
Oracle
(Gulftown)
Westmere 6C
Ivy Bridge

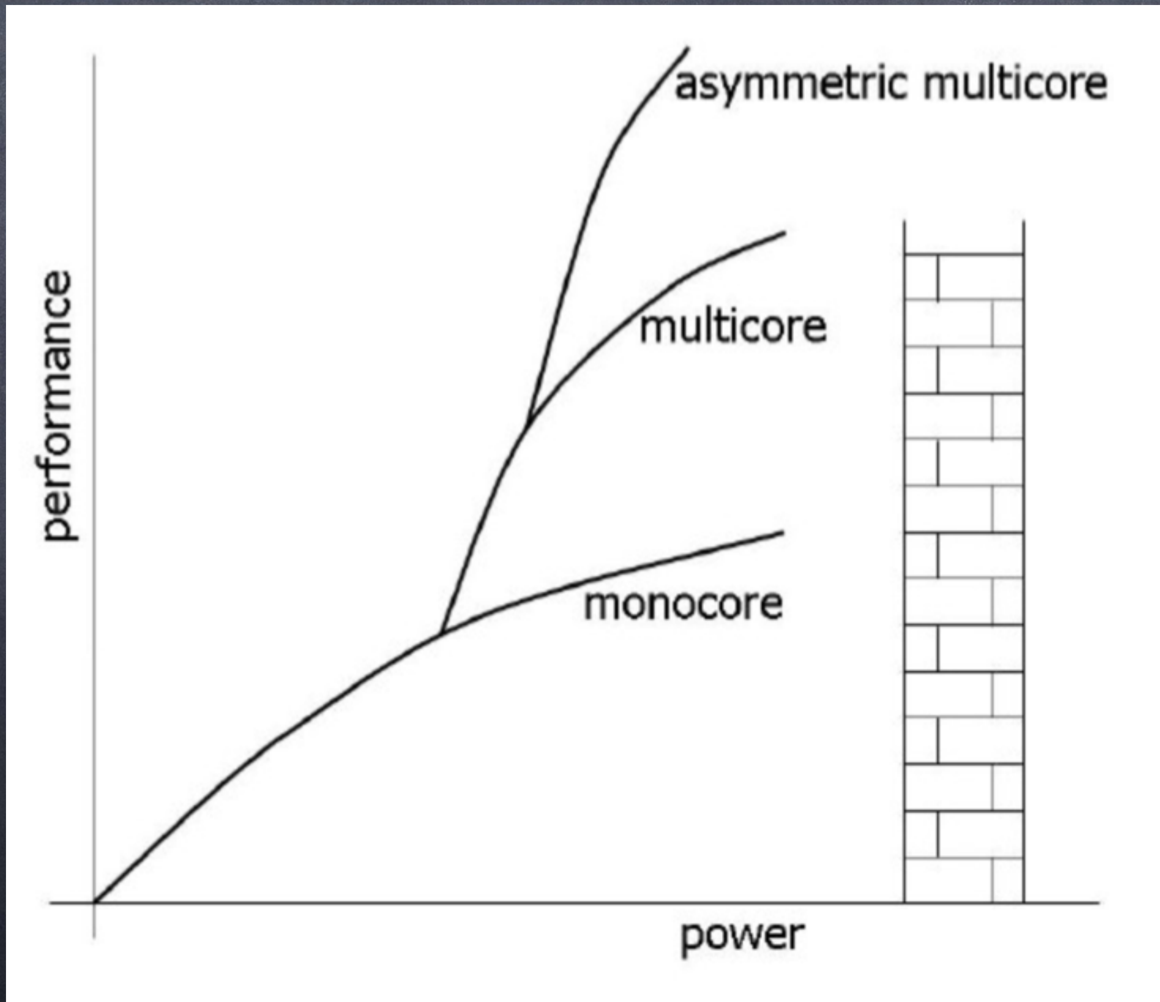
Aspectos de desempenho

- Densidade dos transistores (processadores) cresce de forma quadrática, enquanto o desempenho cresce de forma linear.
- Projetistas precisam usar mais transistores para aumentar o desempenho.



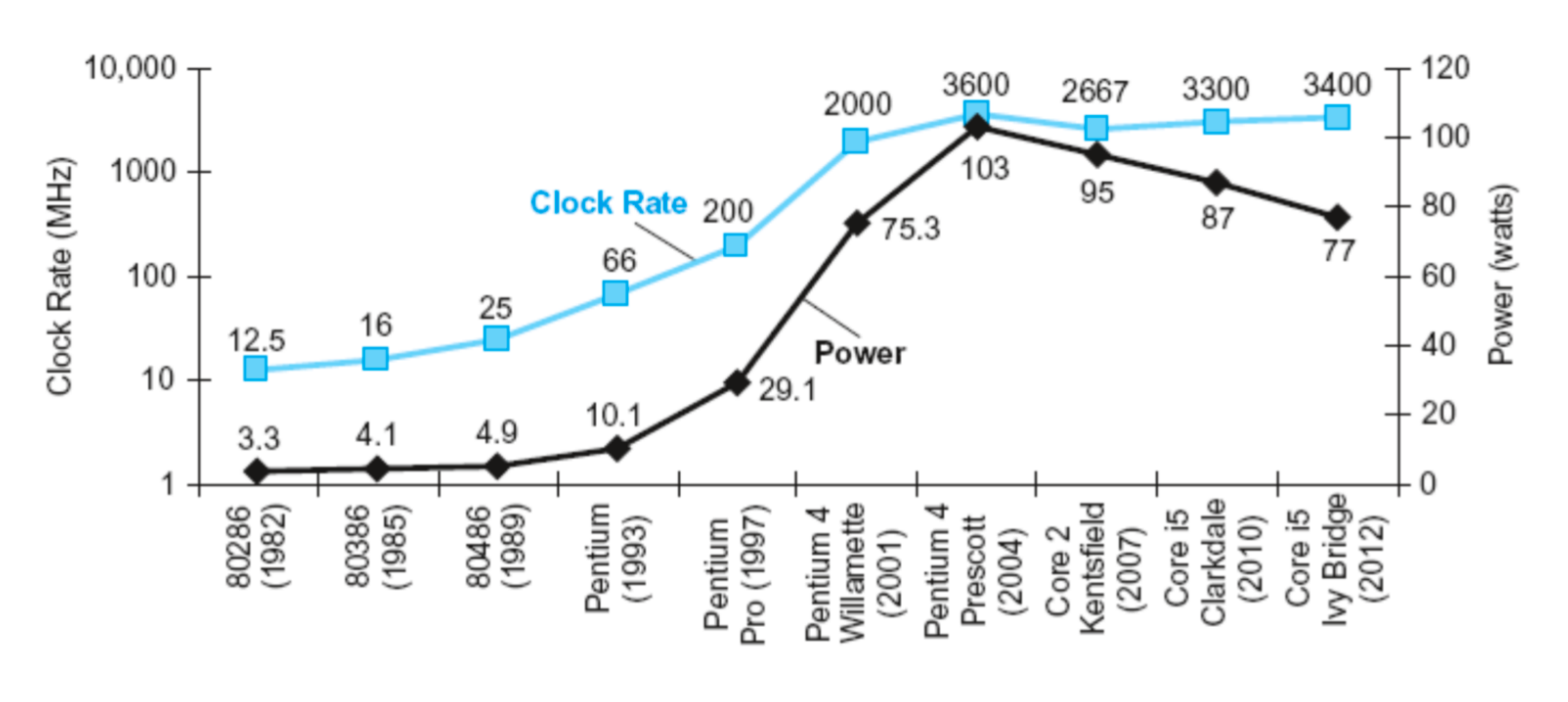
Aspectos de desempenho

- The power wall
    - Como equilibrar desempenho X consumo de energia X dissipação do calor (refrigeração)?



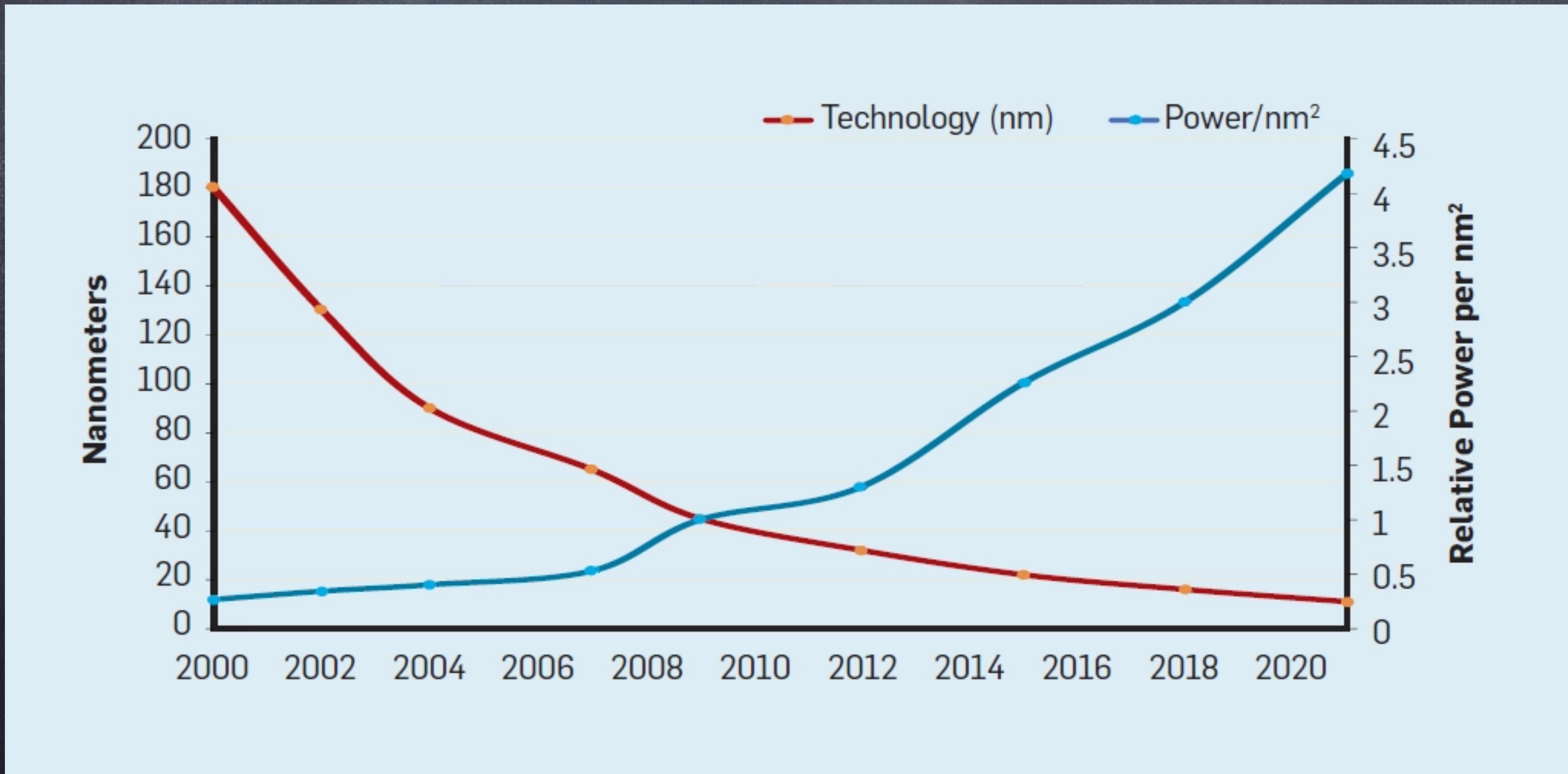
Aspectos de desempenho

- The power wall
  - Evolução da frequência e do consumo dos processadores Intel (8 gerações, 30 anos)



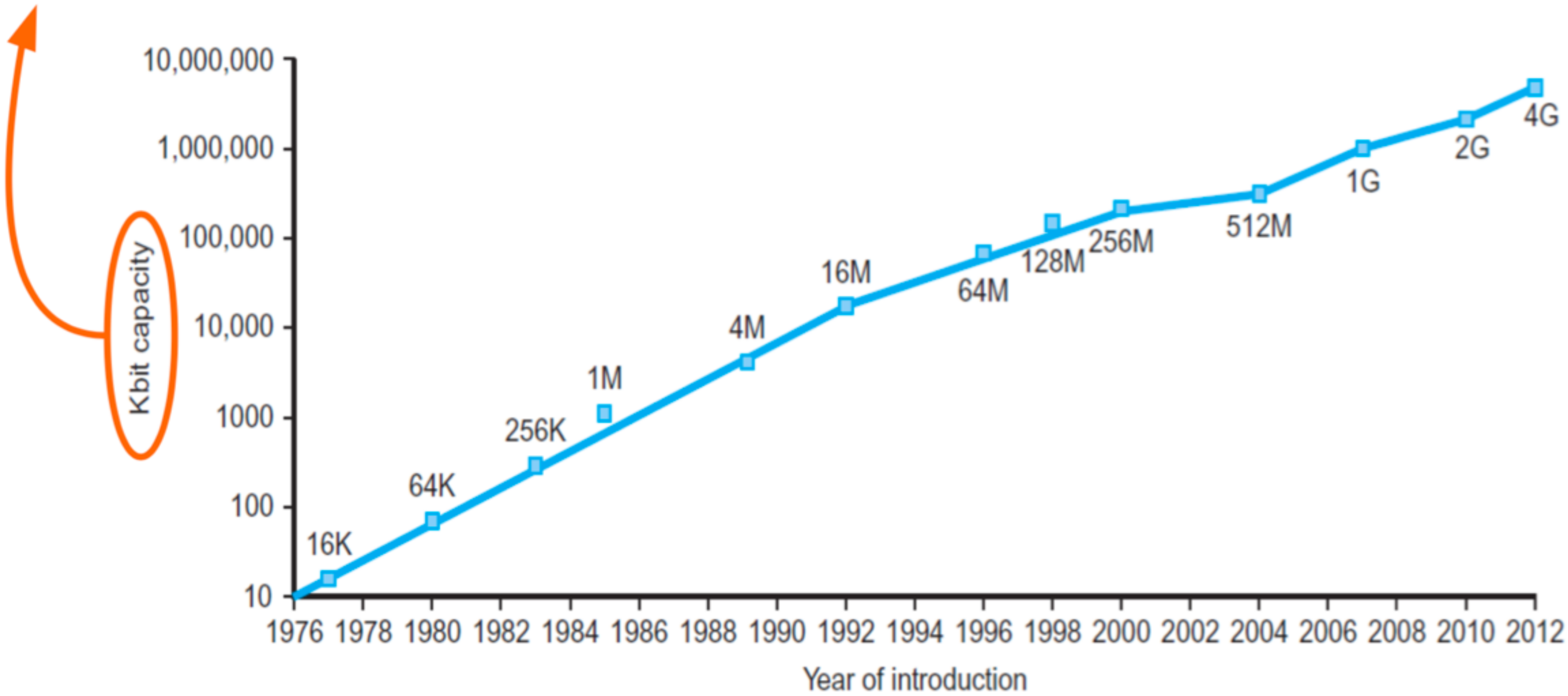
Aspectos de desempenho

- Dennard Scaling (1974)
  - A medida que a densidade do processador aumenta, o consumo de energia por transistor diminui, levando a uma potência por mm2 quase constante.

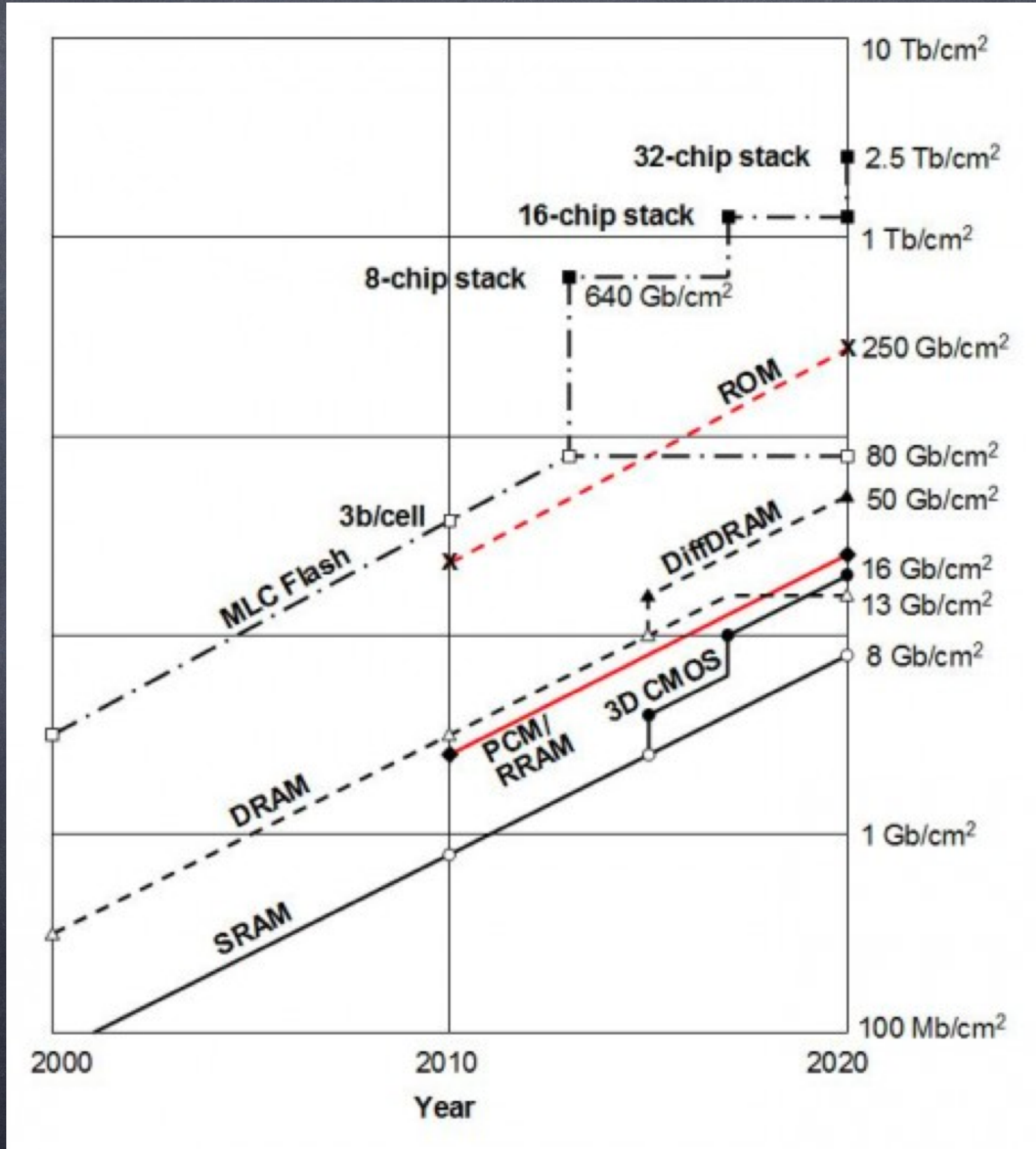


Evolução na capacidade da memória (DRAM)

Decimal term	Abbreviation	Value	Binary term	Abbreviation	Value	% Larger
kilobyte	KB	10^3	kibibyte	KiB	2^{10}	2%
megabyte	MB	10^6	mebibyte	MiB	2^{20}	5%
gigabyte	GB	10^9	gibibyte	GiB	2^{30}	7%
terabyte	TB	10^{12}	tebibyte	TiB	2^{40}	10%
petabyte	PB	10^{15}	pebibyte	PiB	2^{50}	13%
exabyte	EB	10^{18}	exbibyte	EiB	2^{60}	15%
zettabyte	ZB	10^{21}	zebibyte	ZiB	2^{70}	18%
yottabyte	YB	10^{24}	yobibyte	YiB	2^{80}	21%



Towards Terabit Memories



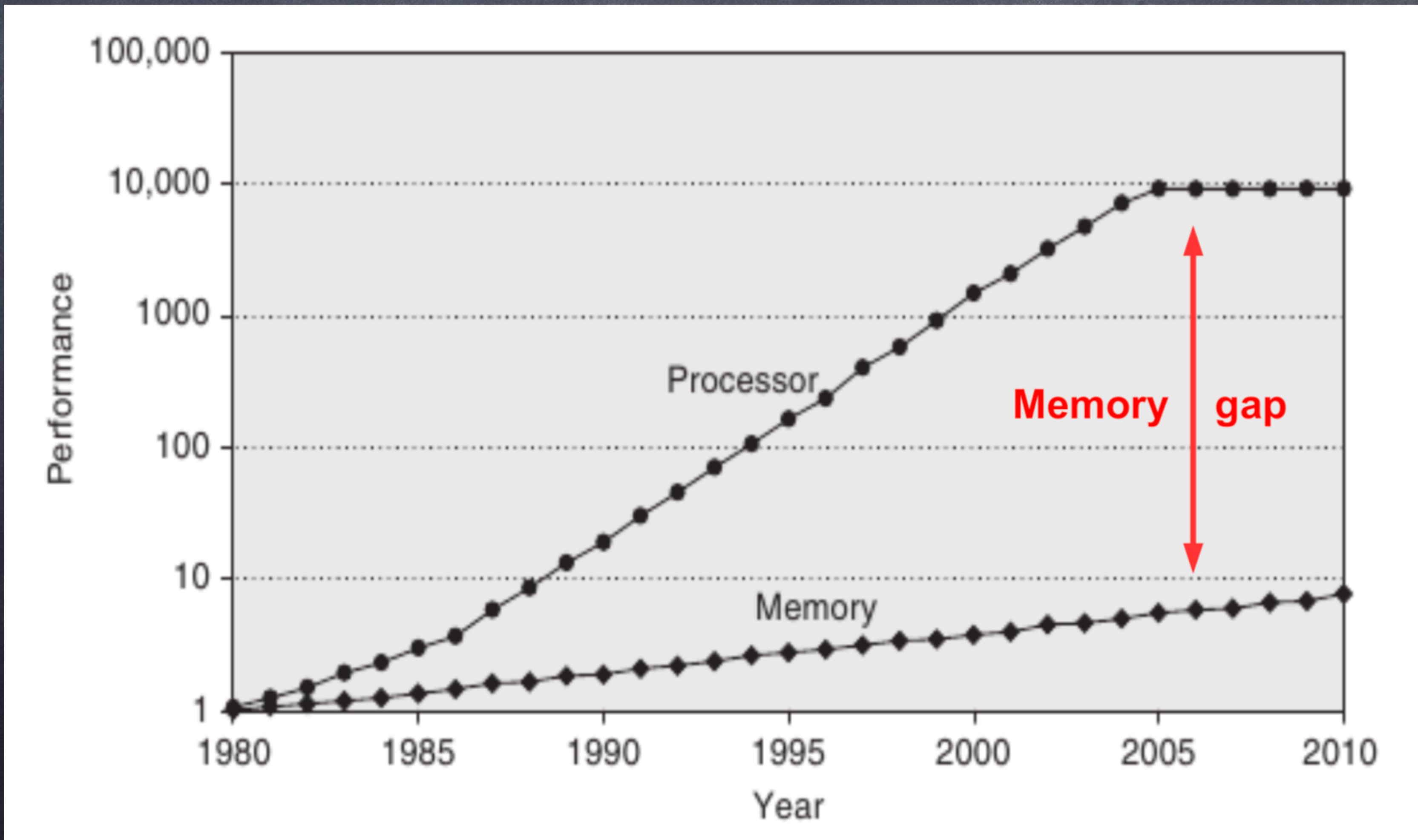
Evolution of CMOS memory density for SRAM, DRAM and Flash memory. MLC multi-level per cell, PCM phase-change memory, RRAM resistive RAM

Aspectos de desempenho

- Memory gap

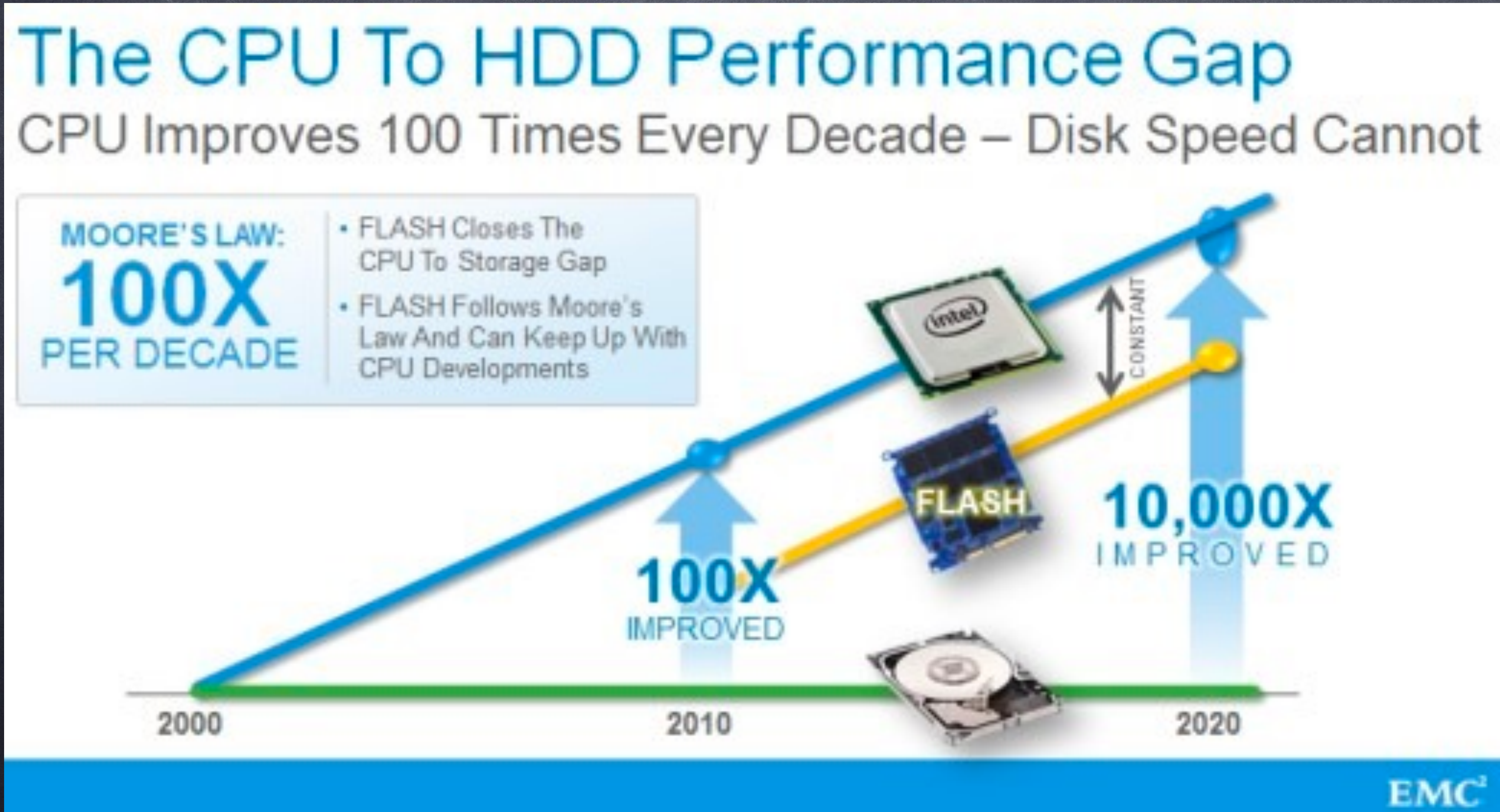
Lembrar do Gargalo de von Neumann!

  - Diferença entre velocidade do processador e da memória.

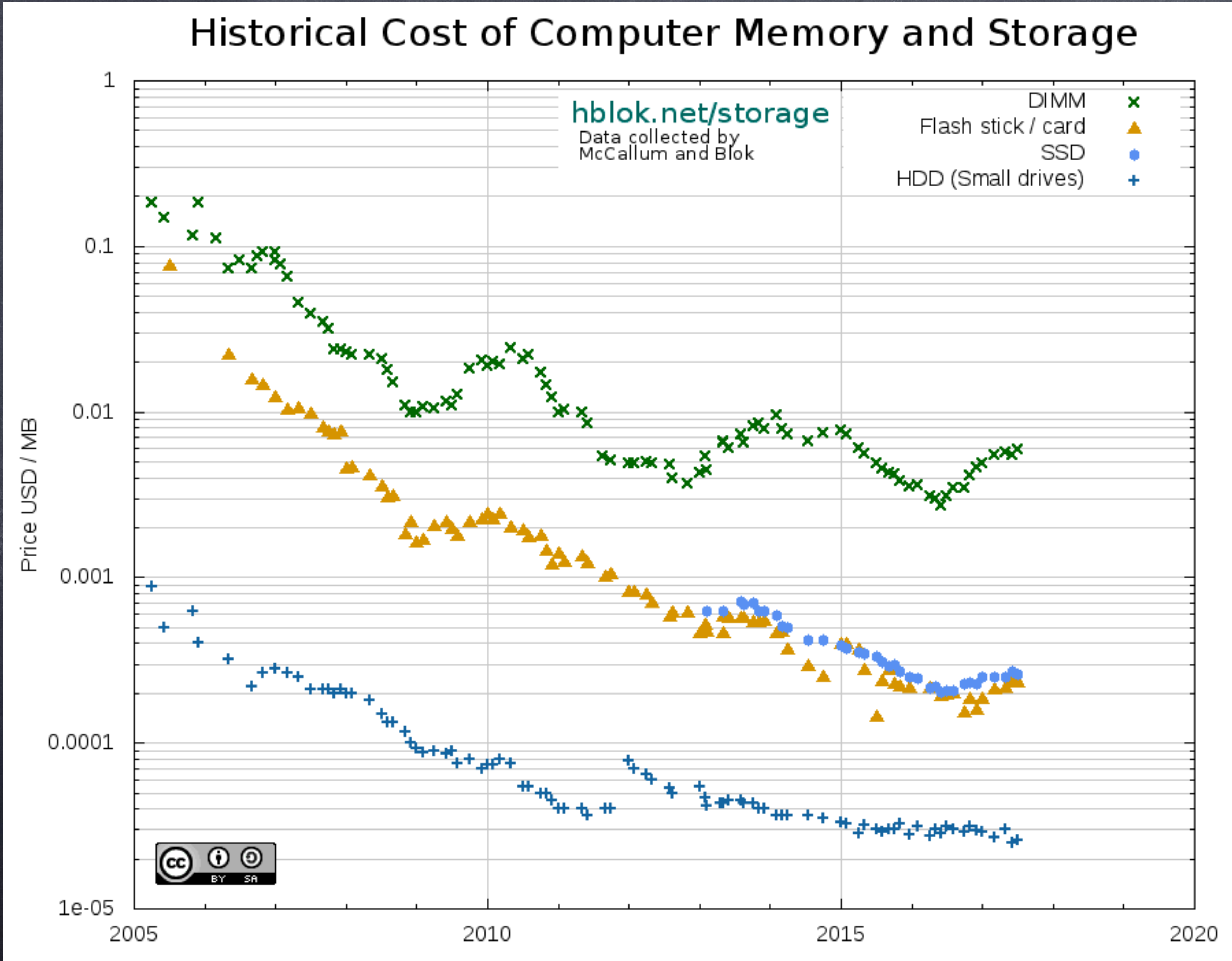


Aspectos de desempenho

- CPU x HDD performance

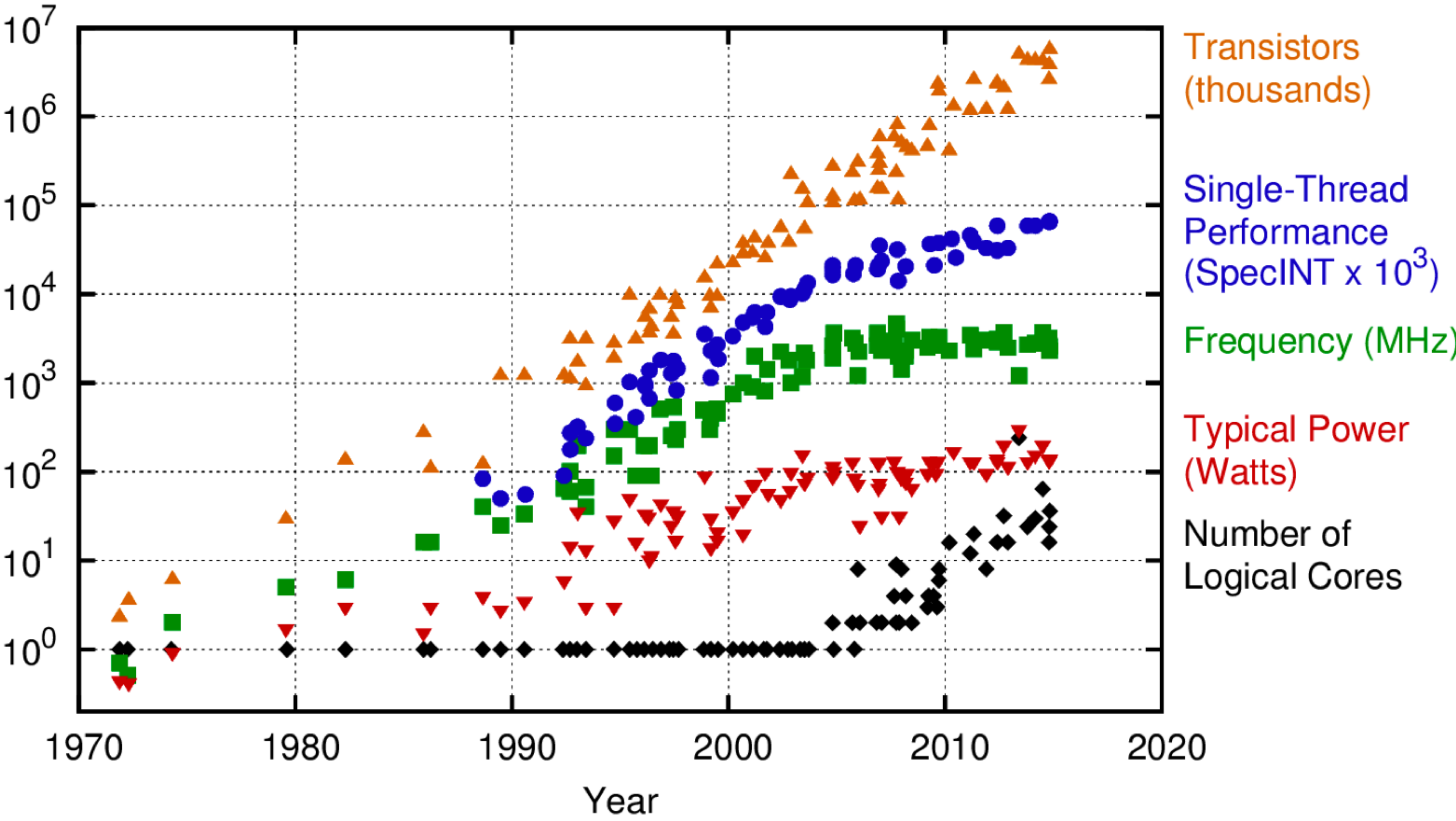


Custos de memória e armazenamento secundário



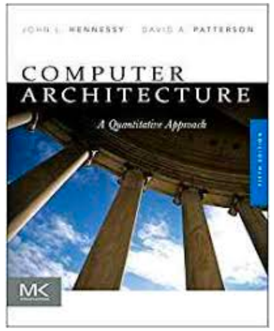
Evolução - capacidade, consumo, potência, desempenho

40 Years of Microprocessor Trend Data



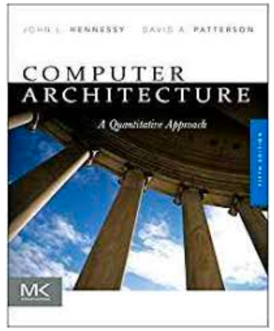
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

Trends in Technology



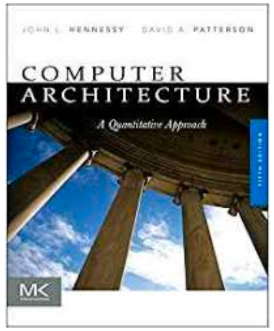
- Integrated circuit technology (Moore's Law)
  - Transistor density: 35%/year
  - Die size: 10-20%/year
  - Integration overall: 40-55%/year
- DRAM capacity: 25-40%/year (slowing)
  - 8 Gb (2014), 16 Gb (2019), possibly no 32 Gb
- Flash capacity: 50-60%/year
  - 8-10X cheaper/bit than DRAM
- Magnetic disk capacity: recently slowed to 5%/year
  - Density increases may no longer be possible, maybe increase from 7 to 9 platters
  - 8-10X cheaper/bit then Flash
  - 200-300X cheaper/bit than DRAM

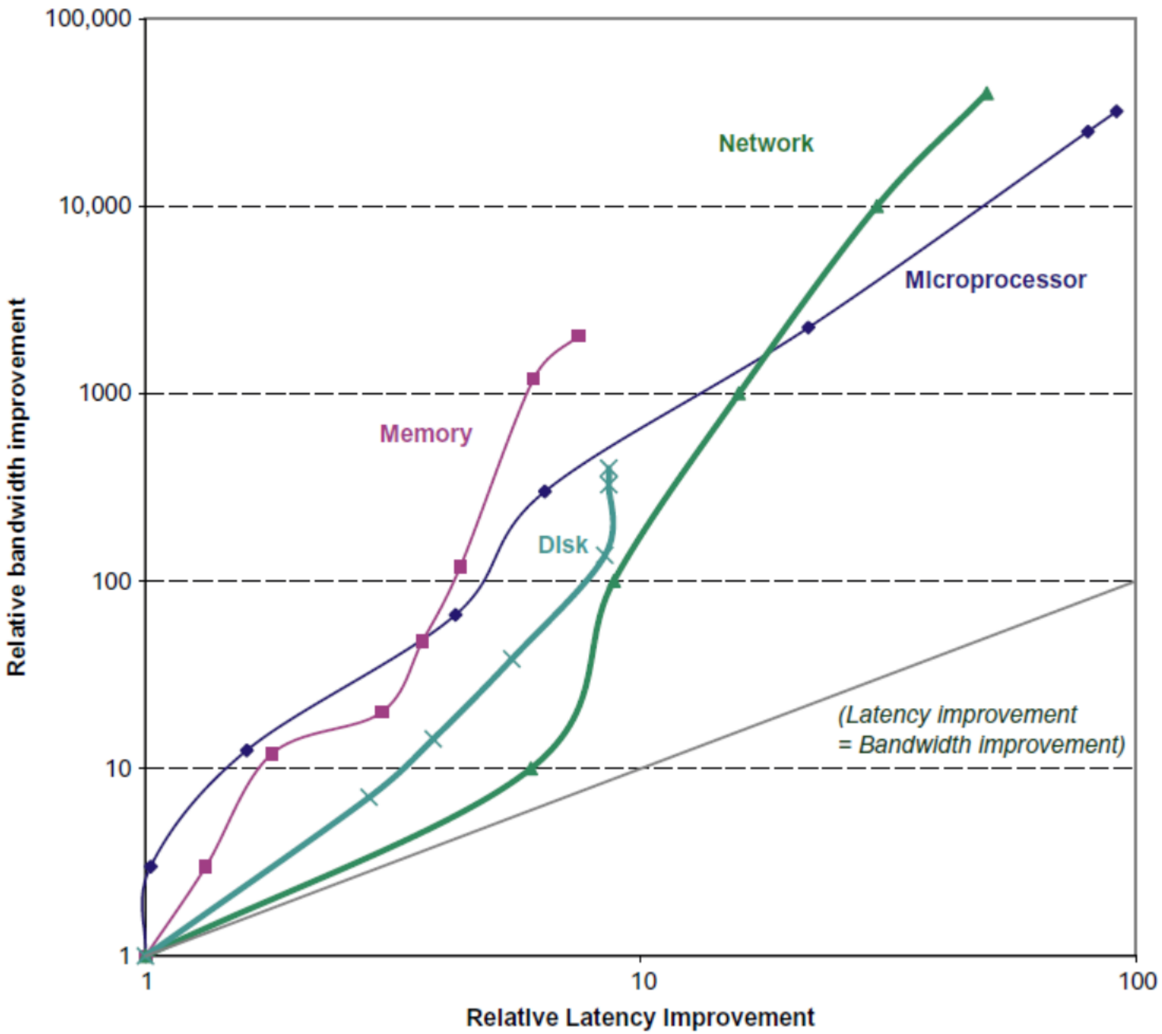
Bandwidth and Latency



- Bandwidth or throughput
  - Total work done in a given time
  - 32,000-40,000X improvement for processors
  - 300-1200X improvement for memory and disks
- Latency or response time
  - Time between start and completion of an event
  - 50-90X improvement for processors
  - 6-8X improvement for memory and disks

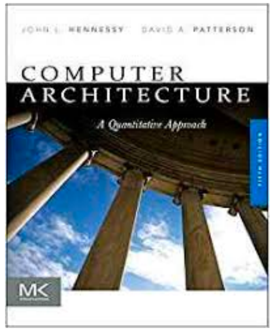
Bandwidth and Latency





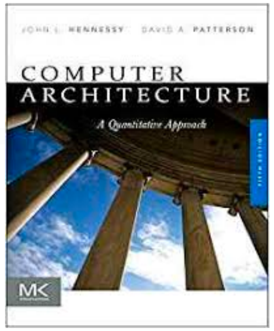
Log-log plot of bandwidth and latency milestones

Transistors and Wires



- Feature size
  - Minimum size of transistor or wire in x or y dimension
  - 10 microns in 1971 to .011 microns in 2017
  - Transistor performance scales linearly
    - Wire delay does not improve with feature size!
  - Integration density scales quadratically

Power and Energy



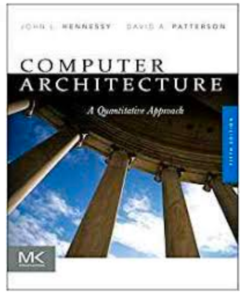
- Problem: Get power in, get power out

- Thermal Design Power (TDP)
  - Characterizes sustained power consumption
  - Used as target for power supply and cooling system
  - Lower than peak power (1.5X higher), higher than average power consumption

- Clock rate can be reduced dynamically to limit power consumption

- Energy per task is often a better measurement

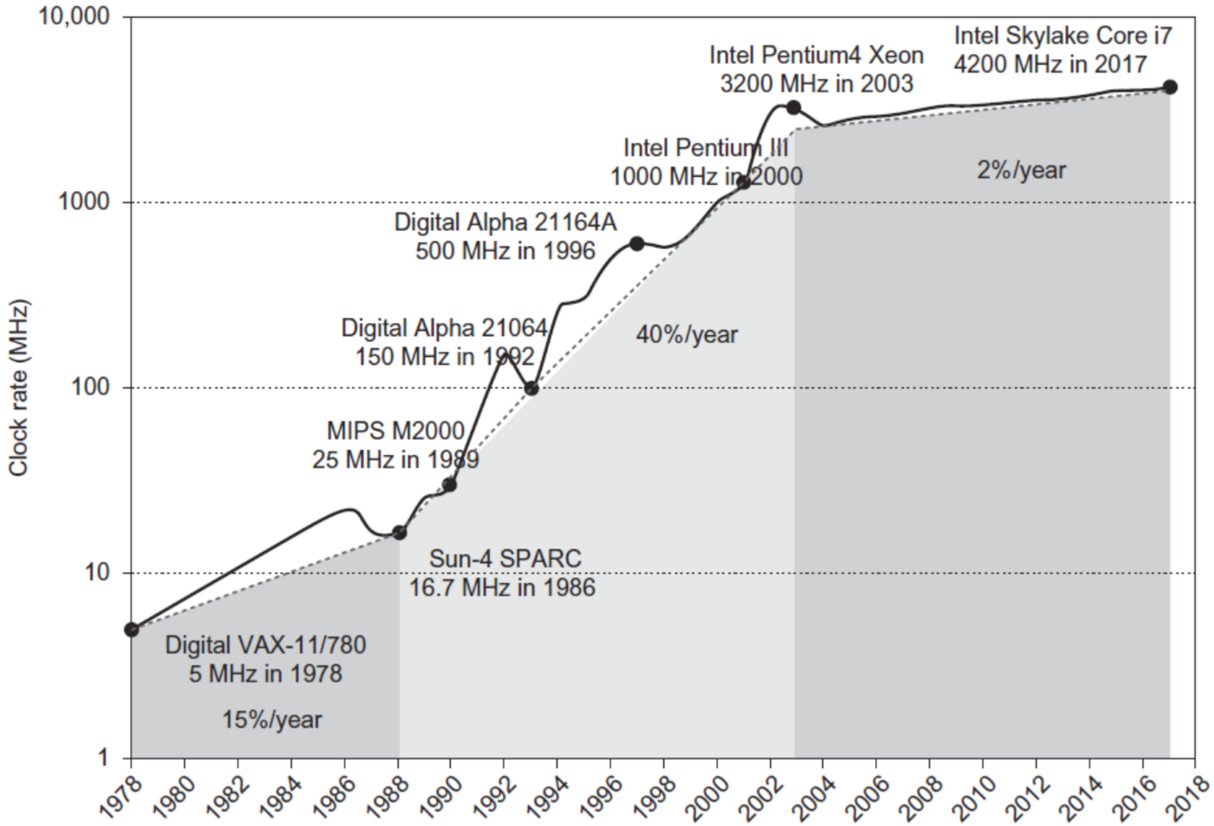
Dynamic Energy and Power



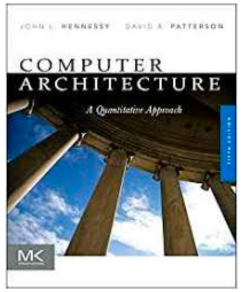
- Dynamic energy
  - Transistor switch from 0 -> 1 or 1 -> 0
  - ½ x Capacitive load x Voltage^2
- Dynamic power
  - ½ x Capacitive load x Voltage^2 x Frequency switched
- Reducing clock rate reduces power, not energy

Power

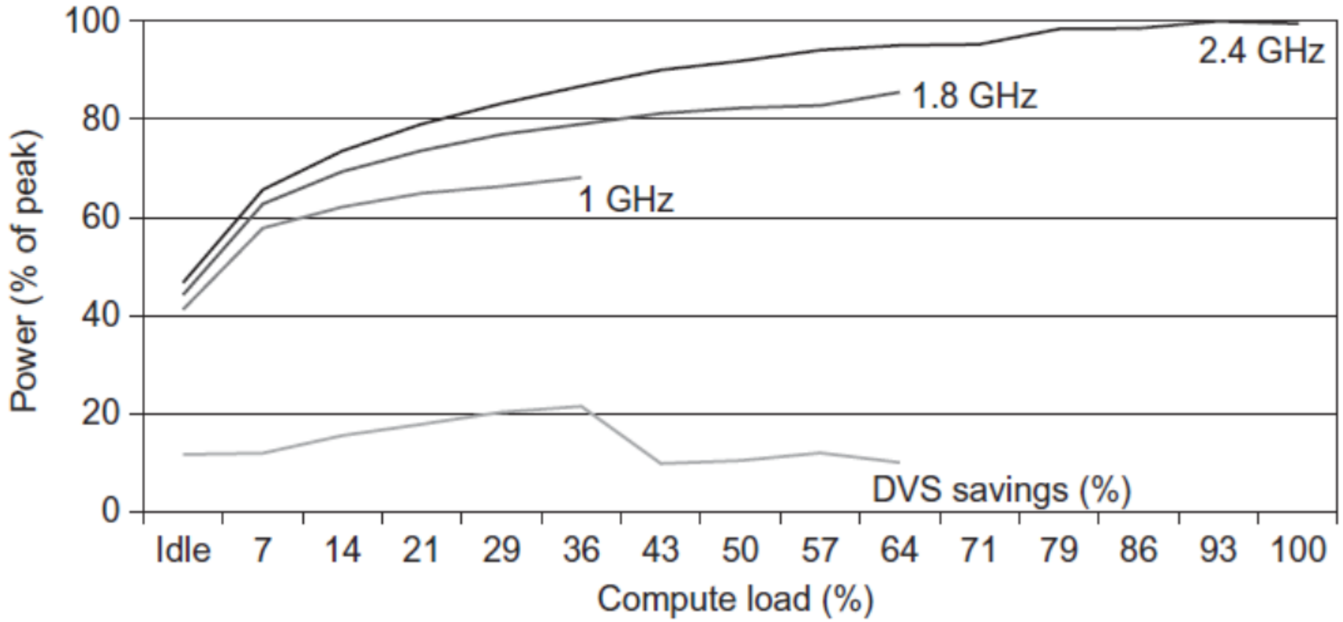
- Intel 80386 consumed ~ 2 W
- 3.3 GHz Intel Core i7 consumes 130 W
- Heat must be dissipated from 1.5 x 1.5 cm chip
- This is the limit of what can be cooled by air



Reducing Power

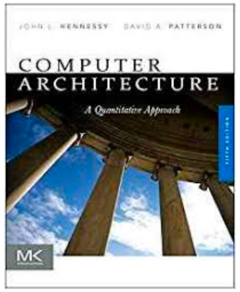


- Techniques for reducing power:
  - Do nothing well
  - Dynamic Voltage-Frequency Scaling



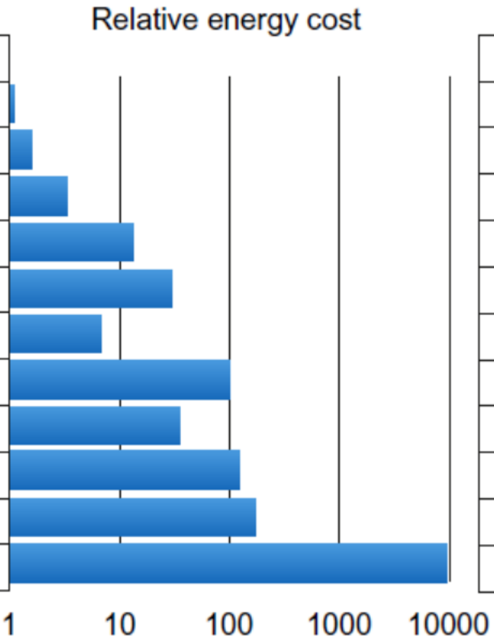
  - Low power state for DRAM, disks
  - Overclocking, turning off cores

Static Power

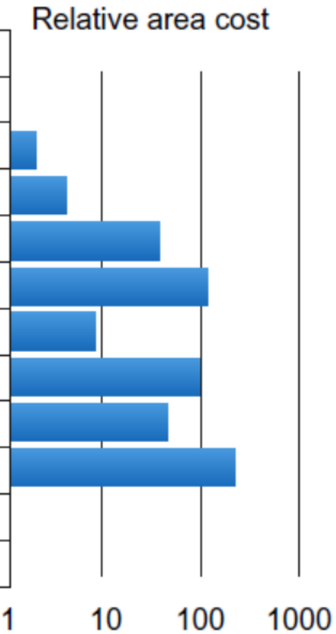


- Static power consumption
  - 25-50% of total power
  - $\text{Current}_{\text{static}}$ x Voltage
  - Scales with number of transistors
  - To reduce: power gating

Operation:	Energy (pJ)
8b Add	0.03
16b Add	0.05
32b Add	0.1
16b FB Add	0.4
32b FB Add	0.9
8b Mult	0.2
32b Mult	3.1
16b FB Mult	1.1
32b FB Mult	3.7
32b SRAM Read (8KB)	5
32b DRAM Read	640



Area (μm^2)
36
67
137
1360
4184
282
3495
1640
7700
N/A
N/A



Integrated Circuit Cost

- Integrated circuit

$$\text{Cost of integrated circuit} = \frac{\text{Cost of die} + \text{Cost of testing die} + \text{Cost of packaging and final test}}{\text{Final test yield}}$$

$$\text{Cost of die} = \frac{\text{Cost of wafer}}{\text{Dies per wafer x Die yield}}$$

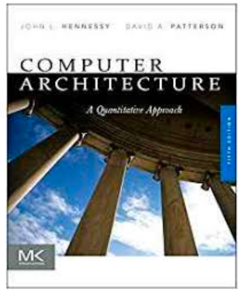
$$\text{Dies per wafer} = \frac{\pi \times (\text{Wafer diameter}/2)^2}{\text{Die area}} - \frac{\pi \times \text{Wafer diameter}}{\sqrt{2 \times \text{Die area}}}$$

- Bose-Einstein formula:

$$\text{Die yield} = \text{Wafer yield} \times 1/(1 + \text{Defects per unit area} \times \text{Die area})^N$$

- Defects per unit area = 0.016-0.057 defects per square cm (2010)
- N = process-complexity factor = 11.5-15.5 (40 nm, 2010)

Dependability



- Module reliability
  - Mean time to failure (MTTF)
  - Mean time to repair (MTTR)
  - Mean time between failures (MTBF) = MTTF + MTTR
  - Availability = MTTF / MTBF