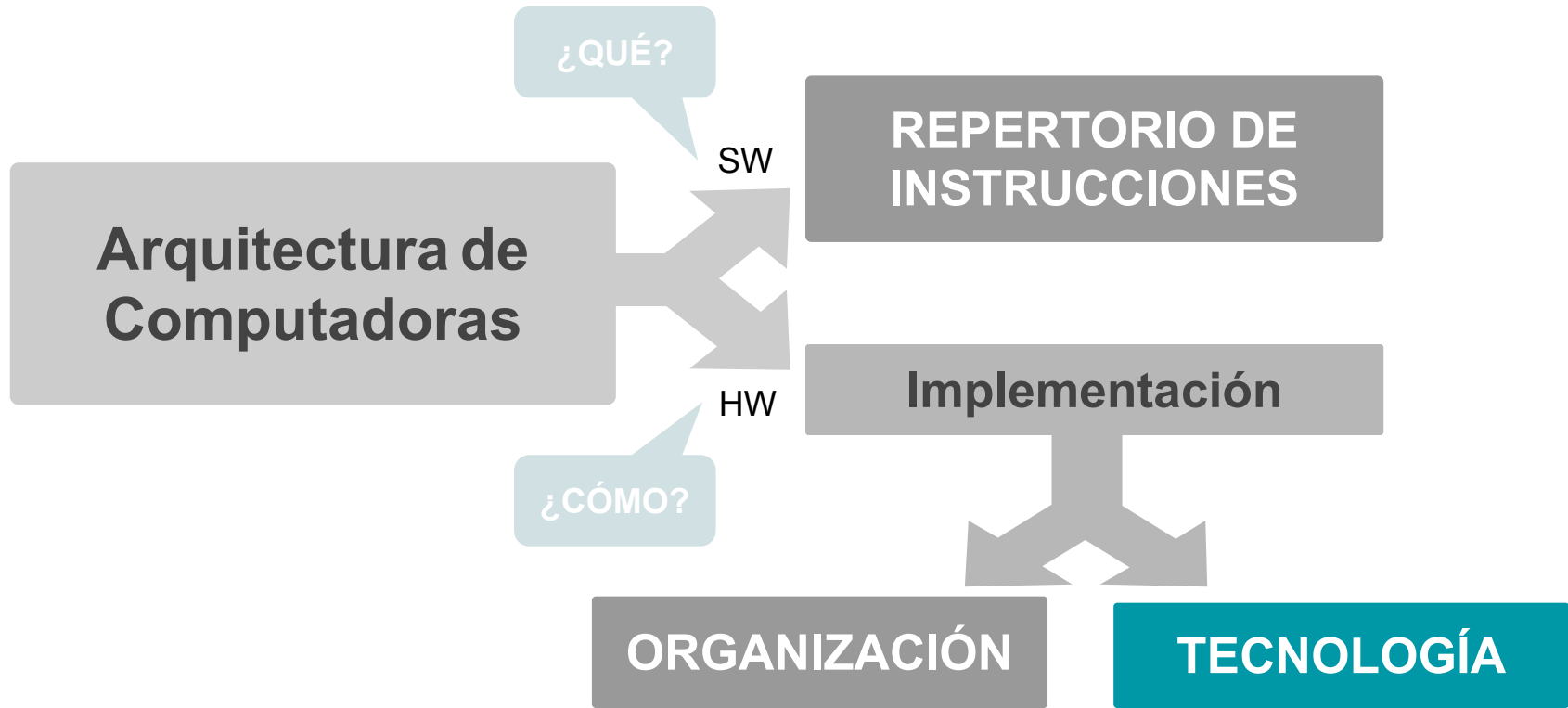


# CLASE 7

## TECNOLOGÍA DE IMPLEMENTACIÓN



## Mejoras del REPERTORIO DE INSTRUCCIONES (ISA)

- Aumento de la cantidad de memoria
- Aumento del número de registros de propósitos generales: disminución de los accesos a memoria.
- Registros de uso específico
- Mejora de la ALU: operaciones y tipos de datos
- Ampliación del repertorio de saltos condicionales: mejora en las bifurcaciones
- Loops: mejoras en las repeticiones
- Nuevos modos de direccionamiento y registros de uso específico: mejora en el acceso a los datos
- La pila: mejora en la implementación de subrutinas
- El sistema de entrada/salida, interrupciones

## Mejoras de la ORGANIZACIÓN del procesador

Optimización del ciclo de instrucción. Paralelismo. Mayor productividad.

- Arquitectura Harvard
- Segmentación

## **CLASE DE HOY: mejoras en la TECNOLOGÍA DE IMPLEMENTACIÓN del procesador**

- **Tecnología CMOS de fabricación -> Tensión de alimentación, frecuencia máxima de operación, densidad, consumo de potencia, modos de bajo consumo, costo, disponibilidad, encapsulado**

# Principales aspectos relacionados con la tecnología de implementación

## Tecnología CMOS de fabricación (xx nm process)

Fotolitografía, múltiples capas, costo, 6-8 semanas (nuevas 14/10/7 nm 15 semanas), fabs completamente automatizadas, clean-room, wafers, FOUPs, etc. - Wikipedia

<https://www.youtube.com/watch?v=6OeYwfDkiH8> (TSMC 2021)

<https://www.youtube.com/watch?v=0wrsZog8qXg> (2010)

## Parámetros

Tensión de alimentación (rango)

Frecuencia máxima de operación (rango, dependiente de la tensión de operación)

Consumo de potencia (DMIPS/Watt, dependiente de la tensión)

Modos de bajo consumo

Densidad de transistores (tr/mm<sup>2</sup>)

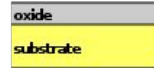
Costo

Disponibilidad

Encapsulado

$$P = f C V^2$$

a. Prepare wafer



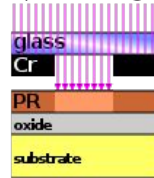
b. Apply photoresist



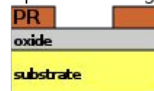
c. Align photomask



d. Expose to UV light



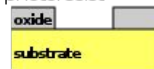
e. Develop and remove photoresist exposed to UV light



f. Etch exposed oxide



g. Remove remaining photoresist

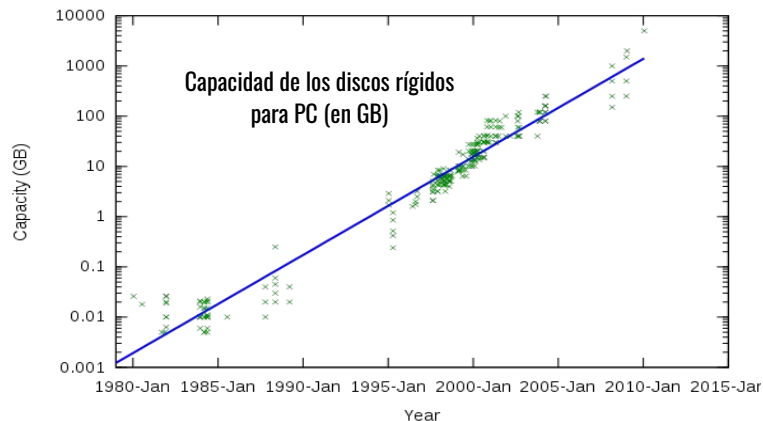
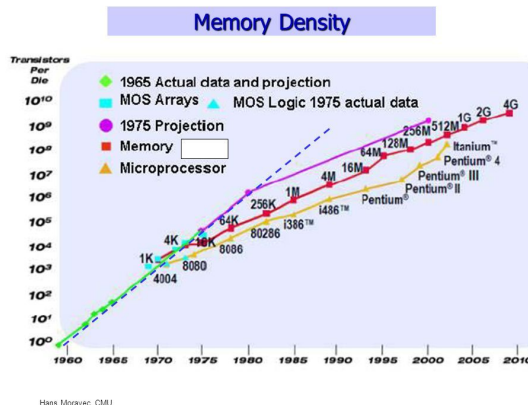
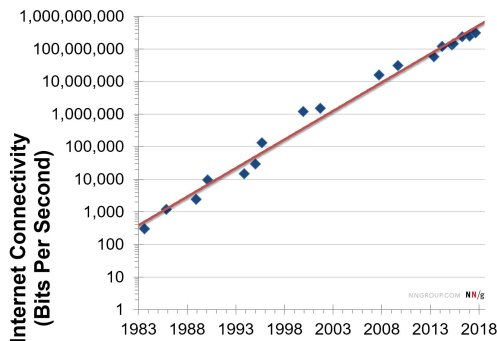


# TECNOLOGÍA DE IMPLEMENTACIÓN

Evolución exponencial sostenida de las tecnologías que condicionan el diseño de un procesador:

- Circuitos integrados (densidad de transistores 50%/año)
- DRAM semiconductor (densidad celdas 40-60%/año)
- Almacenamiento (capacidad 100%/año)
- Networking (ancho de banda 50-60%/año)

La ISA debe sobrevivir a lo largo de ciclos de al menos 5 años  
(2 de diseño + 3-2 de producción).



<https://ourworldindata.org/technological-progress>

# Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World  
in Data

## Transistor count

50,000,000,000

10,000,000,000

5,000,000,000

1,000,000,000

500,000,000

100,000,000

50,000,000

10,000,000

5,000,000

1,000,000

500,000

100,000

50,000

10,000

5,000

1,000

Número de transistores por  
integrado.

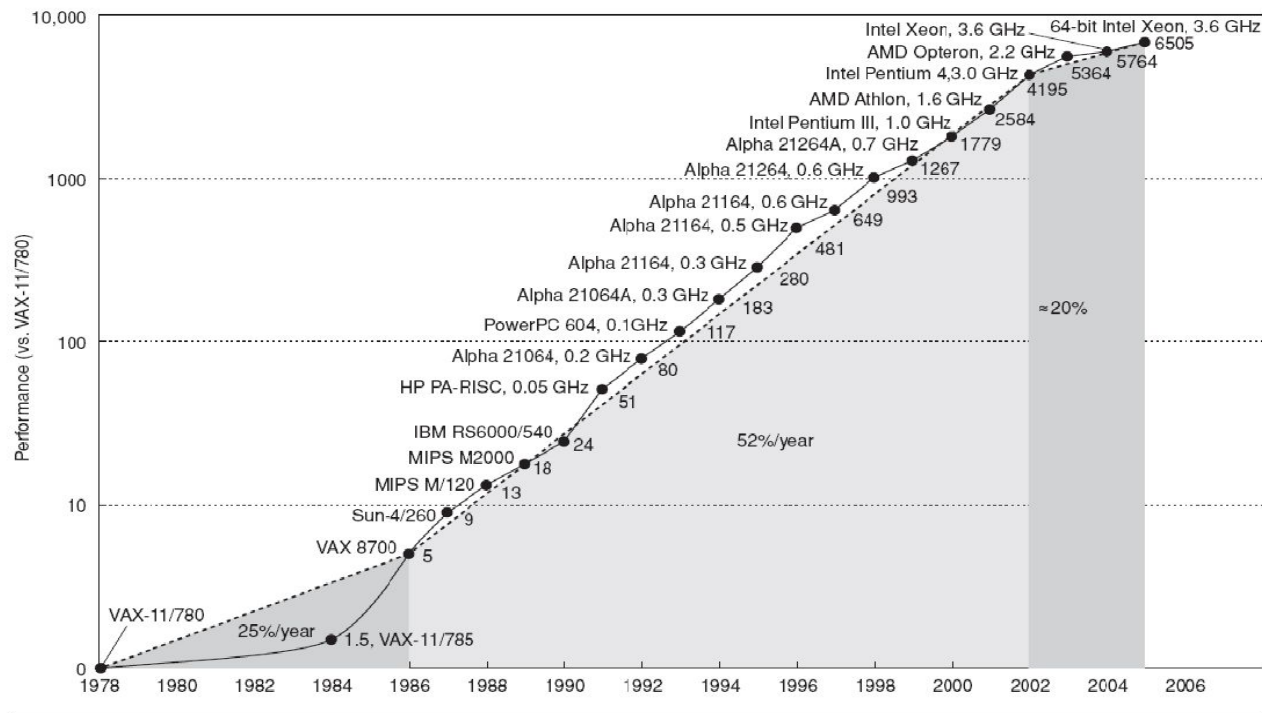
Duplica cada dos años, crecimiento  
exponencial sostenido.

"Profecía autocumplida de la industria"

Data source: Wikipedia ([wikipedia.org/wiki/Transistor\\_count](https://wikipedia.org/wiki/Transistor_count))

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.



## Performance relativa a la VAX-11/780 (1978)

*Crecimiento exponencial sostenido (1986-2002). Luego desaceleración (a pesar de que tanto la densidad de transistores como el resto de las tecnologías acompañaron). El aumento del número de transistores no repercute directamente en la performance. Depende de los avances en la organización del procesador. Límite en el frequency scaling.*

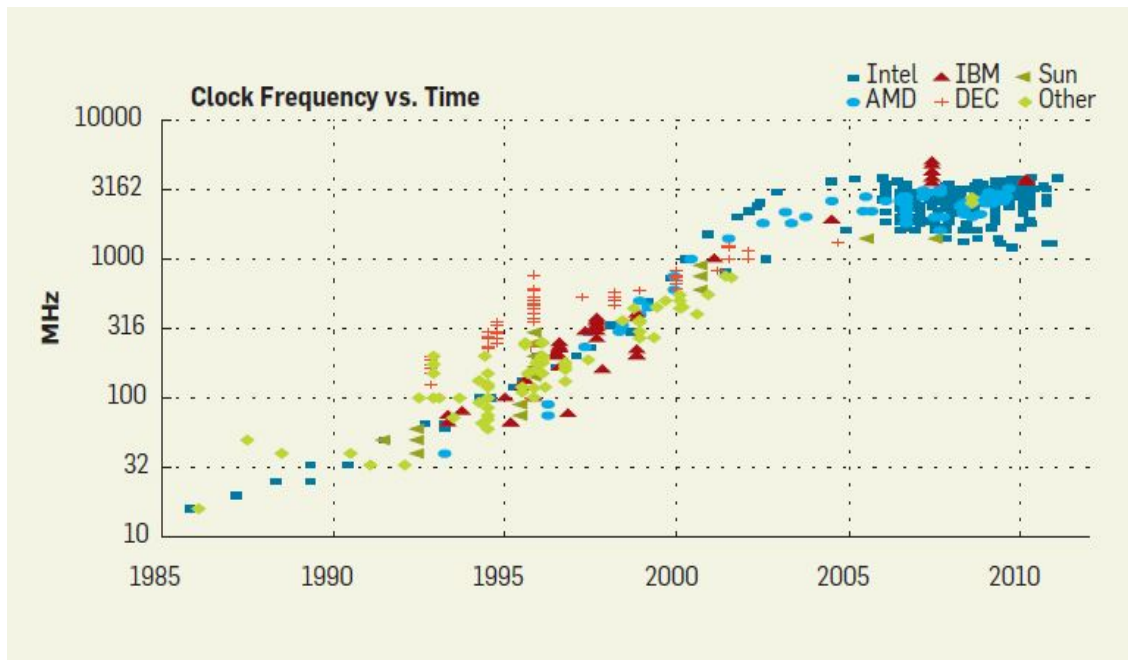
<http://drhart.ucoz.com/index/digital/0-117>



VAX 11/780 (first two cabinet sections), shown with Unibus expansion cabinet (middle cabinet section), two tape drives, two RP05 or RP06 removable pack disk drives, a DECwriter printing terminal, and a VT52 CRT terminal.



## 2004 - El fin del “frequency scaling”



$$t = \frac{N \times CPI}{f_{clock}}$$

Límite en la potencia disipada (que es proporcional a la frecuencia de operación).

$$P = f C V^2$$

(70-100 W)

Sin embargo, a pesar de los problemas de potencia, la Ley de Moore se mantiene.

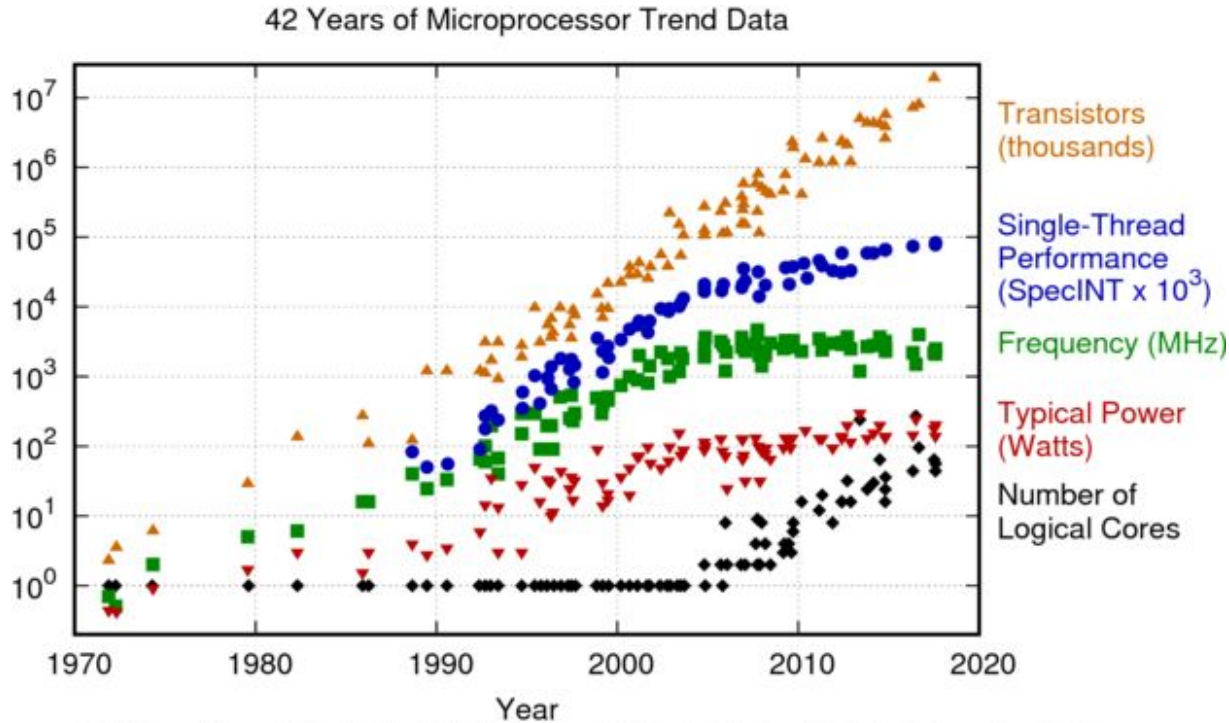
Evolución natural a múltiples núcleos o “parallel scaling”.

NOTA: estrategia de bajar la tensión, deja afuera a los uC

En 2004 Intel cancela sus proyectos monoprocesador Tejas y Jayhawk (server), sucesores del Pentium 4, para reorientar su estrategia a las arquitecturas dual core.

Ver: Wikipedia - Frequency Scaling [https://en.wikipedia.org/wiki/Frequency\\_scaling](https://en.wikipedia.org/wiki/Frequency_scaling)





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 2010-2017 by K. Rupp

## POSIBLES SOLUCIONES:

- multiple cores on one die
- multiple dies in one package
- multiple packages in one system unit

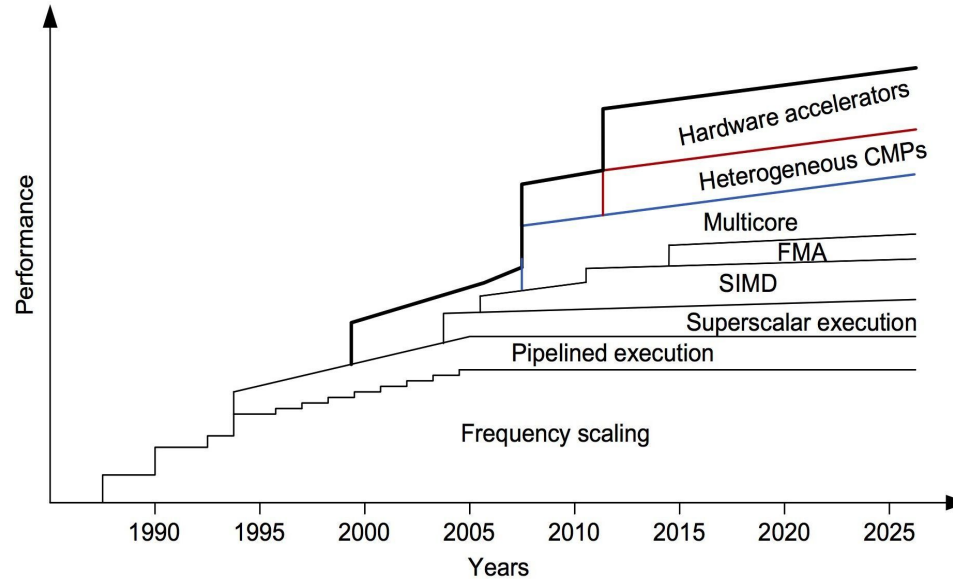
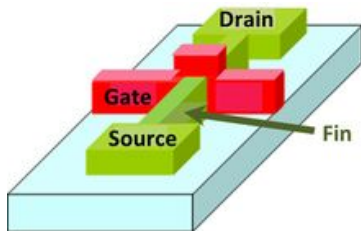
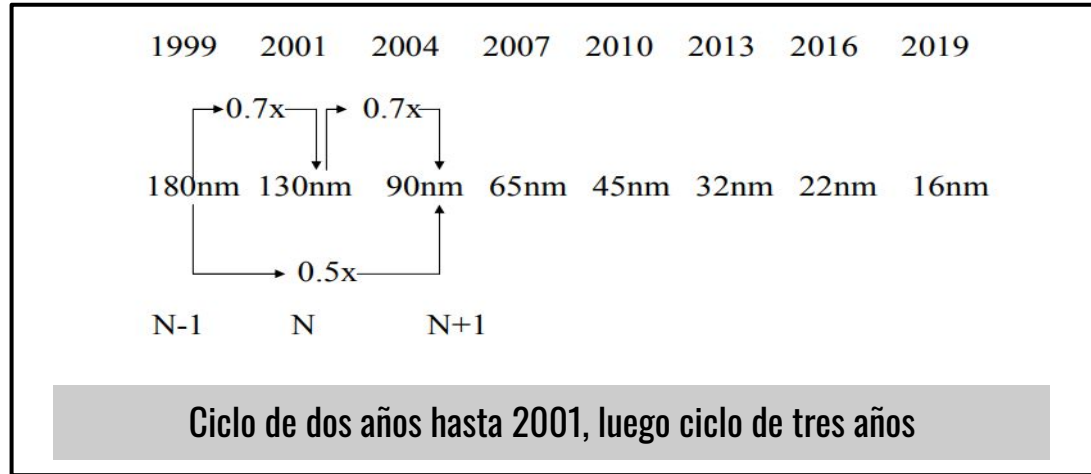


Fig. presents the main drivers that helped improve the performance of computing systems over the last three decades as well as the trends for the upcoming years. Clearly, frequency scaling was the main driver for improving performance until the mid-2000s. Since then, more complex technologies have been employed to increase the computational capacity of computing platforms, including more aggressive pipelining execution, superscalar execution, multicore architectures, single instruction, multiple data (SIMD) support, fused multiply-add (FMA) units, and chip multiprocessors (CMPs). The combination of hardware accelerators with many-core architectures is currently an important source of performance gains for emerging high-performance heterogeneous architectures accelerators, and we believe will continue to be prominent in the upcoming years.

<https://www.embedded.com/design/real-time-and-performance/4459159/High-performance-embedded-computing----Target-architectures>

# Tecnología del proceso CMOS

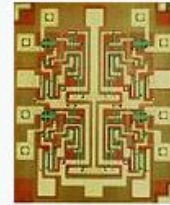
La unidad de medida era  $\frac{1}{2}$  celda DRAM : “expected average half-pitch of a memory cell”



La inversión necesaria para sostener el crecimiento es cada vez mayor: Rock's Law (o segunda ley de Moore): el costo de una planta de fabricación se duplica cada 4 años

Un elemento más pequeño implica mayor cantidad de transistores disponibles, conmutación más rápida y menor potencia disipada.  
LIMITE? gate 5 nm  $\rightarrow$  tunneling  
CMOS  $\rightarrow$  FinFET (3D multigate, 10+nm)

## Semiconductor device fabrication



### MOSFET scaling (process nodes)

10  $\mu\text{m}$  – 1971  
6  $\mu\text{m}$  – 1974  
3  $\mu\text{m}$  – 1977  
1.5  $\mu\text{m}$  – 1981  
1  $\mu\text{m}$  – 1984  
800 nm – 1987  
600 nm – 1990  
350 nm – 1993  
250 nm – 1996  
180 nm – 1999  
130 nm – 2001  
90 nm – 2003  
65 nm – 2005  
45 nm – 2007  
32 nm – 2009  
22 nm – 2012  
14 nm – 2014  
10 nm – 2016  
7 nm – 2018  
5 nm – 2020  
3 nm – 2022

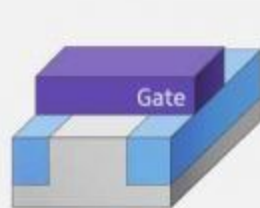
Future  
2 nm ~ 2024

2022

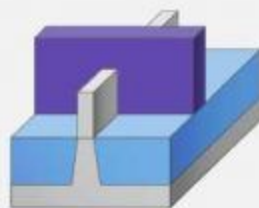
ATmega328

Intel Core i5

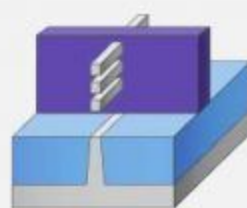
iPhone 12/13



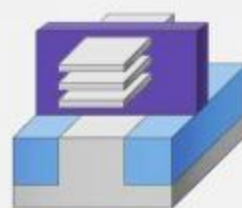
Planar FET



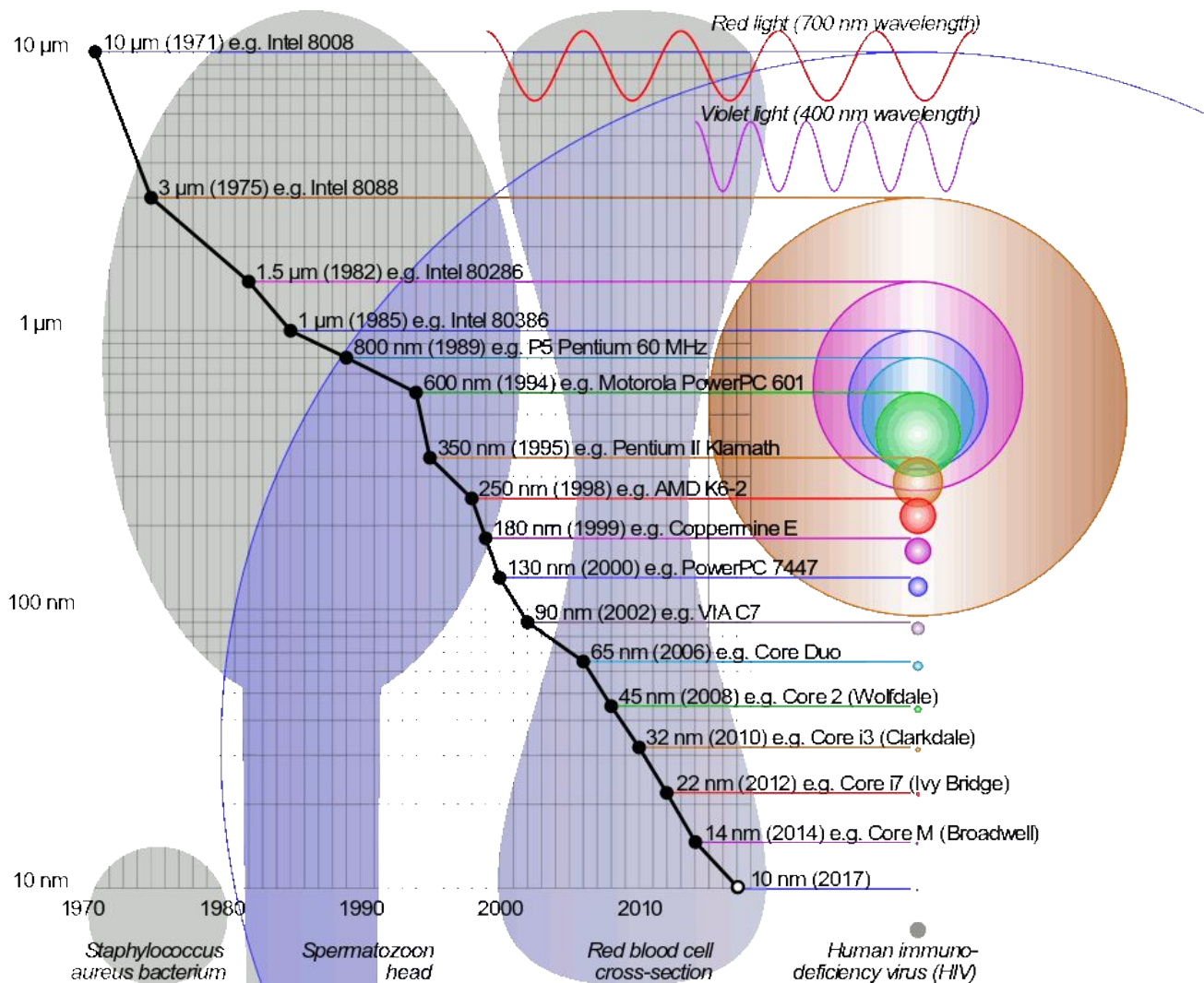
FinFET

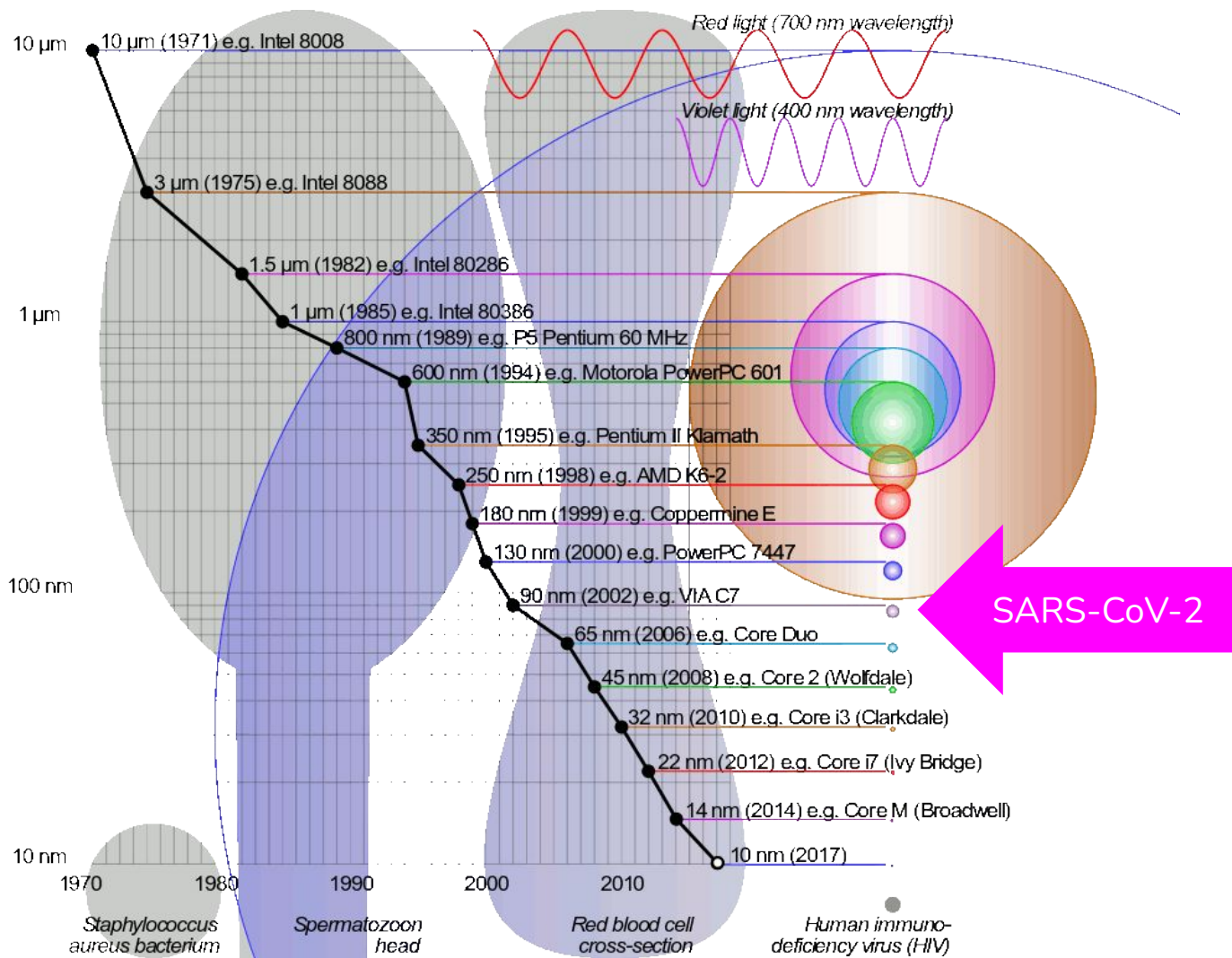


GAAFET  
(Nanowire)



MBCFET™  
(Nanosheet)





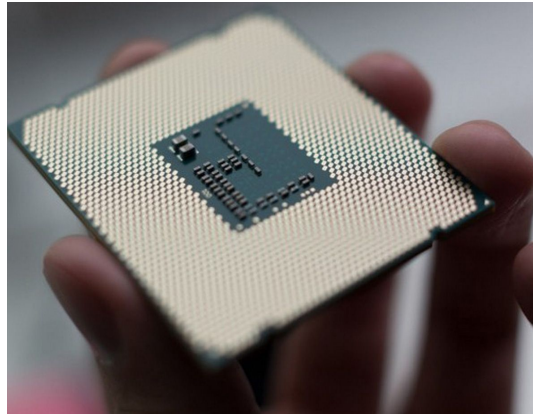
# Tecnología del proceso CMOS

Processor	Transistor count	Date of introduction	Process
Intel 4004	2,300	1971	10,000 nm
Core AVR	<b>48,000</b>	1993	350 nm
Apple A15	15,000,000,000	2021	5 nm



**Evolución +140 dB**

Entran 300,000 AVR en un A15





# Tecnología de microcontroladores

¿Por qué los microcontroladores no utilizan la tecnología de punta?

- Los nuevos procesos requieren bajar la tensión de alimentación, lo cual no es conveniente para los microcontroladores que interactúan con periféricos.
- La mayor parte de la superficie de la pastilla la toman los pads de I/O y sus drivers. Hacer más fino el resto del chip no produce una mejora considerable.
- Costo: La fabricación de una máscara para un proceso de punta cuesta más de 1M, frente a 10K. Priorizan la variedad de productos.
- Velocidad de lectura de la memoria Flash.
- Emisión de ruido.

*Una de las actuales fabs de Microchip fue utilizada en su momento para fabricar memorias de alta gama (Fujitsu). "The baseline N-well CMOS process is released on mature manufacturing lines that support 1.2, 0.9, 0.7, 0.5, 0.4 and 0.25 micron technologies"*

Function	Transistor count
MUX 2-input with TG	6
MUX 4-input with TG	18
NOT MUX 2-input	8
MUX 4-input	24
1-bit Adder full	28
1-bit Adder-subtractor	48
Latch, D gated	8
Flip-flop, edge triggered dynamic D with reset	12
8-bit multiplier	3,000
16-bit multiplier	9,000
32-bit multiplier	21,000





# Plantas de fabricación de semiconductores

Semiconductor fabrication plants (fab/foundry/pure-play) y diseñadores (fabless).  
Integrated device manufacturers (IDM), toman ambas ramas del negocio.

TSMC Taiwan Semiconductors (Taiwan) y Samsung Electronics (Corea del Sur) son los únicos fabricantes de semiconductores que han podido seguir el paso en esta década (sub 14 nm y 10 nm actualmente en producción). Intel (USA) y Global Foundries (USA) vienen atrás.



# Ejemplo



Tecnología **GlobalFoundries** 130 nm

<https://www.globalfoundries.com/sites/default/files/product-briefs/pb-130glpee.pdf>

Las dimensiones del transistor mínimo son las siguientes: largo del canal  $L = 0.12 \mu\text{m}$  y ancho del canal  $W = 0.16 \mu\text{m}$ , tanto para los transistores nMOS como para los pMOS.

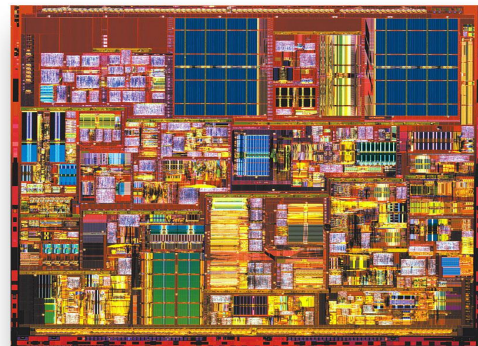
**SAMSUNG**



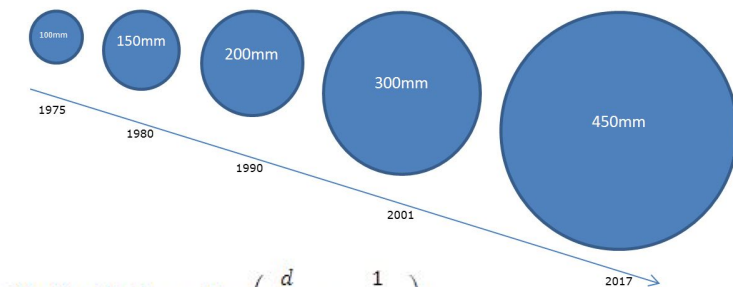
## Tamaño de la pastilla (die size)

Además del límite en la potencia máxima disipada (~70W)

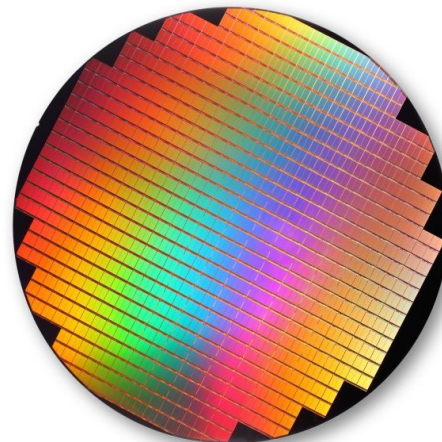
- Yield del wafer (rendimiento, cuántos fallan y desperdicio)
- Encapsulado (número de patas)

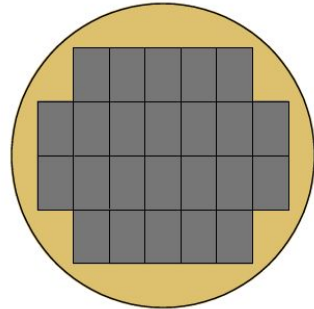
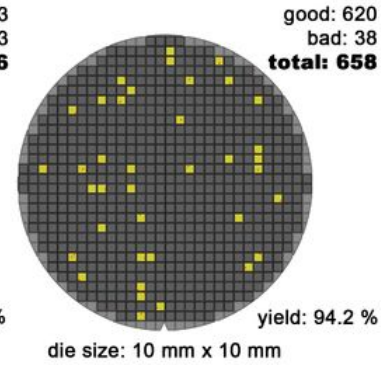
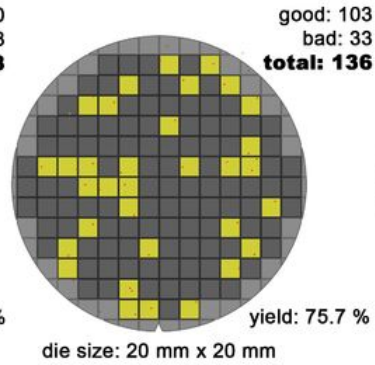
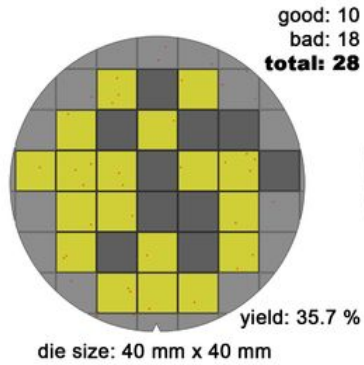


## Tamaño de la oblea (wafer size)

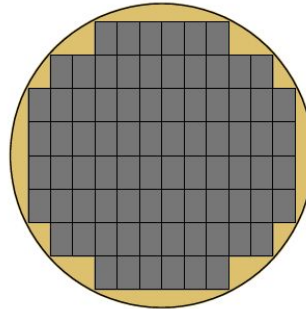


$$\text{Die Per Wafer} = d * \pi \left( \frac{d}{4 * S} - \frac{1}{\sqrt{2} * S} \right)$$

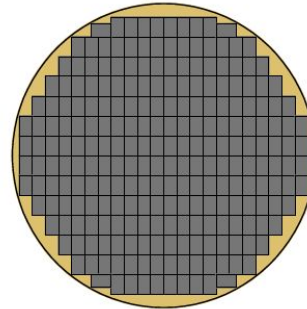




FULL FRAME SENSOR (24mm x 36mm)  
ON 8" SILICON WAFER (20.32 cm)  
Maximum yield = 24 Sensors  
Waste = 36%



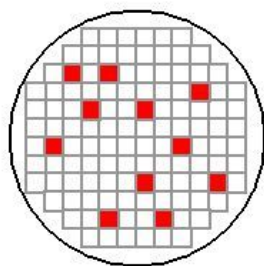
7D Sensor (22.3 mm x 14.9 mm)  
ON 8" SILICON WAFER (20.32 cm)  
Maximum yield = 80 Sensors  
Waste = 18%



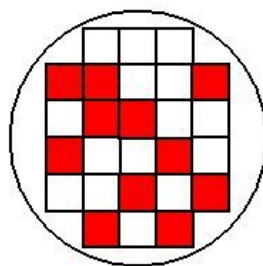
Nikon 1 Sensor (13.2 mm x 8.8 mm)  
ON 8" SILICON WAFER (20.32 cm)  
Maximum yield = 244 Sensors  
Waste = 12.6%



## Effect of Die Size on Yield



120 dies, 109 good



26 dies, 15 good

□ Good Die

■ Defective Die

**Dramatic decrease in yield with larger dies**

$$\text{Yield} = (\text{Number of Good Dies}) / (\text{Total Number of Dies})$$

$$\text{Yield} = \frac{1}{(1 + (\text{Defect per area} \times \text{Die area} / 2))^2}$$

$$\text{Die Cost} = (\text{Wafer Cost}) / (\text{Dies per Wafer} \times \text{Yield})$$

Intel 8086	29,000	1978	Intel	3 $\mu\text{m}$	33 $\text{mm}^2$
Intel 8088	29,000	1979	Intel	3 $\mu\text{m}$	33 $\text{mm}^2$
<b>Motorola 68000</b>	<b>68,000</b>	<b>1979</b>	<b>Motorola</b>	<b>4 <math>\mu\text{m}</math></b>	<b>44 <math>\text{mm}^2</math></b>
Intel 80286	134,000	1982	Intel	1.5 $\mu\text{m}$	49 $\text{mm}^2$
Intel 80386	275,000	1985	Intel	1.5 $\mu\text{m}$	104 $\text{mm}^2$
Intel 80486	1,180,235	1989	Intel	1 $\mu\text{m}$	173 $\text{mm}^2$
Pentium	3,100,000	1993	Intel	0.8 $\mu\text{m}$	294 $\text{mm}^2$
<b>ARM 7</b>	<b>600,000</b>	<b>1994</b>	<b>ARM</b>	<b>...</b>	
Pentium II	7,500,000	1997	Intel	0.35 $\mu\text{m}$	195 $\text{mm}^2$
Pentium III	9,500,000	1999	Intel	0.25 $\mu\text{m}$	128 $\text{mm}^2$
Pentium 4	42,000,000	2000	Intel	180 nm	217 $\text{mm}^2$
<b>Itanium 2</b>	<b>592,000,000</b>	<b>2004</b>	<b>Intel</b>	<b>130 nm</b>	<b>432 <math>\text{mm}^2</math></b>
<b>Cell</b>	<b>241,000,000</b>	<b>2006</b>	<b>IBM</b>	<b>90 nm</b>	<b>221 <math>\text{mm}^2</math></b>
Core 2 Duo	291,000,000	2006	Intel	65 nm	143 $\text{mm}^2$
<b>ARM Cortex-A9</b>	<b>26,000,000</b>	<b>2007</b>	<b>ARM</b>	<b>...</b>	
<b>Xeon Phi</b>	<b>5,000,000,000</b>	<b>2012</b>	<b>Intel</b>	<b>22 nm</b>	
<b>GK110 Kepler</b>	<b>7,080,000,000</b>	<b>2012</b>	<b>NVIDIA</b>	<b>28 nm</b>	<b>561 <math>\text{mm}^2</math></b>
<b>Virtex-7</b>	<b>6,800,000,000</b>	<b>2011</b>	<b>Xilinx</b>	<b>28 nm</b>	
<b>A15 Bionic</b>	<b>15,000,000,000</b>	<b>2021</b>	<b>Apple</b>	<b>5 nm</b>	

#### EJEMPLO

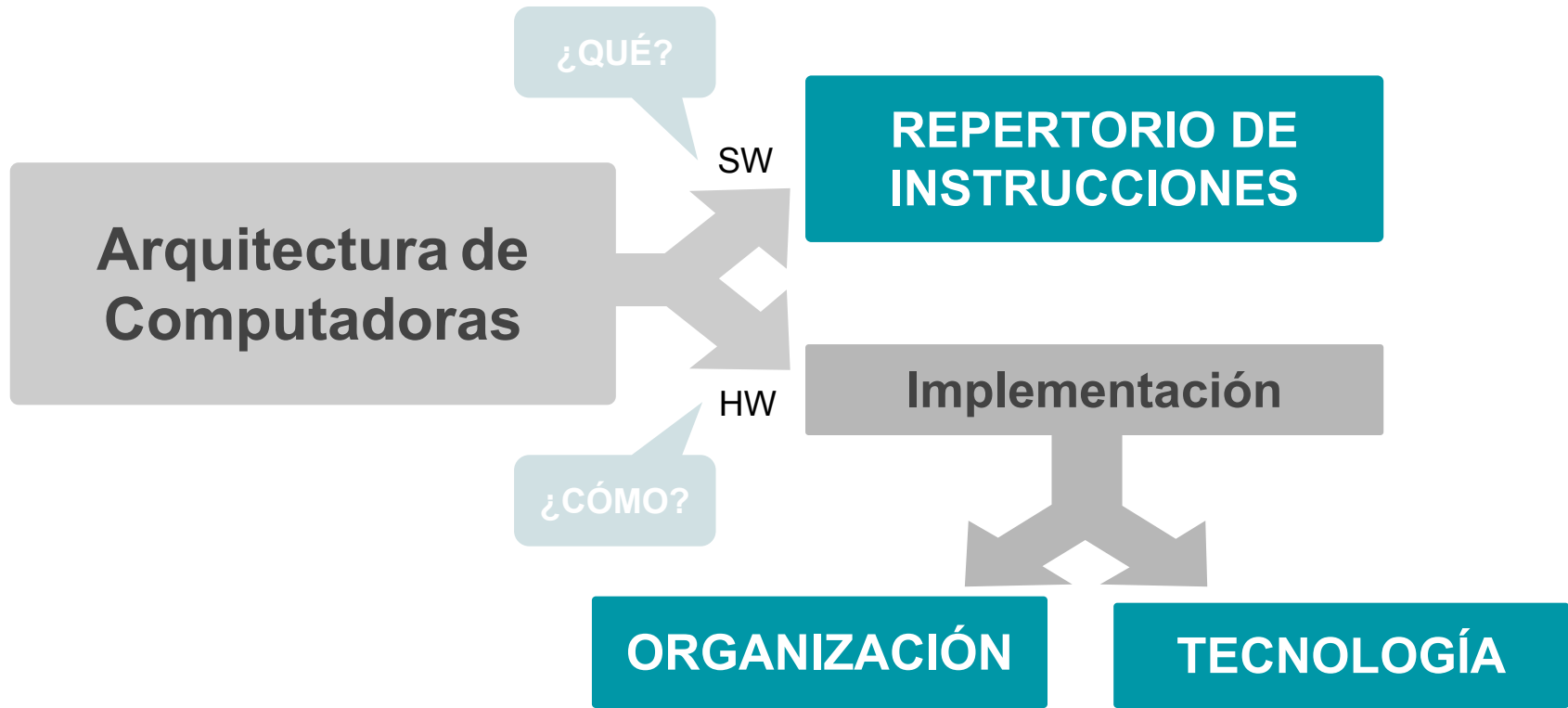
Sumador segmentado en Punto Flotante  
23347 transistores

COMPLETAR LA TABLA CON  $\mu\text{C}$  MÁS  
COMUNES

# Consolas de juego

4ta Generación	5ta Generación	6ta Generación	7ma Generación	8va Generación
1990	1995	2000	2005	2013
16b	32b	64b	128b	64b
Sega Génesis Motorola 68000	Sega Saturn Hitachi SupeH RISC	Sega Dreamcast Hitachi SuperH RISC		
Super Nintendo WDC W65C816	Nintendo 64 MIPS R4200	Nintendo Gamecube POWER Gekko	Nintendo Wii POWER Broadway	Nintendo Wii U POWER7
	PlayStation MIPS 3000	PlayStation II Emotion Eng. (MIPS)	PlayStation III POWER Cell	PlayStation 4 AMD Jaguarx86-64 AMD Radeon
		Xbox Pentium III	Xbox 360 POWER Xenon	Xbox One AMD Jaguarx86-64 AMD Radeon
Cartridge	CD	DVD	Bluray	Bluray/Internet
		Ethernet	WiFi	Gigabit Ethernet
	PII/PowerPC	PIII/PIV/AMD K7	Core/AMD64	Athlon





¿QUÉ?

SW

REPERTORIO DE INSTRUCCIONES

Arquitectura de Computadoras

HW

Implementación

¿CÓMO?

ISA

$$t = \frac{N \times CPI}{f_{clock}}$$

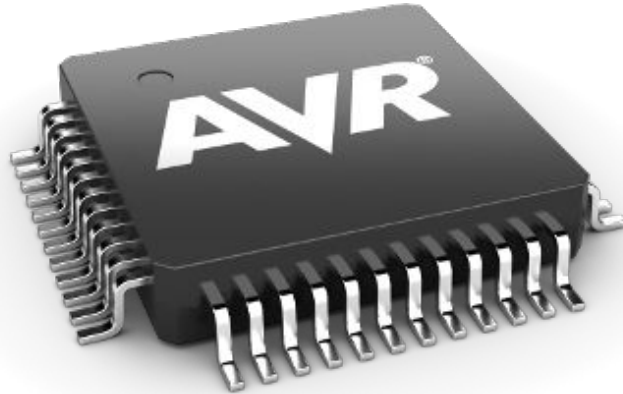
ORG

ORGANIZACIÓN

TECNOLOGÍA

**Intervalo 15'**

**A continuación: introducción a la arquitectura AVR**



## Microcontroladores ATMEL ATtiny/ATmega

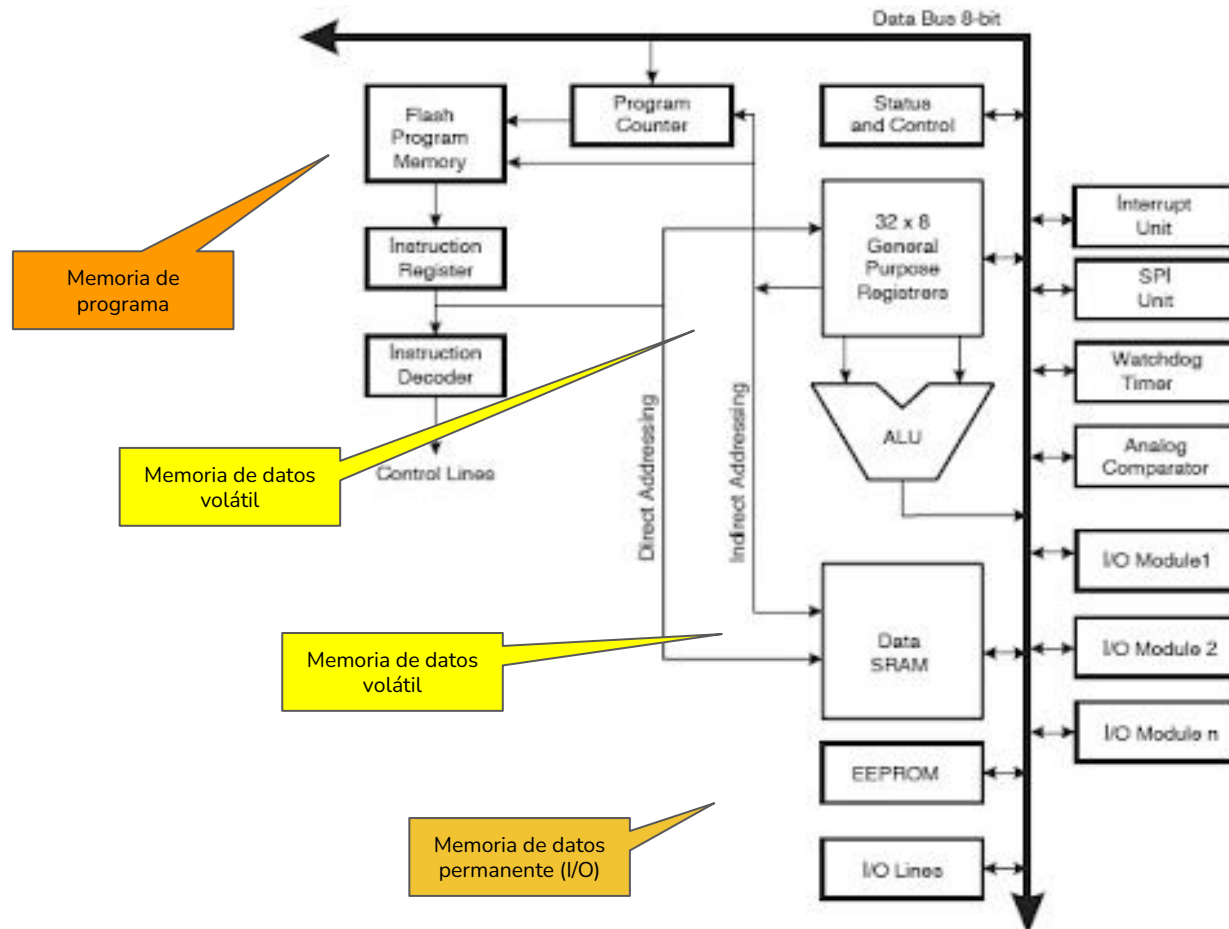
- AVR CPU ←
- Clock
- Memoria Flash (programa)
- Memoria DRAM (datos)
- Modulos de entrada/salida
  - I/O ports
  - Interfaces de comunicación
  - Temporizadores/contadores
  - ADC, etc.
- Single supply, low cost, low power

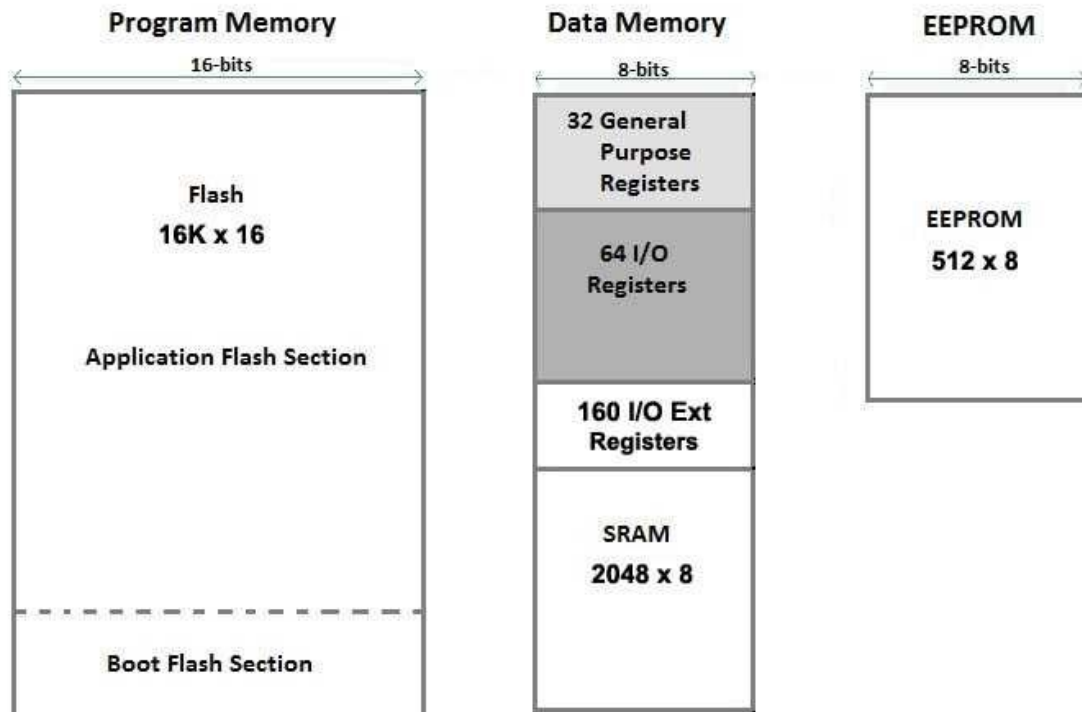


AVR: <https://microchipdeveloper.com/8avr:start>

PIC: <https://microchipdeveloper.com/8bit:start>

AVRfreaks forum  
<https://www.avrfreaks.net/>





# Modos de direccionamiento: inmediato, directo e indirecto

Immediate Addressing: `andi r16, $0F` (el operando viene de la instrucción)

Register Direct Addressing: `and r16, r0` (el operando viene de un registro)

`in r25, PINA` (el operando viene de i/o)

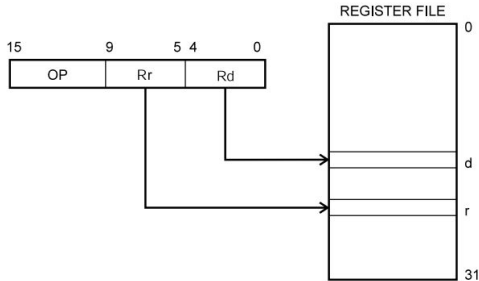
Data Direct Addressing: `lds r5, $F123` (la dirección está en la instrucción)

Indirect Addressing: `ld r11, X` (la dirección está guardada en un puntero)

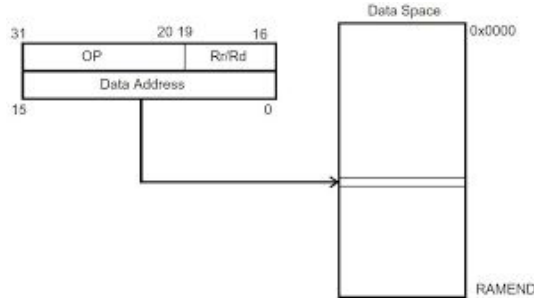
with displacement: `std Y+10, r14`

with Pre-decrement: `std -Y, r14`

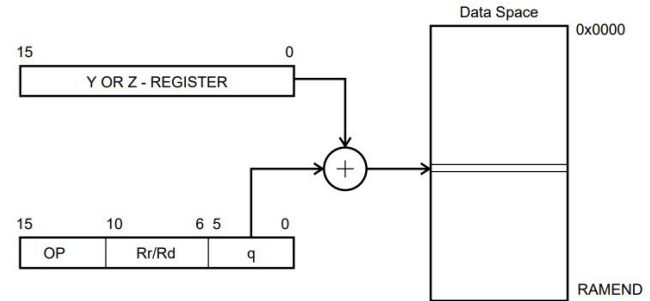
with Post-increment: `std Y+, r14`



Direct Register Addressing



Direct Data Addressing



Data Indirect with Displacement

# Arquitectura **AVR**<sup>®</sup>

## REPERTORIO DE INSTRUCCIONES

32 registros de 8 bits, ALU de 8 bits enteros, load/store tipo (0,2).  
IO mapeada en memoria.  
Dir. 16 bits (registros dobles) y pila.  
126 instrucciones, 7 modos de direccionamiento, interrupciones.

## Implementación

### ORGANIZACIÓN

Harvard (datos de 8 bits, instrucciones de 16 bits).  
Segmentación en dos etapas (F-E).

### TECNOLOGÍA

130 nm, core 48,000 tr, SMD plástico, 3.3V, 16 MHz.  
32K FLASH, 2K SRAM



ATmega328