

Dipartimento di Scienze Fisiche, Informatiche e Matematiche

2. I Computer: Astrazioni e Tecnologia

Architettura dei calcolatori [MN1-1143]

Corso di Laurea in INFORMATICA (D.M.270/04) [16-215] Anno accademico 2020/2021 Prof. Andrea Marongiu andrea.marongiu@unimore.it

È vietata la copia e la riproduzione dei contenuti e immagini in qualsiasi forma.

È inoltre vietata la redistribuzione e la pubblicazione dei contenuti e immagini non autorizzata espressamente dall'autore o dall'Università di Modena e Reggio Emilia."



- I computer sono pervasivi
- La loro continua evoluzione rende possibili applicazioni sempre nuove
 - Dispositivi personali
 - Auto a guida autonoma, droni
 - Human brain project
 - World Wide Web/Internet, Cloud, IoT
- Ciò è reso possible dal costante avanzamento della tecnologia
 - Legge di Moore

«the number of transistors in an IC doubles every two years»

Moore, G.E., Cramming more components onto integrated circuits. Electronics, 38(8), April 1965



Gordon E Moore (1929–)

Classi di Computer

Personal computers

- General purpose, varietà di software
- Trade-off costi/prestazioni



Server computers

- Network
- Alta capacità (storage), performance, affidabilità
- Da piccolo server privati ad interi edifici





Classi di Computer

Supercomputers

- Calcoli scientifici e ingegneristici ad alta richiesta di prestazioni
- Prestazioni più alte in assoluto
 - Una frazione relativamente piccolo del totale mercato, ma sta crescendo...



Embedded computers

- Nascosti come component di un sistema
- Vincoli di potenza/performance/costo stringenti
- Requisiti di tipo Real-time e di affidabilità



L'era Post PC

Personal Mobile Device (PMD)

- Dispositivi a batteria (basso consumo)
- Alta connettività (Internet and more)
- Centinaia di euro (basso costo)
- Smart phones, tablets, electronic glasses



Cloud computing

- Warehouse Scale Computers (WSC)
- Software as a Service (SaaS)
- Portion of software run on a PMD and a portion run in the Cloud
- Amazon, Microsoft, Google



Data Center



Microsoft Data Center eastern US - 2017



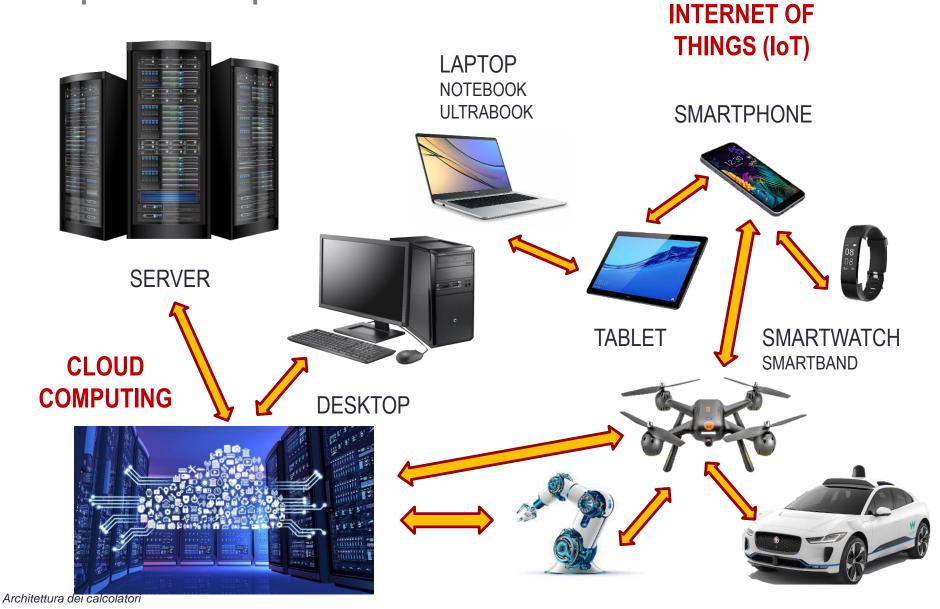
Planned expansion: 2km long...

Tipi di Computer

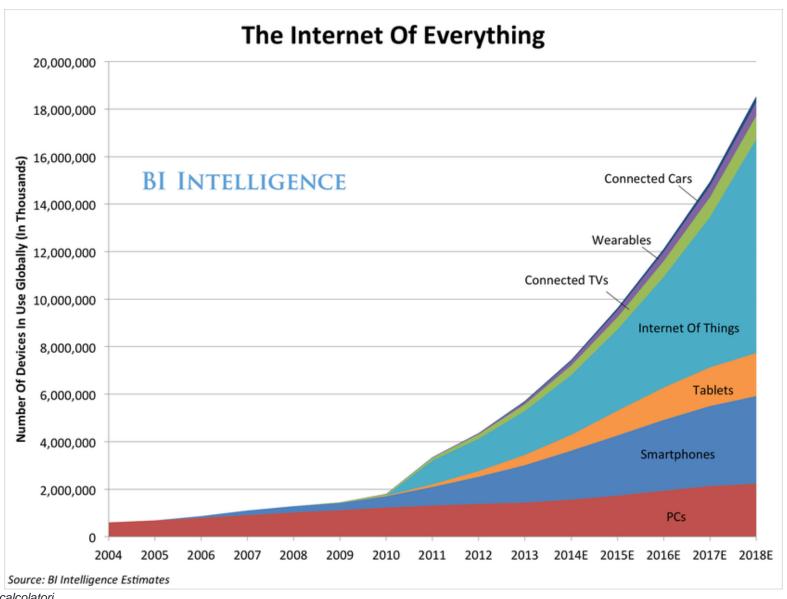


Dimensioni maggiori Alti consumi Maggiore performance Dimensioni ridotte Basso consumo Minore performance

Tipi di Computer

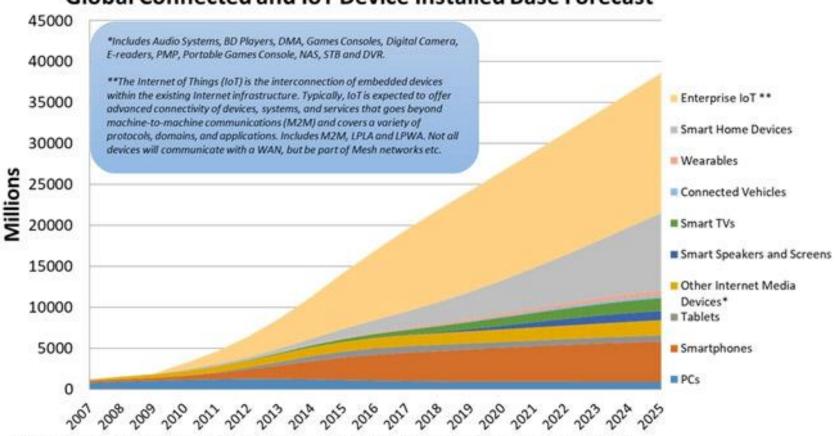


L'era Post PC – la visione del 2013



L'era Post PC – la visione odierna

Global Connected and IoT Device Installed Base Forecast



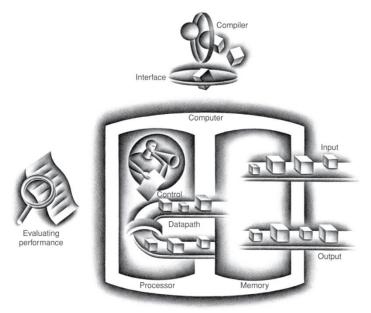
Source – Strategy Analytics research services, May 2019: IoT Strategies, Connected Home Devices, Connected Computing Devices, Wireless Smartphone Strategies, Wearable Device Ecosystem, Smart Home Strategies

- In breve...
- I computer sono pervasivi e onnipresenti in tutti gli aspetti della nostra vita quotidiana
- Lo studio di come sono progettati e programmati i computer è fondamentale in un mondo (e un mercato) che è dominato da questa tecnologia

- Ma cosa c'è dentro a un computer?
- E i vari tipi di computer sono tanto diversi tra loro?

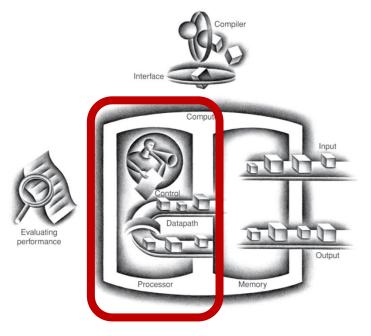
Ma cosa c'è dentro a un computer?

E i vari tipi di computer sono tanto diversi tra loro?



Componenti di un computer

- **[SPOILER]** Gli stessi per tutti i tipi di computer
 - Desktop, server, embedded
- Processore
- Memoria
- Input/Output (I/O)

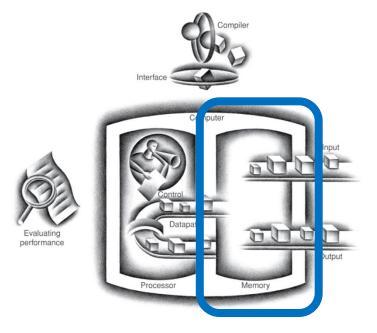


Componenti di un computer

- **[SPOILER]** Gli stessi per tutti i tipi di computer
 - Desktop, server, embedded

Processore

- Central Processing Unit (CPU)
- Il cuore di un computer: processa i dati in *input* per produrre l'*output*
- Datapath: esegue le operazioni sui dati
- Control: controlla il funzionamento del datapath, memoria, ...



Componenti di un computer

- **[SPOILER]** Gli stessi per tutti i tipi di computer
 - Desktop, server, embedded

Memoria

- Organizzata come una gerarchia
 - Performance/capacity/cost tradeoff
 - La DRAM è capiente, ma leeenta...
 - I *registri* (SRAM¹) sono veloci quanto la CPU, ma piiiccoli...
- La cache memory sta in mezzo
 - SRAM

1. SRAM is faster but less dense, and hence more expensive, than DRAM

E se spengo il computer?

- La memoria principale (DRAM) è volatile
 - Perde tutta l'informazione quando spengo il PC



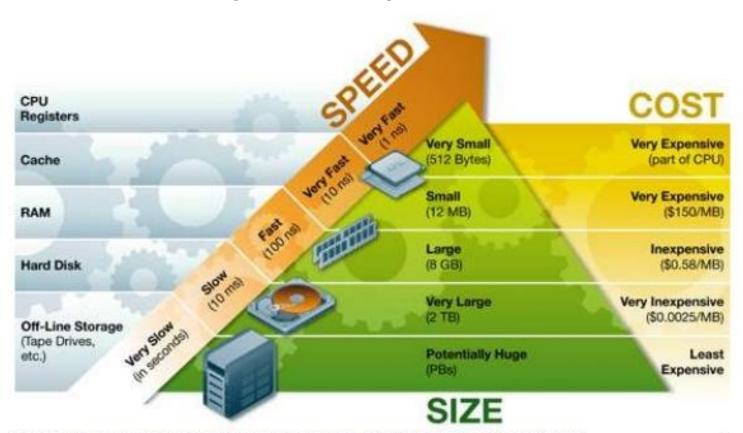
- HARD DISK Memoria secondaria non-volatile
 - Magnetic disk (HDD)
 - Flash memory (SD, SSD)
 - Optical disk (CDROM, DVD)



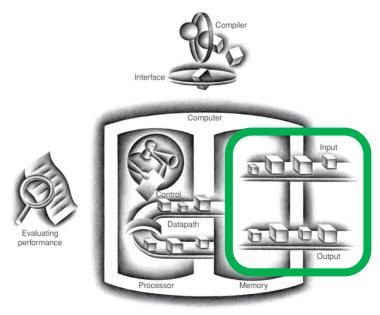


Architettura dei calcolatori HDD SSD

The memory hierarchy



Source: http://www.ts.avnet.com/uk/products and solutions/storage/hierarchy.html



Componenti di un computer

- **[SPOILER]** Gli stessi per tutti i tipi di computer
 - Desktop, server, embedded
- Input/output (I/O)
 - User-interface devices
 - Display, keyboard, mouse
 - Network adapters
 - For communicating with other computers

Network adapters

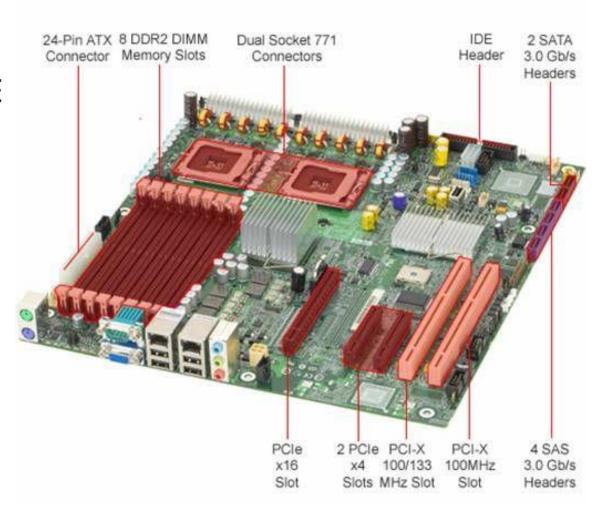
- Communication, resource sharing, nonlocal access
- Local area network (LAN)
 - > Ethernet (10/100 Gbit/s)
- Wide area network (WAN): the Internet
- Wireless network (IEEE 802.11)
 - → WiFi, Bluetooth → 1-100 Mbit/s



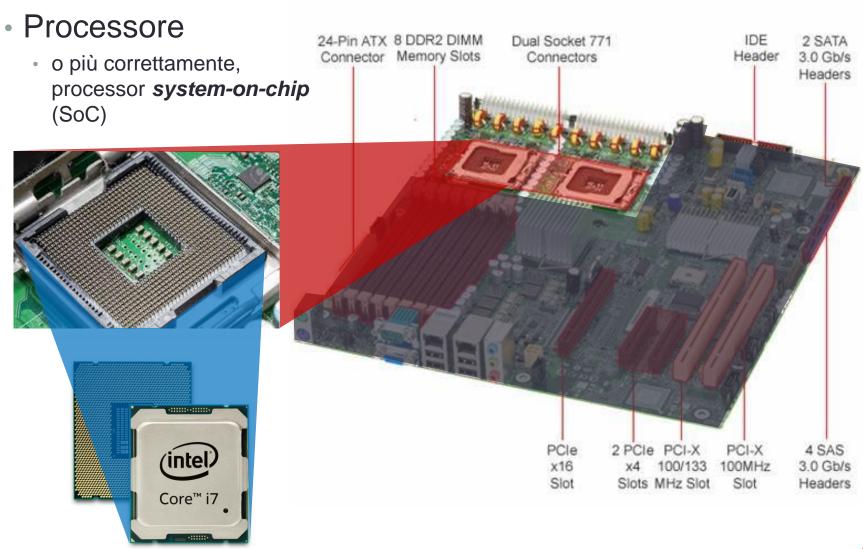


- Ma cosa c'è dentro a un computer?
- E i vari tipi di computer sono tanto diversi tra loro?

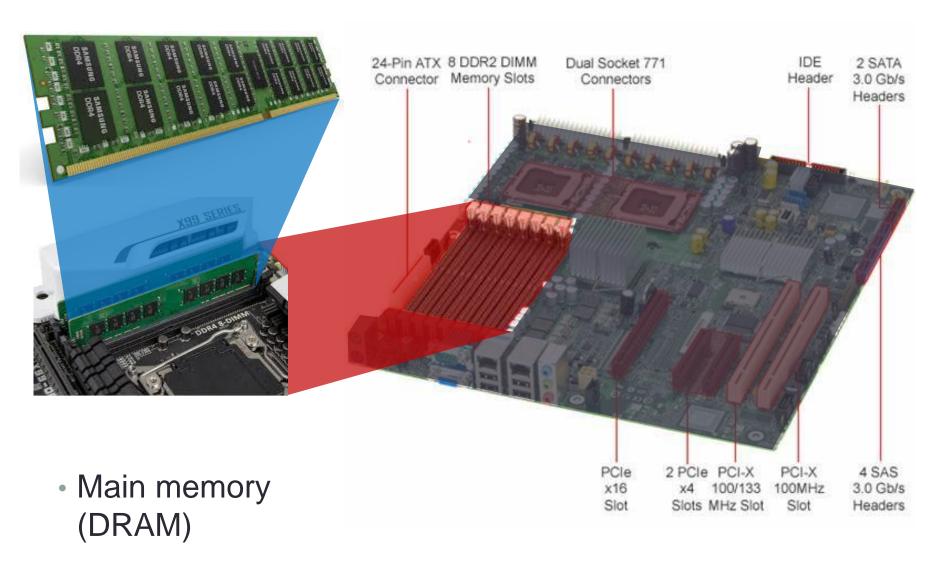
- Desktop PC
- SCHEDA MADRE



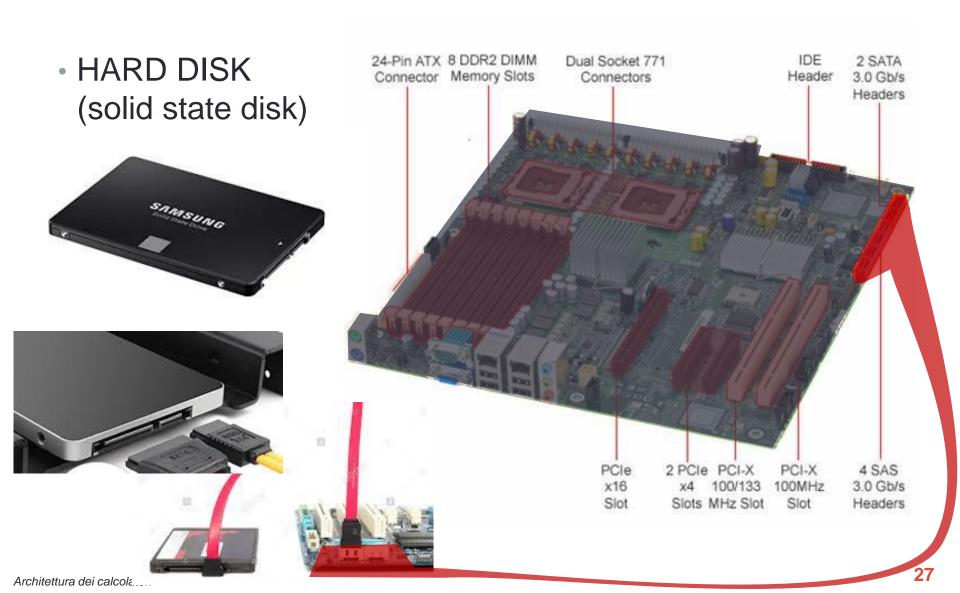
Architettura dei calcolatori

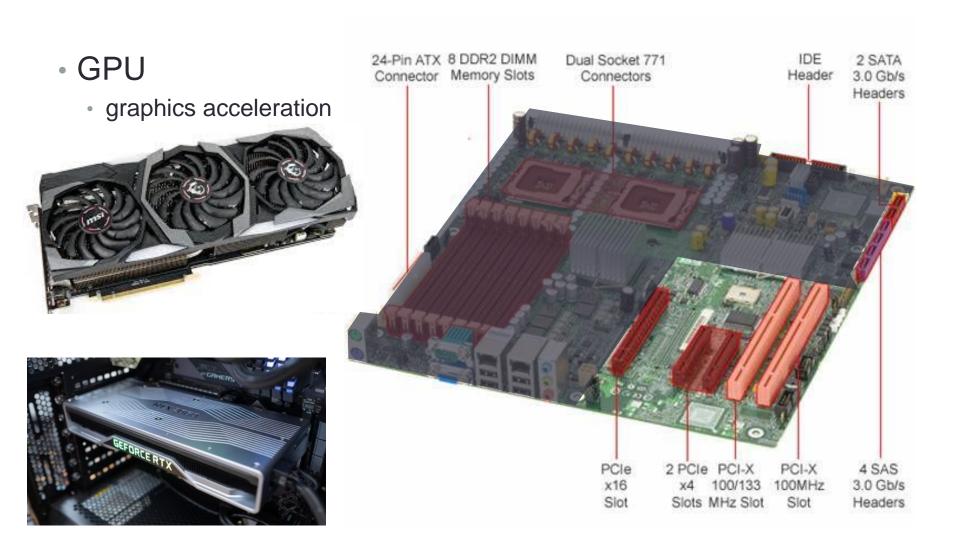


25



Architettura dei calcolatori

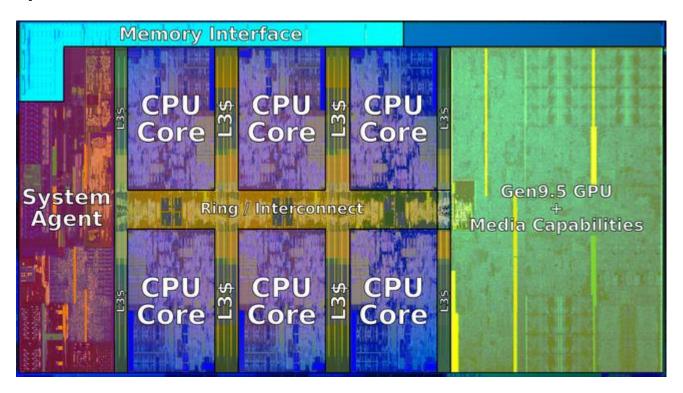




Architettura dei calcolatori 28

E cosa c'è dentro al **Processor SoC?**

- Coffee Lake system-on-chip (2018)
- Processo a 14nm
- Multi-Core (2 to 8) CPU
- GPU
- Media

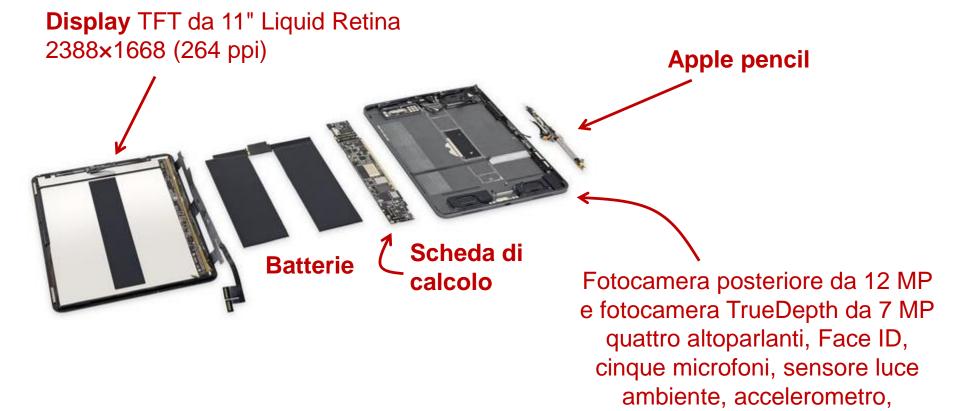


29 Architettura dei calcolatori

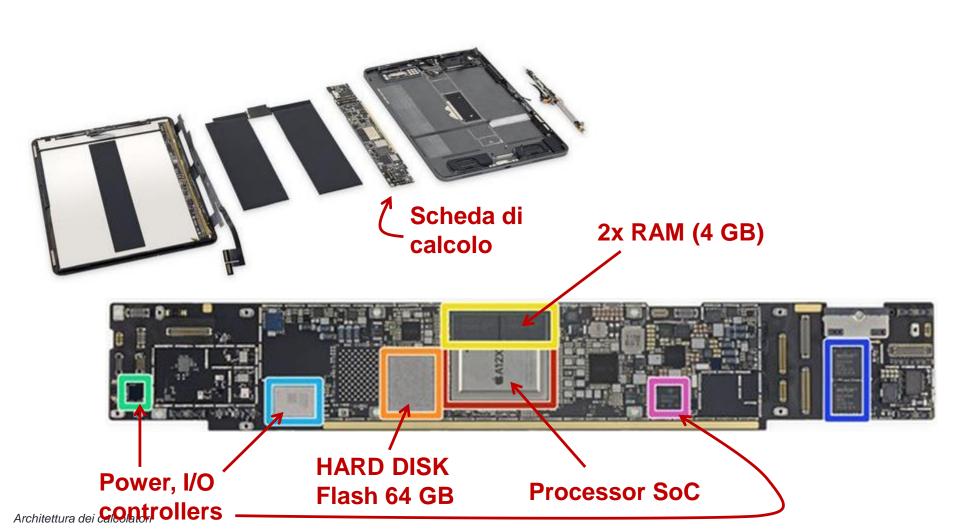
- Un dispositivo dell'era PostPC
- Touchscreen
 - rimpiazza keyboard e mouse
 - di tipo Capacitivo
 - consente multipli punti di "tocco" simultaneamente



iPad Pro 11", 2018



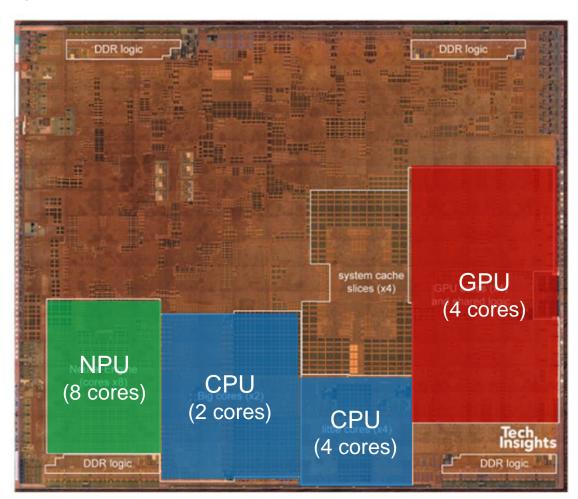
barometro e giroscopio a tre assi Wi-Fi + Bluetooth 5.0



E cosa c'è dentro al **Processor SoC?**

- Apple A12 Bionic system-on-chip
 - 83,27 mm²
 - (8,42 mm x 9,89 mm)
 - Processo a 7nm
 - 6,9 miliardi di transistor
 - Hexa-Core (6) CPU
 - 2 (Vortex) @2,49GHZ
 - 4 (Tempest) @1,49GHz
 - 4-Core GPU
 - 8-Core NPU





Architettura dei Calcolatori

- Quindi in questo corso si studia l'anatomia dei calcolatori?
 - Sì, ma non solo...

L'astrazione nei calcolatori









(C/C++, JAVA, OpenCV, CUDA...)

linguaggio macchina



IL COMPUTER

LE DISCIPLINE

Programmazione

- Compilatori
- sistemi operativi
- Architettura dei calcolatori

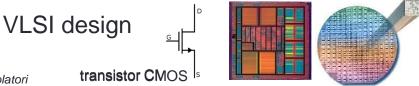
compiler

L'astrazione nei calcolatori





IL COMPUTER

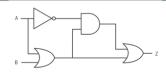


 Livello della progettazione su silicio (CMOS)





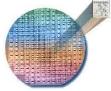
Logic circuits



VLSI design

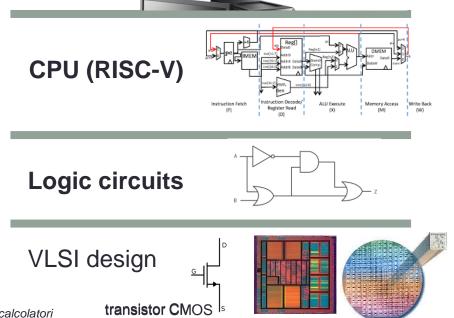




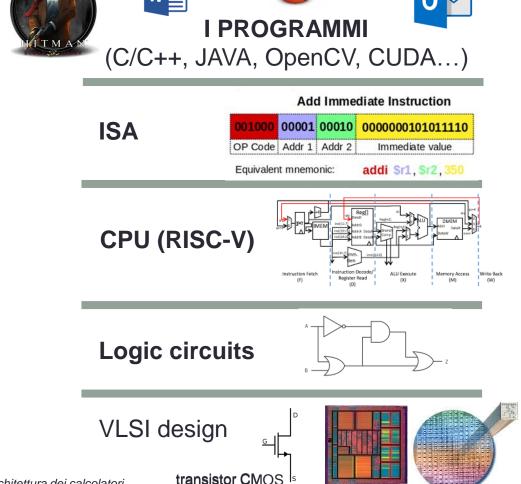


- Livello della progettazione con logica binaria
- Livello della progettazione su silicio (CMOS)



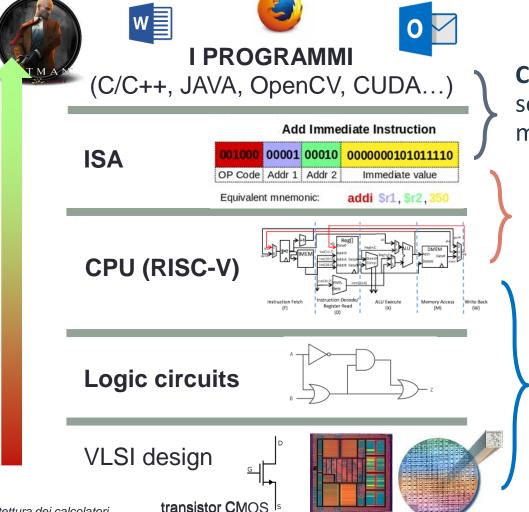


- Livello della progettazione architetturale
- Livello della progettazione con logica binaria
- Livello della progettazione su silicio (CMOS)



- Livello della progettazione di interfaccia col software
- Livello della progettazione architetturale
- Livello della progettazione con logica binaria
- Livello della progettazione su silicio (CMOS)

Dai transistor ai programmi



Compilatori: Come i programmi sono tradotti in linguaggio macchina

L'interfaccia hardware/software

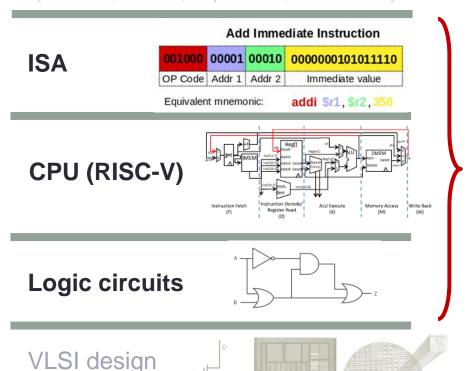
- L'Instruction Set Architecture (ISA)
 - Come l'hardware esegue il programma
- Ciò che determina la performance del programma e del sistema
- Come I progettisti hardware migliorano la performance

Dai transistor ai programmi

O, limitandoci al programma del corso, dalle reti logiche all'ISA

I PROGRAMMI

(C/C++, JAVA, OpenCV, CUDA...)



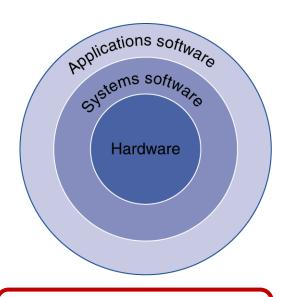
transistor CMOS

Il contenuto principale del corso di Architettura dei Calcolatori

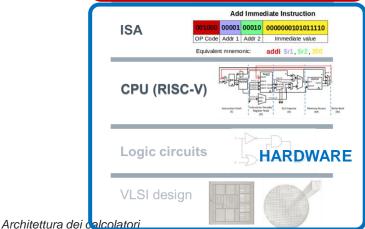
Ma prima di immergerci nel contenuto del corso...

Un po' di contesto

La dicotomia Hardware/Software







Il vostro programma

- Application software
 - Written in high-level language
- System software
 - **Compiler**: translates HLL code to binary (machine) code
 - Operating System: service code
 - Handling input/output
 - Managing memory and storage
 - Scheduling tasks & sharing resources

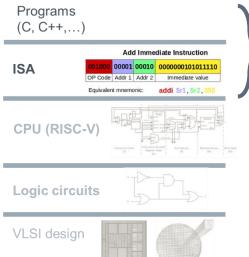
Hardware

Processor, memory, I/O controllers

Il codice del programma

High-level language program (in C)

- Linguaggio di alto livello (C, C++, Java)
 - Livello di astrazione prossimo al dominio del problema
 - Facilita la produttività e la portabilità
- Linguaggio assembly
 - Una rappresentazione delle istruzioni macchina comprensibile agli umani
- Il binario (o eseguibile)
 - Istruzione e dati codificati come stringhe di bit



Assembly language program (for RISC-V)

> ialr x0, 0(x1)Assembler

Binary machine language program (for RISC-V)

language of the CPU

0000000001101011001001100010011 0000000011001010000001100110011 000000000000011001100101000011 0000000011100110011000000100011 0000000010100110011010000100011

swap(int v[], int k)

v[k] = v[k+1]: $v\lceil k+1 \rceil = temp$:

slli x6. x11. 3

x6, x10, x6

x5, 0(x6)

x7, 8(x6)

x7.0(x6)x5.8(x6)

temp = v[k]:

{int temp;

Compiler

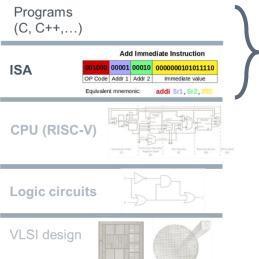
swap:

ISA

001000	00001	00010	0000000101011110
OP Code	Addr 1	Addr 2	Immediate value

L'Instruction Set Architecture (ISA)

- Una instruction set architecture (ISA) è un modello astratto di un computer
- Una realizzazione di una ISA, ovvero una CPU, è detta una implementazione di quell'ISA.
- Una ISA definisce i tipi di dato supportati, i registri, il supporto hardware per interagire con la memoria principale e l'I/O.



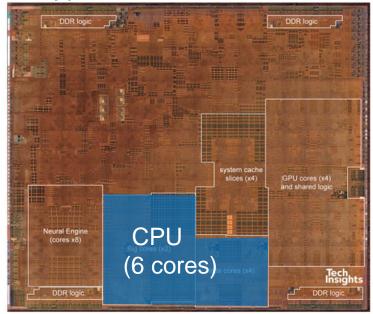
L'interfaccia HW/SW

ISA

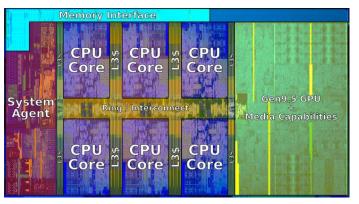
001000	00001	00010	0000000101011110
OP Code	Addr 1	Addr 2	Immediate value

- Vi ricordate i Processor SoC del PC e del Tablet?
 - Ognuno con tante CPU

Apple A12 Bionic SoC



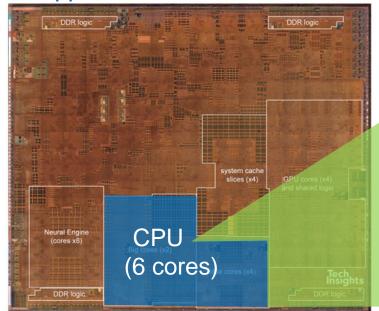
Intel i7 Coffee Lake SoC



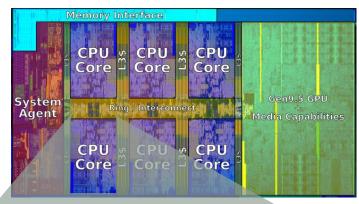
45

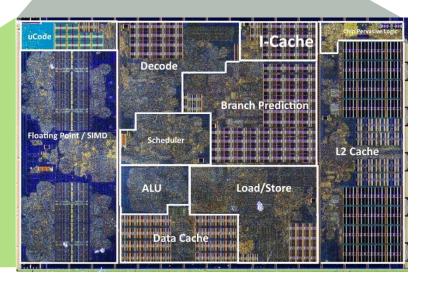
- Vi ricordate i Processor SoC del PC e del Tablet?
 - Ognuno con tante CPU

Apple A12 Bionic SoC

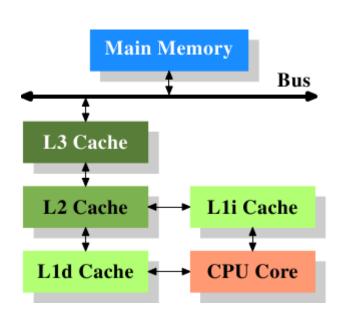


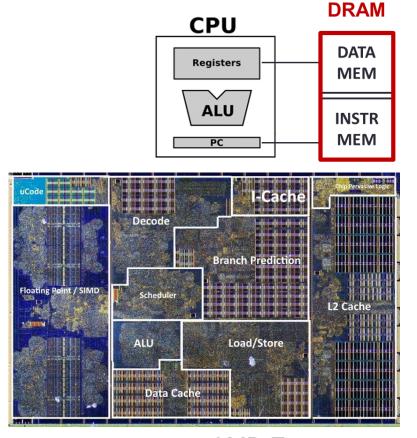
Intel i7 Coffee Lake SoC



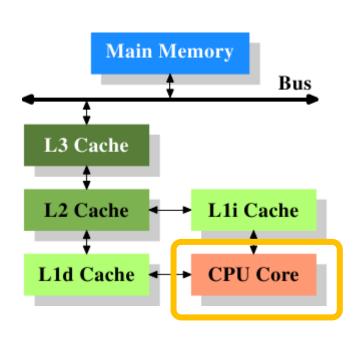


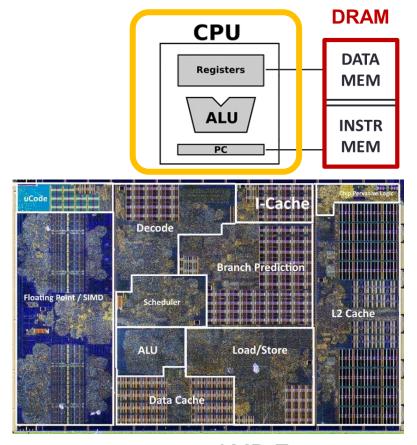
 E vi ricordate la gerarchia di memoria?

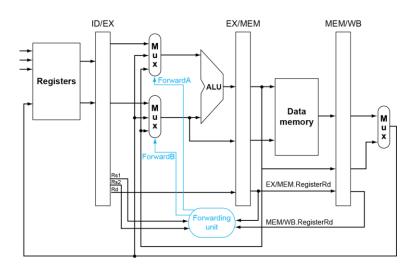




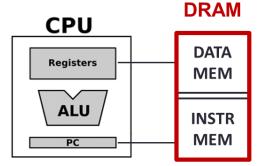
 E vi ricordate la gerarchia di memoria?

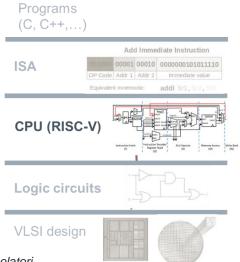


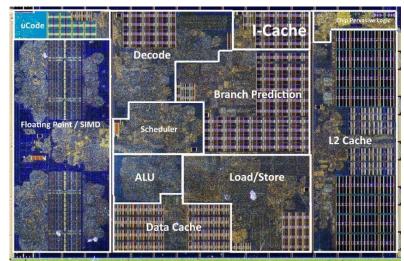




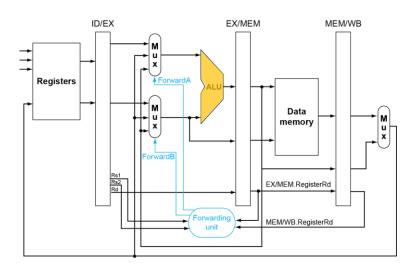
- ALU
- Registers
- Data mem



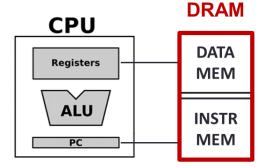


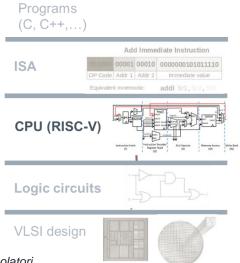


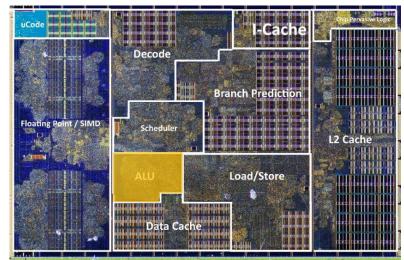
AMD Zen2 core



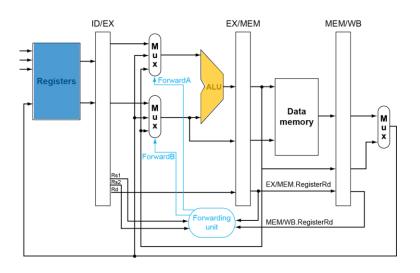
- ALU
- Registers
- Data mem



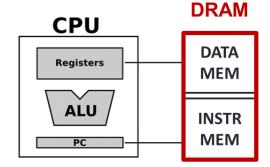


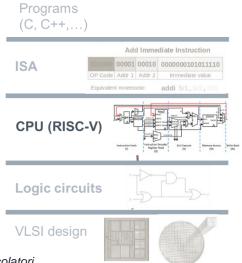


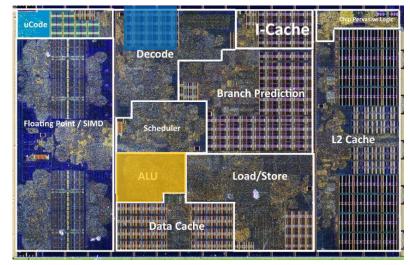
AMD Zen2 core



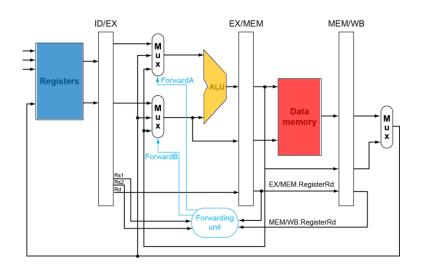
- ALU
- Registers
- Data mem



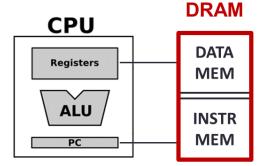


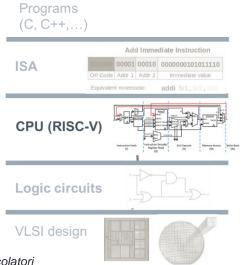


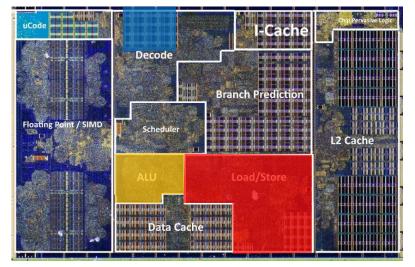
AMD Zen2 core



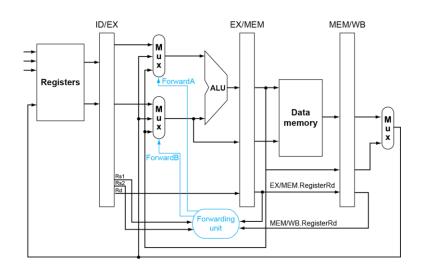
- ALU
- Registers
- Data mem





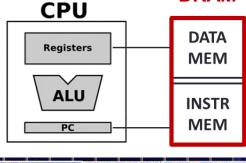


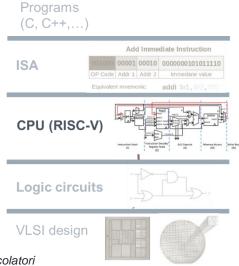
AMD Zen2 core

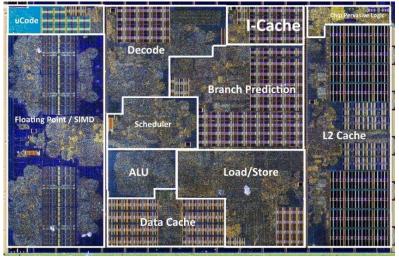


CPU PIPELINE

 Costituita da blocchi progettati come reti logiche



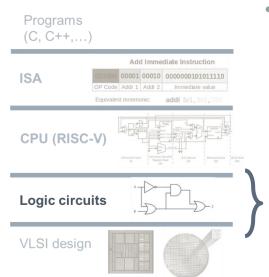




AMD Zen2 core

Reti logiche

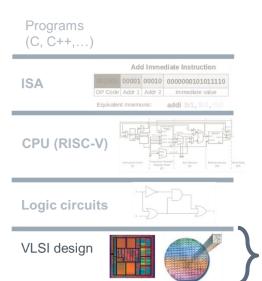
- Le Reti Logiche forniscono un'astrazione funzionale di come funziona la tecnologia sottostante (vedi sotto), e sono composte dall'interconnessione di numerose porte logiche
- Porta logica = dispositivo ideale che realizza una funzione booleana
 - Un'operazione logica eseguita su uno o più input binari che produce un singolo output binario.



- Le porte logiche sono implementate con transistor che fungono da interruttori elettronici
 - ON/OFF → TRUE/FALSE

Tecnologia CMOS per le reti logiche

- Le Reti Logiche forniscono un'astrazione funzionale di come funziona la tecnologia sottostante (vedi sotto), e sono composte dall'interconnessione di numerose porte logiche
- Porta logica = dispositivo ideale che realizza una funzione booleana
 - Un'operazione logica eseguita su uno o più input binari che produce un singolo output binario.

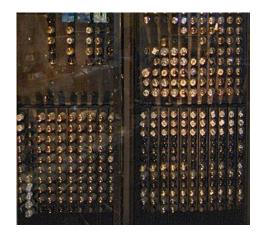


- Le porte logiche sono implementate con transistor che fungono da interruttori elettronici
 - ON/OFF → TRUE/FALSE
 - La maggior parte dei chip esistenti utilizza logica
 CMOS¹ per realizzare i transistor.
 - Ma si possono realizzare con valvole termoioniche, nanotubi di carbonio, logica fluidica, pneumatica, ottica, molecolare...

- Electronics technology continues to evolve
 - Increased capacity and performance
 - Reduced cost



The first triode, the De Forest Audion, invented in 1906



The 1946 ENIAC computer used 17,468 vacuum tubes and consumed 150 kW of power

Year	Technology	Relative performance/cost
1951	Vacuum tube	1

- Electronics technology continues to evolve
 - Increased capacity and performance
 - Reduced cost



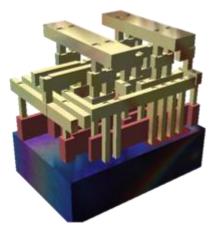
Assorted discrete transistors



The 1955 CADET computer used 324 point-contact transistors

Year	Technology	Relative performance/cost
1951	Vacuum tube	1
1965	Transistor	35

- Electronics technology continues to evolve
 - Increased capacity and performance
 - Reduced cost



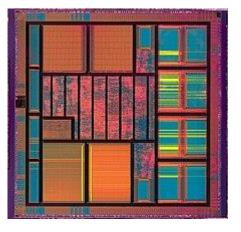
Electronic circuits are gradually created on a wafer made of pure semiconducting material. Silicon is almost always used



Robert Noyce invented the first monolithic IC chip in 1959. It was made from silicon and was fabricated using Hoerni's planar process and Atalla's surface passivation process.

Year	Technology	Relative performance/cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit (IC)	900

- Electronics technology continues to evolve
 - Increased capacity and performance
 - Reduced cost



A VLSI integrated-circuit die

VLSI is the process of creating an integrated circuit (IC) by combining millions of MOS transistors onto a single chip

Year	Technology	Relative performance/cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit (IC)	900
1995	Very large scale IC (VLSI)	2,400,000

- Electronics technology continues to evolve
 - Increased capacity and performance
 - Reduced cost

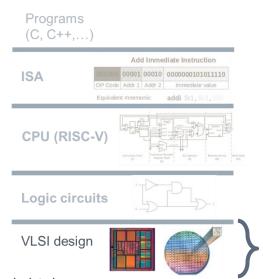
Name	Signification	Year	Transistors	Logic gates
SSI	small-scale integration	1964	1 to 10	1 to 12
MSI	medium-scale integration	1968	10 to 500	13 to 99
LSI	large-scale integration	1971	500 to 20.000	100 to 9999
VLSI	very large-scale integration	1980	20.000 to 1.000.000	10.000 to 99.999
ULSI	ultra-large-scale integration	1984	> 1.000.000	>= 100.000

Year	Technology	Relative performance/cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit (IC)	900
1995	Very large scale IC (VLSI)	2,400,000
2013	Ultra large scale IC	250,000,000,000

Semiconductor Technology

Silicon: semiconductor

- A semiconductor material has an electrical conductivity value falling between that of a conductor, such as metallic copper, and an insulator, such as glass.
- Its conducting properties may be altered in useful ways by introducing impurities ("doping") into the crystal structure



Atom size is 20-200 pm

Semiconductor Manufacturing Process

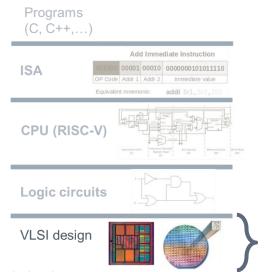
```
-1971
10 µm
6 µm
       -1974
3 µm
       -1977
1.5 \, \mu m - 1982
1 um
       -1985
800 nm - 1989
600 nm - 1994
350 nm - 1995
250 nm - 1997
180 nm - 1999
130 nm – 2001
90 nm - 2004
65 nm
      -2006
45 nm
       -2008
32 nm
       -2010
       -2012
22 nm
14 nm
       -2014
       -2017
10 nm
       -~2019
7 nm
```

- ~2021

5 nm

Semiconductor Technology

- Silicon: semiconductor
- Add materials to transform properties:
 - Conductors
 - microscopic copper or aluminum wire
 - Insulators
 - plastic sheathing or glass



Atom size is 20-200 pm

Semiconductor Manufacturing Process

```
10 µm
       -1971
6 µm
       -1974
3 µm
       -1977
1.5 \, \mu m - 1982
1 \mu m
       -1985
800 nm - 1989
600 nm - 1994
350 nm - 1995
250 nm - 1997
180 nm - 1999
130 nm - 2001
90 nm - 2004
65 nm
       -2006
45 nm
       -2008
32 nm
       -2010
       -2012
22 nm
14 nm
       -2014
       -2017
10 nm
       -~2019
7 nm
```

5 nm

-~2021

Manifacturing ICs



Manifacturing ICs





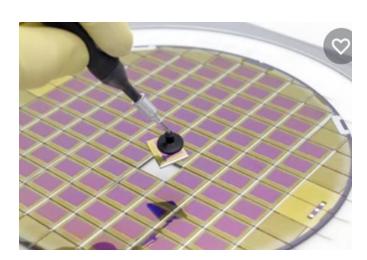
From a silicon ingot (99.999999999 pure)...

Manifacturing ICs





From a silicon ingot (99.999999999 pure)...



...to a silicon wafer with lcs...

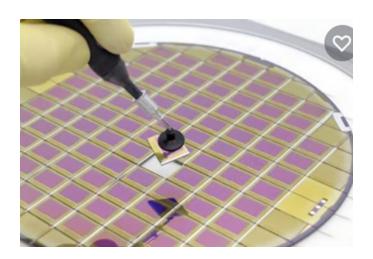
Manifacturing ICs













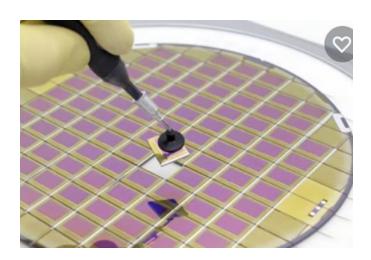
...to a silicon wafer with lcs...

Manifacturing ICs



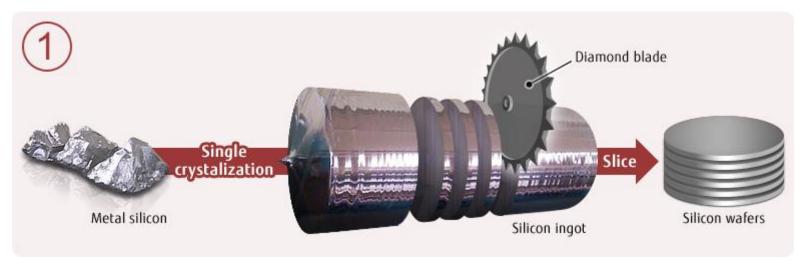
(intel)
Core™ i7

From a silicon ingot (99.999999999 pure)...

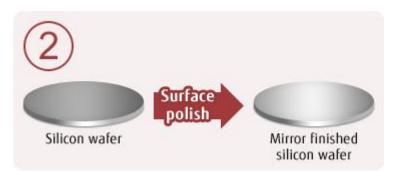


HOW DO WE GET THERE?

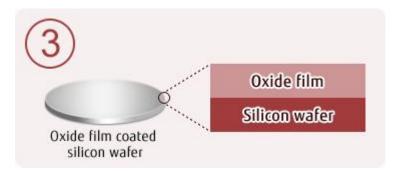
...to a silicon wafer with Ics...



The silicon ingot is sliced into thin silicon wafers.



The surface of the silicon wafer is polished to create a mirror finish.

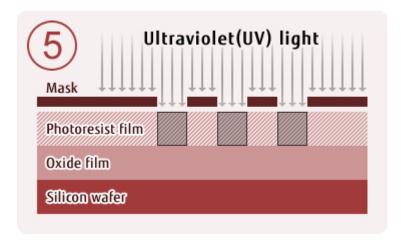


An oxide film is grown onto the wafer.

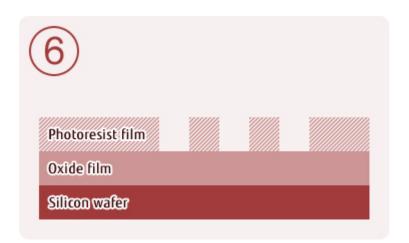
https://www.fujitsu.com/ph/about/businesspolicy/tech/k/whatis/processor/cpu.html



A photoresist film is coated on the wafer surface.

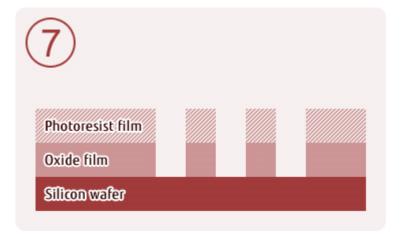


The photoresist film is exposed to ultraviolet (UV) light through the pattern on the mask.

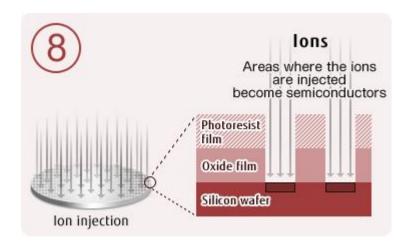


The photoresist region that was exposed to the UV light is removed using a developer. (The exposed area changes to a substance that is dissolved by the developer.)

The oxide film is then removed using a caustic agent. This exposes the silicon surface.



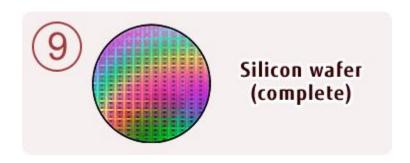
https://www.fujitsu.com/ph/about/businesspolicy/tech/k/whatis/processor/cpu.html



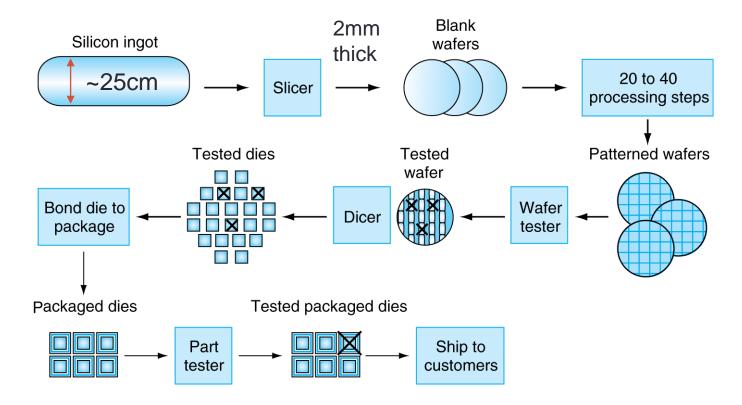
The required ions are then injected into the silicon surface. The character of the silicon then changes into a semiconductor, which is a state where elements with electrical characteristics can be created.

Repeat steps 4 to 8 many times...



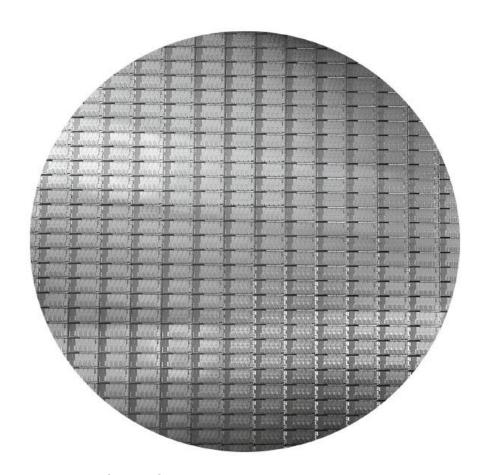


https://www.fujitsu.com/ph/about/businesspolicy/tech/k/whatis/processor/cpu.html



- One layer of transistors and 2-8 levels of metal conductor, separated by layers of insulators
- Yield: proportion of working dies per wafer

Intel Core i7 Wafer



- 300mm wafer, 280 chips, 32nm technology
- Each chip is 20.7 x 10.5 mm

Integrated Circuit Cost

Cost per die =
$$\frac{\text{Cost per wafer}}{\text{Dies per wafer} \times \text{Yield}}$$

Dies per wafer ≈ Wafer area/Die area

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

- The cost of an integrated circuit rises quickly as the die size increases
 - lower yield and fewer dies that fit on a wafer.
 - nonlinear relation to area and defect rate
 - Wafer cost and area are fixed
 - Defect rate determined by manufacturing process
 - Die area determined by architecture and circuit design