



**Lecture**  
**9\_5.1**

# ***The Memory Hierarchy***



**Test**  
*Group*

**Paolo PRINETTO**

**Politecnico di Torino (Italy)**

**University of Illinois at Chicago, IL (USA)**

[Paolo.Prinetto@polito.it](mailto:Paolo.Prinetto@polito.it)

[prinetto@uic.edu](mailto:prinetto@uic.edu)

[www.testgroup.polito.it](http://www.testgroup.polito.it)

[www.comitato-girotondo.org](http://www.comitato-girotondo.org)

# *License Information*

This work is licensed under the  
**Creative Commons BY-NC**  
License



To view a copy of the license, visit:  
<http://creativecommons.org/licenses/by-nc/3.0/legalcode>

# ***Disclaimer***

- **We disclaim any warranties or representations as to the accuracy or completeness of this material.**
- **Materials are provided “as is” without warranty of any kind, either express or implied, including without limitation, warranties of merchantability, fitness for a particular purpose, and non-infringement.**
- **Under no circumstances shall we be liable for any loss, damage, liability or expense incurred or suffered which is claimed to have resulted from use of this material.**

## **Goal**

- This lecture presents the basic concepts of *memory hierarchy* in a typical computer system

# *Prerequisites*

- **Module 0\_4 : Digital Systems: Definitions and Taxonomies**

# *Homework*

- **None**

## ***Further readings***

- **Students interested in making a reference to a text book on the arguments covered in this lecture can refer, for instance, to:**
  - ***G. Conte, A. Mazzeo, N. Mazzocca, P. Prinetto: “Architettura dei calcolatori”, Città Studi, 2015 (Chapter 7 - Sistema delle memorie)***



# Outline

- Introduction
- The concept of memory hierarchy

# Outline

- Introduction
- The concept of memory hierarchy

# Introduzione

- Ogni sistema di elaborazione contiene dispositivi per la memorizzazione di dati e istruzioni.
- L'insieme di tali dispositivi, e degli algoritmi per la loro gestione, costituisce il **sistema di memoria**.

# ***Introduzione***

- **Un sistema di calcolo ideale dovrebbe essere dotato di un organo di memoria caratterizzato da**
  - **uno spazio di indirizzi grande a piacere**
  - **tempi di accesso piccoli a piacere**
  - **costo basso e indipendente dalla tecnologia impiegata.**

# **Introduzione**

- **I dispositivi di memoria reali utilizzati nei sistemi di calcolo non godendo di tali proprietà sono caratterizzati da**
  - **soluzioni legate a specifiche tecnologie e associati costi**
  - **classi di memorie caratterizzate principalmente dalla velocità di accesso e dalla capacità.**
- **Si deve quindi cercare un compromesso tra il costo del sistema di memoria e le sue prestazioni.**

# ***Strategia Generale***

- **Nella progettazione di un sistema di memoria vengono tenuti in conto i seguenti punti:**

# Strategia Generale

1. conviene che il sistema complessivo sia composto da memorie di tipo e costo diversi, rispondenti ai diversi usi che della memoria vengono fatti (*gerarchia di memoria*)
2. la gestione della memoria deve essere il più possibile trasparente per il programmatore e l'utente (*memoria virtuale*)
3. se il sistema è di tipo multiprocessore, ogni processore deve poter lavorare con la memoria al massimo della velocità e senza interferire con il lavoro degli altri.

# ***Memoria: Aumento delle Prestazioni***

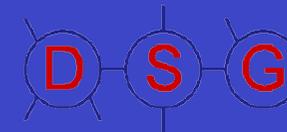
- A causa del ruolo critico che la memoria assume quando si cerca un incremento nel throughput del sistema essa viene detta: ***The Von Neumann Bottleneck***

# Outline

- Introduction
- **The concept of memory hierarchy**

# ***Livelli gerarchici di memorie***

- **La memoria di un calcolatore è normalmente organizzata in livelli gerarchici**
- **Ogni livello è caratterizzato da un diverso tipo di memoria (in termini di velocità, costo e dimensione)**



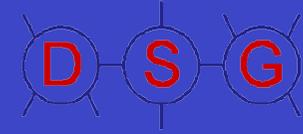
# LA GERARCHIA

Coppia di strutture di memoria  $M_1$  e  $M_2$  con:

- ⇒ costo per bit:  $c_1 > c_2$
- ⇒ tempi di accesso:  $t_{A1} < t_{A2}$
- ⇒ dimensioni:  $S_1 < S_2$

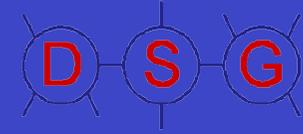


$M_1$  e  $M_2$  realizzati con tecnologie diverse.



## CRITERI DI GESTIONE

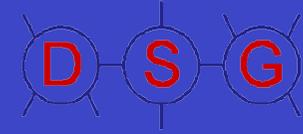
- ➔ I dati utilizzati più spesso vanno posti in memorie facilmente accessibili
- ➔ I dati sono allocati dinamicamente per utilizzare gli spazi disponibili con la massima efficienza
- ➔ Spostamento automatico dei dati tra i livelli
- ➔ Canali di comunicazione veloci fra i livelli



## CRITERI DI GESTIONE

La politica di gestione tende a mimare una memoria che abbia:

- ⇒ i tempi di accesso della più veloce
- ⇒ le dimensioni della maggiore
- ⇒ i costi della più economica

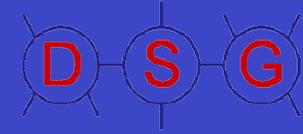


## HIT E MISS RATIO

Le prestazioni del sistema dipendono dal tasso di successo o **Hit ratio** =  $H$  definito come la probabilità che la richiesta sia soddisfatta al livello  $M_1$

Si definisce tasso di insuccesso o **Miss ratio** la probabilità che la richiesta non sia soddisfatta al livello  $M_1$

$$\text{Miss ratio} = 1 - H$$



# TEMPO ACCESSO

Tempo di accesso medio globale:

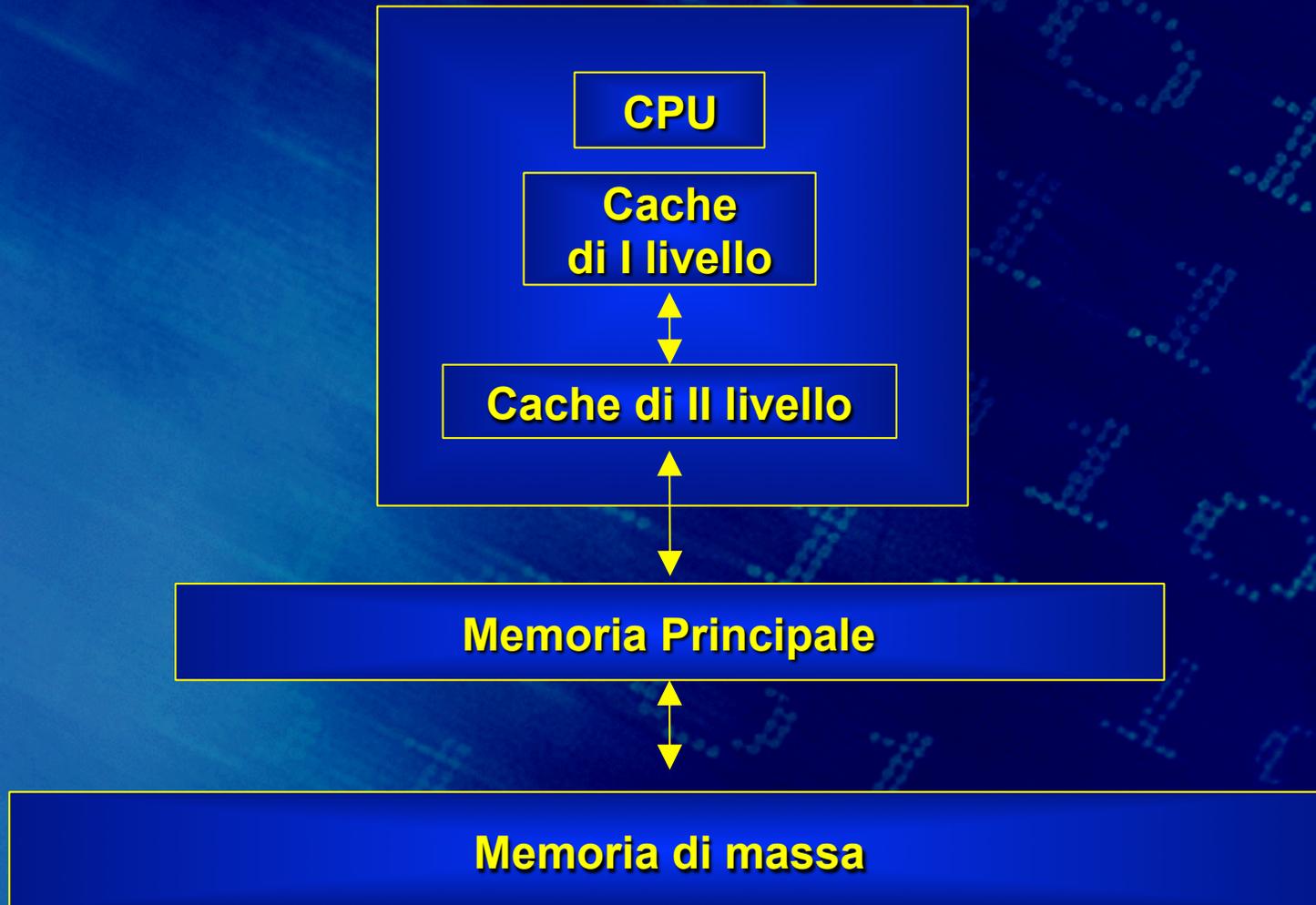
$$t_A = H t_{A1} + (1 - H) t_{A2}$$

dove:

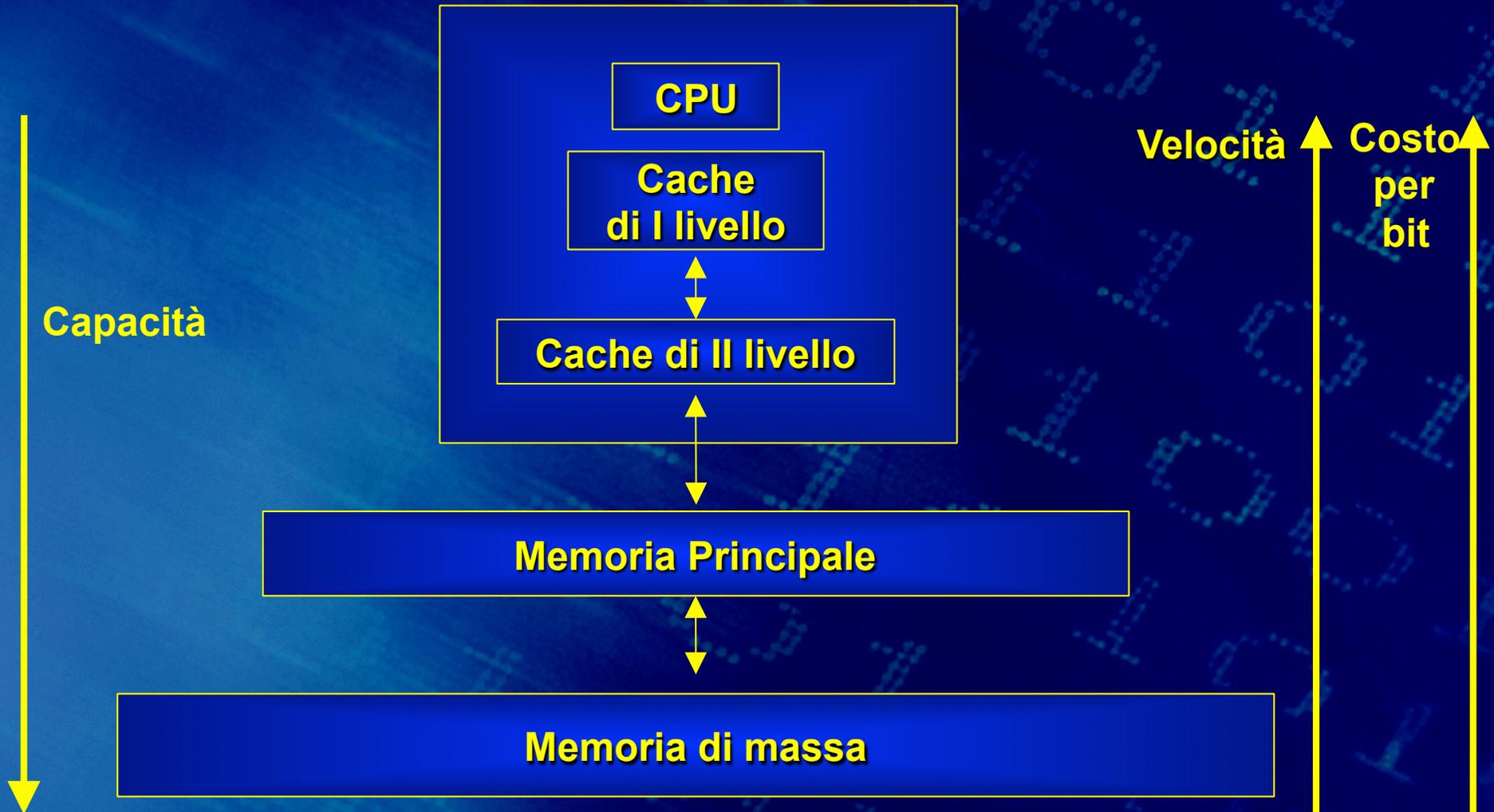
- $t_{A2} = t_{A1} + t_B = r t_{A1}$
- $t_B$  è il tempo di accesso a un blocco di  $M_2$



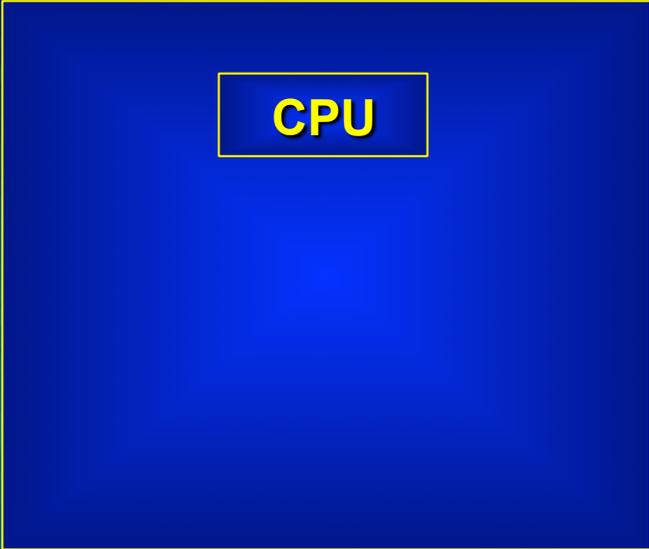
# Gerarchia di Memorie



# Gerarchia di Memorie



# Gerarchia di Memorie

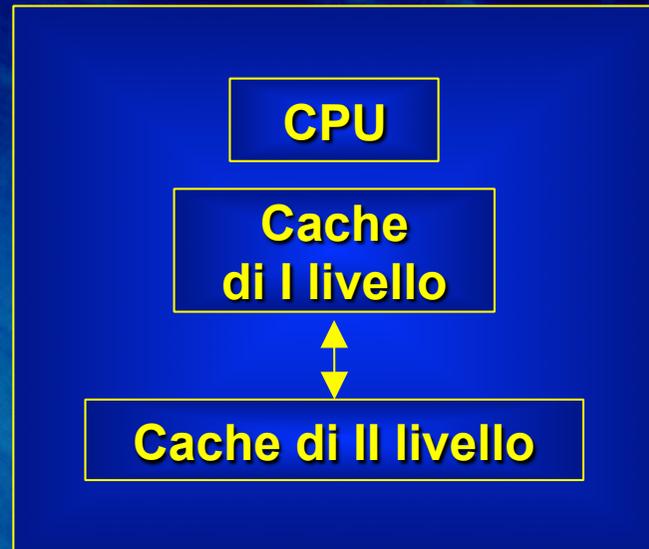


CPU

# ***Memoria interna alla CPU***

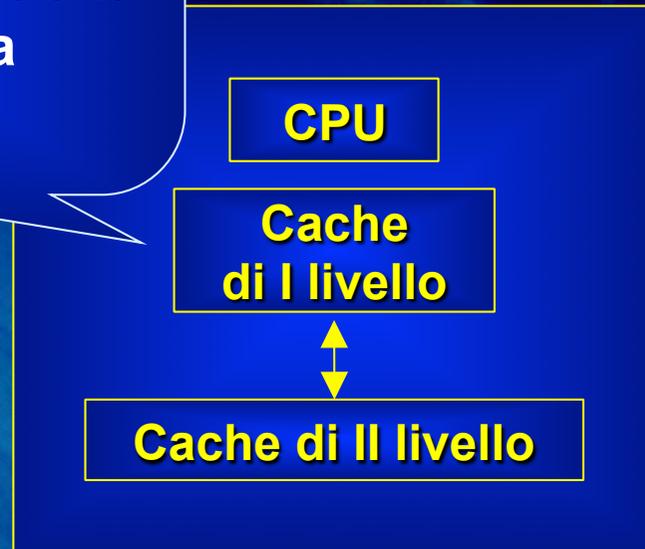
- **Corrisponde ai registri interni del processore.**
- **È caratterizzata da:**
  - **alta velocità (comparabile con quella del processore)**
  - **limitate dimensioni (al più qualche KB)**
- **È di solito realizzata tramite SRAM**

# Gerarchia di Memorie



# Gerarchia di Memorie

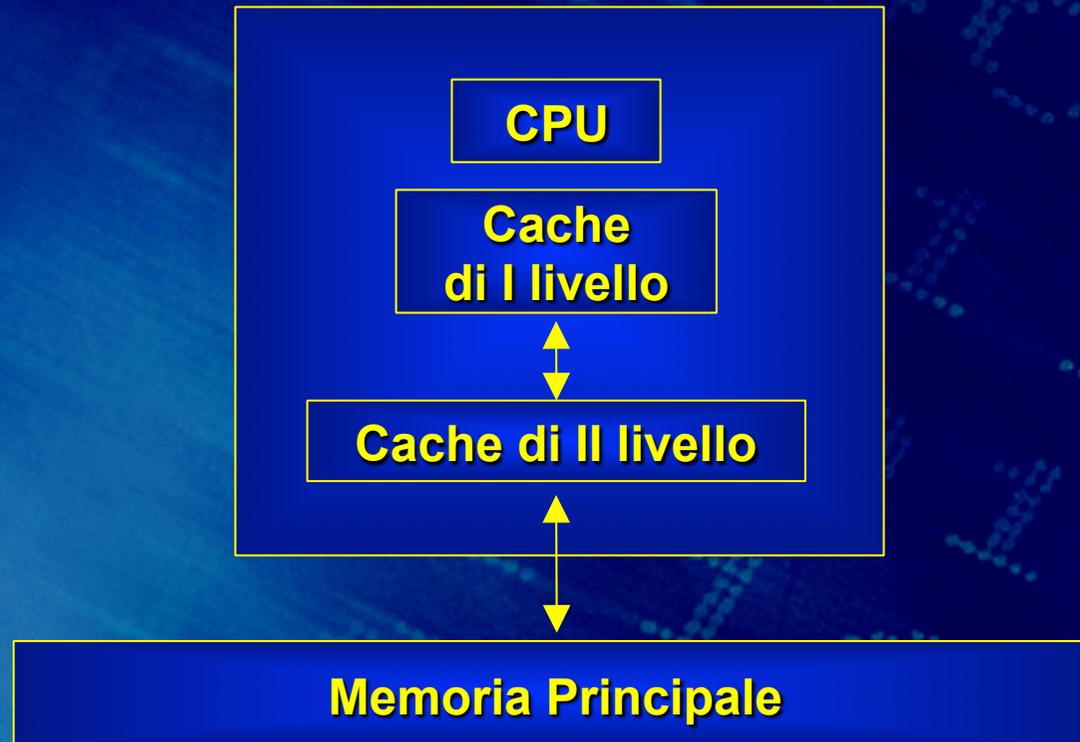
La Cache, costituita da memorie estremamente veloci, rappresenta il livello della gerarchia tra la CPU e la Memoria principale



# Cache

- **All'interno di una cache risiedono temporaneamente i dati e/o i programmi utilizzati in quel momento**
- **Il loro uso è trasparente al programmatore e al Sistema Operativo**
- **Le cache permettono di aumentare la velocità di accesso ai dati nella memoria principale senza ricorrere a memorie di tipo più costoso**

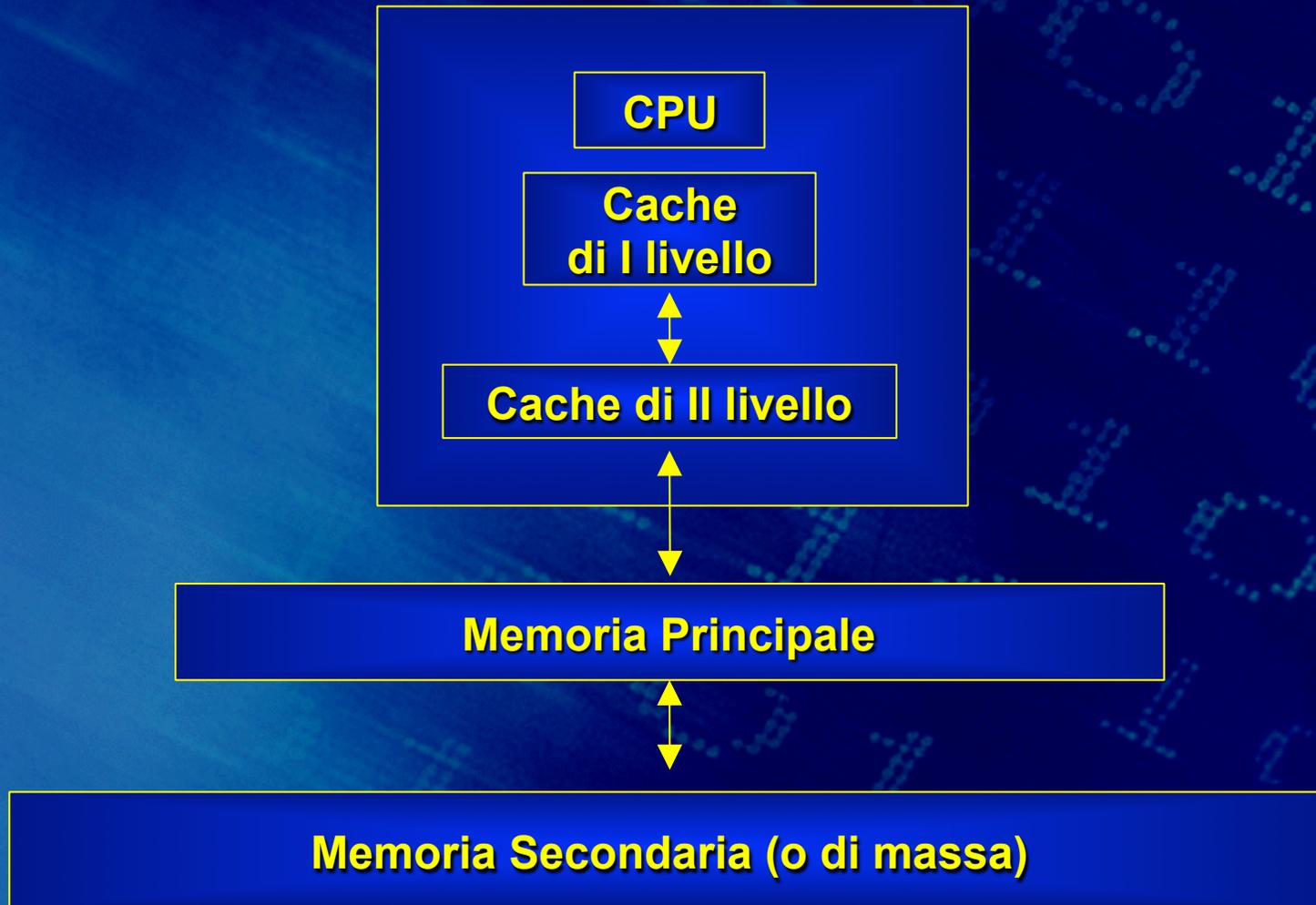
# Gerarchia di Memorie



# ***Memoria Principale***

- **Ha dimensioni molto maggiori (fino a qualche GB) ma tempi di accesso più elevati**
- **È accessibile in modo diretto tramite indirizzi**
- **È normalmente realizzata sotto forma di circuito integrato**
- **È di solito realizzata tramite DRAM**

# Gerarchia di Memorie



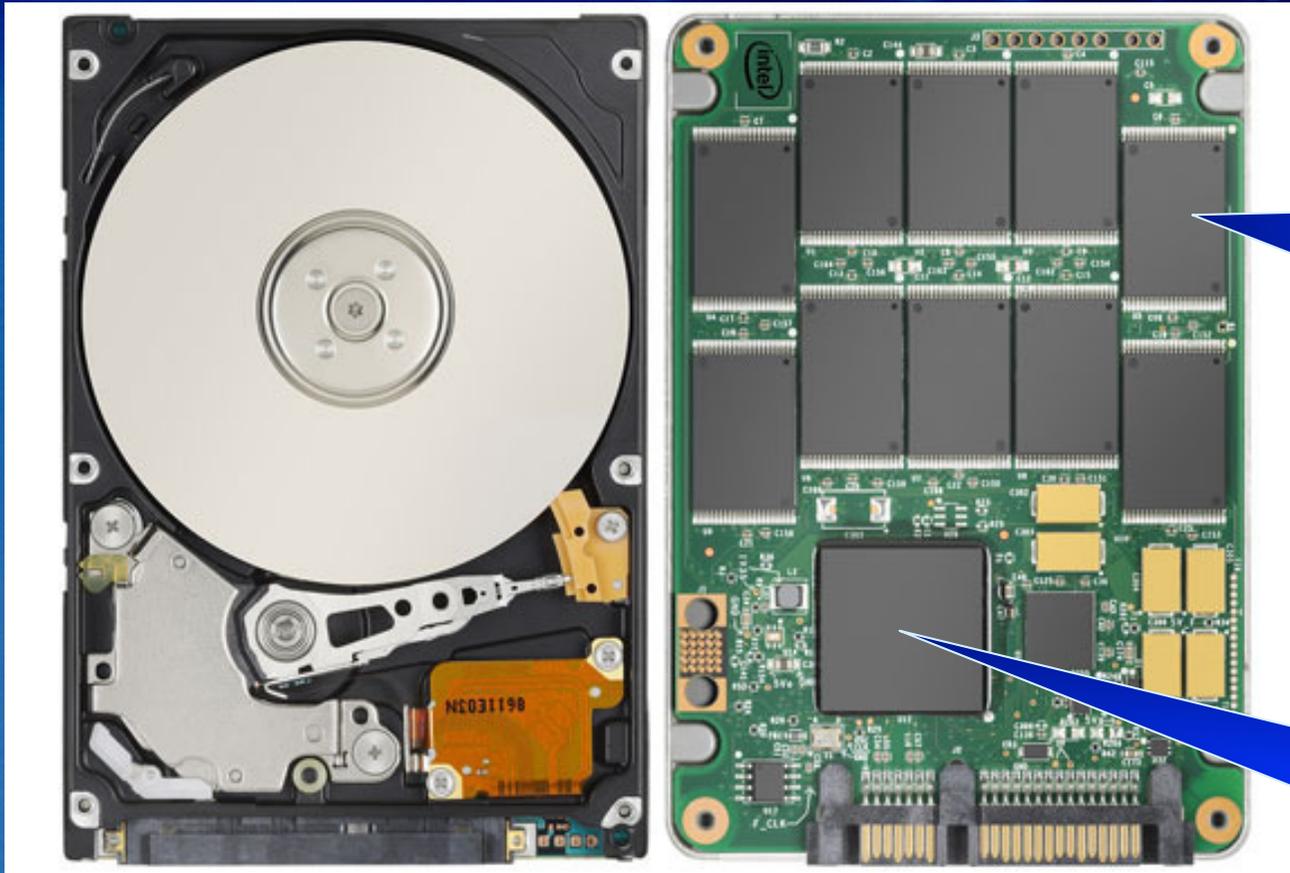
# **Memoria Secondaria**

- **Ha dimensioni e tempi di accesso ancora maggiori**
- **Viene usata per memorizzare dati e programmi non immediatamente richiesti dal processore**
- **L'accesso è gestito da appositi programmi di interfaccia**
- **È normalmente realizzata sotto forma di dischi magnetici o a stato solido**

# Solid State Drive (SSDs)

Yesterday

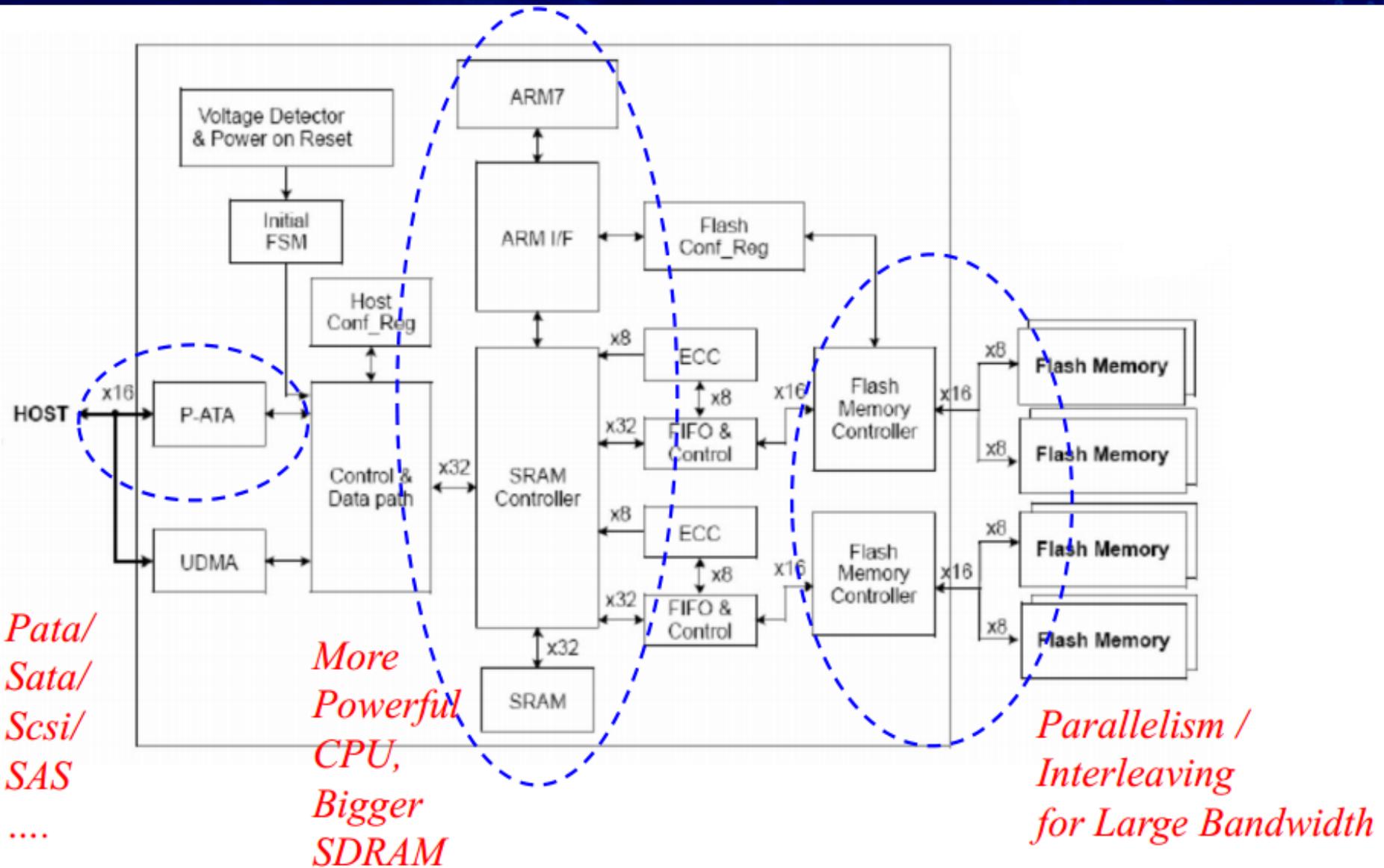
Today



**NAND Flash  
Memories**

**Memory  
Controller  
(ARM-based)**

# SSD = computing platform



Малые Автюхи, Калинковичский район, Республики Беларусь

