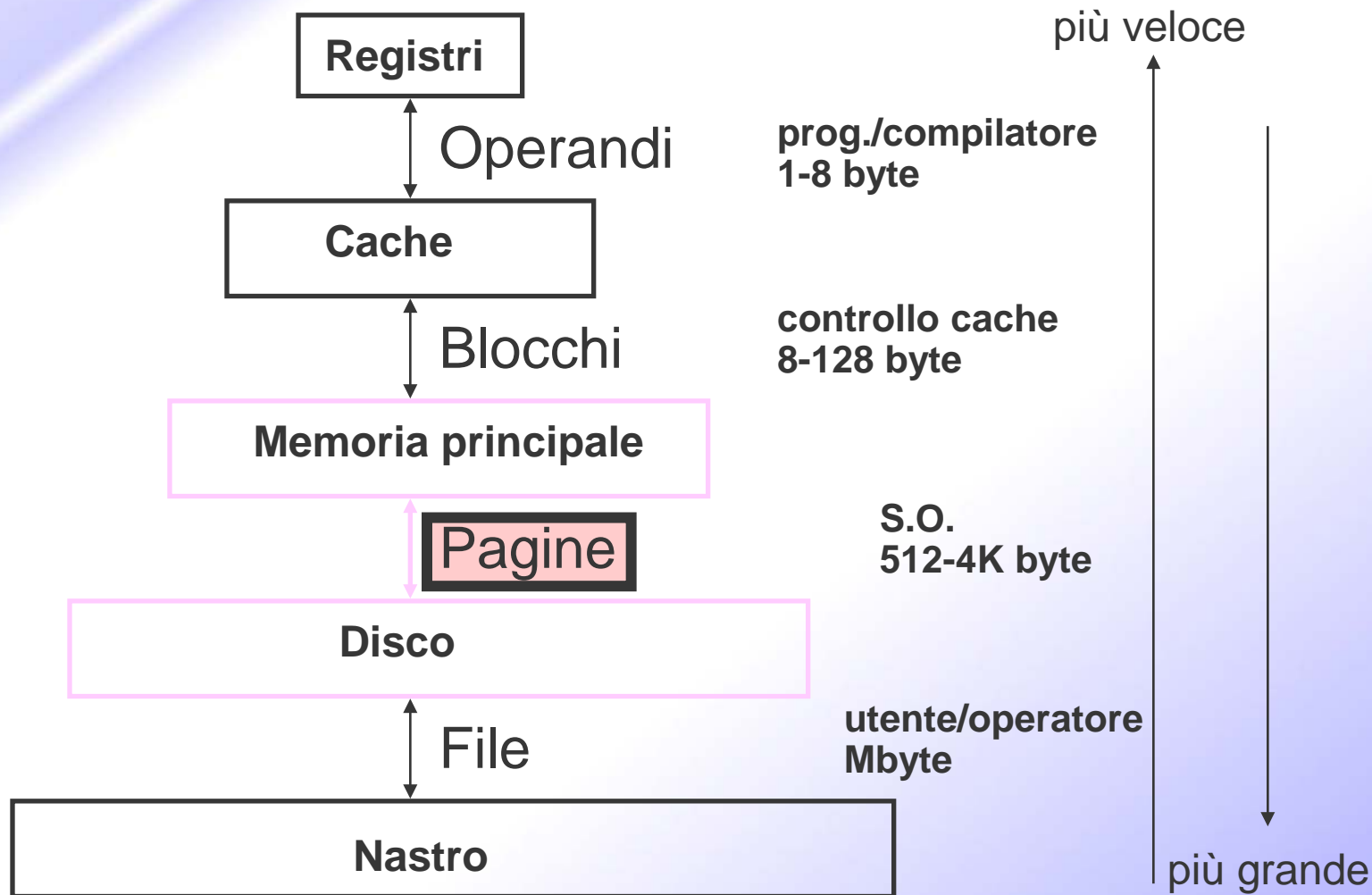


Memoria Virtuale

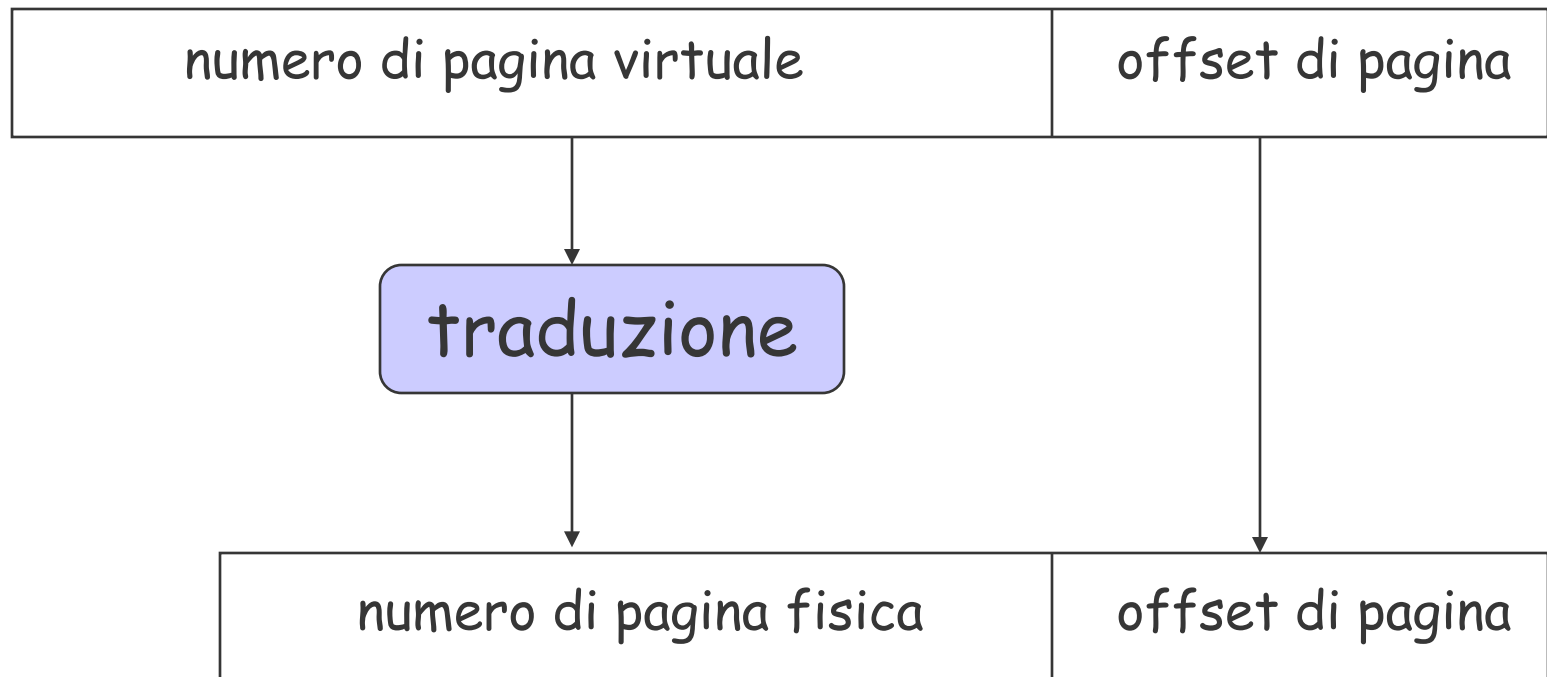
La gerarchia di memoria



Memoria virtuale

- La memoria principale può agire come una cache per la memoria secondaria (come il disco)
- Vantaggi:
 - illusione di avere più memoria fisica
 - riallocazione dei programmi
 - protezione

Trasformazione degli indirizzi

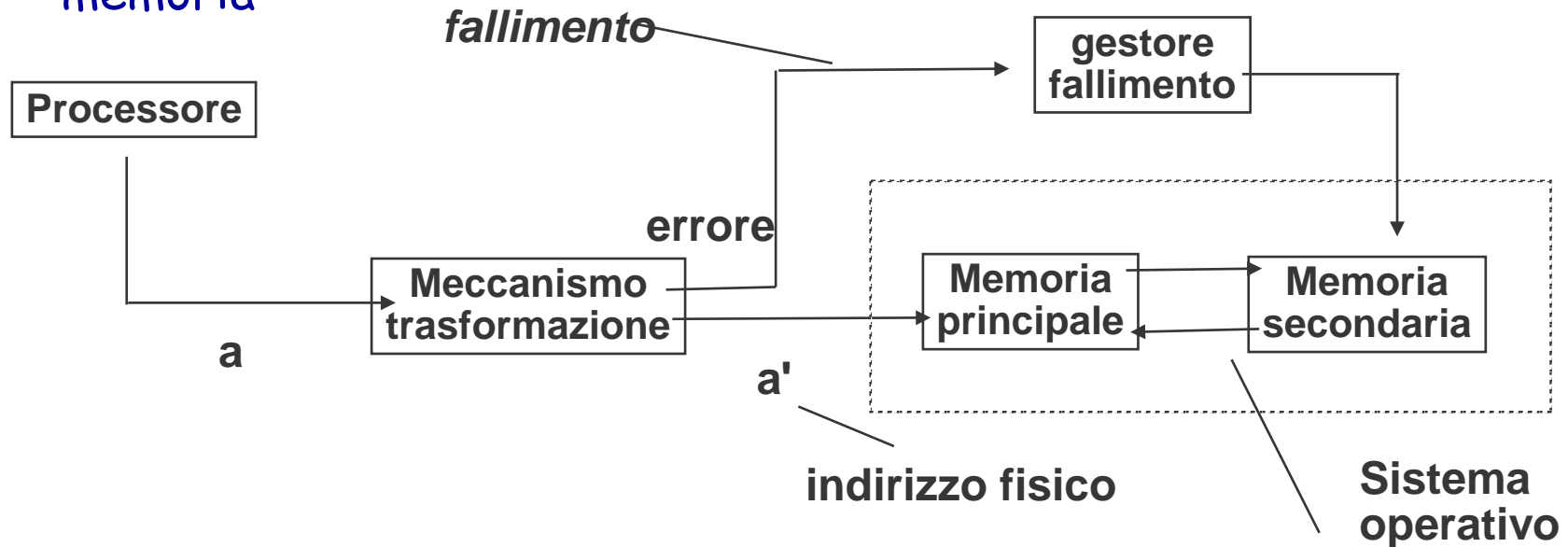


Problemi fondamentali

- Penalità di fallimento molto alta
 - accesso a memoria secondaria (milioni di cicli)
- Dimensione delle pagine trasferite da memoria secondaria a memoria principale (in breve, M)
 - pagine grandi abbastanza da ammortizzare la penalità di fallimento
 - 32KB e 64KB
- Ridurre il tasso di fallimenti
 - associatività completa
- Gestione dei fallimenti via software
 - strategie più intelligenti
- Uso della tecnica write-back
 - write-through troppo costosa

Trasformazione degli indirizzi

- $V = \{0, 1, \dots, n - 1\}$ spazio indirizzi virtuali
- $M = \{0, 1, \dots, m - 1\}$ spazio indirizzi fisici
 - $m \ll n$
- MAP: funzione di trasformazione degli indirizzi
 - $MAP(a) = a'$, se pagina con indirizzo virtuale a è presente in memoria all'indirizzo fisico a'
 - $MAP(a) = \text{errore}$, se pagina con indirizzo virtuale a non è presente in memoria



Pagine e quadri (frames)

- Illusione che la memoria fisica sia più grande di quanto lo è in realtà

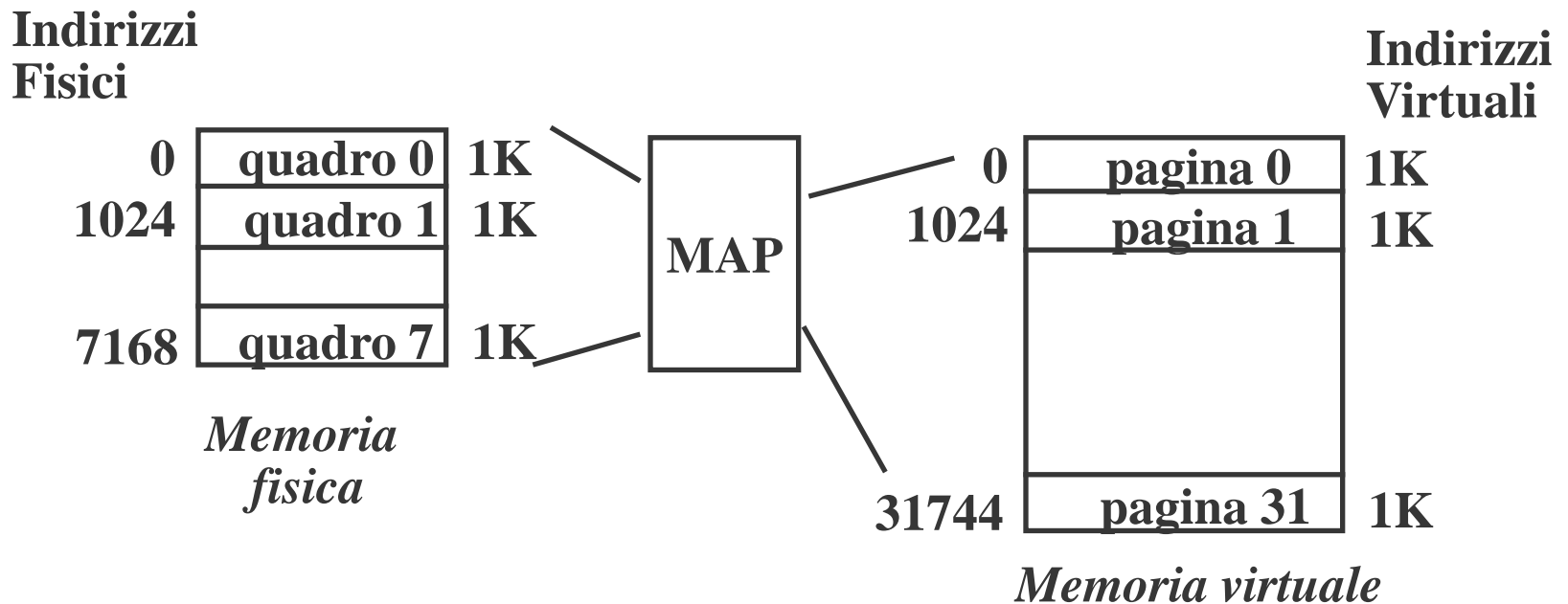
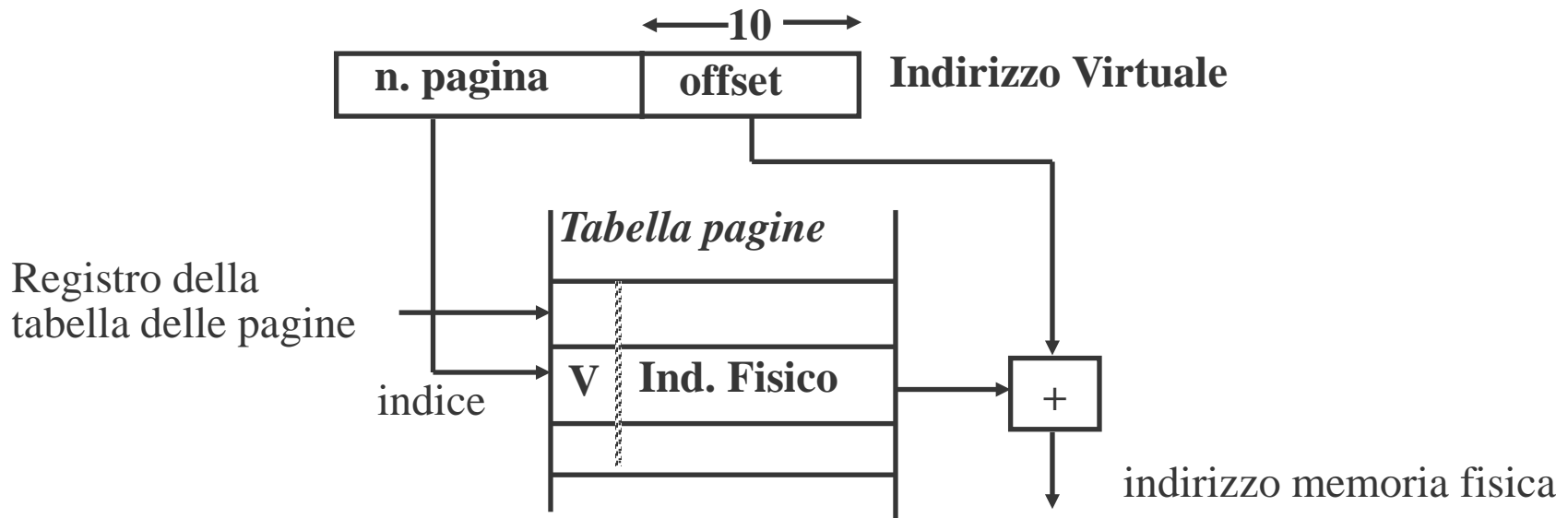


Tabella delle pagine

- Una tabella delle pagine per ogni programma
 - contiene una linea per ogni pagina virtuale
 - contenente il corrispondente indirizzo fisico
 - risiede in memoria principale
 - puntata dal registro della tabella delle pagine



Indirizzi fisici su disco

- Se la pagina virtuale non si trova in memoria fisica il controllo passa al sistema operativo che la carica da disco
- Per sapere dove si trova su disco, la tabella delle pagine contiene
 - indirizzo di memoria principale per le pagine virtuali presenti
 - **indirizzo di disco per le pagine virtuali assenti**
 - calcolati dal sistema operativo al momento in cui il programma viene avviato

Strategia di sostituzione

➤ LRU

- pagine utilizzata il meno recentemente

➤ LFU

- pagina utilizzata il meno frequentemente

➤ Approssimazione di LRU (detto 2° chance)

- bit di utilizzo associato ad ogni pagina
 - azzerato periodicamente
 - asserito quando una pagina viene utilizzata
- sceglie pagina con bit di utilizzo non asserito

Dimensione della tabella

➤ Esempio

- indirizzo virtuale a 32 bit
- pagine di 4KB: offset di 12 bit
- 4 byte per ogni elemento della tabella
- numero di elementi: $2^{32}/2^{12} = 2^{20}$
- dimensione tabella: $2^{20} \times 4 \text{ byte} = 4\text{MB}$ (per programma)

➤ Se molti programmi sono attivi non è gestibile

- soluzioni:
 - aumentare la dimensione della tabella quando serve
 - uso di funzioni hash
 - tabelle delle pagine a più livelli
 - paginazione delle tabelle delle pagine

- Per velocizzare la trasformazione si usa una **cache speciale** i cui elementi indicano le pagine usate recentemente
 - **Translation Lookaside Buffer (TLB)**
 - Tempo di accesso alla TLB confrontabile a quello della cache (molto inferiore all'accesso alla memoria principale)
 - La TLB può essere completamente associativa, N-associativa, oppure a mappa diretta
 - Le TLB sono tipicamente piccole: ciò permette un'organizzazione completamente associativa.
 - La maggior parte delle macchine usano una strategia di sostituzione casuale.

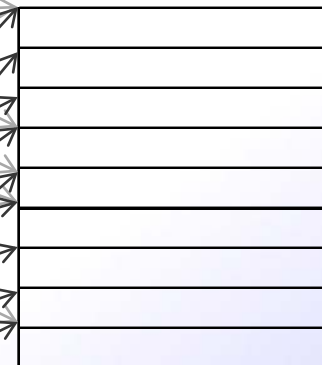
TLB, cache e memoria principale

Numero di
pagina
virtuale

TLB

Val.	Mod.	Rif.	Etichetta	Indirizzo di pagina fisica
1	0	1		●
1	1	1		●
1	1	1		●
1	0	1		●
0	0	0		
1	0	1		●

Memoria
Fisica



Pagina fisica o
Val. Mod. Rif. indirizzo di disco

1	1	1	●
1	0	0	●
1	0	0	●
1	0	1	●
0	0	0	●
1	0	1	●
1	0	1	●
0	0	0	●
1	1	1	●
1	1	1	●
0	0	0	●
1	1	1	●

Memoria
Disco

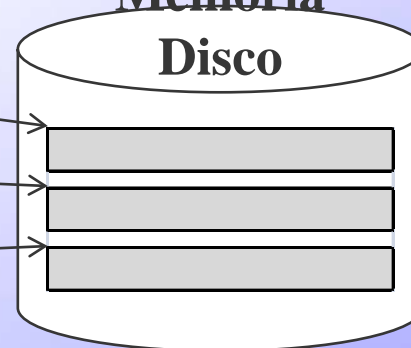


Tabella delle pagine

➤ Write-through

- troppo alto il costo di scrivere in memoria secondaria

➤ Write-back

- scelta obbligata
- ulteriore vantaggio
 - più efficiente trasferire pagine piuttosto che blocchi in quanto il tempo di accesso è più elevato del tempo di trasferimento
- utilizzo di *dirty bit*
 - indica se la pagina è stata modificata e quindi deve essere trasferita dalla memoria principale al disco quando viene sostituita

- Dove può essere messo un blocco?
 - Un posto (mappa diretta), pochi posti (associatività a più vie), qualunque posto (associatività completa)
- Come viene identificato un blocco?
 - Indicizzazione (mappa diretta), ricerca limitata (associatività a più vie), ricerca completa (associatività completa), tabella di ricerca separata (tabella delle pagine)
- Quale blocco va sostituito in caso di fallimento?
 - Casuale, LRU, LFU, 2° chance
- Cosa succede in scrittura?
 - Write-through, write-back
 - Buffer di scrittura tra cache e memoria principale: processore scrive dati in cache e buffer. Problema della saturazione del buffer.