

Cache a blocchi

Per sfruttare la località spaziale, le cache operano a blocchi di b parole. I trasferimenti tra i livelli di cache (compreso $L3 \leftrightarrow DM$) sono sempre a blocchi, soltanto tra $L1$ e la CPU si scambia una parola alla volta (per via della dimensione dei registri).

Per implementare efficientemente una cache a blocchi sono necessari alcuni accorgimenti nella progettazione dell'hardware:

- i bus di collegamento tra le cache devono supportare l'invio *contemporaneo* di b parole (utilizzare un ciclo di bus per parola annulla i vantaggi dell'uso di blocchi);
- analogamente, la memoria deve supportare la lettura contemporanea di indirizzi adiacenti (memoria interallacciata).

I compilatori possono allineare il primo elemento di un array al blocco, in modo che leggendolo vengano caricati i $b - 1$ elementi successivi (e nessun indirizzo precedente, che non fa parte dell'array).