

Figure 3.1: A Typical Computer System.

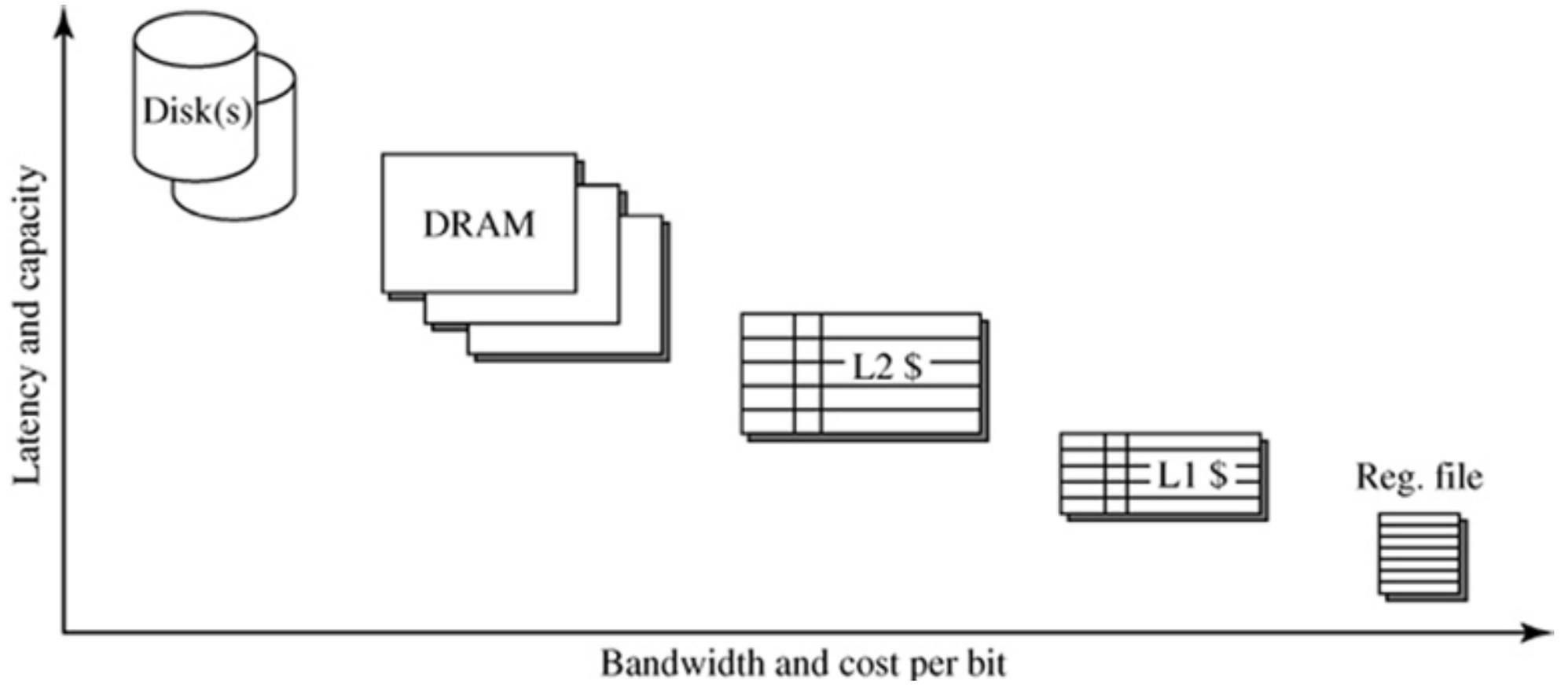
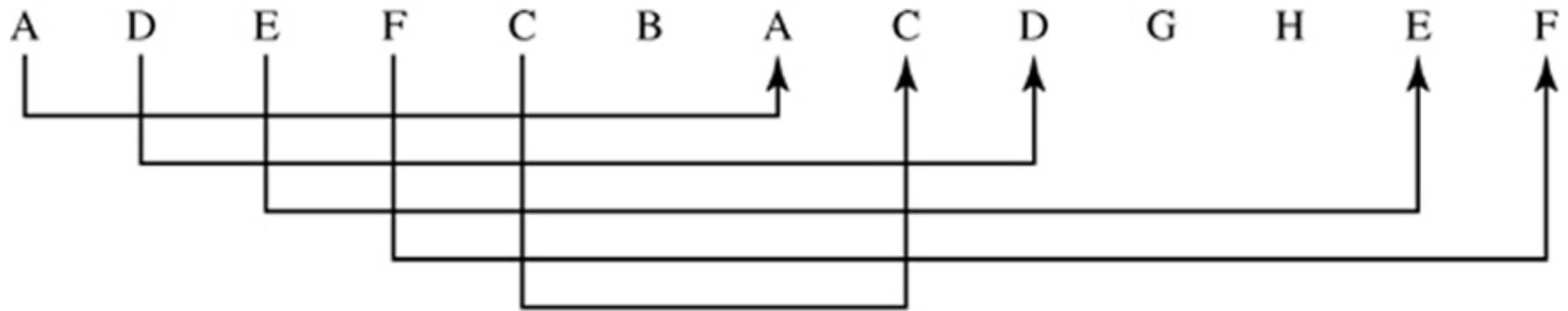
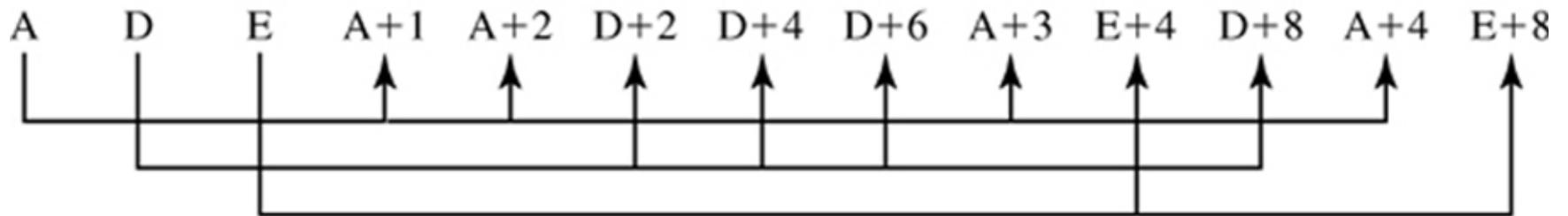


Figure 3.2: Memory Hierarchy Components.



(a)



(b)

Figure 3.3: Illustration of (a) Temporal and (b) Spatial Locality.

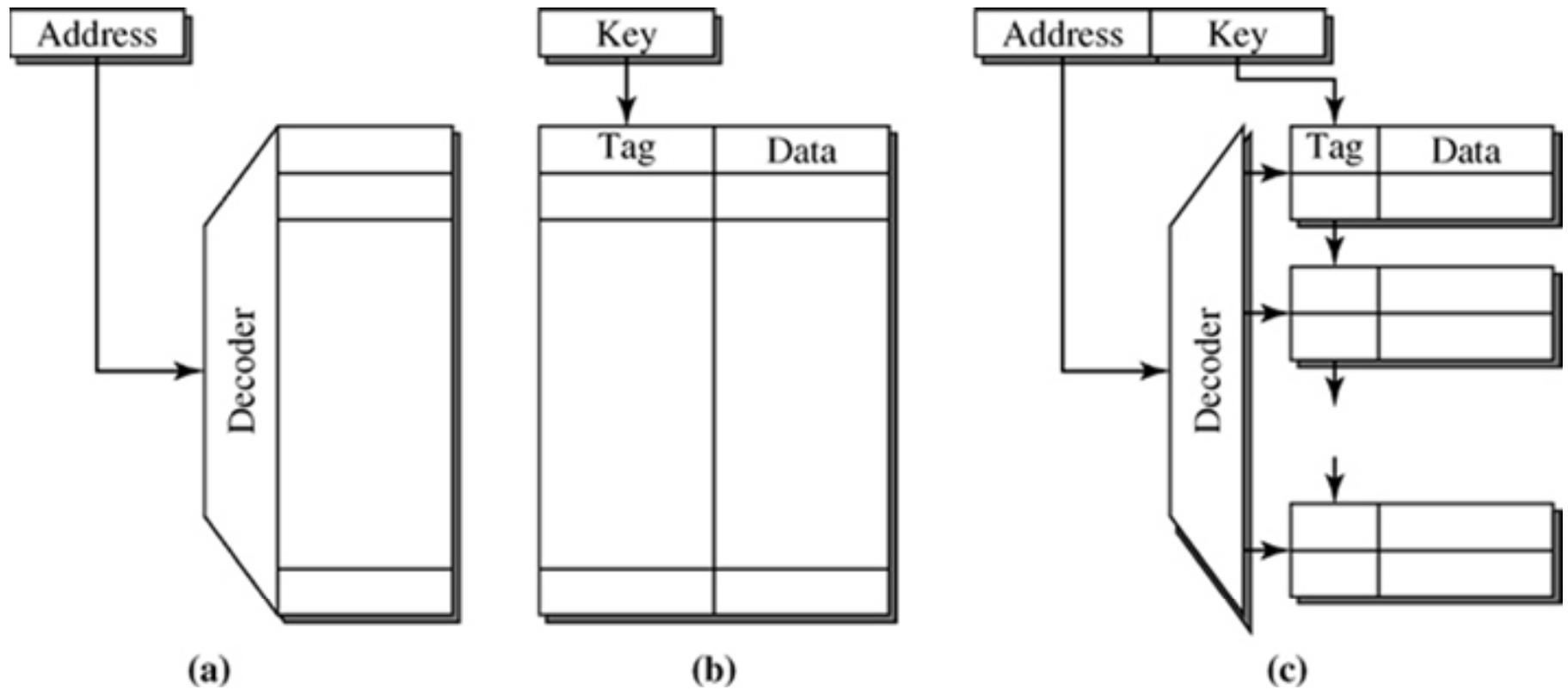


Figure 3.4: Block Placement Schemes: (a) Direct-Mapped, (b) Fully Associative, (c) Set-Associative.

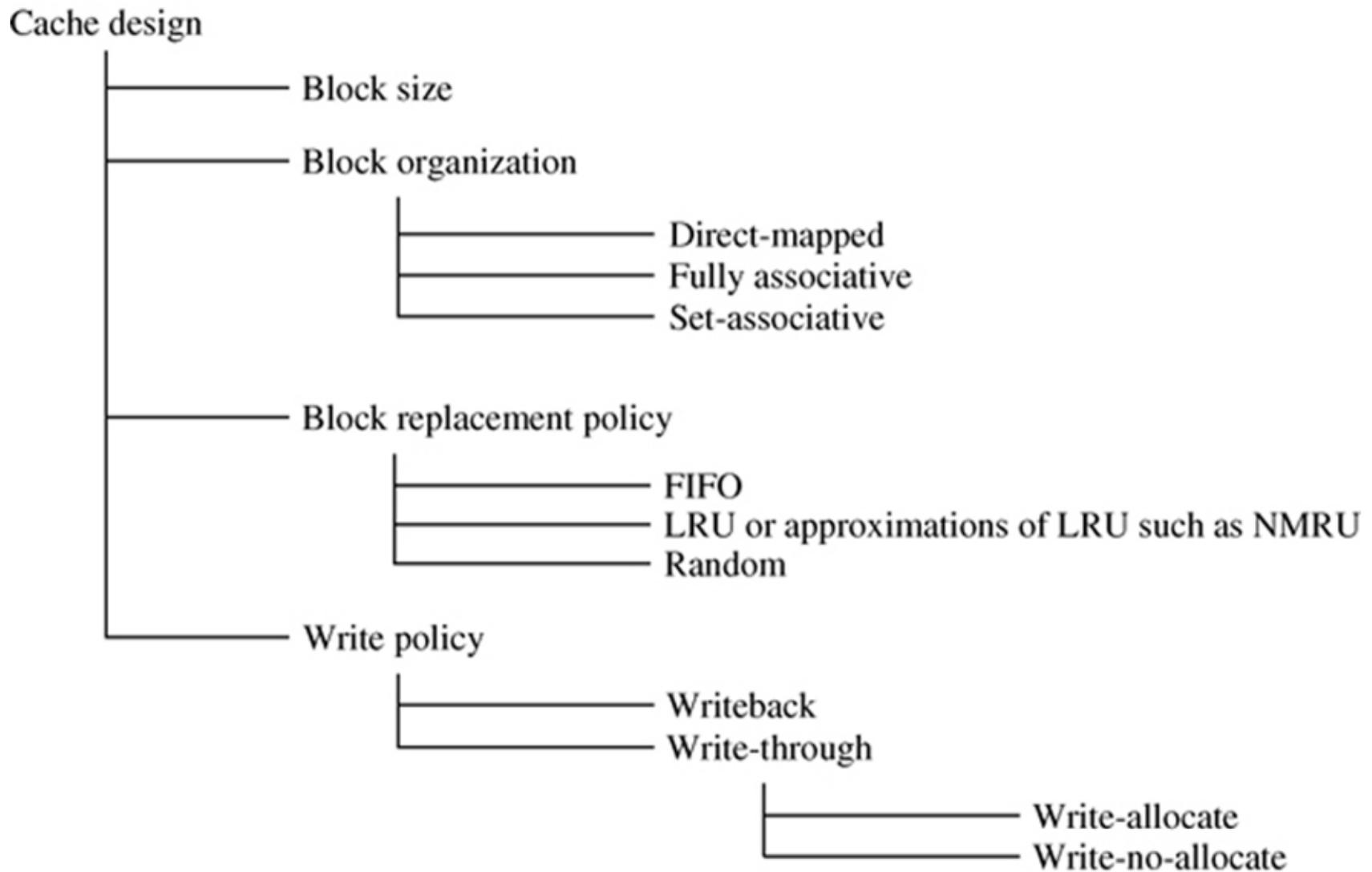


Figure 3.5: Cache Design Parameters.

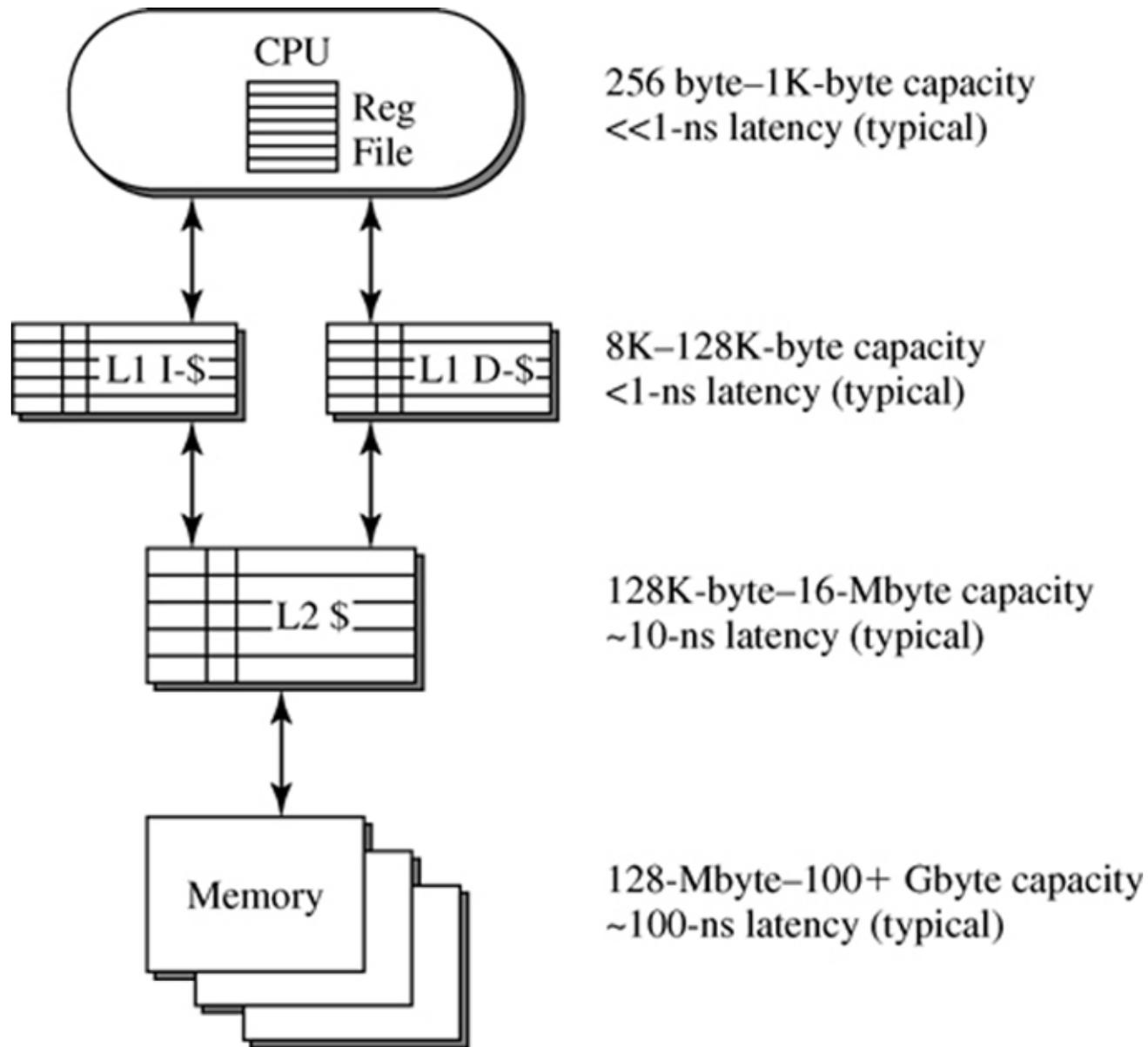


Figure 3.6: A Typical Memory Hierarchy.

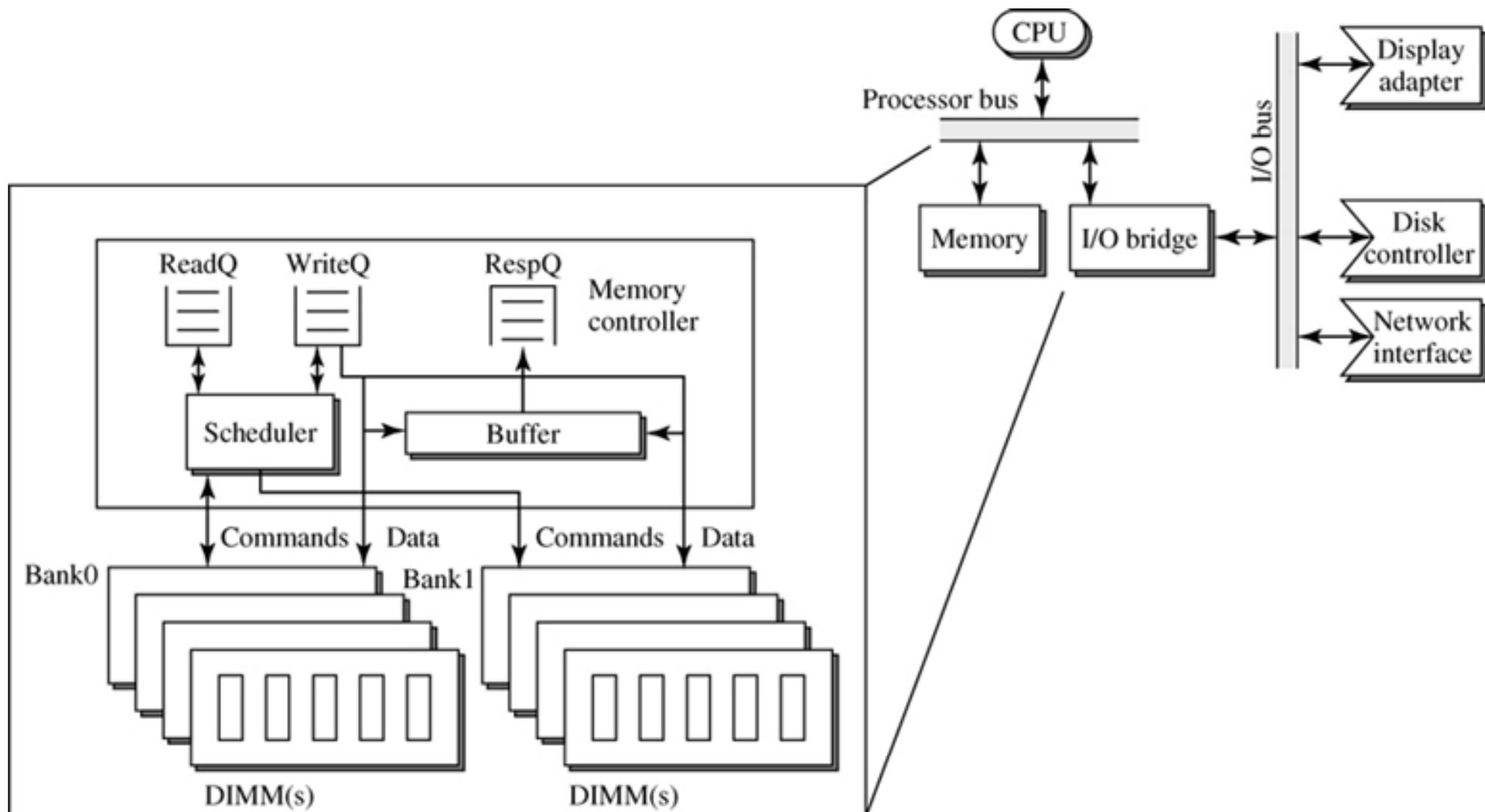


Figure 3.7: Typical Main Memory Organization.

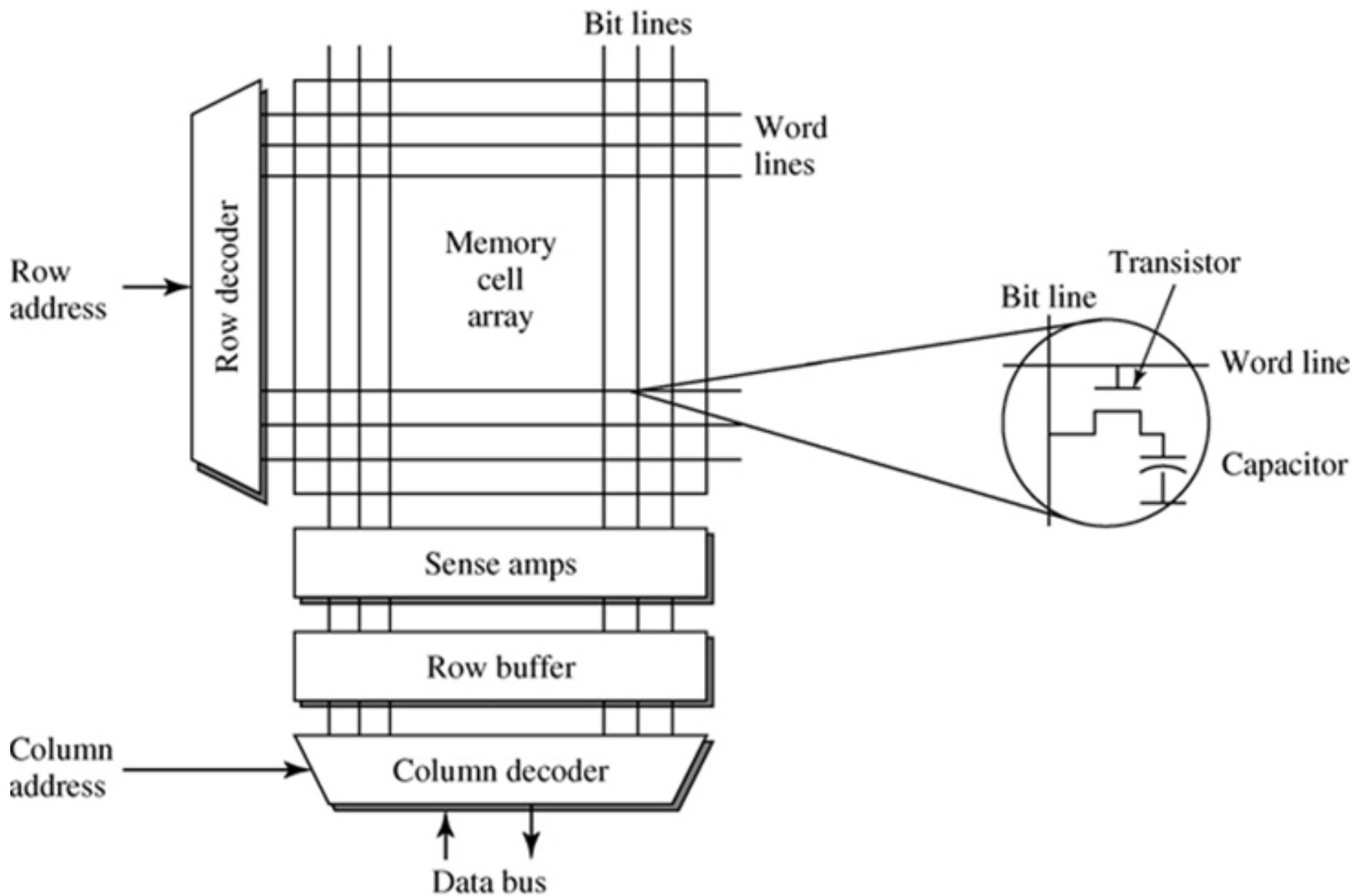


Figure 3.8: DRAM Chip Organization.

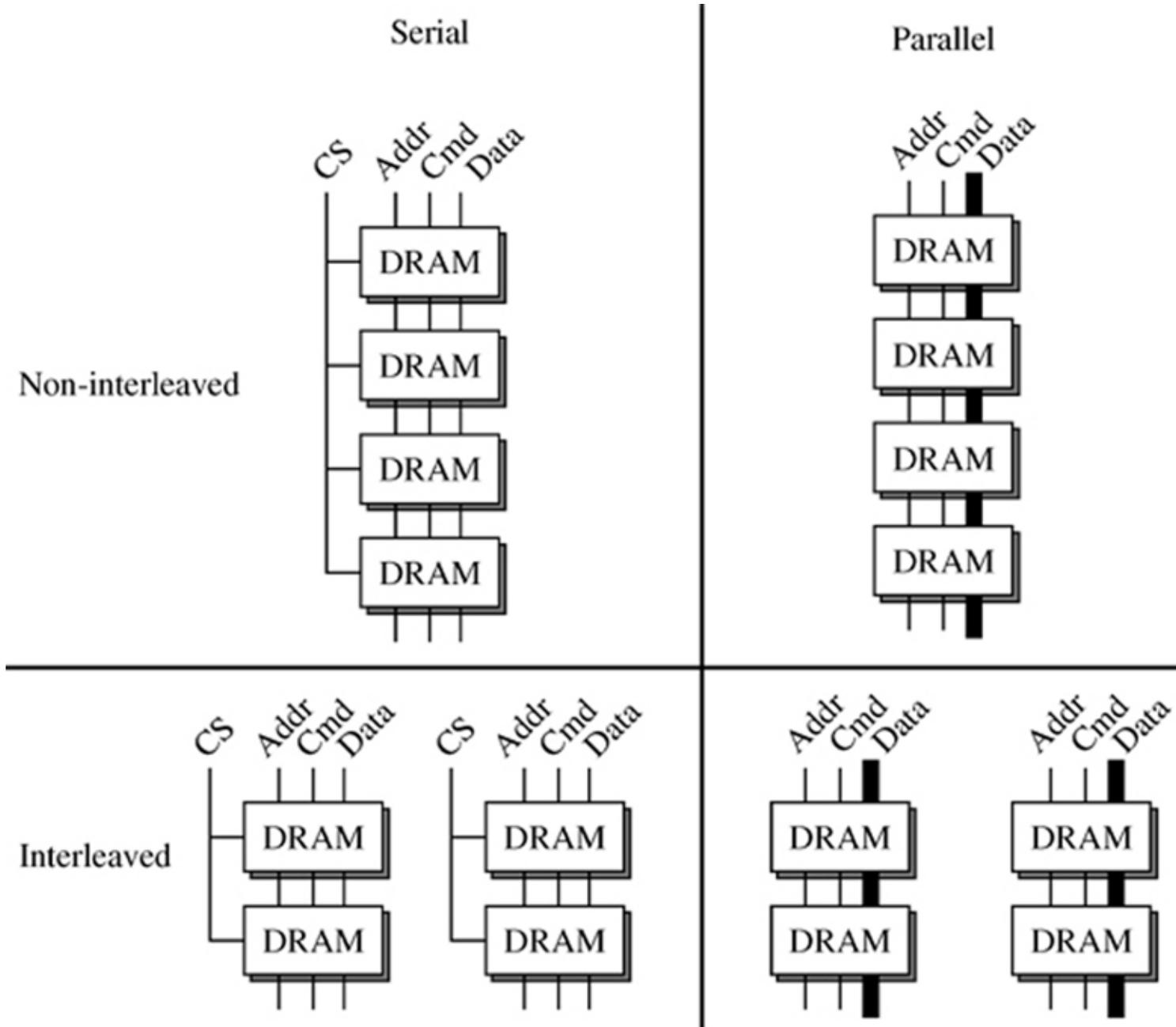


Figure 3.9: Memory Module Organization.

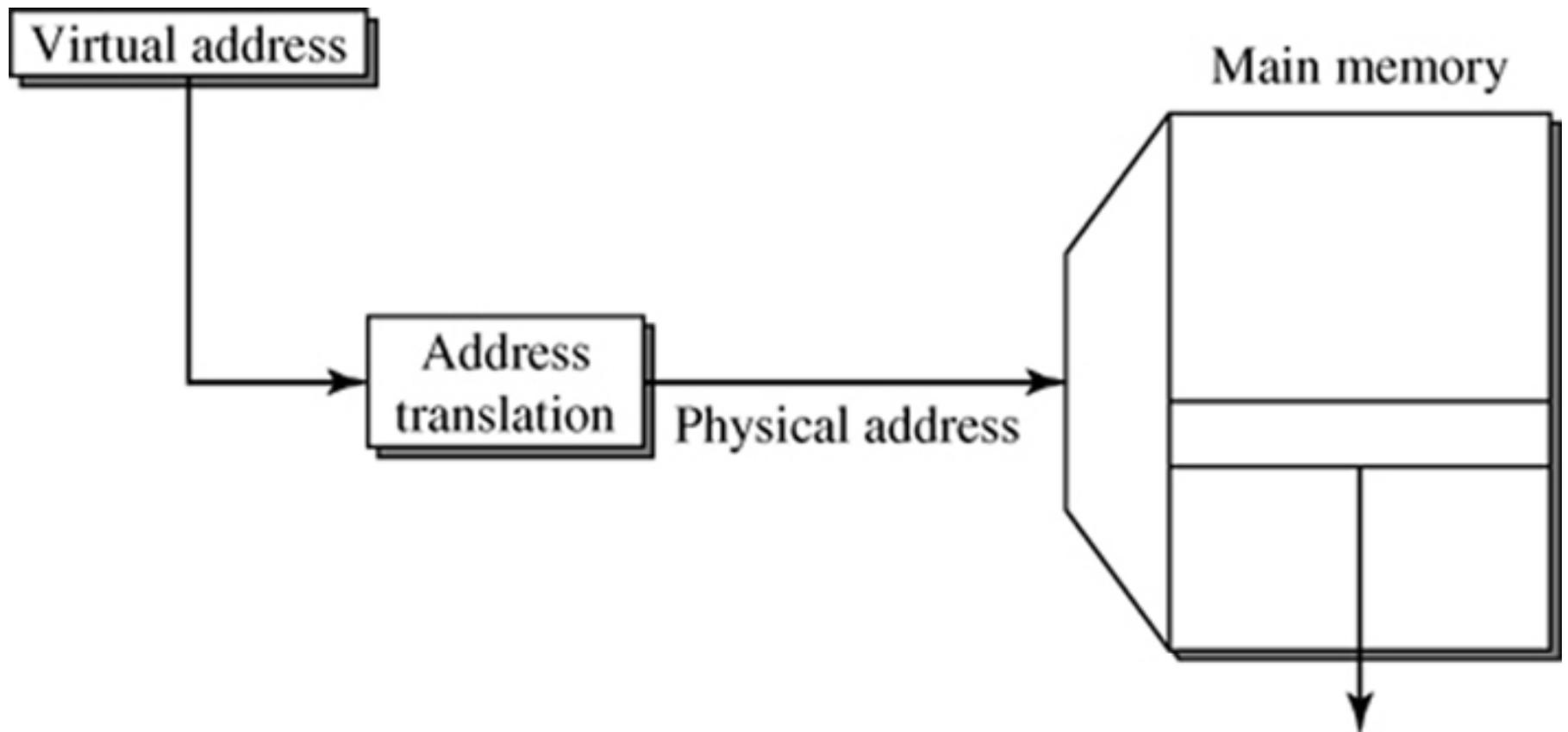


Figure 3.10: Virtual to Physical Address Translation.

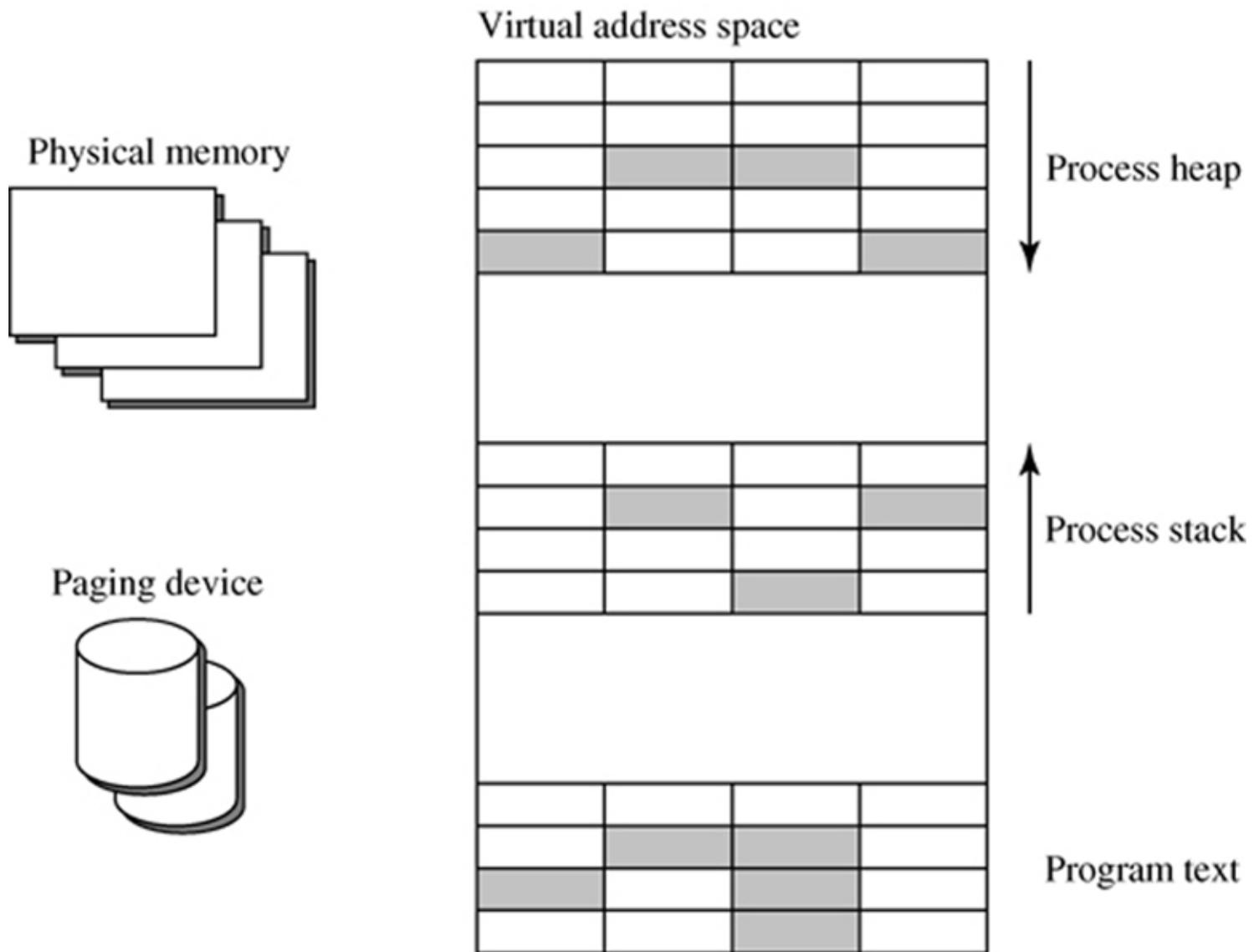


Figure 3.11: Virtual Memory System.

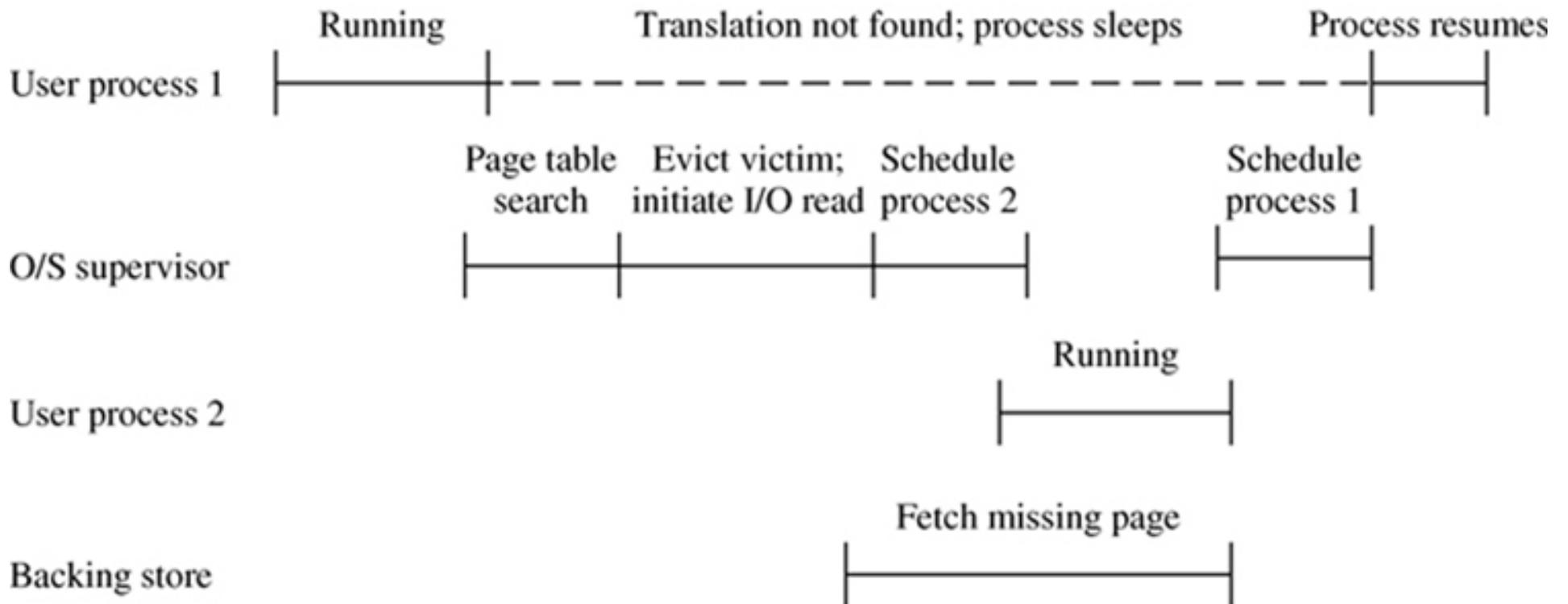


Figure 3.12: Handling a Page Fault.

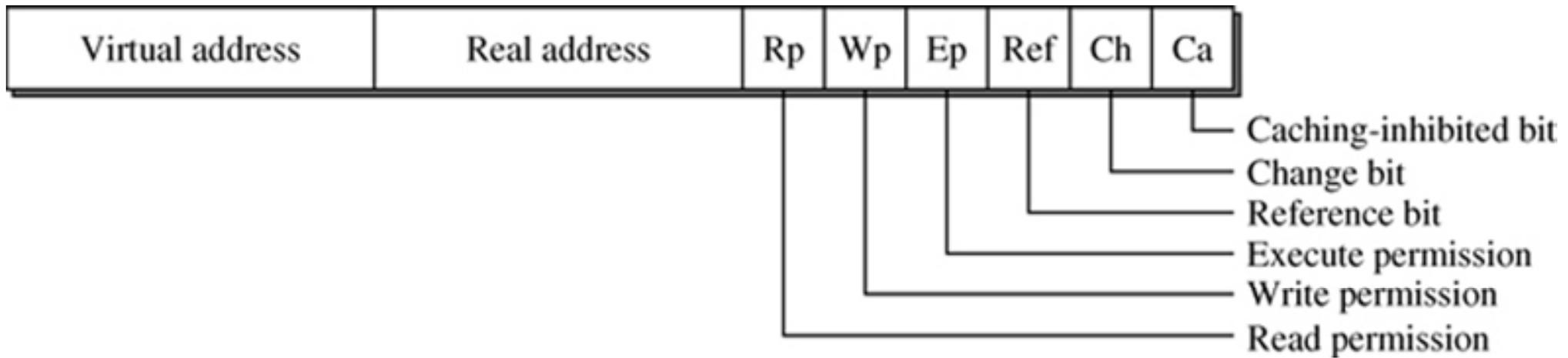


Figure 3.13: Typical Page Table Entry.

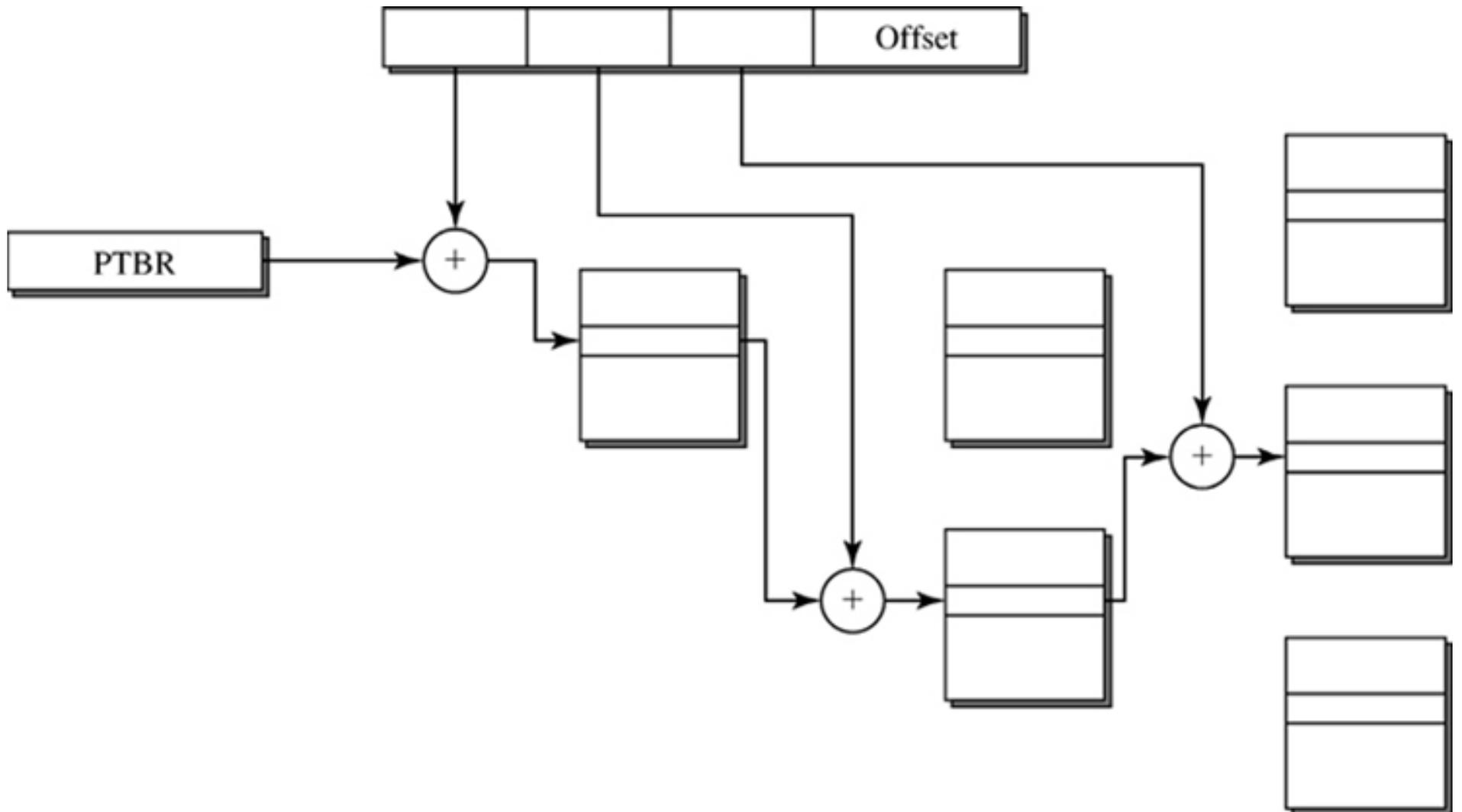


Figure 3.14: Multilevel Forward Page Table.

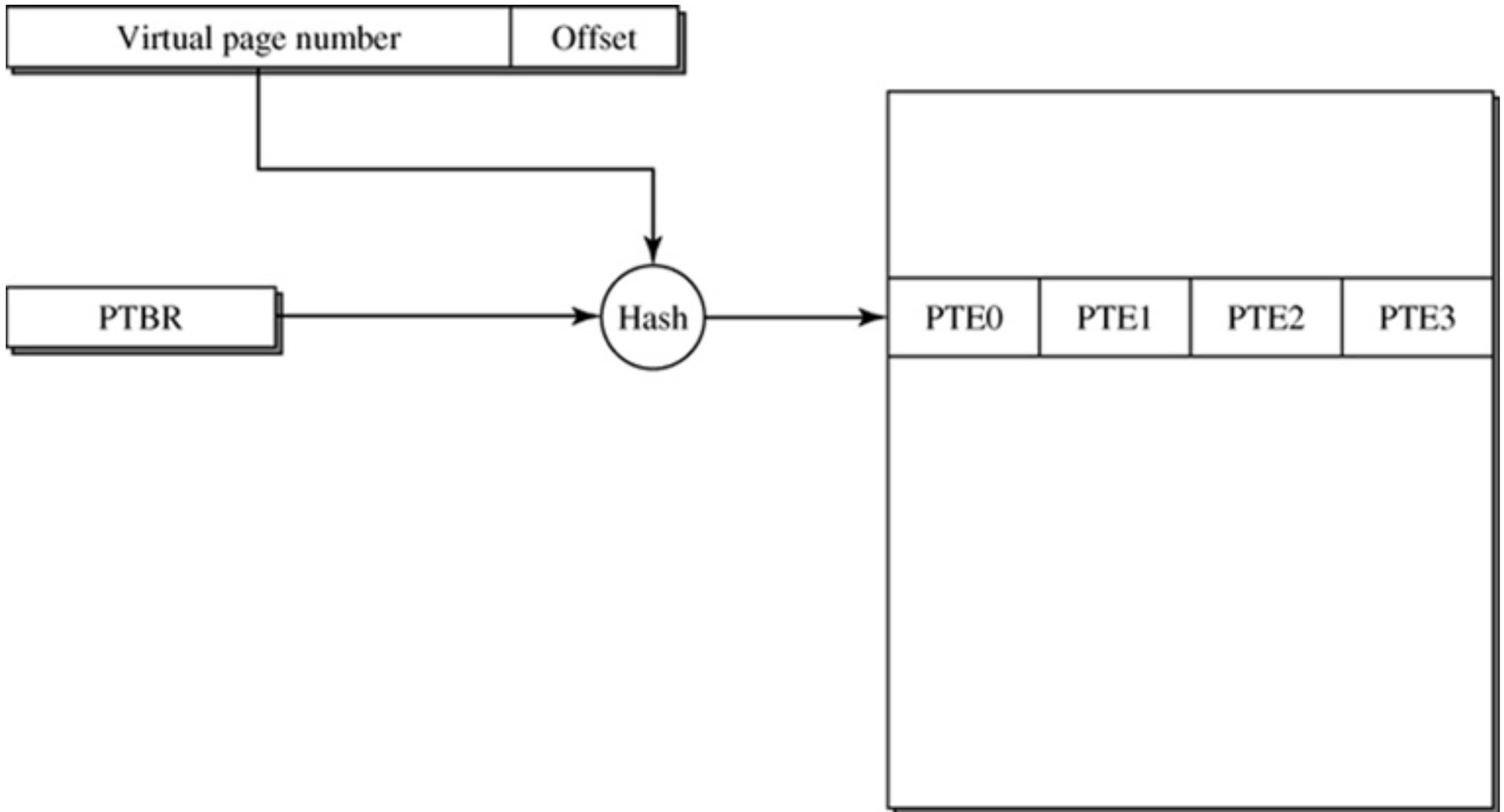


Figure 3.15: Hashed Page Table.

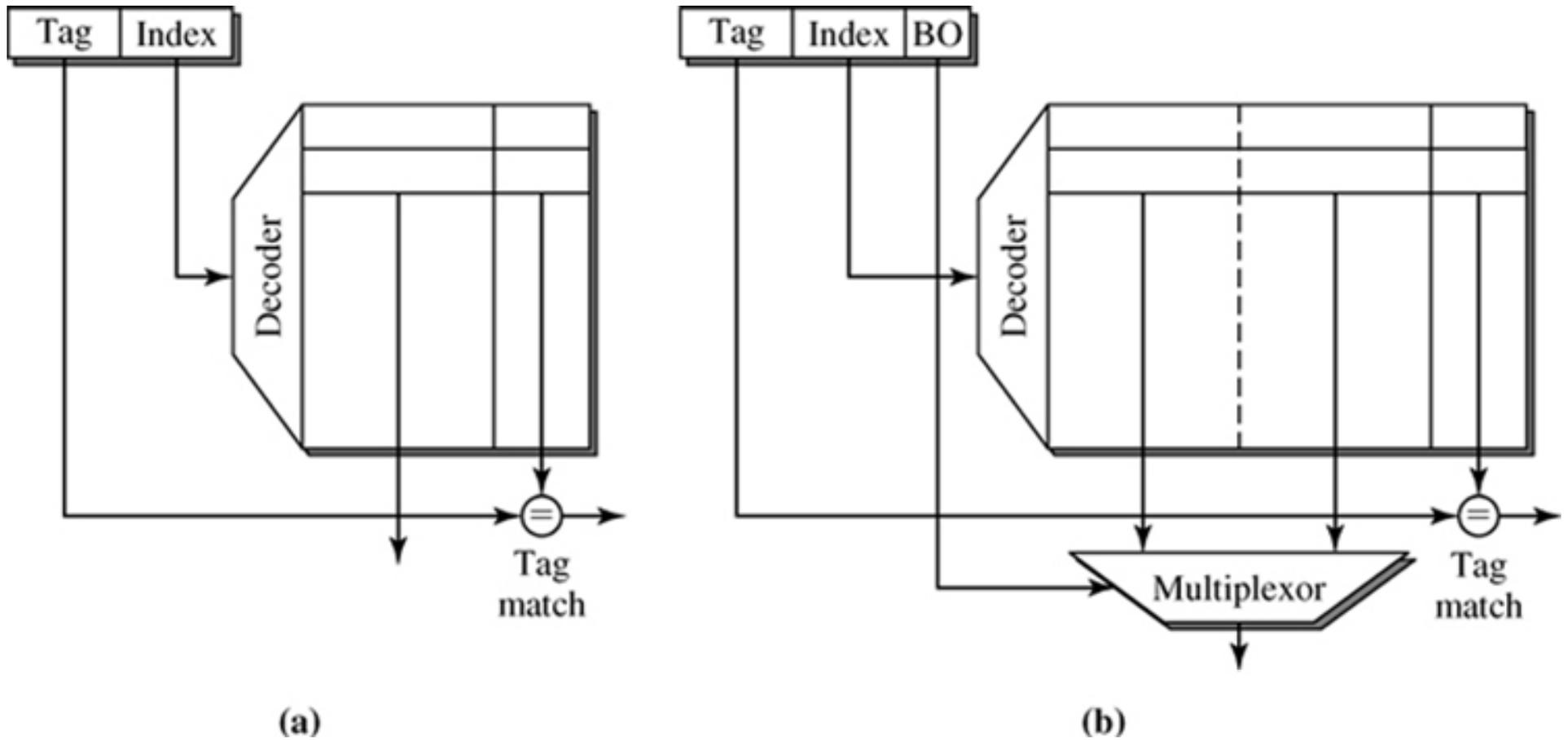


Figure 3.16: Direct-Mapped Caches: (a) Single Word Per Block; (b) Multiword Per Block.

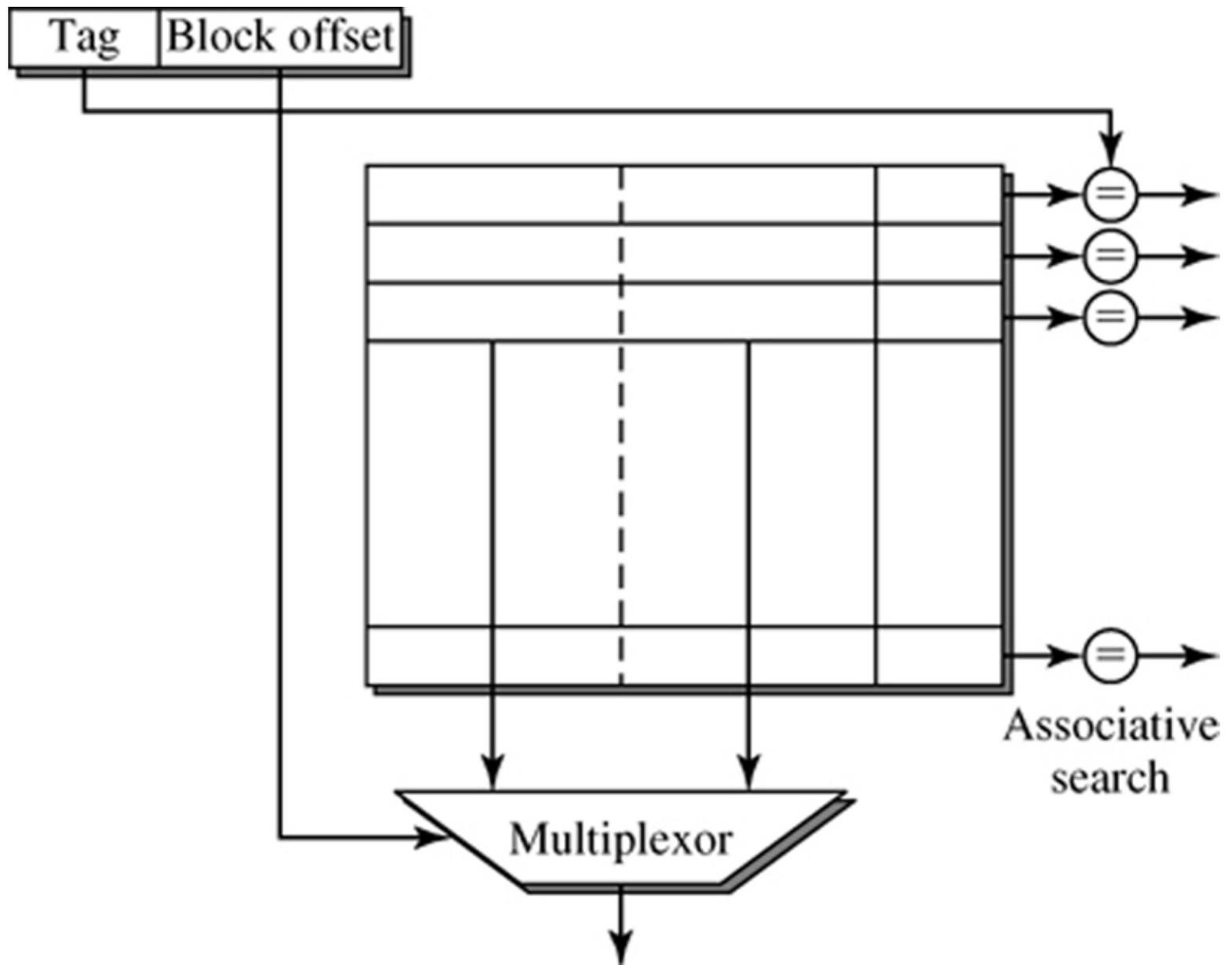


Figure 3.17: Fully Associative Cache.

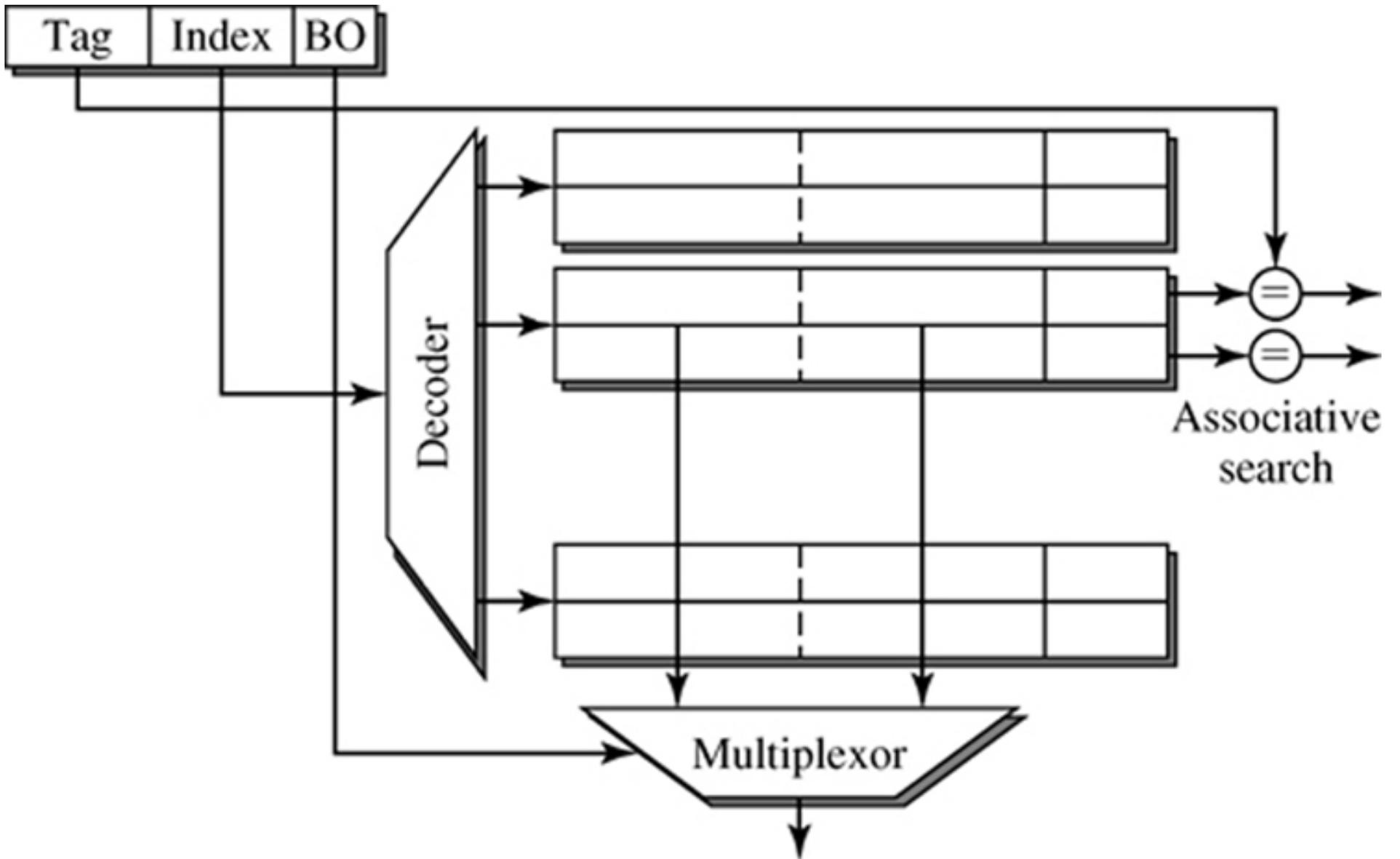


Figure 3.18: Set-Associative Cache.

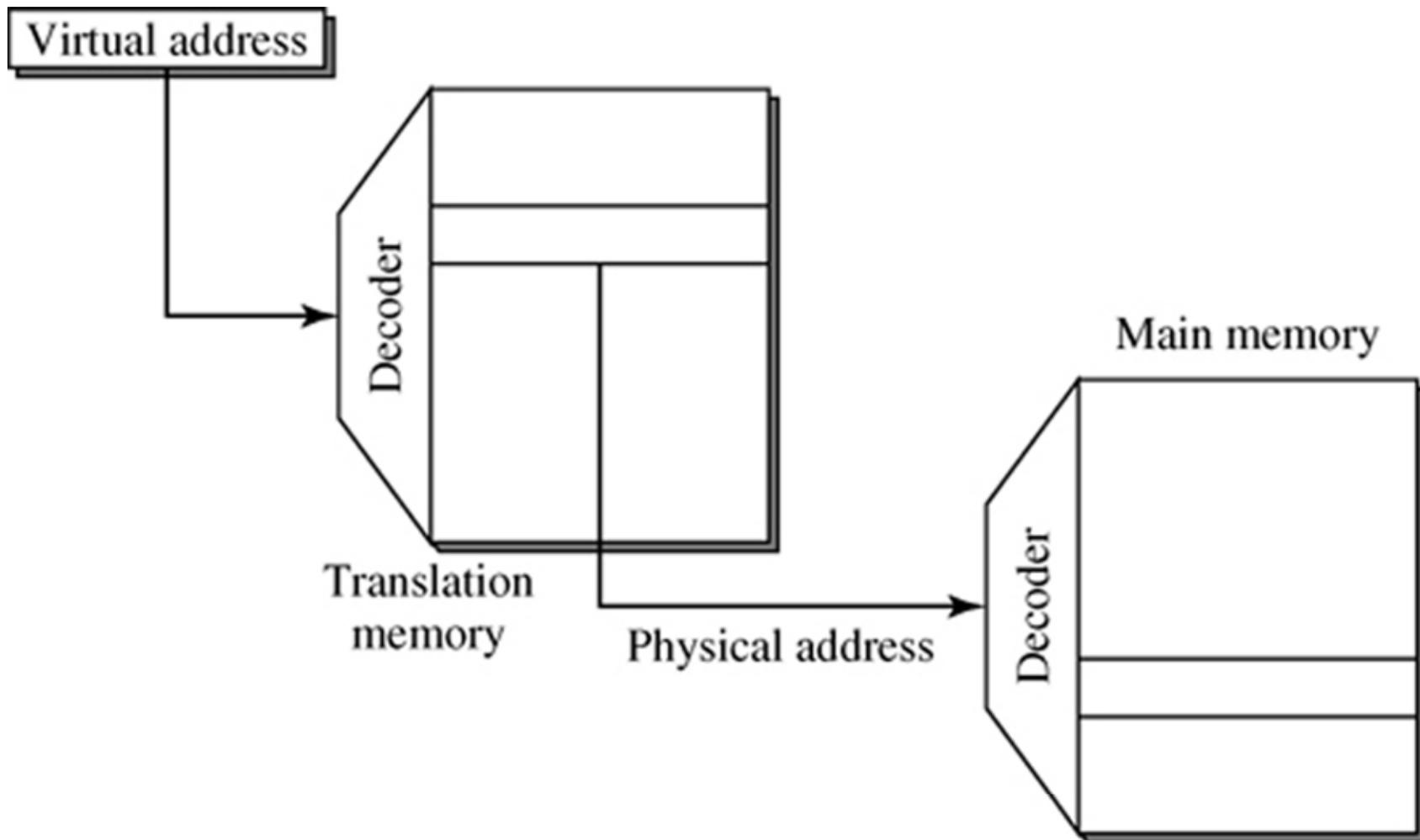


Figure 3.19: Translation of Virtual Word Address to Physical Word Address Using a Translation Memory.

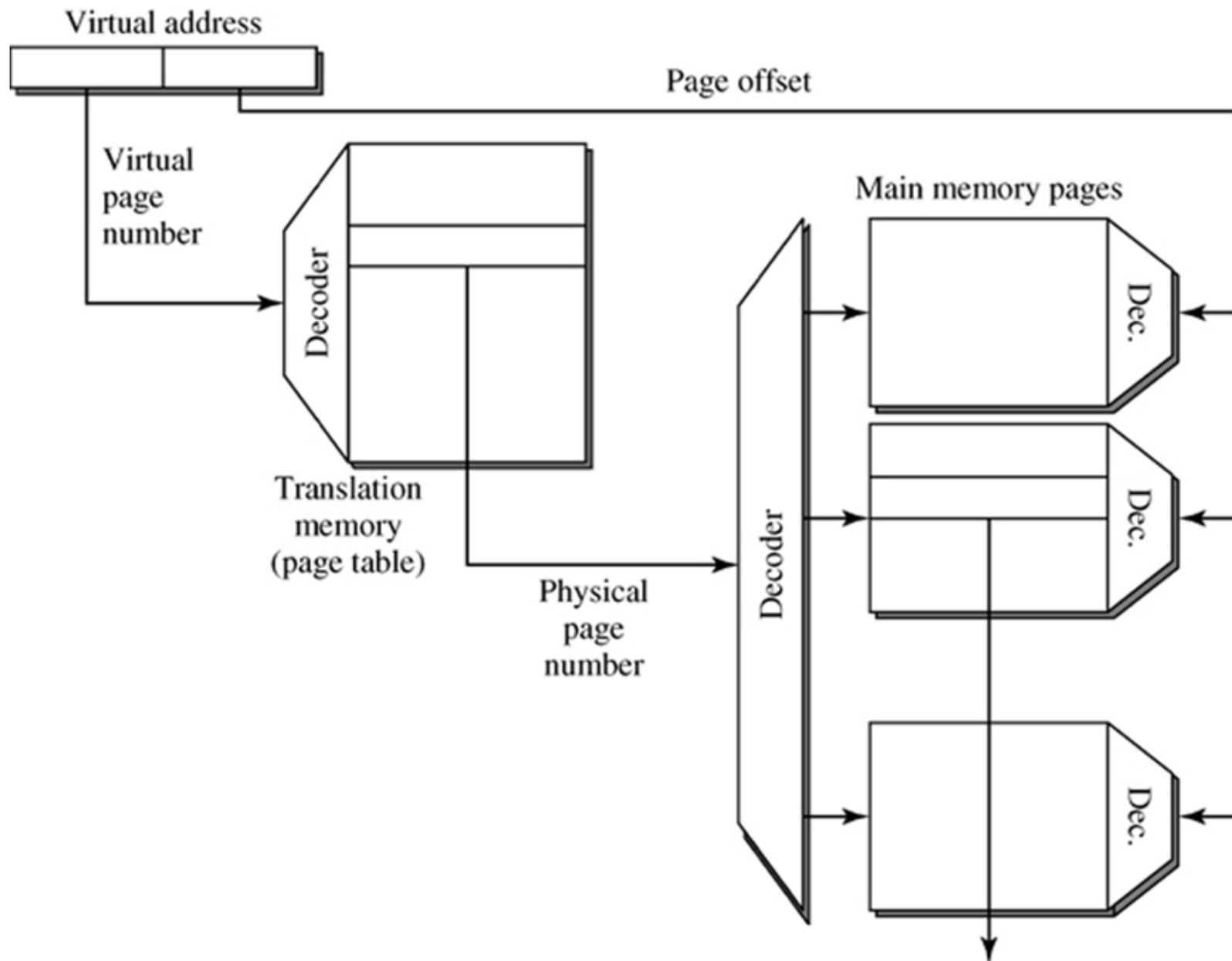


Figure 3.20: Translation of Virtual Page Address to Physical Page Address Using a Translation Memory.

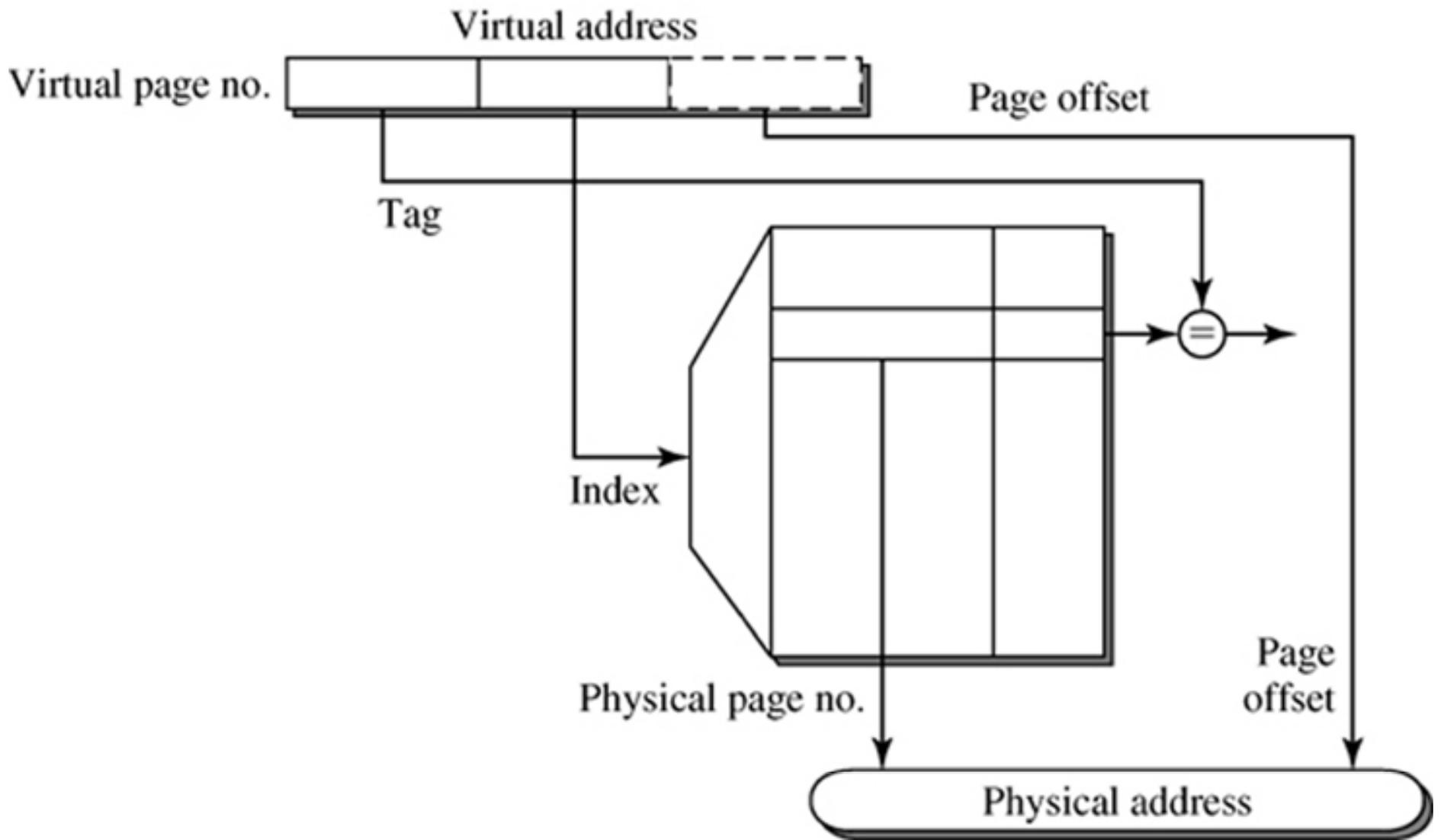
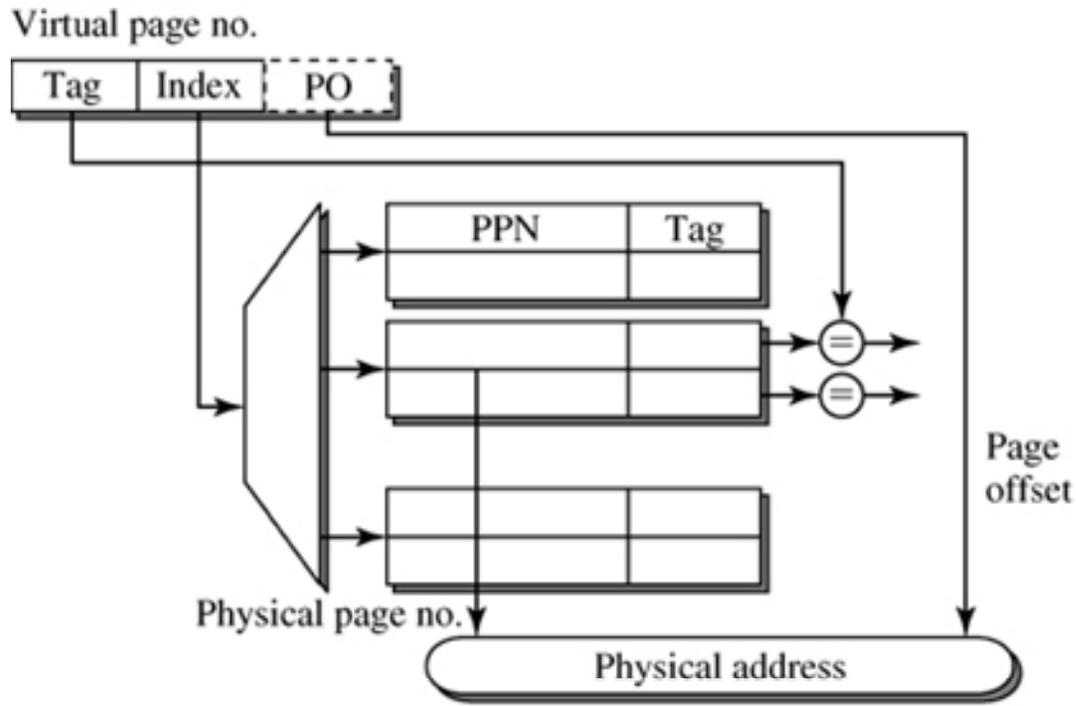
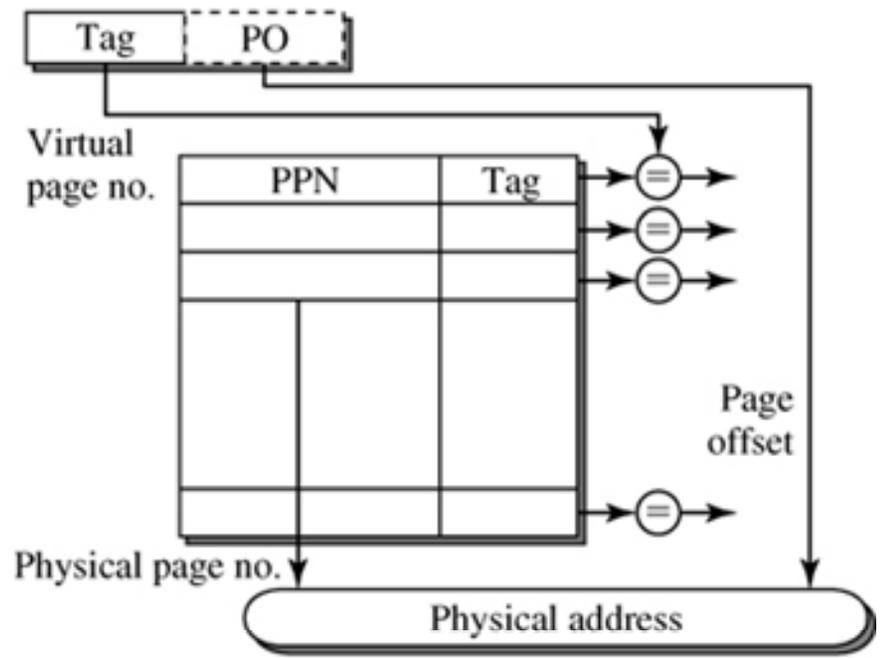


Figure 3.21: Direct-Mapped TLB.



(a)



(b)

Figure 3.22: Associative TLBs: (a) Set-Associative TLB; (b) Fully Associative TLB.

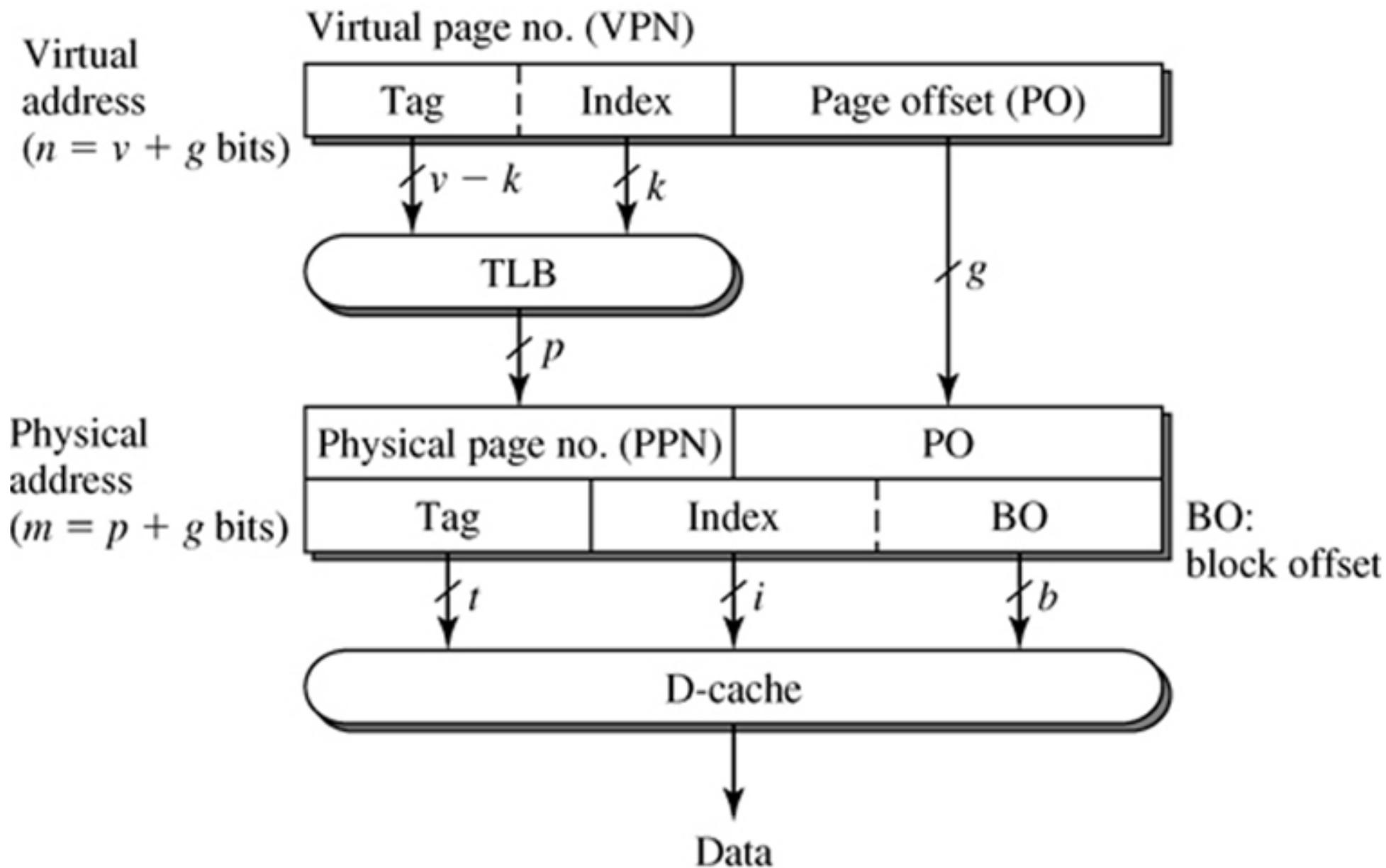


Figure 3.23: Interaction Between the TLB and the Data Cache.

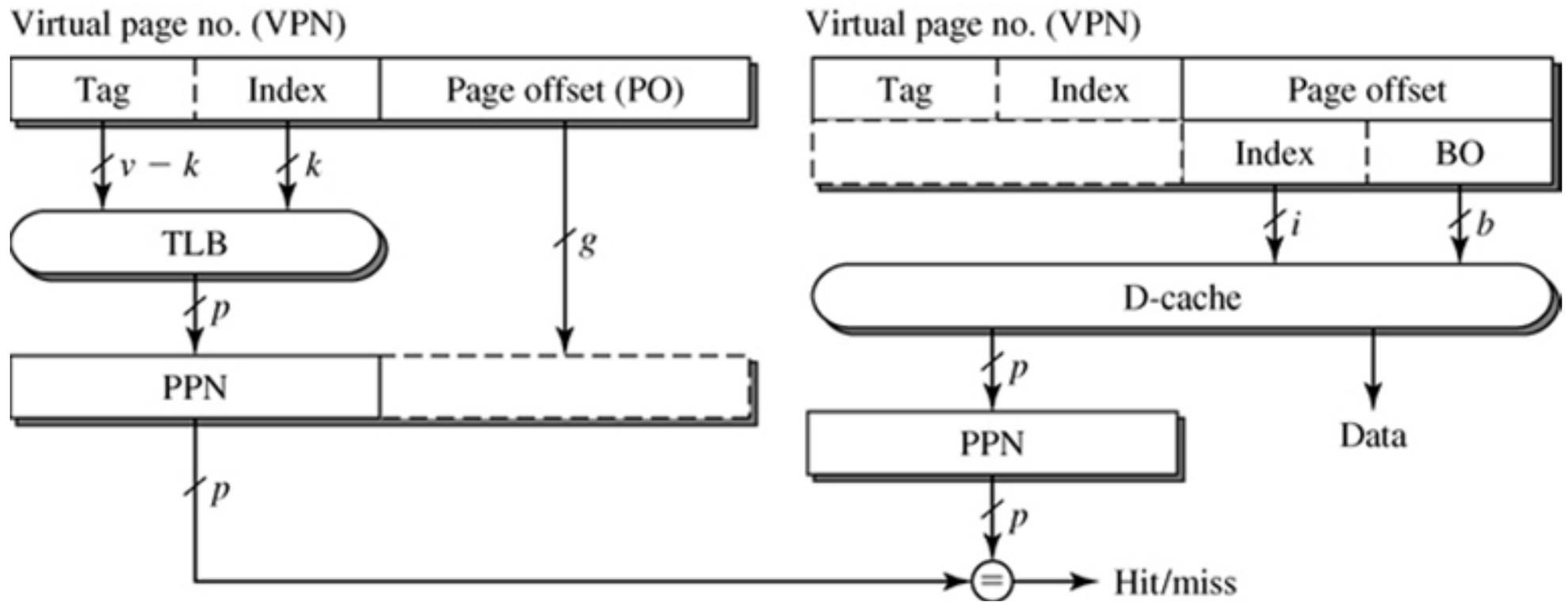


Figure 3.24: Virtually Indexed Data Cache.

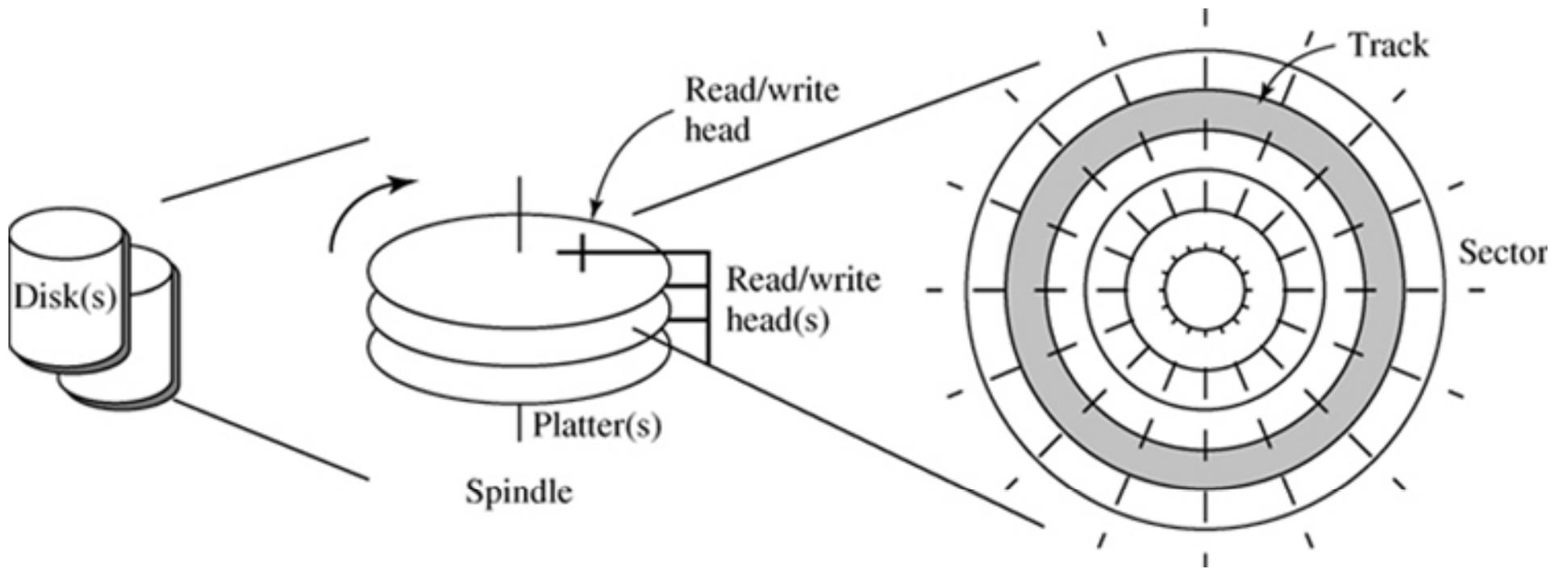


Figure 3.25: Disk Drive Structure.

A0	B0	C0	D0
A1	B1	C1	D1
A2	B2	C2	D2

Independent

A0	A0	A0	A0
A1	A1	A1	A1
A2	A2	A2	A2
B0	B0	B0	B0
B1	B1	B1	B1
B2	B2	B2	B2
C0	C0	C0	C0
C1	C1	C1	C1
C2	C2	C2	C2

Fine-grained

A0	A1	A2	A3
B0	B1	B2	B3
C0	C1	C2	C3

Coarse-grained

Each disk is represented by a column, each block is represented by a name (A0, A1, A2, etc.), and blocks from the same file are named with the same letter (e.g., A0, A1, and A2 are all from the same file). Independent disk arrays place related blocks on the same drive. Fine-grained interleaving subdivides each block and stripes it across multiple drives. Coarse-grained interleaving stripes related blocks across multiple drives.

Figure 3.26: Striping Data in Disk Arrays.

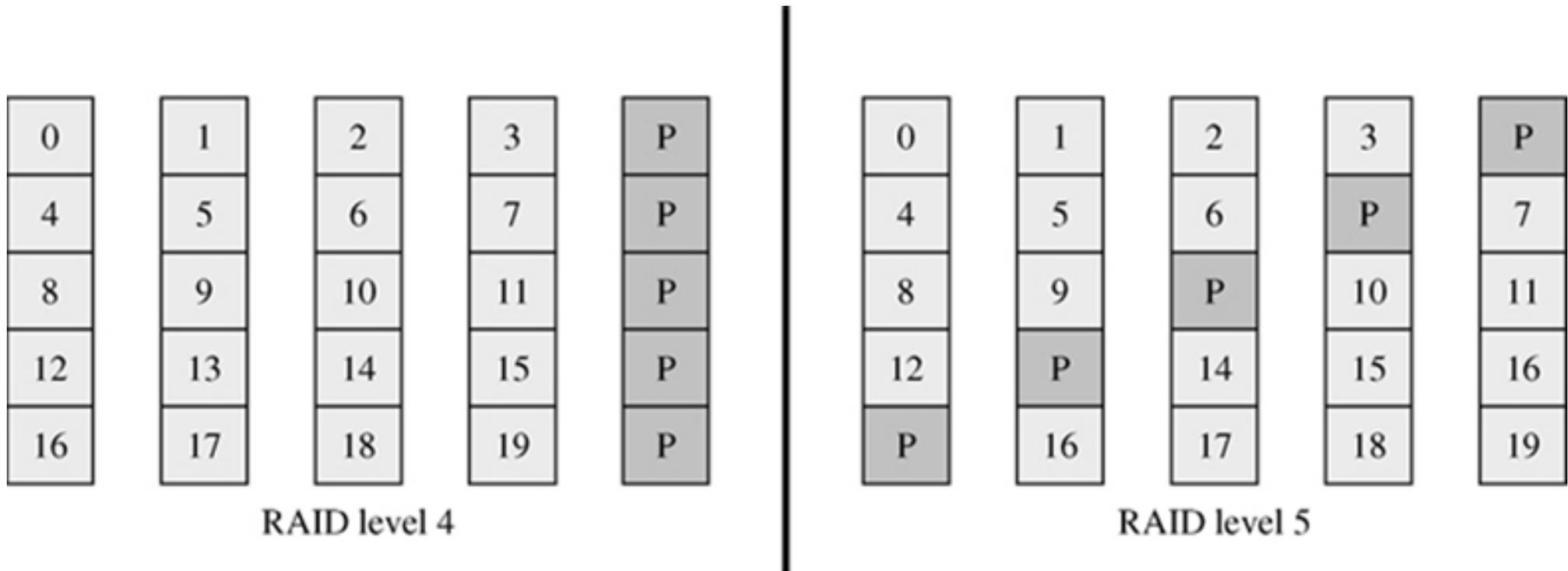


Figure 3.27: Placement of Parity Blocks in RAID Level 4 (Left) vs. RAID Level 5 (Right).

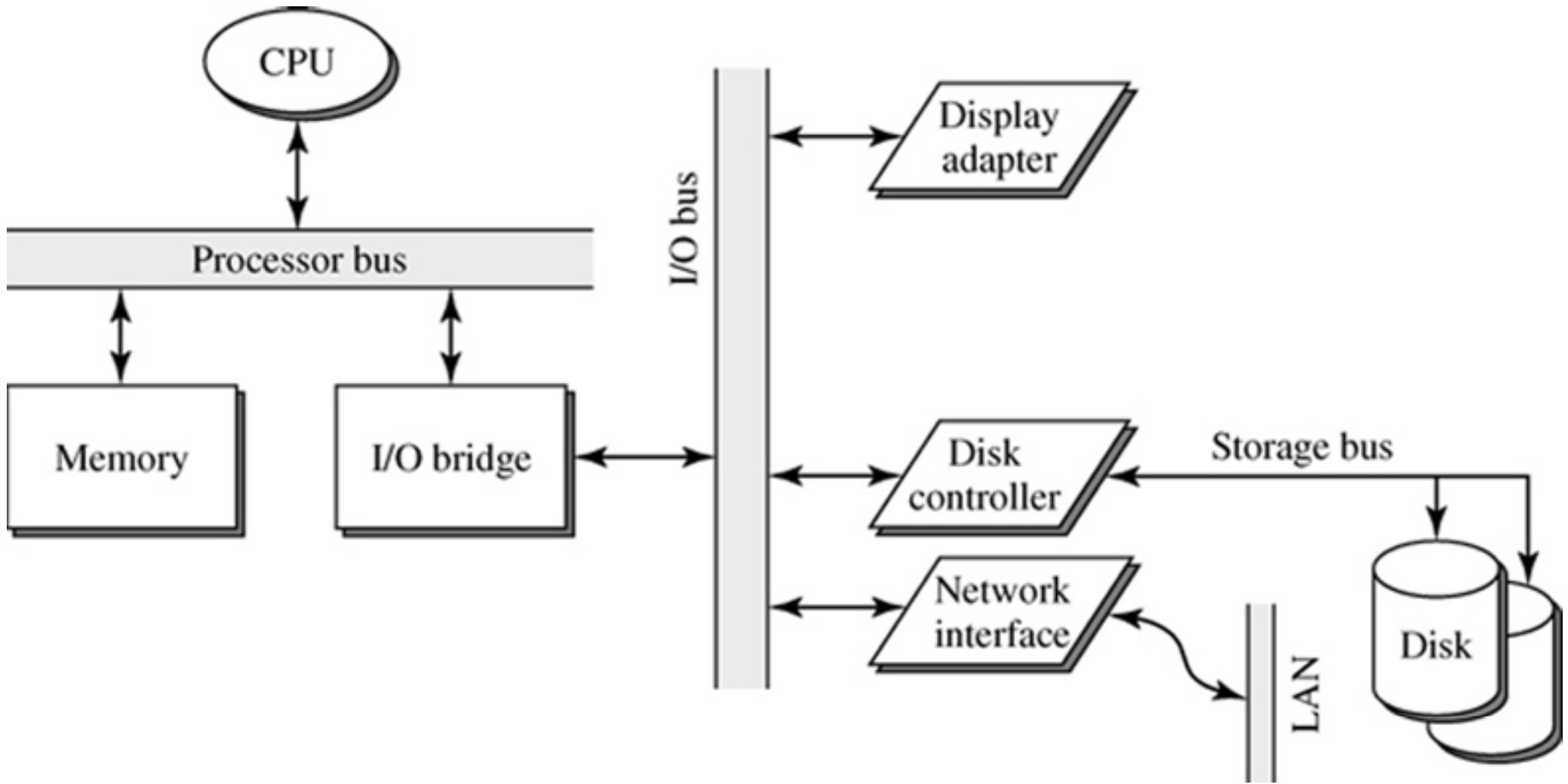


Figure 3.28: Different Types of Computer System Busses.

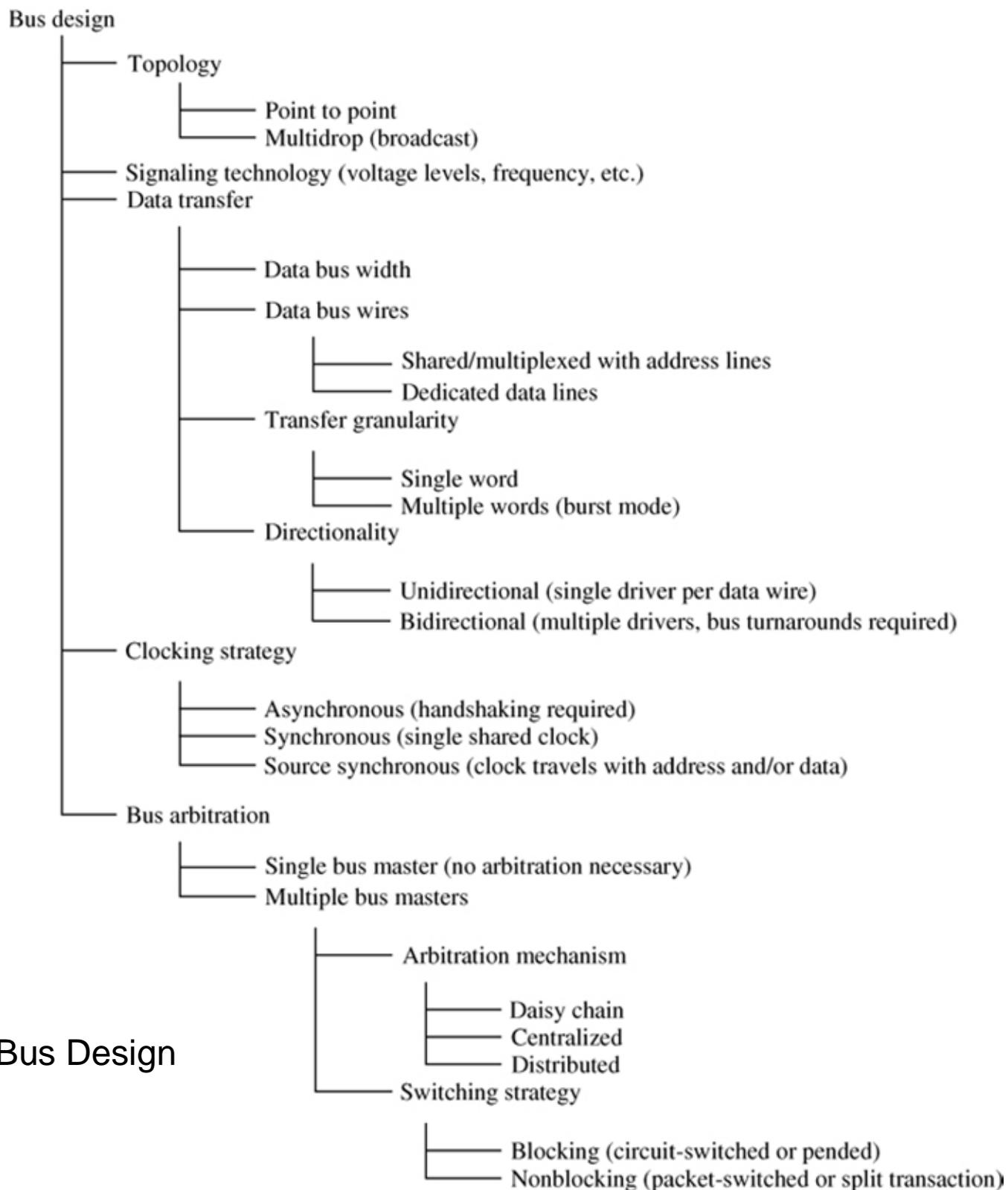


Figure 3.29: Bus Design Parameters.

I/O device communication

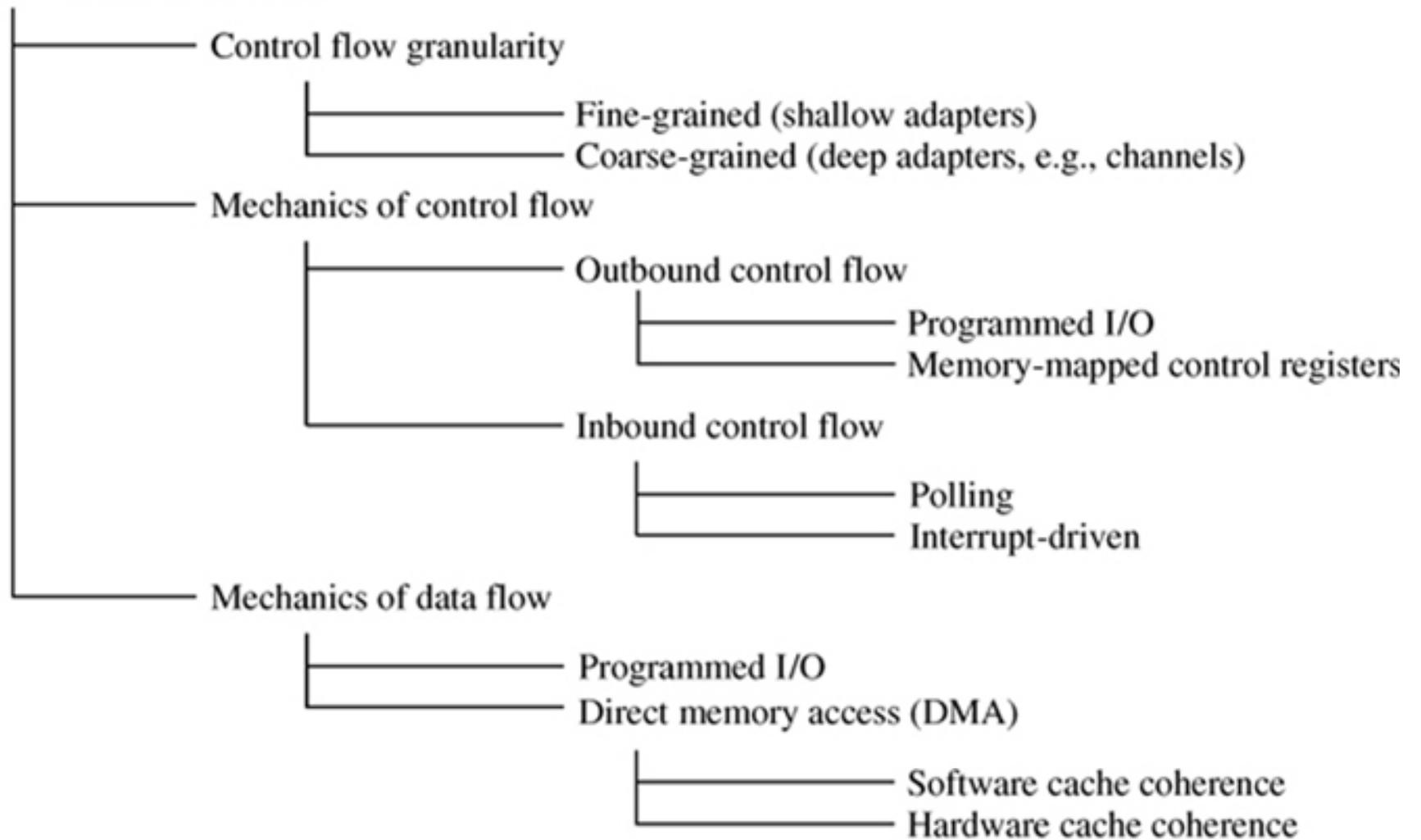
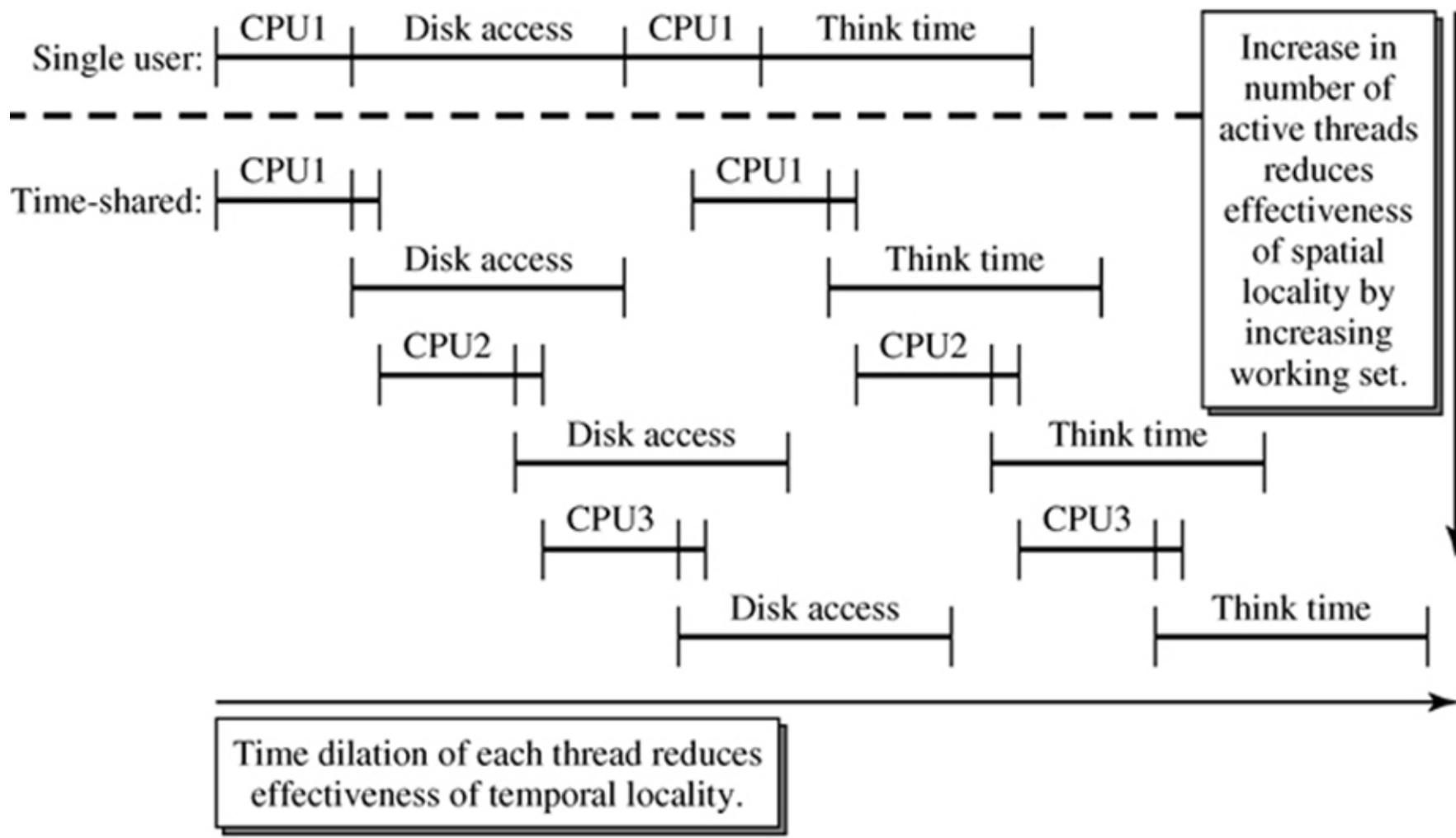


Figure 3.30: Communication with I/O Devices.



In this example, three users time-share the CPU, overlapping their CPU usage with the disk latency and think time of the other interactive users. This increases overall throughput, since the CPU is always busy, but can increase the latency observed by each user. Latency increases due to context switch overhead and queuing delay (waiting for the CPU while another user is occupying it). Temporal and spatial locality are adversely affected by time-sharing.

Figure 3.31: Time-Sharing the CPU.