Chapter 6: The Memory Hierarchy

Locality

- **programs with good locality run faster than programs with poor locality**

Temporal Locality

- in a program with good **temporal locality**, a memory location that is referenced once is likely to be referenced again multiple times in the near future

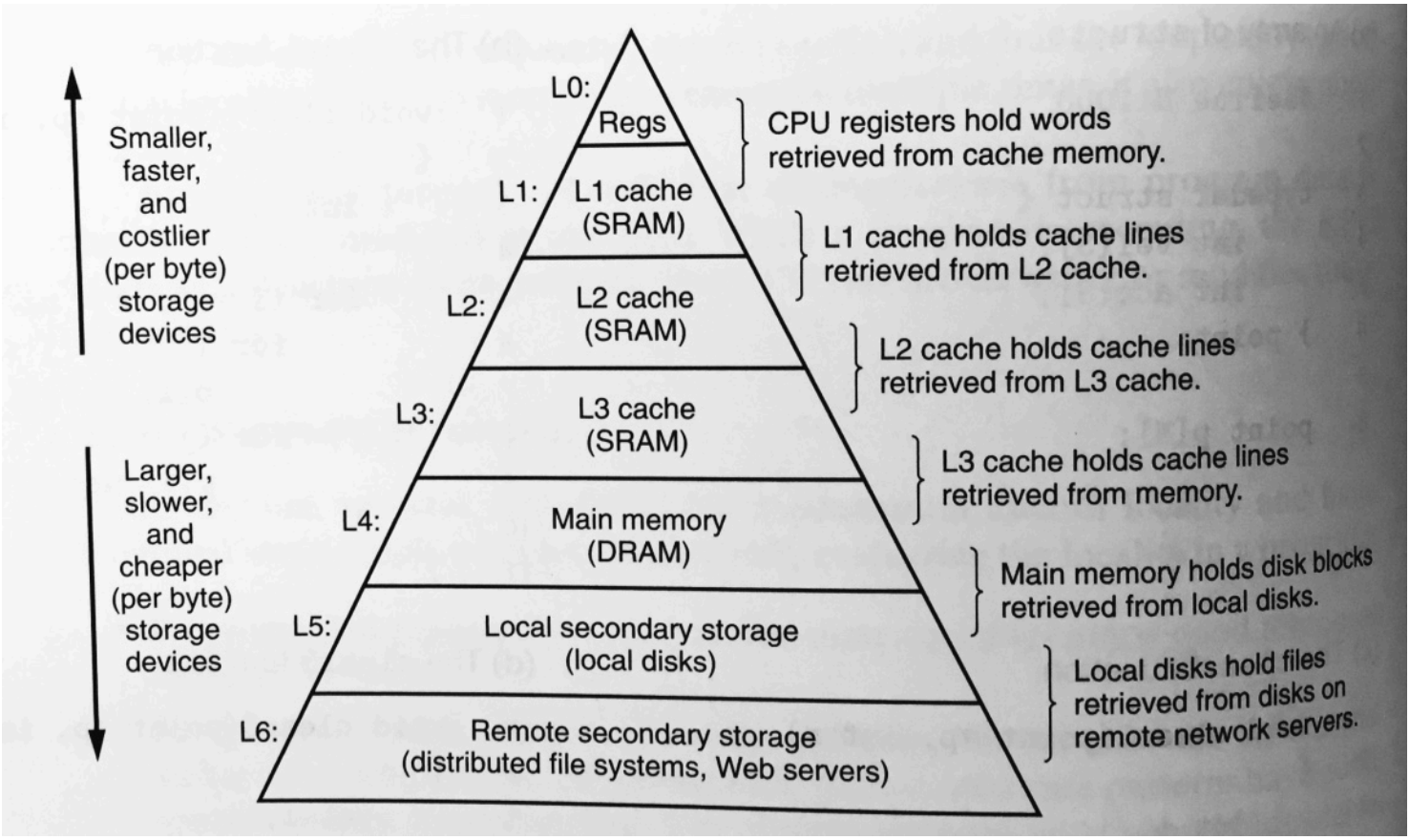
Spatial Locality

- in a program with good **spatial locality**, if a memory location is referenced once, then the program is likely to reference a nearby memory location in the near future

Comparisons

- programs that repeatedly reference the same variables enjoy good temporal locality
- visiting every *k*th element of a contiguous vector is called a **stride-k reference pattern**
- for programs with stride-*k* reference patterns, the smaller the stride, the better the spatial locality
    - program with stride-1 reference patterns have good spatial locality
    - programs that hop around memory with large strides have poor spatial locality
- loops have good temporal and spatial locality with respect to instruction fetches
    - the smaller the loop body and the greater the number of loop iterations, the better the locality

Memory Hierarchy


Smaller, faster, and costlier (per byte) storage devices
Larger, slower, and cheaper (per byte) storage devices
L0:
Regs
CPU registers hold words retrieved from cache memory.
L1:
L1 cache (SRAM)
L1 cache holds cache lines retrieved from L2 cache.
L2:
L2 cache (SRAM)
L2 cache holds cache lines retrieved from L3 cache.
L3:
L3 cache (SRAM)
L3 cache holds cache lines retrieved from memory.
L4:
Main memory (DRAM)
Main memory holds disk blocks retrieved from local disks.
L5:
Local secondary storage (local disks)
Local disks hold files retrieved from disks on remote network servers.
L6:
Remote secondary storage (distributed file systems, Web servers)