Cache memory

- Accessing main memory is relatively slow, as it is not located on the CPU
- Recently accessed memory, and memory that is likely to be accessed soon, is stored in the CPU cache
- Accessing cache is on the order of 1 ns (nano second)
- Accessing main memory is on the order of 100 ns
- Main memory is on the order of GB
- CPU cache is on the order of MB
- Locality principle
  - Memory references made in any short time interval tend to use only a fraction of total memory
  - The next memory reference will likely be an address near the last
  - Cache lines - blocks of memory stored in cache, typically 64 bytes

- What caches well?
  - Arrays - stored in a contiguous area of memory
  - Local variables - stored on the stack, another contiguous area of memory

- What doesn't cache as well
  - Linked lists - not stored in contiguous memory, could be scattered in the heap

Secondary memory

- Main memory is too small to store everything
  - Media - images, audio, video
  - DataBases
  - etc.
- Main memory does not persist when the computer is powered off
- Secondary memory is cheaper, larger, slower, and persists even without power

Magnetic disks

- Data is stored on spinning platters
- A platter has many tracks, which are rings of data (bits) organized into sectors

Solid state drive

- Stores data using flash memory
- No moving parts
- Faster and more energy efficient than magnetic disks
- Still more expensive than magnetic disks