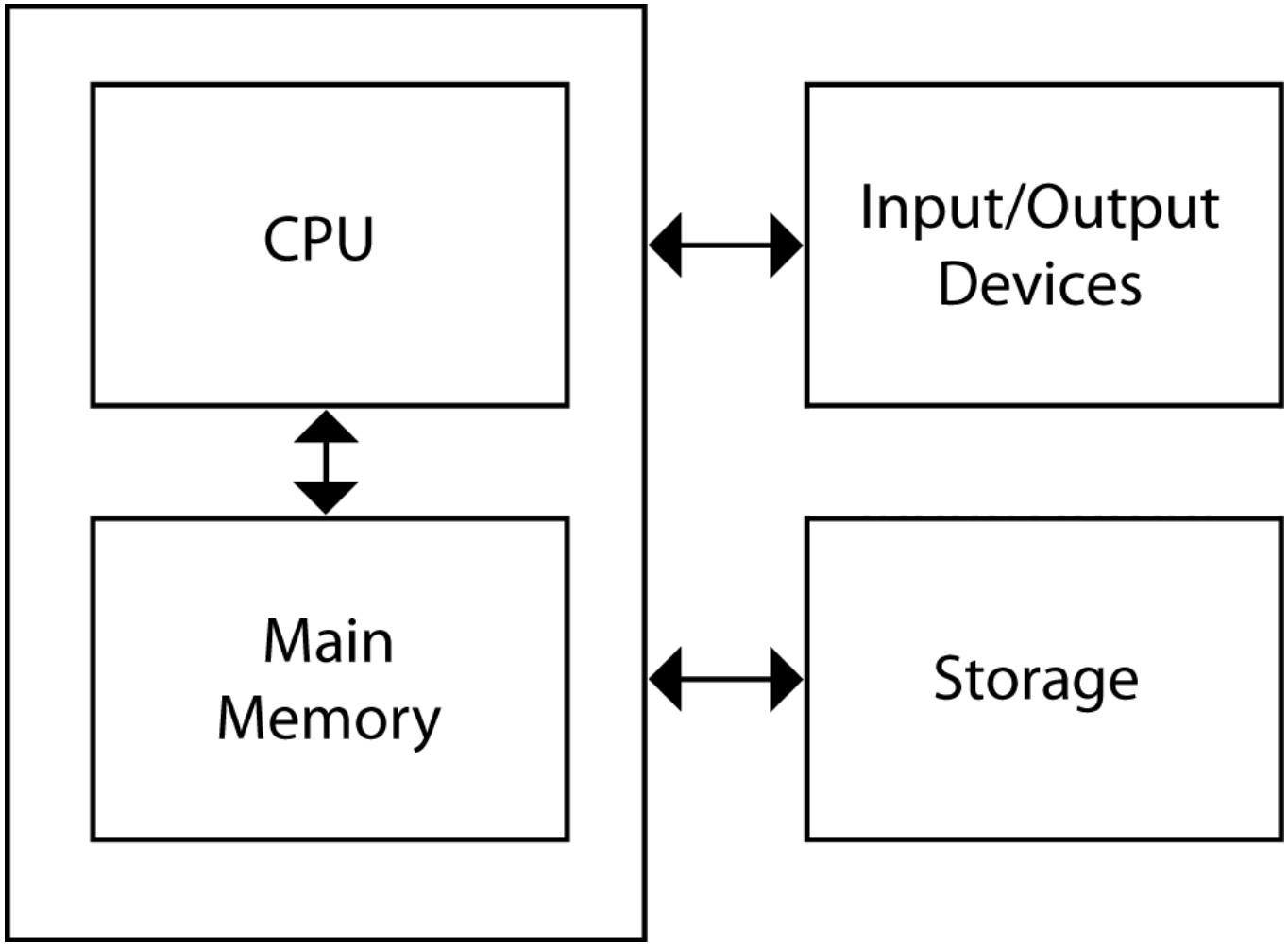
Lecture 02: Computers

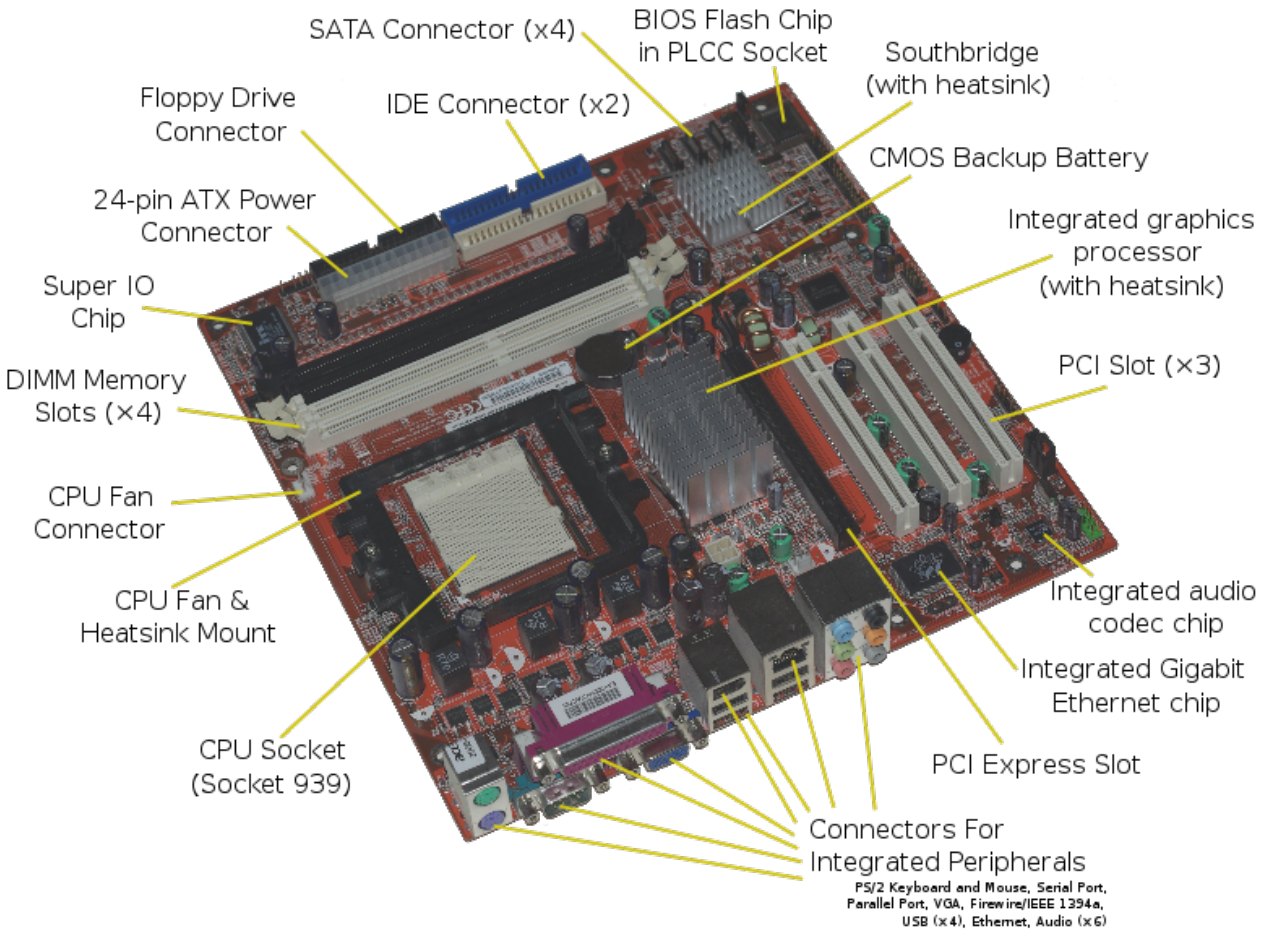
Christopher Simpkins
chris.simpkins@gatech.edu
http://www.cc.gatech.edu/∼simpkins/

Computers



Functional view of the major subsystems of a computer.

Computer Hardware



[1]

A computer motherboard housing the CPU, main memory, I/O controllers, and I/O ports.

[1] Source: http://en.wikipedia.org/wiki/File:Acer_E360_Socket_939_motherboard_by_Foxconn.svg

Central Processing Unit

The “brain” of the computer. In the von-Neumann architecture, the CPU fetches an instruction from memory, executes the instruction, possibly writes something to memory, then repeats the cycle. Core components of a microporocessor are arithmetic-logic unit (ALU) and control unit. CPU is implemented as a microprocessor

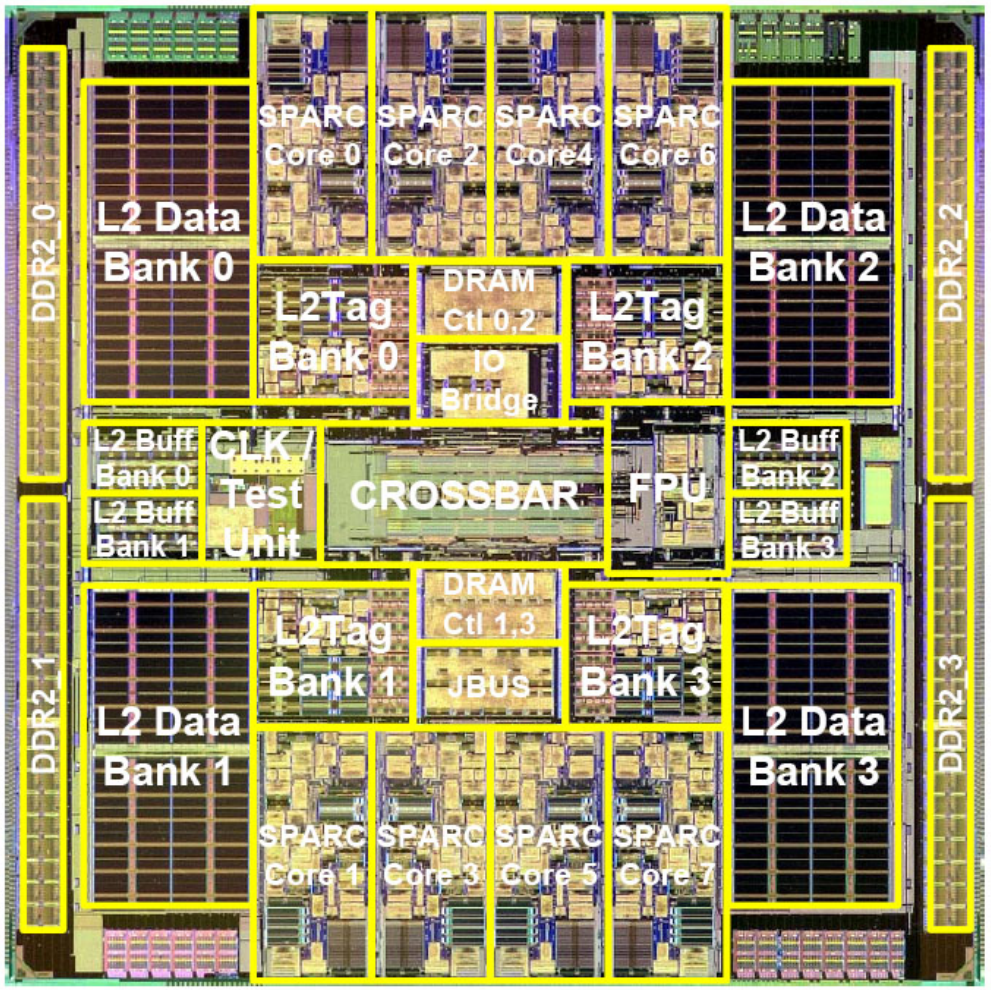
- RISC (Reduced Instruction Set Computing) processors have small, fast-executing instruction sets. All modern microprocessors are RISC
- CISC (Complex Instruction Set Computing) processors have large instruction sets and complicated addressing modes that slow all instructions down. Notably, the Intel x86 architecture was originally CISC but is now RISC

CPU Performance

How to compare CPUs?

- Clock speed, typically measured in GHz, indicates the number of instructions that can be executed in a second
- Modern pipelined processors are superscalar, meaning they can execute more than one instruction per clock cycle.
- Benchmarks such as FLOPS (floating-point operations persecond) can help, but hard to isolate non-CPU factors such as memory speed, cache performance, etc.
- Best benchmarks are suites of real-world application tests using wall-clock time - these tests compare computers, not necessarily isolating CPUs

Multicore Processors



a

- Trend is toward multi-core processors, which have multiple processing cores on a single chip. Each core is like a CPU, with an ALU, control unit, registers, and optionally dedicated Level 1 cache memory
- The Sun UltraSPARC T1 pictured here has 8 processing cores

[a] Source: http://jinsatoh.jp/ennui/ultrasparcT1_overlay_die.jpg

32-Bit and 64-Bit Processors

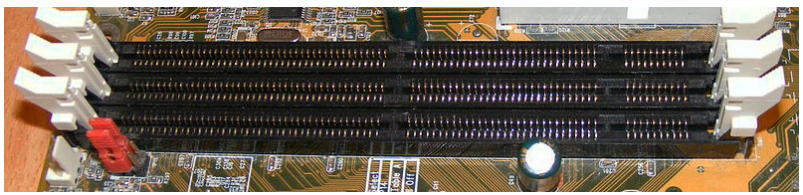
What does it mean to have a 32-bit or 64-bit processor?

- Processor registers hold operands, intermediate computation results, and values that represent memory locations
- Word size is the size of the processor registers. A 32-bit processor has 32-bit registers (4 bytes - a byte is 8 bits)
- Memory address busses are typically the same width as the word size, which makes memory access very fast
- Wider memory busses mean more addressable memory.
- 32-bit processors can typically address 4 GB of RAM (2^{32})
- 64-bit processors can typically address (in theory) 2^{64} = 18 EB (exabytes - $18x10^{18}$ bytes) of RAM
- Addressing more RAM means holding bigger data structures, like massive arrays, in memory

2

[2] Note that memory is byte-addressable, regardless of word size

Main Memory



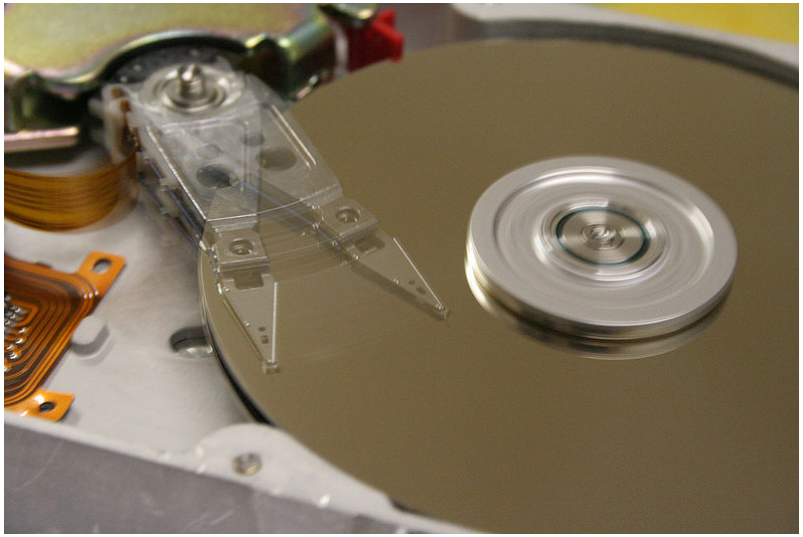
3 4

- Random access memory (RAM)
- Where running programs and data are stored
- Typically housed in dual-inline memory modules (DIMMs) that insert easily into slots, pictured to the left

[3] Source: http://en.wikipedia.org/wiki/File:DIMMs.jpg

[4] Source: http://en.wikipedia.org/wiki/File:3SDRAM-DIMMs.jpg

Storage



A spinning hard disk with the read-write heads moving over the platters. Data stored in magnetic particles on platters. “Head crashes” are possible, especially if HDD is jarred.

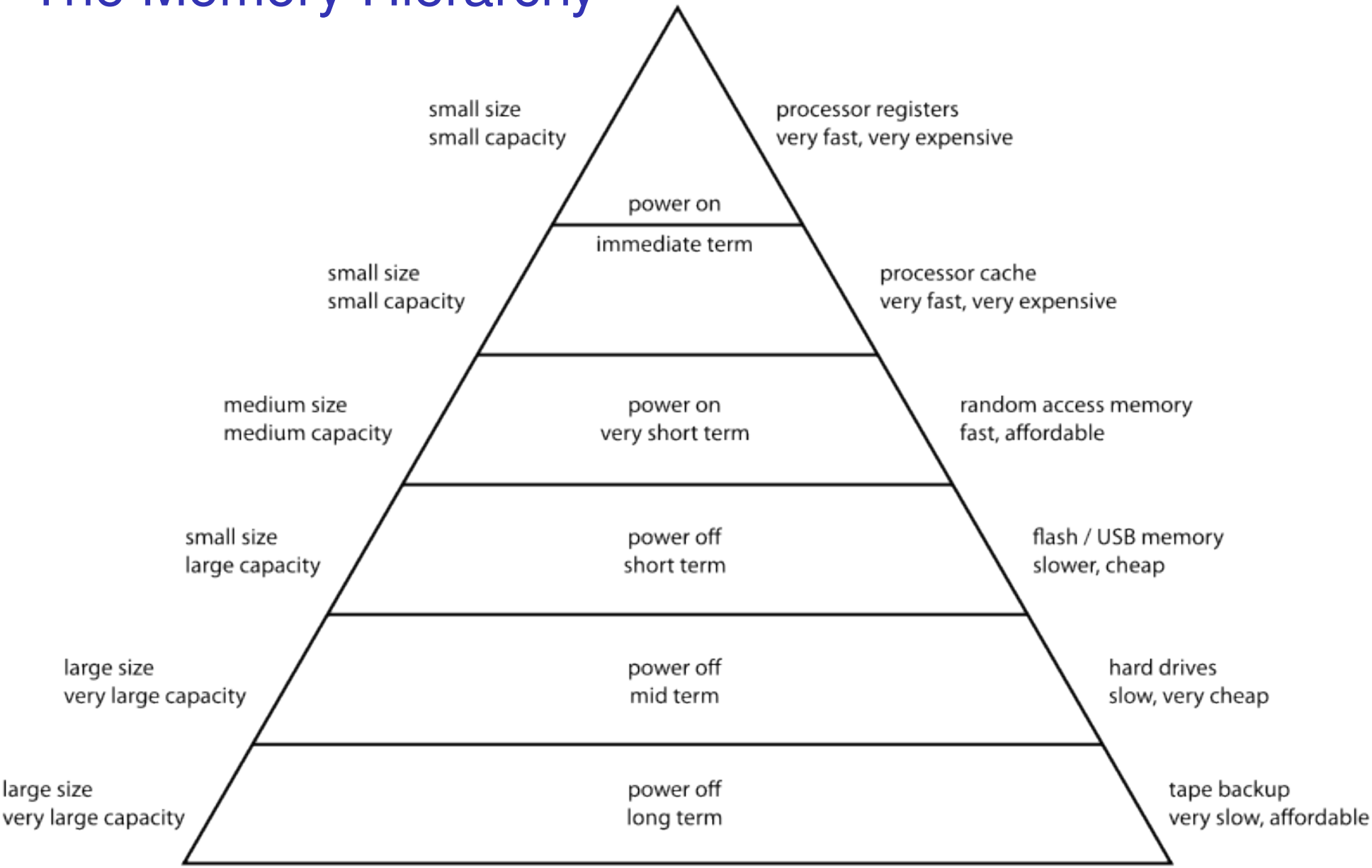
5

Where programs and data are stored when not in use.

- Hard disk drives (HDDs) most common. Good trade-off between cost and performance
- Solid-state drives (SDDs) becoming more common. Much faster than HDDs, more reliable (no moving parts), but much more expensive
- Tapes and optical drive slow and sometimes write-only. Used for long-term archiving

[5] Source: http://www.flickr.com/photos/alphasix/158829630/

The Memory Hierarchy



http://en.wikipedia.org/wiki/File:ComputerMemoryHierarchy.svg

Input/Output Devices

Input devices convert physical phenomena into electrical signals that are then converted to readable bits in the computer (see Sensors class material for more):

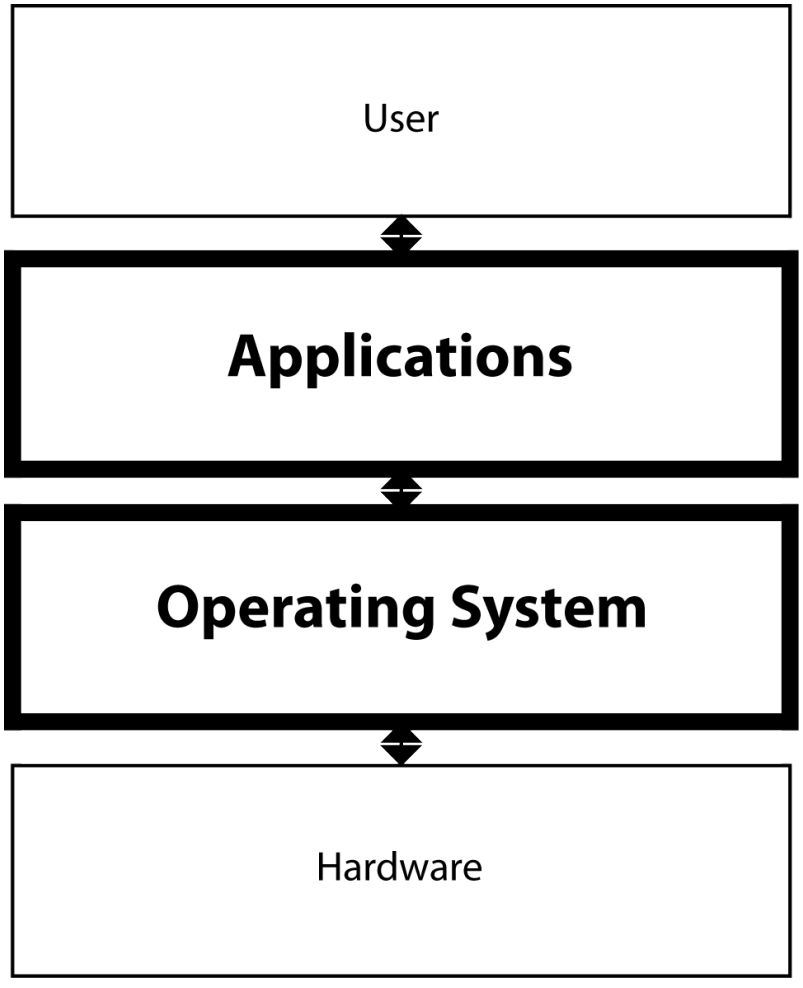
- Keyboard
- Camera
- Microphone
- Accelerometer
- Gyroscope

Output devices convert bits from the computer into forms perceptible to humans:

- Display monitor
- Speakers

Note that input/output (I/O) also refers to communication with anything external to the CPU and memory, for example, disk I/O or network I/O

Software



Two kinds of software:

- Application software: what the users interact with. Web browsers, email clients, spreadsheets, text editors, etc.
- System software: what the applications interact with. File systems, device drivers, network services, etc.

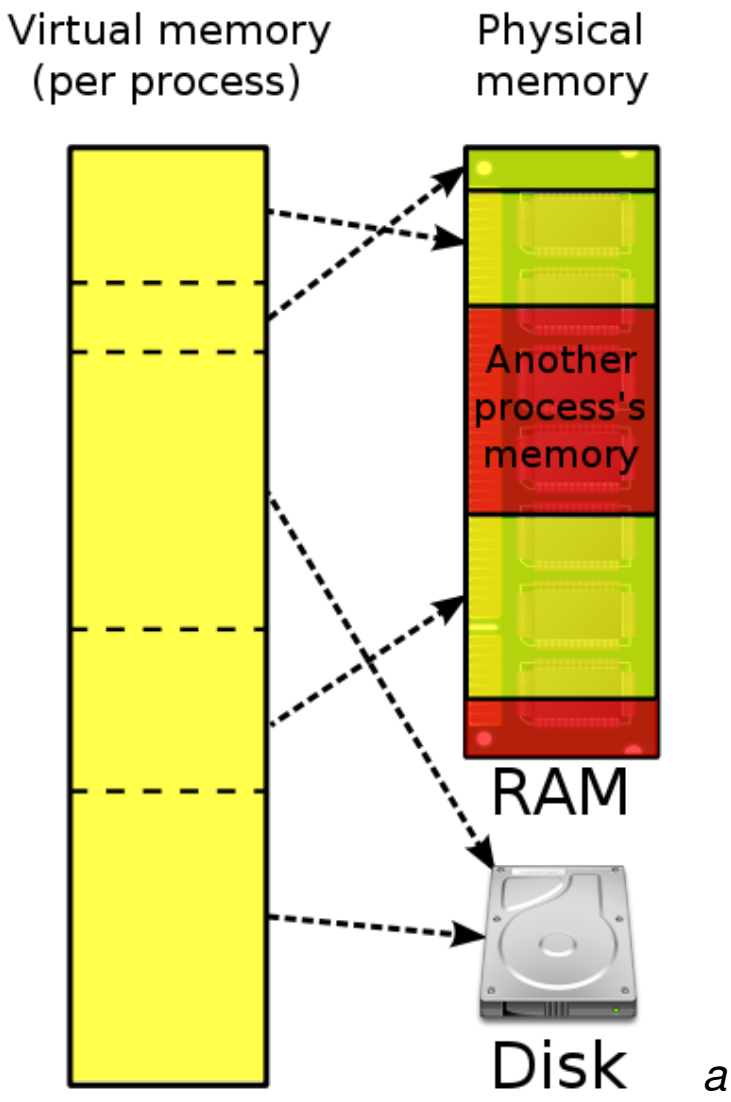
Each layer of the computer system abstracts the layer below

Operating Systems

Rough definition of operating system: the software that controls the hardware. Operating systems provide:

- a file system, so application programs can deal with files and directories instead of sectors and tracks;
- process scheduling, so multiple programs can run concurrently and share resources such as the processor;
- device drivers, so programs can interact with devices uniformly rather than having to know details of many different versions of devices; and
- virtual memory management, so that programs can be given a simple, large enough memory space to run even when physical RAM is limited.

Virtual Memory



[a] http://en.wikipedia.org/wiki/File:Virtual_memory.svg

Programs and their data rarely fit in physical RAM

- OS swaps pages not currently used to disk, reloads them when needed
- When a memory location that's on disk is accessed, the page that contains it is swapped back into physical RAM
- This process gives rise to the "clicking" you hear from your disk drive when you launch a program
- Disks are slow compared to RAM

Take-away lesson: if you're swapping a lot (lots of disk activity), adding physical RAM will improve your computer's performance

Consequences of the Memory Hierarchy

- Typical CPUs run at 2 GHz or more, RAM is typically 1 GHz, meaning that when the processor must access RAM, it waits for data for at least one instruction cycle
- RAM can be accessed on the order of nanoseconds (10^{-9})
- Hard disks typically accessed on the order of milliseconds (10^{-3})
- When programs access the disk, or the operating system's virtual memory manager performs paging operations, performance suffers
- Conclusion: large caches and lots of RAM mean higher performance

Applications

Now we're familiar with the hardware and software platform that applications run on.

- Hardware houses all the software and constrains performance characterstics by processor speeds and memory architectures
- Operating systems provide a unified view of the hardware and simplified execution environment for applications

Most of the remainder of this course will be primarily concerned with application software