Memory Hierarchy
& Data Locality

CSE 373
Data Structures & Algorithms
Ruth Anderson

Today's Outline

- **Admin**:
  - HW #5 due Thursday, June4th at 11:45pm

- **Memory Hierarchy and Locality**
- **B-trees**

5/29/2009   2

Why do we need to know about the memory hierarchy/locality?

- One of the assumptions that Big-Oh makes is that all operations take the same amount of time.
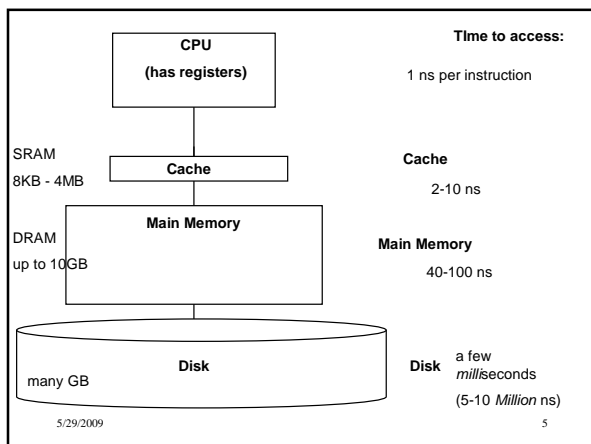- Is that really true?

5/29/2009   3

Definitions

Cycle – (for our purposes) the time it takes to execute a single simple instruction. (ex. Add 2 registers together)
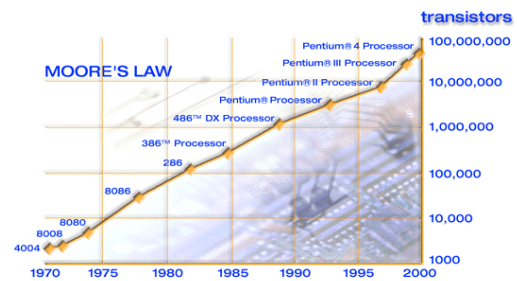
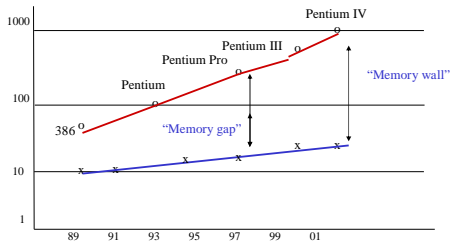**Memory Latency** – time it takes to access memory

5/29/2009   4

	Time to access:
CPU (has registers)	1 ns per instruction
SRAM 8KB - 4MB **Cache**	**Cache** 2-10 ns
DRAM up to 10GB **Main Memory**	**Main Memory** 40-100 ns
Disk many GB	**Disk** a few *milli*seconds (5-10 *Million* ns)

5/29/2009   5

Moore's Law

5/29/2009   6

Processor-Memory Performance Gap

- x86 CPU speed (100x over 10 years)

What can be done?

- **Goal**: Attempt to reduce the number of accesses to the slower levels.
- How?

Locality

Temporal Locality (locality in time) – If an item is referenced, it will tend to be referenced again soon.

Spatial Locality (locality in space) – If an item is referenced, items whose addresses are close by will tend to be referenced soon.

Caches

- Each level is a **sub-set** of the level below.

Cache Hit – address requested is in cache

Cache Miss – address requested is NOT in cache

Cache line size (chunk size) – the number of contiguous bytes that are moved into the cache at one time

---

Examples

```
x = a + 6;        x = a[0] + 6;
y = a + 5;        y = a[1] + 5;
z = 8 * a;        z = 8 * a[2];
```

Locality and Data Structures

- Which has (at least the potential for) better spatial locality, arrays or linked lists?