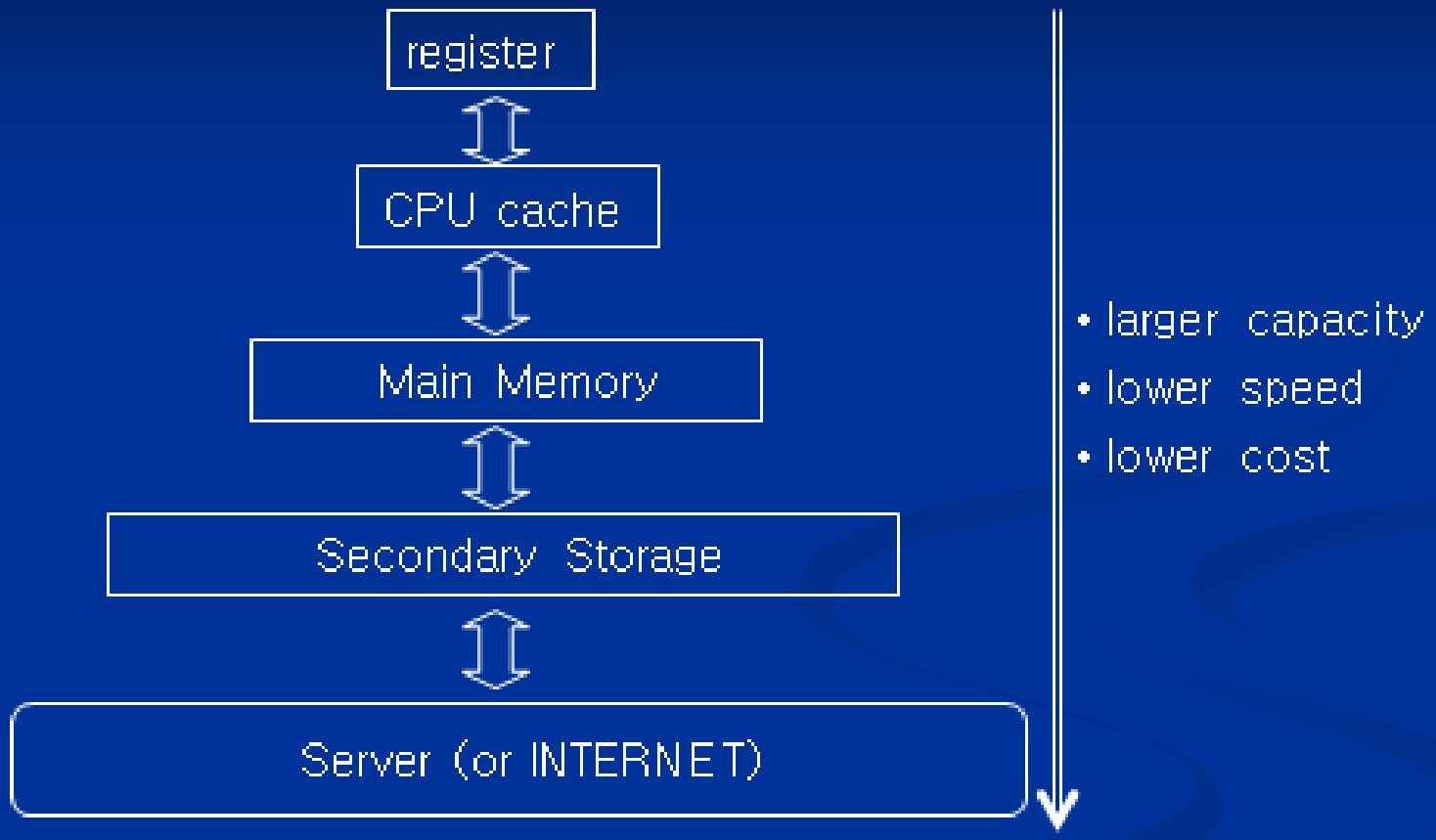
Virtual Memory

หน่วยความจำเสมือน

Memory



http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

What is…

- Virtual memory as an alternate set of memory addresses.
- Programs use these virtual addresses rather than real addresses to store instructions and data.
- When the program is actually executed, the virtual addresses are converted into real memory addresses.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

History

- virtual memory was developed in approximately 1959 – 1962, at the University of Manchester for the Atlas Computer, completed in 1962.

- In 1961, Burroughs released the B5000, the first commercial computer with virtual memory.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Object…

- When a computer is executing many programs at the same time, Virtual memory make the computer to share memory efficiently.

- Eliminate a restriction that a computer works in memory which is small and be limited.

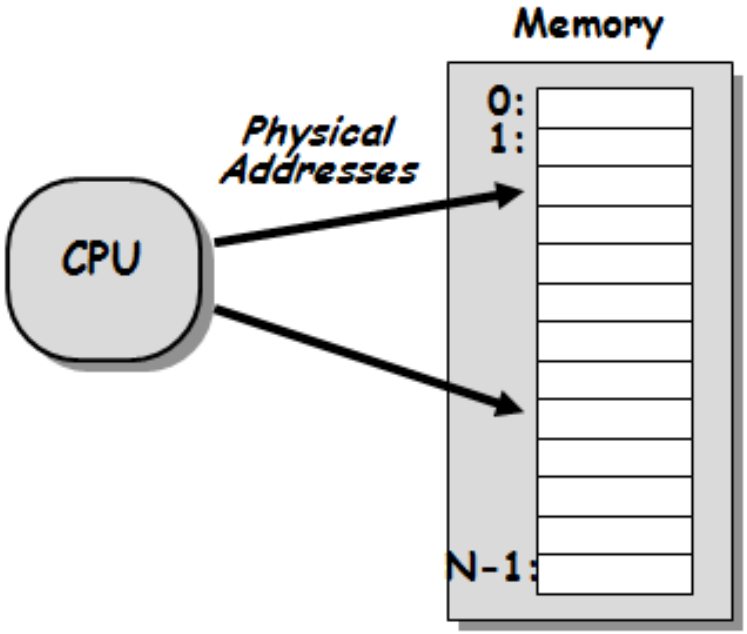
- When many programs is running at the same time, by distributing each suitable memory area to each program, VM protect programs to interfere each other in each memory area.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

A System with Physical Memory Only

Examples:

- most Cray machines, early PCs, nearly all embedded systems, etc.



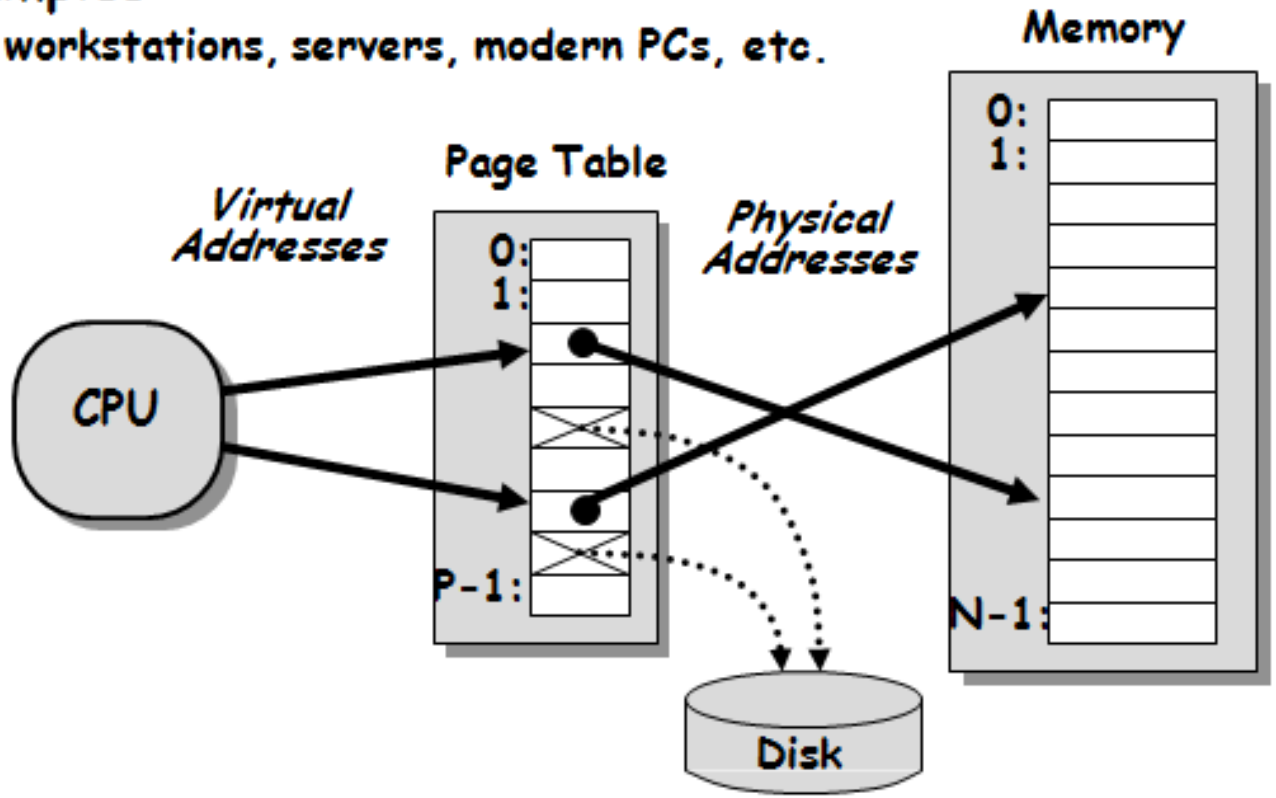
Addresses generated by the CPU point directly to bytes in physical memory

http://www.cs.cmu.edu/afs/cs/academic/class/15213-s00/lectures/class20.ppt

A System with Virtual Memory

Examples:

- workstations, servers, modern PCs, etc.



Address Translation: the hardware converts *virtual addresses* into *physical addresses* via an OS-managed lookup table (*page table*)

http://www.cs.cmu.edu/afs/cs/academic/class/15213-s00/lectures/class20.ppt

How does it work…

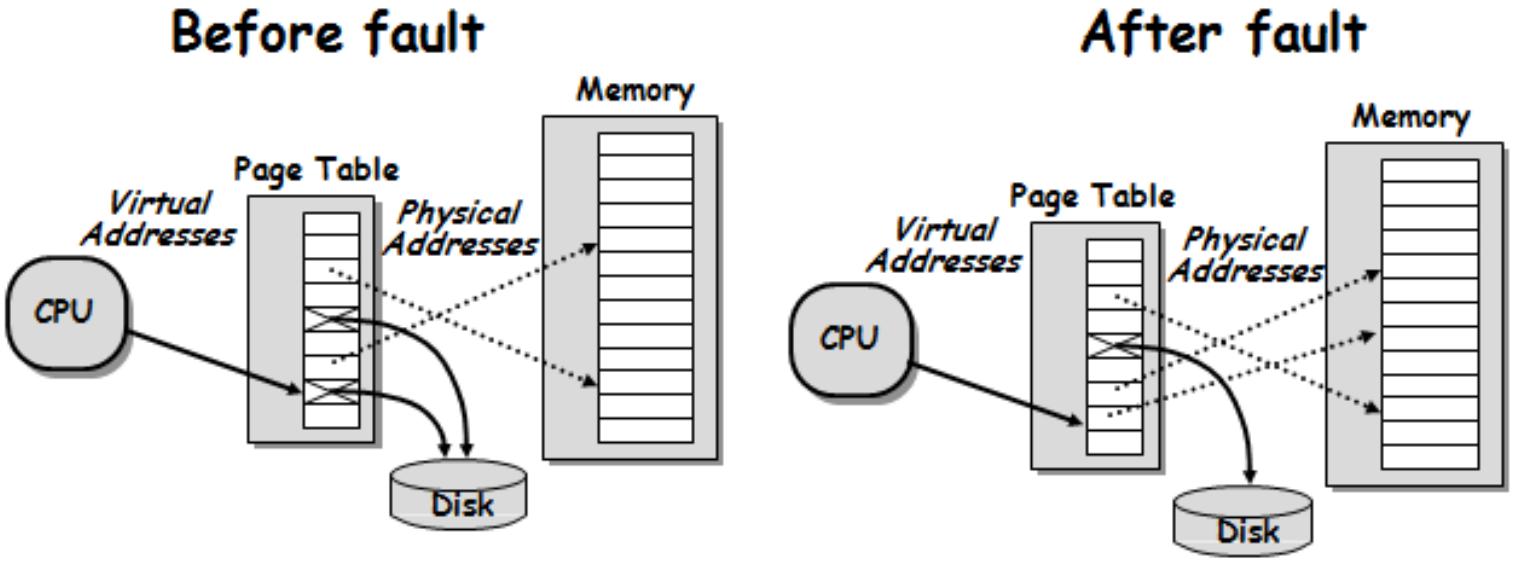
- To facilitate copying virtual memory into real memory, the operating system divides virtual memory into pages, each of which contains a fixed number of addresses.
- Each page is stored on a disk until it is needed.
- When the page is needed, the operating system copies it from disk to main memory, translating the virtual addresses into real addresses.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Page Faults (Similar to "Cache Misses")

What if an object is on disk rather than in memory?

- Page table entry indicates that the virtual address is not in memory
- An OS exception handler is invoked, moving data from disk into memory
  - current process suspends, others can resume
  - OS has full control over placement, etc.



http://www.cs.cmu.edu/afs/cs/academic/class/15213-s00/lectures/class20.ppt

Servicing a Page Fault

Processor Signals Controller

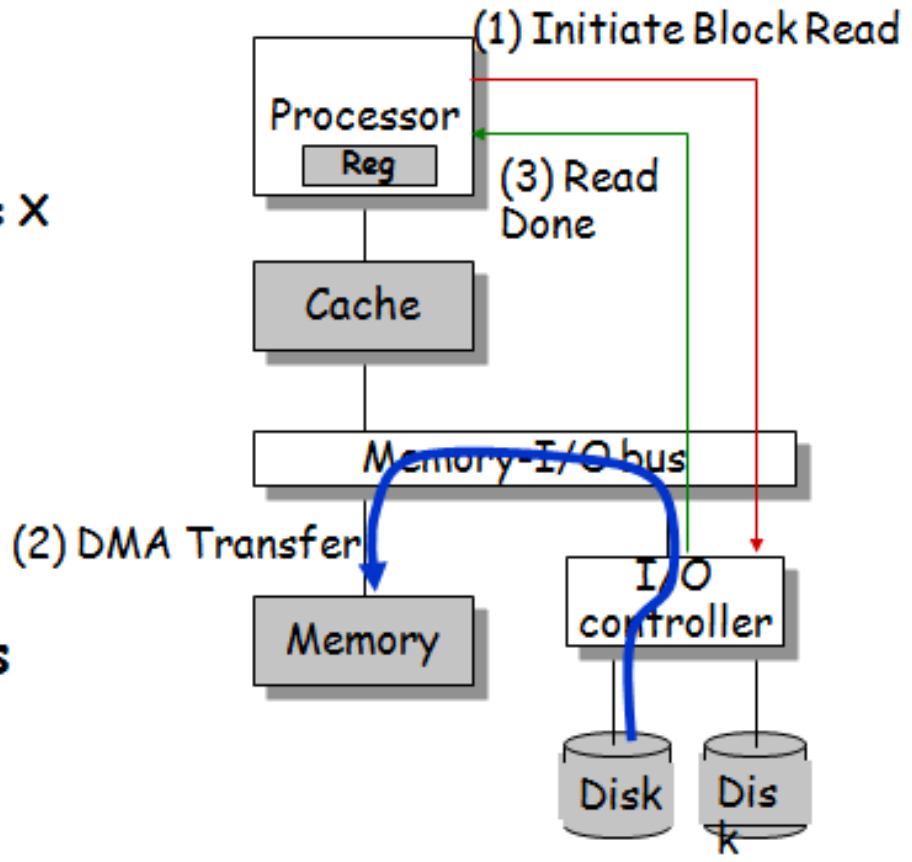
- Read block of length P starting at disk address X and store starting at memory address Y

Read Occurs

- Direct Memory Access (DMA)
- Under control of I/O controller

I/O Controller Signals Completion

- Interrupt processor
- OS resumes suspended process



http://www.cs.cmu.edu/afs/cs/academic/class/15213-s00/lectures/class20.ppt

MMU (Memory Management Unit)

- The hardware base that makes a virtual memory system possible.
- Allows software to reference physical memory by virtual addresses, quite often more than one.
- It accomplishes this through the use of page and page tables.
- Use a section of memory to translate virtual addresses into physical addresses via a series of table lookups.
- The software that handles the page fault is generally part of an operating system and the hardware that detects this situation.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Segmentation
VS
Paging

Segmentation……

- Segmentation involves the relocation of variable sized segments into the physical address space.
- Generally these segments are contiguous units, and are referred to in programs by their segment number and an offset to the requested data.
- Efficient segmentation relies on programs that are very thoughtfully written for their target system.
- Since segmentation relies on memory that is located in single large blocks, it is very possible that enough free space is available to load a new module, but can not be utilized.
- Segmentation may also suffer from internal fragmentation if segments are not variable-sized, where memory above the segment is not used by the program but is still “reserved” for it.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Paging……

- Paging provides a somewhat easier interface for programs, in that its operation tends to be more automatic and thus transparent.
- Each unit of transfer, referred to as a page, is of a fixed size and swapped by the virtual memory manager outside of the program's control.
- Instead of utilizing a segment/offset addressing approach, as seen in segmentation, paging uses a linear sequence of virtual addresses which are mapped to physical memory as necessary.
- Due to this addressing approach, a single program may refer to series of many non-contiguous segments.
- Although some internal fragmentation may still exist due to the fixed size of the pages, the approach virtually eliminates external fragmentation.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Paging……(cont'd)

- A technique used by virtual memory operating systems to help ensure that the data you need is available as quickly as possible.
- The operating system copies a certain number of pages from your storage device to main memory.
- When a program needs a page that is not in maim memory, the operating system copies the required page into memory and copies another page back to the disk.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Page fault

- An interrupt to the software raised by the hardware when a program accesses a page that is not mapped in physical memory.

- when a program accesses a memory location in its memory and the page corresponding to that memory is not loaded
- when a program accesses a memory location in its memory and the program does not have privileges to access the page corresponding to that memory.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Paging replacement algorithms

- OPT(MIN) : eliminate the page that be not expected to be used.
- FIFO(first input/first output) : rather than choosing the victim page at random, the oldest page is the first to be removed.
- LRU(Least Recently used) : move out the page that is the least rarely used.
- LFU(Least Frequently used) : move out the page that is not used often in the past.

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt

Summary…

- **Virtual memory** is a common part of most operating systems on computers.
- It has become so common because it provides a big benefit for users at a very low cost.
- benefits of executing a program that is only partially in memory.
- program is no longer constrained by the amount of physical memory.

  ⇒ user would be able to write programs for an extremely large virtual address space.
- more programs could be run at the same time

  ⇒ increase CPU utilization and throughput.
- less I/O would be needed to load or swap each user program

  ⇒ run faster

http://www.eng.auburn.edu/~agrawvd/COURSE/E6200_06/STUDENT_TALKS/NOV11%20virtual%20memory.ppt