



College of Information Studies

University of Maryland Hornbake Library Building College Park, MD 20742-4345

Physical Infrastructure

Week 1

INFM 603

Agenda

- Computers
- The Internet
- The Web
- About the course

A COMPUTER WANTED.

WASHINGTON, May 1.—A civil service examination will be held May 18 in Washington, and, if necessary, in other cities, to secure eligibles for the position of computer in the Nautical Almanac Office, where two vacancies exist—one at \$1,000, the other at \$1,400..

The examination will include the subjects of algebra, geometry, trigonometry, and astronomy. Application blanks may be obtained of the United States Civil Service Commission.

The New York Times

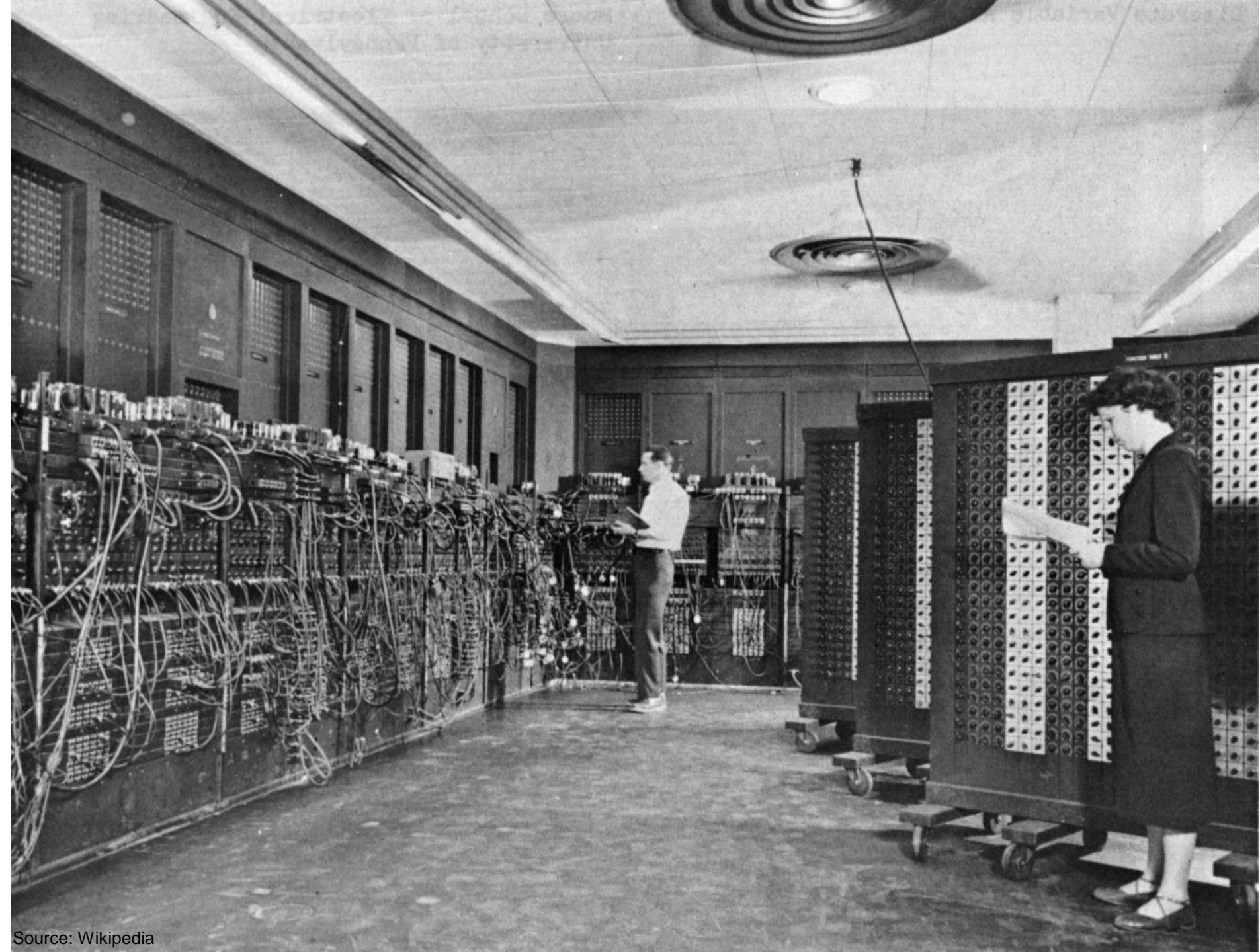
Published: May 2, 1892

Copyright © The New York Times

A Very Brief History of Computing

- Hardware
 - Mechanical: essentially a big adding machine
 - Analog: designed for calculus, limited accuracy
 - Digital: early machines filled a room
 - Microchips: designed for missile guidance
- Software
 - Numeric: computing gun angles
 - Symbolic: code-breaking











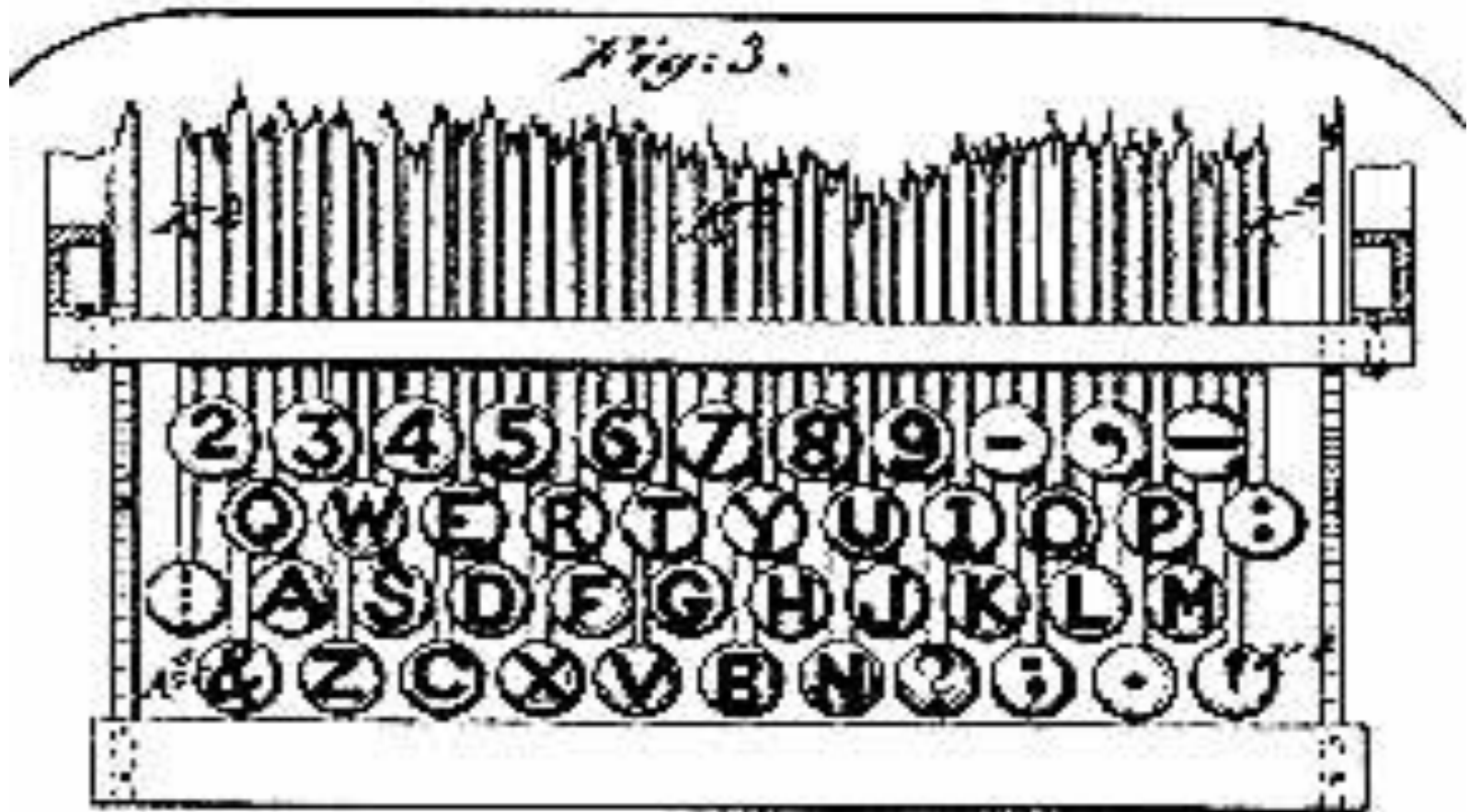


Source: Wikipedia

Input Devices

- Text
 - **Keyboard**, optical character recognition
 - Speech recognition, handwriting recognition
- Direct manipulation
 - 2-D: **mouse**, trackball, touchpad, touchscreen
 - 3-D: wand, data glove
- Remote sensing
 - Camera, speaker ID, head tracker, eye tracker

Input Example: QWERTY Keyboard



From <http://home.earthlink.net/~dcrehr/whyqwerty.html>

Dvorak Keyboard



Dvorak Keyboard Layout

From <http://www.mwbrooks.com/dvorak/>

Binary Data Representation

Example: American Standard Code for Information Interchange (ASCII)

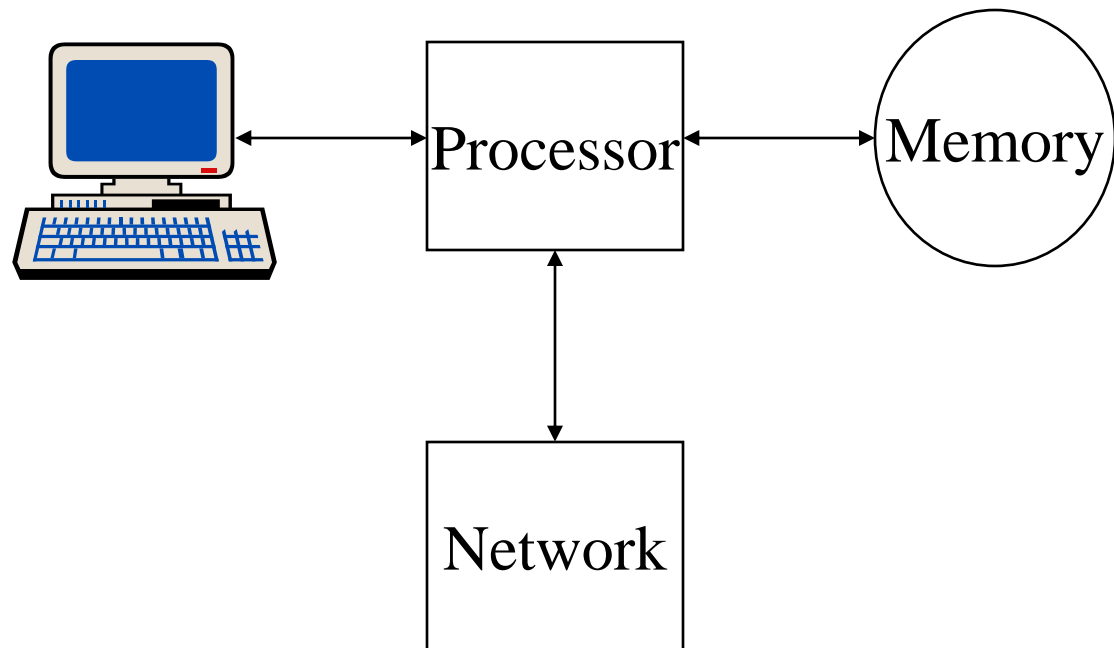
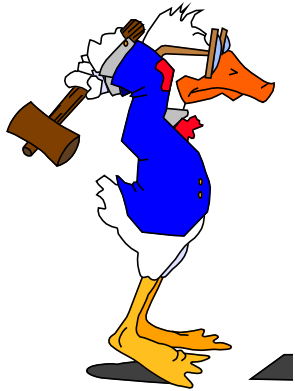
01000001 = A	01100001 = a
01000010 = B	01100010 = b
01000011 = C	01100011 = c
01000100 = D	01100100 = d
01000101 = E	01100101 = e
01000110 = F	01100110 = f
01000111 = G	01100111 = g
01001000 = H	01101000 = h
01001001 = I	01101001 = i
01001010 = J	01101010 = j
01001011 = K	01101011 = k
01001100 = L	01101100 = l
01001101 = M	01101101 = m
01001110 = N	01101110 = n
01001111 = O	01101111 = o
01010000 = P	01110000 = p
01010001 = Q	01110001 = q
...	...

Output Devices

- Visual
 - Screen, projector, head-mounted display, CAVE
- Acoustic
 - Speakers, headphones
- Physical
 - Tactile (vibrotactile, pneumatic, piezoelectric)
 - Force feedback (pen, joystick, exoskeleton)
 - Thermal
- Vestibular (motion-based simulators)
- Locomotive (treadmill, stationary bicycle)
- Olfactory



The Big Picture



Hardware Processing Cycle

- Input comes from somewhere
 - Keyboard, mouse, microphone, camera, ...
- The system does something with it
 - Processor, memory, software, network, ...
- Output goes somewhere
 - Monitor, speaker, robot controls, ...

Computer Hardware

- Central Processing Unit (CPU)
 - Intel Xeon, Motorola Power PC, ...
- Communications “Bus”
 - FSB, PCI, ISA, USB, Firewire, ...
- Storage devices
 - Cache, RAM, hard drive, floppy disk, ...
- External communications
 - Modem, Ethernet, GPRS, 802.11, ...



What's that?

Units of Frequency

Unit	Abbreviation	Cycles per second
hertz	Hz	1
kilohertz	KHz	$10^3 = 1,000$
megahertz	MHz	$10^6 = 1,000,000$
gigahertz	GHz	$10^9 = 1,000,000,000$

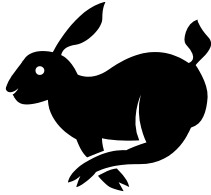
Units of Time

Unit	Abbreviation	Duration (seconds)
second	sec/s	1
millisecond	ms	$10^{-3} = 1/1,000$
microsecond	μs	$10^{-6} = 1/1,000,000$
nanosecond	ns	$10^{-9} = 1/1,000,000,000$
picosecond	ps	$10^{-12} = 1/1,000,000,000,000$
femtosecond	fs	$10^{-15} = 1/1,000,000,000,000,000$

The Storage Hierarchy

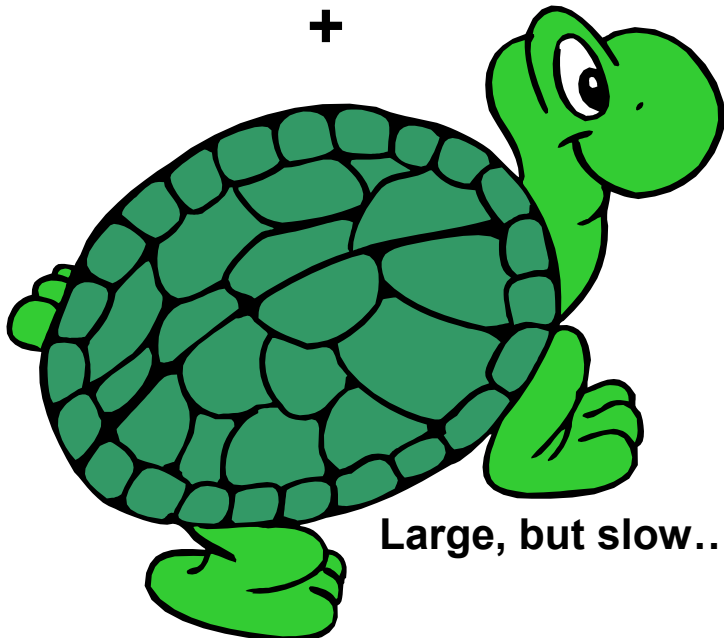
- Speed, cost, and size:
 - You can easily get any 2, but not all 3
- Fast memory is expensive
 - So large memory is slow!
 - But fast access to large memories is needed
- Solution:
 - Keep what you need often in small (fast) places
 - Keep the rest in large (slow) places
 - Get things to the fast place before you need them

Best of Both Worlds



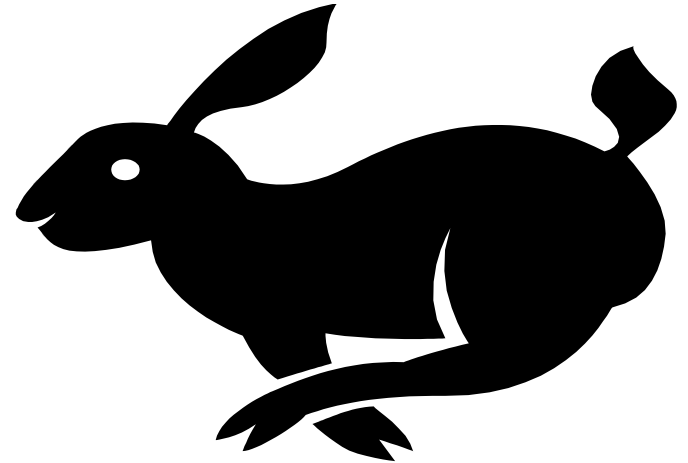
Small, but fast...

+



Large, but slow...

=



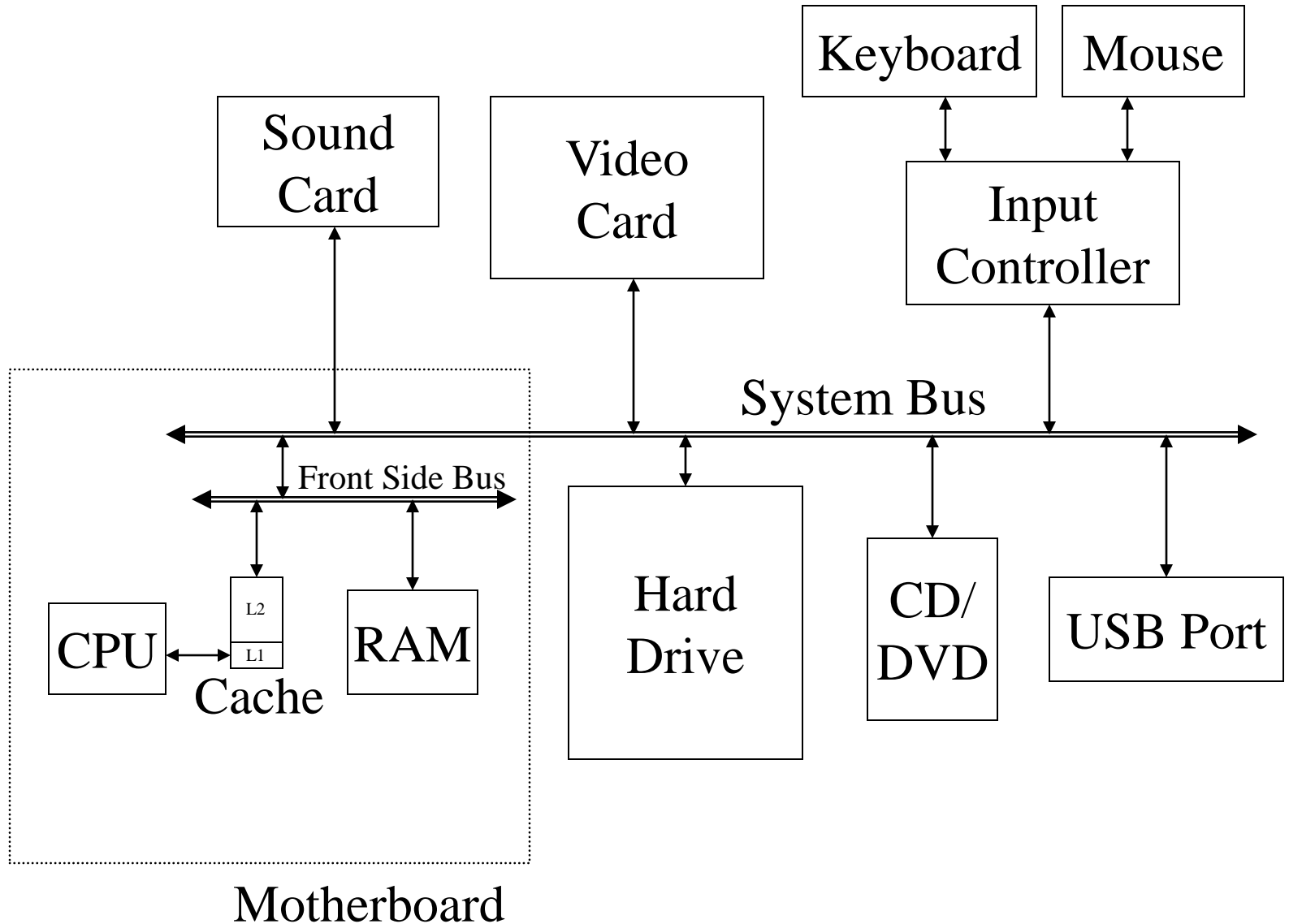
Is Large and seems fast

Think about your bookshelf and the library...

Locality

- **Spatial locality:**
 - If the system fetched x , it is likely to fetch data located near x
- **Temporal locality:**
 - If the system fetched x , it is likely to fetch x again

System Architecture



Everything is Relative

- The CPU is the fastest part of a computer
 - 3 GHz Core 2 Duo = 6,000 MIPS
 - 3 operations per processor every nanosecond
- Cache memory is fast enough to keep up
 - 128 kB L1 cache on chip (dedicated, CPU speed)
 - 4 MB L2 cache on chip (shared, CPU speed)
- RAM is larger, but slower
 - 1 GB or more, ~6 ns

Units of Size

Unit	Abbreviation	Size (bytes)
bit	b	1/8
byte	B	1
kilobyte	KB	$2^{10} = 1024$
megabyte	MB	$2^{20} = 1,048,576$
gigabyte	GB	$2^{30} = 1,073,741,824$
terabyte	TB	$2^{40} = 1,099,511,627,776$
petabyte	PB	$2^{50} = 1,125,899,906,842,624$

The Storage Hierarchy

Type	Speed	Size	Cost
Registers	~300 ps	256 B	Very expensive
Cache	~1 ns	4 MB	Expensive
RAM	~10 ns	1 GB	Cheap
Hard drive	~10 ms	100 GB	Very cheap

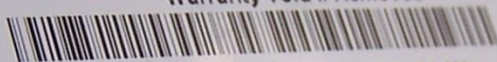
“Solid-State” Memory

- ROM
 - Does not require power to retain content
 - Used for “Basic Input/Output System” (BIOS)
- Cache (Fast low-power “Static” RAM)
 - Level 1 (L1) cache: small, single-purpose
 - Level 2 (L2) cache: larger, shared
- (“Dynamic”) RAM (Slower, power hungry)
 - Reached over the “Front-Side Bus” (FSB)
- Flash memory (fast read, slow write EEPROM)
 - Reached over USB bus or SD socket
 - Used in memory sticks (“non-volatile” storage)



elixir

M2U51264DS8HC3G-5T
512MB DDR-400MHz-CL3
PC3200U-30331
Warranty Void If Removed



0542.MN05A1107A.5YM0410N08.XX.CN

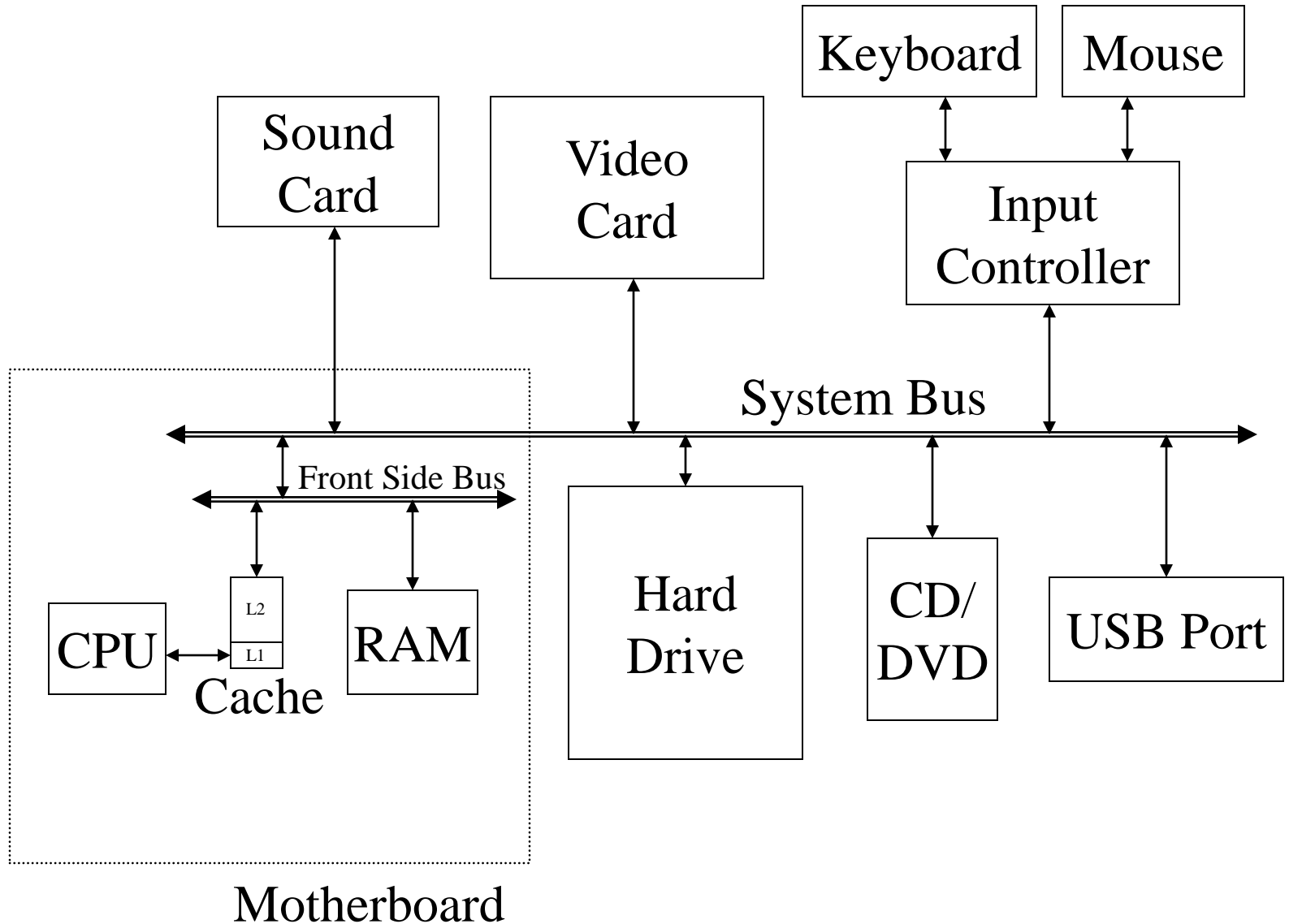
elixir

M2U51264DS8HC3G-5T
512MB DDR-400MHz-CL3
PC3200U-30331
Warranty Void If Removed



0542.MN05A1107A.5YM0410N08.XX.CN

System Architecture



“Rotating” Memory

- Fixed magnetic disk (“hard drive”)
 - May be partitioned into multiple volumes
 - In Windows, referred to as C:, D:, E:, ...
 - In Unix, referred to as /software, /homes, /mail, ...
- Removable magnetic disk
 - Floppy disk, zip drives, ...
- Removal optical disk
 - CDROM, DVD, CD-R, CD-RW, DVD+RW, ...

How Disks Work

Step 1:

The circuit board controls the movement of the head actuator and a small motor.

Step 2:

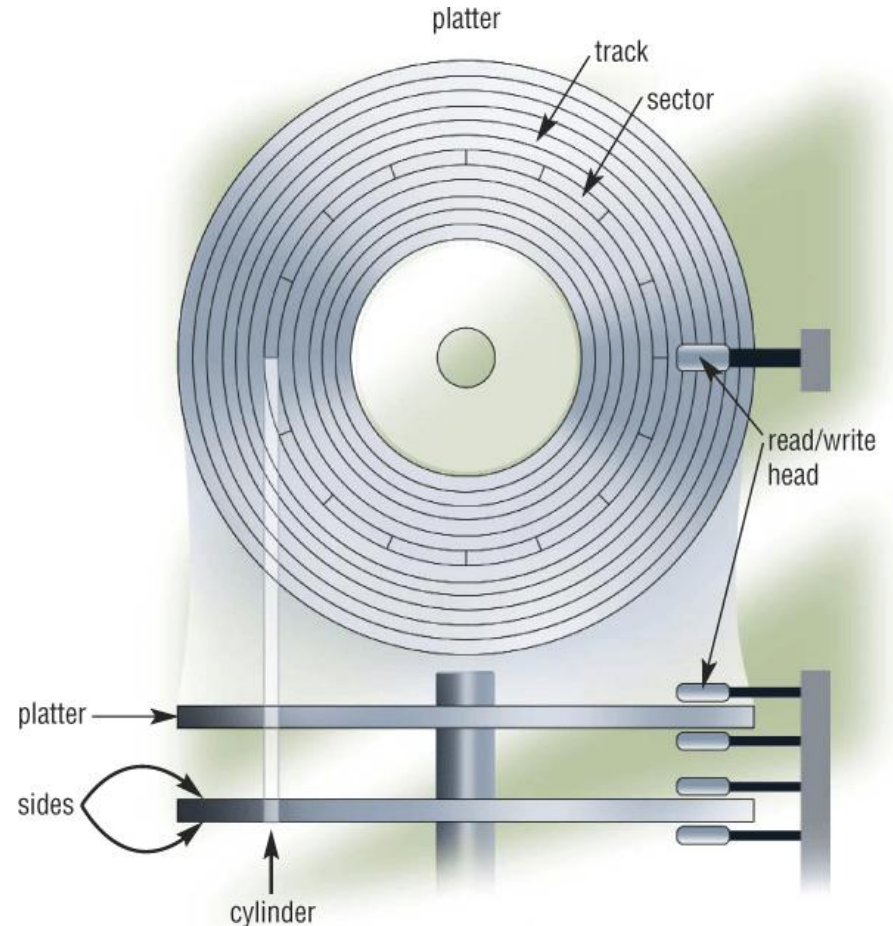
A small motor spins the platters while the computer is running.

Step 3:

When software requests a disk access, the read/write heads determine the current or new location of the data.

Step 4:

The head actuator positions the read/write head arms over the correct location on the platters to read or write data.



RAID-5

- Disks can fail in two ways:
 - Bad sectors (data sectors, directory sectors)
 - Mechanical failure
- RAID-5 arrays “stripe” blocks across disks
 - “Parallel” data transfer is faster than “serial”
 - ~30% “parity” allows reconstruction if one disk fails

transistors

MOORE'S LAW

10,000,000,000

1,000,000,000

100,000,000

10,000,000

1,000,000

100,000

10,000

1,000

1970 1975 1980 1985 1990 1995 2000 2005 2010

Dual-Core Intel® Itanium® 2 Processor

Intel® Itanium® 2 Processor
Intel® Itanium® Processor

Intel® Pentium® 4 Processor
Intel® Pentium® III Processor

Intel® Pentium® II Processor

Intel® Pentium® Processor
Intel 486™ Processor

Intel 386™ Processor

286

8086

8080

8008

4004

Moore's Law

- Processing speed doubles every 18 months
 - Faster CPU, longer words, larger cache, more cores
- Cost/bit for RAM drops 50% every 12 months
 - Less need for “virtual memory”
- Cost/bit for disk drops 50% every 12 months
 - But transfer rates don't improve much




More cores!

Agenda

- Computers
- The Internet
- The Web
- About the course

Network

- Computers and devices connected via
 - Communication devices
 - Transmission media
- 



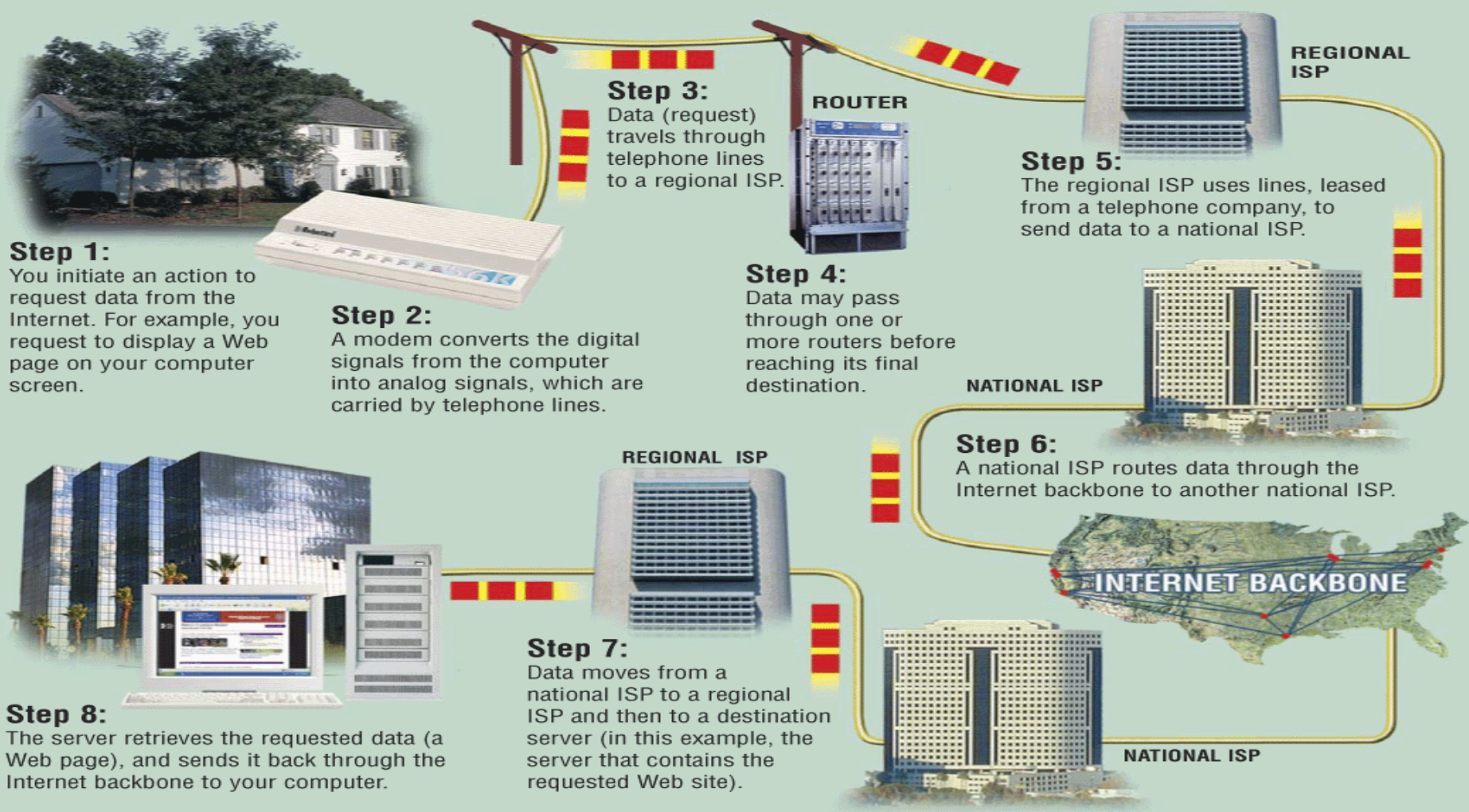
Packet vs. Circuit Networks

- Telephone system (“circuit-switched”)
 - Fixed connection between caller and called
 - High network load results in busy signals
- Internet (“packet-switched”)
 - Each transmission is routed separately
 - High network load results in long delays

Packet Switching

- Break long messages into short “packets”
 - Keeps one user from hogging a line
- Route each packet separately
 - Number them for easy reconstruction
- Request retransmission for lost packets
 - Unless the first packet is lost!

Overview



Networks of Networks

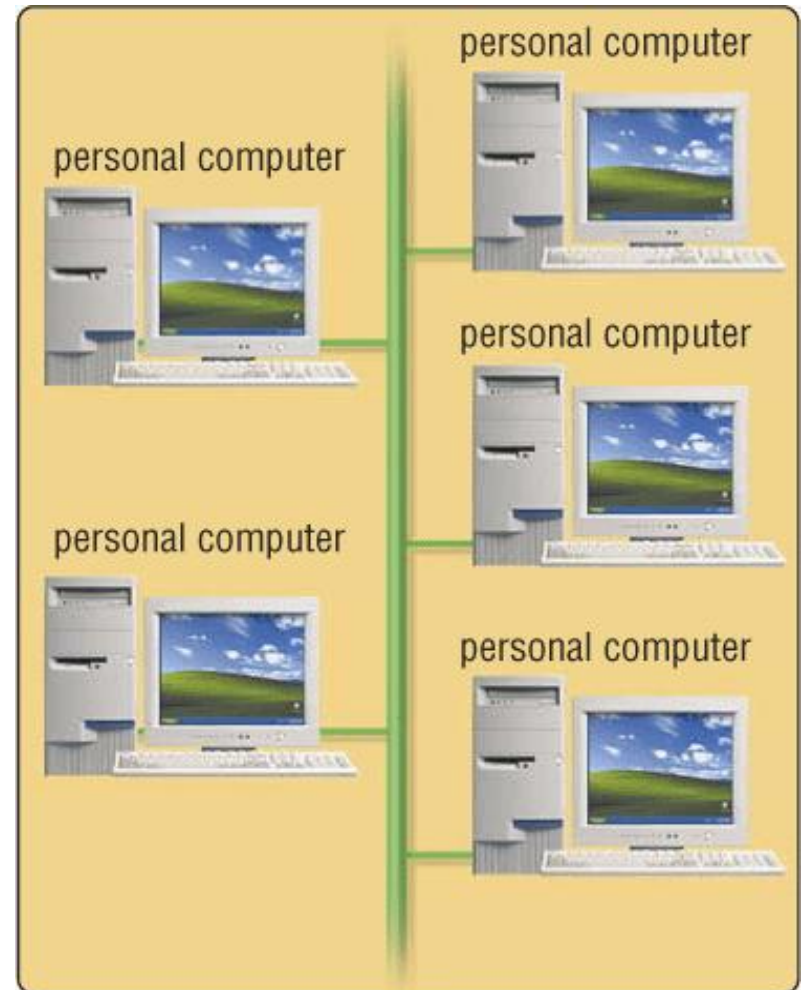
- Local Area Networks (LAN)
 - Connections within a room, or perhaps a building
- Wide Area Networks (WAN)
 - Provide connections between LANs
- Internet
 - Collection of WANs across multiple organizations

Local Area Networks

- Within a campus or an office complex
 - Short-distance lines are **fast** and **cheap**
 - Fast communications makes routing simple
- Ethernet is a common LAN technology
 - All computers are connected to the same cable
 - Ordinary phone lines can carry 10 Mb/sec
 - 100 Mb/s connections require special cables
 - 1 Gb/s connections require special switches
 - Every host broadcasts everything to all others
 - Collisions limit throughput to about 50% utilization

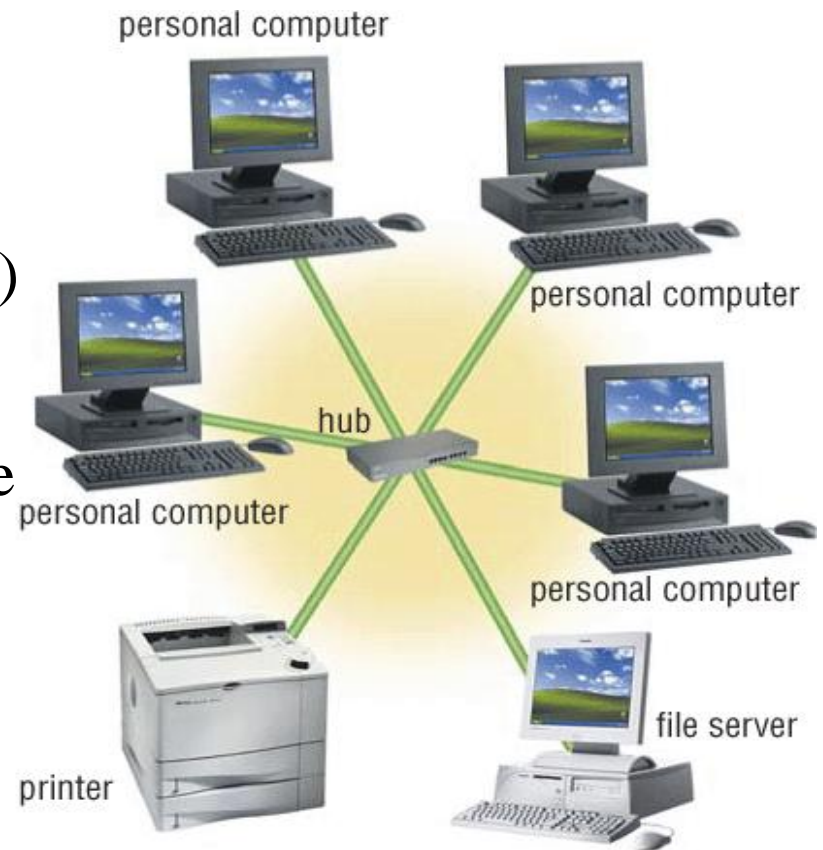
Shared Network

- All attach to the same cable
 - Ethernet and “cable modems”
- Transmit anytime
 - Collision detection
 - Automatic retransmission
- Inexpensive and flexible
 - Easy to add new machines
 - Robust to computer failure
- Practical for short distances
 - Half the bandwidth is wasted



Switched (“Star”) Network

- All attach directly to a hub
 - Switched Ethernet
 - Digital Subscriber Lines (DSL)
- Higher cost
 - Line from hub to each machine
 - Hub must handle every packet
 - Hub requires backup power
- Much higher bandwidth
 - No sharing, no collisions
 - Allows disks to be centralized



Wireless Networks

- Radio-based Ethernet
 - Effective for a few rooms within buildings
- “Access Point” gateways to wired networks
 - Available throughout most of the Maryland campus
 - Commercial providers offer “hot spots” in airports, etc.
- “WiFi WLAN” is available in several speeds
 - IEEE 802.11b: 10Mb/s (good enough for most uses)
 - IEEE 802.11g: 54Mb/s (required for wireless video)
 - IEEE 802.11n: 248Mb/s (and longer range)
- Computer-to-computer networks are also possible
 - “Bluetooth” is the most common (very short range)

Wide Area Networks

- Campus, regional, national, or global scale
- Expensive communications must be used well
 - Limiting to two hosts allows 100% utilization
- Routing is complex with point-to-point circuits
 - Which path is shortest? Which is least busy? ...

Types of Digital Channels

- “Backbone”
 - Microwave
 - Satellite
 - Fiber
- “Last mile” wired
 - Telephone modem
 - ADSL
 - Cable modem
 - Fiber
- “Last mile” wireless
 - Wi-Fi (IEEE 802.11)
 - GSM

Thinking About Speed

- Two parts of moving data from here to there:
 - Getting the first bit there
 - Getting everything there
- Fundamentally, there's no difference:
 - Moving data from the processor to RAM
 - Saving a file to disk
 - Downloading music from a server in China

Some Definitions

- **Latency**

- The amount of **time** it takes data to travel from source to destination

- **Bandwidth**

- The amount of data that can be transmitted in a fixed amount of **time**

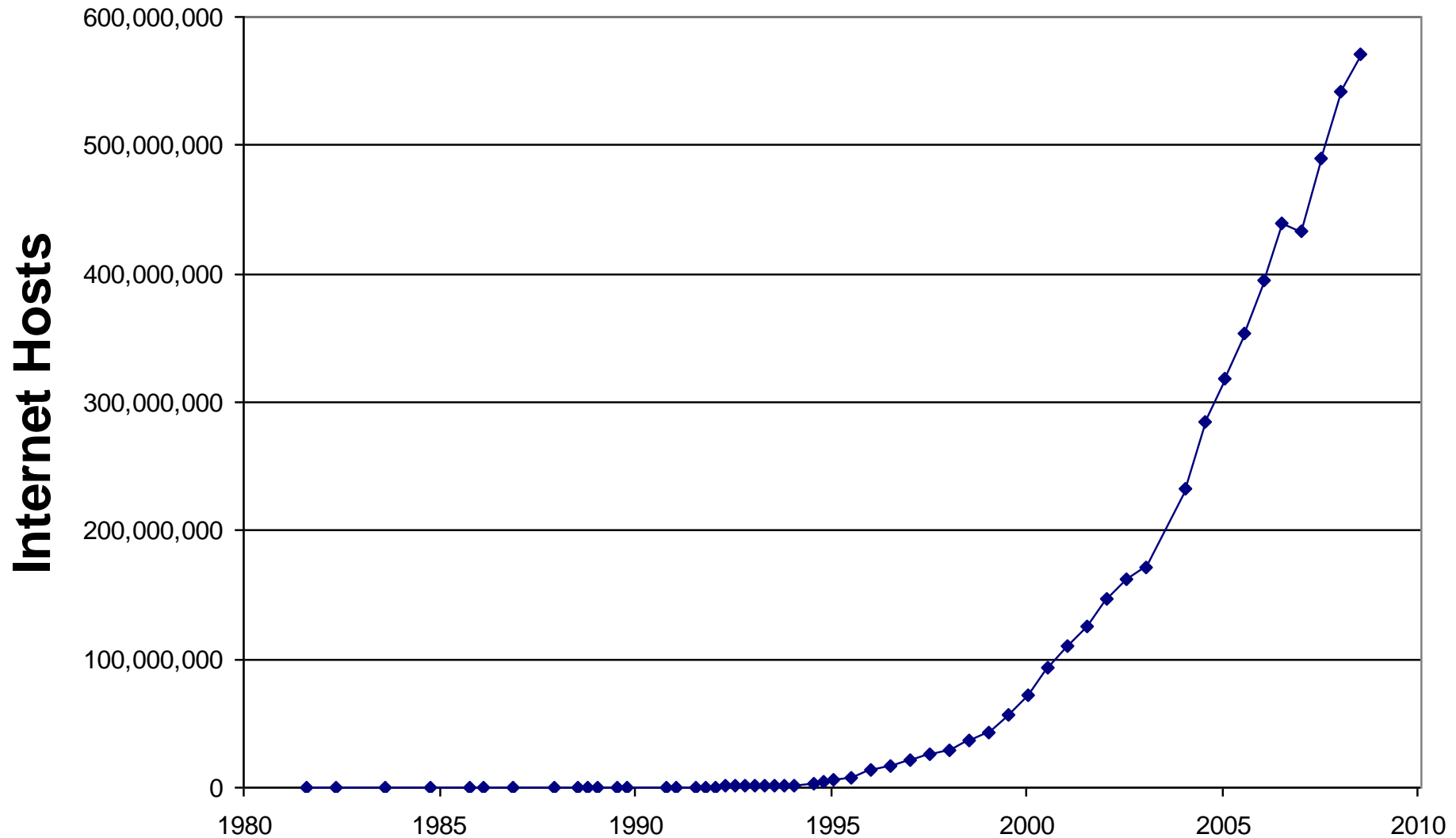
The Internet

- Global collection of public “IP” networks
 - Private networks are often called “intranets”
- Independent
 - Each organization maintains its own network
- Cooperating
 - Internet Protocol (IP) address blocks
 - Domain names
 - World-Wide Web Consortium (W3C)
 - Computer Emergency Response Team (CERT)

A Short History of the Internet

- 1969: Origins in government research
 - Advanced Research Projects Agency (ARPAnet)
 - Key standards: UDP, TCP, DNS
- 1983: Design adopted by other agencies
 - Created a need for inter-network connections
 - Key standards: IP
- 1991: World-Wide Web added point-and-click
 - Now 571 million Internet “hosts” (August 2008)
 - Key standards: HTTP, URL, HTML, XML

What Changed in 1994?



Types of Internet “Nodes”

- Hosts
 - Computers that use the network to do something
- Routers
 - Specialized computers that route packets
- Gateway
 - Routers that connect two networks
- Firewall
 - Gateways that pass packets selectively

IP Address

- Every host (and every router) is identified by an “Internet Protocol” (IP) address
- 32 bit number, divided into four “octets”

128.8.11.33

216.239.39.99

199.181.132.250

Example: point your browser at “<http://66.249.93.99/>”

An Internet Protocol (IP) Address

IP address:

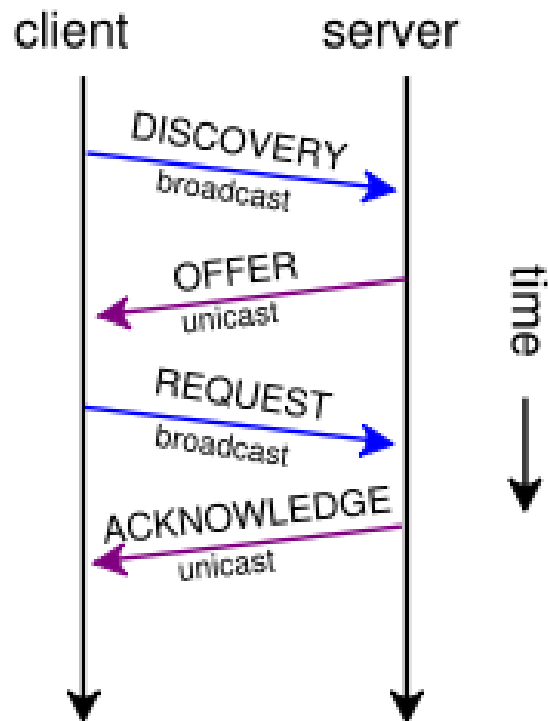
216.183.103.150

Identifies a LAN

Identifies a specific computer

Dynamic IP Addresses

- Dynamic Host Configuration Protocol (DHCP)



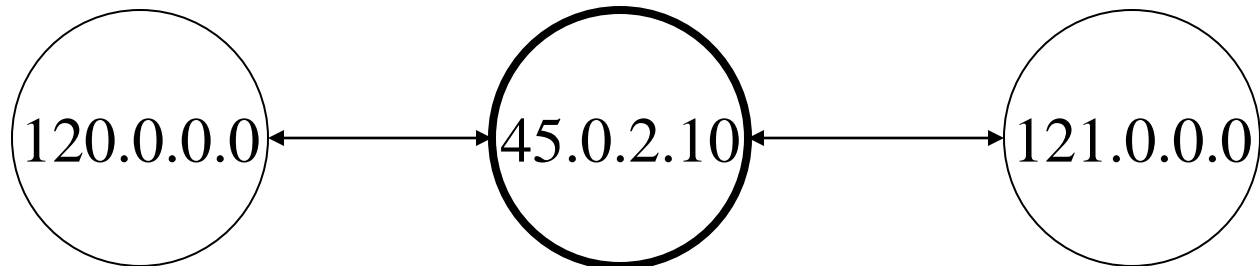
Hands-on:

Learn About Your IP Address

- Find your IP address
 - Select “start” on the taskbar, then “Run”
 - Type in “cmd” and click “OK”
 - Type “ipconfig /all” (and press enter)
- See who “owns” that address
 - Use <http://remote.12dt.com/>
- See where in the world it (probably) is
 - <http://www.geobytes.com/ipLocator.htm>

Routing Tables

IP Prefix	Next Router	Estimated Delay
216.141.xxx.xxx	120.0.0.0	18 ms
216.xxx.xxx.xxx	121.0.0.0	34 ms
101.42.224.xxx	120.0.0.0	21 ms
xxx.xxx.xxx.xxx	121.0.0.0	250 ms



Hands On: TraceRoute

- See how packets get from South Africa to you
 - Use <http://utl-lnx1.puk.ac.za/cgi-bin/webutil>
- Look at the same data visually
 - <http://visualroute.visualware.com/>

Domain Name Service (DNS)

- “Domain names” improve usability
 - Easier to remember than IP addresses
 - Written like a postal address: specific-to-general
- Each “name server” knows one level of names
 - “Top level” name servers know .edu, .com, .mil, ...
 - .edu name server knows umd, umbc, stanford, ...
 - .umd.edu name server knows wam, ischool, ttclass, ...
 - .wam.umd.edu name server knows rac1, rac2, ...

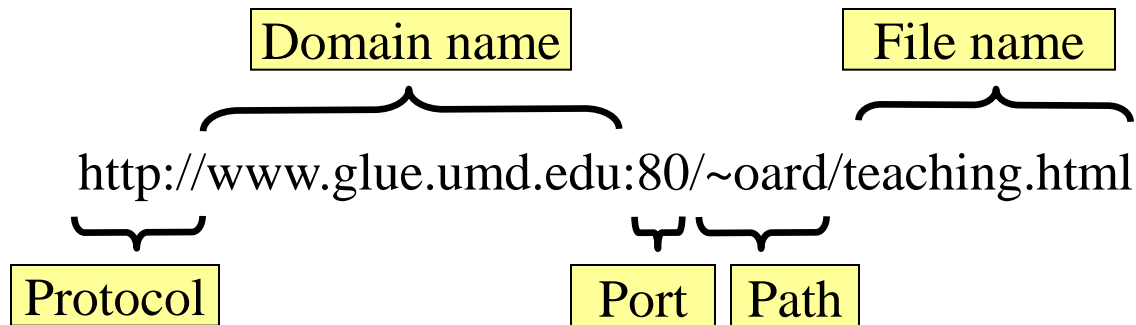
IP Addresses and Domain Names

IP address: 128.8.10.142

Domain Name: wam.umd.edu

Uniform Resource Locator (URL)

- Uniquely identify Web pages



Ports

- Well-known ports
 - 22 Secure Shell (for SSH and SFTP)
 - 25 Simple Mail Transfer Protocol (SMTP)
 - 53 Domain Name System (DNS)
 - 68 Dynamic Host Configuration Protocol (DHCP)
 - 80 Hypertext Transfer Protocol (HTTP)
 - 143 Internet Message Access Protocol (IMAP)
 - 554 Real-Time Streaming Protocol (RTSP)
- Registered Ports
 - 8080 HTTP server run by ordinary users
- Ephemeral Ports

Port Mapping

- Internet Service providers lease one IP address
 - But home networks may contain many machines
- Network Address Translation (NAT)
 - Each internal machine gets a private IP address
 - Ports on internal machines are mapped both ways
- Port forwarding
 - Permits public server to run in the local network

Paths

- Specify location of files on a hard drive
- Folder metaphor

- Hierarchically nested directories

`/afs/wam.umd.edu/home/wam/j/i/jimmylin/home`

`C:\Documents and Settings\Jimmy Lin\My Documents`

- Absolute vs. relative paths

`../pub`

`..\Desktop`

`~/oard`

Hands On:

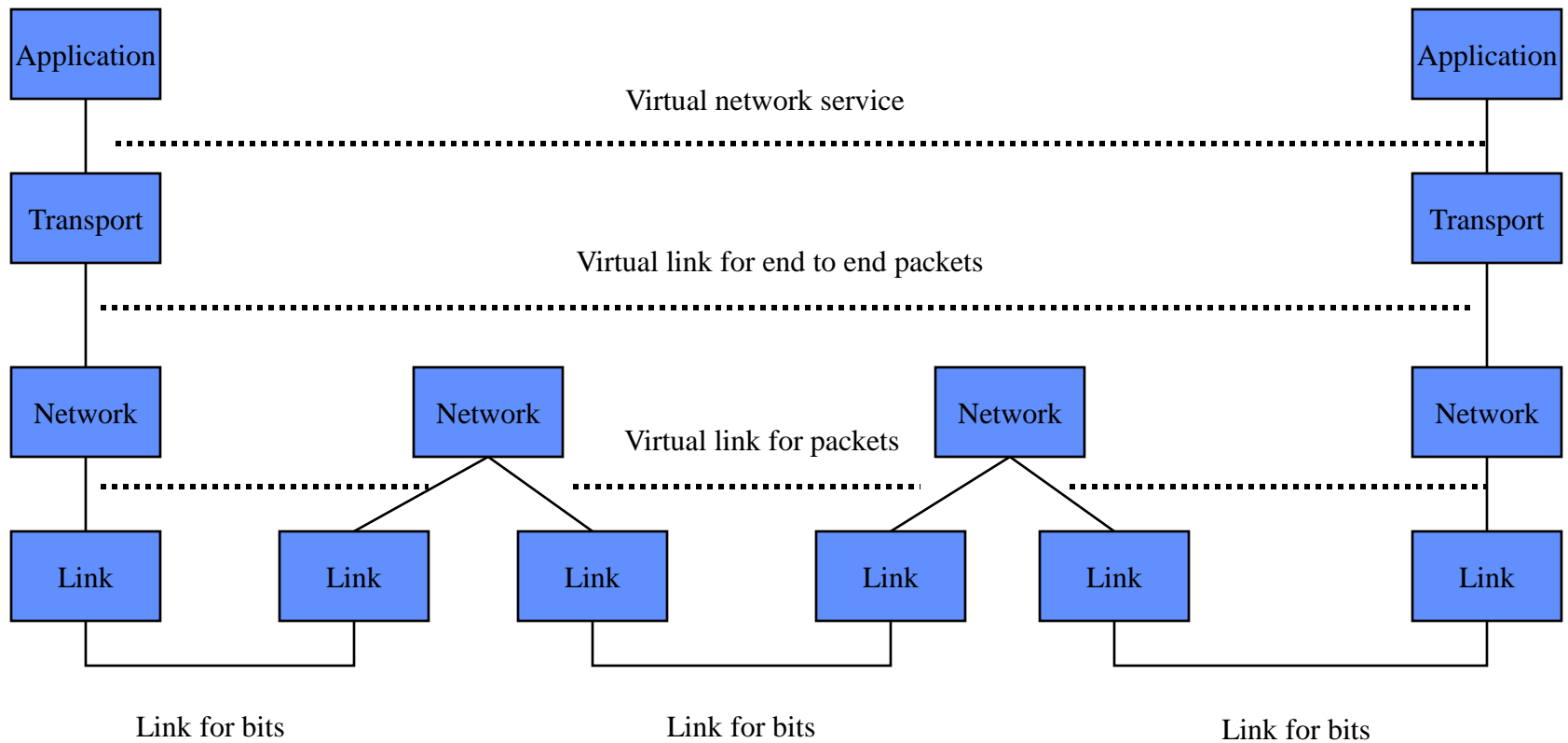
The Directory Tree

- First, use Windows Explorer to visually explore the directory tree
- Now launch a “shell” with Start->Run->cmd
 - “c:” takes you to Drive C
 - “dir” lists the present “directory”
 - “cd WINDOWS” takes you “down” to the WINDOWS directory
 - cd .. takes you “up” in the tree

The TCP/IP “Protocol Stack”

- Link layer moves bits
 - Ethernet, cable modem, DSL
- Network layer moves packets
 - **IP**
- Transport layer provides services to applications
 - UDP, **TCP**
- Application layer uses those services
 - DNS, SFTP, SSH, ...

TCP/IP layer architecture



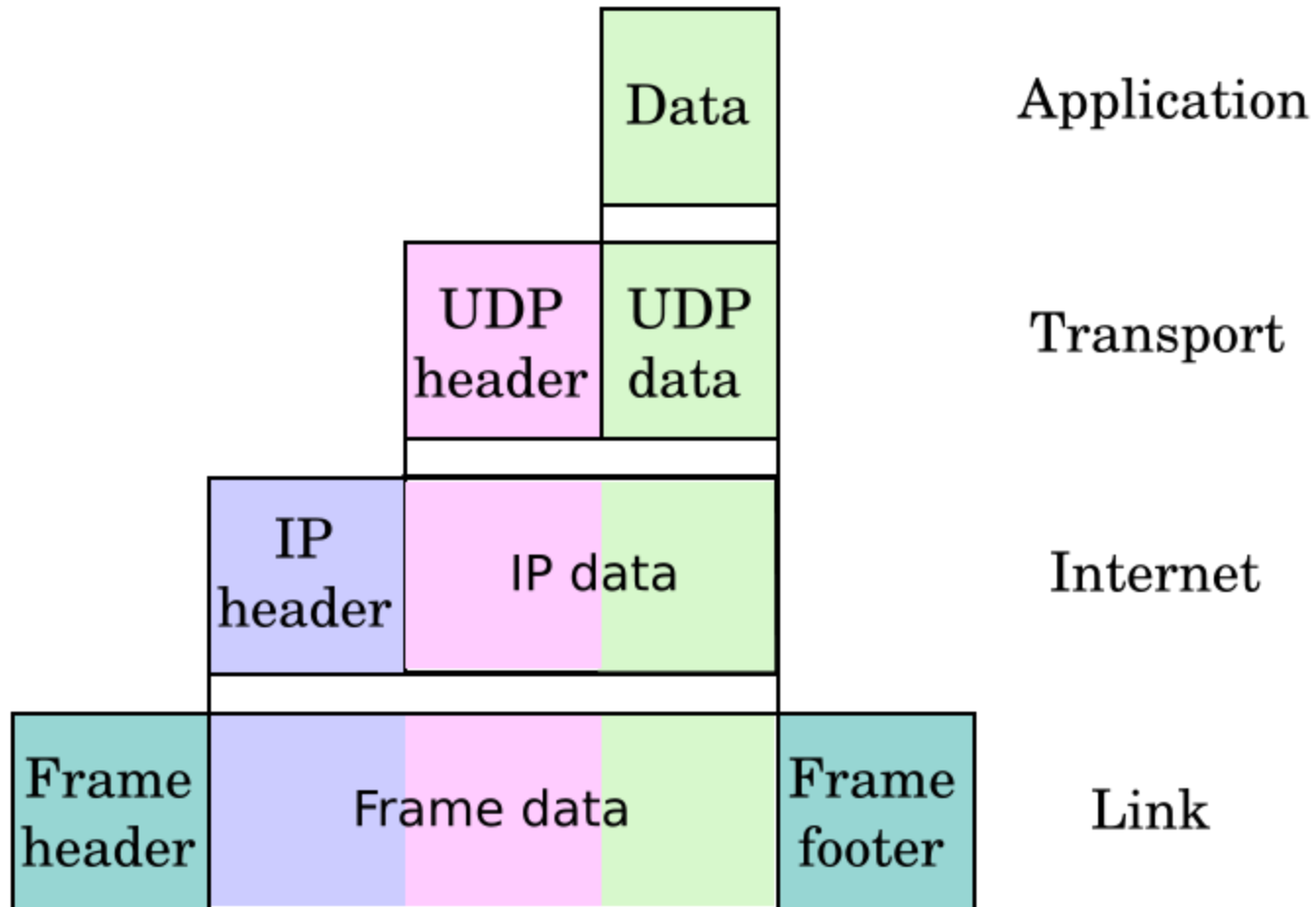
Transmission Control Protocol (TCP)

- Built on the network-layer version of UDP
- Guarantees delivery all data
 - Retransmits missing data
- Guarantees data will be delivered in order
 - “Buffers” subsequent packets if necessary
- No guarantee of delivery time
 - Long delays may occur without warning

User Datagram Protocol (UDP)

- The Internet's basic transport service
 - Sends every packet immediately
 - Passes received packets to the application
- No delivery guarantee
 - Collisions can result in packet loss
- Example: sending clicks on web browser

UDP/IP Protocol Stack



File Transfer Program (FTP)

- Used to move files between machines
 - Upload (put) moves from client to server
 - Download (get) moves files from server to client
- Both visual and command line interfaces available
- Normally requires an account on the server
 - Userid “anonymous” provides public access

Hands On:

Graphical Secure FTP

- SFTP to “terpconnect.umd.edu”
- Change directory to “/pub/**USERID**”
- Upload or download files
- You can see these files at:
[http://www.wam.umd.edu/~**USERID**/](http://www.wam.umd.edu/~USERID/)

Hands On:

Unsecure Command Line FTP

Start->Run->cmd

“ftp umiacs.umd.edu”

Login in as user “anonymous”

Download a file

- “cd pub/gina/lbsec690/”
- “binary”
- “get hwOne.ppt”

Exit

- “quit”

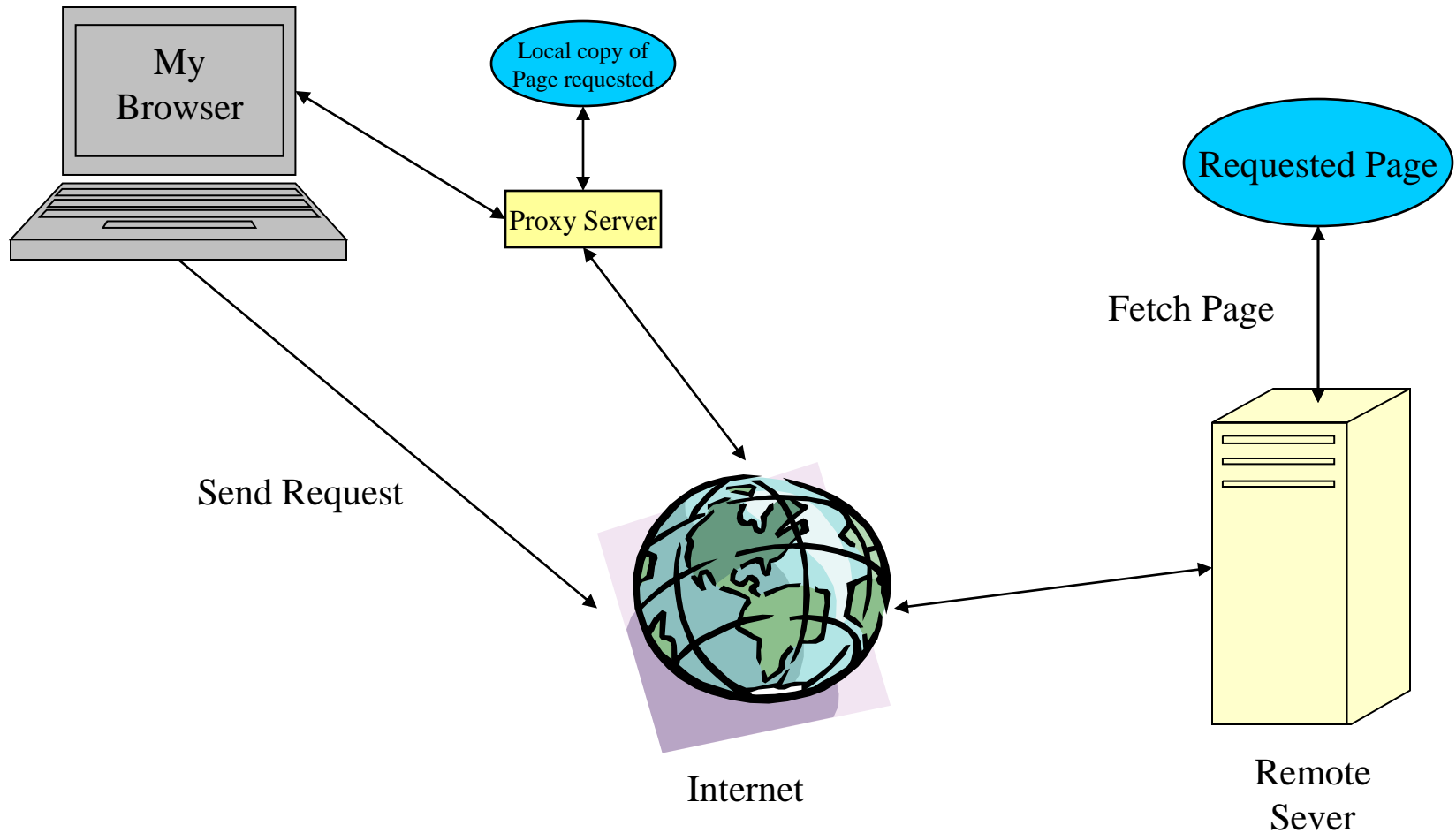
Agenda

- Computers
- The Internet
- The Web
- About the course

Internet \neq Web

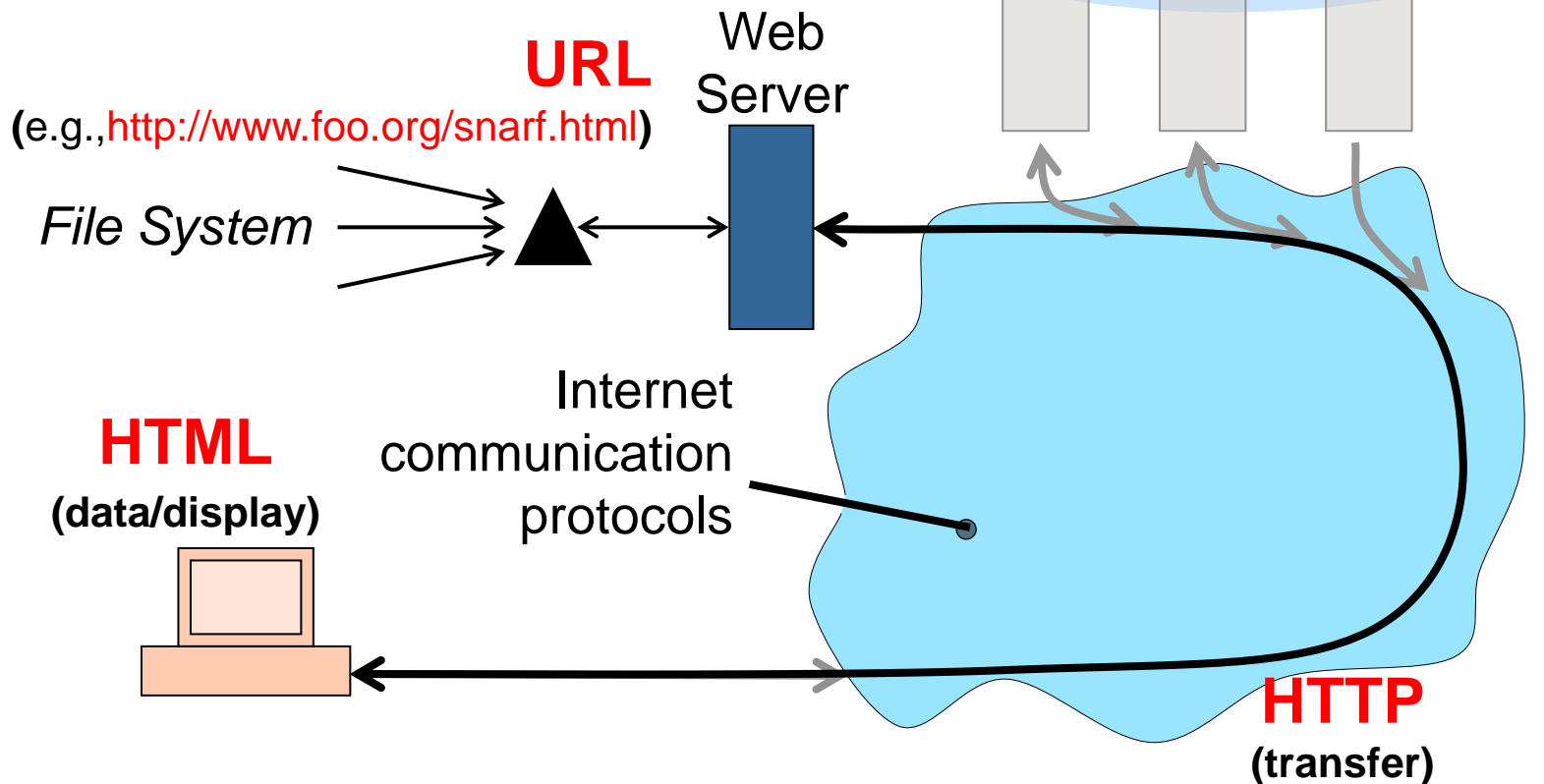
- Internet: collection of global networks
- Web: way of managing information exchange
- There are many other uses for the Internet
 - File transfer (FTP)
 - Email (SMTP, POP, IMAP)

The World-Wide Web



"The Web"

HTML
HTTP
URL



HyperText Transfer Protocol (HTTP)

- Send request

GET /path/file.html HTTP/1.0

From: someuser@jmarshall.com

User-Agent: HTTPTool/1.0

- Server response

HTTP/1.0 200 OK

Date: Fri, 31 Dec 1999 23:59:59 GMT

Content-Type: text/html

Content-Length: 1354

<html><body> <h1>Happy New Millennium!</h1> ... </body> </html>

HyperText Markup Language (HTML)

- Simple document structure language for Web
- Advantages
 - Adapts easily to different display capabilities
 - Widely available display software (browsers)
- Disadvantages
 - Does not directly control layout

“Hello World” HTML

This is the header

```
<html>  
<head>  
<title>Hello World!</title>  
</head>
```

```
<body>  
  
<p>Hello world! This is my first webpage!</p>  
  
</body>  
</html>
```

This is the actual content of the HTML document

Hands On:

Learning HTML From Examples

- Use Internet Explorer to find a page you like
 - <http://terpconnect.umd.edu/~oard>
- On the “Page” menu select “View Source” (in IE)
 - Opens a notepad window with the source
- Compare HTML source with the Web page
 - Observe how each effect is achieved

Hands On: “Adopt” a Web Page

- Modify the HTML source using notepad
 - For example, change the page to yours
- Save the HTML source on your “M:” drive
 - In the “File” menu, select “Save As”
 - **Select “All Files”** and name it “test.html”
- FTP it to your ../pub directory on terpconnect
- View it
 - [http://www.wam.umd.edu/~\(yourlogin\)/test.html](http://www.wam.umd.edu/~(yourlogin)/test.html)

Tips

- Edit files on your own machine
 - Upload when you're happy
- Save early, save often, just save!
- Reload browser to see changes
- File naming
 - Don't use spaces
 - Punctuation matters

HTML Document Structure

- “Tags” mark structure
 - `<html>a document</html>`
 - `an ordered list`
 - `<i>something in italics</i>`
- Tag name in angle brackets `<>`
 - Not case sensitive
- Open/Close pairs
 - Close tag is sometimes optional (if unambiguous)

Logical Structure Tags

- Head
 - Title
- Body
 - Headers: <h1> <h2> <h3> <h4> <h5>
 - Lists: , (can be nested)
 - Paragraphs: <p>
 - Definitions: <dt><dd>
 - Tables: <table> <tr> <td> </td> </tr> </table>
 - Role: <cite>, <address>, , ...


Physical Structure Tags

- Font
 - Typeface: ``
 - Size: ``
 - Color: ``
 - http://webmonkey.wired.com/webmonkey/reference/color_codes/Emphasis
 - Bold: ``
 - Italics: `<i></i>`

(Hyper)Links

index.html

```
<html>
<head>
<title>Hello World!</title>
</head>
<body>
<p>Hello world! This is my first webpage!</p>
<p>Click <a href="test.html">here</a> for another page.</p>
</body>
</html>
```



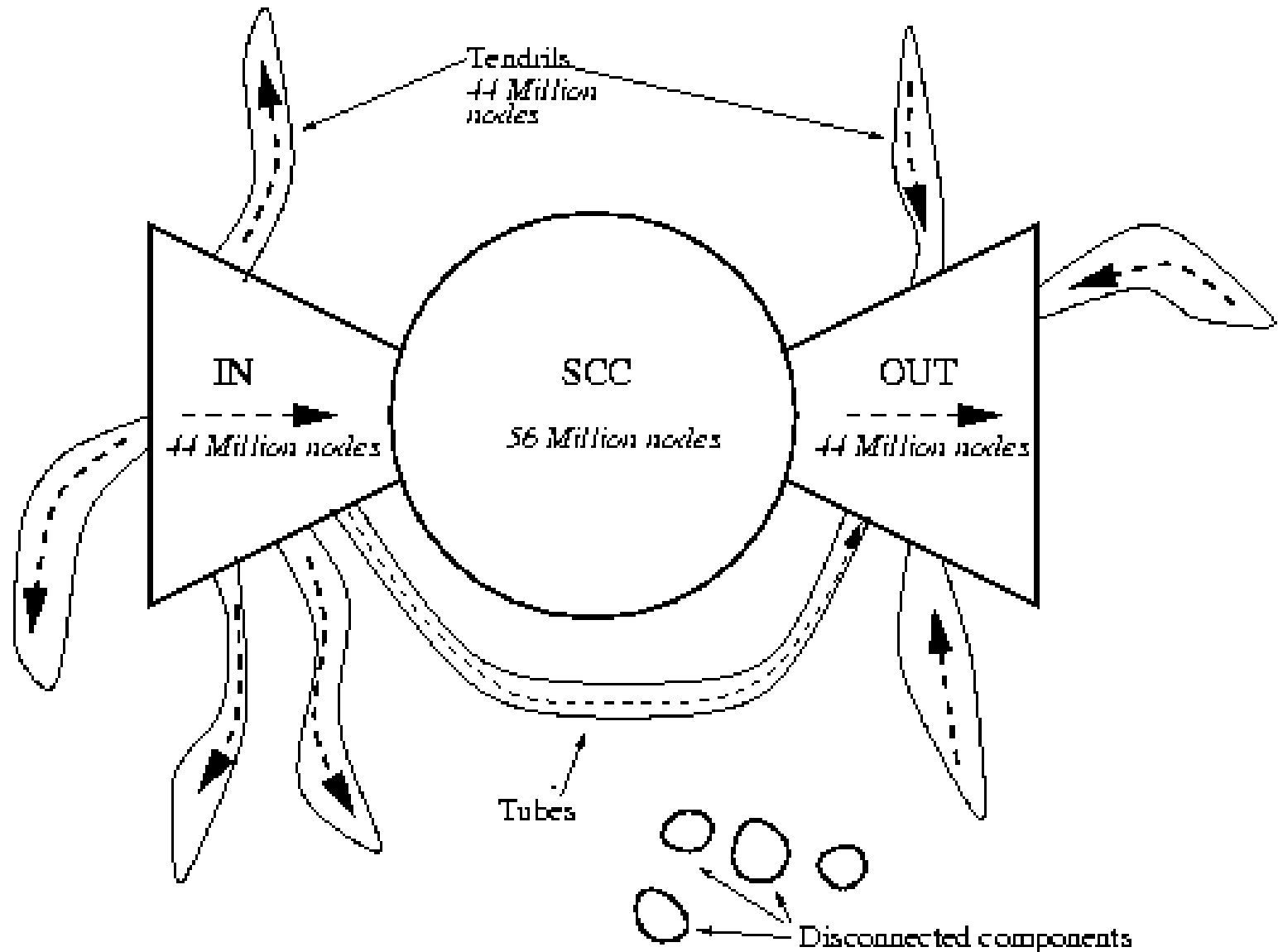
test.html

```
<html>
<head>
<title>Another page</title>
</head>
<body>
<p>This is another page.</p>
</body>
</html>
```

Hypertext “Anchors”

- Internal anchors: somewhere on the same page
 - ` Students`
 - Links to: `Student Information`
- External anchors: to another page
 - `CLIS`
 - `CLIS students`
- URL may be complete, or relative to current page
 - `2`
- File name part of URL is case sensitive (on Unix servers)
 - Protocol and domain name are not case sensitive

Link Structure of the Web



Images

- `` *or* ``
 - ``
 - SRC: can be url or path/file
 - ALT: a text string
 - ALIGN: position of the image
 - WIDTH and HEIGHT: size of the image
- Can use as anchor:
 - ``
- Example:
 - <http://www.umiacs.umd.edu/~daqingd/Image-Alignment.html>

Tables

eenie	mennie	miney
mo	catch	a tiger
by	the	toe

Table Example

```
<table align="center">
```

```
<caption align="right">The caption</caption>
```

```
< tr align="LEFT">
```

```
    <th> Header1 </th>
```

```
    <th> Header2</th>
```

```
</tr>
```

```
<tr><td>first row, first item </td>
```

```
    <td>first row, second item</td></tr>
```

```
< tr><td>second row, first item</td>
```

```
    <td>second row, second item</td></tr>
```

```
</table>
```

See also: <http://www.umiacs.umd.edu/~daqingd/Simple-Table.html>

Hands on:

The Internet Archive

- alexa.com Web crawls since 1997
 - <http://archive.org>
- Check out the CLIS Web site from 1998!
 - <http://www.clis.umd.edu>

Agenda

- Computers
- The Internet
- The Web
- About the course

A Personal Approach to Learning

- Work ahead, so that you are never behind
- Find new questions everywhere
 - Then find the answers somewhere
- Enrich your practical skills relentlessly
- Pick topics you want to learn more about
- Start thinking about your project soon
 - Pick partners with complementary skills

The Fine Print

- Group work is encouraged on homework
 - But you must personally write what you turn in
- Deadlines are firm and sharp
 - Allowances for individual circumstances are included in the grading computation
- Academic integrity is a serious matter
 - No group work during the exam!
 - Scrupulously respect time limits

Before You Go

On a sheet of paper, answer the following (ungraded) question (no names, please):

What was the muddiest point in today's class?