CT2108 – Networks and Data Communications 1

Dr Des Chambers

Introduction

1

Content

- Computer Networks vs. Distributed Systems
- Uses of computer networks
- Network Hardware
- Network Software
- Network Technologies

2

Computer Networks

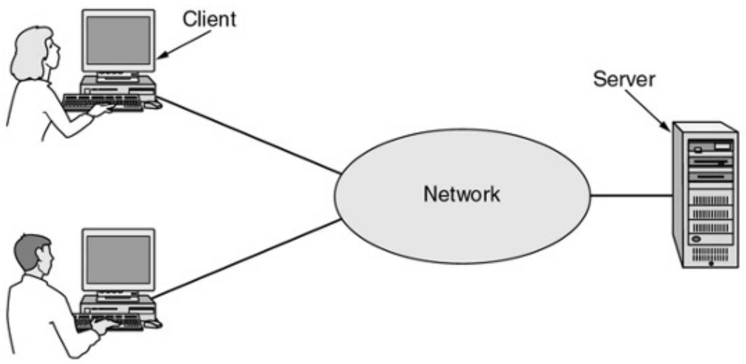
- A Computer Network is a collection of autonomous devices interconnected by some type of network technology
- Networks come in many shapes and sizes
- The Internet is a network of networks
  - World Wide Web is not a physical network. It is a distributed Client-Server application that runs over the Internet

3

Two computers are said to be INTERCONNECTED if they are able to exchange information. The connection can be copper wire, optical fiber, wireless, etc…

Clients - Server

- Client – Server Model
- Example - A network with two clients and one server.

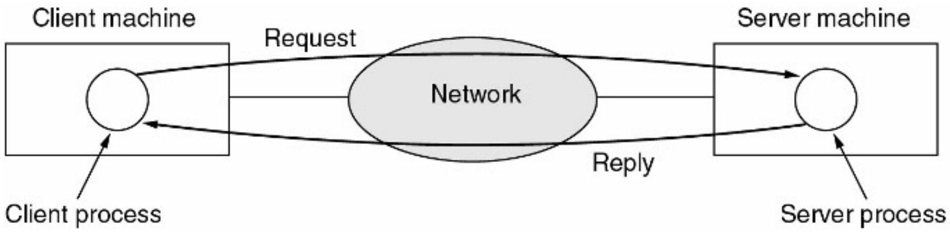


4

CLIENT – SERVER model is employed in most network applications. The Server is a powerful machine that can have multiple concurrent clients accessing its resources at the same time. Clients are usually simpler devices that run apps to interpret or display information provided by a server.

Client - Server

- The client-server model involves requests and replies.



- Client – Server model employs at least two processes: one running on the server and one running on the client

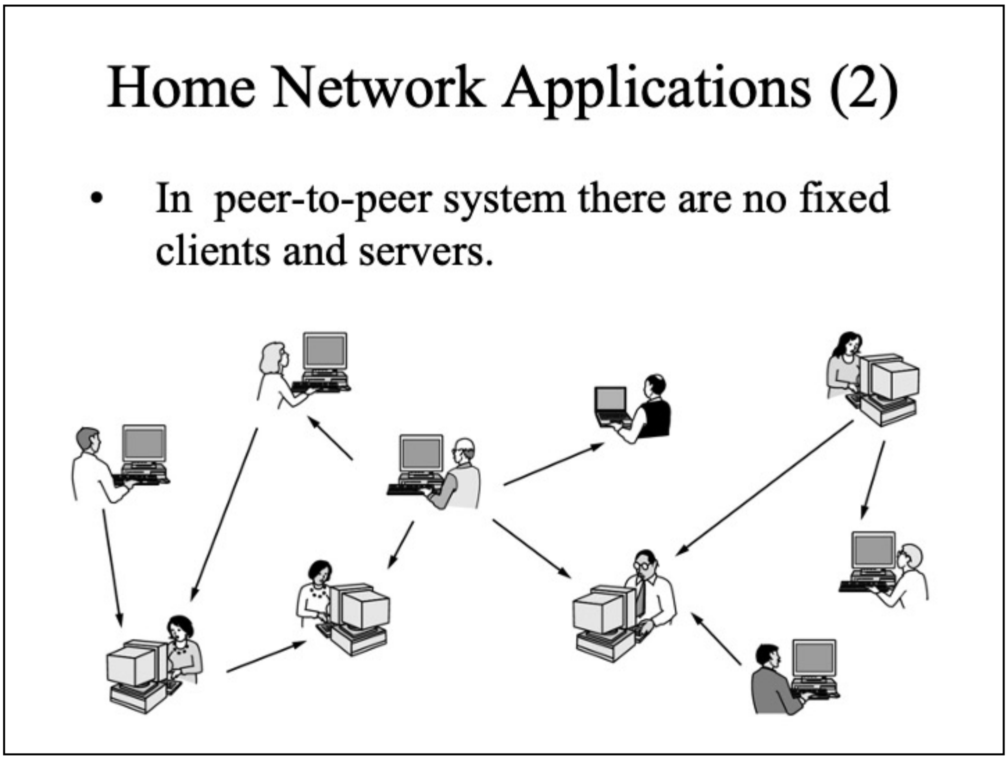
5

Home Network Applications (1)

- Access to remote information
  - Newspapers, publications, etc..
- Person-to-person communication
  - Instant messaging, e-mail, chat rooms, peer to peer communication
  - Interactive entertainment
  - Network games, video on demand, audio on demand, etc…
- Electronic commerce

6

KEN OLSEN – president of Digital (second big computer vendor in the world, after IBM, in 1977) said " There is no reason for any individual to have a computer in his home". History proved otherwise, and Digital went out of business in the 1990s.



Newer peer to peer systems (like BitTorrent) don't have a centralized db. Lookup of the content comes from a local db-s maintained by each of the members. Besides content, each user maintains a list of other users as well.

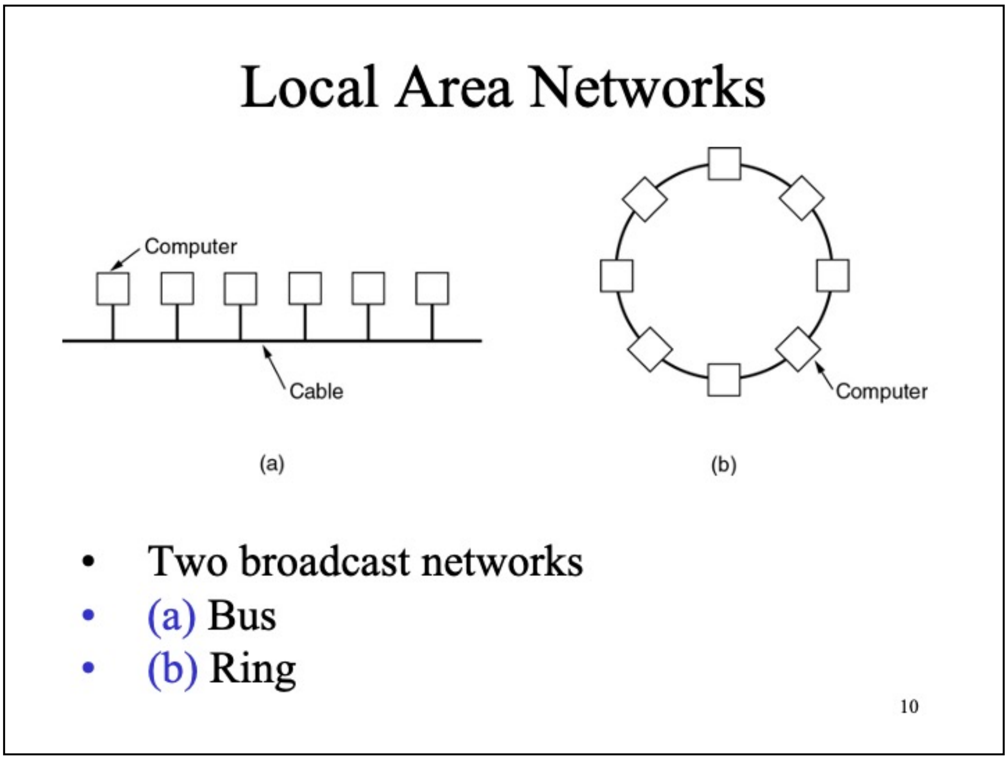
Network Types

- Local Area Networks
- Metropolitan Area Networks
- Wide Area Networks
- Wireless Networks
- Home Networks
- The Internet

Network Types

- Classification of interconnected processors by scale.

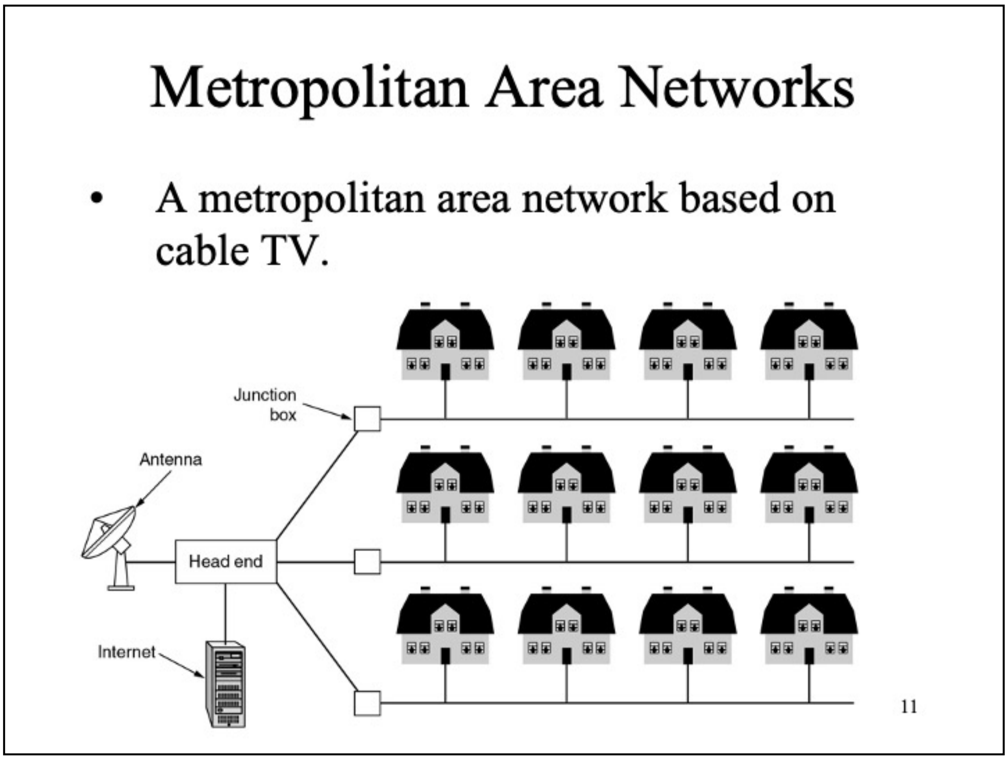
Interprocessor distance	Processors located in same	Example
1 m	Square meter	Personal area network
10 m	Room	Local area network
100 m	Building	Local area network
1 km	Campus	Local area network
10 km	City	Metropolitan area network
100 km	Country	Wide area network
1000 km	Continent	Wide area network
10,000 km	Planet	The Internet



Local Area Networks are privately owned networks within a single building or camps, up to a few km in size. Are restricted in size, which means that worst case transmission time is bounded and known in advance.

LANs use very often same cable to which all the machines are attached. Speeds ranging from 100Mb/s to about 10Gb/s. Various **TOPOLOGIES** are possible for broadcast LANs: **BUS**, **STAR** (most popular here is Ethernet) and **RING** are mostly used.

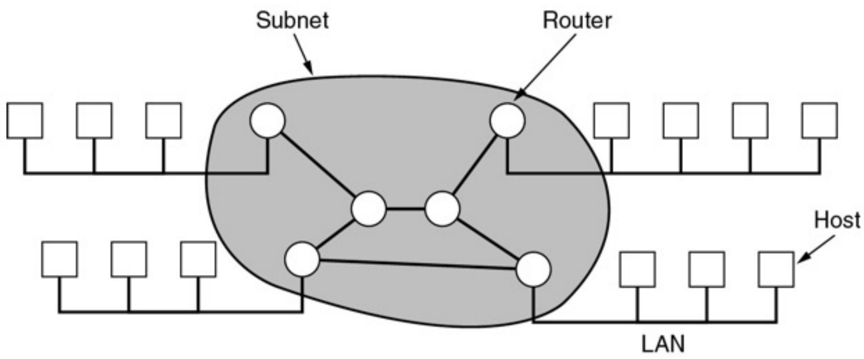
ETHERNET – bus-based network, bus and/or star topology, broadcast decentralized network, usually operating at 100Mb/s to 10Gb/s. Computers in Ethernet can transmit whenever they want; if two packets collide, each computer just waits a random time and tries again later.



Metropolitan Area Network (MAN) covers a city. Best known example of a MAN is the cable TELEVISION NETWORK available in many cities. Until late 1990's they were intended for television only. After that, the cable providers realized that they could offer two-way Internet in the unused parts of the spectrum. This transformed the TV network in a metropolitan area network.

Wide Area Networks (1)

- Relation between hosts on LANs and the subnet.



A Wide Area Network (WAN) spans over a large area, often a country or a continent. It contains a number of machines (called **HOSTS** in the networking context) that are connected by a communication **SUBNET**. The hosts are usually owned by people, while the subnet is owned by the telecom providers or Internet providers.

The job of the subnet is to carry messages from host to host. Separation of the pure communication aspects of the network (the subnet) from the application aspects (hosts) simplifies the complete network design.

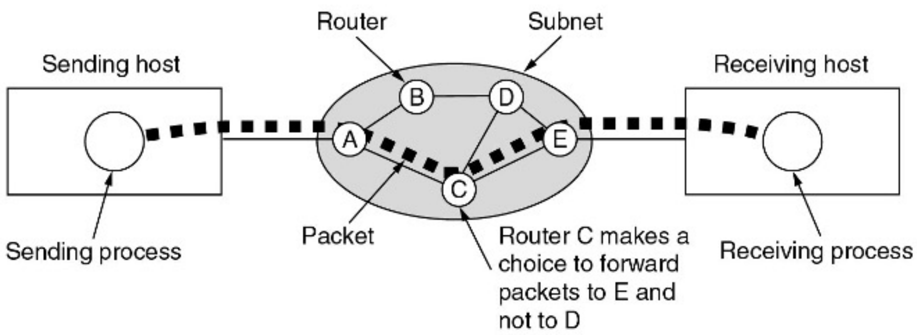
Subnets contain two components:

•**TRANSMISSION LINES** – move bits between machines (copper, optical fiber, radio, etc)

•**SWITCHING COMPONENTS** – specialized computers that connect three or more transmission lines. When data comes on one of the lines, the switching element must choose an ongoing line to forward the data. **ROUTER** is the technical name for those switching elements.

Wide Area Networks (2)

- A stream of packets from sender to receiver.



13

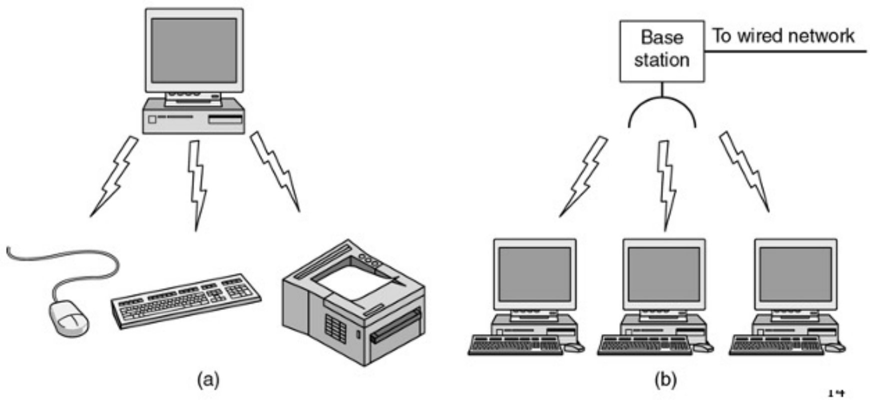
A **STORE-AND-FORWARD** or **PACKET SWITCHED** subnet is one where the packets are received entirely at intermediate routers, stored until some outgoing transmission line is free and then forwarded to the next router. When packets are small and all the same size, they are called **cells**.

When a process on a host wants to send a message to another host in the network, the sending host cuts the message into packets, each one caring some sort of sequence number. Those packets are then injected into the network, one at a time, in quick succession. The packets are delivered over the network and delivered to the receiving host, where they are reassembled and delivered to the receiving process.

In this figure, all the packets from sender to receiver followed same route ACE. In some subnets, the packets *must* follow always same path, in other subnets, the packets can follow different paths (they are routed separately). When the packet is getting to router A, the decision to follow path C or path B is made locally. This decision is made by A and how this decision is made is calling routing algorithm.

Wireless Networks

- (a) Bluetooth configuration
- (b) Wireless LAN



BLUETOOTH is an example of system interconnection network, and it refers to interconnecting computer components (monitor, mouse, keyboard, etc…). It is a master slave topology. The master tells the slaves what addresses to use, when they can broadcast, how long they can transmit, what frequencies they can use and other information.

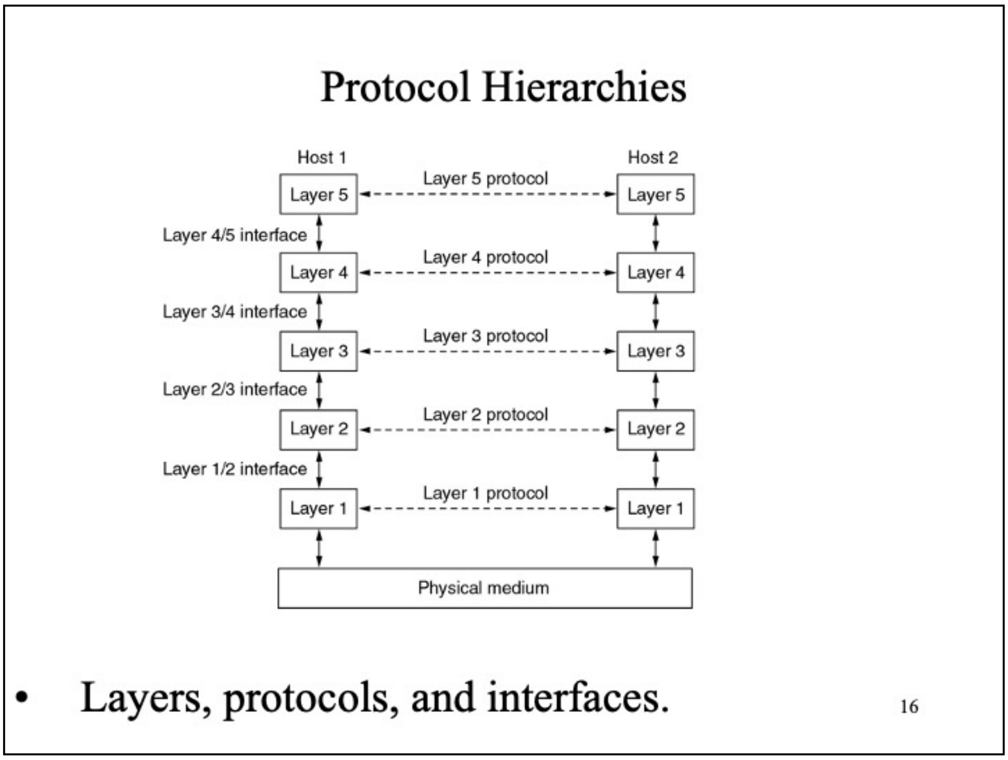
WIRELESS LANs – each computer has a radio modem and antenna with which it can communicate with other systems. IEEE 802.11 is a basic standard for wireless LAN. A number of newer, derivate standards are in place now.

WIRELESS WANs – cellular phone networks: 3G/4G/5G

Network Software

- Protocol Hierarchies
- Design Issues for the Layers
- Connection-Oriented and Connectionless Services
- Service Primitives
- The Relationship of Services to Protocols

15



To reduce complexity of design, networks are organized as layer, each one build upon the one below it. The number of layers, the name of each layer, the contents and function of each layer differ from network to network.

The purpose of each layer is to create services for the layers above, hiding to those layers the details of how those services are actually implemented. The fundamental idea is that a particular piece of software (or even hardware) provides a service to its users but keeps the details of its internal state and algorithms hidden from them. Layer n on a machine carries a conversation with layer n on another machine. The rules and conventions used in this conversation are known as layer n protocol. In essence, a **PROTOCOL** is an agreement between the communicating parties on how communication is to proceed. The entities that implement the protocol at different layers level are called **PEERS**. It is peers that communicate using the protocol.

In reality, no data is directly transferred from layer n on one machine on layer n on the other machine. In effect, each layer passes data and control information to the layer below it, until the lowest layer is reached. Below Layer 1 is the **physical medium**, through the communication occurs. Between each pair of adjacent layers there is an interface. The interface defines which primitive operations and services the lower layer makes available to the upper one. Most difficult design issue is to define clean interfaces between layers. A set of layers and protocols is called a **NETWORK ARCHITECTURE** (it has to contain enough information to allow hardware and software engineers to design hardware and software that would obey the right protocol). A list of protocols used by certain systems, one protocol per layer, is called a **PROTOCOL STACK**.

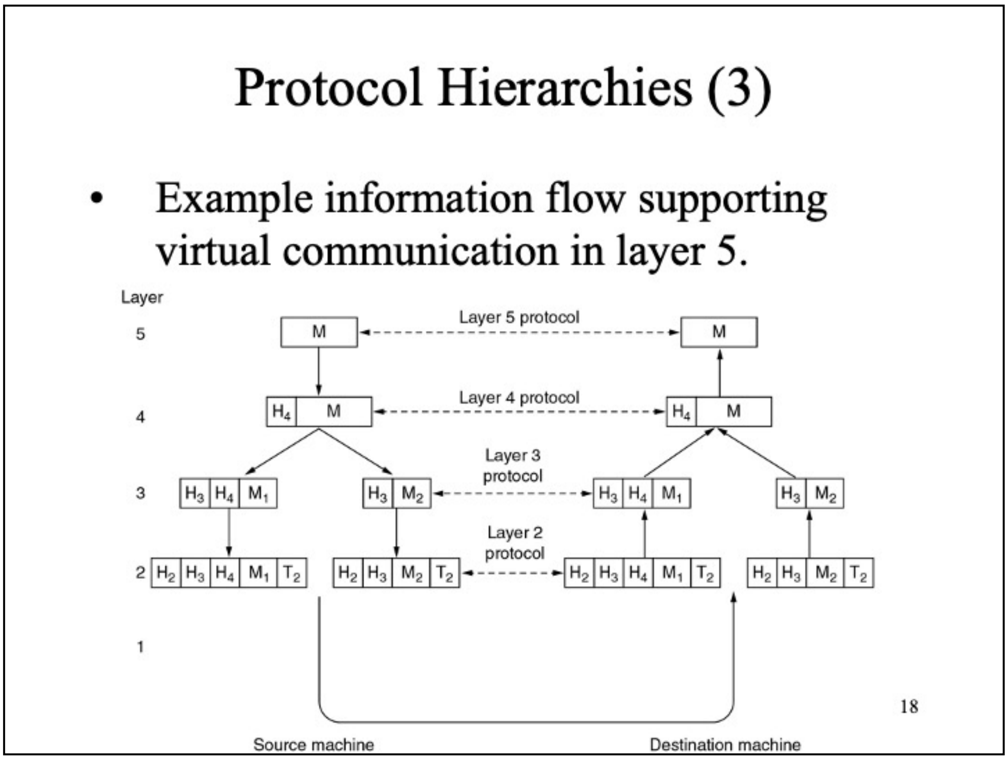
Protocol Hierarchies (2)

- The philosopher-translator-secretary architecture.

17

Layer3 – two philosophers (peer processes), one speaks English and the other one speaks French. Since they have no common language, they engage a translator (peer processes at Layer 2). The translators, each contact a secretary (peer process Layer 1). Translators have agreed on a common language that both know (Dutch).

Note that each protocol is completely independent of the other ones. If at any time, the translators decide to change the language, all they have to do is to agree between each other. None of the interfaces with layer 3 or layer 1 will be changed. Similarly, the secretary could choose to use a different transmission medium (say e-mail), without disturbing or even informing the other layers.



A message is produced by an application process running at layer 5. Message M is then given to layer 4 for transmission. Layer 4 puts a header in the front of the message to identify the message and pass the result to layer 3. The header includes control information, such as sequence numbers to allow layer 4 on destination to deliver messages in the right order, if the lower layers do not maintain sequence. In some layers, headers can also contain sizes, times and other control fields.

At layer 3 there is a limit on the size of the packet that can be transmitted. So layer 3 will break the incoming message into smaller parts, packets, adding header H3 corresponding to layer 3 on each packet. In this example, message M is split into M1 and M2.

Layer 3 decides which outgoing lines to use and passes the message to layer 2. Layer 2 adds not only a header to each piece, but also a trailer and gives the resulting units to layer 1 for physical transmission.

At the receiving machine, the message moves upwards, from layer to layer, with headers being stripped off as it progresses.

Design Issues for the Layers

- **Addressing**
  - consequence of having multiple destinations
- **Error Control**
  - The receiver should be able to inform sender which data was received correctly
- **Flow Control**
  - Keep sender from swamping slow receiver with data
  - Keep the sender from swamping with data slow networks
- **Multiplexing**
  - Use same communication channel for multiple, unrelated conversations
- **Routing**
  - When multiple paths between source and destination, one path must be chosen

Connection-Oriented and Connectionless Services

- Layers can offer two types of services to the layers above: connection oriented services and connection-less services
- Connection oriented services
  - Reliable message stream (sequence of pages)
  - Reliable byte stream (remote login, file transfer, etc..)
  - Unreliable connection (digitized voice or video)
- Connection-less
  - Datagram service (in analogy with telegram service)
  - Acknowledged datagram service
  - Request-reply service

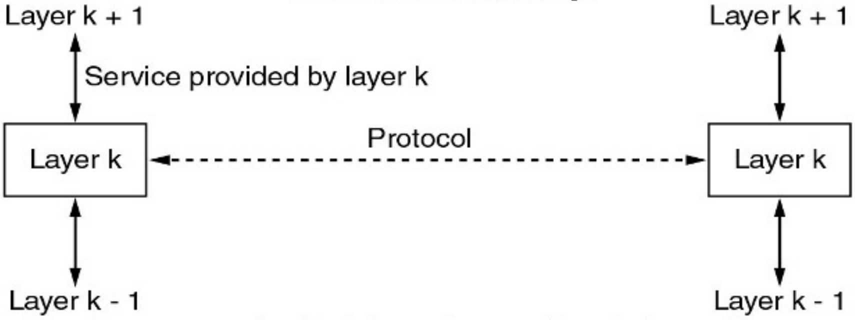
20

CONNECTION ORIENTED SERVICE – modeled after the phone systems. The service users establishes a connection, uses the connection and then releases the connection. The main idea, is that the connection acts as a PIPE, at one end data is pushed and at the other end data is received. In most of the cases, the order is preserved. Sometime, during the connection establishment phase, a **NEGOTIATION** is employed (for establishing some parameters of the connection)

CONNECTION-LESS SERVICE – modeled after the postal system. Each message (letter) carries the full destination address, each one being routed through the system independent of the others. It is possible that the messages will arrive at the destination in out of order.

Each service is characterized by **QUALITY OF SERVICE**. Some services are reliable in the sense that they never loose data. Usually, reliability is implemented with acknowledgements from the receiver that it received data. This introduces overhead and delays in the communication, which sometime is OK, but sometime is not (in real time voice and video communication).

Services to Protocols Relationship

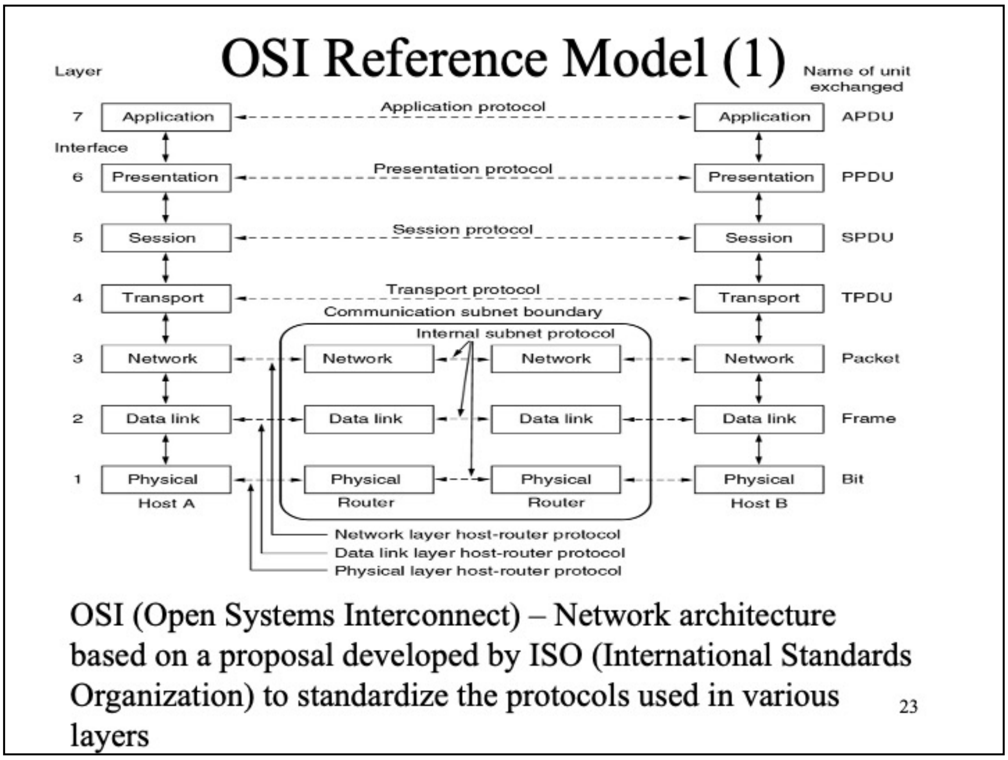


- Service – set of primitives (operations) that a layer provides to the layer above it; relate to the interfaces between layers
- Protocol – set of rules governing the format and meaning of packets exchanged by peer entities within a layer; relate to the packets that are sent between peer entities between different machines

21

Reference Models

- The OSI Reference Model
- The TCP/IP Reference Model
- A Comparison of OSI and TCP/IP
- A Critique of the OSI Model and Protocols
- A Critique of the TCP/IP Reference Model



Design principles that lead to the seven-layer design are as follows:

1. A layer should be created where a **different abstraction is needed**
2. Each layer should perform a **well-defined function**
3. The function of each layer should be chosen with an eye toward defining **internationally standardized protocols**
4. The layer boundaries should be chosen to **minimize the information flow across the interfaces**
5. The number of layers should be large enough to avoid throwing together separate, **distinct functions** out of necessity and small enough to avoid inefficiency

OSI Reference Model (2)

- Physical Layer
  - Transmitting raw bits over communication channel
  - Typical questions that are addressed:
    - How many volts used to represent a "1" and how many for "0"
    - How many nanoseconds a bit last
    - Full duplex transmission or not (both directions)
    - How initial connection is established and how is torn down when both sides are finished
    - How many pins the network connector will have and what is each pin used for
  - Design issues – sending one bit "1" on one side has to get in the other side as "1" not as "0"
    - Mechanical, electrical and timing interfaces
    - Physical transmission medium

OSI Reference Model (3)

- Data Link Layer
  - Transform the raw transmission facility (offered by the physical layer) into a line that appears free of undetected transmission errors to the network layer
  - Design issues
    - **Error detection and correction**
      - The sender breaks up the input data into **data frames** (typically a few hundred or thousands bytes) and transmits the frames sequentially. If the service is reliable, the receiver has to confirm the correct receipt of each frame
    - **Flow control** – keep a fast transmitter drowning a slow receiver with data
      - Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Usually, this is integrated with the error handling mechanism
    - Broadcast networks have an additional issue in the data-link layer: how to control **access to the shared channel**. A special sub-layer of the data-link layer, the medium access control sub-layer deals with this problem

OSI Reference Model (4)

- Network Layer
  - Controls the operation of the subnet
  - Design issues
    - **Routing** - how packets are routed from source to destination
    - **Congestion control** – if too many packets are present in the subnet at the same time
    - Allow heterogeneous networks to be interconnected
    - In broadcast networks, the routing problem is thin or non existent

26

OSI Reference Model (5)

- Transport Layer
  - Accepts data from the above layer, split it into smaller units and pass them to the network layer. Ensures that those pieces arrive correctly at the other end
  - Determines what type of service to provide
    - Most popular type of transport connection is error free, point to point channel that delivers messages or bytes in the order in which they were sent
    - Other type of transport services: delivering datagrams with no guarantee about the order of delivery, broadcasting messages
  - It is a true end-to-end layer, all the way from source to the destination

27

The Transport layer has to perform its function in a way that isolates the upper layers from the inevitable changes in the hardware technology

Layers one to three are chained, while layers four to seven are END to END layers.

OSI Reference Model (6)

- Session Layer
  - Allows users on different machines to establish sessions between them
  - Sessions offer different services:
    - **Dialog control** – keeping track of those whose turn is to transmit
    - **Token management** – preventing two parties from attempting same critical operation at the same time
    - **Synchronization** – marking long transmissions to make sure they can be resumed from where they were when a crash happened

OSI Reference Model (7)

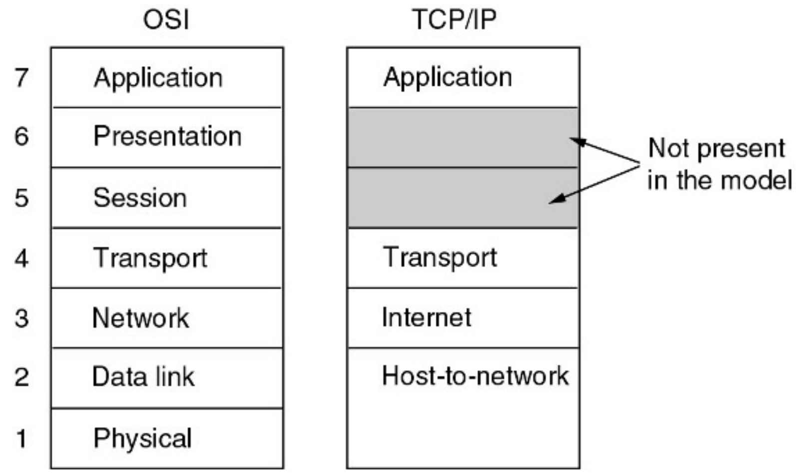
- Presentation Layer
  - It is not concerned with moving bits around, but with checking the syntax and semantics of data that is being moved by the layers below
  - In order to make it possible for computers with different data representations to communicate, the data structures to be exchanged can be defined in an abstract way, along with an standard encoding to be used on the wire.
  - It manages these abstract data structures and allows higher-level data structures to be defined and exchanged

29

OSI Reference Model (8)

- Application Layer
  - The application layer contains a number of different protocols and applications that are needed by the users.
  - A good example of widely used application protocol is HTTP (HyperText Transfer Protocol), which is the basis for Wide World Web distributed system.
  - When a browser wants a page, it sends the name of the page, to a web server, using HTTP protocol
  - Other application: FTP, e-Mail, news, etc…

TCP/IP Reference Model (1)



- Used by Internet, packet switching network (of networks) based on a connectionless internetwork layer

31

TCP/IP Reference Model (2)

- Internet Layer
  - Permits the hosts to inject packets into any network and have them travel independently to the destination (potentially using different paths or networks). The packets may arrive in a different order. It is the job of the higher layer to rearrange them
  - It defines an official packet format and protocol, called IP (**Internet Protocol**). The job of internet layer is to deliver IP packets where they want to go
    - Packet routing is one of the biggest issues
    - Avoiding congestion is another big issue

32

Analogy with the snail mail system

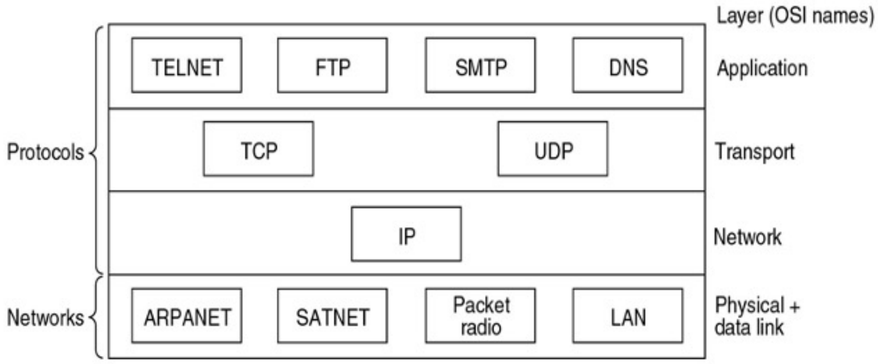
TCP/IP Reference Model (3)

- The Transport Layer
  - Designed to allow peer entities on the source and destination to carry on a conversation
  - Two end to end protocols: TCP and UDP
    - Transmission Control Protocol – end to end reliable connection oriented protocol that allows a byte stream originating from one machine to be delivered with no error on another machine in the Internet
    - User Datagram Protocol – unreliable connectionless protocol for applications that don't want TCP's sequencing flow control and want to provide theirs (or to apps that don't want connection overhead)

33

TCP/IP Reference Model (3)

- Application Layer – contains all the high-level protocols: http, file transfer, e-mail, domain name system, etc…



TCP/IP Reference Model (3)

- Host to Network Layer
  - Below the Internet Layer, in TCP/IP reference model is a great void
  - The model doesn't say much about it, except that the host has to connect to the network using some protocol, so it can send IP packets to it
  - This protocol is not defined and varies from host to host and from network to network

35

Comparing OSI and TCP/IP (1)

- Concepts central to the OSI model
  - Services
  - Interfaces
  - Protocols
- TCP/IP model didn't make a clear distinction between services, interfaces and protocols

36

We are only comparing the models, not the corresponding protocol stacks. Both OSI and TCP/IP models have much in common. Both are based on a concept of a stack of independent protocols. Also, the functionality of the layers is somehow similar. **OSI** has three concepts that are central, and perhaps the biggest contribution of OSI was to make the explicit distinction between the three concepts:

SERVICES – tells what the layer does, not how entities above it access it, nor how the layer works. It defines the layer's semantics.

INTERFACES – tells the processes above it how to access it. It specifies what the parameters are and what result to expect. It says nothing about how the layer works inside.

PROTOCOLS – are peer layers own business. A layer can use any protocol if it gets the job done (provides the offered services). The protocols can change without affecting the software in the higher layer.

TCP/IP Model did not originally make a clear distinction between services, interfaces and protocols. People have tried to retrofit after the specifications, to make it look more like OSI. i.e., the only real services offered by the IP layer are: SEND_IP_PACKET and RECEIVE_IP_PACKET. Consequently, protocols in the OSI model are better hidden than in TCP/IP model and can be replaced a lot easier (as the technology changes), without disturbing the layers above.

Comparing OSI and TCP/IP (2)

- OSI reference was described **before** the protocols were invented
  - It means that the model is not biased towards a set of protocols, but is rather generic
  - Downside is that designers didn't have much experience (what function to put in which layer)
- TCP/IP protocols came first. The model was done just as a description of the protocols
  - The protocols did fit perfectly the model
  - The problem was that the model didn't fit any other protocols, useless to describe other types of networks than TCP/IP based.

Comparing OSI and TCP/IP (3)

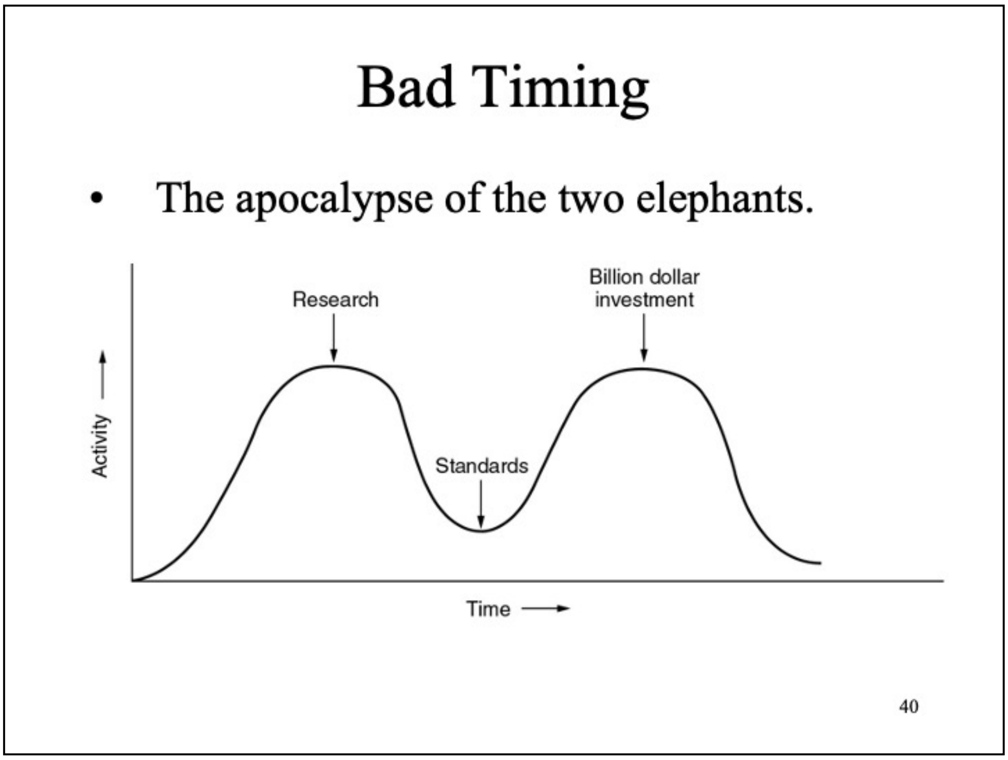
- OSI has 7 layers, TCP/IP has only 4 layers
- Connection less vs. connection-oriented communication
  - OSI model – supports connection less and connection-oriented services at the network layer, while at transport layer supports only one type of connection-oriented service
  - TCP/IP model – supports only connection less services at the network layer, while at the transport layer offers both connection oriented and connection less services, giving the users a choice

38

A Critique of the OSI Model and Protocols

- Why OSI did not take over the world
  - Bad timing
  - Bad technology
  - Bad implementations
  - Bad politics

39



Standards have to be written between the research phase and investment phase. With OSI this didn’t happen. Partially because the two phases were too close, partially because the competing TCP/IP model was already used in the research institutions, and companies already started to offer TCP/IP based products.

Bad Technology

- Two of the layers in OSI were nearly empty (session and presentation) while two others were overcrowded (network and data-link)
- The protocols and service definitions are very complex. They were almost incomprehensible.
- Some of the functions (such as addressing, flow control and error control) reappear again and again at different layers. This is unnecessary and inefficient

41

Bad Implementation

- Given the complexity of the protocols, the implementations were huge and inefficient
- People started to associate OSI with poor quality because the implementations were slow on the available equipment
- In contrast, one of the first implementations of TCP/IP was part of Berkeley UNIX and was quite good (also free). People begun to use it -> improvements -> large community -> more improvements

42

Bad Politics

- OSI was created by European Telecommunications ministries and USA Government – this has been seen (perceived) as a bunch of bureaucrats trying to push an inferior standard on the throats of researchers and scientists, that already had a working solution (TCP/IP)
- Of course, this was only partially true, but enough for TCP/IP model to get a lot of supporters

43

A Critique of the TCP/IP Reference Model

- Problems:
  - Service, interface, and protocol not distinguished
  - Not a general model
  - Host-to-network “layer” not really a layer
  - No mention of physical and data link layers
  - Minor protocols deeply entrenched, hard to replace

44

Hybrid Model

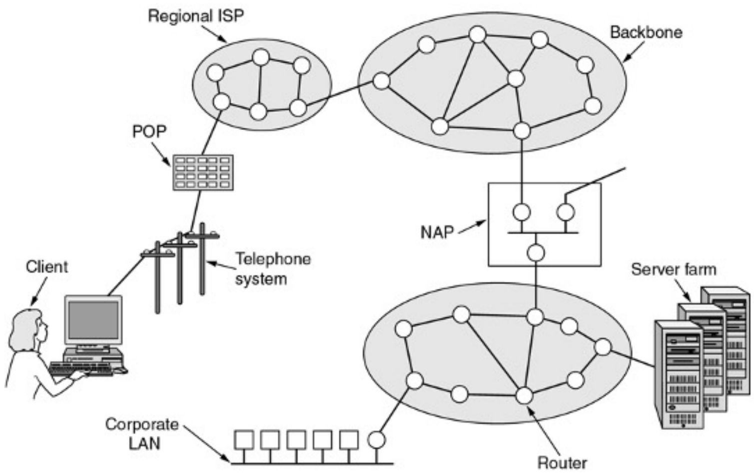
- The hybrid reference model to be used in this course

5	Application layer
4	Transport layer
3	Network layer
2	Data link layer
1	Physical layer

45

Architecture of the Internet

- Overview of the Internet.



46

The **INTERNET** is not a network at all, but a vast collection of different networks that use common protocol and provide common services. Internet was not planed nor controlled by anyone. It all started back in late 1950's when DoD realized that all the communication is based on telephony systems with little or non redundancy. Therefore, a highly distributed, fault tolerant system needed to be in place. Late 1950's ARPA (Advanced Research Project Agency) was formed. In 1967 ARPA's interest turned into networking … ARPANET, first network of computers has been built. The subnet would consist of minicomputers (called **IMPs – INTERNET MESSAGE PROCESSORS**), interconnected by 56kb/s transmission lines. For reliability, each IMP would be connected to at least two another IMPs. The subnet was a datagram subnet, so if some lines and IMPs were destroyed, the messages could automatically be rerouted along the alternative paths.

The packets carried over the modem are transferred to the ISPs POP (**POINT OF PRESENCE**) where they are removed from the telephone system and injected into the ISP regional network. From this point on, the system is fully digital, and packet switched. The regional ISP consists of several routers, located in main cities where the ISP operates. The regional ISP subnet is connected to a backbone, that usually run between states or countries (even continents). The backbones are usually made of high bandwidth optical fiber lines, interconnected by routers. Interconnecting backbones often belonging to different competing ISPs.

Network Standardization

- Telecommunications World
  - ITU (International Telecommunication Union)
    - Radiocommunication Sector (ITU-R)
    - Telecommunications Standardization Sector (ITU-T)
    - Development Sector (ITU-D)
- International Standards World
  - ISO (International Standards Organization)
    - ANSI (American National Standards Institute), BSI (British Standards Institute), etc…
  - IEEE (Institute of Electrical and Electronics Engineers)
    - i.e. IEEE 802 group standardize LAN standards
- Internet Standards World
  - ITB (Internet Architecture Board)
    - RFC (Request for comments) for different standards
  - Got divided into:
    - IRTF (Internet Research Task Force)
    - ITEF (Internet Engineering Task Force)

47

References

- Andrew S. Tanenbaum – Computer Networks, ISBN 0-13-066102-3

48

CT2108 – Nets and Comms 1

Physical Layer

Content

- Theoretical basis for data communications
- Transmission media
  - Guided, wireless and satellite
- Examples of telecommunicating systems used in practice for WANs:
  - Fixed phone system
  - Mobile phone system
  - Cable television

The Theoretical Basis for Data Communication

- Fourier Analysis
- Bandwidth-Limited Signals
- Maximum Data Rate of a Channel
- Channel Organization
- Types of data communication

Fourier Analysis (1)

- A periodic signal with period T can be constructed as a sum of a number of sine and cosines

$$g(t) = \frac{1}{2}c + \sum_{n=1}^{\infty} a_n \sin(2\pi nft) + \sum_{n=1}^{\infty} b_n \cos(2\pi nft)$$

- Where:
  - **f**=1/T
  - $\mathbf{a_n}$ and $\mathbf{b_n}$ are the sine and cosine amplitudes of the n^{th} harmonics
  - **c** -s a constant

Such a decomposition is called FOURIER SERIES. The original function of time can be reconstructed (by performing the sum) if the period T is known and the amplitudes an, bn and the constant c are given.

Fourier Analysis (2)

- The $\mathbf{a_n}$ amplitudes can be computed for any g(t) by multiplying the previous formula (Fourier series) by sin (2πkft) and integrating from 0 to T

$$a_n = \frac{2}{T}\int_0^T g(t)\sin(2\pi nft)dt$$

SINA * SINB = ½[COS(A-B) - COS(A+B)]

SINA * COSB = ½[SIN(A+B) + SIN(A-B)]

COSA * SINB = ½[SIN(A+B) – SIN(A-B)]

COSA * COSB = ½[COS(A+B) + COS(A-B)]

Ssin = -cos

Scos= sin

Fourier Analysis (3)

- The bn amplitudes can be computed by multiplying the Fourier series by cos(2πkft) and integrating from 0 to T

$$b_n = \frac{2}{T}\int_0^T g(t)\cos(2\pi nft)\,dt$$

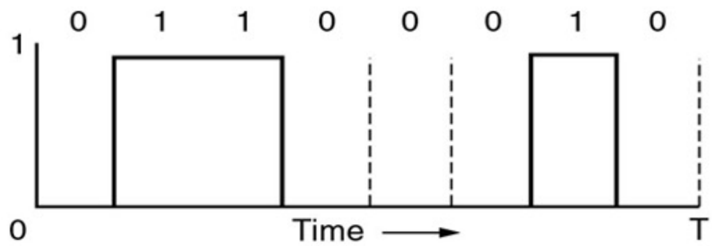
Fourier Analysis (4)

- The constant c can be computed by just integrating form 0 to T

$$c = \frac{2}{T} \int_0^T g(t)\,dt$$

Bandwidth limited signals (1)

- Let's consider ASCII character 'b' encoded in a byte. The bit pattern to be sent is 01100010
- The way to deal with data signal that has a fixed duration (like our example) is to imagine that the entire pattern repeats over again forever



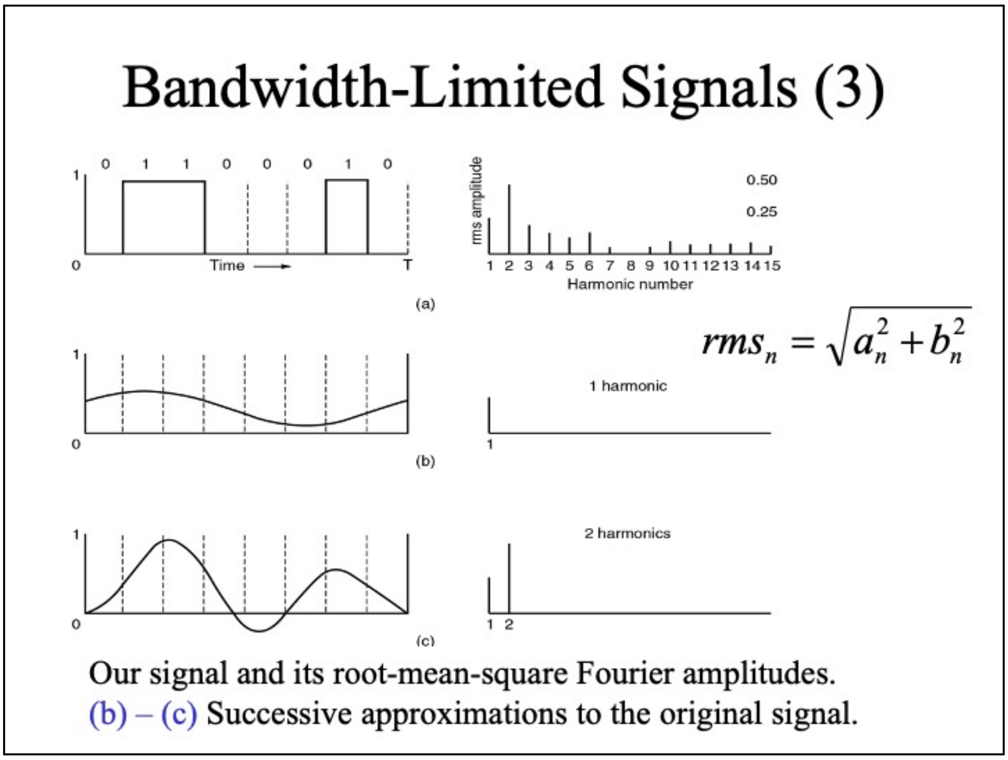
Bandwidth limited signals (2)

- The Fourier analysis of the signal yields to the coefficients:

$$a_n = \frac{1}{\pi n}[\cos(\frac{\pi n}{4}) - \cos(\frac{3\pi n}{4}) + \cos(\frac{6\pi n}{4}) - \cos(\frac{7\pi n}{4})]$$

$$b_n = \frac{1}{\pi n}[\sin(\frac{3\pi n}{4}) - \sin(\frac{\pi n}{4}) + \sin(\frac{7\pi n}{4}) - \sin(\frac{6\pi n}{4})]$$

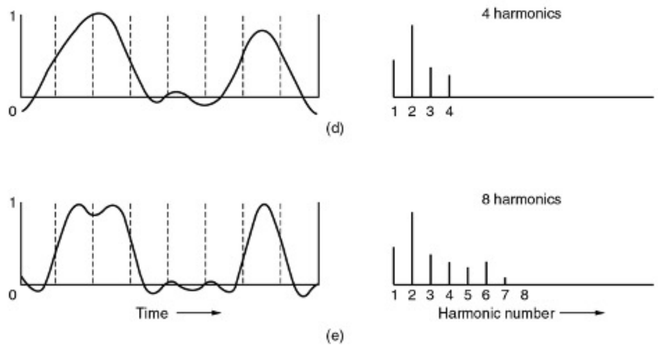
$$c = \frac{3}{4}$$



The root mean square amplitudes **(RMS)** are of interest because those values show the **ENERGY** transmitted at the corresponding frequency. No transmission facility can transmit signals without loosing some power in the process. If all Fourier components were equally diminished than the resulted signal would be reduced in amplitude, but not distorted. Unfortunately, all the transmission medium diminish different Fourier components by different amounts, resulting in a distortion of the signal at the other end. Usually, the amplitudes are transmitted undiminished between 0 and a frequency Fc (F cut, measured in Hz), with all frequencies above this Fc attenuated. The range of frequencies transmitted without being attenuated it is called **BANDWIDTH.** In practice the cut of frequency is not really sharp, so usually this bandwidth is the range from 0 to the frequency where half of the power of the signal gets through.

The bandwidth is a physical property of the transmission medium and usually depends on the construction, thickness and length of the medium. Consider our signal (ASCII 'B', (a)) and look how the signal would look if only the lowest frequencies where transmitted (i.e. if the function was approximated by only a few terms).

Bandwidth-Limited Signals (4)



(d) – (e) Successive approximations to the original signal.

Bandwidth-Limited Signals (5)

Bps	T (msec)	First harmonic (Hz)	# Harmonics sent
300	26.67	37.5	80
600	13.33	75	40
1200	6.67	150	20
2400	3.33	300	10
4800	1.67	600	5
9600	0.83	1200	2
19200	0.42	2400	1
38400	0.21	4800	0

Given a rate of b bits per second. The time to send one byte is 8/b, which is the period for the first harmonic. The frequency of the first harmonic would be b/8 Hz.

A (non-DSL) phone line has an artificially introduced cut off frequency of 3000Hz. That means that the number of the highest harmonics passed through is about 3000/b/8 = 24000/b. It is clear that on a (non-DSL) phone line signals can't be sent beyond 9600 bps, since the reconstruction of the signal at the other end would be extremely difficult. It is obvious that higher data rates, for binary signals, it is impossible to receive and reconstruct the original signal without somehow increasing the available bandwidth. DSL enabled phone lines now provide much wider wider bandwidth than the original 3KHz that was available, so much higher data rates are now possible. Sophisticated coding and modulation do exist and can achieve higher data rates.

Maximum rate of a channel (1)

- Noiseless channel – Nyquist theorem:

$$MaxDataRate = 2B\log_2 V \qquad [bits/\sec]$$

- Where:
  - B – noiseless channel bandwidth
  - V – number of discrete levels of the signal transmitted through the channel

Maximum rate of a channel (2)

- If random noise is present, then the situation deteriorates rapidly.
- In reality, there is always random noise, due to the motion of the molecules in the system
- The amount of the thermal noise is measured by the ratio of the signal power to the noise power, called **signal-to-noise ratio**

$$SNR_{dB} = 10\log_{10}(\frac{S}{N}) \qquad [db]$$

- Signal to noise ratio is given in decibels (dB) and the ratio itself is not usually quoted.
- A ratio of 10 is 10dB, a ratio of 100 is 20dB, a ratio of 1000 is 30dB and so on

Maximum rate of a channel (3)

- Shannon theory:

$$MaxDataRate = B\log_2(1+\frac{S}{N}) \qquad [bits/\sec]$$

- Where:
  - B – bandwidth of the noisy channel
  - S/N – signal to noise ratio of the channel
- Shannon's theory demonstrates that the maximum data rate through a channel is limited by the amount of noise present, no matter how many or few signal levels are used and no matter how often or infrequently samples are taken
- This limit is the upper limit and real systems rarely achieve it

Example

- A channel of 3000Hz bandwidth (phone line) has a SNR ratio of 30dB. The transmitted signal has 8 levels. What is the maximum data rate that this channel can accommodate?
- What if the SNR of the channel would be only 10dB?

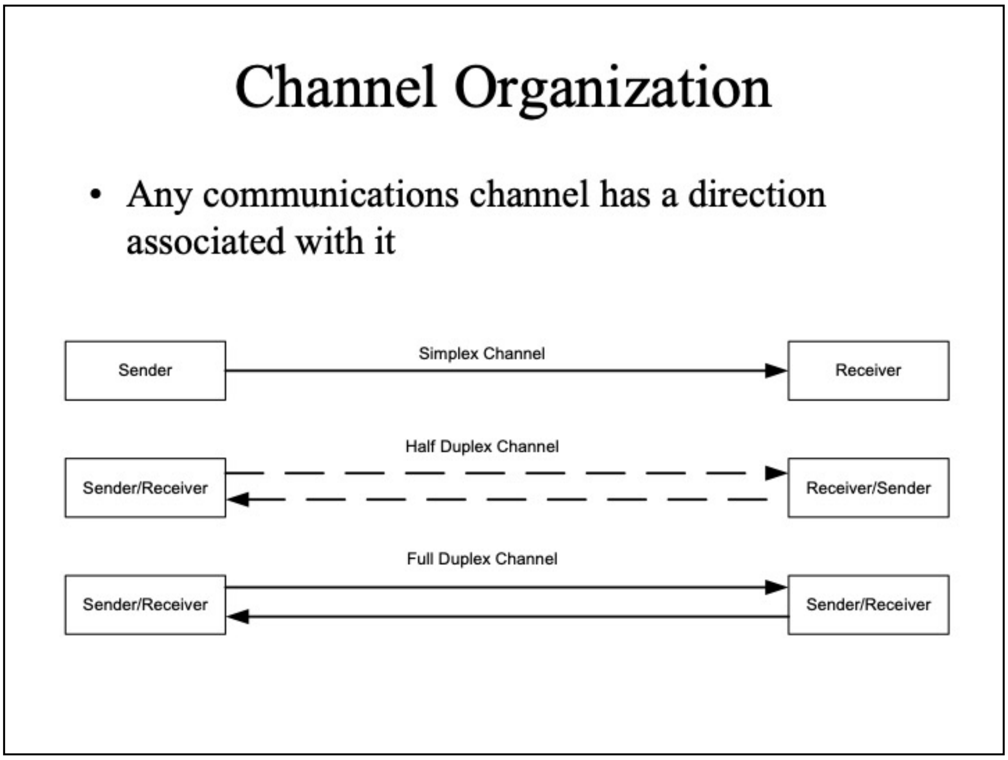
Compute both Nyquist and Shannon:

Nyquist – max = 18 kbps

Shannon – max = 30kbps

Take the lowest, in our case is the Nyquist limit. So the maximum throughput through the channel would be 18kb/s

If the channel would have 10dB, than the Shannon theory gives aabout max = 10kbps, therefore the maximum throughput would be 10kbps. Always take the smallest between the two (Nyquist and Shannon)



The message source is the transmitter, and the destination is the receiver. A channel whose direction of transmission is unchanging is referred to as a simplex channel. For example, a radio station is a simplex channel because it always transmits the signal to its listeners and never allows them to transmit back.

A half-duplex channel is a single physical channel in which the direction may be reversed. Messages may flow in two directions, but never at the same time, in a half-duplex system. In a telephone call, one party speaks while the other listens. After a pause, the other party speaks and the first party listens. Speaking simultaneously results in garbled sound that cannot be understood.

A full-duplex channel allows simultaneous message exchange in both directions. It really consists of two simplex channels, a forward channel and a reverse channel, linking the same points. The transmission rate of the reverse channel may be slower if it is used only for flow control of the forward channel

Channel Organization

- Synchronization
  - the receiver should know the exact moment when data is valid
- Synchronous channels
  - Data and timing information are sent separately (through separate channels or same channel)
  - The timing channel transmits clock pulses to the receiver
  - Upon receive of clock pulse, the receiver reads the data and latches it
  - The data is not read again until next clock pulse arrives
- Asynchronous channels
  - No separate timing information is used
  - Transmitter and receiver must agree in advance on timings
  - Start and stop conditions are used
  - Accurate oscillators will measure the bit widths.

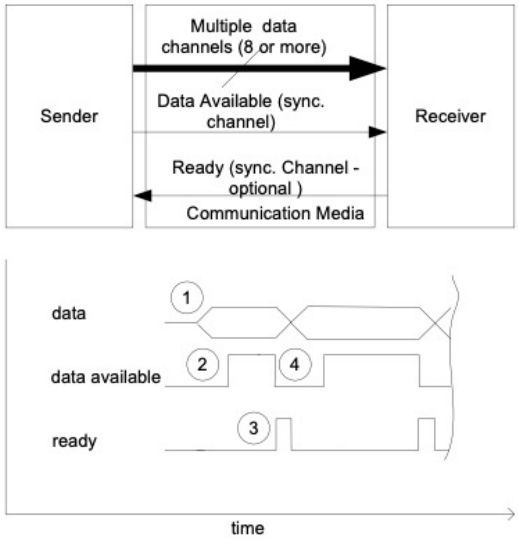
Data is not generally sent at a uniform rate through a channel. Instead, there is usually a burst of data followed by a pause, after which the data flow resumes. Packets of binary data are sent in this manner, possibly with variable-length pauses between packets, until the message has been fully transmitted. In order for the receiving end to know the proper moment to read individual binary bits from the channel, it must know exactly when a packet begins and how much time elapses between bits. When this timing information is known, the receiver is said to be synchronized with the transmitter, and accurate data transfer becomes possible. Failure to remain synchronized throughout a transmission will cause data to be corrupted or lost.

Two basic techniques are employed to ensure correct synchronization: synchronous and asynchronous communication

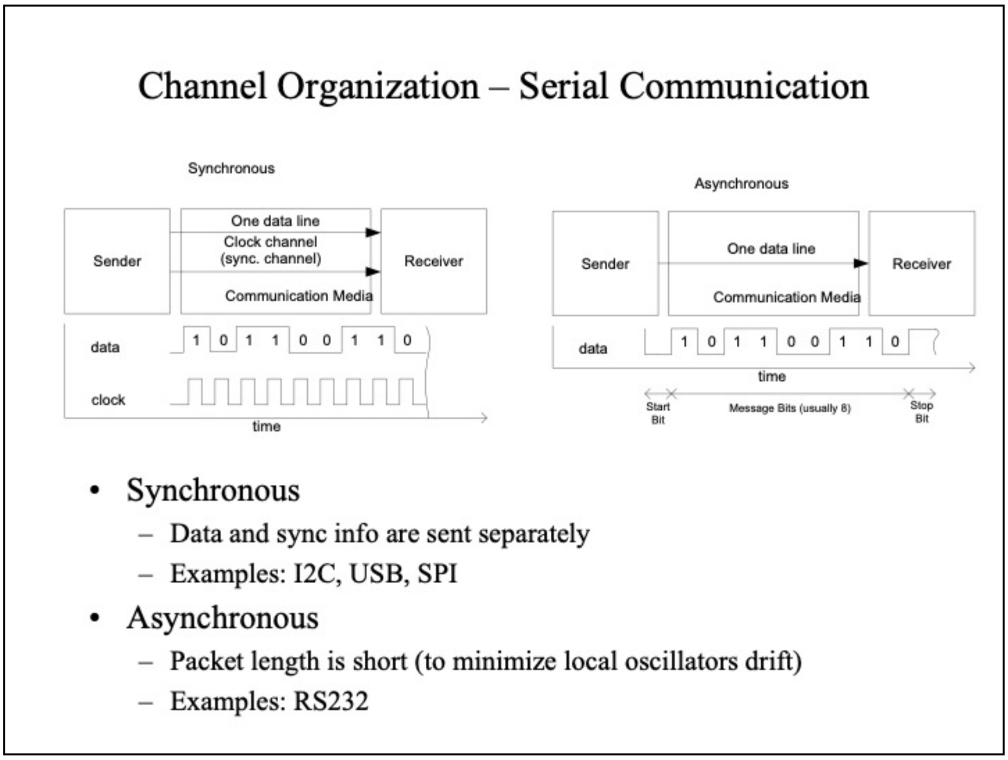
In synchronous systems, separate channels are used to transmit data and timing information. The timing channel transmits clock pulses to the receiver. Upon receipt of a clock pulse, the receiver reads the data channel and latches the bit value found on the channel at that moment. The data channel is not read again until the next clock pulse arrives. Because the transmitter originates both the data and the timing pulses, the receiver will read the data channel only when told to do so by the transmitter (via the clock pulse), and synchronization is guaranteed. Techniques exist to merge the timing signal with the data so that only a single channel is required. This is especially useful when synchronous transmissions are to be sent through a modem. Two methods in which a data signal is self-timed are non-return-to-zero and biphase Manchester coding. These both refer to methods for encoding a data stream into an electrical waveform for transmission.

In asynchronous systems, a separate timing channel is not used. The transmitter and receiver must be preset in advance on timings. A very accurate local oscillator within the receiver will then generate an internal clock signal that is equal to the transmitter's within a fraction of a percent. For the most common serial protocol, data is sent in small packets of 10 or 11 bits, eight of which constitute message information. When the channel is idle, the signal voltage corresponds to a continuous logic '1'. A data packet always begins with a logic '0' (the start bit) to signal the receiver that a transmission is starting. The start bit triggers an internal timer in the receiver that generates the needed clock pulses. Following the start bit, eight bits of message data are sent bit by bit at the agreed upon baud rate. The packet is concluded with a parity bit and stop bit

Channel Organization – Parallel Communication



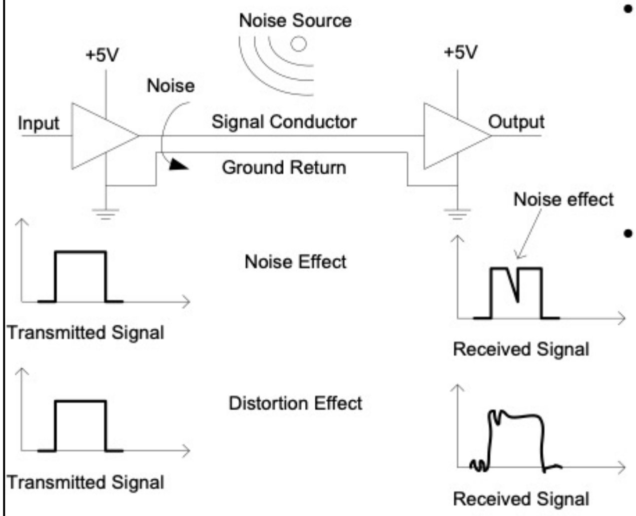
- 1 – Sender places data on the channel
- 2 – Sender asserts "data available"
- 3 – receiver reads the data and asserts "ready" signal
- 4 – Sender de-asserts "data available" signal and it will be ready for a new data transfer



In synchronous systems, separate channels are used to transmit data and timing information. The timing channel transmits clock pulses to the receiver. Upon receipt of a clock pulse, the receiver reads the data channel and latches the bit value found on the channel at that moment. The data channel is not read again until the next clock pulse arrives. Because the transmitter originates both the data and the timing pulses, the receiver will read the data channel only when told to do so by the transmitter (via the clock pulse), and synchronization is guaranteed. Techniques exist to merge the timing signal with the data so that only a single channel is required. This is especially useful when synchronous transmissions are to be sent through a modem. Two methods in which a data signal is self-timed are nonreturn-to-zero and biphase Manchester coding. These both refer to methods for encoding a data stream into an electrical waveform for transmission.

Asynchronous serial transmission: for the most common serial protocol, data is sent in small packets of 10 or 11 bits, eight of which constitute message information. When the channel is idle, the signal voltage corresponds to a continuous logic '1'. A data packet always begins with a logic '0' (the start bit) to signal the receiver that a transmission is starting. The start bit triggers an internal timer in the receiver that generates the needed clock pulses. Following the start bit, eight bits of message data are sent bit by bit at the agreed upon baud rate. The packet is concluded with a parity bit and stop bit. The packet length is short in asynchronous systems to minimize the risk that the local oscillators in the receiver and transmitter will drift apart. When high-quality crystal oscillators are used, synchronization can be guaranteed over an 11-bit period. Every time a new packet is sent, the start bit resets the synchronization, so the pause between packets can be arbitrarily long.

Noise and Distortion



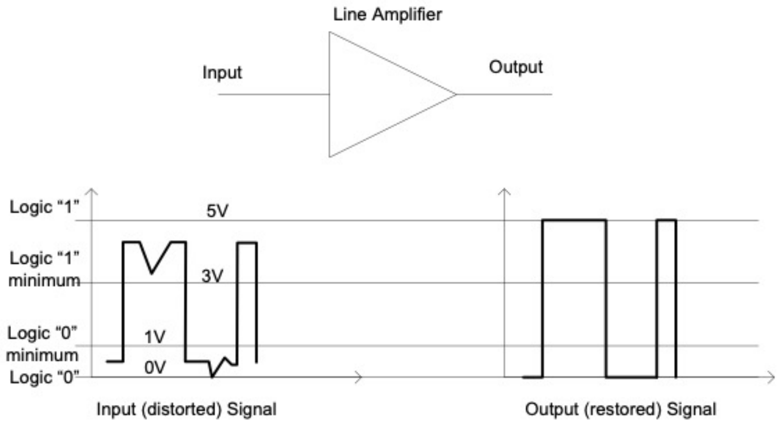
- Long conductors act like receiving antennas for electrical noise radiated by motors, switches and other electrical circuits
- Electrical signals get distorted when passing through metallic conductors due to limitations of bandwidth and internal thermal noise of the transmission media

Because of the very high switching rate and relatively low signal strength found on data, address, and other buses within a computer, direct extension of the buses beyond the boudaries of the main circuit board or plug-in boards would pose serious problems.

First, long runs of electrical conductors, either on printed circuit boards or through cables, act like receiving antennas for electrical noise radiated by motors, switches, and electronic circuits.

A second problem involves the distortion of electrical signals as they pass through metallic conductors. Signals that start at the source as clean, rectangular pulses may be received as rounded pulses with ringing at the rising and falling edges.

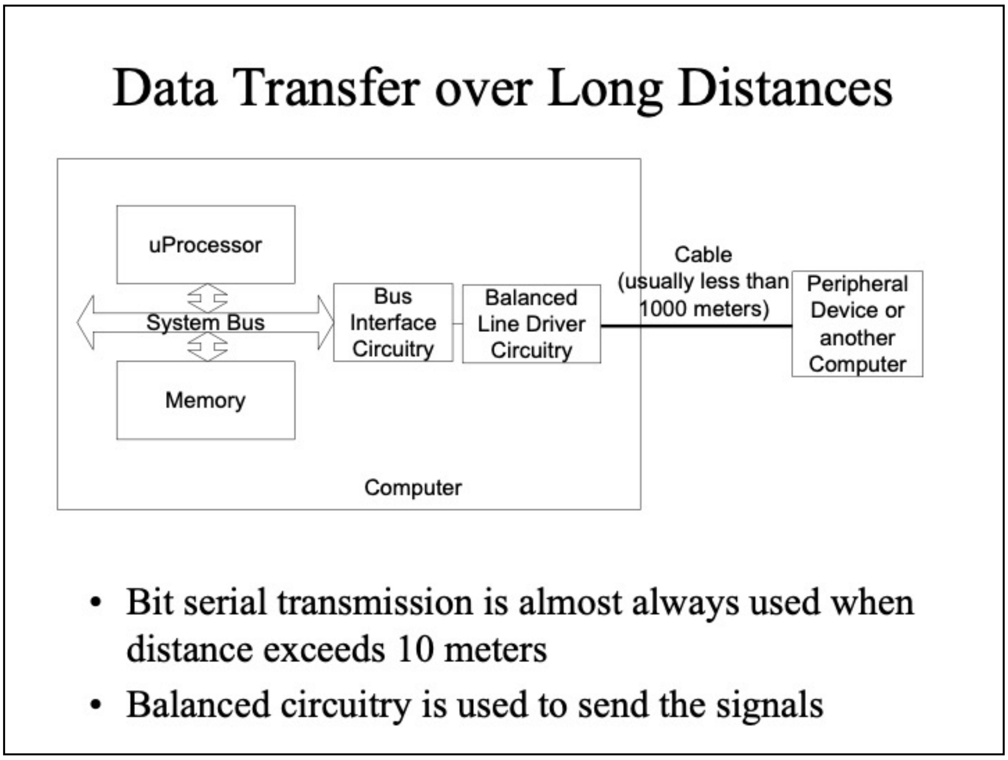
Noise Compensation



- For digital signals transmitted over short and medium distances

These effects are properties of transmission through metallic conductors, and become more pronounced as the conductor length increases. To compensate for distortion, signal power must be increased or the transmission rate decreased.

Special amplifier circuits are designed for transmitting direct (un-modulated) digital signals through cables. For the relatively short distances between components on a printed circuit board or along a computer backplane, the amplifiers are in simple IC chips that operate from standard +5v power. The normal output voltage from the amplifier for logic '1' is slightly higher than the minimum needed to pass the logic '1' threshold. Correspondingly for logic '0', it is slightly lower. The difference between the actual output voltage and the threshold value is referred to as the noise margin, and represents the amount of noise voltage that can be added to the signal without creating an error.

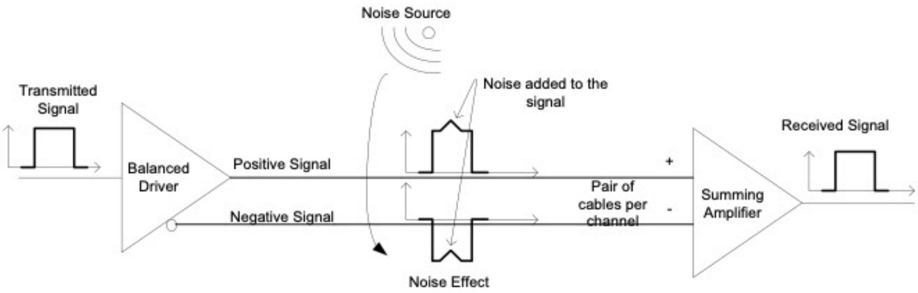


When relatively long distances are involved in reaching a peripheral device, driver circuits must be inserted after the bus interface unit to compensate for the electrical effects of long cables (noise and distortion).

This is the only change needed if a single peripheral is used. However, if many peripherals are connected, or if other computer stations are to be linked, a local area network (LAN) is required, and it becomes necessary to drastically change both the electrical drivers and the protocol to send messages through the cable. Because multi-conductor cable is expensive, bit-serial transmission is almost always used when the distance exceeds 20 feet.

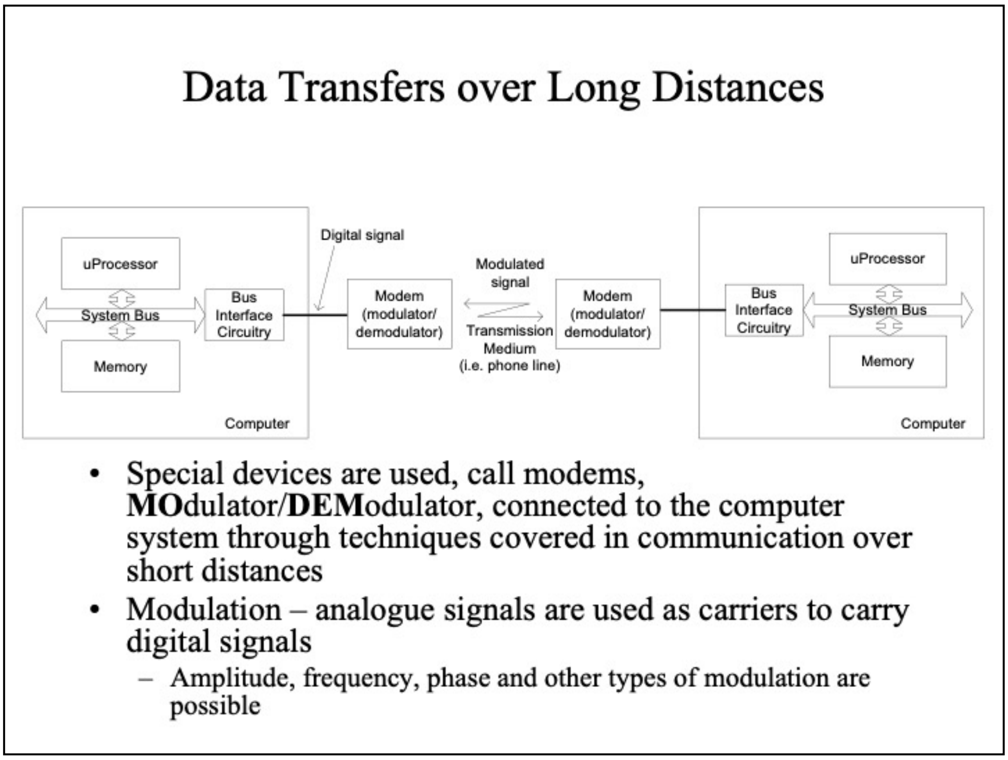
In either a simple extension cable or a LAN, a balanced electrical system is used for transmitting digital data through the channel. This type of system involves at least two wires per channel, neither of which is a ground. Note that a common ground return cannot be shared by multiple channels in the same cable as would be possible in an unbalanced system.

Balanced Line Transmission



- The signal is sent over two wires simultaneously
  - One wire is carrying the positive signal
  - One wire is carrying the negative (inverted) signal
- At the receiver, the signals are subtracted by a summing amplifier producing a stronger signal than any of the signals swinging on the individual lines
- When the signals are subtracted, any induced noise cancels

The basic idea behind a balanced circuit is that a digital signal is sent on two wires simultaneously, one wire expressing a positive voltage image of the signal and the other a negative voltage image. When both wires reach the destination, the signals are subtracted by a summing amplifier, producing a signal swing of twice the value found on either incoming line. If the cable is exposed to radiated electrical noise, a small voltage of the same polarity is added to both wires in the cable. When the signals are subtracted by the summing amplifier, the noise cancels and the signal emerges from the cable without noise.



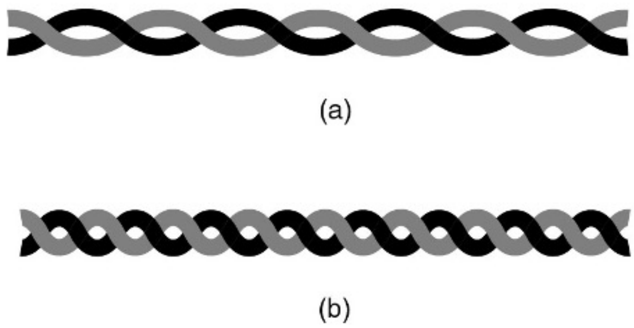
A good example of data communications over longer distance using copper wire are the use of the telephone network for your home internet connection. Transmissions over such distances are not generally accomplished with a direct-wire digital link, unless you have fibre to the home, but rather with digitally-modulated analog carrier signals, as these are easier to transmit intact over a bandwidth limited link to the ISP. This technique makes it possible to use existing phone lines for digital data, although at possibly reduced data rates compared to a direct digital link. Transmission of data from your home over phone lines requires that data signals be converted to modulated carrier waves by a modem. One or more sine wave carriers are used, and, depending on the baud rate and protocol, the modem will encode data by varying the frequency, phase, or amplitude of the carrier. The receiver's modem accepts the modulated sine wave and extracts the digital data from it. Several modulation techniques are typically used in encoding digital data for transmission, some of these will be looked at latter in this presentation.

Guided Transmission Data

- Twisted Pair
- Coaxial Cable
- Fiber Optics

Magnetic media is another (not related to networking) media used to transfer data.

Twisted Pair



(a) Category 3 UTP.
(b) Category 5 UTP.

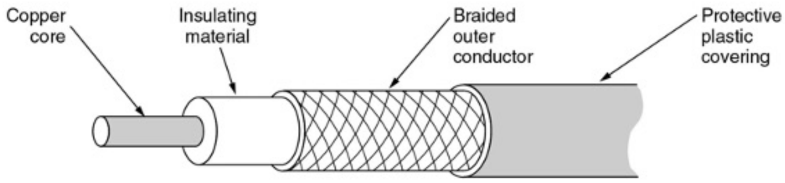
Consists of two insulated copper wires, typically about 1mm thick. The wires are twisted together in a helical form (just as DNA molecules). Twisting is done because two parallel wires constitute a fine antenna. When the wires are twisted, the waves radiated from different twists are canceling each other. The wires are usually bundled together and encased together in protective shields, when coming from a block of apartments to a phone company. If the wires wouldn’t be twisted, then the interference between the wires part of same bundle would be big.

Twisted pair wires can be used for transmission of both digital and analog signals. The bandwidth depends on the thickness of the wire and the distance, but usually several Mbps can be easily achieved.

Cat 3 UTP (16 MHz) – used until 1988 to wire telephone systems. Latter replaced by Cat 5 UTP and later variants (more twists per centimeter, resulting in less cross-talk and better-quality signals over longer distances), able to handle about 100Mhz. Later categories are 6 and 7, able to handle signals of 250MHz and 600 MHz and therefore higher data rates over longer distances.

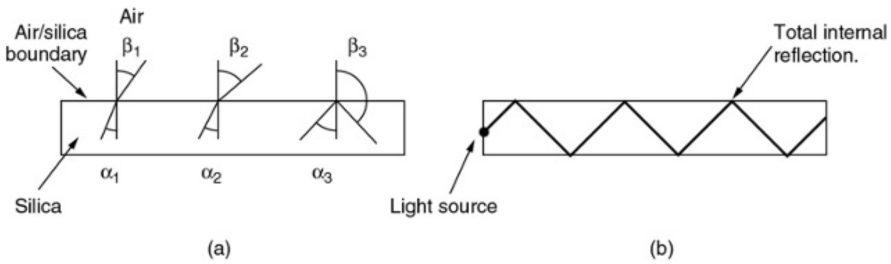
Coaxial Cable

A coaxial cable.



It is better shielded than twisted pair cable and it can span over longer distances at higher speeds. 50 ohms (for digital transmission) and 75 ohms (for analog transmission) are available. Due to the construction and the shielding process, the coax cables can have large bandwidth (up to 1GHz). They used to be largely used by phone companies for long distance lines, but now they have been replaced with fiber optics.

Fiber Optics



(a) Three examples of a light ray from inside a silica fiber impinging on the air/silica boundary at different angles.

(b) Light trapped by total internal reflection.

The optical fiber achievable bandwidth is in excess of 50 000 Gbps (50 Tbps). The current practical signaling limit is 10Gbps and it is not limited by the characteristics of the optical fiber but by our inability to convert electrical signals into optical signals any faster.

OPTICAL TRANSMISSION SYSTEM has three components: light source, the transmission medium and the detector. Conventionally a pulse of light indicates a 1 and absence of light indicates 0. Transmission medium is thin optical fiber, and the detector generates an electrical pulse when light falls on it. By attaching the source of light at one end of an optical fiber and the detector at the other end, we have a unidirectional data transmission system, that accepts electrical signal, converts it into light, transmits it over the optical fiber, it is received by the detector and transformed back into electrical signal.

The angle at which the light is injected into the optical fiber is very important. Because of refraction, the light can escape the optical fiber, or it can be “trapped” inside, with virtually no loss.

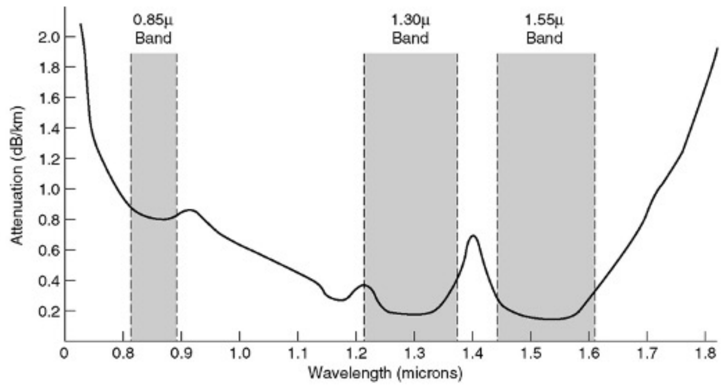
MULTIMODE FIBER – many different rays could bounce inside of the fiber, at different angles.

SINGLE MODE FIBER – the diameter of the fiber is reduced to a few wavelengths of light, which causes the light to travel (propagate) only in straight line.

Single mode fibers are more expensive and are used for transmission on very long distances. 50Gbps for over 100Km are possible, without any amplification.

Transmission of Light through Fiber

$$Attenuation_{dB} = 10\log_{10}\frac{TransmitedPower}{ReceivedPower}$$



Attenuation of light through fiber in the infrared region.

The attenuation of light through glass (the raw material used for optical fiber cable) depends on the wavelength of the light but also on the physical properties of the glass.

For example, an attenuation of 2 is given by 10log2 = 3dB

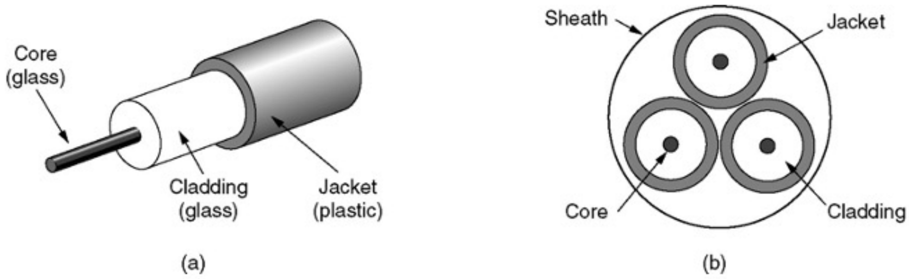
Visible light is from 0.4 to 0.7 microns (400 to 700 nm).

Three wavelength bands are used for communication, centered on: 0.85microns, 1.30microns and 1.55microns .

As the ray lights travel down the fiber, CHROMATIC DISPERSION is happening (the process of spreading of the wavelength). By making the pulses of a special shape, nearly all the dispersion effects can be canceled out. These pulses are called SOLITONS.

Fiber Cables (1)

(a) Side view of a single fiber.
(b) End view of a sheath with three fibers.



Multimode fibers, the core is 50 microns. For single mode fibers, the diameter of the core is 8 to maximum 10 microns.

Fibers can be connected in three ways: connectors, mechanically connected (spliced) and fused (melted) together.

Fiber Cables (2)

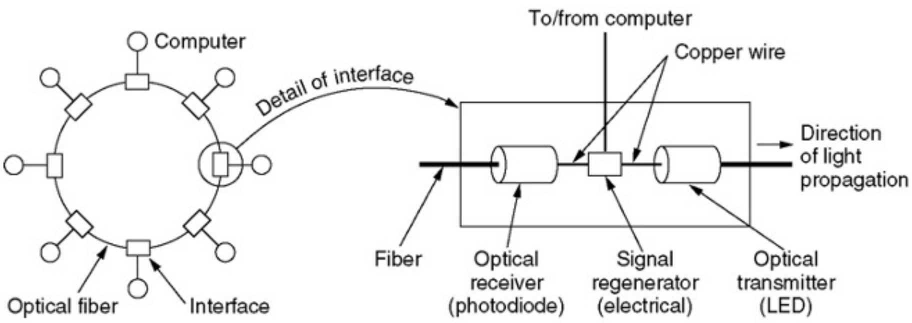
A comparison of semiconductor diodes and LEDs as light sources.

Item	LED	Semiconductor laser
Data rate	Low	High
Fiber type	Multimode	Multimode or single mode
Distance	Short	Long
Lifetime	Long life	Short life
Temperature sensitivity	Minor	Substantial
Cost	Low cost	Expensive

Two kind of light sources can be used: LED (Light Emitting Diode) and semiconductor lasers. The receiver is a photodiode, which gives and electrical pulse when stroke by light. The response time is usually 1ns, which limits the data rates at about 1Gbps.

Fiber Optic Networks

A fiber optic ring with active repeaters.



A ring network is just a collection of point to point links. The interface on each computer passes the light pulse stream through to the next link and also serves as a T junction to allow the computer to send and accept messages. The interfaces could be passive (tapping an LED and a photodiode on the fiber) or active repeater (presented in the figure). If an active repeater fails, the ring is broken and the network goes down. The passive interface is loosing signal, therefore, the light can’t travel to far, so it limits the size of the network.

Wireless Transmission

- The Electromagnetic Spectrum
- Radio Transmission
- Microwave Transmission
- Infrared and Millimeter Waves
- Lightwave Transmission

The Electromagnetic Spectrum (1)

- The number of oscillations per second of a wave is called its **frequency** (f), measured in Hz
- The distance between two consecutive maxima (or minima) is called wavelength (λ), measured in meters.
- In vacuum all electromagnetic waves travel at same speed – speed of light (300×10^6 m/s)
- In copper all waves travel at about 2/3 of the speed of light.

The Electromagnetic Spectrum (2)

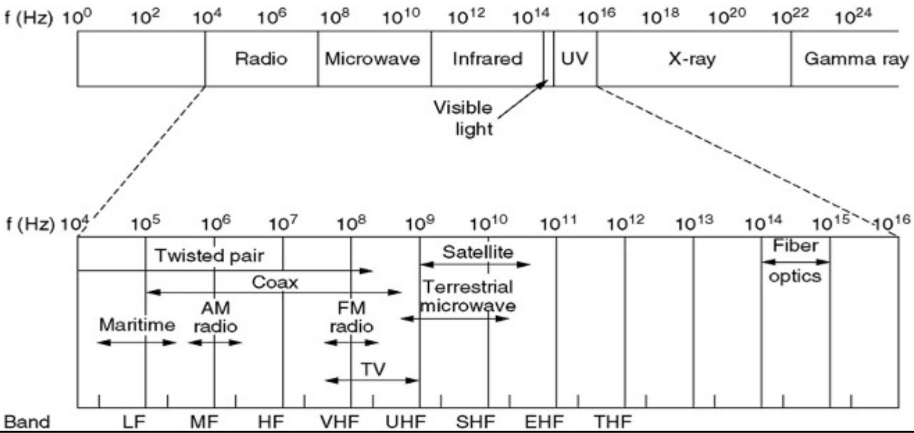
- The relation between the frequency and wavelength of an electromagnetic wave is given by the relation (in vacuum):

$$\lambda * f = c$$

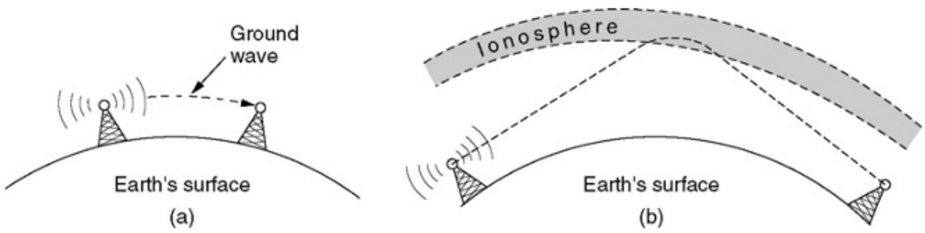
- Since c is a constant, if we know the frequency we can compute the λ and vice versa
- Example: compute the wavelength of a signal having 100MHz.
  - $100*10^6 * \lambda = 300 * 10^6$
  - $\lambda = 3$ meters

The Electromagnetic Spectrum (3)

The electromagnetic spectrum and its uses for communication.



Radio Transmission



(a) In the VLF, LF, and MF bands, radio waves follow the curvature of the earth.

(b) In the HF band, they bounce off the ionosphere.

Almost all transmission use a narrow frequency band (with few exceptions – frequency hopping spread spectrum and direct sequence spread spectrum). For the moment we assume all transmission do use a narrow frequency band.

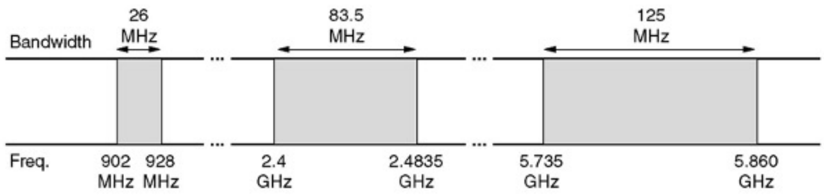
Microwave transmission

- Above 1000MHz the waves travel in straight line
- Repeaters are needed between two communication points, if too far apart
- Microwaves don’t pass through buildings
- Multipath fading problem
- Over 5GHz – the absorption by water problem
- It is widely used for long distance communication therefore a shortage of the available bandwidth has occured

Multipath fading refers to the fact that delayed waves may arrive out of phase with the direct wave and thus cancel the signal. Some operators keep their channels idle as spares to switch on when multipath fading wipes out some frequency band temporarily. It is a phenomena that is dependent on weather and on the operating frequency.

Politics of the Electromagnetic Spectrum

The ISM bands in the United States.



Due to radio waves ability to travel long distances, interference between users is a problem. For this reasons, governments license the use of radio transmitters.

ISM – Industrial, Scientific and Medical. The ISM bands is different somehow from country to country.

Infrared and Millimeter Waves

- Widely used for short range communication (i.e. used in remote controls)
- Directional and easy to build but don't pass through objects
  - This is also a plus – it means no interference with similar systems sitting in next room, no security issues, etc..
- No government license is needed to operate infrared systems
- It has some usage in interconnecting some devices as well (i.e. mice to computers, mobile phones to other equipment, etc…)

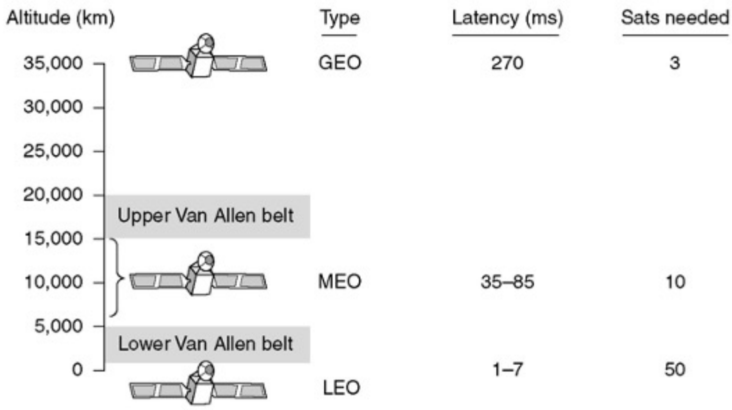
The more we go from long wave radio towards visible light, the electromagnetic wave behaves more and more like light and less and less like radio.

Communication Satellites

- Geostationary Satellites
- Medium-Earth Orbit Satellites
- Low-Earth Orbit Satellites

In its most simple form, a communication satellite can be seen as a big microwave repeater up in the sky. It contains several transponders, each of which listens to some portion of the spectrum, amplify the incoming signal and then rebroadcasts it to another frequency to avoid interference with the incoming signal. The downward beam can be broad (covering a large area of Earth's surface) or can be narrow (covering only a few hundreds of KM in diameter).

Communication Satellites (1)



Communication satellites and some of their properties, including altitude above the earth, round-trip delay time and number of satellites needed for global coverage.

Issues with satellites:

1. Orbital period – the highest the satellite, the larger the period (at an altitude of about 36 800 KM, the period is about 24 hours, at 384 000 the period is about one month)
2. Presence of the van Allen belts – layers of highly charged particles trapped by earth's magnetic field. Any satellite flying inside of those belts would be destroyed quickly.

GEO – Geostationary Earth Orbit satellites that sit at about 35 800 in a circular equatorial orbit (Arthur C. Clarke – science fiction author – computed that at this altitude, the satellite would appear to remain motionless in the sky).

MEO – Middle Earth Orbit satellites – between the two van Allen belts

LEO – Low Earth Orbit satellites

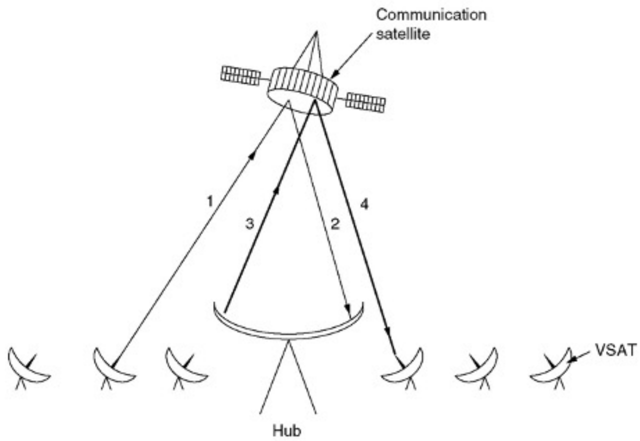
Communication Satellites (2)

The principal satellite bands.

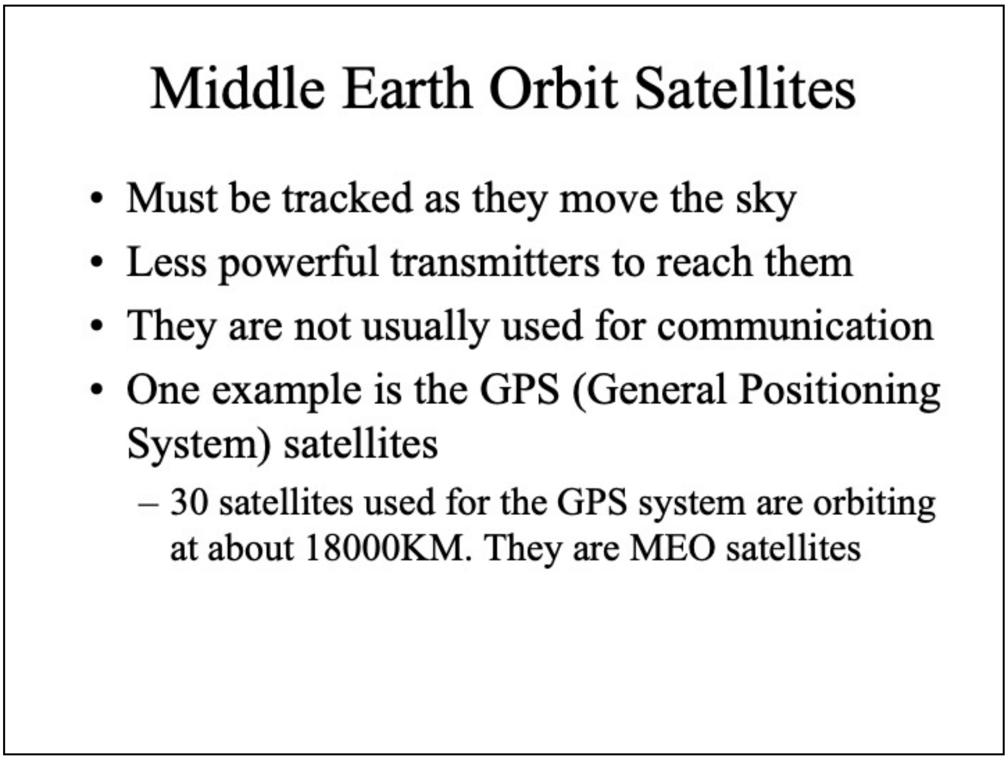
Band	Downlink	Uplink	Bandwidth	Problems
L	1.5 GHz	1.6 GHz	15 MHz	Low bandwidth; crowded
S	1.9 GHz	2.2 GHz	70 MHz	Low bandwidth; crowded
C	4.0 GHz	6.0 GHz	500 MHz	Terrestrial interference
Ku	11 GHz	14 GHz	500 MHz	Rain
Ka	20 GHz	30 GHz	3500 MHz	Rain, equipment cost

Communication Satellites (3)

VSATs using a hub.



VSAT – VERY SMALL APPERTURE TERMINALS – 1 meter or smaller compared to 10 meters required for GEO satellites communication.

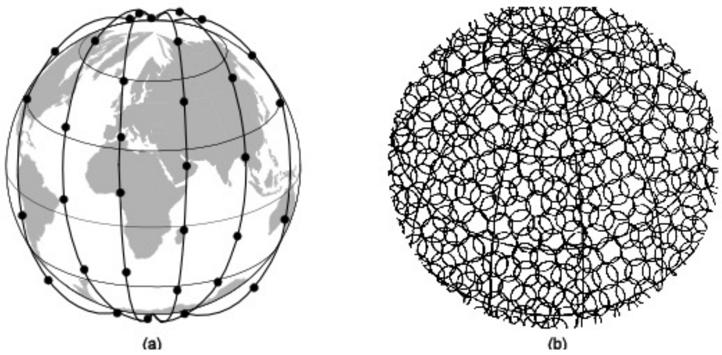


GPS needs a minimum of 24 to operate and the typical number active is about 30. GPS receivers released since 2018 have much higher accuracy, pinpointing to within 30 centimeters.

Low Earth Orbit Satellites

- Due to their motion, large number of them are needed for a complete coverage
- Because are low orbit, the ground stations don't need much power and the round-trip delay is only a few milliseconds
- Examples
  - Iridium
  - Starlink

Iridium



(a) The Iridium satellites from six necklaces around the earth.

(b) 1628 moving cells cover the earth.

In 1990 Motorola started so called Iridium project. Its goal was to use 77 low orbit satellites (the element 77 is Iridium) to cover completely the whole surface of the Earth. When one satellite went out of view, another one would replace it. Latter, the project was revised and instead 77 satellites, only 66 where used.

The satellites were launched in 1997 and the service begun in 1998. Iridium wasn’t profitable, being one of biggest fiascos in history (this partially because of the cellular telephony developing so rapidly).

The Iridium service was restarted in March 2001, providing worldwide communication services using hand-held devices communicating directly with satellites. It provides voice, data, fax and navigation services everywhere on land, water or air.

Iridium satellites are positioned at 750KM altitude, in circular polar orbits. They are arranged in North-South necklaces, with one satellite very 32 degrees of latitude. With six satellite necklaces, the entire earth is covered.

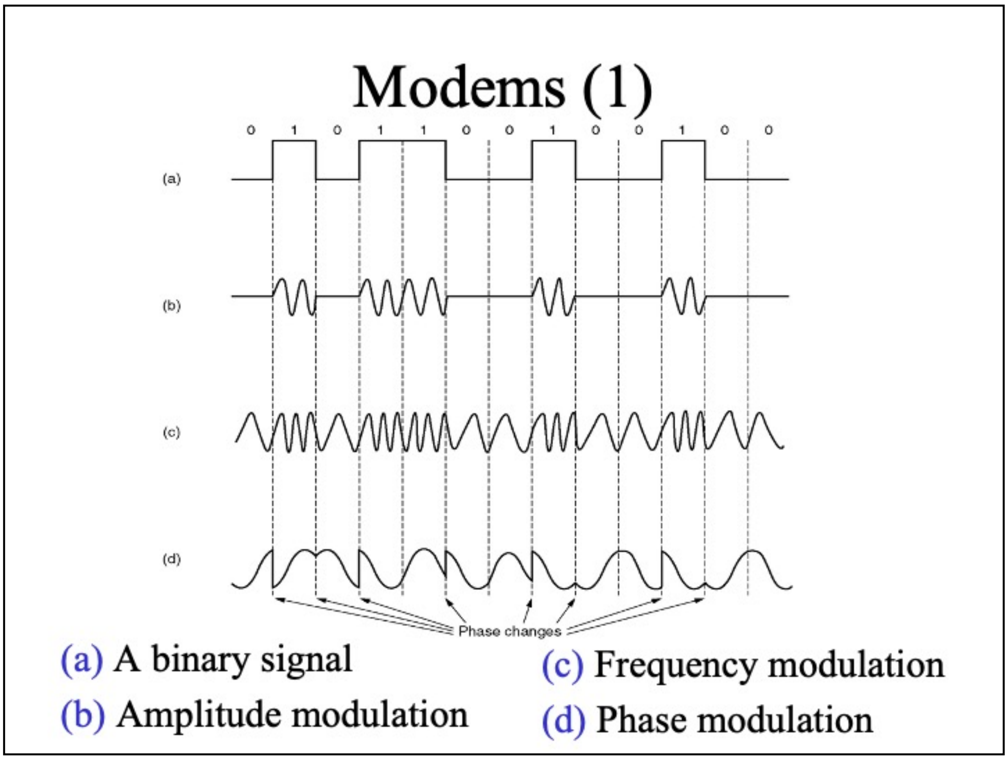
Each satellite has a maximum of 48 cells (spot beams) with a total of 1628 cells over the surface of the earth. Each satellite has a capacity of 3840 low bandwidth phone channels.

Modulation (1)

- Due to attenuation (and other discussed problems) square waves (digital signal) can't be used for long distance transmission
- AC signaling is used instead (continuous tone, around 1000 – 2000 Hz, called sin wave carrier) is introduced
- Its amplitude, frequency or phase can be **modulated** to transmit information

Modulation (2)

- **Amplitude modulation**
  - Two different amplitudes are used to represent 0 or 1
- **Frequency modulation** (or frequency shift keying)
  - Two (or more) different tones are used
- **Phase modulation**
  - The carrier wave is shifted 0º or 180º at bit intervals, to show a transition.
  - A better scheme is to use shifts of 45, 135, 225 or 315 º to transmit two bits of information



A modem is a device that accepts a serial stream of bits as input and produces a carrier modulated by one (or more) of these methods (modulator/demodulator). The modem is inserted between the (digital) computer and the (analog) telephone system.

- – binary signal
- - amplitude modulation using absence of carrier to encode a ‘0’ and presence to encode a ‘1’
- - frequency modulation using waves with different frequencies to encode ‘0’ and ‘1’
- - phase modulation using phase shifts to show a change from ‘0’ to ‘1’ and ‘1’ to ‘0’. No phase shift means the signal stays the same (0 or 1 during the bit intervals).

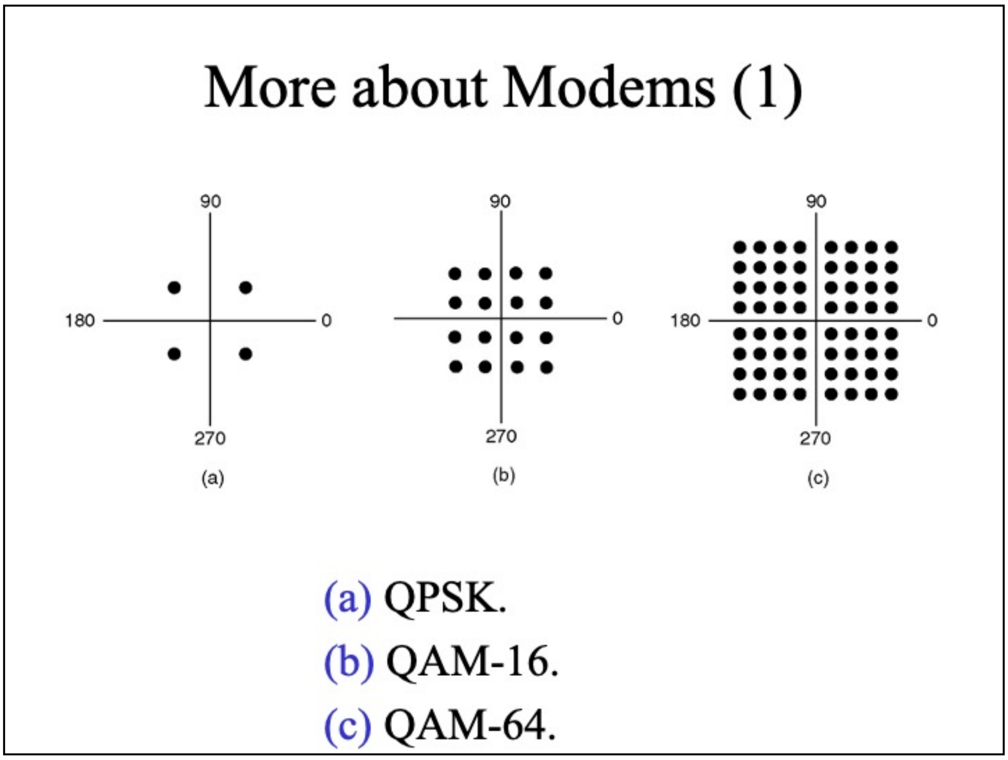
Modems (2)

- A phone line has a band of 3000Hz, no point in sampling the line at faster rates than 6000 times per second (according to Nyquist theorem)
  - In practice most modems sample at 2400 times per seconds and focus in getting more bits per sample
- The number of samples per second is measured in **baud**. During one baud one symbol is sent
- One symbol can consist of one or more bits

N baud line transmits n symbols per second. A 2400 bps line transmits one symbol about every 416.667 microseconds. If the symbol consists of one bit, than the line is 2400 bps. If the symbol consists of 2 bits per symbol, then the bit rate is 4800 bits per second.

Bandwidth, baud rate and bit rate

- Bandwidth of a medium is the range of frequencies that pass through with minimum attenuation
  - Physical property of the medium (0 to some maximum frequency, measured in Hz)
- Baud rate – number of samples made per second
  - Per sample is sent out one symbol, therefore the baud rate and symbol rate are the same
- Bit rate – amount of information sent over the channel and is equal to the number of symbols/sec times the number of bits/symbol



QPSK – QUADRATURE PHASE SHIFT KEYING – four phases are used

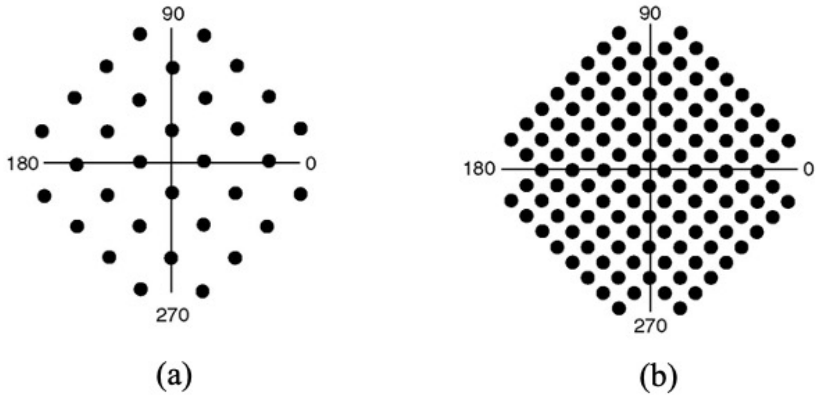
QAM-16 – QUADRATURE AMPLITUDE MODULATION – four amplitudes and four phases to form a total of 16 possible combinations. In other words, a symbol can encode 4 bits. 9600 bits per second over a 2400 baud line

QAM-64 – QUADRATURE AMPLITUDE MODULATION – a total of 64 possible combinations. In other words, a symbol can encode 6 bits.

Higher order QAMs are also used.

Those diagrams are called CONSTELLATION DIGARAMS – show the legal combinations of amplitude and phase. Each modem is using its own constellation diagram and can talk only with modems implementing same constellation diagrams.

More about Modems (2)



(a) V.32 for 9600 bps.

(b) V32 bis for 14,400 bps.

With many points in the constellation diagram, even a small amount of noise would result in transmission errors.

TCM (TRELLIS CODED MODULATION) is used, where beside the data, a symbol contains parity checking.

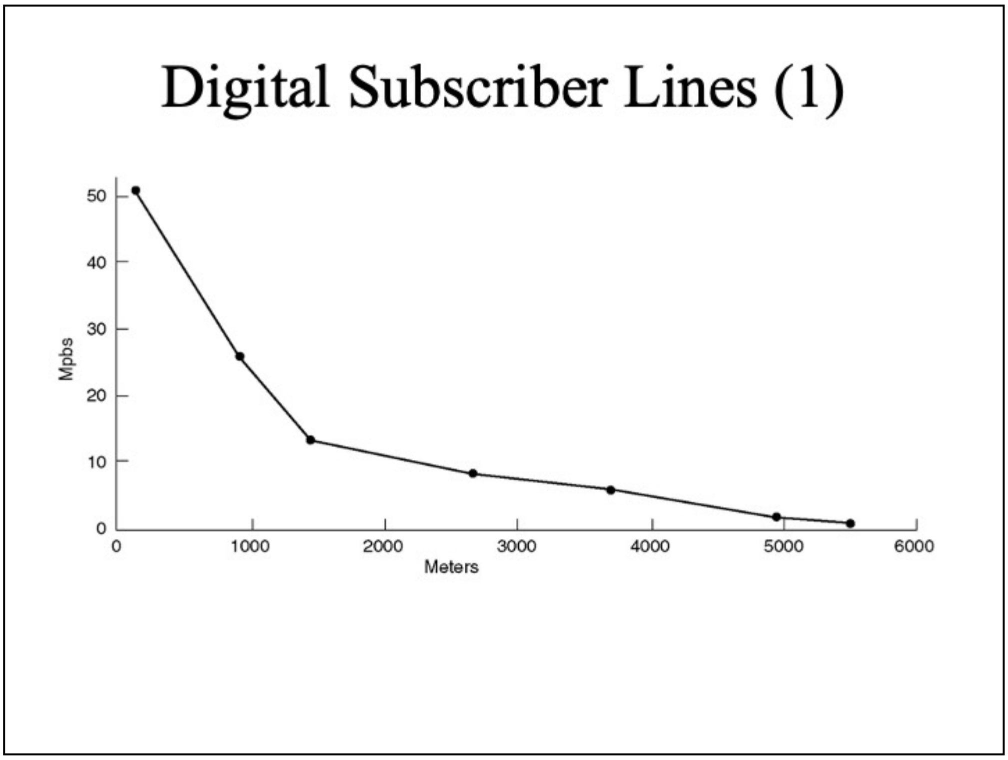
The V32 modem is using 32 points constellation to send 4 bits of data and 1 bit of parity per each symbol, at an effective bit rate of 9600 bps.

Next step up from **V32** is **V32.bis**, where 128 points constellation is used. 6 bits for data and one for parity in a 7 bit symbol, at a line of 2400 baud. It achieves 14400 bps effective data bit rate.

V34 runs at 28800 at 2400 baud channel with 12 data bits/symbol.

V34.bis runs at 33600 over a 2400 baud line, using 14 data bits/symbol.

Modems are bidirectional devices, by using different frequencies for different directions. Bidirectional simultaneously connections are called full duplex connections.



Services with more bandwidth than standard telephone services are called BROADBAND (more marketing than technical).

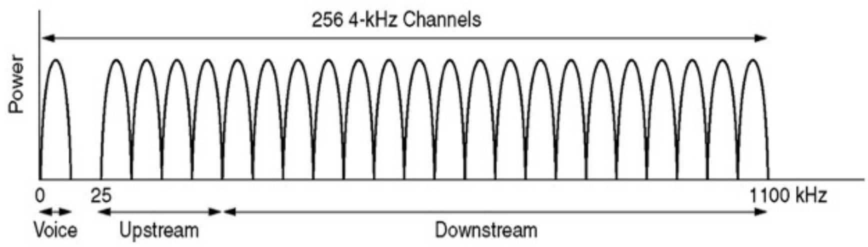
DSL (DIGITAL SUBSCRIBER LINE) services started to be offered by phone companies, to be competitive on the broadband services market (TV and satellite companies were already offering broadband services).

The reason that modems are so slow is because phone lines were invented to carry human voice not data. Therefore, frequencies bellow 300 HZ and over 3100Hz were artificially filtered out. Cutoff filters were added on the local loops to filter out other frequencies than voice. The cutoff is not sharp, so the bandwidth is usually quoted to be 4000Hz even thought the distance between the 3dB points is 3100 Hz. Data is thus restricted to this narrow band.

The trick that made DSL (or ADSL – Asymmetric DSL) work is that when a customer subscribes to it, the incoming line is connected to a different kind of switch (in the end office), one that doesn't have this voice filter installed, thus making the entire capacity of the local loop available.

The capacity of the local loop depends on the length and thickness of the cable and an plot of it is presented here. Therefore the ADSL services are limited to customers that are within certain radius from the end office (phone company switch).

Digital Subscriber Lines (2)

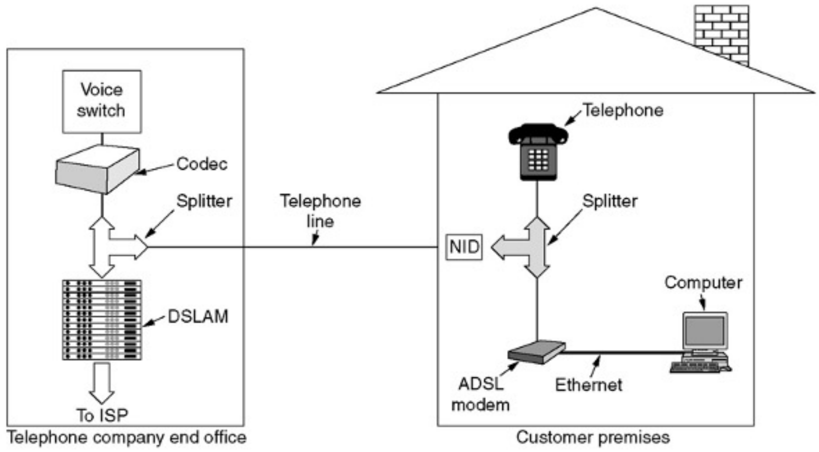


Operation of ADSL using DMT (**Discrete Multi Tone**) modulation.

The available spectrum of the local loop (which is 1.1 MHz for a decent service) is divided into 256 channels of 4312.5 Hz each. Channel 0 is used for POTS (Plain Old Telephone Service). Channels 1 to 5 are not used, to make sure that there is enough guard between voice and data. The remaining 250 channels are used for data and stream control (one is used for upstream control and one is used for downstream control). It is up to the provider how many channels will be used for upstream and how many for downstream. A 50% 50% mix for upstream and downstream is possible, but usually the providers allocate more for downstream than for upstream (80% to20%, since most of the subscribers are using the line for Internet, so they pull pages down onto their system. This is why DSL service turned into ADSL service). A common split is 32 channels for upstream and the rest for downstream.

Within each channel, a modulation schema similar to V.34 is used, although the sampling rate is 4000 baud instead of 2400. The line quality in each channel is constantly monitored and the data rate adjusted continuously as needed, so different channels may have different data rates. The data is sent with QAM modulation with a maximum of 15 bits per symbol. With 224 downstream channels, the downstream data bit rate is about 13.4 Mbps. In practice, the signal to noise is never good enough to achieve this bit rate, but 8 Mbps is possible on short local loops, and the ADSL standard goes this far.

Digital Subscriber Lines (3)



A telephone company technician must install a NID (NETWORK INTERFACE DEVICE) in the customer's house. Close or combined with the NID, a SPLITER is installed, which is an analog filter that separates the phone signal (0 to 4000Hz) from the data signal (over 26kHz).

The ADSL modem is actually a signal processor that acts as 250 QAM modems, operating in parallel at different frequencies available for each channel. Usually, the connection between the modem and the computer is Ethernet based.

At the other end (End Office) a special DSLAM (DIGITAL SUBSCRIBER LINE ACCESS MULTIPLEXER) receives the data over 26 KHz (separated by the splitter) and recovers the bit stream from the data (same function as the ADSL modem at the customer end). The low frequency signal (POTS signal) is sent to a conventional switch.

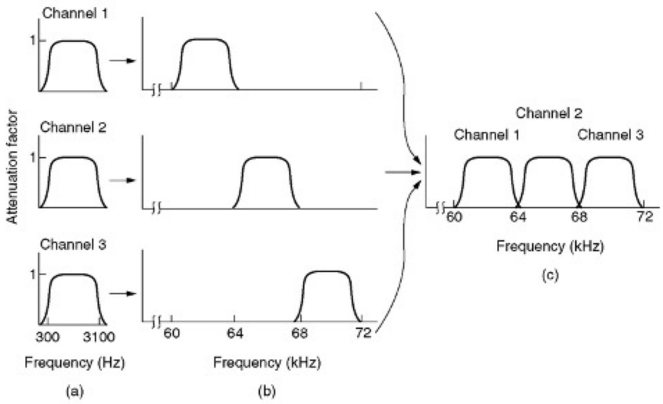
Trunks and data Multiplexing

- FDM (Frequency Division Multiplexing)
  - Wavelength division multiplexing
- TDM (Time Division Multiplexing)

High bandwidth trunks are available between switching offices. The data collected from end loops needs to be multiplexed over those high bandwidth trunks. There are two main categories for data multiplexing: FDM and TDM.

In FDM the frequency spectrum is divided into frequency bands, with each user having exclusive possession of some band, while in TDM, the users take turn (in a round robin fashion), each one getting periodically the entire available bandwidth for a short period of time.

Frequency Division Multiplexing



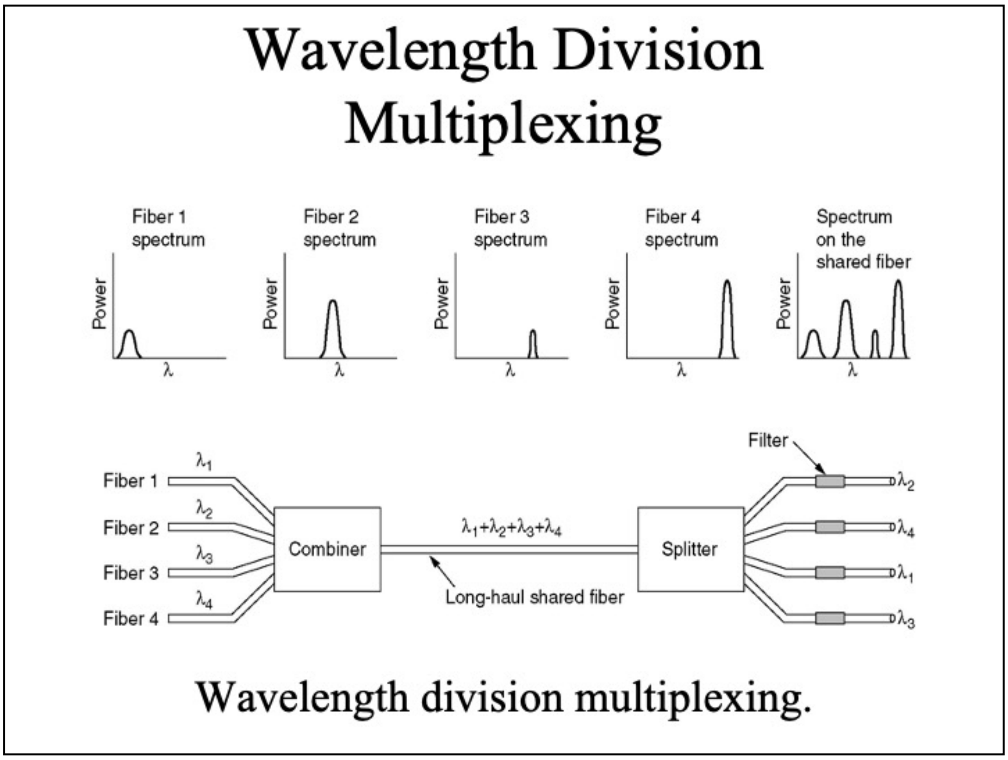
(a) The original bandwidths.

(b) The bandwidths raised in frequency.

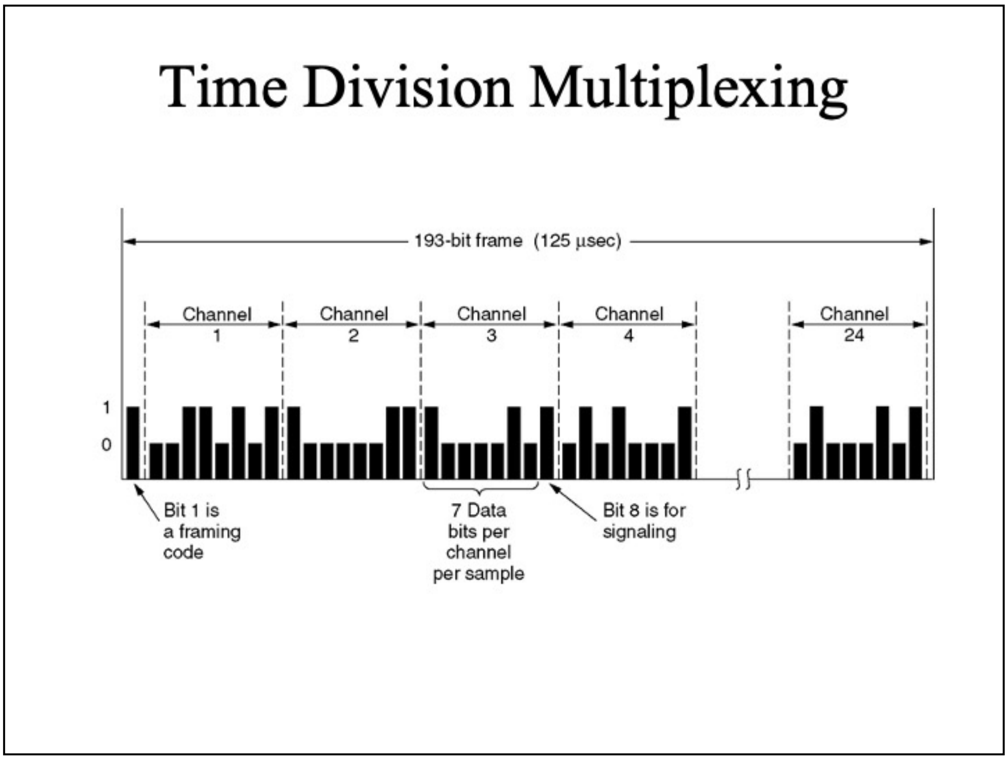
(b) The multiplexed channel.

This slide presents how voice grade telephone channels are multiplexed using FDM. When many channels are multiplexed together, 4KHz bandwidth is allocated to each channel, to keep them separated. First, each channel is raised in frequency, each by a different amount. Then they can be combined, because no channel will occupy the same portion of the spectrum.

FDM schemas are to some degree standardized – a wide spread standard is 12 4KHz voice channels multiplexed into 60 to 108kHz bad. This is called a **GROUP**. Five groups can be multiplexed to form a **SUPERGROUP**. The next group is the **MASTERGROUP** which is made out of 5 supergroups.



It is a variation of the FDM at very high frequencies and it is used over the optical fiber trunks. Systems with 96 channels of 10Gbps each are available in production. 960 Gbps is enough to send 30 full time movies per second…



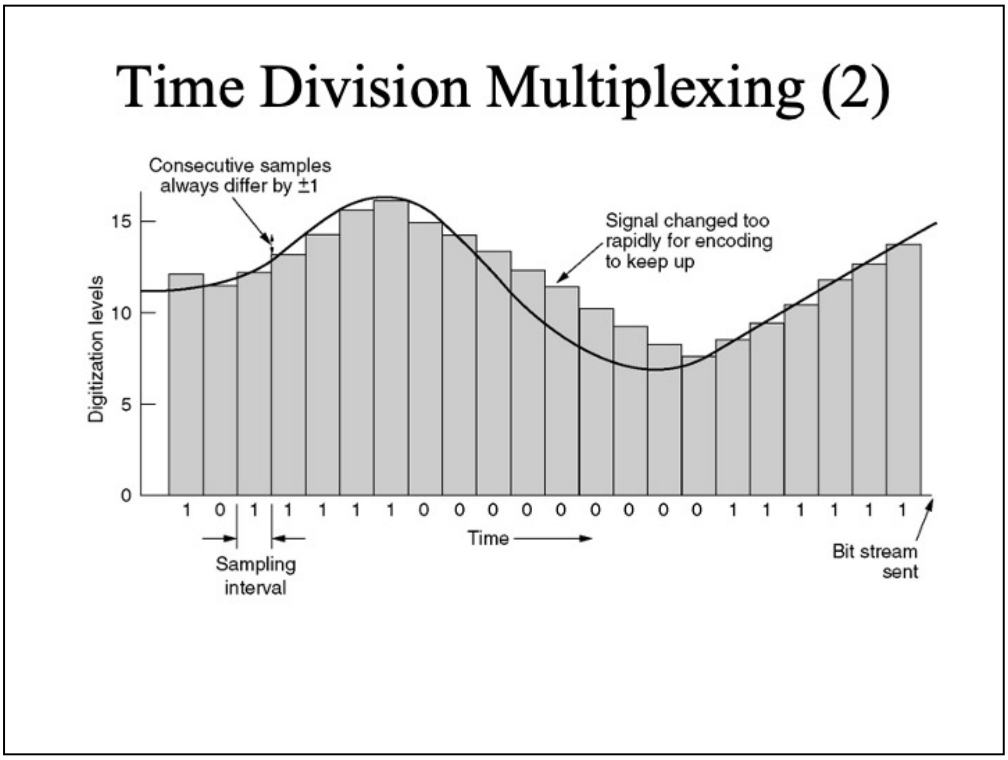
How multiple analog voice telephone channels are digitized and combined onto a single outgoing digital trunk.

The analog data is digitized in the end office by a device called **CODEC**. The codec makes 8000 samples per second (125 us per sample, according to Nyquist theorem it is enough to sample at twice the bandwidth of the signal to capture all the information carried by the signal) and measures the amplitude of each sample with a precision of 8 bits. This process is called PCM (**PULSE CODE MODULATION**). PCM forms the heart of the modern telephone service. As a consequence, all time intervals within a phone system are multiples of 125 us.

Incompatible schemas are in use in US and Europe. Japan and US are using the T1 carrier (presented in this slide) while in Europe E1 carrier is used.

Each T1 frame accommodates 24 channels (sampled each 8000 times/second, 7 bit per sample), giving a gross data rate of 1.544 Mbps. An extra bit is used for framing control. This bit is changing value from 0 to 1 and from 1 to 0 every frame. The changing patter is 01010101… and it is used by the receiver to make sure it is not loosing synchronization. Analog customers can't generate this wave, since it corresponds to a sin wave at 4000 Hz (this frequency is not present in the analog voice channels).

E1 has 32 channels, 8 bit per channel packed into the basic 125 us frame. This yields to a gross data rate of 2.048 Mbps.

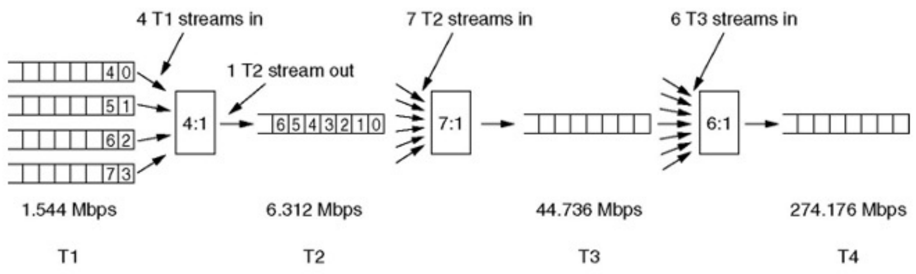


Once the voice channel has been digitized, it is tempting to use statistical techniques to obtain fewer bits per sample (therefore lower bandwidth required to carry the voice information). All the compaction methods are based on the fact that the analog voice signal is changing relatively **SLOWLY** compared to the sampling frequency, so much in the information in the 7th and 8th bit is redundant.

DPCM (DIFERENTIAL PULSE CODE MODULATION) is based on producing as output the difference between the current value and the previous one. Since jumps of +/- 16 in a scale from 0 to 128 are unlikely, only 5 bits are enough to represent the audio signal. If the signal does occasionally jump sharply, than a few samples are required to catch up with the new value. For speech, this kind of error can be ignored.

DELTA MODULATION is a variation of DPCM and requires the current value to differ only with + or -1 from the previous value. Under those conditions, only a bit can be sent to indicate if the next value is over or below the previous value.

Time Division Multiplexing (3)



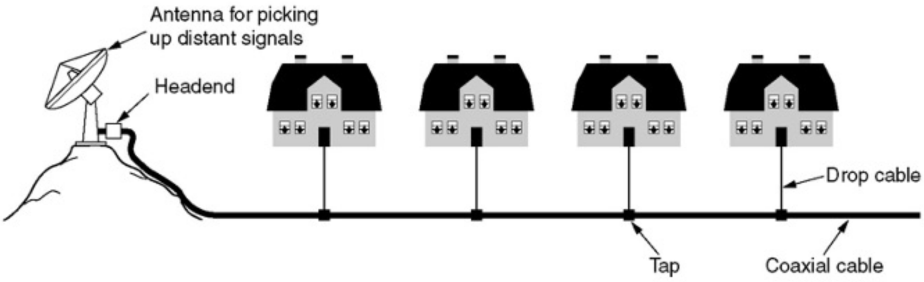
Multiplexing T1 streams into higher carriers.

Cable Television

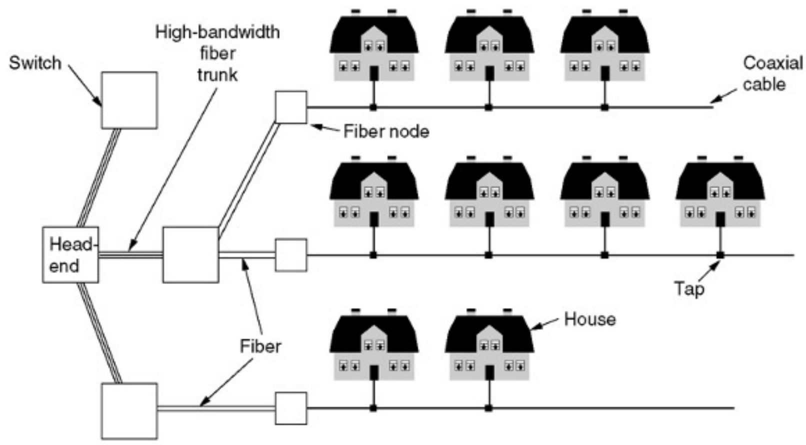
- Community Antenna Television
- Internet over Cable
- Spectrum Allocation
- Cable Modems

Community Antenna Television

An early cable television system.

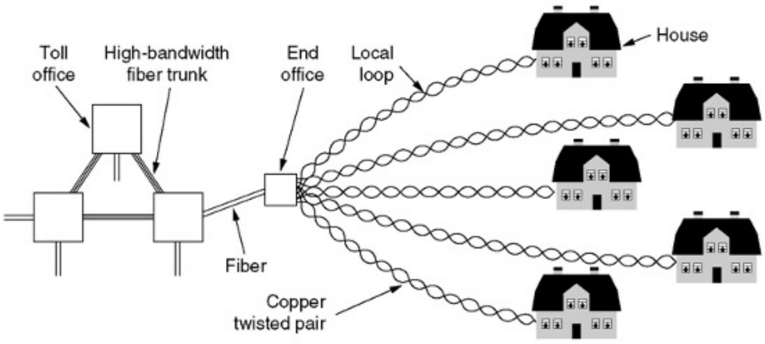


Internet over Cable



(a)

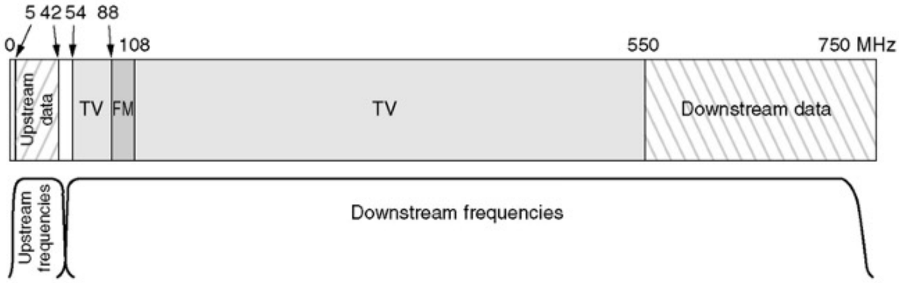
Internet over Cable (2)



(b)

Spectrum Allocation

Frequency allocation in a typical cable TV system used for Internet access



Cable Modems

Typical details of the upstream and downstream channels