

UNIT 2

The Physical Layer

(6 Hours) — Computer Networking (CACS303)

2.1 Functions of the Physical Layer

The Physical Layer is the lowest layer of the OSI reference model. It is responsible for the actual transmission and reception of raw, unstructured bit streams over a physical transmission medium connecting two devices. Everything at this layer is concerned with mechanical, electrical, timing, and procedural characteristics of the physical link — it does not understand or interpret the meaning of the bits it carries.

- Physical characteristics of interfaces and media: defines properties of the connectors, pins, and cables (e.g. shape of a connector, number of wires).
- Representation of bits: decides how 0s and 1s are encoded into signals — e.g. voltage levels for electrical signals, light pulses for optical fiber, or radio wave modulation for wireless links.
- Data rate: defines the transmission rate, i.e. the number of bits sent per second, and how long each bit lasts.
- Synchronization of bits: the sender and receiver clocks must be synchronized at the bit level so that the receiver knows exactly when one bit ends and the next begins.
- Line configuration: defines how devices are connected to the medium — point-to-point (two devices share one dedicated link) or multipoint (several devices share a single link).
- Physical topology: defines how devices are arranged — mesh, star, bus, ring, or hybrid.
- Transmission mode: defines the direction of flow between two connected devices — simplex, half-duplex, or full-duplex.

Key Takeaways — 2.1

- ✓ Physical layer moves raw bits over a physical medium; it has no notion of frames, addresses, or meaning.
- ✓ Its responsibilities: bit representation, data rate, synchronization, line configuration, topology, and transmission mode.
- ✓ It is purely concerned with hardware-level, electrical/optical/radio transmission.

2.2 Data and Signals

2.2.1 Analog and Digital Signals

Data can be represented either as analog or digital quantities, and signals that carry data can likewise be analog or digital.

- Analog data: takes on continuous values within some range (e.g. sound, temperature).
Analog signal: has infinitely many levels of intensity over a period of time; as the wave moves from one value to another, it passes through an infinite number of intermediate values.
- Digital data: takes on discrete values (e.g. text stored as ASCII, or binary computer data).
Digital signal: can have only a limited number of defined values — often just two levels, representing binary 0 and 1.

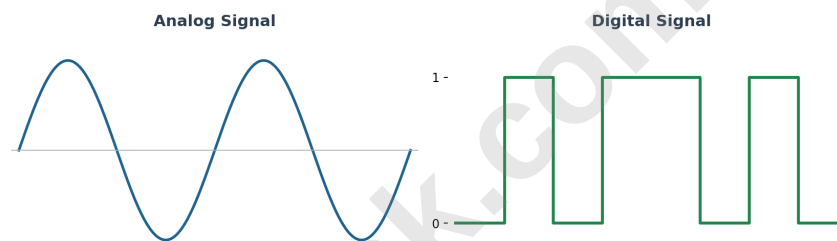


Fig. 2.1 — An analog signal (continuous wave) vs. a digital signal (discrete levels).

Note: A digital signal is a special case of an analog signal with a finite number of levels; an analog signal, at the physical level, always carries a digital signal when it is transmitted through a medium such as wire or air.

Periodic vs. Aperiodic Signals

- Periodic signal: completes a pattern within a measurable time frame, called a period, and repeats that pattern for subsequent identical periods (e.g. a sine wave).
- Aperiodic (non-periodic) signal: changes without exhibiting a repeating pattern; used to send digital data where each bit sequence is generally unique.

Characteristics of a Periodic Analog Signal (Sine Wave)

Parameter	Meaning	Unit
Amplitude	Peak value (height) of the signal at any given point in time	Volts (V)
Frequency (f)	Number of periods (cycles) completed in one second	Hertz (Hz)
Period (T)	Time required to complete one full cycle; $T = 1 / f$	seconds (s)

Parameter	Meaning	Unit
Phase	Position of the waveform relative to time zero (starting point of the cycle)	degrees / radians
Wavelength	Distance a signal travels in one period through a medium	meters (m)

Example 2.1

A sine wave completes one cycle in 0.02 seconds. What is its frequency?

Solution

$$f = 1 / T = 1 / 0.02 = 50 \text{ Hz.}$$

Composite Signals and Fourier Analysis

A single-frequency sine wave is not very useful in data communications; we need to send a composite signal, made of many simple sine waves with different frequencies, amplitudes, and phases. According to Fourier analysis, any composite signal can be represented as a combination of simple sine waves with different frequencies, amplitudes, and phases. The set of frequencies a composite signal contains is its frequency spectrum. The bandwidth of a composite signal is the difference between the highest and lowest frequencies contained in that signal.

2.2.2 Transmission Impairment

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment — the signal at the beginning of the medium is not the same as the signal at the end. Three impairments commonly affect the quality of a signal: attenuation, distortion, and noise.

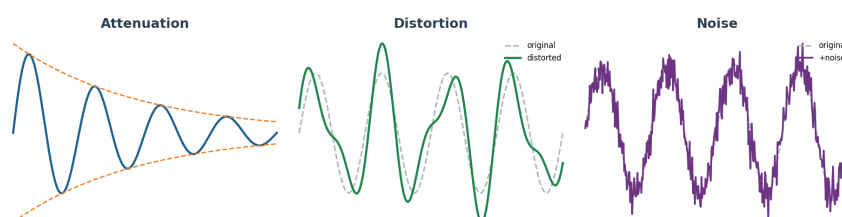


Fig. 2.2 — Attenuation (signal weakening), distortion (change in waveform shape), and noise (random unwanted signal added).

- **Attenuation:** means loss of signal energy as it travels through a medium; when a signal travels a long distance it loses some of its energy to the resistance of the medium. Amplifiers are used to compensate for this loss, and loss/gain is measured in decibels (dB): $\text{dB} = 10 \log_{10}(P_2/P_1)$.
- **Distortion:** means the signal changes its form or shape; it occurs mainly in composite signals made of different frequencies, where each frequency component has its own

propagation speed, causing a difference in delay and hence a change in the shape of the composite signal at the receiver.

- **Noise:** refers to random, unwanted electrical or electromagnetic energy that degrades the quality of signals. Types include thermal noise (random motion of electrons), induced noise (from motors/appliances), crosstalk (effect of one wire on another), and impulse noise (spikes from power lines or lightning).

Signal-to-Noise Ratio (SNR)

SNR is the ratio of the average signal power to the average noise power, expressed in decibels: $\text{SNR(dB)} = 10 \log_{10}(\text{SNR})$. A high SNR means the signal is less corrupted by noise, while a low SNR means the signal is heavily corrupted.

2.2.3 Data Rate Limits

The maximum data rate of a channel depends on three factors: the bandwidth available, the number of signal levels used, and the quality (noise level) of the channel. Two theoretical formulas define this limit for noiseless and noisy channels.

Nyquist Bit Rate (Noiseless Channel)

For a noiseless channel, the Nyquist formula defines the theoretical maximum bit rate: $\text{BitRate} = 2 \times B \times \log_2(L)$, where B is the bandwidth of the channel, and L is the number of signal levels used to represent data.

Example 2.2

A noiseless channel has a bandwidth of 3000 Hz, transmitting a signal with 2 signal levels. What is the maximum bit rate?

Solution

$$\text{BitRate} = 2 \times 3000 \times \log_2(2) = 2 \times 3000 \times 1 = 6000 \text{ bps.}$$

Shannon Capacity (Noisy Channel)

For a noisy channel, Claude Shannon's formula gives the theoretical highest data rate, called the channel capacity: $\text{Capacity} = B \times \log_2(1 + \text{SNR})$, where B is the bandwidth and SNR is the signal-to-noise ratio.

Example 2.3

A telephone line normally has a bandwidth of 3000 Hz and an SNR of 3162. What is the channel capacity?

Solution

Capacity = $B \times \log_2(1 + \text{SNR}) = 3000 \times \log_2(3163) \approx 3000 \times 11.6 \approx 34,860$ bps.

Note: Shannon's formula gives the theoretical upper limit; the Nyquist formula tells us how many signal levels we would need to reach that rate in practice.

2.2.4 Performance

Performance of a network is evaluated using several measures that describe how well the network delivers data.

Measure	Description
Bandwidth	Range of frequencies a channel can carry, or a measure of the capacity of a link (bps)
Throughput	The actual rate at which data is successfully delivered over a channel, usually lower than bandwidth
Latency (Delay)	Time required for a message to travel from sender to receiver; includes propagation, transmission, queuing, and processing delay
Jitter	Variation in packet arrival/delay times, important for real-time traffic such as voice and video
Bandwidth-delay product	Number of bits that can fill the link between sender and receiver (bandwidth $\times$ delay)

Key Takeaways — 2.2

- ✓ Signals may be analog (continuous) or digital (discrete); sine waves are described by amplitude, frequency, phase, and period.
- ✓ Transmission impairments — attenuation, distortion, and noise — degrade signal quality as it travels.
- ✓ Nyquist gives maximum bit rate for a noiseless channel; Shannon gives channel capacity for a noisy channel.
- ✓ Performance is measured through bandwidth, throughput, latency, jitter, and bandwidth-delay product.

2.3 Data Transmission Media

Transmission media carry signals from sender to receiver. They are broadly classified into guided (wired) media, which provide a physical conduit, and unguided (wireless) media, which transmit electromagnetic waves through free space.

2.3.1 Guided Media

Guided media provide a conduit from one device to another and include twisted-pair cable, coaxial cable, and fiber-optic cable.

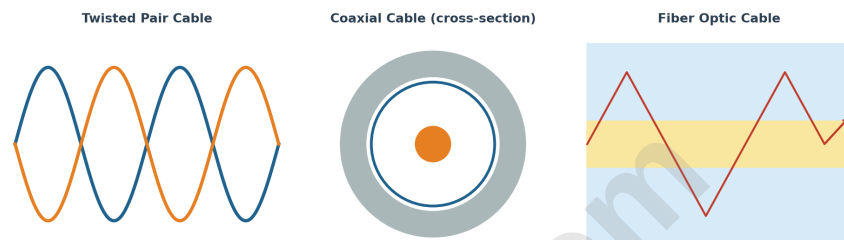


Fig. 2.3 — Twisted-pair cable, coaxial cable cross-section, and fiber-optic cable.

Media	Structure	Characteristics
Twisted-Pair Cable	Two insulated copper wires twisted together to reduce interference (crosstalk)	Low cost, easy to install; used in telephone lines and LANs (UTP/STP); limited bandwidth and distance
Coaxial Cable	A central copper core, insulation, metallic shield, and outer plastic cover	Higher bandwidth and better noise immunity than twisted pair; used for cable TV and older Ethernet
Fiber-Optic Cable	A glass or plastic core that carries light signals, surrounded by cladding and jacket	Very high bandwidth, low attenuation, immune to electromagnetic interference, but costly and fragile

Note: Fiber-optic transmission uses total internal reflection to keep light confined to the core as it travels through the cable, which is why signal loss is far lower than in copper-based media.

2.3.2 Unguided Media (Wireless Transmission)

Unguided media transport electromagnetic waves without confining them to a physical conductor. Signals are broadcast through air (or space) and can be received by any device capable of receiving them. Common types include radio waves, microwaves, and infrared.

Type	Frequency Range	Typical Use
Radio Waves	3 kHz – 1 GHz	Omnidirectional broadcasting: AM/FM radio, television, cordless phones

Type	Frequency Range	Typical Use
Microwaves	1 GHz – 300 GHz	Unidirectional, line-of-sight links: satellite and terrestrial point-to-point communication
Infrared	300 GHz – 400 THz	Short-range line-of-sight use: remote controls, some indoor wireless links

2.3.3 Satellite Communication

A communication satellite is essentially a microwave relay station placed in orbit that receives, amplifies, and retransmits signals over very large distances between widely separated ground stations. Satellites are classified by their orbital altitude.

Satellite Type	Altitude	Notes
GEO (Geostationary)	~35,786 km	Appears fixed relative to Earth; used for TV broadcast and long-distance links; needs only 3 satellites for global coverage but has high latency
MEO (Medium Earth Orbit)	~5,000–20,000 km	Used for GPS and navigation systems; moderate latency
LEO (Low Earth Orbit)	~500–2,000 km	Used for modern broadband and imaging (e.g. Starlink); low latency but requires many satellites

Key Takeaways — 2.3

- ✓ Guided media (twisted pair, coaxial, fiber) physically confine signals; fiber offers the highest bandwidth and lowest loss.
- ✓ Unguided media (radio, microwave, infrared) send electromagnetic waves through free space.
- ✓ Satellites relay microwave signals and are classified as GEO, MEO, or LEO by altitude.

2.4 Bandwidth Utilization

Bandwidth utilization techniques allow a link's available capacity to be used more efficiently by combining multiple signals or by spreading a signal across a wider band of frequencies for security and robustness. The two broad categories are multiplexing and spreading.

2.4.1 Multiplexing

Multiplexing is the technique of combining and sending multiple data streams over a single medium; the technique divides the capacity of a link into several logical channels, one for each stream.

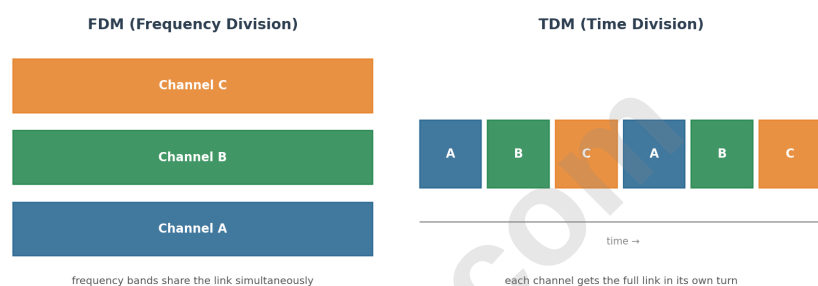


Fig. 2.4 — Frequency Division Multiplexing (FDM) vs. Time Division Multiplexing (TDM).

Type	Principle	Typical Use
FDM (Frequency Division)	Divides the available bandwidth into separate non-overlapping frequency bands, one per channel; signals are transmitted simultaneously	Radio and TV broadcasting, analog telephone trunk lines
WDM (Wavelength Division)	A version of FDM used in fiber-optic systems, where different wavelengths (colors) of light carry separate signals	High-capacity fiber-optic backbone links
TDM (Time Division)	Divides transmission time into slots; each channel is allotted a slot in a repeating cycle (synchronous TDM: fixed slots; statistical TDM: slots allotted on demand)	Digital telephone systems, T-1/E-1 lines

Note: In FDM, guard bands (unused frequency gaps) are placed between channels to prevent overlap and interference; in TDM, guard times serve a similar purpose between time slots.

2.4.2 Spreading

Spread spectrum techniques take a signal and spread it over a much wider bandwidth than strictly necessary, making the transmission more resistant to interference, jamming, and eavesdropping. Two common spreading techniques are used in wireless networks.

- FHSS (Frequency Hopping Spread Spectrum): the signal hops rapidly between many different carrier frequencies in a pattern known to both sender and receiver, making the

transmission harder to intercept or jam.

- DSSS (Direct Sequence Spread Spectrum): each data bit is replaced by a sequence of bits (a chip code), spreading the signal across a wider band; the receiver uses the same code to recover the original data.

Key Takeaways — 2.4

- ✓ Multiplexing (FDM, WDM, TDM) allows several signals to share one physical link efficiently.
- ✓ FDM/WDM separate signals by frequency/wavelength; TDM separates them by time slot.
- ✓ Spread spectrum (FHSS, DSSS) trades bandwidth for security and resistance to interference.

bcalink.com

2.5 Switching

A network is made of a set of switches connected by physical links, in which a packet or message travels from one device to another by passing through these intermediate switches. Switching refers to the way data is transferred from an input port to an output port of a switch (node).

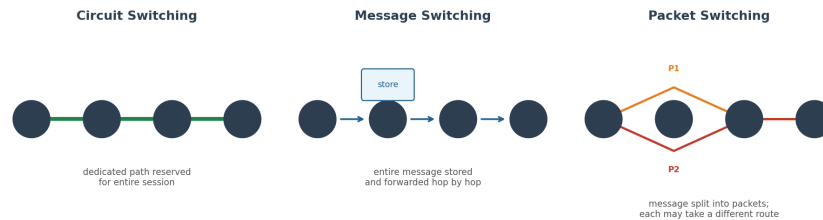


Fig. 2.5 — Circuit switching (dedicated path), message switching (store-and-forward of whole messages), and packet switching (independently routed packets).

2.5.1 Circuit Switching

In circuit switching, a dedicated physical path (circuit) is established between sender and receiver before actual data transfer begins, and this path remains reserved for the entire duration of the communication session, even during idle periods. It involves three phases: connection establishment, data transfer, and connection teardown. The traditional telephone network is the classic example.

2.5.2 Message Switching

In message switching, there is no dedicated path between sender and receiver. Instead, the entire message is sent as a single unit and stored at each intermediate switch (a technique called store-and-forward) before being forwarded to the next node when the outgoing link becomes free. Because full messages (which can be very large) must be stored at each node, this method requires large storage capacity and introduces significant delay; it is largely obsolete today but historically used in telegram and early email systems.

2.5.3 Packet Switching

In packet switching, the message is divided into smaller, fixed- or variable-size units called packets, each carrying header information (such as source and destination address). Packets belonging to the same message may travel independently and even take different routes through the network, and are reassembled at the destination. This is the switching method used by the modern Internet, and it is implemented in two forms:

- Datagram approach (connectionless): each packet is treated independently and may follow a different path; packets can arrive out of order (used by IP).
- Virtual-circuit approach (connection-oriented): a logical, pre-agreed path is set up before data transfer, and all packets follow that same path in order (used by technologies such as Frame Relay and ATM).

Feature	Circuit Switching	Message Switching	Packet Switching
Path	Dedicated, reserved	No dedicated path	No dedicated path (or virtual path)
Setup delay	Present (call setup)	None	None (datagram) / Present (virtual circuit)
Resource use	Reserved even if idle	Shared, but full message stored	Efficient, shared per packet
Example	Telephone network	Old telegram systems	Internet (IP)

Key Takeaways — 2.5

- ✓ Circuit switching reserves a dedicated path for the whole session — simple but wasteful of idle capacity.
- ✓ Message switching stores and forwards whole messages hop by hop — flexible but slow and storage-heavy.
- ✓ Packet switching splits data into packets that share the network efficiently and is the basis of the modern Internet.

2.6 Telephone, Mobile, and Cable Networks for Data Communication

2.6.1 Telephone Network

The traditional telephone network was originally designed for voice, using analog circuit-switched connections. It is organized hierarchically:

- **Local loops:** analog twisted-pair lines connecting individual subscribers to the local end office (exchange).
- **Trunks:** digital links that carry multiplexed voice channels between switching offices.
- **Switching offices:** hierarchical exchanges (end offices, tandem offices, regional/international offices) that route calls between subscribers.

For data communication, a modem (modulator/demodulator) is used to convert digital computer data into analog signals suitable for the local loop, and back again at the receiving end. Later, DSL (Digital Subscriber Line) technology allowed much higher-speed digital data transmission over the same twisted-pair local loop by using frequencies above the voice band.

2.6.2 Mobile (Cellular) Network

Mobile telephone networks divide a geographic region into small areas called cells, each served by a base station (cell tower). This cellular structure allows frequency reuse across non-adjacent cells, greatly increasing overall capacity. As a mobile device moves from one cell to another during a call, the network performs a handoff (handover), transferring the connection to the new cell's base station without dropping the call. Mobile data communication has evolved through several generations:

Generation	Approx. Technology	Key Feature
1G	Analog cellular (AMPS)	Voice only, analog signaling
2G	GSM, CDMA	Digital voice; basic data (SMS)
3G	UMTS, CDMA2000	Higher-speed packet data; mobile internet access
4G	LTE	High-speed all-IP packet-switched data, video streaming
5G	New Radio (NR)	Very high bandwidth, low latency, massive device connectivity

2.6.3 Cable Network for Data Communication

Cable television networks, originally built with coaxial cable to distribute analog TV signals, have been adapted to carry data using a technology called DOCSIS (Data Over Cable Service

Interface Specification). A cable modem at the subscriber's premises communicates with a Cable Modem Termination System (CMTS) at the provider's head-end, allowing high-speed internet access over the same cable infrastructure that delivers television channels. Modern cable networks typically use a Hybrid Fiber-Coaxial (HFC) architecture, where fiber-optic cable carries signals from the head-end to a neighborhood node, and coaxial cable completes the final connection to individual homes — combining the high capacity of fiber with the low cost of reusing existing coaxial drops.

Key Takeaways — 2.6

- ✓ The telephone network's local loops and trunks were adapted for data using modems and later DSL technology.
- ✓ Cellular networks reuse frequencies across cells and hand off calls as devices move; mobile data has evolved through 1G to 5G.
- ✓ Cable networks use DOCSIS and a Hybrid Fiber-Coaxial architecture to deliver broadband data alongside television service.

— *End of Unit 2* —