Physical Layer

The physical layer deals with transporting bits between two machines. How do we communicate 0's and 1's across a medium? By varying some sort of physical property such as voltage or current.

Moreover, by representing the property as a function of time, we can analyze it mathematically. Our goal is to understand what happens to a signal as it travels across some physical media. That is, will the receiver see the exact same signal generated by the sender? Why or why not?

Theoretical Basis for Data Communication

Fourier Analysis

Fourier showed that a periodic function $g(t)$ can be represented mathematically as an infinite series of sines and cosines:

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

1. f is the function's *fundamental frequency*
2. $T = \frac{1}{f}$ is the function's *period*
3. a_n and b_n are the amplitudes of the nth *harmonics*

The series representation of $g(t)$ is called its *Fourier series* expansion.

In communications, we can always represent a data signal using a Fourier series by imagining that the signal repeats the same pattern forever.

Moreover, we can compute the coefficients a_n and b_n:

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

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

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

For instance, suppose we use voltages (on/off) to represent "1"s and "0"s, and we transmit the bit string "011000010'. The signal would look as follows:

Recall (from calculus):

1. the derivative of sin(x) = cos(x)dx
2. the derivative of cos(x) = -sin(x)dx

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

$$= \frac{2}{T}[\int_1^3 sin(2\pi nft)dt + \int_6^7 sin(2\pi nft)dt]$$

$$= \frac{2}{T}\frac{-1}{2\pi nf}cos(2\pi nft)|_{1,6}^{3,7}$$

$$= \frac{-1}{\pi nfT}[cos(2\pi nf3) - cos(2\pi nf) + cos(2\pi nf7) - cos(2\pi nf6)]$$

$$f = 1/8$$

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

Similarly,

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

And

$$c = \frac{2}{T}\int_0^T g(t)dt = \frac{2}{T}3 = \frac{6}{T} = \frac{3}{4}$$

Points to note about the Fourier expansion

1. The more terms in the expansion, the more exact our representation becomes.
2. The expression $\sqrt{a_n^2 + b_n^2}$ represents the amplitude or energy of the signal (e.g., the harmonics contribution to the wave).

   In our example, the amplitude consists of a_n and continually gets smaller. (The b_n term is always zero.) Here, as in most cases, the first harmonics are the most important ones.

So what does this have to do with data communication? The following facts are important:

1. Signals *attenuate* (strength of signal falls off with distance) during transmission. How much attenuation occurs? The exact amount is dependent on physical properties of the medium.
2. *Distortion* results because attenuation is non-uniform across the frequency spectrum; some frequencies distort more than others. That is, the signal doesn't distort uniformly. If every component decreased by the same amount, the signal would be weaker, but not distorted, and amplifying the signal would restore it. Because the received signal is distorted, however, amplification simply magnifies the distortion and probably won't help.
3. A transmission medium carries signals lying within in a *spectrum* or range of frequencies; the absolute width of the spectrum is called the *bandwidth* of the channel. In other words, most channels completely attenuate (e.g. chop off) frequencies beyond some threshold value.

   What does this mean in terms of Fourier harmonics? In terms of fundamental frequencies of a Fourier representation, higher harmonics get completely chopped off during transmission and are not received at the receiving end!

Conclusion: it's essentially impossible to receive the *exact* signal that was sent. The key is to receive enough of the signal so that the receiver can figure out what the original signal was.

Note: "bandwidth" is an overloaded term. Engineers tend to use bandwidth to refer to the spectrum of signals a channel carries. In contrast, the term "bandwidth" is often also used to refer to the data rate of the channel, in bps.

Factors determining the rate of data transmission

1. The *baud rate* (also known as the *modulation rate*) refers to the maximum rate at which the signal changes value (e.g., switches voltages). For example, if "0"s and "1"s were represented as +5V, -5V, respectively, the baud rate would refer to the number of times per second the signal switches as its transmitting a string of alternating 0's and 1's. Note that we can potentially achieve a higher data rate by switching the voltage faster.[1]

2. The *encoding method* determines the amount of information carried in one baud.

   In our example we encoded only one bit of information (0 or 1). How can we encode 2 bits worth of information in one baud? Use 4 different voltage levels. For example, 0, 1, 2, 3 could be represented as -10, -5, +5 and +10 volts respectively.

Note: baud rate is not the same thing as the data rate. For a given baud rate, we can increase the data rate by changing the encoding method (subject to Nyquist and Shannon limits, of course.)

[1]Note, however, that Nyquist or Shannon may have something to say about this.

Voice Grade Lines

What kind of data rate can we achieve using voice-grade phone lines?

The phone system is designed to carry human voices (not data!), and its bandwidth line is limited to about 3 kHz.

Suppose that we have a bit rate of b bits/sec (assume only encode one bit of data per baud).

1. For 8 bits of data, the fundamental frequency F would be $b/8$ Hz.
2. Because the phone line attenuates frequencies above 3 kHz, the number of the highest harmonic passed through is $3000/F = 3000/(b/8) = 24000/b$.
3. At 1200 baud, the fundamental frequency is $1200/8 = 150Hz$, and the highest numbered harmonic passed is $24000/1200 = 20$. That is, only the first 20 terms of the Fourier series are relevant; the phone line will chop off all higher numbered terms.

The following table gives more values (Figure 2-2)

Baud Rate	Fundamental Harmonic (Hz)	Number of Harmonics sent
1200	150	20
2400	300	10
4800	600	5
9600	1200	2.5
19200	2400	1.25
38400	4800	.625

Will we be able to send data at 38,400 baud? No! It should be clear that sending data at 38400 baud over a voice grade line simply won't work. Even at 9600 baud only the first and second harmonic are transmitted, and the signal will be severely distorted. It is unlikely that the receiver will be able to recognize the signal as intended.

Must use better encoding schemes for higher data rates.

Maximum Data Rate of a Channel

Nyquist (1924) studied the problem of data transmission for a fine bandwidth noiseless channel. Nyquist states:

1. If a signal has been run through a low-pass filter of bandwidth H, the filtered signal can be completely reconstructed by making $2H$ samples.

   The important corollary to Nyquist's rule is that sampling more often is pointless because the higher frequencies have been filtered out.

2. If the encoding signal method consists of V states:

   maximum data rate $= 2H \log_2 V$ bps

What's the maximum data rate over phone lines? Going back to our telephone example, Nyquist's theorem tells us that a one-bit signal encoding can produce no better than:
$2 \times 3000 \times \log_2 2 = 6000$bps.

But there is a catch. In practice, we don't come close to approaching this limit, because Nyquist's rule applies only to *noiseless* channels.

Noise on a Channel

In practice, every channel has background noise. Specifically:

1. *Thermal* noise results from thermal agitation of electrons in a conductor. It cannot be eliminated, and depends on the temperature, bandwidth, and Boltzman's constant K. Is uniformly distributed across the frequency spectrum and thus called white noise.
2. *Intermodulation* noise results when different frequencies share the same transmission medium; unwanted signals often appear at frequencies that are the sum or differences of the two frequencies.
3. *Crosstalk* noise results from unwanted coupling between signal paths. Hearing another conversation (faintly) on a telephone connection is an example of crosstalk.
4. *Impulse* noise consists of sharp, shortlived disturbances, from such sources such as lightning. Stallings Fig 2.15.

How do we measure (or quantify the amount of) background noise? The *signal-to-noise* ratio is a measure of the unwanted noise present on a line. It is expressed in decibels (db) and given by:

$$S/N_{(db)} = 10 \log_{10} \frac{signal\ power}{noise\ power}$$

Shannon's Theorem

Shannon's theorem gives the maximum data rate for channels having noise (e.g., all real channels). Shannon's theorem states that the maximum data rate of a noisy channel of bandwidth H, signal-to-noise ratio of S/N is given by:

max data rate $= H \log_2(1 + S/N)$

Note: the signal to noise ratio S/N used in Shannon's theorem refers to the ratio of signal power to noise power, not the ratio expressed in dbs (decibels). Unlike Nyquist's limit, Shannon's limit is valid regardless of the encoding method.

Let's consider a phone line again. A typical value for the S/N ratio for phone lines is 30db.

$S/N = 10^{\frac{S/N_{(db)}}{10}} = 10^3 = 1000.$

Thus, the maximum data rate $= 3000 \times \log_2(1 + 1000) \approx 30,000$ bps.

But wait — don't modems deliver data at 38.4 and 56 kbps? Many modem companies advertise that their modem deliver higher data rates, are they lying? Not necessarily. Read the fine print. Most likely, the modem uses data compression, and the high data rate is achieved only with text data.

Let's summarize what Nyquist and Shannon say:

- Nyquist: sampling a received signal more frequently than $2H$ (where H is the bandwidth of the channel) is pointless.
- Nyquist: maximum data rate = $2Hlog_2V$ bps, where H is the bandwidth of the channel, and V is the number of distinct encodings for each baud. This result is a theoretical upper bound on the data rate *in the absence of noise*.
- Shannon: maximum data rate = $Hlog_2(1 + S/N)$, where S/N is the ratio of signal power to noise power. Note that Shannon's result is independent of the number of distinct signal encodings. Nyquist's theorem implies that we can alway increase the data rate by increasing the number of distinct encodings; Shannon's limit says that is not so for a channel with noise.

Transmission Media

Guided (a physical path) vs. unguided media (waves propagated, but not in a directed manner).

Summarized in Fig 2-11 (Tanenbaum) and Table 3.1 (Stallings).

Physical Storage Devices

What's the most cost effective way to transmit large quantities of data? Federal Express!

Still magnetic tapes.

Also DAT digital audio tape storage devices.

Main drawback? High delay in accessing data. It takes minutes to hours to days to physically transport the cassette from one location to another.

Twisted Pair

In twisted pair technology, two copper wires are strung between sites:

- The two wires are typically "twisted" together in a helix to reduce interference between the two conductors.
- Can carry analog or digital signals. Actually, they carry only analog signals. However, the "analog" signals can very closely correspond to the square waves representing bits, so we often think of them as carrying digital data.
- Data rates of several Mbps common.
- Spans distances of several kilometers.
- Data rate determined by wire thickness and length. In addition, shielding to eliminate interference from other wires impacts S/N, and ultimately, the data rate.
- Good, low-cost communication. Indeed, many sites already have twisted pair installed in offices — existing phone lines!

Baseband Coaxial

With "coax", the medium consists of a copper core surrounded by insulating material and a braided outer conductor.

The term *baseband* indicates digital transmission (as opposed to *broadband* analog).

Physical connection consists of metal pin touching the copper core. There are two common ways to connect to a coaxial cable:

1. With *vampire taps*, a metal pin is inserted into the copper core. A special tool drills a hole into the cable, removing a small section of the insulation, and a special connector is screwed into the hole. The tap makes contact with the copper core.
2. With a *T junction*, the cable is cut in half, and both halves connect to the T junction. A T-connector is analogous to the signal splitters used to hook up multiple TVs to the same cable wire.

Data rate depends on physical properties of cable, but 10 Mbps is typical.

Broadband Coax

The term *broadband* refers to analog transmission over coax. (Note, however, that the telephone folks use broadband to refer to any channel wider than 4 kHz). The technology:

- Typically bandwidth of 300 MHz, total data rate of about 150 Mbps.
- Operates at distances up to 100 km (metropolitan area!).
- Uses analog signaling.
- Technology used in cable television. Thus, it is already available at sites such as universities that may have TV classes.
- Total available spectrum typically divided into smaller channels of 6 MHz each. That is, to get more than 6MHz of bandwidth, you have to use two smaller channels and somehow combine the signals.
- Requires amplifiers to boost signal strength; because amplifiers are one way, data flows in only one direction.

Two types of systems have emerged:

1. Dual cable systems use two cables, one for transmission in each direction:
   - One cable is used for receiving data.
   - Second cable used to communicate with *headend*. When a node wishes to transmit data, it sends the data to a special node called the *headend*. The headend then resends the data on the first cable. Thus, the headend acts as a root of the tree, and all data must be sent to the root for redistribution to the other nodes.
2. *Midsplit* systems divide the raw channel into two smaller channels, with each subchannel having the same purpose as above.

Which is better, broadband or baseband? There is rarely a simple answer to such questions. Baseband is simple to install, interfaces are inexpensive, but doesn't have the same range. Broadband is more complicated, more expensive, and requires regular adjustment by a trained technician, but offers more services (e.g., it carries audio and video too).

Fiber Optics

In fiber optic technology, the medium consists of a hair-width strand of silicon or glass, and the signal consists of pulses of light. For instance, a pulse of light means “1”, lack of pulse means “0”.

Three components are required:

1. Fiber medium: Current technology carries light pulses for tremendous distances (e.g., 100s of kilometers) with virtually no signal loss.
2. Light source: typically a Light Emitting Diode (LED) or laser diode. Running current through the material generates a pulse of light.
3. A photo diode light detector, which converts light pulses into electrical signals.

Advantages:

1. Tremendously high data rate, low error rate. 1000 Mbps (1 Gbps) over distances of kilometers common. Error rates are so low they are almost negligible.
2. Difficult to tap, which makes it hard for unauthorized taps as well.

   How difficult is it to prevent coax taps? Very difficult indeed, unless one can keep the entire cable in a locked room!
3. Much thinner (per logical phone line) than existing copper circuits.Because of its thinness, phone companies can replace thick copper wiring with fibers having much more capacity for same volume. This is important because it means that aggregate phone capacity can be upgraded without the need for finding more physical space to hire the new cables.
4. Not susceptible to electrical interference (lightning) or corrosion (rust).
5. Greater repeater distance than coax.

Disadvantages:

- Difficult to tap. It really is point-to-point technology. In contrast, tapping into coax is trivial. No special training or expensive tools or parts are required.
- One way channel. Two fibers needed to get full duplex (both ways) communication.

Fiber Uses

- long-haul trunks—increasingly common in telephone network (Sprint ads)
- metropolitan trunks—without repeaters (ave 8 miles in length)
- rural exchange trunks—link towns and villages
- local loops—direct from central exchange to a subscriber (business or home)
- local area networks—100Mbps ring networks.

Direction of the future.

Line-of-Sight (Wireless) Transmission

Line-of-sight technology is used when running a physical cable (either fiber or copper) between two end points is not possible. For example, running wires between buildings is probably not legal if the buildings are separated by a public street.

Infrared signals typically used for short distances (across the street or within same room),

Microwave signals commonly used for longer distances (10's of km). Sender and receiver use some sort of dish antenna.

Difficulties:

1. Weather interferes with signals. For instance, clouds, rain, lightning, etc. may adversely affect communication.
2. Radio transmissions easy to tap. A big concern for companies worried about competitors stealing plans.
3. Signals bouncing off of structures may lead to out-of-phase signals that the receiver must filter out.

Satellites

Satellite communication is based on ideas similar to those used for line-of-sight. Characteristics:

1. Satellite typically placed in *geosynchronous orbit* 36,000 km above earth; satellite never "moves" relative to ground stations. This is important because if the satellite moves, the ground antenna must follow its movements.
2. Satellite typically acts as a *repeater*, receiving signals from earth on one channel and rebroadcasting them on another. Satellite may rebroadcast data over an area covering large fraction of the earth's surface.
3. Number of geosynchronous satellites limited (about 90 total, to minimize interference).
4. International agreements regulate how satellites are used, and how frequencies are allocated.
5. Weather effects certain frequencies.
6. Satellite transmission differs from terrestrial communication in another important way: One-way *propagation delay* is roughly 270 ms. In interactive terms, propagation delay alone inserts a 1 second delay between typing a character and receiving its echo.

Comparison/contrast with other technologies:

1. Propagation delay very high. On LANs, for example, propagation time is in nanoseconds — essentially negligible.
2. One of few alternatives to phone companies for long distances.
3. Uses broadcast technology over a wide area — everyone on earth could receive a message at the same time!
4. Easy to place unauthorized taps into signal.

Satellites have recently fallen out of favor relative to fiber.

However, fiber has one big disadvantage: no one has it coming into their house or building, whereas anyone can place an antenna on a roof and lease a satellite channel.