Lecture 3

Physical layer: Data communications

We switch to the basics of communication: the physical layer. We will learn what are the services it provides to the Datalink layer.

References:

- CN chapter 3.2, 3.5–3.8.

The general scheme

The physical layer is responsible for converting the bit stream obtained from datalink to the physical medium. This is achieved through a few steps.

1. **Collect** one or a few bits at a time from the datalink to form groups.
2. Optionally, **spread** the group of bits into groups of "chips". This process is usually used when the channel is very noisy.
   We'll come back to this later.
3. **Encode** or **modulate** each bit or chip group into **symbols**, which is a physical phenomenon, e.g., voltage, light intensity, etc.
4. Send the symbols through the physical **medium** one by one at a supported speed, forming a **signal** of some physical phenomenon.

This lecture gives enough details of each of the process, so that we can understand the layers above.

Communication media in common use

Every commonly used medium uses a phenomenon that involves the movement of electrons or photons.

… which are different aspects of the same physical phenomenon. Somehow people think using sound waves is a joke (see RFC 1926).

- **A pair of copper wires**. The **current** flowing in the wires or the **voltage** between the wires can be manipulated by the sender and detected by the receiver to achieve communication.

  To **reduce the pick-up of magnetic field** in the surrounding, wires are usually **shielded**, i.e., one of the "wires" is a tube which surrounds the other ("Coaxial cable"); or are **twisted** ("Unshielded Twisted Pair", UTP); or both ("Shielded Twisted Pair", STP).

  The most common form is UTP, since it is easier to bend arbitrarily. UTP of different "categories" have different amount of twisting, more twist would make it more noise resistant and attenuate signals less. Category-3 and -5 are typically used (Category-5 is better).

Communication media in common use (cont'd)

- **A strand of glass fiber**, called **optical fiber**. The sender can generate flashes of light and that receiver can detect. The glass fiber guides the photons to move in a predefined path.

  Light is generated by LED or better, laser (which switches faster, and also makes sure all photons are of similar direction within the fiber). Sometimes the fiber is made extremely thin ("**single mode** fiber") to make sure all photons travel at exactly the same direction.

- **The empty space**. Electromagnetic waves (i.e., photons) can be generated by the sender without any media, and the receiver can detected the photon (using an antenna or other detector).

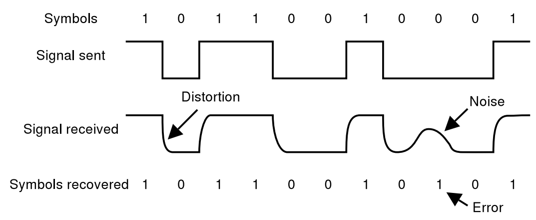
  Communication is usually **omni-directional** at low frequency (i.e., sent to all directions. E.g., radio), while directional at high frequency (e.g., micro-wave, infra-red, visible light).

Baud rate: Basic digital communication

The low level communication abstraction is the same for all the media.

- At any time, one of two, or more, **symbols** can be contained in a medium. E.g., presence or absence of light, one of many voltages, etc.
- The sender has a **clock** to generate symbols at **regular intervals**. The rate at which symbols are generated is called the **baud rate**.
- The stream of symbols, or "signal", is then **put into the medium** for delivery. The delivery will introduce **systematic distortion** and **random noise**, changing the signal slightly.
- The receiver either **uses its own clock of exactly the same rate** to extract the symbols at regular interval, or **monitors the change of the symbols in the medium** to extract the symbols at the correct times.
- When there is noise, the recovered symbol may not be the same as the symbol generated by the sender. We call this an **error** of the signal.

Illustration



Note that the signal generated is a function of time. Sometimes we write $f(t)$ to denote the time-dependent signal function.

Time-independent Linear systems

Two characteristics hold for many media (within reasonable limit, ignoring noise), which makes the complexity of recovering symbols affordable.

- **Linear**, i.e., **IF** the signals $f(t)$ and $g(t)$ sent by the sender causes the receiver to receive the signal $F(t)$ and $G(t)$, **THEN** the signal $f(t) + g(t)$ sent by the sender will be received as $F(t) + G(t)$.
- **Time invariant**, i.e., **IF** the signal $f(t)$ sent by the sender is received as $F(t)$, **THEN** the signal $f(t + dt)$ sent will be received as $F(t + dt)$.

Any medium with these two characteristics has a very useful property:

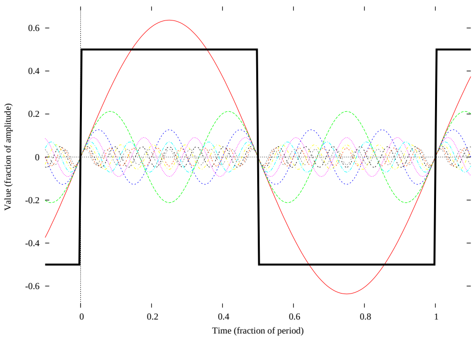
- **IF** a pure sine wave $A \sin(\omega t)$ is sent, **THEN** a pure sine wave $A' \sin(\omega t - \phi)$ is received.

The received signal may be attenuated (by A'/A) or delayed (by ϕ/ω), and the attenuation and delay may be dependent on the frequency of the sine wave. But it is **always** a pure sine wave.

Fourier transform

- Given the property of linear systems, we usually **split any signal** into (perhaps infinitely) **many sine waves** of different frequencies, each has its own **amplitude** and **phase** (start angle in the sin function). The collection is called a **frequency spectrum**.
- The splitting can be achieved through the **Fourier transform**.
  The mathematics involved is not within the scope of this course.
- The **behavior** of each of these sine waves is analyzed **independently**, according to its **frequency**. According to the property we have just seen, each of them produces a sine wave of the same frequency.
- The resulting sine wave can then be added together to predict the final, distorted signal from the medium, since the medium is linear.
- For our purpose, what is important includes (1) **we want linearity**, (2) **attenuation depends on frequency**, and (3) **every signal has a frequency spectrum**.

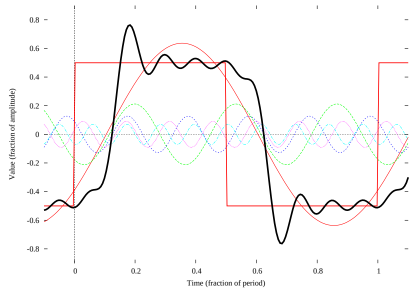
Example: spectrums of pure periodic square wave



Most signal energy are in the sine wave with frequency matching that of the square wave; amplitude of higher frequency (frequency=$n \times$ base frequency) falls approximately as $1/n$, so power of higher frequency falls approximately as $1/n^2$.

Transfer through linear medium

Through a medium in which (1) high frequency traval a bit faster, and (2) frequencies larger than 10 are completely lost, we receive this…

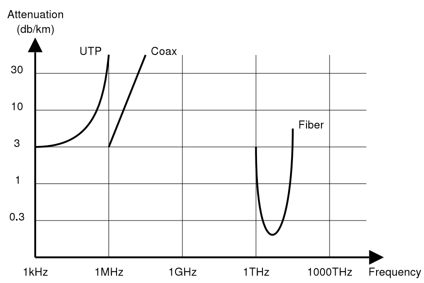


Note that not all sine waves zero at the same time, unlike the sent signal. Since waves of different frequencies are attenuated and delayed differently, distortion occurs.

The role of distance

- The **longer the distance** that the signal must travel, the **more the attenuation**, and the **more the delay**.
- Electrons travel at around 0.8c; light in glass travels at around 0.7c, where $c = 3 \times 10^8\ ms^{-1}$ is the speed of light in vacuum. E.g., to travel 1km it takes around 5 µs.
  This can be a problem because in 5 µs we have already sent 50 bits even if we are as slow as 10Mbps. This is especially problematic on multiple access medium.
- In guided media, signal strength (amplitude) attenuates **exponentially with respect to distance** (i.e., $A_{recv} = A_{send} a^{-d}$). Exactly how much the attenuation depends on how "transparent" is the medium.
  Optical fiber must be extremely pure in order to reduce such attenuation.
- Noise continuously gets added. So it is nearly independent of distance. **When distance is larger, signal/noise is smaller**. At some limit we have to add a **repeater** to regenerate the signal and remove the noise.

Attenuation of various medium



From "Data and Computer Communications", William Stallings. This is an (imprecise) approximation, considering that there are many types of UTPs and many type of Coax.

Here 10dB = a factor of 10, 20dB = a factor of 100, 30dB = a factor of 1000, etc.

Extra attenuation for omni-directional communication

There is an extra factor for communication that is omni-directional.

- The **further** the distance, the **more receivers** can be placed, so each single receiver gets **less part of the sent energy**.
- Usually, **the received energy drops as** d^2 (i.e., $A_{recv} = A_{send}\ a^{-d}/d^2$). , where d is the distance between the sender and the receiver.
  This is natural, as the surface area of the sphere at distance d grows as d^2.
- Since the received energy would normally drop exponentially anyway, this additional factor usually doesn't cause major problem.
- But there is a surprising problem of the opposite direction… at a very close distance to the sending site, the amplitude is **very** large.
- So **whenever one sends**, it is **difficult to receive** anyone else.
  Why this is not as much a problem for directional communication?

Bandwidth

- For all media, **attenuation is small enough** for communication **only for a small portion of the frequency spectrum**.
- Given an **acceptable attenuation**, we can look at the plot between attenuation and frequency to find **what frequencies can be used**.
- E.g., if we use UTP of 500m and we want to keep attenuation to be within 5dB (i.e., 10dB per 1km), then we might be restricted to use a frequency spectrum where most signal energy is between 0Hz to 100kHz.
- The "width" of the frequency spectrum is called the **bandwidth** of the medium. E.g., here the bandwidth is 100kHz - 0Hz = 100kHz.
- Maximum achievable baud-rate given this band is 200kHz, since if you send 101010… at 200kHz you are essentially generating 100kHz square wave—even if you can afford losing all higher frequencies, you need 100kHz.
  Other sequence uses lower frequencies. E.g., sending 0000… gives 0Hz.

Bit-rate and Nyquist signaling rate

Baud-rate has a profound effect on the bit-rate at which data can be transferred.

- If bandwidth is W, then baud-rate is at most $2W$.
- If the medium can contain only 2 symbol, i.e., 1-bit of information at a time, then the bit-rate also can only be $2W$.
- If the medium can contain S **symbols** (i.e., $\log_2 S$ bits of information) **at a time**, then we can **send at** $\boxed{2W\log_2 S}$ **bits per second**.
- We call this the **Nyquist signaling rate** achievable using a fix edbandwidth and number of symbols.

Nyquist signaling rate, example

- So to achieve **higher bit-rate**, we can (1) **increase the bandwidth** (which is usually impossible), or (2) use **more symbols**.
- E.g., we might **use more than 2 symbols** (e.g., voltage levels). If we use 16, we can transfer 4 times as much data per second. So if our baud rate is 10kHz, our bit-rate can be 40k bits per second.
  Actual bit-rate would be a little lower, because we usually want to leave a little bit of time for the sender and receiver to synchronize their clocks.
- Ideally, we can increase the number of symbols infinitely to achieve higher and higher data rate. In practice this is impossible for two reasons:
  - The **more symbols** to send and receive, the **more expensive** to build the system.
  - The more symbols that the receiver must be able to differentiate, **the less the tolerance is the system against noise**.

Signal-to-Noise ratio and Shannon's capacity

- Of course, if the signal is stronger, the effect of noise is less. So we measure the effect of noise by the **signal to noise ratio**, SNR.
- Shannon shows that given a medium of bandwidth W and "thermal" noise (i.e., one that follows Gaussian distribution), the bit rate cannot be higher than $\boxed{W\log_2(1+SNR)}$.
- This is called the **Shannon capacity** of a medium. E.g., phone line: W=3.4kHz, SNR=10000, so bit rate is at most 45.2k bits/second.
  Why W is so small? Each underlying cable is shared by many telephone users.
- To improve the bit rate achievable, we can thus increase the signal to noise ratio by (1) reducing noise level, or (2) increasing signal level.
- Both cannot be done infinitely: part of the noise is a completely random process that happens all the time, and if we increase the signal level too much the medium ceases to be linear.

Line coding

If we want to maximize bit-rate given a bandwidth, we would squeeze as many ($\log_2 m$) bits of data as possible for each symbol.

At other times (especially, for local area networks) when bandwidth is abundant, we instead try to **reduce cost** by sending less bits.

- To ease synchronization, one might try to make sure that **symbol change(s) occurs every, or every few, bits**.
- To avoid sending a DC current when using a pair of wires, one might make sure that **positive and negative voltage is used equally often**.
  Needed if the current must pass through a transformer.
- One might want to have system that works even if the signal is **inverted**. This can be achieved using **differential** coding, where a change in the "form" indicates a bit-1, no change indicates a bit-0.
  This only requires a simple preprocessing that turns every 0 (resp., 1) bit to a bit that is the same as (resp., different from) the previous. E.g., 00110=>00100.

Some line coding schemes

- **Non Return to Zero** (NRZ): send voltage V for 1 and voltage 0 for 0.

  Polar NRZ: sends V/2 and -V/2 resp. to reduce power consumption. Bipolar NRZ: sends V/2 and -V/2 alternatively for 1 and voltage 0 for 0 to remove DC. Differential NRZ is also called NRZI (I for inverted).

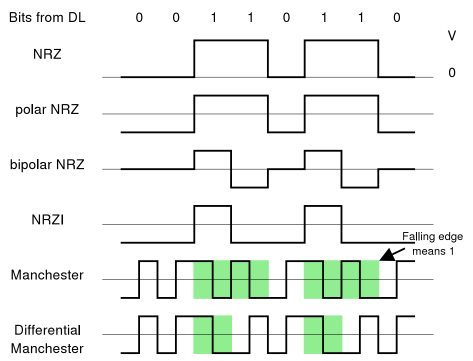
- **Manchester encoding**: send two binary symbols for every bit: (-V, V) for 0, (V, -V) for 1. Doubles the bandwidth, buying us no DC and easy synchronization. Voltage 0 can be used to mean "no data".

- **mBnB**: send n binary symbols for every m bits (n>m). This is a generalization of Manchester (which is 1B2B). If designed properly it can have most of the benefit of Manchester, without wasting as much bandwidth.

- **mBnT**: send n triary symbols for every m bits. This is used when bandwidth becomes tight (e.g., fast Ethernet).

  10Mbps Ethernet uses 1B2B, 100Mbps Ethernet uses 4B5B or 8B6T, Gigabit Ethernet uses 8B10B.

Example line codings



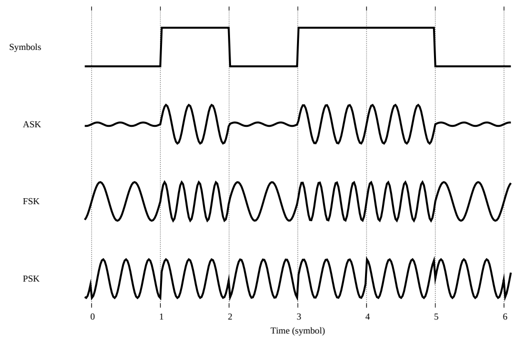
Note that polar NRZ vs. Manchester is exactly the same as NRZI vs. differential Manchester.

Modulation

- Life is a bit harder if the channel is **band-pass** rather than **low-pass**. I.e., it does not pass low frequency at all (e.g., radio).
- Then all "symbols" must be similar to a sine wave, called the **carrier**. Each symbol changes the carrier slightly:
  - **Frequency Shift Keying** (FSK): Slightly different frequency represent different symbols.
  - **Amplitude Shift Keying** (ASK): Different amplitude represent different symbols.
  - **Phase Shift Keying** (PSK): There is a **phase shift** at each cycle, different amount of shift represent different symbols.
- This technique is called **modulation**, or **frequency shifting**: we "shift" our low-frequency bits into the high-frequency symbols.

  In contrast, we call communication without modulation to be "base-band".

Examples of various kinds of modulations



NB: in PSK shown above, whenever a -1 is sent, the phase "goes back" 90 degrees; and whenever a 1 is sent, the phase "go forward" 90 degrees. Since there is always a shift, it is "self-clocking": you know the baud rate by just looking at the signal.

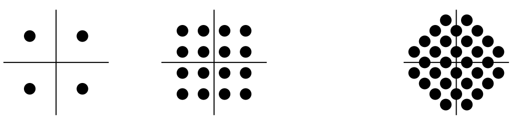
Signal Constellations

Most actual schemes uses both ASK and PSK. I.e., different symbols may have different phase shift or different amplitude or both.

A **signal constellation** is a drawing to show which of the combinations of amplitudes and phase shifts are used, in a **polar coordinate**.

Shannon capacity is still valid: signal to noise level restricts how many points can be put into the signal constellation.

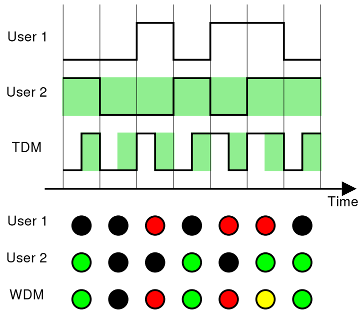
QPSK: PSK using 4 phase shifts — QAM-16: 16 points in constellation to support 9600bps with 2400baud — V.32: 32 points in constellation to support 9600bps with 2400baud, leaving half the points unused for error detection



Multiplexing

- It is expensive to build a cable or fiber between two locations.
- If there are **many pairs of users** who desire communication between the **same two locations**, one can build a single high bandwidth cable, and have all users share the cable (and of course the cost).
- Each pair of users should then have a **small share** of bandwidth. How to "split" a large physical medium into many small channels?
- For baseband communication, the easiest way is to **allocate each pair of users a share of symbols**. So the users "round-robin" to use the medium. We call it **time-domain multiplexing** (TDM).
- With fiber ,there is another possibility: let the senders use light of different **color** (i.e., wavelength). At the receiving end the different color is separated (by prism or diffraction grating). We call this **wavelength-domain multiplexing** (WDM).

Illustration: TDM and WDM



Here, for simplicity we draw WDM such that the users send at exactly the same time. But this is not necessary—although TDM requires the users to be more or less synchronized.

FDM and ADSL

- For modulated signals, it is troublesome to use TDM. A technique similar to WDM is used instead: **different channels use a carrier of different frequency**. We call it **Frequency Domain Multiplexing** (FDM).
- Each channel is a modulated signal, so uses more than just the carrier frequency. Instead, each channel must be allocated a small band, and the **modulated signal must sit within the band**.
- At the receiver, the desired band is **filtered** out using a band-pass filter .
- E.g., When **phone companies** provide **Internet** service, they want to reuse the phone lines between the user and the end-office of the phone companies (which was under-utilized to carry only 3.4KHz signals).

  The Asymmetric Digital Subscriber Line (ADSL) scheme uses FDM to split the 1.1MHz bandwidth of the wire to 256 bands each being 4KHz. The lowest channel supports normal phone service, while the remainder uses different modulation to send and receive digital signals.

Some physical layers in common use

- **Phone line with 33.6kbps (V.34bis) modem**: UTP filtered to the band between 100Hz and 3.5kHz, using voltages of at most 48V. Modulation is done using a 1959Hz carrier, running 3429 baud with a constellation with 960 points.
- **100MHz Ethernet on UTP**: 2 pairs of UTP category-5, one for each direction. They run at 25 Mbaud, baseband. 4B5B allows 4 bits (and some framing information) per symbol, with a voltage between 0.85V and -0.85V. It runs at most 250m (for a reason that we will examine later).

  Note: standard Cat-5 UTP have 4 pairs, so 2 pairs are unused. A PoE (Power over Ethernet) system uses them to deliver 12V power to devices.
- **Gigabit Ethernet on single-mode optical fiber**: light of wavelength 0.85 micron is generated by laser and transferred using 10 micron fiber , running at most 5km. Signal is simply presence or absence of light, coded using 8B10B over 1.25 Gbaud. The 8B10B scheme makes sure no code has 4 consecutive 0's or 1's, to ease synchronization.

Some physical layers in common use (cont'd)

- **802.11b 11Mbps Wireless LAN**: Use a 22MHz channel of the radio band between 2400 MHz and 2484 MHz. Outgoing data is grouped into 8-bit symbols.

  This allows at most 3 non-overlapping channels, although overlapping channels continue to work (with higher error rate).

  Each symbol is mapped into 8 phases, each being one of 4 angles: 0, 90, 180 and 270 degrees. They are sent using QPSK, at 11 Mbaud.

  Each phase is 2 bits, so 8 phases has 16 bits. The mapping doubles the number of bits. Each bit influences many phases, so that even if a couple of phases are received wrong the output can still be correct. Such technique to combat noise is called **spreading**.

  The mapping is chosen according to a scheme called "CCK modulation". For details, see http://www.eetkorea.com/ARTICLES/2001MAY/2001MAY25_NTEK_DSP_AN.PDF.