Role of the PL

The PL controls the behaviour of the physical interface between the transceiver and the physical medium (i.e. the flow of bits). It is implemented entirely (or almost entirely) in hardware; the PL operates purely mechanically with its actions driven by a clock; any decision–making is left to the DLL.

The PL works with a single link.

The PL always treats the link as point–to–point, thinking that there is only one receiver. If the link is a broadcast (“multipoint”) link, it is up to the DLL to solve the problem of making the several PL nodes cooperate.

A link may be either **duplex**, **simplex** or **half–duplex**.

Physical Layer

Order of topics:

- Signalling.
- Media.
- Sources of errors.
- Encoding.
- Error detection and correction.

Transmission

Transmission is the act of signalling using a medium.

Transmission can be either analog (continuous) or digital (discrete). In both cases the transmitter must convey to the receiver timing information; the usual method is to divide (continuous or discrete) time into fixed–length **time slots** (periodic signal) and incorporate into the signal some sort of timing marker that indicates the beginning of the next slot.

It is also possible to use variable–length slots (aperiodic signal); this interesting topic is left aside.

The length of a time–slot or period of a signal determines the bandwidth of a link but one also has to take into consideration the number of bits per symbol transmitted (in real–life networks it varies from $\frac{1}{2}$ to 10 or maybe even 66).

Digital transmission

The oldest networks used discrete transmission mode with symbols being transmitted for a time slot followed by a period of silence after which the next symbol was displayed. Examples: Archimedes ($A\rho\chi\iota\mu\acute{\eta}\delta\eta\varsigma$)[a], Napoleon[b], British Navy[c], American Indians[d], telegraph[e], etc.

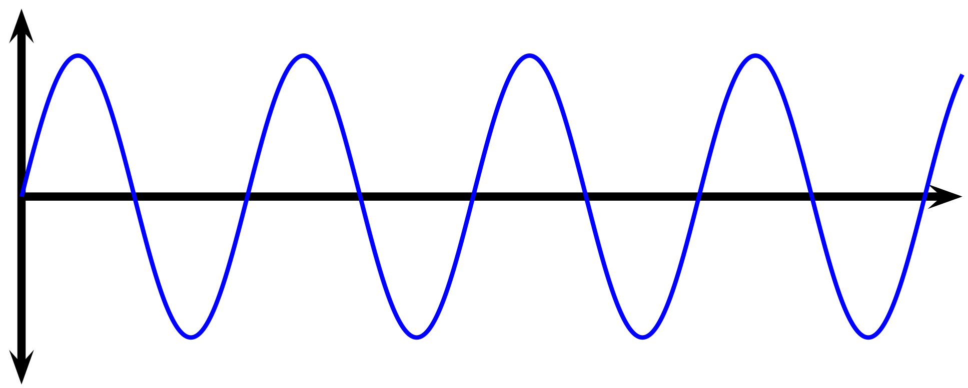
Analog transmission became popular with the discovery of AC current and Electromagnetism. Telephone systems, radios, TV were analog devices.

Digital transmission returned to supremacy shortly after the development of the computer.

[a] mirrors
[b] semaphores
[c] flags
[d] smoke
[e] DC

Continuous signalling

In a continuous signal the signal intensity has the same properties as a mathematical continuous function.



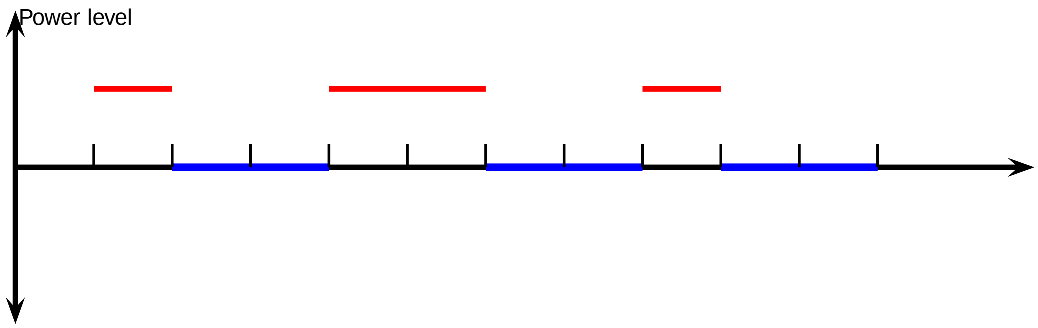
The absence of any signal, called silence, was a clear indication that “there is nothing out there” which could indicate one of two very different things:

- The sender has nothing to say.
- The sender is broken (or does not exist).

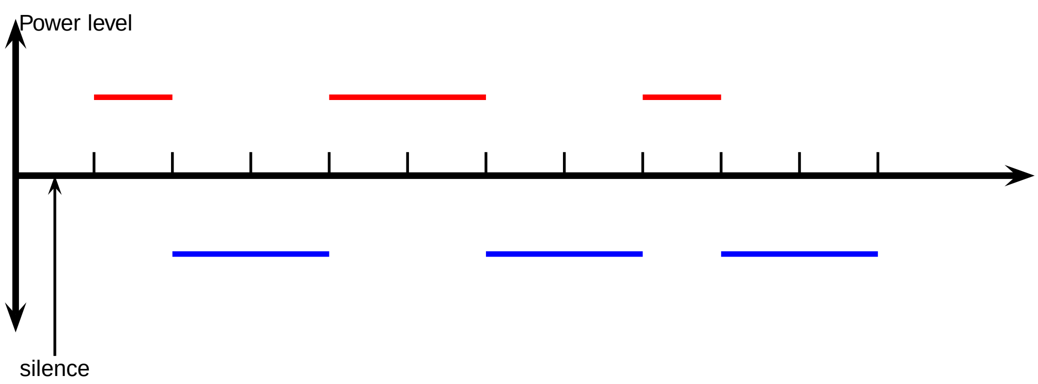
To distinguish between the two, an analog signal is (almost always) added to a wave, called **carrier wave** or just carrier, that is there to indicate the presence of the other side (the carrier is also used for modulation).

Discrete signalling

The signal has no derivative at some points (this is true at the sender).



Note that the use of a zero–amplitude signal is unwise, so that a more real–life discrete signal would look like this:



Comparison

- Discrete signals survive distortion much better: the receiver only needs to be able to determine the approximate signal level.
- Digital signals are repeated (by repeaters) differently than analog: a repeater sends a copy of the original signal, not of the signal it received.
- Analog signals do not require precise timer synchronisation. The receiver can pick the signal at any time.

Comparison

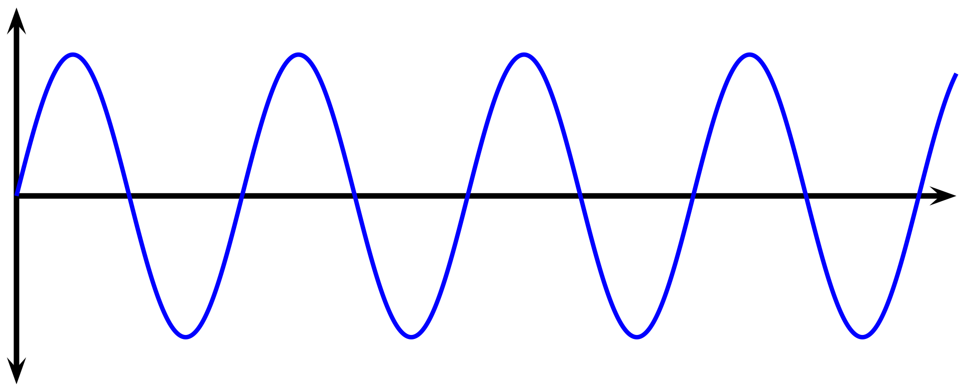
- Analog signals must arrive at the receiver almost intact.
- An analog repeater amplifies the signal it receives, increasing the volume of distortion as a byproduct.[a] Particularly relevant over long distances.
- The receiver of a digital signal must synchronise its clock to bit boundaries which is hard. In some digital codes, if the receiver does not pick the signal from the very beginning, it may be impossible to decipher the rest (differential codes).

---

[a] This is less relevant if a Fourier transform is applied to clean the signal.

Analog transmission

An analog signal is continuous by nature (it has an infinite number of levels of intensity over a period of time).

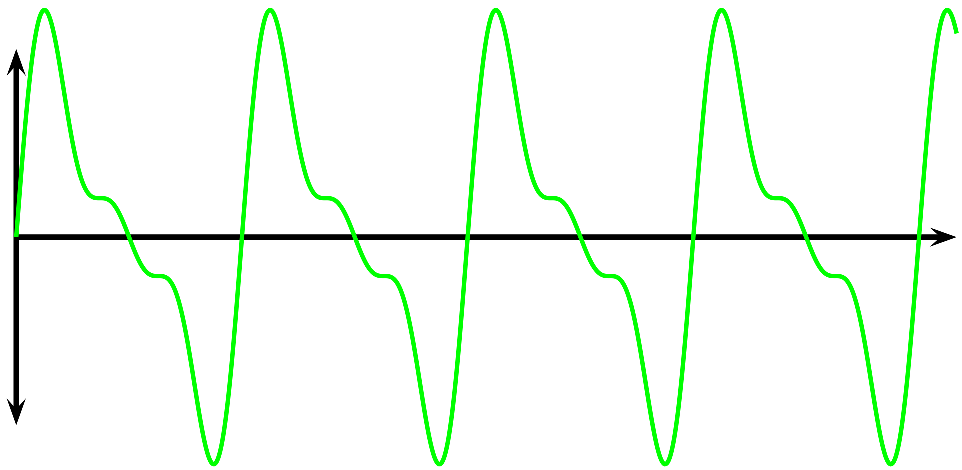


An AC power supply comes as a sine wave with a frequency of 60Hz (50Hz elsewhere). A frequency of 60Hz implies a period of $\frac{1}{60Hz} = 16.67ms$. Old computer hands will instantly recognise this period, because the power supply used to be employed as an external clock in old computers (the clock device would interrupt the CPU 60 times per second).

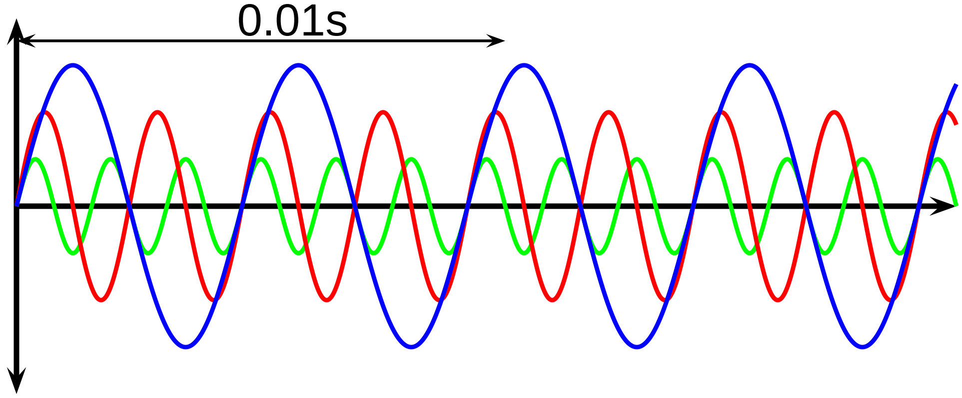
Since the days of **Fourier** (1768–1830) we consider any analog signal to be a composition of sine waves of different frequencies. Each sine wave has a specific amplitude (“intensity” or “volume”) and its own frequency; these two values carry the information (or data) to the receiver.

Any periodic signal can be decomposed into a combination of a countable number of different sine waves (countable in the $\aleph_0$ sense). An aperiodic signal has a non–countable number of components.

The signal:



is a combination of 3 sine waves (A3, A4, E5):



A periodic signal has a finite **pitch period** which is the smallest time interval after which the signal repeats itself. We normally use the inverse of the pitch, the fundamental frequency measured in Hz.

Clarification for those not following:

> Suppose a “perfect” singer sings the tune C4-A4 holding each note for 0.2 seconds.[a]
>
> What we observe is a period of time made of 52 pitch periods of fundamental frequency 262Hz followed by a short period of random noise, followed in turn by a period of time made of 88 pitch periods of f.f. 440Hz.
>
> In reality, no singer can avoid impurities, so that the actual sound will have a number of other frequencies, albeit with much smaller amplitudes.

[a] C4 is the “middle C” Also known as MIDI–60; it has a frequency of 262Hz. A4 is “middle A” (MIDI–69) and it is the famous 440Hz note.

Fourier visited briefly

Every periodic signal can be represented as a Fourier series (see MathWorld).

Consider a periodic signal v. Its amplitude (i.e. magnitude) $v(t)$ at time t is given by:

$$v(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos 2\pi n f_0 t + \sum_{n=1}^{\infty} \sin 2\pi n f_0 t$$

where

$v(t)$ is the signal level (e.g. in volts) as a function of time.

f_0 is the **fundamental frequency** component of the signal measured in Hz. All other components of the series have frequencies that are multiples of f_0; they are called **harmonics**.

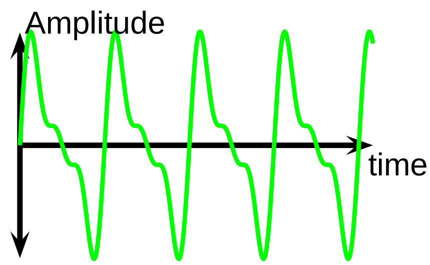
a_0 is the **constant component** of the signal (in electricity, the DC component, e.g. the mean voltage level over time).

a_i, b_i are Fourier coefficients (expressed in the same units as a_0).

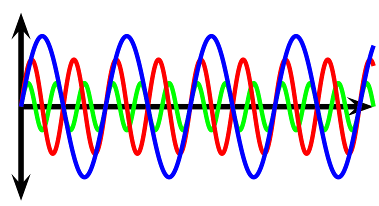
Domains

The common terms for the two basic representations of a signal are:

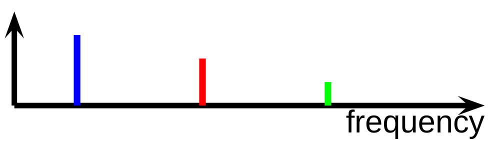
Time/Spatial domain



made of



Frequency domain



Amplitude can be expressed in various units, such as units of pressure (*atm*, *Pa* or *psi*), units of voltage (*V*) or whatever.

Fourier transform

Given a signal in time domain we can represent it in the frequency domain by applying to it the Fourier transform:

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x} dx$$

The inverse Fourier transform goes in the opposite direction:

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi i k x} dk$$

when x is a variable representing time in s, the transform variable k is represents frequency in Hz.

Fortunately, a discrete counterpart exists: the Discrete Fourier Transform. Its common form, the Fast Fourier Transform (FFT) is an efficient algorithm running in $O(n \log n)$ time for a sample of size n (these are measurements of the values of function f at n equidistant points).

Frequency domain and noise

The Fourier transform produces a frequency domain description of a signal; graphically it gives the magnitude of the signal by frequency used. If the signal is not flawless, it its frequency domain will show low–amplitude random noise; the usual way is to view noise as a function of signal frequency and define classes of noise based on the nature of that function.

If the independent variable (frequency) is f, noise can be called:

white noise i.e. uniformly distributed random noise with constant density (independent of f). White noise is an imaginary concept but exists in reality in approximated forms (note that the power of a white–noise only signal is infinite).

pink noise with a density proportional to $\frac{1}{f}$.

brown/red noise has a power density proportional to $\frac{1}{f^2}$

Noise frequencies can be removed digitally or using a special filter.

Bandwidth in analog signalling

The bandwidth of a composite periodic signal is the difference between the highest and the lowest frequencies contained in that signal. Note that a signal normally has an infinite number of frequency components, hence the highest frequency of a periodic signal is often ∞.

To transmit an analog signal faithfully one must have a transmission channel capable of transmitting all the possible frequencies (up to ∞). That is not realistic.

Fortunately

The total energy carried by a signal must be finite. This implies that if there are infinitely many non–zero values in the frequency–domain representation of the signal, all but a finite number must be infinitely small (this is mathematical rubbish but it is equivalent to a correct statement).

Hence, past a certain value, as frequencies increase, their amplitude decreases fast. This implies that ignoring these frequencies will have a small effect on the signal.

So, we reduce the infinite bandwidth needed for a signal to a finite one by discarding the higher harmonics, i.e. replacing the series

$$v(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos 2\pi n f_0 t + \sum_{n=1}^{\infty} \sin 2\pi n f_0 t$$

with an approximation:

$$v(t) = a_0 + \sum_{n=1}^{c} a_n \cos 2\pi n f_0 t + \sum_{n=1}^{c} \sin 2\pi n f_0 t$$

where c is cutoff of a low–pass filter applied to the signal before transmission.[a]

[a] Normally, the cutoff is not given in the number of harmonics, but as the cutoff frequency, the filter deleting anything with a frequency higher than the cutoff.

Filters

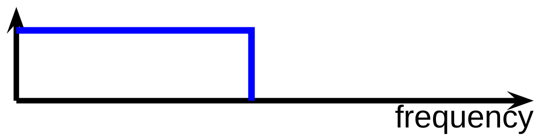
An electronic filter is a device that reduces the amplitude of signal of unwanted frequencies.

Filters can be analog or digital, active or passive; they can also be linear or non–linear, etc.

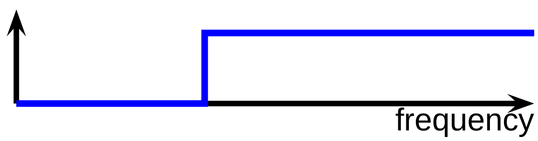
- A passive filter does not require a source of power.
- An active filter is a passive filter with an amplifier (of sorts) added; thus, it requires an external supply of power.
- A digital filter is not a filter–it is a processor that takes an input digital signal and produces another digital signal based on some algorithm. Digital filters are used for analog signals with the help of A/D and D/A converters.
- Non–linear filters are too complex for this course.

There are two basic passive filters:

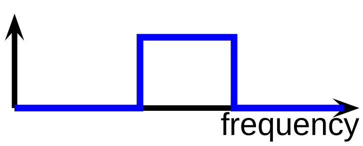
Low–pass filters that cut frequencies above a cutoff.

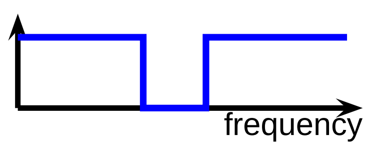


High–pass filters that cut frequencies below a cutoff.



These two can be combined to create pass–band filters and block–band filters.



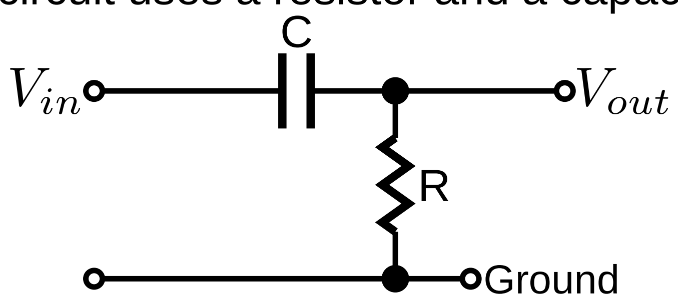


A simple filter

Passive filters are extremely simple devices; a most basic filter is made of one resistor and either one capacitor or one inductor.

A high–pass filter

This /red RC circuit uses a resistor and a capacitor:



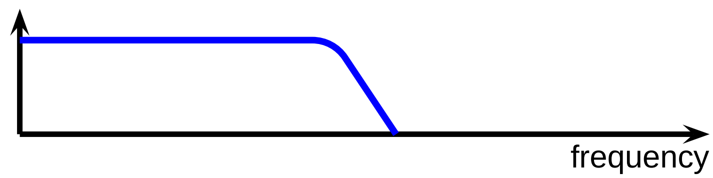
The cutoff frequency f is defined as the frequency for which the output power is half of the input (-3dB "gain"):

$$f = \frac{1}{2\pi RC}$$

f is in Hz when R is in ohms and C is in farads.

An RC low–pass filter has the capacitor and the resistor switched around.

Filters are not ideal; they attenuate gradually. For example, a low–pass filter attenuates all the frequencies above a cutoff frequency in a manner similar to:



Filters attenuate at a rate of 6 dB per octave (double/halved frequency) for first–order filters, 12 dB for second–order, and so on.

Some real–life data for a third–order high–pass 2.1 MHz filter:

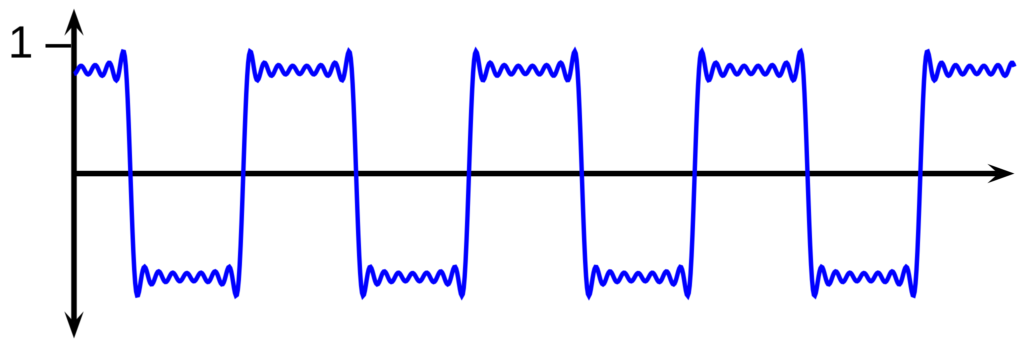
Freq.	100 kHz	500 kHz	1 MHz	1.6 MHz	2.1 MHz
Gain	-120 dB	-66 dB	-38 dB	-15 dB	-3 dB

A real signal

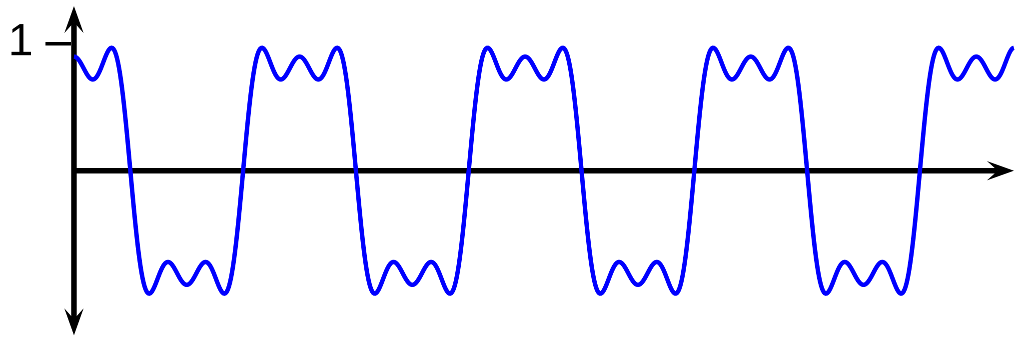
Suppose we are transmitting a signal that has a Fourier expansion of:

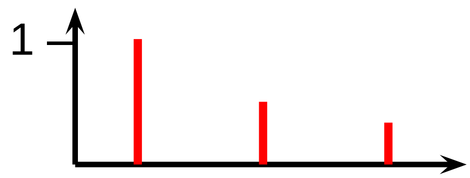
$$v(t) = \sum_{i=0}^{\infty} \frac{(-1)^i}{2i+1} cos((2i+1)t)$$

(this is a “classic” 0–1 sequence in a coding scheme that has no DC component).

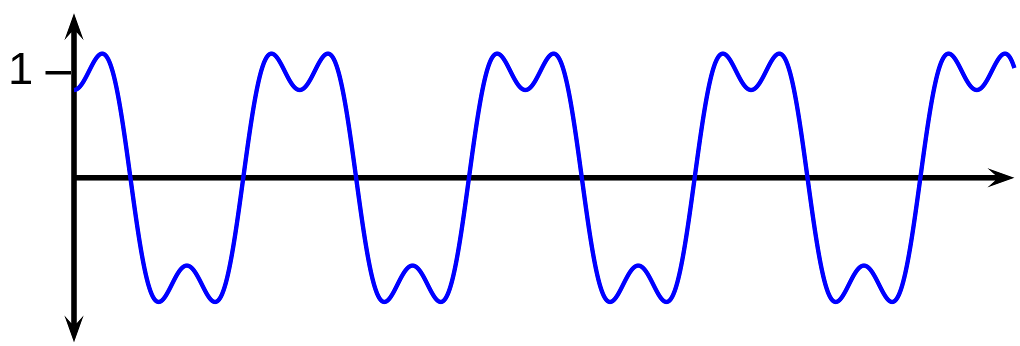


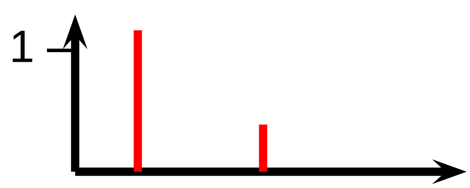
$$\cos x - \frac{1}{3}\cos 3x + \frac{1}{5}\cos 5x$$





$$\cos x - \frac{1}{3}\cos 3x$$

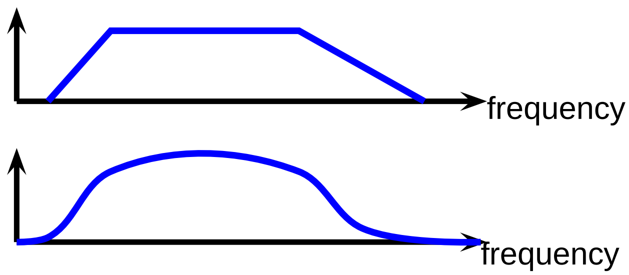




Aperiodic analog signal

A periodic signal has a frequency–domain representation made of a number (infinite but countable) of discrete spikes, each denoting a frequency that contributes to the signal.

The frequency–domain representation of an aperiodic signal is a continuous shape usually resembling somewhat a trapezoid:



The second example has an infinite tail of very small values (remember that the whole area of the figure must be finite).

Aperiodic analog signal

Analog TV is an example of analog transmission of aperiodic signals. The North American standard called for 525 horizontal lines of 700 pixels each (525×700 is a 3:4 ratio). Each pixel was either black or white which was represented by a low or high amplitude signal (simulating a digital signal).

If black and white pixels alternated in the whole screen, a bandwidth of 4.2MHz would be needed (taking into account interleaving and 42 lines lost in vertical retrace). Since it can always be prevented, TV stations were given a bandwidth of 4MHz per channel.

Likewise, AM radio stations which broadcast a combination of periodic and aperiodic signals were allocated 10kHz per station and FM stations were allocated 200kHz per station.

Transporting a signal from place to place must involve some physical phenomenon. Until quantum effects are finally employed in information transfer, we are restricted to traditional physical phenomena which all are, to some degree at least, analog.

So, digital and analog are not antonyms as much as they represent different ways to prepare data for transmission.

Digital transmission

The basic attributes of an electromagnetic wave, frequency and amplitude can be used to convey data in an analog setup (as will be seen, another attribute, phase, can be used as well).

In a digital world these attributes can be used as well, albeit in a different way. The attributes are labelled differently:

Bit duration (or bit/symbol period corresponds to the pitch period.

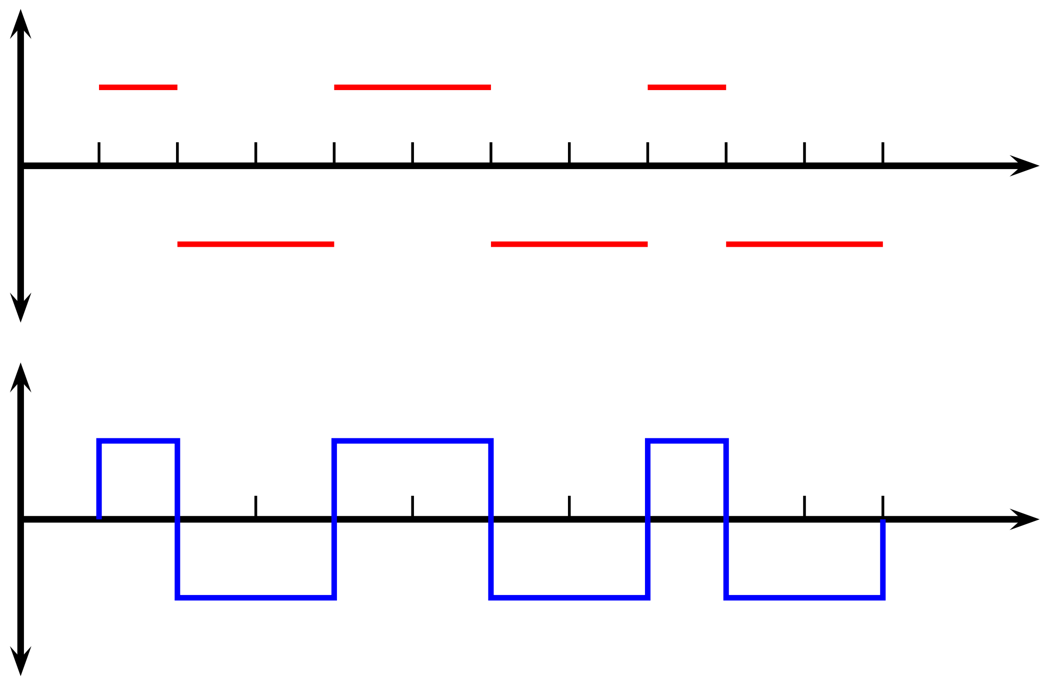
Signal level corresponds to amplitude.

Bit rate corresponds to fundamental frequency.

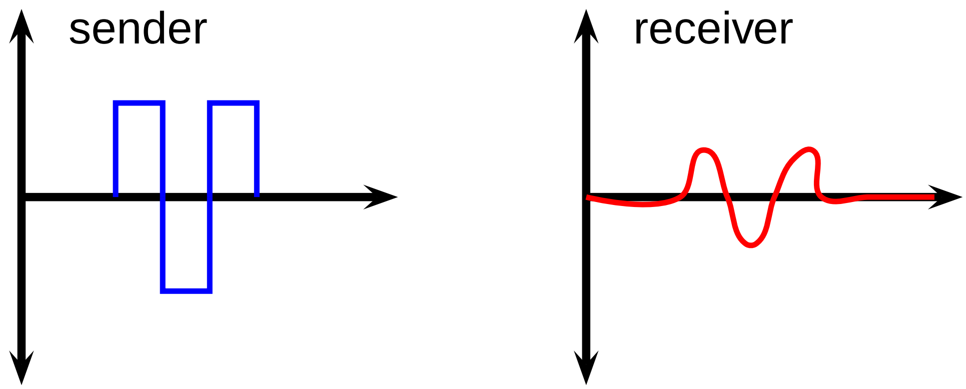
Bandwidth retains the name but the definition loses its accuracy.

Bit length is a new concept (blasphemous to physicists). It is the distance that a bit occupies in/on the transmission medium.
Bit length = propagation speed $\times$ bit duration
so that in traditional Ethernet, the bit length was $2{\times}10^8 m/s{\times}10^{-7} s{=}20m$.

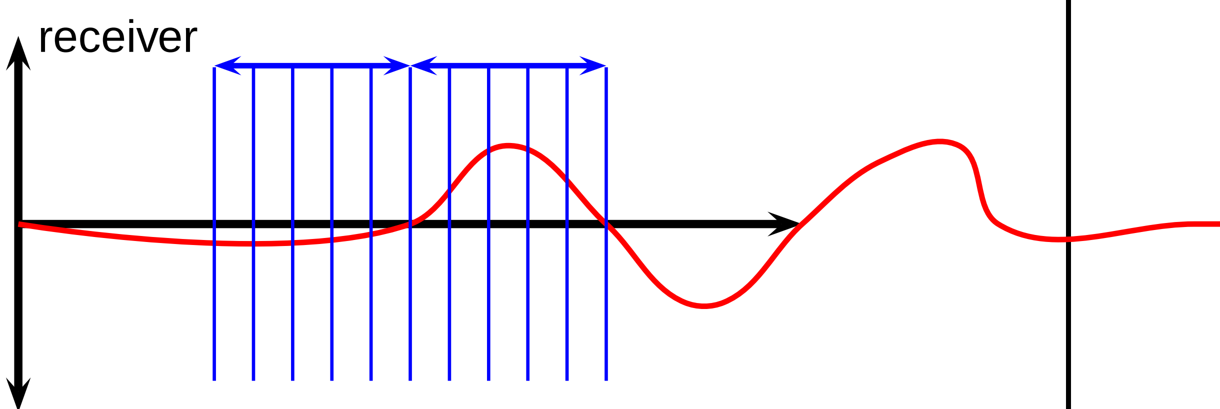
The simplest indicator of a digital value is amplitude and it remains the basis of all digital signalling methods.



Not so square



The receiver tries to figure out what it received by sampling the signal several times per bit duration:



The measurements made during a bit period are added and compared to a threshold value.

The key difficulty is to guess the bit boundary.

Multilevel signals

A digital signal can have any number of levels of magnitude hence more than one bit can be transmitted per period.

This requires more accurate instruments that sample the signal level; such more accurate instruments could also be used to reduce the duration of the bit period which has the same effect of increasing the effective bit rate.

Any increase in bit rate also increases the bit error rate, calling for fancier methods of error detection and recovery.

Baseband transmission

Baseband is a loose term that implies that a digital signal is sent directly (without being converted to analog using a D/A converter).

A low–pass filter reduces the bandwidth to fit a low–pass channel, a transmission medium capable of transmitting all the frequencies in a range $< f_0, f_1 >$ where f_0 must be very close to 0 Hz (ideally it should be 0). The value of f_1 is called the bandwidth and its value is related to the desired bit rate:

> If a bit rate N is desired, the bandwidth required by a baseband transmission is not less than $k \times \frac{N}{2}$ Hz where k is the highest harmonic that the low–pass filter will leave unattenuated.

Hence, Nyquist law for a noiseless channel:

$$Bit\, rate \leq 2B \log_2 L$$

where B is the bandwidth and L is the number of signal levels used. The bit rate reaches a maximum when only the first harmonic (fundamental) is passed by the filter.

A baseband signal could be sent *as is* but this is impractical for many reasons, one of them being that low frequencies suffer from distortion in a disproportionate manner.

Hence a baseband signal is typically added to a carrier frequency simultaneously (modulating that frequency) which is often chosen from a range called loosely radio frequencies RF (note that RF has nothing to do with radio broadcasting).

The use of modulation to transmit baseband signals is commonly called line coding although line coding is just one of many possibilities of using modulation.

Modulation

A signal is merged with a modulated carrier signal to give a signal in the RF band. The resulting signal is transmitted and eventually demodulated at the receiver. Two modulation techniques are in use: amplitude and frequency modulation

In amplitude modulation two symmetric copies of the baseband signal are added to the carrier signal:
see AM + baseband.

In frequency modulation a baseband signal modulates the frequency of a carrier signal resulting in a modulated signal.

Note that the signal being sent can be in digital or analog form.

Multiplexing shared media

When a network link is used concurrently by many (virtual) circuits, sharing is a necessity. It is done by dividing the channel among the circuits using it (circuit switching) or by encoding differently the transmissions sent to different receivers.

FDM (frequency–division multiplexing). Each circuit is assigned a frequency band within the range available in the link. The band is used continuously by the circuit.

TDM (time–division multiplexing). At regular intervals, the whole link is made available to each circuit for a fixed amount of time, called a **slot**.

CDMA (code division multiple access). All circuits transmit simultaneously, with each circuit assigned a unique code which spreads the signal.

Multiplexing in telephony

In a T1 1.544 Mb/s link sets 8 kb/s for control, with the remaining 1,536 Mb/s shared by 24 circuits, each circuit will get:

- One continuously available virtual link of 64 kb/s when FDM is used.
- One slot of size 8 bits every $\frac{1}{8000}$ of a second when TDM is used[a].

Telephone networks used FDM until the 1960s when PCM using TDM was introduced.

[a] It could just as well be 1 slot of size slot of size $\frac{64}{24}$ kb every $\frac{1}{24}$ second. Etc.

CDMA

CDMA (code division multiple access. Only a receiver possessing the same code can recover this signal; all others will hear it as noise. CDMA limits the number of circuits to the number of available codes, but does not waste bandwidth when moments of silence occur in some circuits. (See details) The method is used in wireless telephony (e.g. 3G).

Statistical multiplexing is not used at the Physical Layer although it is the method of choice within the layers 3+.

Unlike channel division multiplexing, statistical multiplexing does not **guarantee** an advertised bandwidth but gives a **best effort** bandwidth. CDMA is somewhere in between.