# *Redes de Computadores* (RCOMP)

Lecture 01

2017/2018

- Digital data transmission.
- Signals and transmission mediums.
- Digital signals and line coding.
- Analogue signals and digital modulation.
- Network communication.
- Network nodes and node address.
- Switching networks and shared medium networks.
- The packet concept. Payload and control information.
- Virtual circuits.

# Digital data

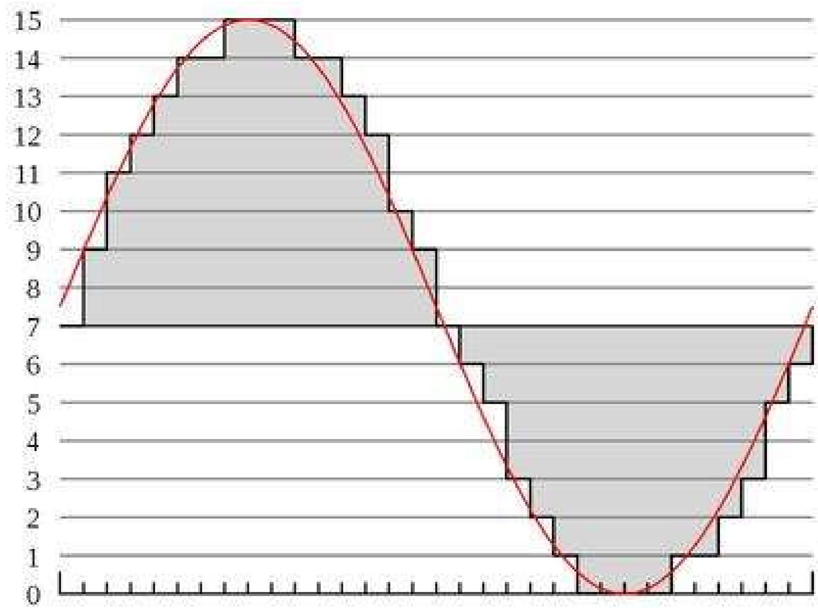
As we already know, current computers operate with digital data. Even though the real world is analogue, inside a computer it is represented as digital.

Real world analogue data is continuous, as close as two values are, there is always a value in between them, for instance, space and time are continuous. Storing analogue data would require an infinite storage capacity.

Digital data, on the other hand, is discrete. There is only a finite set of possible values, between two successive digital values nothing exists.

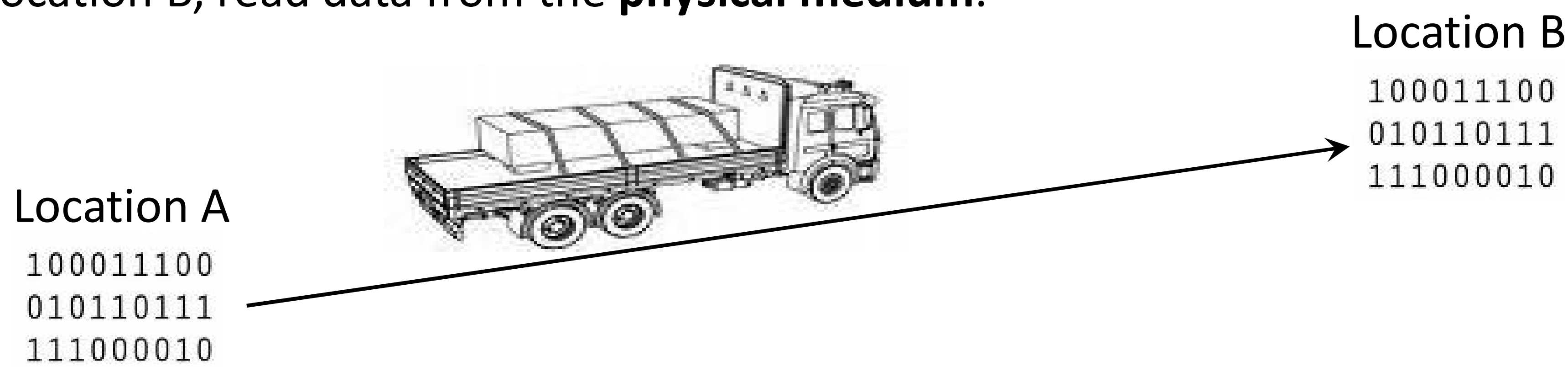
This means when we measure real world analogue data and convert it into digital data we lose a lot of information.

In the image we can see the conversion of a real world analogue data in red to digital data in grey, in this case, digital data can only take sixteen different values. Therefore, it could be represented using only four bits ($2^4$=16).



# Digital data transmission

Digital data transmission is the base stone of computer networks. We want to transport digital data from a location **A** to a physically apart location **B**. One approach would be, in location A, write data to a **physical medium**, then transport the **physical medium** from location A to location B and, finally, in location B, read data from the **physical medium**.

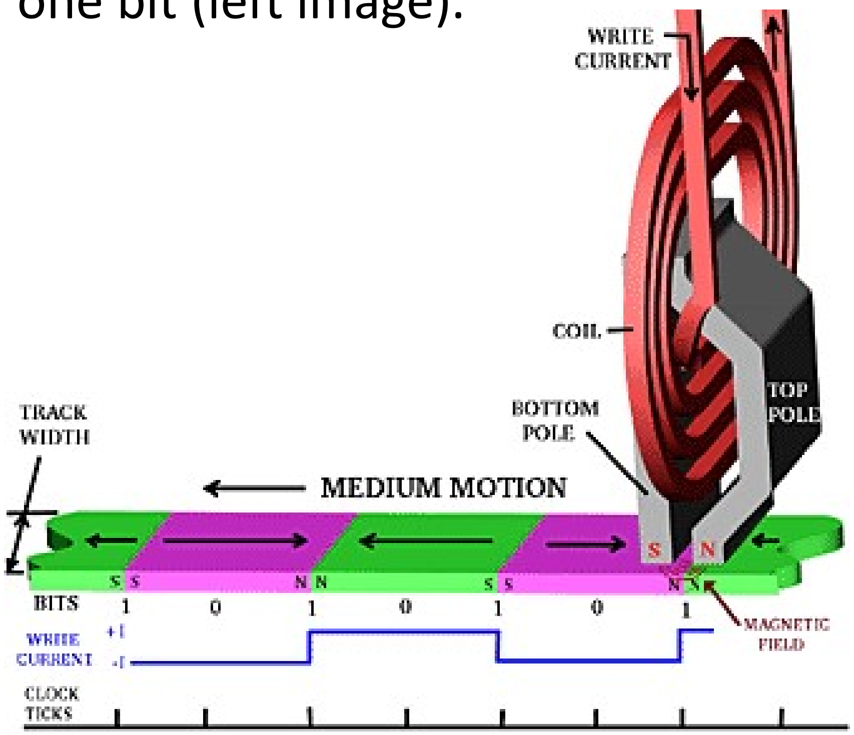


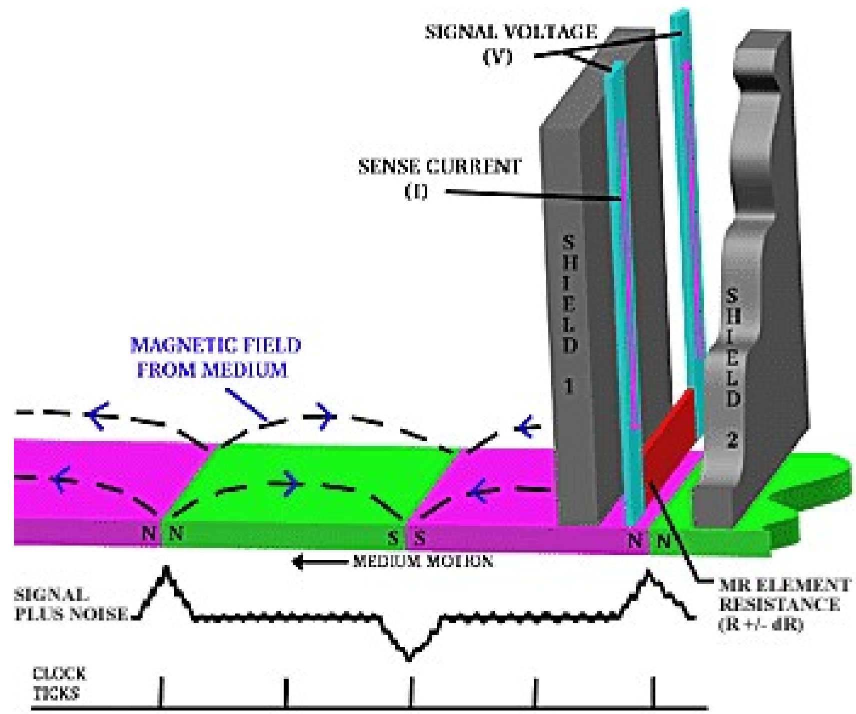

The so-called **physical medium** is some kind of material that, somehow, keeps on it (stores), a data representation (it could be a sheet of paper). Generally speaking, storing data in a physical medium involves changing a lasting physical property to represent the data.

Even though this is not a reasonable solution for data transmission, it provides a hint on how data transmission is achieved. So, let’s see some cases.

# Data storing case – magnetic fields

Over a magnetizable surface, typically a tape or a disk, a coil is used to create magnetic fields whose polarity depends on whether we want to store a zero or one bit (left image).



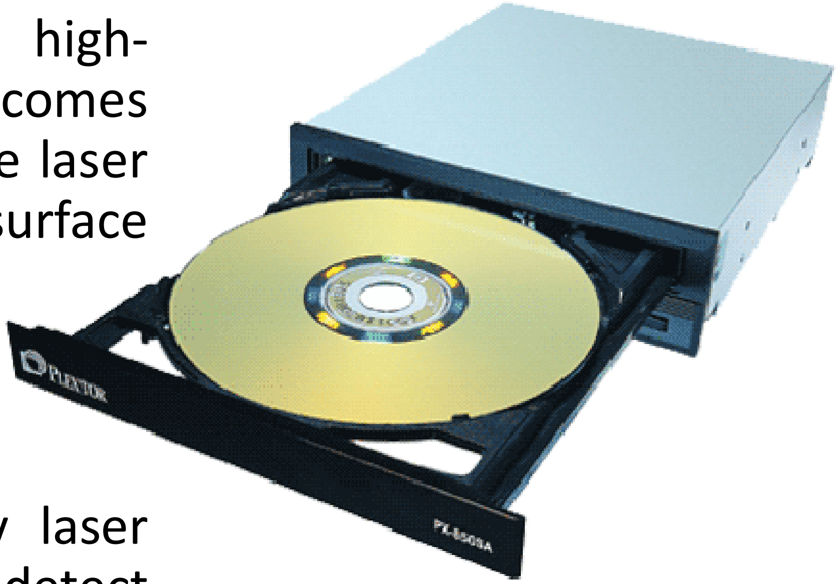

Later the recorded magnetic fields can be read to obtain the original information (right image).

# Data storing case – light reflection

In a physical surface point, we can use the light reflection, or lack of it, to represent a bit of data. This is rather similar to writing and erasing dots with a pencil on a sheet of paper.

A CD-RW (Compact Disc-Re-Writable) has a special surface whose reflective properties change according to the melting temperature.

An intense laser beam leads to an high-temperature melting and the surface becomes non-reflective on that point. A less intense laser beam leads to crystallisation, making that surface point reflective.




To read the stored data a low-intensity laser beam is used together with a sensor to detect the reflecting light or its absence.

# From data storing to data transmission

The use of physical materials properties is well suited for data storage, but it's not the best solution for the data transmission as it would require the transport of the physical materials where the data is stored.

Instead of physical materials, an **immaterial physical phenomenon** is used which also have physical properties that can be changed.

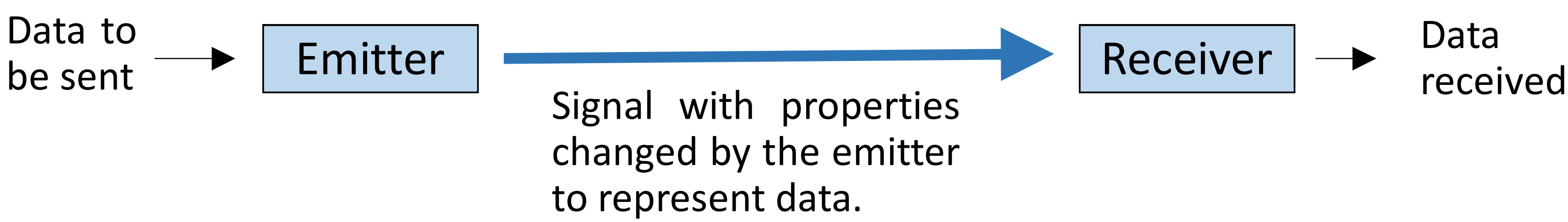
We look for a physical phenomenon that has the **ability to propagate itself in space** as this will solve our problem of data transmission.

This physical phenomenon will be called **signal**.

(usually it is also called **carrier** because it is used to carry data)

# Data transmission through a signal

The **emitter** produces a signal capable of propagating through a transmission medium and reach the **receiver**. However, the signal is not static, one or several signal properties are changed by the emitter in such a way they represent data to be sent.




The **receiver** listens for the signal, measures the properties the **emitter** is changing, and thus, gets the originally sent data.

Several types of signals can be used, like sound, light, electric current and electromagnetic radiation (radio waves). In practice, sound or inaudible ultrasound are rarely used.

Each of these signal types has properties that can be used to represent data, the most often used is intensity. For instance, a less intense signal can represents a zero bit and a more intense signal a one bit.

# Signals and transmission mediums

For each signal type, there is an ideal physical environment (transmission medium) that favours its propagation:

LIGHT - empty space, propagation is obstructed by any opaque object that may arise along the way. However, very thin glass fibres can be used to conduct light throughout meandering paths, also known as **optical fibres**.

ELECTRIC CURRENT – materials classified as good electrical conductors, notably metals. Given its properties (conductivity, mechanical characteristics, and cost), the most widely used is copper.

ELECTROMAGNETIC RADIATION (radio waves) – as with light, propagates better in empty space. However, unlike light, in some conditions is able to pass through solid materials without major problems.

# Radio waves

Electromagnetic signals are produced when there is a cyclical change of electrical current in an electronic circuit, typically attached to an antenna.

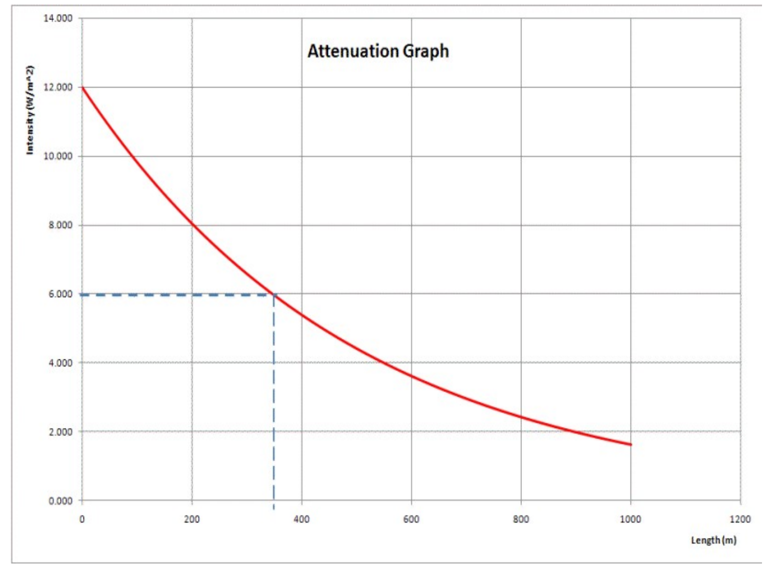
The number of cycles that occur every second (frequency) has a major impact on the signal propagation. Frequency is measured in Hertz (Hz), one Hertz represents one cycle per second.

Depending on their frequency they can be classified as:

- Low frequency (< 1 kHz). On air and empty space, the range distance is rather limited. However, it propagates well over electrical conductors (copper).
- Radio frequency (RF). Spreads in all directions and passes through most solid materials.
- Microwaves (> 2 GHz). Propagation begins to resemble light, not passing through and being reflected by solid materials.

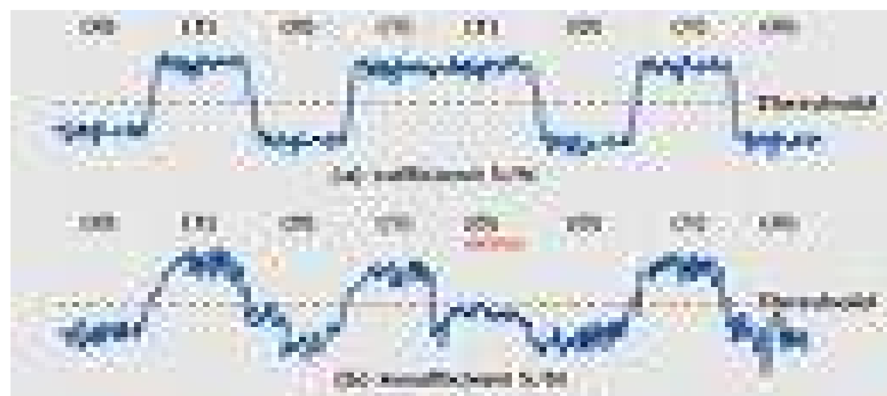
# Signals and attenuation

Any signal loses power as it propagates, this is called attenuation. The further away is the receiver is from the emitter, the less will be the received signal intensity.




Ultimately, at some distance the received signal intensity will be so low it becomes too hard for the receiver to measure the property used to transport data. This dictates the maximum reach of the transmission system.

In addition, there is noise. Noise is made of naturally or artificially introduced random signals that are mixed with the emitter's signal.



The key value to watch for is the signal/noise ratio (SNR or S/N) because this is an indicator of how difficult will be for the receiver to distinguish the emitter signal among the noise.

# Signals, noise and attenuation

The origin of noise can be internal or external, internal noise is produced by the transmission system itself, for instance signal amplifiers and transmission mediums are not perfect, they change slightly the signal shape, this must be regarded as noise as well. External noise comes from external sources, electrical and electromagnetic signals are greatly exposed to external noise due to the **Electromagnetic Interference** (**EMI**) phenomena.

The noise level can be higher on specific locations along the transmission medium, this depends on the local surrounding environment and is hard to anticipate. In a pragmatic approach, we will expect noise level to be roughly the same along the entire transmission medium.

From this we can conclude **the further away is the receiver from the emitter, the less will be the signal/noise ratio** (S/N).

This happens because the signal becomes less intense due to attenuation, and noise keeps approximately the same level.

Noise and attenuation work together to make the receiver's life harder. More than the signal intensity in itself, this is what settles the transmission system maximum practical reach.

# Signal and medium bandwidth

Electric and electromagnetic signals are subjected to further issues.

A steady sine waveform (sinusoidal) signal is pure because it’s made of a single signal frequency.

Any non-sinusoidal electric or electromagnetic signal is in fact the sum of several sinusoidal signals with different frequencies and amplitudes.

The set of frequencies a non-sinusoidal signal is made of is called the signal´s spectrum. Of course, the spectrum of a sinusoidal signal has a single point.

Issues around the signal´s spectrum arise from the fact that **transmission mediums do not behave the same way on all signal frequencies**. Namely:

- A transmission medium’s signal **attenuation** is not the same for all signal frequencies.

- A transmission medium’s signal **propagation delay** is not the same for all signal frequencies.

# Baseband and passband mediums

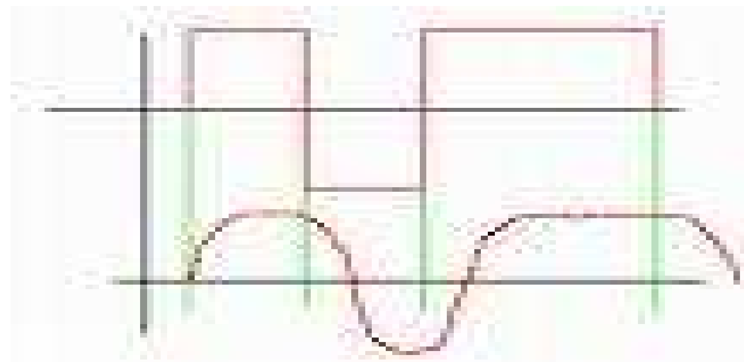
In a given transmission medium signal attenuation is not the same for all signal frequencies. Two types of communication channel can be named according to their behaviour on signal's attenuation:

**Baseband medium** – has a low attenuation for ranges of frequency that go from zero to a maximum value, thus a zero frequency (static) electric signal will have low attenuation. Signals with frequencies from zero up to the maximum value can be used. This is the case of most copper transmission mediums.

**Passband medium** – has low attenuation for ranges of frequency that go from a minimum value (greater than zero) to a maximum value, thus zero frequency (static) electric signals will not pass-through. Only signals with frequencies from the minimum value up to the maximum value can be used. This is the case of radio waves, however, some wired transmission mediums also have this behaviour, notably the public switched telephone network (**PSTN**), originally designed to transport human voice only.

# Inter Symbol Interference (ISI)

Different frequency signals propagate at slightly different speeds along the transmission medium. If the signal is not sinusoidal, different frequency parts of the signal will not reach the receiver with the same offset they had on the emitter, and thus parts of the signal from a symbol will overlap the next symbol at the receiver. This is called ISI.



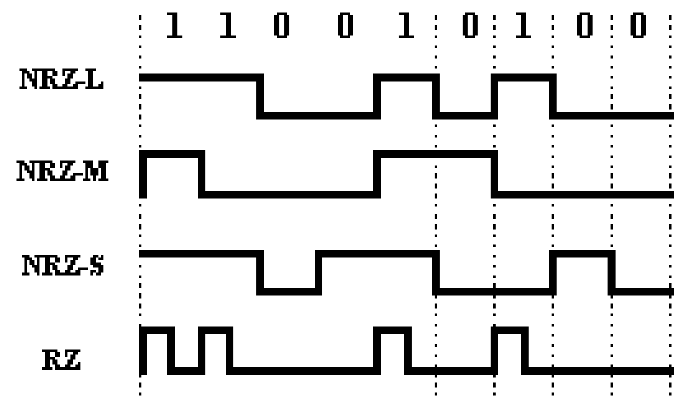
Both attenuation and propagation delay depend on frequency, so whenever a non-sinusoidal signal is sent the received signal will be rather distorted. **To overcome this problem, emitters apply a passband filter to transform the non-sinusoidal signal into a fairly sinusoidal signal before sending it.**

On the image above we can see a emitted digital signal and the received signal after passing thru a transmission medium. A digital signal has some fixed possible values and jumps between those values. This is the worst case scenario, a digital signal has frequencies from zero (when it is stable at a level) up to infinite (when it jumps instantly between levels). Of course, in reality, there are no level transitions in zero time, so the maximum frequency is not really infinite.

# Digital signals and line coding

The most straightforward approach for representing digital data in a signal is creating a digital signal directly from data. Techniques to achieve this are called **line coding**.

The image shows some of them. NRZ-L (Non-Return-to-Zero Level) means there are no automatic returns to the zero level, and also, each signal level represents a data symbol, zero bit is represented by a low level and the one bit by a high level.




In NRZ-M each one symbol is represented by a level transition, contrariwise for NRZ-S, each zero symbol is represented by a level transition.

One advantage of NRZ-L is more than two signal levels can be established, by doing that, more bits can be transmitted each time. With four levels a 2-bit symbol is transmitted each time, for 3-bit symbols 8 levels would be required. Generalizing, the number of levels required is $\mathbf{2^{(number\ of\ bits)}}$.

One other NRZ line coding advantage is it produces a signal with a maximum frequency equal to only **half of the data rate**.

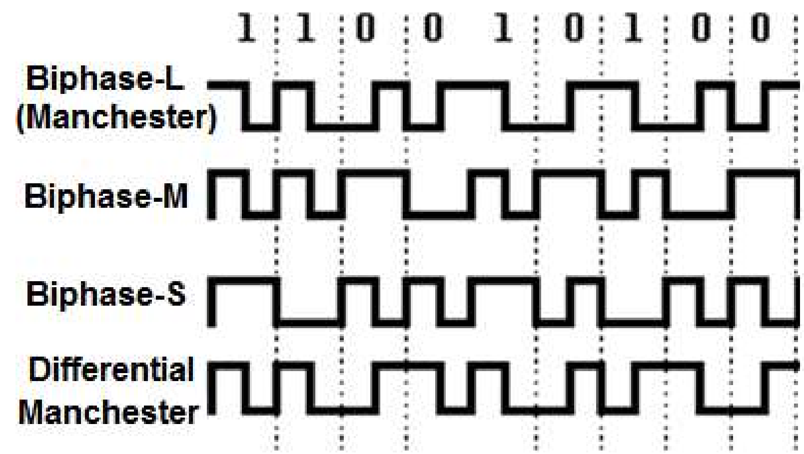
# Symbol synchronisation

We have already discussed why the digital signal arriving at the receiver is going to be rounded up to approach a sinusoidal form. This makes it even more imperative for the receiver to read the signal value at the centre of the symbol.

In other words, the receiver's reading clock must match the rate and offset of the emitter's clock. The best way to effectively guarantee that is providing synchronisation points to the receiver by changing the signal level. When the receiver detects a signal level transition, that is used to establish the local clock offset.

Under this point of view, previously described line coding schemes have a major issue because some symbol sequences keep the signal at the same level.

The image at right, show other line coding schemes to overcome this problem, they force level transitions, either in the middle of the symbol or between symbols. **The flipside is that now the signal will have a maximum frequency equal to the data rate.**



# Asynchronous and synchronous transmission

Oldest digital signal transmissions use NRZ and RZ line coding. Due to the synchronisation issue, the number of symbols that could be sent each time was very limited, only 7 or 8 bits of data (aka **character**).

Before real data bits, a start bit is sent, it changes the signal level and by doing so it allows the receiver to settle its clock offset. Because the time required to transmit so few data bits is very short, significant de-synchronisation never happens. This is known as **asynchronous transmission**.

**Synchronous transmissions**, on the other hand, use biphasic line coding. Since the signal level changes on every symbol, synchronisation is not an issue and very long sequences of symbols can be transmitted. In practice, for other reasons they are normally limited to some hundreds or thousands of bytes, called **frames**.

As increasingly higher data rates are required, the higher frequency limit of the transmission medium is reached. One approach to overcome this limit is turning back to NRZ, the signal frequency is only half the data rate and multiple levels can be used. The NRZ symbol synchronisation issue is usually overcome by inserting additional symbols among data symbols to ensure signal level transitions.

# Analogue signals and digital modulation

Creating a digital signal is the ideal and more straightforward way to transmit digital data, though, it can be used only on baseband transmission mediums. When using a passband medium like radio waves, zero and low frequency digital signals will not pass through.

In a passband medium, the signal must have the appropriate frequency and vary only slightly. The solution is creating a perfect sinusoidal signal with a frequency matching the passband medium, and then, change signal properties. This is called modulation, because we are using digital data, it's called **digital modulation**. Three properties of a sinusoidal signal can be changed by the emitter to represent data:

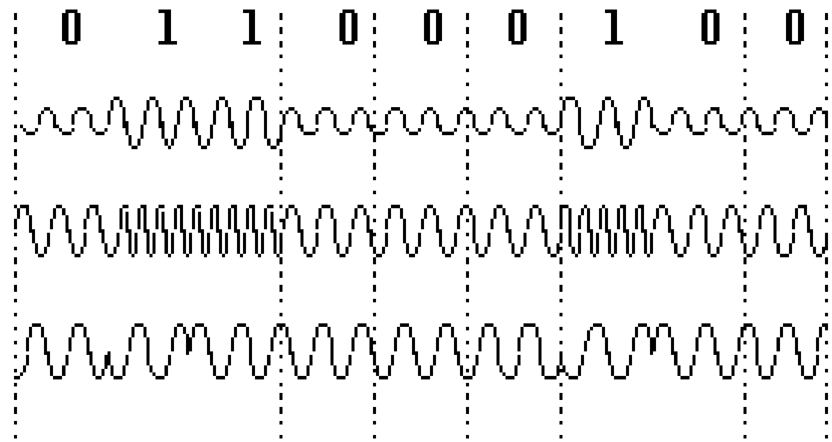
**ASK** (Amplitude Shift Keying) – the signal amplitude (intensity) is used to represent symbols.

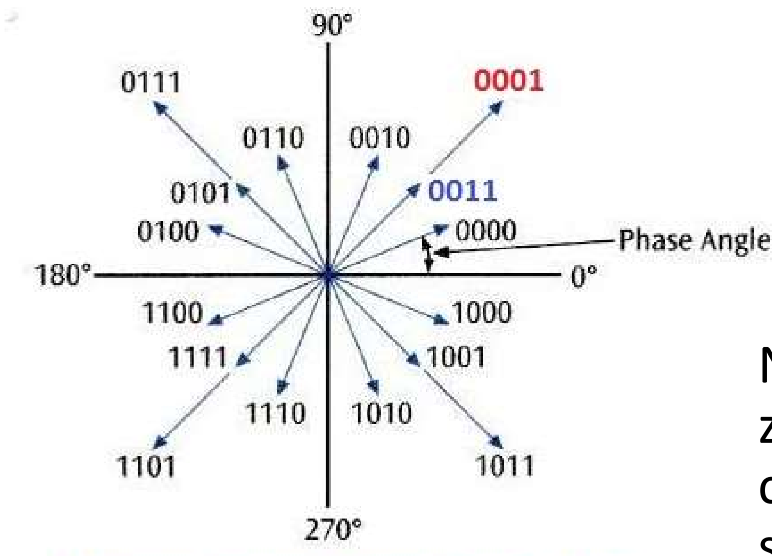
**FSK** (Frequency Shift Keying) – the signal frequency is slightly changed to represent symbols.

**PSK** (Phase Shift Keying) – the signal phase (offset) is changed to represent symbols, the phase shift is measured in the unit circle, zero or 360 degrees represent therefore no phase shift.

# Analogue signals and digital modulation

The image on the right represents signals with the three types of digital modulation. Each symbol to be transmitted is represented by the value for one property. Normally these modulation technics are used together to obtain a more extensive set of possible values, and thus, being able to transmit more bits in each symbol.



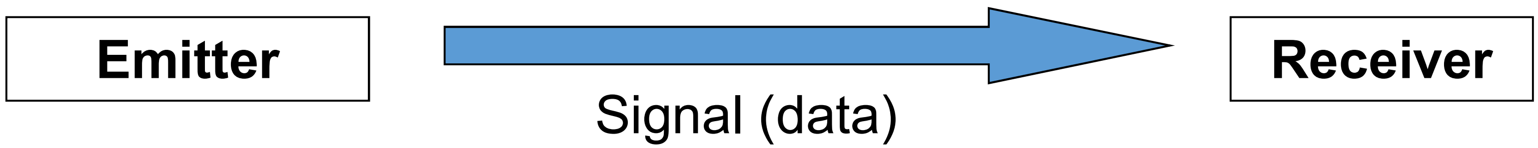

12 Phase Angles, 2 Amplitudes, 4 bits

The image on the left shows 16 QAM (Quadrature Amplitude Modulation) combining 12 different phase shifts with 2 amplitude values to obtaining 16 different signal values, hence, transmits 4 bits at each symbol.

Notice there is no symbol with phase shift zero, therefore, the phase will always be changing between symbols. This allows symbol synchronisation by the receiver.

# Network communication

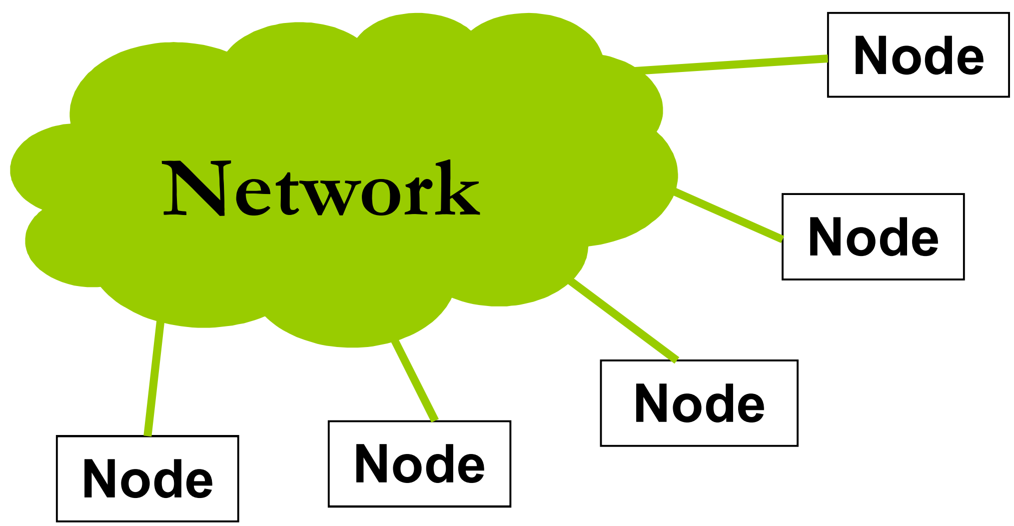
The goal of a computer network goes far beyond the transmission of data between a emitter and a receiver.




- In a computer network, players are both transmitters and receivers, they are known as **network nodes**.

- In a computer network, there may be a large number of nodes, any of them should be able to transmit data to any other node in the network.

Nodes are data emitters and receivers connected to a data transmission infrastructure called network.



# Network nodes and node address

A network can have many nodes (the internet has more than 100 million nodes). Each node is a potential data sender and receiver, so it must have an associated identifier, unique throughout the whole network.

**The unique identifier each network node must have is known as node address.**

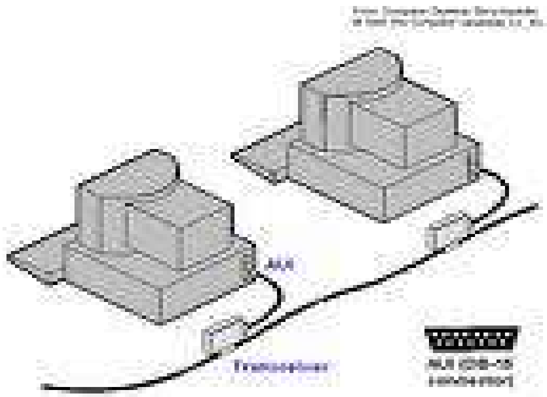
When a node sends data to the network, it must identify the destination node address. The network role is ensuring sent data is received at the correct destination node.

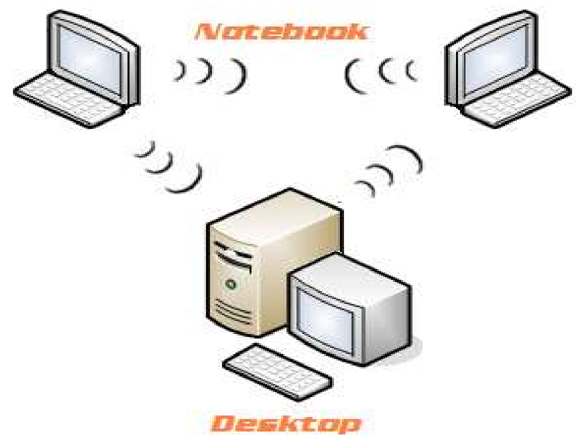
A network can be very simple, it can be just a shared transmission medium like a shared copper cable, also wireless networks use a shared transmission medium. They are called **shared medium networks**.

Other networks are more complex and are made of a set of interconnected intermediate nodes, they are able to determine the path to reach the destination node. These are called **switching networks**.

# Shared medium networks

If the transmission medium is shared among all nodes, then the network role is already accomplished: ensure data is received on the destination node.






This is not, though, a perfect solution:

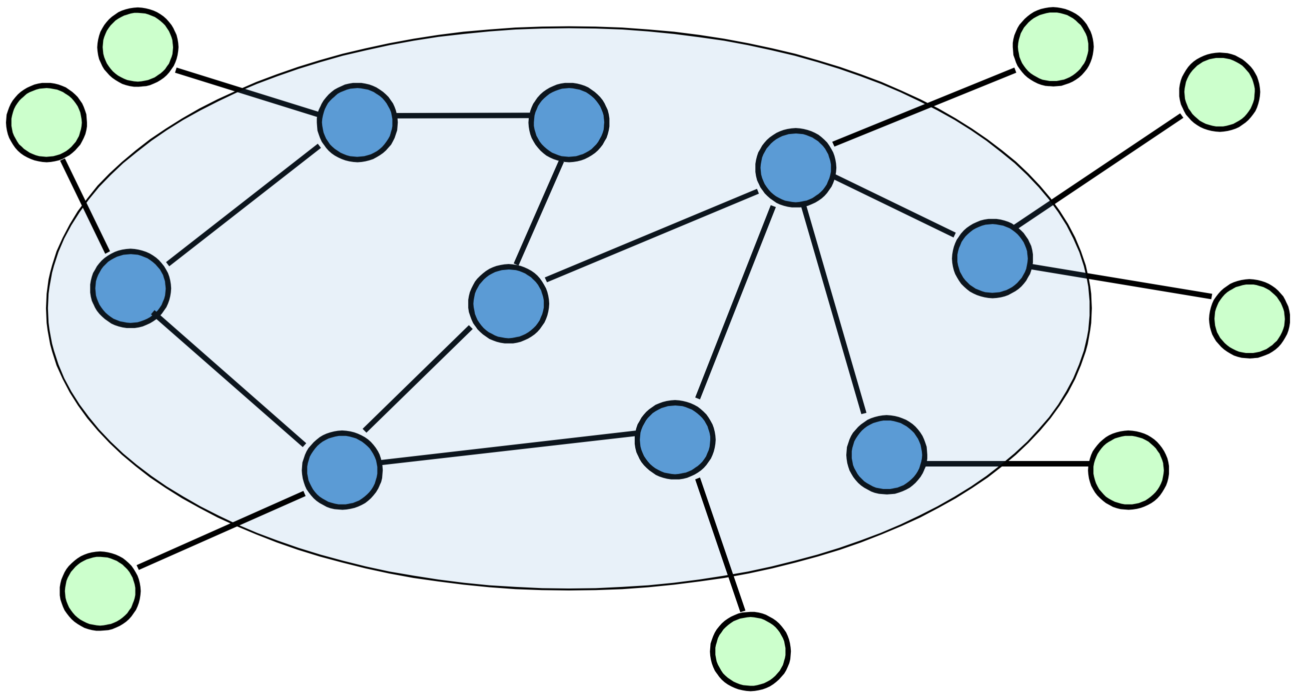
- Data reaches all nodes, it's up to each node to check if it's addressed to it or not. Beyond privacy issues this is also deteriorates network performance because the network will be busy transmitting data to places where that data is not required.
- Only one node can emit at a time, therefore the medium rate capacity must be dived by the nodes. A MAC (Media Access Control) mechanism must be deployed to ensure only one node emits at each time, otherwise, there will be a collision.

# Switching networks

Shared medium networks have important issues. Even assuming media access control works perfectly, the network transmission rate will always be divided by the number of nodes. Although cabled shared medium networks are now rare, wireless shared medium networks are widely used. The rule to follow for efficiency on these networks is keep the number of nodes as small as possible.

Switching computer networks are more complex, they are based on a set of **intermediate nodes** with multiple links between them:

Intermediate nodes have the important mission of making routing decisions based on the data destination address and decide which is the next node data should be sent to.

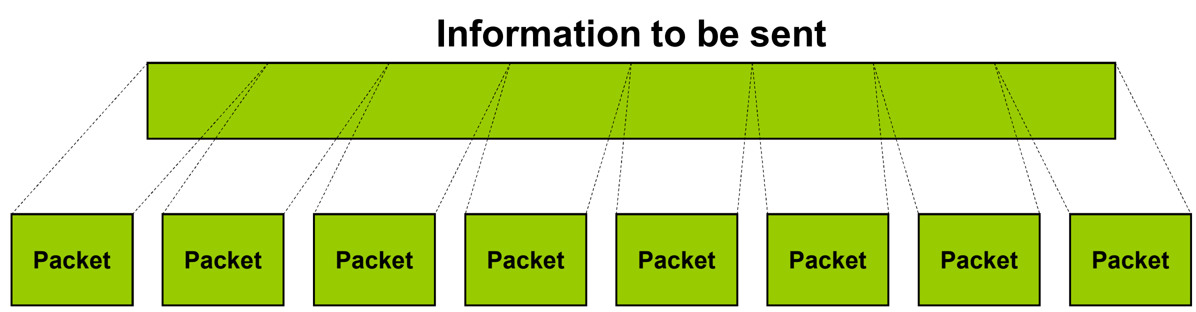


# Packets

Networks impose certain rules to nodes. A network is an infrastructure shared by multiple nodes, these rules are important to avoid unevenness in its use.

**One of the rules to almost every type of network impose to their nodes is a limit on the <u>maximum amount of data sent each time.</u>**

If unlimited amounts of data were allowed, this could disrupt the other nodes network infrastructure usage. This limit, far below the needs of most applications (e.g. a file transfer), forces the division of the information to be sent into smaller pieces usually called packets.



# Control information

Packets, however, are required to carry more info than the useful data coming from the application, this is called control information. Main control information is sent before data and is called control **header**, in some cases, additional control information is sent after data, this is usually called the **tail**.

| Control Header | **Data (payload)** | TAIL |
|---|---|---|

The packet payload is the useful data the application wants to send.

Packet overhead is defined as: **(packet size – payload size) / (packet size)**

Packet efficiency is defined as: **(payload size) / (packet size)**

Among the information present on the control header we usually have: the **source node address** (this is required for the receiver being able to send back a reply), the **destination node address** and a **label** to identify the type of data being transported on the packet payload.

The tail is in some cases used to carry an error detection code, this makes sense because the error code is useful only after receiving the payload.

# Packet switching with virtual circuits

In a switching network, each packet follows its own path, thus, packets may follow different paths and overtake each other. <u>Packet sequence is not assured</u>.

When a data transaction is divided into several packets, the network is not aware those packets belong to the same transaction. The receiver will end up with the additional task of reordering received packets.

Virtual circuits, also called virtual channels, avoid this problem.

**A virtual circuit is a path between the source and destination node, it's settled before sending data packets.**

1. The source node asks the network to create a virtual circuit to the destination node whose address it provides.
2. Intermediate network nodes settle the path and assign it a virtual circuit identifier. The network returns the virtual circuit identifier to the source node.
3. With the virtual circuit identifier, the source node can start sending data packets. Each packet header carries the **virtual circuit identifier instead of the destination node address** as would happen on normal packet switching.
4. Intermediate nodes forward the packets throughout the pre-established virtual circuit, all packets follow the same path, and sequence is therefore guaranteed.