9 Data Transmission and Storage

9.1 Preliminary Discussion

This subject is so vast that any attempt to address it within the limits of an introductory book has great chances to fail. A way to increase our chances of success is to reduce the subject to a few topics strictly linked to the instrumentation used in environmental sciences, and to the description of some examples of applications. In order to better understand this subject it is recommended that the reader be familiar with the way analog signals are converted into digital ones (Section (3.6)).

All measuring systems produce data as a result, and these data must be evaluated sooner or later, either by an automatic system or a human being; "the final aim of a measurement process is taking a decision" (Ferrero, 2005). The immediacy with which the information should be assessed depends on the purpose for which it was collected. This purpose defines what to do with the information in the steps following the measurement: basically the information generated by an instrument can be stored, transmitted or both.

The storage of information is treated at the end of the chapter, the transmission of data being the backbone of this chapter. It begins with a general introduction and goes on with generic issues on digital data with the purpose of establishing some concepts and the vocabulary frequently used. This generic approach to data communication concepts would hopefully help the reader to understand the transmission methods used at present in scientific instruments. Some characteristics of the transmissions systems, such as transmission delay, local and remote transmissions, network topologies, etc. are described.

A concise mention of analog transmission is done followed by digital data transmission which, due to its importance, is the matter developed most extensively. Several digital transmission concepts, such as signal encoding, transmission modes, serial and parallel transmission, asynchronous and synchronous transmission, error detection and correction, etc., are developed with the aim of providing the minimum needed information to understand how instruments can transmit data.

Finally, some effort is devoted to describe the three most commonly used media to transmit data from a field instrument (or group of instruments) to a central station, namely, private networks, digital telephony and satellite communications.

It has to be stressed that in order to keep the explanations as simple as possible, the descriptions found in this chapter are basic. We do not delve into the details that some communications systems employ to transport data more safely. At present, complex communications systems employ very refined strategies and algorithms to protect the data and correct errors. These approaches, which increase transmitted data immunity and integrity, are merely delineated.

9.1.1 Standardization

Communications in industrial hydraulics have well defined temporal and spatial scales (they need few samples per second, and instruments are commonly disposed in a relatively small area). These characteristics have made it possible to develop communication standards (Modbus, Fieldbus, Profibus, etc.) that are followed by instrument manufacturers. Therefore, to solve the problem of data transmission, the industrial user has only to connect the instrument to the network and setup some parameters in the software. Data transmission in industry is well documented in specific literature and its description is not addressed in this chapter.

As for the transmission of data in environmental sciences, there is a tendency to standardize data management and archiving in specific areas of sciences where data have common characteristics (i.e. marine geology and geophysics). However, due to the very different kinds of data generated by diverse instruments it is difficult to agree on a unique way of data storage and transmission. Moreover, different communication systems have very unlike capabilities of data transportation, which would impede adopting a single transmission way for environmental data transmission. Therefore, in environmental sciences there exist diverse ways of data management that will be addressed below.

9.2 Introductory Concepts

9.2.1 Data Storage

Sometimes the information about a natural phenomenon or an industrial process is collected only for statistical studies or for further analysis. For example, rain precipitation measured for agriculture purposes is averaged over days or weeks and there is no need for the collected data to be processed instantaneously. Also, when directional wave energy is measured for studies on coastal erosion it is necessary to collect information during some big storms before arriving at some conclusion; data is thus not immediately needed. In the above examples it is only necessary to store the information in some kind of storage media for being retrieved and analyzed at a later stage. The process of safekeeping the information for a future use is referred to as **data storage**. The connotation of storage is very broad, even if we reduce its application to the field of instrumentation. Data storage may be performed either in a memory on board an instrument or on a database in an information processing center, which might be located very far from the data collection or measurement point.

9.2.2 Data Transmission

On some occasions real time knowledge of the information is required. For example, when rain is measured in a city with the aim of forecasting floods to evacuate its inhabitants, the data has to be evaluated in real time; when directional waves are measured to decide the transit of ships at the entrance of a harbor, the information has to be available within a few minutes.

In the above examples data has to be known soon, and most of the time measuring systems are not at the same places as the processing and decision maker systems. Therefore, the information has to be conveyed from one place to the other. Distances may be from some meters to thousands of kilometers; the process of sending the information from one point to the other is referred to as **data transmission**.

Also, the significance of transmission is very broad. It may be called the mere action of transferring data stored in an instrument memory to a computer memory adjacent to the instrument by means of a cable. But it can also mean the process of transferring data from a remote field instrument to a processing center through different media; for example via telephone, satellite and Internet networks.

There are different reasons why data generated by instruments has to be transmitted. Very often it is essential for measuring instruments to have the ability of transmitting the information that they generate because they are not accessible and real time knowledge of the information is required, such as in the forecasting of tropical storms. Instruments for measuring snow accumulation in high mountain areas or rain in oceanographic buoys, equipment attached to animals to study their displacements and habits, and obviously instruments onboard satellites, all need to transmit data in order to render the measurements useful.

Besides technical reasons, there may be economic reasons that compel the transmission of the measured data. Simply, it could be cheaper to transmit data than to go to the place where the instruments are installed to collect it. Multiple circumstances may call for data transmission.

9.2.3 Transmission Delay

Transmission of information on a physical medium has a minimum **transmission delay** between the emission and reception instants, which depends on the speed of wave propagation. We stated that measurements are finally used in a decision-making process, and then the transmission delay becomes significant only if it slows down this process. For many purposes the transmission delay due to wave propagation is negligible. In a metallic wire the information is transmitted as an electrical quantity (voltage or current); on a radio link it is transported by an electromagnetic wave that propagates in air or vacuum; and in an optical fiber the information is carried by light.

In all the above examples the propagation speed is close to the speed of light (3×10^8 m/s). Delays are quite short even in long radio links as those through satellites.

It should be noted that even when the wave propagation speed is fast, the data transportation system might introduce additional delays. For example, the time delay between a cellular phone set and a relay station placed 10 km away may be 33 μs, but we all know that a text message between two phone sets may suffer a considerable delay (minutes). It is due to the priority assigned by the phone system to the message. Then in some cases delays introduced by the communication systems have to be avoided. For example, because tide information changes slowly, the transmission system delay would perhaps be unimportant and text messages may be used to transmit it. But if a tsunami warning has to be sent, we would wish no delay introduced by the communication system, and other media should be used.

So far we have focused on communication systems using electromagnetic waves. Acoustic communications are performed by means of a mechanical wave (vibration) that travels generally on a fluid (air or water). The propagation velocity in water is about 1,500 m/s. Thus, underwater acoustic transmissions of information from an instrument to the receiver would introduce a much higher delay than electromagnetic transmission does. Also, the bandwidth of an acoustic link is very restricted, and then the amount of data that can be transferred per unit time is quite limited (about 30 kbit/s). Remember that the bandwidth describes the ability of a system to allow the passage of a signal without significant attenuation and distortion (Section (2.4.4)), and that bits are materialized by pulses of a given frequency. Then, to allow high frequency signals (great amount of data per unit time) to be transmitted, a great bandwidth is needed (http://www.evologics.de). In summary, the capacity of an acoustic communication channel is much less than that of an electromagnetic channel and the delay is much higher. Anyway, in spite of these drawbacks, acoustic links are one of the few ways to communicate underwater instruments, and autonomous underwater vehicles between them or with the surface.

9.2.4 Data Quality and Digital Information

There is a direct relationship between the amount of digital information to be stored or transmitted and the quality requirements for the information being measured and acquired. The first question to answer, even before buying an instrument is: what is the quality of data we need? The answer to this question will define part of the cost of the measuring process. Overestimating data quality could lead to misuse of resources.

In order to know the resources they will need when preparing a measuring campaign, researchers should evaluate the amount of bytes they will store, transmit or process. This amount of digital information is proportional to the wanted resolution and the desired samples per second. Once again, transmitting slow varying phenomena like tide requires less byte than transmitting ocean wave information.

High quality data (high resolution and high sampling rate) requires more bytes per second than a low quality data.

At present, the bottleneck of the field data acquisition flow is the capability of field data processing and transmission. Nowadays memories have great storage capability, and sometimes this leads to the acquisition of more information than needed. Occasionally, it is worth acquiring more information than considered necessary because redundancy prevents data looses due to instrument failures or an incorrect pre evaluation of the temporal scale of the phenomenon being studied. Redundancy in the sense of the field deploying more instruments than are strictly needed is a good practice because a research effort could be incomplete due to the interruption of temporal series if just the minimum amount of required instruments is rigorously used. Also, acquisition of redundant data or high resolution data may be useful for future reprocessing of the information by means of new algorithms.

However, the exaggerated acquisition of information may lead to an unnecessary waste of time and money. When a great amount of data is acquired, a price is paid when the data has to be transmitted. Transmitting large files using satellite channels or cellular networks, which use small-bandwidth radio links, takes long times and generally increases costs. One possibility is to store high quality data but transmitting just that which is necessary. In this case the price is paid in going to the remote site to collect the data. Also, acquiring data with more samples and higher resolution drains more current from batteries and more energy is required. In addition, frequent recharge of batteries may shorten their working lives. Therefore, there is a compromise between quality of the data, storage and transmission, which has to be solved for each particular case.

If it is known that a certain phenomenon has, for example, a maximum frequency of 1 Hz, it would be reasonable to take from 5 to 10 samples per second (sps) to reconstruct the signal easily, but it would be unprofitable to take 100 sps "just in case". Also, if the error of the measuring system is on the order of 5 % it could be useless to store a 10 bits data word for each sample, it would be enough to use only 8 bits (which has a resolution ≈ 0.4 %, Section (3.6.6)). These options (reducing sampling from 100 sps to 10 sps and the resolution from 10 to 8 bits) could reduce the storage, transmitting and processing resources between 12 and 20 times.

The selection of adequate requirements of quality for the data may be crucial in real time processing. The time delay introduced by running models could be decreased by a clever selection of the quality of the information to be acquired.

9.2.5 Local and Remote Transmissions

In the past, local systems were confined to a reduced area such as an industry or a city, where most of the communications were conveyed through copper wires. Because these wires introduce signal attenuation as their length increases, when attenuation becomes significant the signal can be disturbed by noise, thus losing valuable information. Therefore, long distance communications were handled by radio frequency. Then in general, copper wires and radio frequency were synonyms of local and remote transmission systems, respectively.

The arrival of fiber optics (f.o.) with the capacity for transmitting large amounts of data, made it convenient to install very long f.o. cables to replace radio links in remote transmissions. The capacity of optical fibers for transporting huge amounts of data is so great that it became profitable to install transoceanic cables for voice and data communication services. Thus, at present we have communication services over very long distances (remotes) through optical cables. Fiber optics are not affected by electromagnetic noise and do not conduct electricity. Thus, they are very useful even in short distances, when signals have to be transmitted in an electrically noisy environment, or electrical isolation is needed between transmitter and receiver.

Laboratory networks where many instruments simultaneously collect information in a computer are examples of local data transmission. Hundreds of instruments may also be connected in industry to a control process unit through a local network. Networks of computers in universities are also local networks. In these cases short transmission distances are involved (tens to hundreds of meters) and they can be materialized by means of cables, either electrical or optical. Sometimes they are called local area networks (LAN); they have very high data-transfer rates and generally the network is the property of the users, meaning that they do not need to use public service networks. A known representative of the technology employed for these networks is the **Ethernet standard**, which transmits over twisted pair, wireless and optical fiber cable.

9.2.6 Examples of Data Transmission System Selection

The selection of a data transmission system depends on the distance between transmitter and receiver, the amount of data to be transmitted, the noise that could compromise the signal integrity, the maximum time allowed between the data is sent and received, the economic or symbolic value of the data, the transmission service costs, etc. Even when some general rules exist, each transmission problem has it particular solution because the communication's media availability depends on the geographic location of the transmitter and receiver.

There are some remote transmission cases with particular characteristics that constrain the communication solutions. For example, assuming an application case in which distances between transmitting instruments and the receiving unit are of hundreds of kilometers, measuring points may be suppressed or added periodically, and the amount of data to be transmitted is low. This case has a preferential solution based on radio links, either with radio stations spread over the surface of the Earth or through satellite links. A typical example of this case could be a meteorological network, which may be spread all around a country on land and sea. In this case the amount of data to be transmitted may be kept low because instruments can process the information on board, sending only the results.

A quite different case is the remote communication of an intelligent highway that may be hundreds of kilometers in length, and may use many measuring points installed at fixed places along the road. For this application, in general, colored images of the road are needed, which produce great amount of data to be transmitted. Usually, to manage the transit on a highway, the information is required in real time in a control center, and because of the huge amount of data, radiofrequency is not a good option. Since all the instruments are on the trace of the highway, a protection trench along the road may be constructed, and thus this problem is frequently solved by means of an optical cable.

As can be observed from the previous examples, in order to select a data transmission media, many considerations must be taken into account.

9.2.7 Network Topologies

Network topology refers to the structure of a communication network of equipments (Fig. 9.1). It is independent of the physical medium used to materialize the communication, and is related to the way in which each device "sees" the others. Each communication point in the network is called a node and each node has an address. Messages have a header with the address that indicates the node the message is intended for. The most well known communication physical network structures are the topologies called bus, star, ring, mesh and tree.

In the **bus** topology (Fig. 9.1a) all the nodes are connected to the same common transmission medium that can be as simple as a cable. The received header is analyzed by all the nodes, but only the specific recipient, i.e. that whose address agrees with the sent one, will take the message. If the common cable (bus) breaks, the system breaks.

The main characteristic of the **star** topology (Fig. 9.1b) is that there is a central node through which all the traffic passes, and each peripheral node is directly connected to the central node. If the central node fails, all the system fails.

The **ring** topology (Fig. 9.1c) circulates the information around the ring; each node receives the information and retransmits it until it arrives at the corresponding node. As in the bus topology, each node has an address, which permits the recipient node to be identified. If one node fails, the system can change the direction in which the information circulates in the ring, and it can arrive at the corresponding node by a different path.

The **mesh** topology (Fig. 9.1d) refers to a network where all nodes are directly connected to all the nodes of the network. It is a very robust network because if a node fails, the rest of the network continues working due to its great redundancy in communication channels, but the disadvantage is that it is expensive to implement.

The **tree** topology (Fig. 9.1e) has a hierarchical structure with a root node from which the other nodes are connected. There is a direct connection between the root node and the second level nodes. These second level nodes have a direct connection with third level nodes. The root node has a map with the address of the whole tree. Each node has the address of those nodes to which they are connected. If root node fails, all the system fails.

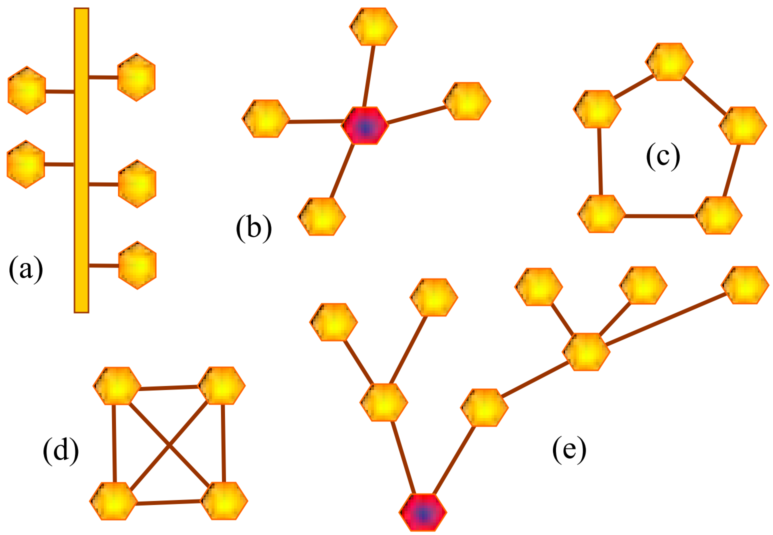

Fig. 9.1: Network topologies. (a) bus, (b) star, (c) ring, (d) mesh, (e) tree. Topologies (b) and (e) have central nodes which organize the network.

9.3 Data Transmission

9.3.1 Analog Data Transmission

Nowadays most of the information collected by measuring instruments is transformed to digital format before stored or transmitted, but some instruments still have analog outputs. Data storage and transmission may be done in an analog way; sending a current through a pair of wires or recording a signal on a paper chart are two analog examples, but today's methods are preponderantly digital.

The voltage output of a measuring instrument may be transmitted by means of an electrical cable just as it is delivered by the instruments' output amplifiers, but if electrical noise exists and the distance between the instrument and the receiver is of several meters, the information at the receiver could be noisy, so transmitted data might be modified. The simplest way to reduce the influence of the electric noise on the transmitted information is to use a cable less prone to noise, such as a twisted pair or a shielded cable. If this is not enough, the instrument's output voltage must be transformed into current before sending it through the cable. The current loop is highly immune to the electrical noise present in the surroundings, such as that generated by large electrical motors deployed in industrial buildings.

A very widespread way of transmitting analog data as an electric current is the 4-20 mA standard, often used in the past. It establishes that when an instrument sends a current of 4 mA it indicates that there is no signal, and a current of 20 mA indicates that the signal is at its full range. For example, a level sensor with a range between 0 and 10 m will send 4 mA to mean that the level is null, 20 mA for a 10 m level and 12 mA for 5 m:

$$4\ \text{mA} + \left(\frac{16\ \text{mA}}{10\ \text{m}}\right) \times 5\ \text{m} = 12\ \text{mA}$$

In general, a 250 Ω resistor is placed in series with the current loop in the receiving unit to convert the standard current to voltage from 1 to 5 V. This voltage is finally converted to digital format through an analog to digital converter at the receiver, thus facilitating control decisions or storage.

Most instruments used in environmental sciences have digital outputs, and some of them also offer additionally the 4-20 mA option as an analog output. The question arises quite logically: why use analog outputs when digital outputs are less sensible to noise? There is a case in which this facility could be advantageous over digital outputs.

When it is needed to synchronize the recording of two fast changing time series, sample to sample, and they are measured by two different digital instruments the existence of analog outputs in both instruments can be useful. In general two instruments with digital outputs acquire their analog inputs independently; samples of two instruments are thus not synchronized at a sample to sample level. For example, when studying rapid changes in pressure and velocity in a hydraulic model, it could be of interest that each pressure sample be taken at the same instant as the velocity sample. In general, it is impossible to collect samples taken exactly at the same instant using different instruments because they control their respective data acquisition and conversion processes with two different internal clocks.

The analog outputs of the instruments offer an opportunity of synchronization. For that to happen, analog outputs from different instruments have to be connected to the inputs of an independent external data acquisition system with a unique clock, thus samples from all the instruments can be simultaneously acquired and stored in the external data logger memory (Fig. 9.2).

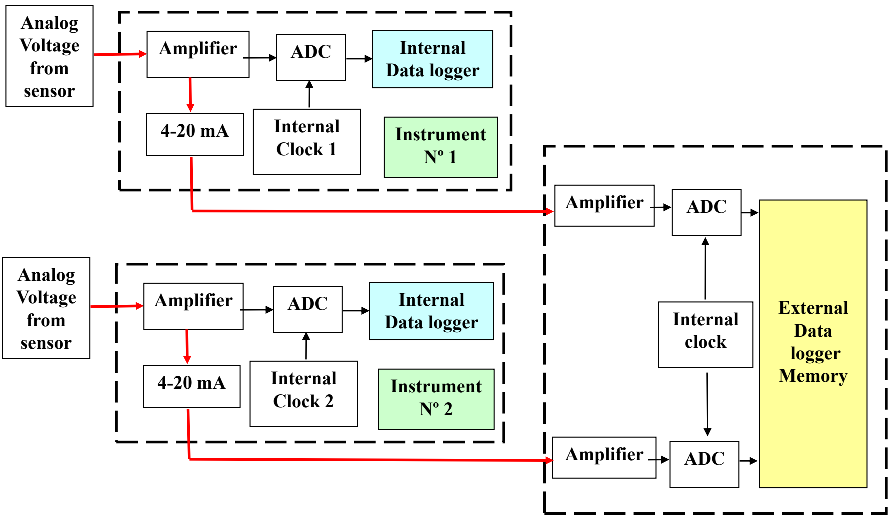

Fig. 9.2: Two instruments with their own internal clock. The data stored in each internal data logger are thus not synchronized. The analog outputs are acquired by an external data logger with a unique internal clock which permits the samples to be synchronized.

9.3.2 Digital Data Transmission

Digital data transmission is the most common way of transmitting data nowadays because digital signals are less prone to noise, and easily amplified and retransmitted. Digital information is coded as 0's and 1's; these bits form strings of bytes, an octet of bits being a byte (Section (3.6.6)). In a given string of bytes, the least significant one is the byte that has the least potential value.

Symbols 0 and 1 must be transmitted as two distinct states over a physical communication medium; for example: two current levels on a wire, the presence or absence of light or sound, etc. The physical medium is also called a **communication line** or a **transmission channel,** which is the pathway for the information and can be materialized for example by a wire, a radiofrequency link, an acoustic link, an optical link, etc.

There is a device that transforms binary data into a physical signal, and vice versa; it is called **Data Communication Equipment** (DCE). Figure 9.3 shows how the DCE interfaces with the physical medium and the digital data at both ends of the communication link. The device that produces the digital data is usually called **Data Terminal Equipment** (DTE). Both devices may be embedded in the measuring instrument. For example, at one end of the physical medium we could think of a field instrument, while at the other end, of a PC.

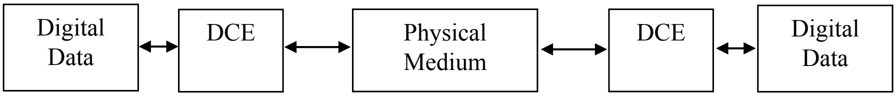

Fig. 9.3: Digital data is converted into a physical signal by the Digital Communication Equipment and vice versa. The physical medium conveys the signal between both DCE.

In order to evaluate the transmission media needed to transport certain information it is important to quantify it on terms of digital units. Storage and transmission of data into computers are done in groups of bits. The most known group being the **byte** (symbol B) which is a word of 8 bits, already defined in Section (3.6.6).

The hardware architecture of old microprocessors used to process 8 bits (one byte) at a time; mathematical and logical operations, as well as memory read and write operations, were based on steps of one byte. More recently, microprocessors extended their internal and external architectures to manage 16, 32 and 64 data bits simultaneously (managing words of 2, 4 or 8 bytes). These processors' capabilities make them more capable of dealing with precision calculations.

In order to optimize data handling by digital systems that are based on microprocessors, each datum should have the amount of bits managed by the digital processing device. For example, handling data with 9 bit resolution in an 8 bit microprocessor requires using two data bytes (16 bits), because the information is handled in modules of 8 bits, thus 7 bits with no information were stored or transmitted, which implies a misuse of resources.

Besides the length of the digital data word, other equally important concepts in data transmission are the **bandwidth of a transmission channel** (B) and the **capacity of a digital channel**. As in the case of the bandwidth of a sensor (Section (2.4)), **the bandwidth of a transmission channel** (B) is the frequency range over which the signal passes through the system suffering low attenuation. For example, a telephone line has approximately a bandwidth from 300 to 3400 Hz for 3 dB attenuation (which corresponds to a signal loss of 50%). This means that the central frequencies of the voice spectrum will pass without attenuation, but low and high frequencies will result somewhat attenuated. As shown in Section (2.4.4) the bandwidth limits the quality of the signal passing through the system.

The **capacity of a digital channel** is the amount of information per second that can be transmitted. In general, for binary symbols, it is the amount of bits per second that the channel can carry, which is also known as **baud rate**. As previously stated, the channel is the media through which the information travels; thus a noisy channel can corrupt the traveling information. Also, as already mentioned, a reduced bandwidth may disturb the quality of the signal. Therefore, the capacity of a digital channel is directly proportional to the channel bandwidth and to the signal-to-noise ratio. That is, the greater the signal and bandwidth, and the lower the noise, the grater the channel capacity. In other words, channels with large bandwidths and low noise are able to transport more information by second, i.e. the channel has a greater capacity.

Part of the energy injected by the transmitter is lost in the medium and then the signal results attenuated as it advances in its way; **attenuation** is proportional to the length of the transmission channel. In addition, the channel will result disturbed by noise; thus the signal degrades in its path through the transmission channel and some information could be lost. Perturbations induced by noise are called **interferences**.

9.3.3 Signal Encoding

In early times, data transmission was analog; signals were varying voltage or current, or modulated radio frequency carriers. Analog signals have the disadvantage that they are prone to noise present in the communication channel. An example will be developed to show the convenience of digital signals.

Let us assume a laboratory water tank in which surface waves with an elevation h varying from -50 to 50 cm are represented by an analog voltage signal that takes continuous values between -5 and 5 V. Suppose that the signal is transmitted on a pair of wires and due to electromagnetic interference a voltage noise of ±0.5 V is superimposed on the signal. When the signal plus the noise arrive at the receiver there will be an uncertainty of ±5% in the wave height due to the noise.

When the analog signal is converted into a digital one before transmitted, it is less disturbed by noise. Using the same example, and keeping the same voltage range, let us assume that the digital signal takes values of -5 V for a digit “0” and 5 V for a digit “1”; suppose also that the signal is sent by the same previously mentioned channel. Due to the channel noise, the digits “0’s” will be represented by voltages varying between -5.5 and - 4.5 V at the receiver end, whereas the digits “1’s” will be represented by voltages between 4.5 and 5.5 V. Thus, at the receiver we could state that any voltage less than -4.5 V is a digital “0” and that any voltage greater than 4.5 V is a digital “1”. Following these criteria, the digital signal is easily rescued from the noise.

This ability to easily distinguish digital signal from noise, makes the transmission of digital signal attractive. In the example we have associated digital values with voltages, but they can be associated with light or any other physical magnitude. In order to transport the digital signal on a real communication channel, the 0’s and 1’s have to be transformed into a physical magnitude representing them, and this is done by means of the above-mentioned DCE.

Different encoding systems were developed with the aim of better adapting the digital signal to the physical medium. At the beginning of digital communications, the physical medium for transmission was copper wires. When electrical cables are used, one of the goals of encoding is that the average value of the physical signal be null, because it renders the transmission more efficient. For this to happen, the value of the physical magnitude (*S*) for each symbol (0's and 1's) should be of opposite sign (negative and positive). Hopefully, if the average the number of 0's and 1's are the same the average of the physical magnitude will be zero. In the case of electrical cables *S* is a voltage or a current.

One type of signal encoding method utilizes symbols with only two levels (-*S* and +*S*), while there is another that uses three levels (-*S*, 0 and +*S*). The simplest encoding using two levels is known as the NRZ (No Return to Zero) whereas the three-level code is known as the bipolar encoding. In the NRZ encoded 1's are represented by +*S* and 0's with -*S*. In the bipolar encoding, 0's are represented by 0 level, and 1's are alternatively represented by +*S* and -*S*, as depicted in Figure 9.4. The top drawing in Figure 9.4a shows the digital symbols at the input of the DCE while the following drawings represent the DCE output encoded as NRZ (Fig. 9.4b) and bipolar (Fig. 9.4c).

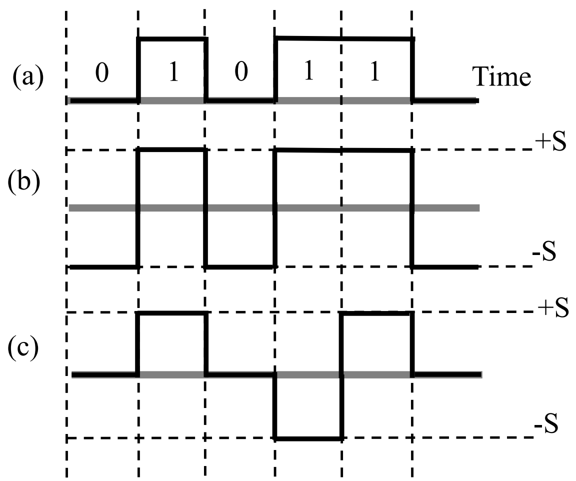

Fig. 9.4: (a) Digital symbols at the DCE input are converted into output levels on the physical medium; (b) NRZ code; (c) bipolar code.

9.3.4 Transmission Modes

Transmission modes describe the way in which the data flow between two points. The DCE that sends the data is defined as a **transmitter**, and that which receives the data is called **receiver**. Data flow can be managed either from the transmitter or the receiver, as will be seen below.

It is our intention to keep the description that follows as simple as possible, and for this reason some sophistication introduced by today's intelligent transmission systems are avoided. Problems inherent to each transmission mode will be mentioned below. However, it should be recognized that modern communication systems have some capabilities that help to decrease the impact of these problems.

Three possible cases are represented in Figure 9.5. In order to fix concepts assume that instruments collecting field data are reporting them to a central station. The Digital Data on the left of the figure is produced by the instrument and sent to the DCE. The right side represents the central station, composed by a DCE and the Digital Data, the latter being materialized by a PC data input.

In this example, the top drawing corresponds to the **simplex** mode in which the information travels in only one direction, from the instrument to the central station in our example (Fig. 9.5a). Thus, the remote station defines when to transmit. The transmission decision could come simply from a clock that periodically activates the instrument and the transmitter (time activated), or in the case of an intelligent instrument, when some event is detected, for example when the measured parameter surpasses certain predefined level (event activated).

The simplex mode requires only one transmitter and one receiver, but if a datum is lost in the transmission process, it cannot be recovered. There is no way to ask the instrument to resend the information lost. Data may be lost due to noise in the communication channel, as could be due to another transmission in the same channel, or to an electric impulse as produced by lightning, electric motor, etc.

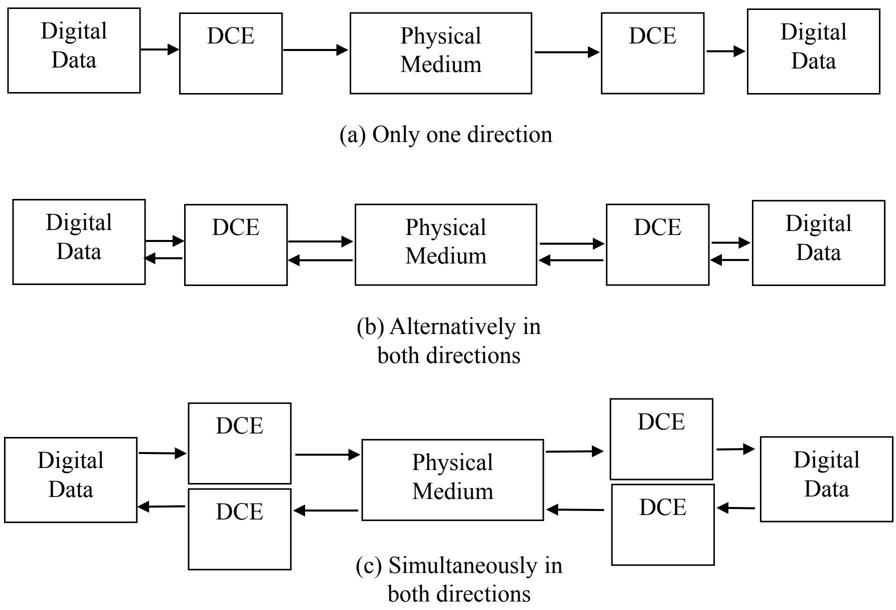

Fig. 9.5: Transmission modes: (a) Simplex; (b) Half-duplex; (c) Full-duplex.

Because each remote station decides the instant in which the transmission is done, if a network of several instruments reporting to a unique central station is implemented using the simplex mode, two remote stations could transmit simultaneously. As the receiver will not be able to manage two concurrent transmissions in the same channel, a collision of the information could appear and the receiver would loose one or both of the two data sets.

In a **half-duplex** connection (Fig. 9.5b) data flows in both directions, but not at the same time. There are one transmitter and one receiver at each end of the communication channel, but there is only one channel available (e.g. in radio frequency both transmitters and receivers share the same frequency).

Following the same example of a remote instrument, this kind of connection makes it possible for the central station to ask the remotes when they have to transmit the data. This kind of controlling action from the central station is usually called **pooling**. The central station sends a message to each remote station, one at a time, asking whether the remote stations want to use the communication line. In this way conflicts between stations transmitting simultaneously can be avoided.

Generally, by means of error detection codes, the central station has the capability of detecting whether some data has been missed in the transmitting process. Therefore, the half-duplex capability allows the central station asking the remote station for a repetition of the missed data.

When a remote station is asked by the central station to transmit a large amount of data, the communication channel will be occupied by the remote station and is useless until the transmission ends; the channel cannot be simultaneously used by the other stations. Thus the central station would loose the control of the communication channel during the transmission from one of the remote stations.

In a **full-duplex** connection (Fig. 9.5c) data can flow in both directions simultaneously because two logical channels are used. In this context, logical channels means that they can be theoretically conceived as two channels, but in practice these channels can be physically implemented, for example, just on a pair of wires (using special modulating techniques). In radio frequency, two frequencies can be employed; the central station transmits in the frequency used by all the remote stations' receivers; and the remote transmitters work at the same frequency that the central station receiver. Each end of the line can thus transmit and receive at the same time.

A full-duplex connection allows sending orders to the remote stations while they are transmitting and can order the remote to stop its transmission. Then the central station always has control over the communication channel and could interrupt the data transmission to attend other more urgent tasks, for example, to interrogate another station which has urgent information to transmit.

9.3.5 Serial Transmission

So far we have described how transmission systems work and how channels are managed. In the following paragraphs it will be analyzed how the information is transmitted inside the channel. A transmission channel as shown in Figures 9.3 and 9.5 sends only a bit at a time; this kind of connection is called serial transmission. A communication standard for wire serial connection, very used in the recent past and still in use in some industrial computers and scientific instruments, is the RS-232 standard. It states the physical characteristics that the hardware should accomplish. For example, it establishes voltage levels, amount of wires, maximum length of cables, connectors' sizes and wiring, pin assignment and signals, data flow control, etc.

RS-232 can send information over a maximum length of cable at a given speed; as a general rule, when the length increases the speed decreases. The maximum length of cable for RS-232 is quite limited; then, when long cables are needed, the standards RS-422 or RS-485 have to be used. An instrument with a serial port could solve communication needs with only three wires (transmit, receive and ground) plus the necessary wires for power supply; in this way handy and low cost cables result.

At present, computers use a serial port called USB (Universal Serial Bus) standard, which has replaced the RS-232. USB can work at higher speeds than RS-232, but over short distances. There are some converters that allow using the USB port of portable computers with instruments that use RS-232 port. Approximated lengths and speeds for the abovementioned standards are shown in Table 9.1.

Table 9.1: Lengths and speeds for some serial connections

Serial connection	Maximum length (m)	Speed (kbit/s)
RS-232	15	115.2
RS-422 and RS-485	1,200	100
USB	From 2 to 5	1,500-12,000

9.3.6 Parallel Transmission

A parallel connection (Fig. 9.6) can transmit a number of *N* bits simultaneously. That is, there are *N* different channels that permit a bit to be sent by each channel. Internal connections in a computer are parallel channels called buses that work at very high speeds but over short distances. They are made of tracks on a printed circuit board, each track being a channel.

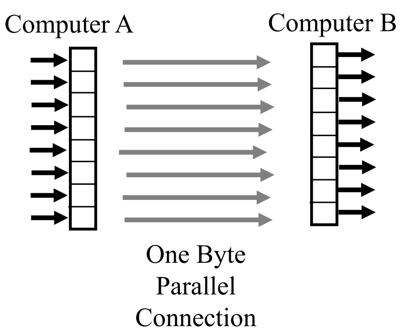

Fig. 9.6: A parallel connection of 8 channels (one byte) allows transmitting 8 bits simultaneously.

In the case of a radio frequency link, each channel needs to use a different frequency. Thus several bits can be sent simultaneously through different frequencies. The same happens with a fiber optic channel.

As is obvious, sending information by means of *N* channels will require less time than sending the same information by means of one channel. Parallel communication is thus faster than serial communication; the higher the number of channels, the more information is sent in the same period of time.

9.3.7 Serial-to-Parallel and Parallel-to-Serial Conversion

In most simple field instruments, their external communications were traditionally solved through serial ports. Perhaps the reasons for this solution are that instruments do not generally require high transmission speeds, computer's serial ports are ease to use, and a few connection wires are needed. If parallel transmissions were used, a considerable amount of wire would be needed. This would result in more heavy and expensive cables and in bigger, and sometimes fragile, connectors.

Modern instruments usually have microprocessors for controlling all the internal activities and to manage the interface with the external world. Instruments handle fast parallel data processing internally, and serial communications externally. When the serial data sent by an instrument arrives at a computer port it has to be transformed again into parallel format, because computers' internal data handling is parallel. Therefore, to perform these transformations some converters from parallel to series and vice-versa are required (Fig. 9.7).

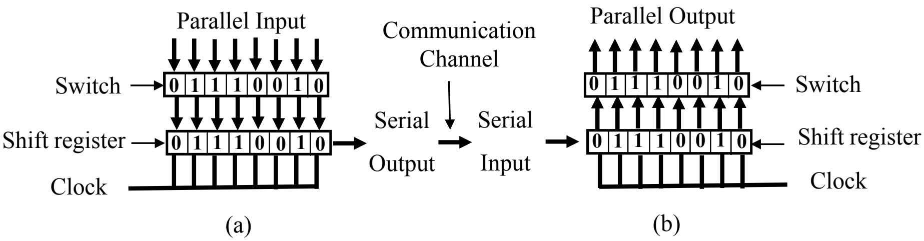

Fig. 9.7: (a) Parallel to serial converter. The parallel input is loaded into the shift register by the switch. The clock shifts the bits to the right, one at a time, into the communication channel. (b) On leaving the channel, the bits enter the serial to parallel converter and the clock accommodates them into the shift register. Once all the bits have arrived they are switched to the parallel output.

9.3.8 Asynchronous Transmission

Usually, when digital information is transported through a channel the speed of transmission (baud rate) is previously agreed between transmitter and receiver. Thus, the receiver knows how much time lasts for each bit. This information is needed by the receiver to recognize the transmitted information, but this is not enough. Suppose that an instrument sends one datum after another, thus it would produce a stream of 0's and 1's on the communication channel, therefore, the receiver needs to know when each datum begins and ends. In order to solve this ambiguity there are basically two transmission strategies: synchronous and asynchronous.

The main characteristic of asynchronous transmission is that each datum is sent in an undefined instant, that is, intervals between data are irregular, as is the case of a person typing on the keyboard of a computer. Each stroke is a datum but the time between strokes is variable. Thus, some extra information has to be sent to the receiver to indicate the start and end of a datum. Asynchronous transmission solves the problem by sending a **start bit** before the datum and one or more **stop bits** at the end. In the case of the keyboard of the computer each datum is called a character; a character is a digital word of 7 bits.

The ASCII Character Set, which stands for the "American Standard Code for Information Interchange", was designed in the 60's, but the character set is still used in modern computers, HTML and Internet. This 7-bit word represents 128 characters. They contain the numbers from 0 to 9, the uppercase and lowercase letters from A to Z, and some special command characters. An example of asynchronous transmission using ASCII characters is presented in Figure 9.8 to clarify ideas. Figure 9.8a shows the start and stop bits.

Assume that when no datum is sent (periods of inactivity, idle) the state of the transmission line is fixed at binary "1" (see right end of Figure 9.8b). Then the start bit will be a binary "0" which warns the receiver that the next bit will be the beginning of a character. Obviously, the receiver knows how many bits compose a character but at the end of the character a binary "1" is sent as a stop bit to reset the state of the transmission line to "1".

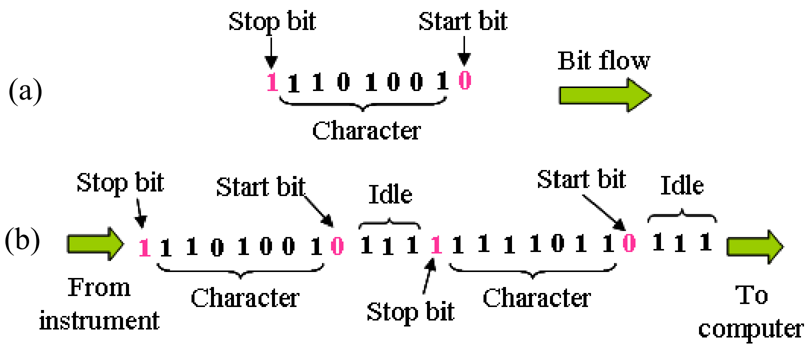

Fig. 9.8: (a) One start bit and one stop bit are added at the beginning and at the end of an ASCII character. (b) Idle periods between characters are at binary "1". The first "0" after an idle period indicates that the next bit is the start of a new character.

9.3.9 Synchronous Transmission

In synchronous transmission, transmitters and receivers are continuously synchronized at the same speed by the same clock. This transmission mode is used when a large amount of data has to be transmitted; data is arranged in blocks called frames or packets. The data frame is headed by special characters called synchronous characters. There are not any gaps between transmitted characters and, in some way, the same timing is supplied at both ends of the channel. Eventually, both ends could go out of synchronism and errors occur. To solve this problem some means are provided to resynchronize clocks and check if the received information has been corrupted.

9.3.10 Error Detection

Special circumstances on transmissions channels can distort the information that travels over them and then the digital word arrived at the receiver would be different to that sent by the transmitter. The information at the receiver could have some bits changed. Thus, it will process or store a wrong data. To reduce this kind of risk some error detection strategies have been developed. All error detection codes are based on redundancy; this means that some extra information has to be added and sent together with the information we want to protect.

The simplest strategy is to add a **parity bit** to the message, which means that an extra bit that has some logical relation with the information we want to protect is added

to the message. For example, if odd parity is desired a parity bit will be added at the end of the message so that the total number of ones in the word to be transmitted (including the parity bit) is odd. For most applications odd parity is used, so that at least one "1" is transmitted.

Figure 9.9 shows an example of a twelve-bit word protected with a parity bit, using odd parity. Because the word has an even number (8) of "ones" an extra "one" has to be added, in order to have a total odd number.

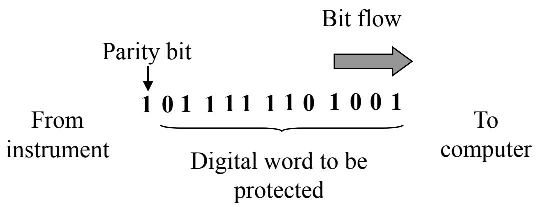

Fig. 9.9: A parity bit "1" is added at the end of the word to be protected to complete an odd number of "ones".

If, due to noise present in the transmission media, an even number of bits is changed in the transmission channel the message will be erroneous but the parity of the message will be unchanged and the error will be undetected. This means that the error detection code is not robust enough. The longer the word to be protected the more the chances that an even number of bits results changed by noise. Furthermore, this method permits an existing error to be known, but nothing is known about which is the erroneous bit.

Another simple method is called the **checksum**; it consists in summing up all the data in a message and, attaching the less significant byte of this sum at the end of the message. Then at the receiver end the same sum is performed and the least significant byte compared with the checksum byte sent.

Let us introduce an example. If a string with 300 bits is to be sent, its sum expressed in binary will have two bytes, namely 00000001-00101100, then, the second of these bytes (00101100, i.e., the least significant byte) is attached to the end of the string. At the receiving end, the checksum byte is separated from the string and the rest of the bits of the string summed. The least significant byte of this sum is compared with the sent checksum. If both agree, there is a great chance that the message be correct.

Also, to detect a corrupted word it could be sent twice. Thus, after a bit to bit comparison of both words at the receiver, a simple algorithm will decide that the word is correct if all bits are identical. This method allows identifying which bit has been changed. In case that a difference appears between the two words, the receiver asks for a retransmission of them. Probabilities that noise corrupts the same bit in both words, keeping words identical are low, but not null.

9.3.11 Error Correction

Previous strategies permit knowing with some degree of probability that a digital word has survived the transmission process unaltered, but do not allow the error to be corrected. In order to have the correct information it is necessary to ask for a retransmission of the information previously identified as corrupted. This requires at least a half-duplex connection.

If a simplex mode connection is used, a simple error correcting code method could consist in sending the same word three times; if the three are identical the chances of being wrong is insignificant; if one of the words is different than the other two, there is a high probability that it is wrong. Both words remaining identical have a great probability of being true and they are selected as the correct word. More efficient methods to correct errors have been developed; the Hamming code (Proakis & Salehi, 2008) is among the most spread correcting codes.

As seen from the previous examples, to increase the probability that the received word be free from error it is necessary to increase the redundancy of the message. In other words, **the price to pay when increasing the certainty of a data is sending more information than the data we really need**.

9.4 Transmission Media

In private networks, users are their owners in contrast with public networks, where users rent a service to the owners (i.e. the owners being telephone or satellite companies). Private radio frequency networks for collecting field measurements from scientific instruments are, in general, of the star or tree topologies.

9.5 Private Networks

It is possible to distinguish two types of private networks: serviced by cables, as in a network of instruments and computers in a laboratory or industry, where most of the time the communication is solved by means of fiber optic and copper cables; and networks where communications are carried by means of radio frequency links (as in field stations).

Private radio frequency networks were frequently used in the past, when satellite communications had a great cost per minute of use, and cell telephony had only services in urban areas. They are still in use in places not covered by the above public services, and also when a big amount of data has to be handled in a medium size area (of a few tens of kilometers in diameter). In these cases the cost of installing the antennas, transmitters and receivers could be less expensive than the cost of a public service.

In some cases where cell telephony and private radio frequency networks could compete, the user should compare both solutions in costs and capability. The advantages of private networks for the users are: (i) they do not pay for the service as in a public network (this could be an important issue if the amount of data to be transported is great); (ii) they do not depend on a carrier company and on the availability or service in the area. The disadvantages are: (i) they have a great initial installation cost; (ii) periodical maintenance service is required; (iii) they need provision of energy (power supply); (iv) in some places it is necessary a license to use the radio frequency channel.

In countries with low population density, digital telephony service in the rural areas does not cover beyond the most important roads and highways; private networks or satellite communications are thus the only alternative solution. There no fixed rule to define when the network has to be public or private. Each case has to be solved on an individual basis, and in general a mixed network using private and public services can result the most efficient solution and has the lowest installation and operative costs.

9.6 Public Networks

9.6.1 Digital Telephony

The evolution of digital telephony was so fast that it would take several pages to explain it, even briefly. Moreover, it would not provide useful knowledge to potential users and any description will be quickly obsolete. There are so many acronyms that only the professional can understand this synthetic language. For these reasons, we will only refer to some features that this technology can provide to transport information from a field instrument to a central station and vice-versa.

In the early times of telephony, voice was transmitted as an analog signal. But because digital signals are less prone to noise and easily regenerated (amplified and retransmitted), in 1962, for special links of commercial telephone systems, voice began to be digitized (Bellamy, 2000). Today almost all telephone services are digital, which facilitates multiplex (placing several transmissions in one physical communication medium) and encryption.

At present, there are basically two telephone networks, one called landline or fixed-line and another called mobile or cellular line. The first uses solid or mechanical media such as copper wires or fiber optic cables as the physical communication medium, while the second uses radio frequency waves propagating in the air. Also, it could be considered that there exists a third telephone network, because many public and private telephone communications are routed through the Internet network.

In fixed telephone networks, the first section from the subscriber (or client) to some point on the network is connected through copper wires, the voice being an analog signal (it is a variable voltage over which the voice is "impressed"). Telephone voice bandwidth is approximately between 0.25 and 3.3 kHz. The copper lines from all the subscribers in a certain area are run to a point in the network called **telephone switch**. It consists of a building where electronic circuits interconnect the lines arriving from the subscribers.

The copper wires used for voice has a transmission bandwidth that much exceeds the required for the voice bandwidth, so the wires have extra capability of transmitting information on them. This additional potential is used to send digital information. The copper lines have a useful capacity as a digital channel (Section (9.3.2)) that can be used for applications such as data transmission.

There are technologies known as ADSL, HDSL, VDSL, etc., that get maximum utility from the copper wires dividing their bandwidth, which enables normal voice telephony **(public switched telephony network – PSTN)** and digital data transfer to be achieved simultaneously. Some telephone companies offer Internet service up to a transfer speed of 54 Mbps, using these technologies.

9.6.2 Global System for Mobile Communications (GSM)

GSM is a cellular mobile communication system adopted worldwide as the standard system for mobile communications (Bellamy, 2000). It is based on radiofrequency transceivers (transmitter and receiver) and the basic unit consists of the central station (or **base station**) and a limited number of remote stations (also called **subscribers** or **mobile units**) working in a restricted geographical area called **cell**. An elementary description of how a cell works to connect two remote stations within the cell will be conceptually described using some concepts explained before.

The example of a **full-duplex** connection shown in Figure 9.5c, refers to a central station (CS) communicated to some remote stations (RS). Let us assume that this network has a star topology (Fig. 9.1b), where the CS is the central node which organizes the network. Two channels (frequencies) are needed to send data simultaneously in both directions between the CS and one RS. With only two channels, the CS can establish a communication with only one remote station at a time. But adding two extra channels to the CS allows it to communicate with another RS simultaneously (a total of four frequencies should be available for this purpose).

If more RS are needed, two new different frequencies are needed for each RS added, because if the same channel (frequency) is used by two transmitters in the same area at the same time, a receiver will take in both transmissions, being unable to distinguish between them. Thus it would not be possible to have a comprehensive message and it is said that there is communication **interference**. This is an unreal oversimplification of the problem, just to give a conceptual idea of how cellular telephony works.

Assume that there exists a set of contacts in the CS (called switch in Figure 9.10), which connects RX1C with TX2C and RX2C with TX1C. Then, both remote stations could communicate between them through the central station. Let us identify the link to send information from remote station 1 to 2, as: TX1, RX1C, TX2C and RX2; and to send information in the opposite direction as: TX2, RX2C, TX1C and RX1.

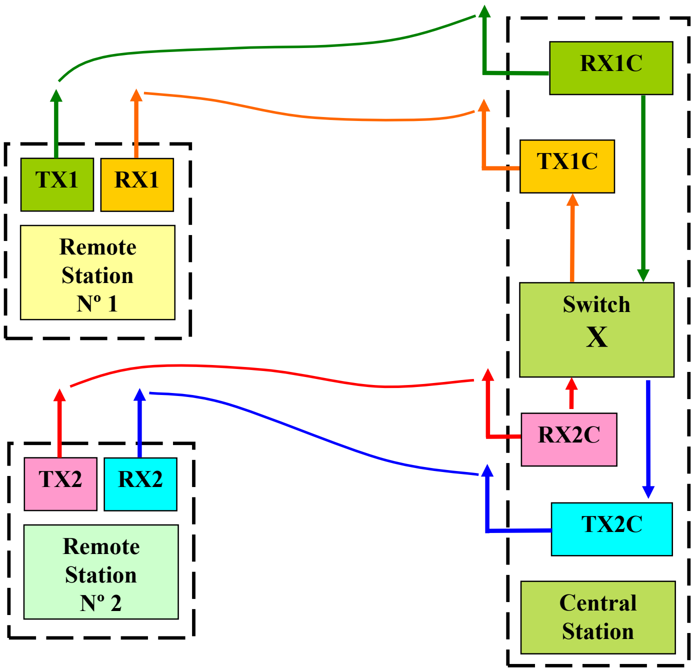

Fig. 9.10: Remote stations 1 and 2 have transmitters TX1 and TX2 and receivers RX1 and RX2. The CS has also transmitters and receivers at the same frequencies as the remote stations in such a way that, for example, TX1 at the RS1 sends data to RX1C at the CS, and TX1C sends data from the CS to RX1 at the RS1. TXs and RXs linked by a solid line indicate that they work in the same frequency. CS switch permits the dialog between RS1 and RS2 to be established.

Now let's think that the RS are mobile phone sets and the data sent is voice. Then two persons could talk through the CS while they are walking in the area covered by the CS radio link.

Each new person (subscriber) added to the network will require two new frequencies, one to send information and another to receive it. Then, because the available bandwidth to allocate the radiofrequency channels is physically limited, the system would have a maximum amount of subscribers in the given area known as **cell**. For the purpose of implementing a realistic system, it is needed to divide the service area in cells reutilizing frequencies used in one cell, into other cells at a certain distance from the first. The size of the cells and the power transmitted by the base station are in accordance with the population density. If the population density increases in a cell, it could be divided into several smaller cells.

A group of cells covering a big geographical area are shown in Figure 9.11. In general, cells have the shape of a hexagon and the serviced area is covered by adjacent hexagons; hexagons having the same letter indicate the same group of frequencies used in it (in the web version of the book hexagons have same colors). The same groups of frequencies are used in areas separated a certain distance to avoid interferences.

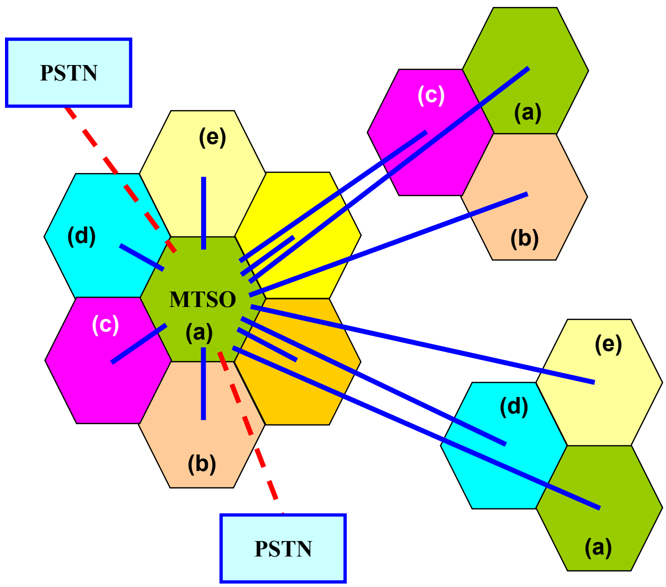

Fig. 9.11: Letters inside the hexagons [(a) to (e)] indicate the group of frequencies used in each cell. Same group should be separated a certain distance to avoid interference. MTSO manages all the functions of the network. It allows communication to be established between mobiles and with the fixed telephone network through the links to the PSTN.

Each cell is connected to the Mobile Telephone Switching Office (MTSO) (solid lines - blue). Cells can be connected to the MTSO by radio frequency, fiber optic or special copper cables (Tomasi, 2001). In this way, the interconnections of cells allow communication between subscribers placed in different cells. Also, the (dashed - red) lines connect the MTSO to the Public Switched Telephone Network (PSTN) to communicate mobile subscribers to subscribers belonging to the fixed or wired network.

The function of the MTSO is to manage and control the calls, to allocate radio frequency channels and to supervise the passage of mobiles from one cell to another. This last function is based in the measurement of the power received by the base stations; when the mobile travels between cells, the power decreases in one base station and increases in another; the MTSO decides which base station will handle the communication.

The mobile station (MS) consists of two components: the terminal, which is a radiofrequency transmitter and receiver, and the Subscriber Identity Module (SIM), which permits the user to access the contracted services at any terminal. The SIM is identified through the International Mobile Subscriber Identity (IMSI) and can be protected by a password to avoid unauthorized use.

GSM has a messaging bidirectional service called Short Message Service (SMS) this is an interesting possibility to send data from field instruments if the amount of data is low and delays are accepted.

9.6.3 Summary on Digital Telephony

Potential users of these technologies, who have scientific applications in mind to be implemented in densely populated areas with a highly developed telephone network, will not face difficulties to find a data transportation service provided by a telephone company. They will have to care only about the way to send the information at the lowest price and to deal with the company taking into account the technical requirements and the service-level agreement offered by it. These users will have to understand how their data are produced and processed to select the most adequate service. They will have to evaluate how much data they will need to transport and the delay they can accept in the transference of the information. The selection of the service to be contracted has to be done with the help of the service provider company because there are services offered by some companies that are not offered by others. Also, services may be charged by time or by the amount of information transported.

9.6.4 Satellites

From the point of view of space aeronautics, an artificial satellite is an object launched by humans into space, describing orbits around the Earth or other celestial objects. Obviously, satellites that will interest us are those that orbit around the Earth. The orbits of the Earth artificial satellites can be circular or elliptical. In the first case, the speed of the satellite is constant; in the second one, the satellite has higher speed when it is close to the Earth and lesser when it is far (Kepler's second law (Feynman et al., 1967)).

Satellites can be passive or active, the first can be used to reflect (or scatter) electromagnetic waves on their surfaces; some tests were performed using the moon as a reflecting satellite. In this way, two points on Earth can communicate impinging a wave on a passive satellite, but to have a detectable signal at the receiver, the impinging wave needs to be of high power.

Active satellites have receivers and transmitters and they work, not merely reflecting, but retransmitting the signal that they receive from the Earth stations. Also, because they have directional antennas, the energy needed by the transmitters and receivers decreases considerably. Commercial satellites are of the active type.

Basically, as regards the orbit, there are two types of satellites: asynchronous and geosynchronous. Asynchronous satellites continuously change their position relative to a fixed point on the Earth, so they can be used only for a short time to transmit and receive data in each pass over that point. For example low-Earth-orbit (LEO) or low-altitude satellites are at around 160 km from the surface of the Earth and it takes them about an hour and a half to turn a full orbit. Then they are visible (available to communicate with a certain ground station) only for 15 minutes in each orbit (Tomasi, 2001). Other more remote satellites employ about 12 hours to complete an orbit and are available for 4 hours at a point.

All the orbits are in a plane passing through the center of gravity of the Earth. If the plane passes through the poles, the orbit is a polar one, if the plane lies on the equatorial plane, the orbit is an equatorial one, and the remaining cases are known as inclined orbits. Equatorial orbiting satellites travel along a circular orbit, and because the distance is about 35,800 km from the surface of the Earth, it takes them 24 hours to turn one orbit; therefore, they stay fixed relative to a point on the Earth surface. Because of this they are called geosynchronous (Tomasi, 2001).

The advantage of geosynchronous satellites is that they remain substantially stationary with respect to a ground station, so they can provide permanent communication between two points that are in the beamwidth of the satellite antenna. The disadvantages are that, by being so far away from the surface of the Earth, they need to use powerful communication equipment. There is a communication delay of about 0.25 s; and propulsion systems are needed to keep satellites in orbit. Moreover, being away from the poles, they are not effective for communication in these areas.

They are ineffective at high latitudes because the Earth's curvature disrupts the transmission path.

The area on the surface of the Earth with which a satellite can communicate is called footprint; it depends on its orbit, communication frequency and antenna. There are maps that allow knowing which zones on the Earth can communicate with the satellite and when.

In the case of a set of instruments for use in environmental sciences that are distributed in a very large area, satellite communication is an interesting, and in some cases, the unique alternative. In environmental studies that include geographic places accessible only for a certain period of the year, or sea places far from the coast (places where there is no cell phone coverage), satellite communication is the only option. As explained in the introduction, the type of satellite to be used depends on the urgency with which the data is needed. In the majority of cases, instantaneous data is not required and asynchronous satellites can be used.

Figure 9.12 depicts an example of a set of instruments distributed in different geographical areas communicating their data by means of an asynchronous satellite. The satellite's receivers and transmitters antennas are focused on a specific area of the Earth's surface from which the satellite can receive or transmit. While the satellite moves along its orbit and the Earth rotates, the antennas' beams sweep a footprint on the Earth's surface.

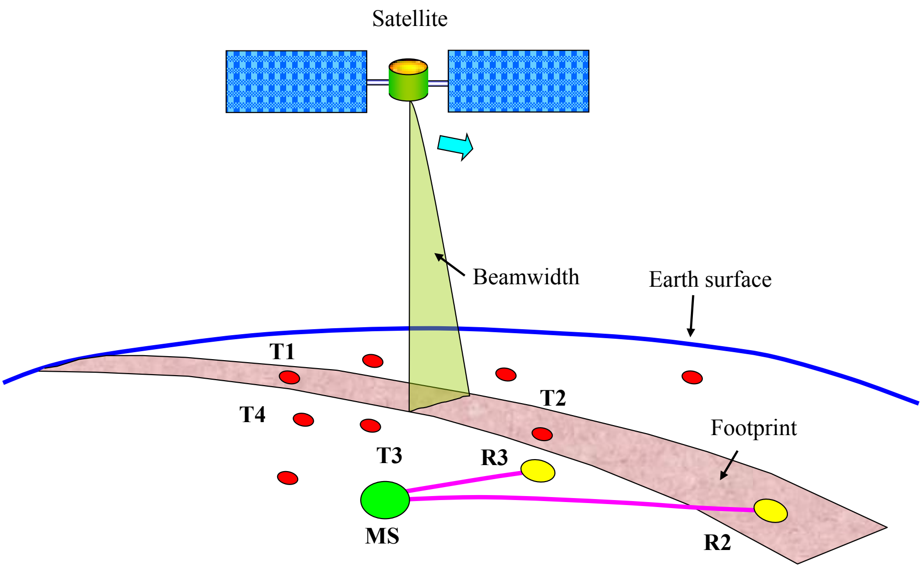

Fig. 9.12: Smallest dots (red) indicate field instruments; R2 and R3 (yellow ones) are Earth stations, and the MS (green one) is the master station. Those instruments on the footprint of the satellite are able to send and receive messages through the satellite when it is passing by. In this example it is assumed that in the next pass the footprint moves to the left passing over T3, T4 and R3.

In Figure 9.12, we assume that the satellite's orbit combined with the Earth rotation will move the footprint from right to left. The instruments from which we want to transmit the data are represented by the smallest dots (red). When the satellite passes over the instrument T1, the instrument has some time available to send data that are stored in the satellite memory. Next, instrument T2 will also send its data. If data are not urgently needed, and the satellite has enough memory to collect the data from all instruments, the information could be downloaded the next time that the satellite will pass over the master station MS.

For a more frequent retransmission of data from the satellite to Earth, there may be other Earth receiving stations R2 and R3 (yellow) such that the data from T1 and T2 are downloaded to the ground station R2. This station could be linked by a public data service (e.g. Internet or telephony) to the master station MS, gathering in this way the information more quickly than if it had to wait the next satellite pass over MS.

In the case of using geosynchronous satellites, the problem is simplified because the communication can be achieved permanently. Although this kind of satellite covers an area on the Earth surface that is well determined, data between satellites can be transferred to cover larger regions. The main problem of these satellites is that they do not cover the areas close to the poles satisfactorily.

9.6.4.1 Some Satellite Constellations

The provision of satellite communication services is such a dynamically changing business that to give names of satellite constellations could make this material become quickly obsolete. Thus, satellite constellations characteristics and services will be described avoiding any commercial names; they will be named as Constellations A, B and C. Because some satellites stop working and new satellites are launched into space, fleets are changing continuously, and thus the amount of satellites composing fleets is offered only with the purpose of giving a rough idea on its size. Communication services also change frequently because new satellites usually have novel technologies that allow offering innovative services.

9.6.4.2 Constellation A

The International Maritime Satellite Organization was originally an international non-profit organization created in 1979 to provide a satellite communication network to enhance safety at sea. Later on it expanded their services to aircraft and land users; in 1999 the organization was converted into a private company (http://www.inmarsat.com).

This constellation is composed of 11 satellites flying in geosynchronous orbit 35,786 km above the Earth. Through this fleet of satellites the constellation provides voice and high speed data services on land, on the sea and in the air.

The constellation has a product that provides access to services such as voice, email and Internet where land networks do not exist. It is based on IP technology and delivers data rates of up to 492 kbps. The service is accessed by means of portable satellite terminals of the size of a laptop.

9.6.4.3 Constellation B

Satellites in this constellation have polar orbits with an average altitude of 780 km; they are distributed in six polar orbital planes, each one having 11 equally spaced satellites. The constellation has 66 active satellites and some spare ones. There is always a satellite visible to all subscribers. Thus, permanent data transmission is possible, unlike what happens with the example shown in Figure 9.12.

This constellation provides a personal communication network for voice and data all over the world. This network provides communication service in both directions between subscribers of this constellation and with those of the public network. Each subscriber has a personal number; service costs are independent of where the communicating points are placed (Tomasi, 2001).

The orbital period of these satellites is 100 minutes and 28 s. Each satellite communicates not only with the Earth's stations but also with satellites in front, behind and in adjacent orbits. This system is of particular interest for the maritime community; ships traveling through any sea may communicate with the coastal stations. The communications follow three steps: the ship communicates with the closest satellite; the message arrives at the satellite nearby the final recipient by means of some bridges between satellites, and this satellite call the recipient to establish the link between both ends. This constellation is becoming popular for transmission of weather and oceanographic data from instruments mounted on marine buoys and ships. Because they use a polar low Earth orbit they can provide truly global satellite communications, including in Polar Regions.

There are manufacturers of instruments which provide modems to connect field instruments to a central station through this constellation (http://www.hko.gov.hk) (http://www.iridium.com). The equipment consists of a transceiver which operates like a standard modem. It has a data input for a RS232 port and they work with 12 VDC; when transmitting they take about 0.6 A from the power supply. In general, a solar panel of about 50 W and 12 VDC is used to keep the battery charged. Obviously, the solar panel has to be selected according with the solar energy available in the place where the remote instrument is located.

9.6.4.4 Constellation C

Another constellation of communication satellites uses about 30 low-Earth-orbit satellites of reduced size and weight (http://www.orbcomm.com). They transmit at a much lower frequency than other commercial satellites (in the 137-150 MHz VHF band), which requires a larger antenna but needs less power to work. Although a satellite may be available for communication during a certain time interval over a given area, subscribers in the area are allowed to use the satellite only 1% of the time the satellite takes to cross the entire area to avoid interferences (about 36 s per hour). Due to this characteristic, this constellation is best suited for small amounts of data such as the case of Global Positioning System reports of vehicles, animal migration or drifting buoys.

There is a delay between the delivery and the reception of messages that for most messages containing less than six bytes may range from 1 to 15 minutes. This delay is usually called latency. Typical data transmission capacity is adequate for sending GPS position data or simple sensor readings. The communication network charges fees are based on the amount of data transmitted. Each field station has an identification number that permits the satellites to be communicated with it.

There are providers of equipment to communicate through this constellation who offer devices for directly connecting instruments to the network (http://www.globalw.com). They consist of modems with RS-232 inputs, power supply and antenna. Some of them include a GPS built in that is used to report ships and trucks positions. Other devices have several analog inputs for instruments with 4-20 mA or voltage outputs. Also, some modems can be used in the reverse sense sending information from the central station to the field station. They have digital outputs which allow activating actuators in the field station, for example, to turn on and off a motor or heater.

In some cases Earth stations receiving satellite communications are connected to Internet storing the received messages in a dedicated web page.

9.6.4.5 Inter Constellation Communications

As Figure 9.12 suggests, low-Earth-orbit (LEO) satellites have a data latency of some minutes to download the information generated by the instruments onboard the satellites because they need to communicate with a ground station network. Until a few years ago, reducing latency meant to increase the number of ground stations, which increased costs. An idea that changed the scene was to transmit the information, not to ground stations but to space stations (McCormick et al., 2010; Lenz et al., 2010).

There is a constellation consisting of three modern geosynchronous satellites spaced at 120° which provides ground, maritime and aeronautical communication services. A terminal type as those used on ground was evolved and adapted to be used onboard a LEO satellite. Thus, this recently developed transceiver allows a LEO satellite to communicate with the ground station in real-time via the existing constellation of geosynchronous satellites.

This new technology, which allows communicating LEO satellites with the space-based platforms, permitted data latencies to be reduced from tens of minutes to milliseconds. It opened new possibilities for the collection and application of data captured by sensors in space.

9.7 GPS (Global Positioning System)

The GPS positioning system has among its objectives the positioning of military platforms in real time with high accuracy and of civilian platforms with moderate accuracy. It provides global coverage 24 hours a day under any climatic conditions and allows synchronizing clocks with accuracy better than 500 ns. It consists of three segments, the space segment, the control segment and the user segment.

9.7.1 The Space Segment

The space segment consists of 24 satellites arranged in six orbital planes (four satellites in each plane); planes are offset approximately 60°. Satellites are in circular inclined orbits (they are not geosynchronous). They are at 20,200 km from the Earth surface and it takes them approximately 12 hours to complete a revolution around the Earth. Satellites are ordered in the orbits so that at any point on the Earth's surface any GPS receiver can always view five to six satellites.

Satellites continuously transmit reference signals. All satellites transmit on the same frequency, but each has a different code; all receivers know each satellite's code. Then they can distinguish which satellite they are receiving. Receivers can determine their positions from the information sent by the satellites. The observation of at least three satellites is needed to calculate the place of the surface on which the observer is.

The user's positioning accuracy depends on the precise knowledge of the position of the satellites' orbits and the measurement of the distance between the satellite and the receiver. To do so, each satellite sends a data packet containing, among other things, the following information: the time at which the packet was sent, measured with an atomic clock on board the satellite; the position of the satellite at the instant of transmission (called ephemeris); summary information on other satellites (called almanac); and a pseudo random code used only by this satellite. A pseudo random code is a digital word (ones and zeros) whose sequence is not repeated for a certain period, so it looks like a random sequence of ones and zeros, but in reality it is a periodic signal (Fig. 9.13). The ephemeris is updated every two hours and is valid for four hours, the almanac is updated every 24 h.

The receiver has stored in its memory the pseudo random sequences of all satellites and seeks what sequence coincides with that of the satellite being received. In the next step the receiver determines the time difference (Δt) from the moment when the sequence was sent from the satellite until the instant it was received. This difference is the time it took for the satellite signal (electromagnetic wave) to travel the distance that separates the satellite from the receiver. As the speed of wave propagation can be assumed to be approximately constant ($c = 3 \times 10^8$ m/s), then the distance (d) between the satellite and the user can be readily known, which is called pseudo range.

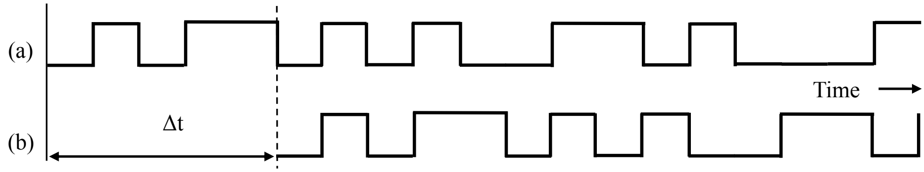

Fig. 9.13: Pseudo random sequences. (a) Sent by the satellite. (b) Receiver generated. Δt is the time needed to travel from satellite to receiver.

$$d = c\,\Delta t \tag{9.1}$$

The solution of Eq. (9.1) allows the receiver to know only the distance *d* from the satellite, but there are infinite points on a sphere of radius *d* that meet that condition. To find the three spatial coordinates of the receiver it would be needed to resolve at least three independent equations, each one corresponding to a different satellite. Because the receiver's own clock is not very accurate, one additional equation is required to set the receiver's clock to the satellite clock.

If (x_i, y_i, z_i, t_i) are the ephemeris and the atomic clock time sent from the satellite (where the subscript *i* denotes the satellite number), we can write the equation of a spherical surface for each satellite as

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = d^2; \quad d = (\tilde{t} + e - t_i)\,c \tag{9.2}$$

where (x, y, z, t) are the unknowns corresponding to the position, $\tilde{t}$ is the approximate time of the receiver and *e* is the offset between the satellite actual time and the approximate time of the receiver. Then, simultaneously solving four equations, the real time and the position of the receiver are obtained.

9.7.2 The Control Segment

The control segment of the GPS is composed of fixed Earth monitoring stations called Operational Control System (OCS) scattered throughout different parts of the Earth's surface, and a Master Control Station (MCS). Monitoring stations are receivers that track the satellites as they pass over them and collect telemetry data. As the positions of the OCS and MCS are accurately known and they have precise clocks, it is possible to verify if satellites are displaced from their theoretical orbits. The MCS receives the data from the OCS and determines if satellites have undergone changes in their clocks or their orbits. Then the ephemerides are recalculated based on this information and they are reloaded to the satellites a number of times per day.

9.7.3 The User Segment

The user segment consists of the receivers of the signals sent by the satellites. They receive the signals from the satellites and calculate their position and the local time.

The signals sent by the satellites give two qualities of services for the absolute positioning of the receiver. The Precise Positioning Service (PPS), exclusive for military use and the Standard Positioning Service (SPS), which is available to all users. For GPS receivers that meet with a set of particular assumptions, the SPS can assure that 95% of the time, the horizontal error is less than or equal to 17 m and the vertical error less than or equal to 37 m (Department of Defense (USA), 2008). For specific locations and with special receivers these errors are significantly reduced (Hughes Technical Center, 2011). But if the above particular assumptions are not satisfied the performance is somewhat degraded.

An additional use of the messages sent by the satellites is the synchronization of time bases. Since the satellite sends the atomic-clock on-board hour, solving the four equations allows the clock time of the receiver to be corrected by using the satellite time that is very accurate. For the SPS, the time transfer domain accuracy is less than or equal to 40 ns (Department of Defense (USA), 2008). This information is used to synchronize systems such as cellular telephony, in which several stations must work synchronized.

9.8 Data Storage

The storage of information is a process quite familiar to most of the readers. Therefore, only some general concepts and problems that could happen with instruments' memories will be mentioned.

9.8.1 Storage Systems

In general, present systems that store information are based on electromechanical magnetic disks like hard disk drives (HDD) used in PC's; on optical-mechanical devices as rewritable compact discs (CD-RW) or DVD, which have a mechanical component to rotate the discs and an optical device to "read / write" it; and on solid-state memory, which has no mobile parts. Rotating disk devices consume a considerable amount of power and occupy a greater volume than solid-state memory, so most measuring instruments use solid-state memories as storage media.

Solid state memories can be either volatile or non-volatile; the first loose the information stored in it when the power supply is turned off, the second keeps the information even without power (as it is the case of a pen drive). The second type is used in instruments as storage media. The non-volatile memories may be of two types: read only memory (ROM) and random access memory (RAM). The first type, once written, can only be read; the second type may be rewritten.

A solid-state drive (SSD) is a solid-state memory that has input/output interfaces compatible to HHD, which facilitates its use. There are two types of SSD, one is of the volatile type (backed up by a battery) and the other is non-volatile. In general, there is a compromise between speed and persistency of the data; memories with a fast access to the data are volatile, non-volatile memories have longer access times. A memory can be visualized as an array of boxes where each data is stored in a particular box (Fig. 9.13). Data writing in memory is the process of placing data in a box. The process of reading the memory is copying the data found in the box to an external device. Data writing alters the content of the memory, data reading does not. The procedures for placing or retrieving data from boxes require knowing the addresses of the boxes. The same happens with writing and reading from memories. Therefore, before to transfer data from and to the memory, it is necessary to identify the address of the memory cell.

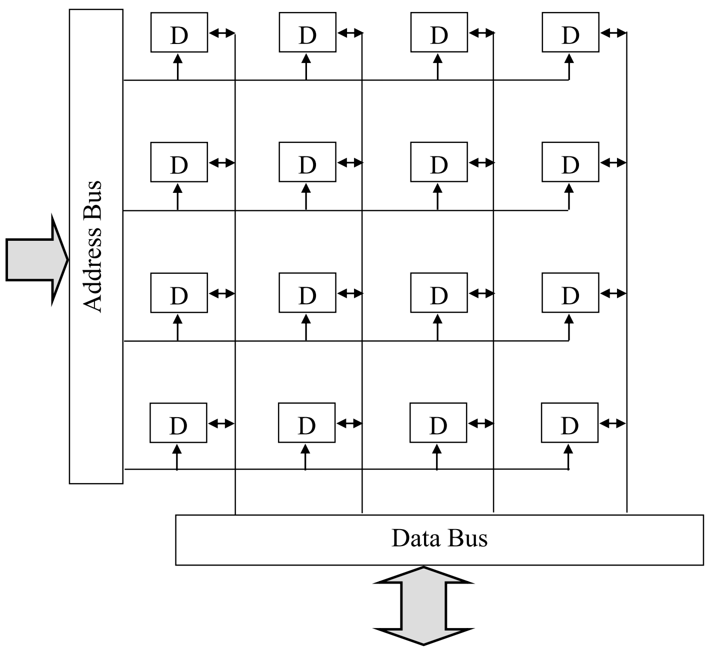

Fig. 9.14: Schematic of a memory

As was explained in Section (9.3.6), internal connections in a computer are parallel buses made of tracks on a printed circuit board, each track carrying a bit. Likewise, solid state memories have internal tracks that form buses with the specific purpose of handling memory information. They are the **address bus** and the **data bus**. The address bus selects a specific cell, and its information is conveyed through the data bus. Put another way, a digital word in the address bus selects which "memory box" will be connected to the data bus. Then if a read process is initiated, information is collected from the bus by other devices, i.e. a register of the processing unit, or if a write process has begun, the information from the bus is introduced into the cell.

An array of 16 cells is depicted in Figure 9.13; each cell has a data (D). The address bus sends the information just in one way, from the bus to the cell; it selects which cell will be connected to the data bus. The data bus is bidirectional, because data can be read (copied from the cell to the data bus) or write (cell data is changed taking the value presented in the data bus).

9.8.2 Comments

Instrument users should be aware that together with the non-volatile memories, where the results of the measurements are stored, instruments might have volatile memories. Because they have short writing and reading access times they are used for some internal applications such as fast signal processing. Sometimes, transference of information from volatile memory to non-volatile is performed in blocks. Then if the battery is disconnected before all the information has been transferred from one memory to the other, some data could be lost. Users should be aware that all instruments' internal processes have to be finished before disconnecting the main battery to avoid data loosing.

In general, an instrument's real-time clock has internal volatile memories and they are supplied by a secondary special long-life battery. Thus, when the main battery is disconnected, the clock keeps working. But when the life of the secondary battery ends and the main battery is disconnected, a reset process of the clock will occurs and the real time is lost. When the reference of time is lost, it is required to set the real time clock before using the instrument again. In general, a manufacturer advises the user about these operational features, but it seems useful to remember them here.

9.9 Tips to Select a Communication Media

These tips could help the user to select a communication media for a field instrument network reporting to a central station. They are based on some questions that the user should answer before selecting a data communication media. Among the parameters to be analyzed are: the period between transmissions, the amount of data to be sent and the accessibility to the remote sites where instruments are installed.

The first question to answer is what real time means for the problem being evaluated. In some control applications, real-time measurements mean to have a set of data inputs for supplying a numerical model at the same speed that the model produces a data output. For example, let us assume that a numerical model is running in a microprocessor on board a model plane to control its trajectory. The numerical model processes the entering information to produce control actions every 0.1 s. In order to obtain meaningful control results, the model should be supplied with a set of data inputs at least ten times per second. If 5 variables, each one with a byte (8 bits) of resolution, are measured onboard the plane, these variables have to be sampled ten times per second to keep the system working in real time; thus, the amount of bits per second resulting from the measurements would be $5 \times 10 \times 8 = 400$ bits/second.

Suppose now that the numerical model is run on a computer placed on land, and that the measured information onboard the plane has to be sent from the plane to the control station and the numerical model results transmitted back to the plane. At least a half-duplex transmission system with a baud rate in excess of 400 bits/s is needed. But if the amount of information to be sent from land to plane is similar to that in the opposite sense, a baud rate about 800 bits/s would be needed or a full-duplex system should be used. This kind of real-time problem is generally solved by means of radiofrequency links.

A real-time system tracking whale migration could be comprised of a GPS receiver attached to a whale and a transmitter to report its position. In this case, perhaps it would be enough to know the whale's position once every 10 minutes. Thus, for this slow speed motion problem, real-time means taking few measurements each hour. Thus, the problem could be solved by transmitting the GPS position through a satellite communication system; a few bits per second are required.

The real time requirement is one of the parameters that most often influence the characteristics of the communication system needed. Then, as already pointed out, it is important to correctly define the time scale of the problem to be solved.

If the maximum delay accepted between the information is sent and received is less than one second, most satellites are not useful as communication media. Perhaps the solution has to be sought in radio frequency links or fiber optic.

The second great question is how much data is really necessary to transmit in real time.

The answer to this question also conditions a lot of the characteristics of the communication system. A field station could have several instruments and all of them store the information on non-volatile memory, and perhaps data from only few of them are needed in real time. Then selecting to send just those really needed could facilitate selection of the communication system, thereby reducing costs.

Another limiting factor is the possibility of reaching the instrument sites at any time, in all seasons and at low cost. If continuous access to the sites is not possible, or if it is expensive to arrive at them, users will be compelled to send the data using any available communication system. Perhaps they should have to modify their measuring expectations, adapting data transmission to the capability of the available communication system.

In selecting a transmission media it should be evaluated whether information has to be sent in both directions. If this is the case, a bidirectional media is required, and simplex mode transmissions are excluded of the selection. This means that a transmitter and a receiver are needed at both ends of the link.

When the information has to be sent in both directions it should also be analyzed whether this has to be done simultaneously. For both ways simultaneous communications, two communication channels are required; in radio frequency, this means that two frequencies are needed. This requirement leaves out some satellite signals that are not bidirectional.

Transmitting large amounts of data precludes the use of some satellite services which are conceived to send small data packets; then, attention should be paid when contracting a satellite service. In these cases, Short Message Service (SMS) of digital telephony cannot be a solution either, because they have been devised for short messages.

For large amounts of data, when possible, a private network should be studied as an alternative, because public networks charge their services by time of use or amount of data transported. Private networks do not have service costs; however they have installation and maintenance costs.

Communication solutions are very different for fixed or mobile instruments. We understand for mobile instrument one which is on board a vehicle or attached to a living moving being. Satellite services are practical solutions for instruments on mobile platforms sending low amounts of data, especially when the platforms are disseminated in a very wide area. In populated areas, perhaps, cellular telephony could be used to solve this kind of communication problem. When mobile platforms are underwater, they require specially designed acoustic networks.

Instruments that simultaneously store and transmit data and can be acceded at any time and low cost, could allow using low quality (generally low cost) communication media, because the data is safe on board the instrument and if the communication fails, the user could recover the data directly acceding to the instrument.

9.10 New Trends in Data Processing and Storage

Since the emergence of Internet the storage and transmission of information have suffered a revolutionary change. This new way of communication gives rise to the availability of a great infrastructure that can be used in measuring systems for the transportation, processing and storing of information in science research activities. An overview on this issue will be addressed below.

9.10.1 Internet

It is a tool well known by all students and it is not necessary to spend time on it. Only a few ideas will be introduces as a first step to the Cloud Computing concept. The Internet is based on a set of protocols which specify how data have to be handled. The most important are known as the Transmission Control Protocol (TCP) and the Internet Protocol (IP). The Internet allows accessing the World Wide Web (WWW) which is a system composed of web pages whose contents are information stored in different formats: images, texts, audio, video, etc. These pages are usually written in the Hypertext Markup Language (HTML) which is a structured language. Such texts contain links to other web pages or other kind of codes inserted inside the texts.

This interlinked system can be navigated by means of web browsers, which are software applications for searching, introducing or retrieving information from the WWW. Each source of information is identified by a Uniform Resource Identifier/Locator (URI/URL) also known as web address. It consists of a string of characters that identifies where to connect, how to connect and what to ask for.

The Internet allows us to access a wide distributed network where we are able to found numerous information resources and services. Individual nodes of the distributed network can share with other nodes their resources such as process capabilities and storage capacity. These abilities give rise to the concept of cloud computing.

9.10.2 Cloud Computing Description

Cloud computing (CC) is a relatively new concept in Information Technology (IT). Since its introduction computing and data storage is moving from desktop and portable PCs to large data centers. It is the latest step of a long way which begun with the concept of parallel computing, followed by distributed computing and grid computing. The main advantage claimed by vendors of cloud services is that customers do not have to pay for infrastructure installation and maintenance (Jadeja & Modi, 2012).

The National Institute of Standards and Technology (NIST) of the United States defines: “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (Jadeja & Modi, 2012).

Cloud computing is also called dynamic computing because it provides computing resources according to the client’s needs. It consists of a virtual group of information resources (networks, storage, central processing unit and memory) to accomplish with the user’s requirement. CC makes it possible parallel batch processing, which allows users to analyze huge amount of data in small periods of time and at low costs. Many computers can work together to perform computing and data processing services that are impossible to do with personal computers (Mollah et al., 2012).

CC refers to the capabilities provided by real-time communication for sharing information and services among computers. For example, by summing up the hardware abilities of several real servers it is possible to simulate a virtual server that does not physically exist but which provide complex information resources or services to the users.

CC is a set of Information Technology (IT) services that meet certain characteristics: multiple clients share a common technology platform and instead of paying fixed costs for a service, they pay according to its consumption, users can scale services according to their needs and, theoretically, there are no limits to growth. This concept of network-based services could be useful in some applications to the environmental science researcher community.

9.10.3 Cloud Computing Types

There are different types of cloud computing according to the kind of users that can access the cloud. A cloud is called a **public cloud** when the network can be accessed by any customer on Internet. The services provided in these clouds can be accessed by any organization (Mollah et al., 2012). In public clouds all the resources and applications are managed by the service provider (Jadeja & Modi, 2012).

When their requirements and policies agree, a group of organizations can merge to construct and share a cloud infrastructure. Such a cloud models are called **community clouds**. They can be used only by members of the community who have common interests. The cloud infrastructure could be hosted by a third-party provider or within one of the organizations in the community (Jadeja & Modi, 2012).

Either public or community clouds are accessed usually via Internet; for this reason security precautions should be implemented. Because they share resources such as infrastructure and services, the network is more efficient. For example, services that are mostly required during day hours (such as email service) can be used in other parts of the world during night hours http://en.wikipedia.org/wiki/Cloud_computing.

Some disadvantages of these clouds are that the permanent availability of the services is not assured, and that the immediate accessibility to the stored information may fail because the reliability of the complete system depends on the reliability of Internet. Furthermore, users run the risk that the information could be disclosed or that the service provider could access the stored information. The service provider could accidentally lose the hosted information as well.

Sometimes cloud computing services have suffered outages and in such cases users are impeded to access their services and data stored; public clouds are more prone to malicious attacks (Jadeja & Modi, 2012).

In contrast, **private clouds** offer IT services to a predefined group of customers; it is a proprietary architecture which provides hosted services to the users within the organization, with access through Internet or private networks. The private cloud can only be used by the members of the specific organization and could have direct connectivity. This kind of cloud is more secure than both previously described because data and service availability depend only on the own organization. Technologically speaking, private clouds are similar to public clouds, and provide similar benefit such as agility and cost savings. Services are pooled together and made available for the users at the organizational level (Jadeja & Modi, 2012). Cloud is protected by firewall barriers that prevent access to hosted services from those outside the cloud (Mollah et al., 2012).

There are also **hybrid clouds** as a combination of a private cloud and one of the two firstly mentioned clouds. This combination could profit from the benefit of both kinds of networks. For example, some critical information could be managed in the private cloud to avoid information disclosure and to have immediate usability, avoiding relying upon Internet connectivity. At the same time some public cloud services could be used for less sensitive information processes. Cloud users should take a look at the works they should place on a controlled private cloud and what other works they can put in the public cloud, where the risk of service interruption and loss of data is higher.

There is also another classification of clouds as **internal** and **external clouds**, the first being a subgroup of the private clouds. They provide services within the same company or corporate group. The second may be public or private and provide services to other companies http://www.networkworld.com.

Cloud computing could decrease initial investments, facilitating future scalability and increasing availability of services. However there are some risks that must be taken into account. There are no safe regulations for storage and backup of the data handled and hosted in the cloud. When outsourcing their own data to a provider, customers are still responsible for the security and integrity of them, even when they are held by a third party provider (Mollah et al., 2012).

Service providers do not assure that interfaces between the services taken from the cloud and those of the company itself will be sustainable over time http://www.networkworld.com. Furthermore, it is only possible to use applications or services that the provider is willing to offer (Mollah et al., 2012).

9.10.4 Cloud Computing Stack

There is a tendency to classify the kind of services that the CC is offering. There are several classifications but we will mention only the simplest and known.

1) Software-as-a-Service (SaaS)
SaaS applications are designed for end-users, it is software which is deployed over Internet, thus eliminating the need to install and run the application on the users system. This is a "pay-as-you-go" model. An advantage of SaaS is that it does not need specific hardware to run the software (Mollah et al., 2012).

2) Platform-as-a-Service (PaaS)
PaaS provide development environment as a service. The provider's equipment can be used by the client to develop its own program and deliver it to the final users through Internet. In other words, PaaS gives the set of tools and services needed to make coding and deploying the generated applications as a service. The advantage of PaaS is that there is no need to buy special hardware and software to develop and deploy enterprise applications (Mollah et al., 2012)

3) Infrastructure-as-a-Service (IaaS)
IaaS is the hardware and software that support all the information system, such as servers, storage space, networks, operating systems, etc. Also, these tools are offered as services.

There is an analogy that could help to understand the stack (www.rackspaceclouduniversity.com). The infrastructure is compared to a road transportation system, the trucks and cars are analog to the platforms, and people and goods could be considered as software and information. Even when each service is shown as a layer of the stack, their differences are not so clear in reality.

9.10.5 Application Example

Researchers working in environmental sciences may need to manage great amounts of data collected by means of field instruments i.e. radars, satellites, etc. Also, some software may be needed to analyze these data, which requires a large amount of computing resources.

In this example, several research institutions could provide and share data over a community cloud where the valuable information (field data) is safely stored and backed up. Some public cloud could be used as provider for the computing services to run particular private owned software.

Another case could be that some commercial software, available among the services that one public cloud provider offers, be useful to analyze the data. Then the use of the software could be bought as a service from the public cloud. Suppose that the results of the data analysis are a great amount of three dimensional maps which require a large amount of memory to be stored. They could be hosted in a public cloud. Therefore, the selection of the best kind of cloud depends on the problem at hand.

References

Bellamy, J. C. (2000). *Digital Telephony*. New York: John Wiley & Sons, Inc.
Department of Defense, United States of America. (2008). *Global positioning system, Standard positioning service, Performance standard, Navstar Global Positioning System,*
Ferrero, A. (2005). *Fundamentals of Metrology and Measurement Science*. IEEE Educational Courses. IEEE Instrumentation & Measurement Society, eLearning.
Feynman, R. P., Leighton, R. B., & Sands, M. (1967). *The Feynman Lectures on Physics*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.
Hughes (William J.) Technical Center. (2011). *Global Positioning System (GPS), Standard Positioning Service (SPS), Performance Analysis Report*. Report #74, NSTB/WAAS T&E Team, Atlantic City International Airport, NJ.
Jadeja, Y., & Modi, K. (2012). Cloud computing - concepts, architecture and challenges. *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*.
Lenz Ch., McCormick, C., Goldsmith, R., & Trachtman, E. (2010). Real time, near global, low earth orbit communications using geostationary Inmarsat BGAN system as a relay. *24th Annual AIAA/USU Conference on Small Satellites*. Pp. 11.
McCormick, C. C., Lenz, C., & Yunck, T. (2010). Real-Time Delivery of Global Environmental Observation Data From Space-Based Sensors using the Inmarsat BGAN System. *American Geophysical Union, Fall Meeting 2010*. Abstract #SM51A-1768.
Mollah, M. B., Islam, K. R., & Islam, S. S. (2012). Next generation of computing through cloud computing technology. *Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on*.
Proakis, J. G., & Salehi, M. (2008). *Digital Communications*. New York: McGraw-Hill Higher Education.
Tomasi, W. (2001). *Electronic Communication Systems: Fundamentals through Advanced*. New Jersey: Prentice Hall, Inc.
Understanding the Cloud Computing Stack SaaS, Paas, IaaS, Available at: www.rackspaceclouduniversity.com
http://www.hko.gov.hk/wservice/tsheet/pms/images/Iridium.pdf
http://www.iridium.com
http://www.orbcomm.com
http://www.globalw.com
http://www.inmarsat.com/services/index.htm
http://www.evologics.de/files/DataSheets/EvoLogics_S2CR_4878_Product_Information.pdf