Syllabus reference

1.2.1 Data transmission

Learners should be able to: show understanding of what is meant by transmission of data; distinguish between simplex, duplex and half-duplex data transmission; show understanding of the need to check for errors.

See also:

1.1 Binary systems (Bits and bytes)

2.1 Data transmission

How data is transmitted

Introduction

In Chapter 1 you learned how data is held in a digital electronic format inside a computer. In this chapter you will learn how data is shared and communicated between computers.

Data transmission

Computers hold data in binary form, using on/off switches. Each on/off signal represents one bit of data. Computers can turn this binary data into a stream of on/off signals which can be transmitted. That means the signals are sent from one place to another. The signals can be:

- electrical pulses that travel down metal cables
- pulses of light that travel down a fibre-optic cable
- wireless signals – radio signals, microwave and infrared waves – when the bits are sent as electromagnetic waves that move through space, and also air and many other materials (including human beings, who are not affected by them).

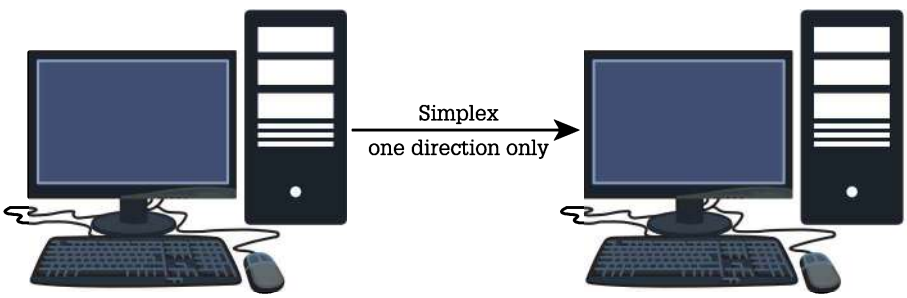
Whatever system is used to carry the on/off signals, it is known as the transmission medium.

Long or short distance

Data transmission is used to link computers. The computers can share and send data. The Internet is based on long-distance communication links.

Data transmission is also used over short distances. The different parts of a computer transmit data to each other. When you send your work to a printer, that requires data transmission.

Bluetooth is an example of a short distance wireless data link. Bluetooth can be used to link an earpiece (headset) to a mobile phone.



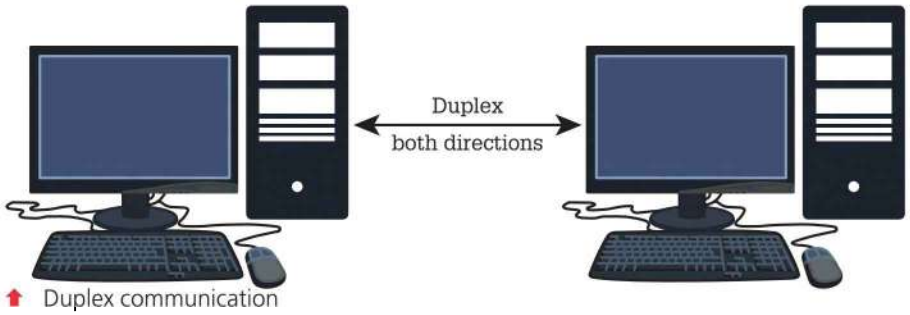
Simplex communication

Simplex and duplex

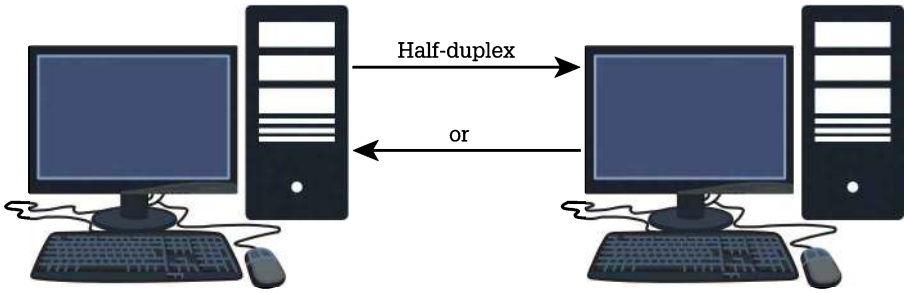
Communication links can be simplex, duplex or half-duplex.

- Simplex communication is a one-way link. The signal can only go in one direction. An example of a simplex communication is the signal from a closed-circuit TV camera to a security guard's monitor. The security guard can see on the monitor what the camera sends, but cannot send anything back to the camera.

- Duplex communication is a two-way link. The signal can go both ways. A phone conversation is an example of duplex communication. Both people can talk. Both people can listen.
- Half-duplex means the link can only carry signals in one direction at a time. The two sides have to take turns to send a signal. A walkie-talkie system is half-duplex.



Duplex communication

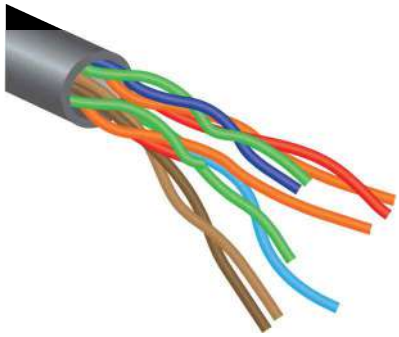


Half-duplex communication

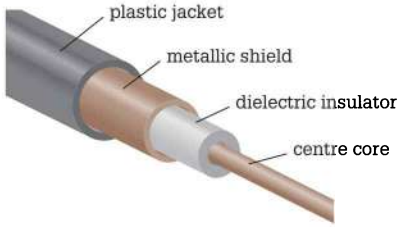
Types of cable

Cables connect devices together. Signals are sent along the cable. These are the main types of cable:

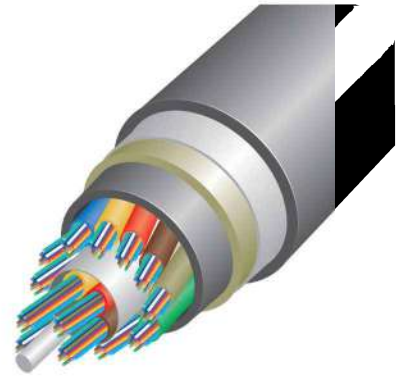
- Twisted pair cable: this is made of pairs of copper wires, individually insulated then twisted together. It is inexpensive, flexible and convenient, but it is not suitable for a long-distance link. Electrical interference can cause errors in the data. It is used for short-distance links.
- Coaxial cable: this is a metal cable, surrounded by a layer of insulation then another layer of metal. It is protected against electrical interference. It is more expensive than twisted pair, and it is not as flexible. It is used where cables need to go close to electrical and radio equipment.
- Fibre-optic cable is fairly expensive but it has many advantages over the other types of cable. It is not affected by electrical interference. It is suitable for long-distance links.



Twisted pair cable



Coaxial cable



Fibre-optic cable

Test yourself

1. What is the alternative to using cables for data transmission?
2. What is the difference between the content of a transmission and the transmission medium?
3. A computer is not connected to the Internet or any other computer. However, it still uses data transmission. Explain why.
4. I have a radio receiver but not a transmitter. What type of communication link is this?
5. Draw a diagram to show the difference between simplex, duplex and half-duplex communication.

Learning activities

You have seen that wireless communication can use different forms of signal: radio, microwave and infrared. Investigate examples of each type of wireless signal being used for communication.

Find examples of how the three types of cable are used.

Syllabus reference

1.2.1 Data transmission

Learners should be able to: show understanding of what is meant by transmission of data; distinguish between serial and parallel data transmission; show understanding of the reasons for choosing serial or parallel data transmission.

See also:

Chapter 1 Data representation

Serial or parallel?

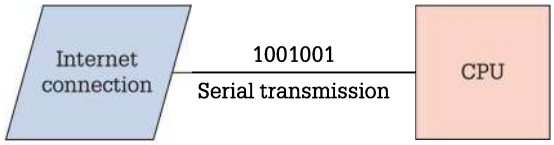
Introduction

You have learned that data transmission happens when bits are turned into signals. The signals are sent from one place to another. Now you will learn about the types of data transmission.

Serial transmission

Most data transmission is serial transmission. In serial transmission, the bits that make up the data are sent one at a time. The bits all travel along the same transmission medium, one after the other, in a series. The signals are sent down a single wire, or as a wireless signal. The bits arrive at the other end one at a time.

Serial transmission is the most reliable method of data transmission. The bits are kept separate from each other. They arrive in the same order that they were sent. Serial transmission is used for long-distance communication, for example an Internet connection.

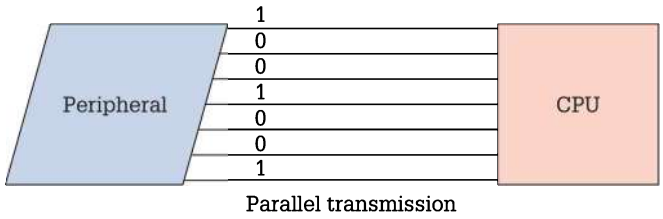


Serial transmission – the eight bits are sent one after the other down the same wire

Parallel transmission

You have learned that bits are held in groups of eight called bytes.

Some communication links use several wires at the same time. Each wire carries one bit, so several bits can be sent at the same time. Some parallel systems have eight wires. This means that all the bits in a byte can be sent at once. Each bit goes down a different wire. All the bits arrive at the same time.



Parallel transmission – the eight bits are sent at the same time along different wires

Advantages and disadvantages

Parallel transmission is quicker than serial transmission. Several bits are transmitted at the same time, so it takes less time to send the data.

However, there are risks when you use parallel transmission. There is more chance of an error in the signal. Microscopic differences in the wires might mean that they transmit signals at slightly different speeds. Over a long distance that might turn into a big difference. The bits won't all arrive at the same time. The signal will not be transmitted accurately. For this reason,

parallel transmission is only used for short-distance communications, for example to connect a monitor to the computer.

Serial transmission can take longer, because the bits are sent one at a time, but it is more reliable over a long distance.

Long or short distance?

Some connections are long distance. For example, Internet connections link computers all over the world. For these connections we would use serial transmission. Some connections are very short distance, for example connections between components inside the computer. For these, we might use parallel transmission.

Other connections are between these extremes, for example the connection between a computer and a printer in the same room. In this case either serial or parallel transmission can be chosen.

Serial and parallel ports

Old personal computers used to be equipped with both serial ports and parallel ports, and you could see the difference between them by looking at the pins on their connectors.

A serial port used only a few pins, with just one reserved for transmitting data and another for receiving. Other pins might be used to control the port, but they weren't used for data.

A serial port

A parallel port

Test yourself

1. What is the main advantage of parallel transmission?
2. Explain why parallel transmission may not be reliable over long distances.
3. What is the visual difference between serial and parallel cable?
4. What is a port on a computer?

Learning activity

Use graphics software to draw a diagram showing both serial and parallel transmission.

Complete this table to show the advantages and disadvantages of the different types of transmission

	Advantages	Disadvantages	Uses
Serial			Long distance
Parallel			Short distance

Syllabus reference

1.2.1 Data transmission

Learners should be able to: show understanding of the use of serial and parallel data transmission, in universal serial bus (USB) and integrated circuit (IC).

See also:

3.3 Inside the CPU

The lines on the circuit board are actually thin copper wires acting as data buses

Data bus

Introduction

You have learned about data transmission between computers. In this section you will learn about data transmission within a single computer system. This is how the different parts of a computer "talk to" each other.

Integrated circuits

Data in the computer is stored using on/off electronic switches. An integrated circuit (IC) is a collection of microscopic electronic circuits, sealed into a single plastic or ceramic package. Different ICs are used for different tasks inside the computer. Many ICs are used for data storage. One of the ICs, the central processing unit (CPU), contains the computer's processor and registers. All these parts must be connected together. We will talk more about the CPU in *3.3 Inside the CPU*.

The different ICs are linked by wired connections called data buses. Some ICs, particularly CPUs, have internal buses too, made from metallic layers within the IC. Each part of the IC works very quickly. The speed of a whole computer is strongly affected by how quickly the buses can transmit data between the different parts.

Parallel data bus

The buses inside the CPU, and between CPU and RAM, use parallel transmission, which has advantages and disadvantages:

- The advantage of parallel transmission is speed. The speed of each data bus strongly affects the performance of the computer system.
- The disadvantage of parallel transmission is that it needs more wires, so it takes up more of the very limited space available inside the IC or on the circuit board.

Connecting peripherals

The processor is at the centre of the computer. That is where the work of the computer takes place. A computer needs other devices such as a screen, a keyboard and a mouse. These additional devices are called peripherals. The peripherals have to be connected to the processor. Buses are used to connect the peripherals to the processor.

There are several ways to join a peripheral to the processor. It can be done using:

- permanent wiring, for example the keyboard of a laptop is permanently wired into the computer casing
- a plug-in cable, for example a monitor can be plugged in to the computer
- a wireless connection, for example a wireless mouse.

In each case, a bus is needed to complete the connection.

Peripherals work more slowly than the processor. For example, when you type data into the computer, your typing comes much more slowly than the computer works. A very fast connection is not so important. A serial bus is often used to connect a peripheral to the processor. A serial bus is slower than a parallel bus, but it is fast enough, and it is less expensive because it needs hardware for only a single signal.

Universal serial bus (USB)

Peripherals are made by many different companies. The manufacturers want people to buy their peripherals. They want to make it easy to connect the peripheral to a computer. They want the peripheral to work with all types of computer.

Nowadays most companies that make peripherals use a standard connection. It is called a USB (which stands for "universal serial bus"). It is a serial connection. It is called universal because it can be used in most modern computers.

Almost all modern computers have one or more USB ports. That means devices from many different manufacturers can be used with those computers.

A USB connection

Test yourself

1. What is meant by a data bus and why is bus speed so important?
2. Each part of the CPU works very fast. Explain why parallel transmission is suitable for use within the CPU.
3. A peripheral works much more slowly than the CPU. Explain why serial transmission is used to connect peripherals to the CPU.
4. A friend sets up a company making keyboards for computers. Write a short note to him, explaining why he should make keyboards that have a USB connection.

Learning activity

Investigate the computer where you are sitting. What peripherals does it have? What connections does it have? Is there a wireless connection? Write a short report giving your findings.

Syllabus reference

1.1.3 Data representation

Learners should be able to identify and describe methods of error detection and correction, such as automatic repeat request (ARQ).

Transmission errors

Introduction

Data transmission means sending bits and bytes from one location to another. It is important that the data is transmitted in full and without errors. In the rest of this chapter you will learn about ways to detect errors.

Transmission errors

Data is transmitted through a medium, which may use a cabled or a wireless connection. The transmission media carries bits in the form of electrical, radio or optical (light) pulses. All types of transmission media can be affected by errors.

Errors can be caused by flaws in the transmission medium, such as imperfections in a copper wire. Errors can be caused by external factors, such as electrical fields. We can design systems to reduce errors. For example, wires can be shielded by an outer conductive layer, to prevent electrical interference. Despite this, errors can still occur.

Transmission errors can have serious effects. Every bit in a signal is important. Changing one bit alters the value of the binary number. The whole signal will be wrong. For this reason it is important to check for errors in transmitted data. If the data has an error, it can be sent again.

Types of error

Errors in transmission can mean that:

- some of the bits are lost from the data stream
- extra bits are added to the data stream
- 1 bits change to 0, or 0 bits change to 1.

If a human operator is involved, for example someone typing the data, the person can also make errors. An error made in copying data, for example when typing it, is called a transcription error. An error where two letters or numbers are in the wrong order is called a transposition error.

Transmitter and receiver

Data transmission involves a transmitter and a receiver.

- The transmitter is the device that has the data to start with, and sends it.
- The receiver is the device that gets the data, after transmission.

Typically these devices are computers, but they could be a computer and its peripherals. The receiver will check the accuracy of the data sent by the transmitter. If an error is found, the receiver will ask the transmitter to send the data again.

Ways to detect errors

On the next few pages you will learn about ways the computer can check a transmission for errors. Parity checks are discussed on pages 40–41. Check digits are discussed on pages 42–43.

Automatic Repeat reQuest (ARQ)

The error checks methods are used in a process called automatic repeat request (ARQ). This is a method to ensure correct transmission of data. It works like this:

- The transmitter sends some data (called a packet).
- When the receiver gets the data packet, it checks it for errors.
- If the receiver finds no errors, it will send an acknowledgement.
- If the transmitter doesn't receive an acknowledgement, it sends the data again.

The transmitter will keep on sending the data packet until it receives an acknowledgement. There is usually a time limit. Once the time is up, the transmitter will stop trying to send the package. The signal has timed out.

Test yourself

1. Give the meanings of transmission errors, transposition errors and transcription errors.
2. What changes can occur to data bits as the result of transmission errors?
3. How does shielding help to prevent transmission errors?
4. Explain the roles of the transmitter and receiver devices in the ARQ process.

Types of error

If two digits or letters are accidentally swapped around, this is called a **transposition error**. It is one of the most common types of **transmission error**.

Syllabus reference

1.2.1 Data transmission

Learners should be able to explain how parity bits are used for error detection.

1.1.3 Data representation

Learners should be able to identify and describe methods of error detection and correction, such as parity checks.

Parity check

Introduction

You have learned that it is important to detect errors in data transmission. In this section you will learn about parity checks. A parity check is one way to check data for errors. Make sure you know the difference between the role of receiver and transmitter.

Parity

"Parity" is a term from mathematics. It means whether a number is odd or even. Parity is used to check whether data has errors in it. The most common type of parity check is known as an even parity bit.

Before the data is sent, the transmitter counts how many 1s there are in each byte. The transmitter then adds an extra bit to the end of each byte:

- If there is an even number of 1s in the byte, the parity bit is set to 0.
- If there is an odd number of 1s in the byte, the parity bit is set to 1.

This extra bit makes sure the number of 1s transmitted is an even number.

After the data is received:

- The receiver counts how many 1s there are in each byte plus its parity bit.
- Each byte plus parity bit should have an even number of 1s.

If any of the bytes with its parity bit has an odd number of bits, the receiver will know there was an error during transmission. The data must be sent again.

Even parity bit

In the example on this page, the even parity bit is shown at the right of the transmitted number. In practice it may not be stored in this location.

Worked example

The transmitter got ready to send this signal.

0	1	1	0	0	0	1	

There are three 1s in the data. That is an odd number, so the parity bit was set to 1. Now there are four 1s in the byte – an even number. The parity bit is highlighted.

0	1	1	0	0	0	1	1

Next the signal was transmitted. There was an error during transmission. One of the bits was altered by an error in transmission. The error is highlighted. The signal has gone wrong.

0	1	0	0	0	0	1	1

The computer that received the data added up the number of 1s in the signal. There were three 1s in the signal. That is an odd number, so there must have been a transmission error.

In conclusion: the error has been spotted. The data has to be sent again.

Odd parity

Some communication systems use "odd parity". In this system the number of 1s in each byte is an odd number. Otherwise it works just the same as even parity. Of course, the transmitter and receiver must both use the same system.

Which bit?

Data is normally stored and sent in groups of eight bits:

- In some cases, seven of the bits are used to send the data. The eighth bit is a parity bit. The parity bit is part of the byte.
- In other cases all eight bits are used to send the data. The parity bit is sent as an extra signal following the byte.

Limitations

The parity method is not perfect:

- If there are two errors in a byte (or any even number of errors) then the parity check will fail.
- If two bits get swapped round (transposition error) then the parity check will not spot the error.

For this reason other data checks are used as well as a parity bit.

Test yourself

Here is a block of data to be transmitted.

0 0 0 1 1 0 0
1 1 1 0 1 1 0
1 0 1 0 1 0 0
1 1 1 1 0 0 0
1 0 0 0 1 1 0
1 0 0 1 1 1 0
0 0 0 1 0 0 1

Assume that you are using even parity and add a parity bit to each row.

Learning activity

This is an extension activity. You will work with the block of data from the last question. You have added an even parity bit in each row. Now look at each column of data. Assume that you are using even parity and add a parity bit to the bottom of each column.

Now the data has a parity bit at the end of each row and the bottom of each column. This gives an additional check. It overcomes the two limitations mentioned on this page.

Syllabus reference

1.2.1 Data transmission

Learners should be able to identify and describe methods of error detection and correction, such as check digits and checksums.

Check digit and checksum

Introduction

Parity checks are error checks for binary data. In this section you will learn about error checks that can be used for denary numbers. These checks can detect transmission errors, and human errors such as typing (transcription) errors.

Check digit

A check digit is similar to a parity bit. It is added to the end of a denary number. The check digit is worked out from the digits in the number.

A check digit is used in the same way as a parity bit:

- A computer works out the check digit before transmission, and sends it with the number.
- The receiving computer works out the check digit after transmission, and compares it to the original.
- The two check digits should match.
- If they do not match there has been an error in transmitting the number.

Calculate the check digit

There is more than one way to work out the check digit from a denary number.

Sum

In mathematics the sum is the result of adding together a group of numbers.

Modulo (mod)

In mathematics the modulo is the remainder that is left after a division has been carried out. It is shortened to "mod".

Simple method

This is the simplest way to work out the check digit:

- Add up all the digits in the number to give the sum of the digits.
- Divide the sum by 10.
- The remainder from this division is used as the check digit.

The remainder when dividing a value by 10 is called the value modulo 10, or the value mod 10. Sometimes a check digit uses modulo 11, where the sum is divided by 11. If the remainder is 10, the letter X is used as the check digit.

A simple check digit cannot identify a transposition error. That is when two digits get swapped around. That is because a transposition of two digits will not change the overall sum of the digits.

Other methods

For this reason, other ways to calculate a check digit have been invented. The digits in different positions in the number are multiplied by different values. Then the numbers are added together and the check digit is calculated in one of the following ways:

- using the Luhn method, every second digit is multiplied by 2 and the total must be an exact multiple of 10.
- using the ISBN-10 method, every digit is multiplied by its position in the number: the first digit by 1, the next by 2 and so on. The check digit is the total mod11.

These methods are more complex, but they have a big advantage. If a digit is accidentally transmitted in the wrong position, the sum will change. The check digit will change. That means these methods will detect transposition errors.

Every credit card has a number on it which includes a check digit

Checksum

Using a checksum is a way of checking a group of numbers:

- The transmitter adds up the total of a group of numbers before transmission.
- The transmitter sends the total along with the numbers.
- The receiver works out the total and compares it to the transmitted total.
- The two totals should match. If the two totals do not match then there has been an error. The data must be sent again.

You can use a checksum even if the sum value does not represent a real total. For example, you can use a checksum when sending a group of phone numbers. A number of this kind – which is not a real total – is called a hash total.

Test yourself

1. A check digit is calculated twice, once each by two different computers. Explain why.
2. What is 32 mod 10?
3. Why is any value mod 10 always a single-digit number?
4. What is the limitation with using a simple sum and mod 10 check digit?
5. When do we call a control sum a hash total?

Learning activity

1. Write down your phone number. Calculate a simple check digit for your phone number, using mod 10.
2. Collect the phone numbers of five friends. Calculate a checksum for this list of numbers.
3. Find a book with a 10-digit ISBN code. Using the first 9 digits, work out the check digit. Check this against the example you see on the real book. It should match.

Use of check digits

A credit card number includes a check digit. Since 1960, the Luhn method has been used to calculate check digits on credit cards.

Every book published is given a code number called the International Standard Book Number (ISBN). These numbers can be either 10 or 13 digits long (ISBN-10 or ISBN-13). Both formats include check digits.