1

Introduction

The field of telecommunications has evolved from a stage when signs, drum beats and semaphores were used for long distance communication to a stage when electrical, radio and electro-optical signals are being used. Optical signals produced by laser sources and carried by ultra-pure glass fibres are recent additions to the field. Telecommunication networks carry information signals among entities which are geographically far apart. An entity maybe a ~~computer, a human being, a facsimile~~ machine, a teleprinter, a data terminal, and so on. Billions of such entities the world-over are involved in the process of information transfer which may be in the form of a telephone conversation or a file transfer between two computers or a message transfer between two terminals, etc. In telephone conversation, the one who initiates the call is referred to as the **calling subscriber** and the one for whom the call is destined is the **called subscriber**. In other cases of information transfer, the communicating entities are known as **source** and **destination**, respectively.

The full potential of telecommunications is realised only when any entity in one part of the world can communicate with any other entity in another part of the world. Modern telecommunication networks attempt to make this idea of 'universal connectivity' a reality. Connectivity in telecommunication networks is achieved by the use of switching systems. This text deals with the telecommunication switching systems and the networks that use them to provide worldwide connectivity.

1.1 Evolution of Telecommunications

Historically, transmission of telegraphic signals over wires was the first technological development in the field of modern telecommunications. Telegraphy was introduced in 1837 in Great Britain and in 1845 in France. In March 1876, Alexander Graham Bell demonstrated his telephone set and the possibility of telephony, i.e. long-distance voice transmission. Graham Bell's invention was one of those rare inventions which was put to practical use almost immediately. His demonstrations laid the foundation for telephony.

Graham Bell demonstrated a point-to-point telephone connection. A network using point-to-point connections is shown in Fig. 1.1. In such a

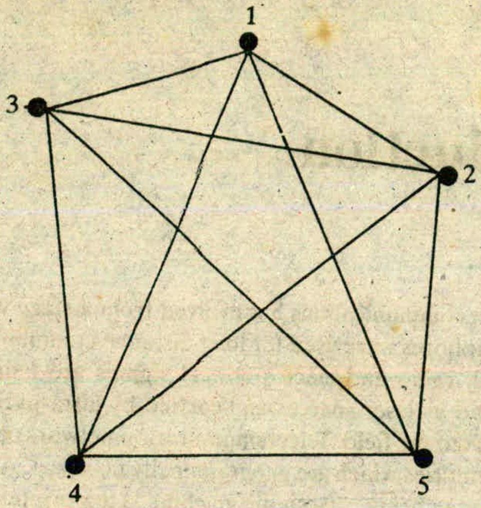


Fig. 1.1 A network with point-to-point links.

network, a calling subscriber chooses the appropriate link to establish connection with the called subscriber. In order to draw the attention of the called subscriber before information exchange can begin, some form of **signalling** is required with each link. If the called subscriber is engaged, a suitable indication should be given to the calling subscriber by means of signalling.

In Fig. 1.1, there are five entities and 10 point-to-point links. In a general case with n entities, there are $n(n-1)/2$ links. Let us consider the n entities in some order. In order to connect the first entity to all other entities, we require $(n-1)$ links. With this, the second entity is already connected to the first. We now need $(n-2)$ links to connect the second entity to the others. For the third entity, we need $(n-3)$ links, for the fourth $(n-4)$ links, and so on. The total number of links, L, works out as follows:

$$L = (n-1) + (n-2) + \ldots + 1 + 0 = n(n-1)/2 \tag{1.1}$$

Networks with point-to-point links among all the entities are known as **fully connected networks**. The number of links required in a fully connected network becomes very large even with moderate values of n. For example, we require 1225 links for fully interconnecting 50 subscribers. Consequently, practical use of Bell's invention on a large scale or even on a moderate scale demanded not only the telephone sets and the pairs of wires, but also the so-called **switching system** or the **switching office** or the **exchange**. With the introduction of the switching systems, the subscribers are not connected directly to one another; instead, they are connected to the switching system as shown in Fig. 1.2. When a subscriber wants to communicate with another,

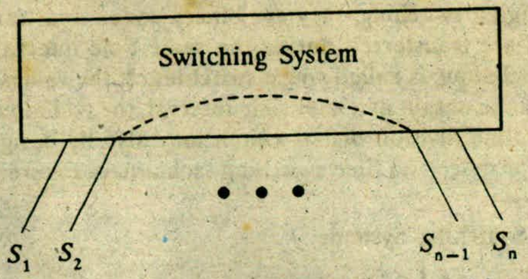


Fig. 1.2 Subscriber interconnection using a switching system.

a connection is established between the two at the switching system. Figure 1.2 shows a connection between subscriber S_2 and S_{n-1}. In this configuration, only one link per subscriber is required between the subscriber and the switching system, and the total number of such links is equal to the number of subscribers connected to the system. Signalling is now required to draw the attention of the switching system to establish or release a connection. It should also enable the switching system to detect whether a called subscriber is busy and, if so, indicate the same to the calling subscriber. The functions performed by a switching system in establishing and releasing connections are known as **control functions**.

Early switching systems were manual and operator oriented. Limitations of operator manned switching systems were quickly recognised and automatic exchanges came into existence. Automatic switching systems can be classified as **electromechanical** and **electronic**. Electromechanical switching systems include **step-by-step** and **crossbar systems**. The step-by-step system is better known as Strowger switching system after its inventor A.B. Strowger. The control functions in a Strowger system are performed by circuits associated with the switching elements in the system. Crossbar systems have hard-wired control subsystems which use relays and latches. These subsystems have limited capability and it is virtually impossible to modify them to provide additional functionalities. In electronic switching systems, the control functions are performed by a computer or a processor. Hence, these systems are called **stored program control (SPC)** systems. New facilities can be added to a SPC system by changing the control program. The switching scheme used by electronic switching systems may be either **space division switching** or **time division switching**. In space division switching, a dedicated path is established between the calling and the called subscribers for the entire duration of the call. Space division switching is also the technique used in Strowger and crossbar systems. An electronic exchange may use a crossbar switching matrix for space division switching. In other words, a crossbar switching system with SPC qualifies as an electronic exchange.

In time division switching, sampled values of speech signals are transferred at fixed intervals. Time division switching may be analog or digital. In analog switching, the sampled voltage levels are transmitted as they are, whereas in digital switching, they are binary coded and transmitted. If the coded values are transferred during the same time interval from input to output, the technique is called **space switching.** If the values are stored and transferred to the output at a later time interval, the technique is called **time switching.** A time division digital switch may also be designed by using a combination of space and time switching techniques. Figure 1.3 summarises

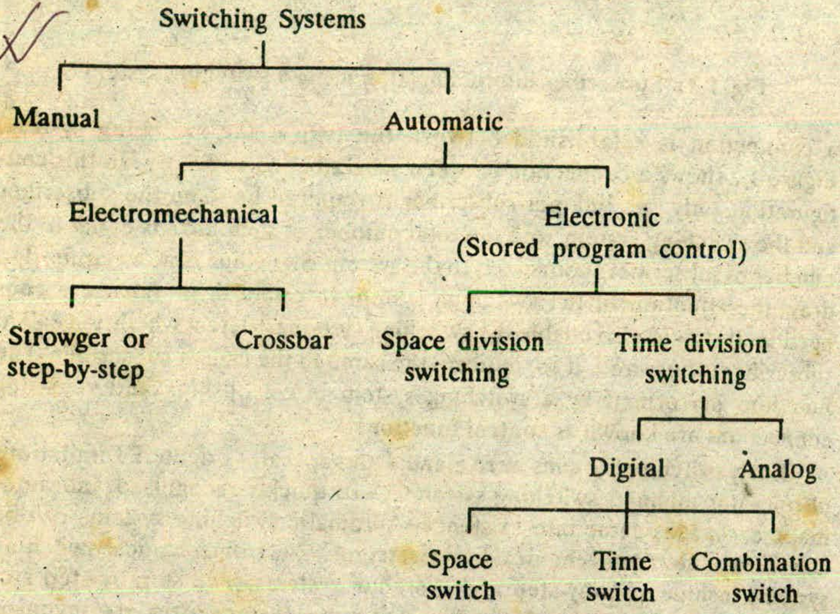
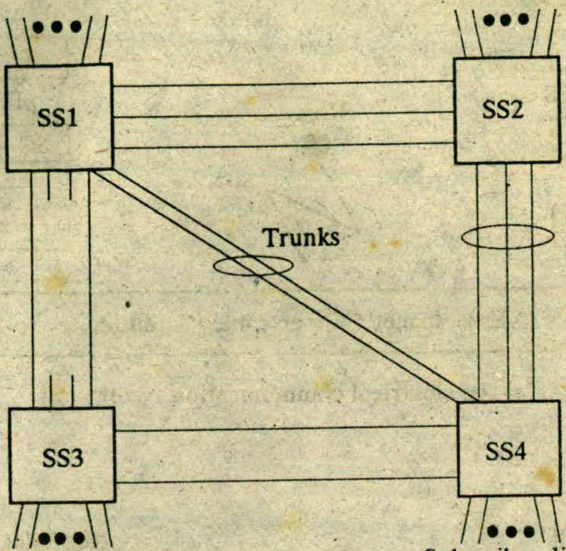


Fig. 1.3 Classification of switching systems.

the classification of switching systems. In Chapters 2 and 3, we deal with electromechanical switching systems. Electronic space division networks are discussed in Chapter 4. Digitisation of speech, which is a fundamental requirement for electronic time division switching networks, is discussed in Chapter 5, and the time division switching techniques in Chapter 6.

Subscribers all over the world cannot be connected to a single switching system unless we have a gigantic switching system in the sky and every subscriber has a direct access to the same. Although communication satellite systems covering the entire globe and low cost roof-top antenna present such a scenario, the capacity of such systems is limited at present. The major part of the telecommunication networks is still ground based, where subscribers are connected to the switching system via copper wires. Technological and engineering constraints of signal transfer on a pair of wires necessitate that the subscribers be located within a few kilometres from the switching system. By introducing a number of stand-alone switching systems in appropriate geographical locations, communication capability can be established among the subscribers in the same locality. However, for subscribers in different localities to communicate, it is necessary that the switching systems are interconnected in the form of a network. Figure 1.4 shows a telecommuni-

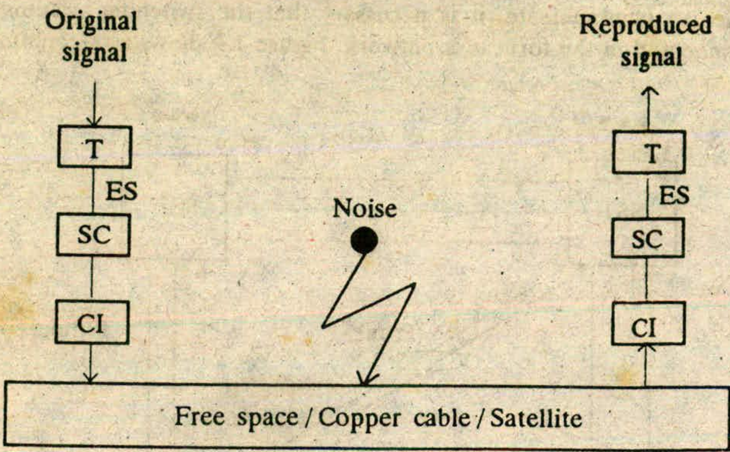


SS = switching system

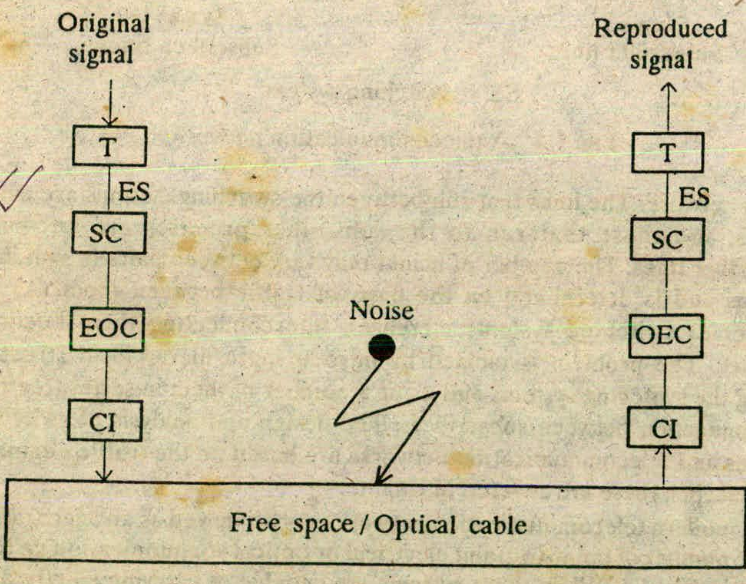
Fig. 1.4 A telecommunication network.

cation network. The links that run between the switching systems are called **trunks**, and those that run to the subscriber premises are known as **subscriber lines**. The number of trunks may vary between pairs of switching systems and is determined on the basis of traffic between them. As the number of switching systems increases, interconnecting them becomes complex. The problem is tackled by introducing a hierarchical structure among the switching systems and using a number of them in series to establish connection between subscribers. The design and analysis of switching systems and telecommunication networks are based on the traffic engineering concepts; these are covered in Chapter 8.

A modern telecommunication network may be viewed as an aggregate of a large number of point-to-point electrical or optical communication systems shown in Fig. 1.5. While these systems are capable of carrying electrical or optical signals, as the case may be, the information to be conveyed is not always in the form of these signals. For example, human speech signals need to be converted to electrical or optical signals before they can be carried by a communication system. Transducers perform this energy conversion. Present day optical sources require electrical signals as input, and the optical



(a) An electrical communication system



(b) An optical communication system

CI = channel interface EOC = electrical to optical converter
ES = electrical signal OEC = optical to electrical converter
SC = signal conditioner T = transducer.

Fig. 1.5 Elements of a communication system.

detectors produce electrical signals as output. Hence, the original signals are first converted to electrical signals and then to optical signals at the transmitting end of an optical communication system and at the receiving end optical signals are converted to electrical signals before the original signal is reproduced. A medium is required to carry the signals. This medium, called the channel, may be the free space, a copper cable, or the free space in conjunction with a satellite in the case of an electrical communication system. An optical communication system may use the line-of-sight free space or fibre optic cables as the channel. Channels, in general, are lossy and prone to external noise that corrupts the information carrying signals. Different channels exhibit different loss characteristics and are affected to different degrees by noise. Accordingly, the chosen channel demands that the information signals be properly conditioned before they are transmitted, so that the effect of the lossy nature of the channel and the noise is kept within limits and the signals reach the destination with acceptable level of intelligibility and fidelity. Signal conditioning may include amplification, filtering, band-limiting, multiplexing and demultiplexing. Fibre optic communication systems are emerging as major transmission systems for telecommunications. Chapter 7 deals with this newly emerging communication system.

The channel and the signal characteristics of individual communication systems in a telecommunication network may vary widely. For example, the communication system between the subscriber and the switching system uses most often a pair of copper wires as the channel, whereas the communication system between the switching systems may use a coaxial cable or the free space (microwave) as the channel. Similarly, the type of end equipment used at the subscriber premises would decide the electrical characteristics of signals carried between the subscriber end and the switching system. For example, electrical characteristics of teleprinter signals are completely different from those of telephone signals. In fact, such wide variations in signal characteristics have led to the development of different service specific telecommunication networks that operate independently. Examples are:

1. Telegraph networks
2. Telex networks
3. Telephone networks
4. Data networks.

We discuss the telephone networks in Chapter 9 and the data networks in Chapter 10. Management and maintenance of multiple networks are expensive. The question then arises: Is it possible to design a single network that can carry all the services? The key to the solution of this problem lies in the digitalisation of services. If all the service specific signals can be converted to a common digital domain, a network capable of transporting digital signals can carry the multitude of services. This approach is leading to the evolution of the integrated services digital network (ISDN) which is discussed in Chapter 11.

1.2 Simple Telephone Communication

In the simplest form of a telephone circuit, there is a one way communication involving two entities, one receiving (listening) and the other transmitting (talking). This form of one way communication shown in Fig. 1.6 is known as

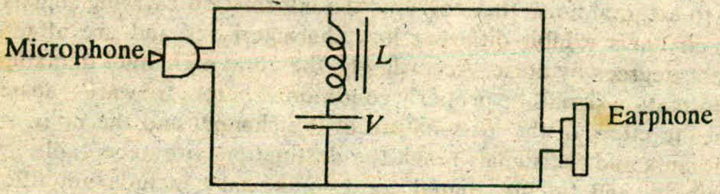


Fig. 1.6 A simplex telephone circuit.

simplex communication. The microphone and the earphone are the transducer elements of the telephone communication system. Microphone converts speech signals into electrical signals and the earphone converts electrical signals into audio signals. Most commonly used microphone is a carbon microphone. Carbon microphones do not produce high fidelity signals, but give out strong electrical signals at acceptable quality levels for telephone conversation. In carbon microphones, a certain quantity of small carbon granules is placed in a box. Carbon granules conduct electricity and the resistance offered by them is dependent upon the density with which they are packed. One side of the box cover is flexible and is mechanically attached to a diaphragm. When sound waves impinge on the diaphragm, it vibrates, causing the carbon granules to compress or expand, thus changing the resistivity offered by the granules. If a voltage is applied to the microphone, the current in the circuit varies according to the vibrations of the diaphragm. The theory of the carbon microphone indicates that the microphone functions like an amplitude modulator. When the sound waves impinge on the diaphragm, the instantaneous resistance of the microphone is given by

$$r_i = r_0 - r \sin \omega t \qquad (1.2)$$

where

r_0 = quiescent resistance of the microphone when there is no speech signal

r = maximum variation in resistance offered by the carbon granules, $r < r_0$

r_i = instantaneous resistance.

The negative sign in Eq. (1.2) indicates that when the carbon granules are compressed the resistance decreases and vice versa. Ignoring impedances external to the microphone in the circuit given in Fig. 1.6, without loss of generality, the instantaneous current in the microphone is given by

$$i = V/(r_0 - r \sin \omega t) = I_0(1 - m \sin \omega t)^{-1} \tag{1.3}$$

where

$I_0 = V/r_0$ = quiescent current in the microphone

$m = r/r_0, \quad m < 1$

By binomial theorem, Eq. (1.3) may be expanded as

$$i = I_0(1 + m \sin \omega t + m^2 \sin^2 \omega t + \ldots) \tag{1.4}$$

If the value of m is sufficiently small, which is usually the case in practice, higher-order terms can be ignored in Eq. (1.4), giving thereby

$$i = I_0(1 + m \sin \omega t) \tag{1.5}$$

which resembles the amplitude modulation (AM) equation. Thus, the carbon granule microphone acts as a modulator of the direct current I_0 which is analogous to the carrier wave in AM systems. The quantity m is equivalent to the modulation index in AM. The higher-order terms in Eq. (1.4) represent higher-order harmonic distortions, and hence it is essential that the value of m be kept sufficiently low. In Eq. (1.5), the alternating current output i is zero if the quiescent current I_0 is zero. Hence, the flow of this steady current through the microphone is essential, and the current itself is known as energising current. In Fig.1.6, the inductor acts as a high impedance element for voice frequency signals but permits the d.c. from the battery to flow to the microphone and the receiver. The voice frequency signals generated by the microphone reach the earphone without being shunted by the battery arm and are converted to audio signals there.

The earphone is usually an electromagnet with a magnetic diaphragm which is positioned such that there is an air gap between it and the poles of the electromagnet. When the electromagnet is energised by passing a current, a force is exerted on the diaphragm. The voice frequency current from the microphone causes variation in the force exerted by the electromagnet, thus vibrating the diaphragm and producing sound waves. Faithful reproduction of the signals by the receiver requires that the magnetic diaphragm be displaced in one direction from its unstressed position. The quiescent current provides this bias. In some circuits, a permanent magnet is used to provide the necessary displacement instead of the quiescent current. The instantaneous flux linking the poles of the electromagnet and the diaphragm is given by

$$\phi_i = \phi_0 + \phi \sin \omega t$$

where

ϕ_0 = constant flux due to the quiescent current or the permanent magnet

ϕ = maximum amplitude of flux variation, $\phi < \phi_0$

ϕ_i = instantaneous flux

Equation (1.6) assumes that the vibrations of the diaphragm are too small to affect the length of the air gap and that the reluctance of the magnetic path is constant. The instantaneous force exerted on the diaphragm is proportional to the square of the instantaneous flux linking the path. Therefore,

$$F = K(\phi_0 + \phi \sin \omega t)^2 \tag{1.7}$$

where K is the constant of proportionality. Expanding the right-hand side of Eq. (1.7), we have

$$F = K(\phi_0^2 + \phi^2 \sin^2 \omega t + 2\phi_0\phi \sin \omega t) \tag{1.8}$$

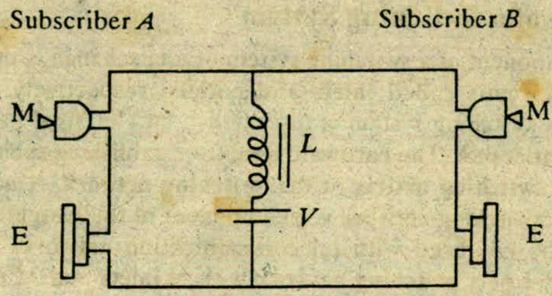
When $(\phi / \phi_0) \ll 1$, we can ignore the second-order term in Eq.(1.8). We then have

$$F = K\phi_0^2(1 + K_1 I_0 \sin \omega t) \tag{1.9}$$

where $I_0 \sin \omega t$ is the current flowing through the coil. We thus see that the force experienced by the diaphragm is in accordance with the signals produced by the microphone.

In a normal telephone communication system, information is transferred both ways. An entity is capable of both receiving and sending although these do not take place simultaneously. An entity is either receiving or sending at any instant of time. When one entity is transmitting, the other is receiving and vice versa. Such a form of communication where the information transfer takes place both ways but not simultaneously is known as **half-duplex communication**. If the information transfer takes place in both directions simultaneously, then it is called **full-duplex communication.**

Figure 1.6 may be modified to achieve half-duplex communication by the introduction of a transmitter and receiver at both ends of the circuit as shown in Fig. 1.7. In this circuit, the speech of A is heard by B as well as in A's own earphone. This audio signal, heard at the generating end, is called **sidetone.** A certain amount of sidetone is useful, or even essential. Human speech and hearing system is a feedback system in which the volume of speech is automatically adjusted, based on the sidetone heard by the ear. If no sidetone is present, a person tends to shout, and if too much of sidetone is present, there is a tendency to reduce the speech to a very low level. In the circuit of Fig. 1.7, the entire speech intensity is heard as sidetone, which is not desirable. Figure 1.8 gives a circuit where a small level of sidetone and the



E = earphone M = microphone

Fig. 1.7 A half-duplex telephone circuit.

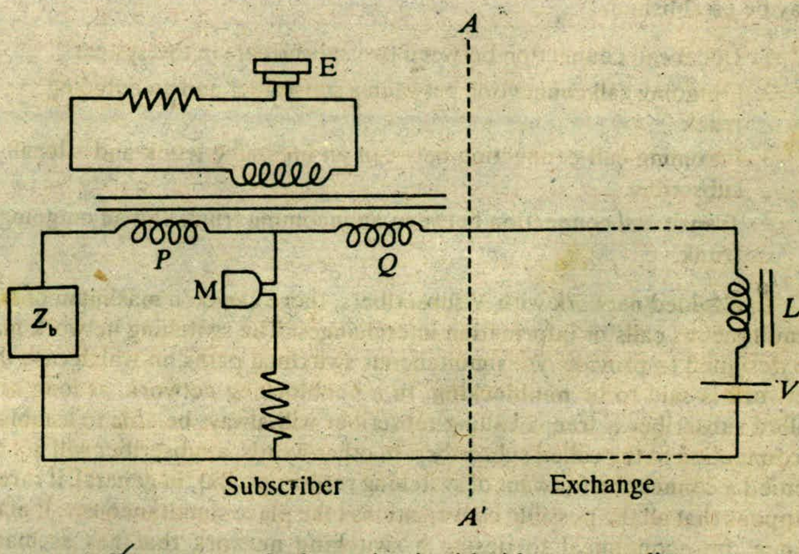


Fig. 1.8 A telephone circuit with sidetone coupling.

full speech signal from the other party are coupled to the receiver. The impedance Z_b is chosen to be more or less equal to the impedance seen by the circuit to the right of section AA'. As a consequence, with proper sidetone coupling the speech signal from the microphone M divides more or less equally in the two windings P and Q. Since the signals in these two windings are in the opposite direction, only a small induced voltage appears in the receiver circuit providing the sidetone. When a signal is received from the other entity, it travels in the same direction in both windings P and Q, inducing a large signal in the receiver circuit.

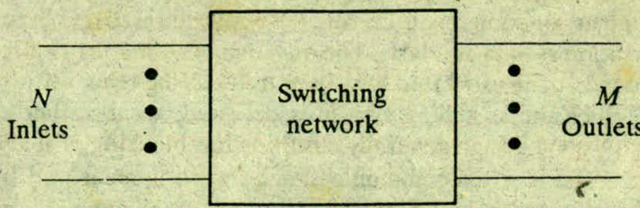
1.3 Basics of a Switching System

A major component of a switching system or an exchange is the set of input and output circuits called **inlets** and **outlets**, respectively. The primary function of a switching system is to establish an electrical path between a given inlet-outlet pair. The hardware used for establishing such a connection is called the **switching matrix** or the **switching network**. It is important to note that the switching network is a component of the switching system and should not be confused with telecommunication network. Figure 1.9(a) shows a model of a switching network with N inlets and M outlets. When $N = M$, the switching network is called a **symmetric network**. The inlets/outlets may be connected to local subscriber lines or to trunks from/to other exchanges as shown in Fig. 1.9(b). When all the inlets/outlets are connected to the subscriber lines, the logical connection appears as shown in Fig. 1.9(c). In this case, the output lines are folded back to the input and hence the network is called a **folded network**. In Fig. 1.9(b), four types of connections may be established:

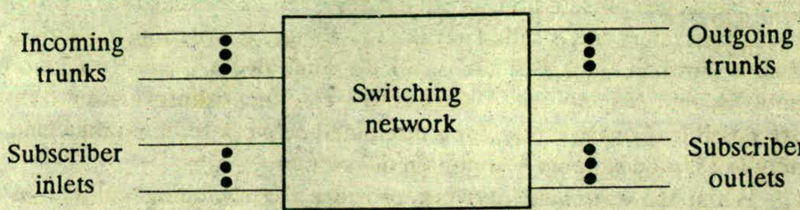
1. Local call connection between two subscribers in the system
2. Outgoing call connection between a subscriber and an outgoing trunk
3. Incoming call connection between an incoming trunk and a local subscriber
4. Transit call connection between an incoming trunk and an outgoing trunk.

In a folded network with N subscribers, there can be a maximum of $N/2$ simultaneous calls or information interchanges. The switching network may be designed to provide $N/2$ simultaneous switching paths, in which case the network is said to be **nonblocking**. In a nonblocking network, as long as a called subscriber is free, a calling subscriber will always be able to establish a connection to the called subscriber. In other words, a subscriber will not be denied a connection for want of switching resources. But, in general, it rarely happens that all the possible conversations take place simultaneously. It may, hence, be economical to design a switching network that has as many simultaneous switching paths as the average number of conversations expected. In this case, it may occasionally happen that when a subscriber requests a connection, there are no switching paths free in the network, and hence he is denied connection. In such an event, the subscriber is said to be blocked, and the switching network is called a **blocking network**. In a blocking network, the number of simultaneous switching paths is less than the maximum number of simultaneous conversations that can take place. The probability that a user may get blocked is called **blocking probability**.

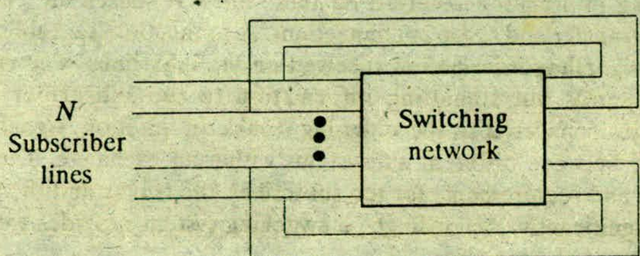
All the switching exchanges are designed to meet an estimated maximum average simultaneous traffic, usually known as **busy hour traffic**. Past records of the telephone traffic indicate that even in a busy exchange, not



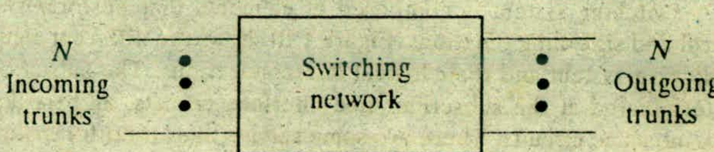
(a) Model of a switching network



(b) Inlets/outlets connections



(c) Folded network



(d) Nonfolded network

Fig. 1.9 Switching network configurations.

more than 20–30 per cent of the subscribers are active at the same time. Hence, switching systems are designed such that all the resources in a system are treated as common resources and the required resources are allocated to a conversation as long as it lasts. The quantum of common resources is determined based on the estimated busy hour traffic. When the traffic exceeds the limit to which the switching system is designed, a subscriber experiences blocking. A good design generally ensures a low blocking probability.

The traffic in a telecommunication network is measured by an internationally accepted unit of traffic intensity known as **erlang** (E), named after an illustrious early contributor to traffic theory. A switching resource is said to carry one erlang of traffic if it is continuously occupied throughout a given period of observation. Teletraffic concepts are discussed in Chapter 8.

In a switching network, all the inlet/outlet connections may be used for interexchange transmission. In such a case, the exchange does not support local subscribers and is called a **transit exchange**. A switching network of this kind is shown in Fig. 1.9(d) and is called a **nonfolded network**. In a nonfolded network with N inlets and N outlets, N simultaneous information transfers are possible. Consequently, for a nonfolded network to be nonblocking, the network should support N simultaneous switching paths.

While the switching network provides the switching paths, it is the control subsystem of the switching system that actually establishes the path. The switching network does not distinguish between inlets/outlets that are connected to the subscribers or to the trunks. It is the job of the control subsystem to distinguish between these lines and interpret correctly the signalling information received on these lines. It senses the end of information transfer and releases connections. A connection is established, based on the signalling information received on the inlet lines. The control subsystem sends out signalling information to the subscriber and other exchanges connected to the outgoing trunks. In addition, signalling is also involved between different subsystems within an exchange. The signalling formats and requirements for the subscriber, the trunks and the subsystems differ significantly. Accordingly, a switching system provides for three different forms of signalling:

1. Subscriber loop signalling
2. Interexchange signalling
3. Intraexchange or register signalling.

A switching system is composed of elements that perform switching, control and signalling functions. Figure 1.10 shows the different elements of a switching system and their logical interconnections. The subscriber lines are terminated at the subscriber line interface circuits, and trunks at the trunk interface circuits. There are some service lines used for maintenance and testing purposes. Junctor circuits imply a folded connection for the local subscribers and the service circuits. It is possible that some switching systems

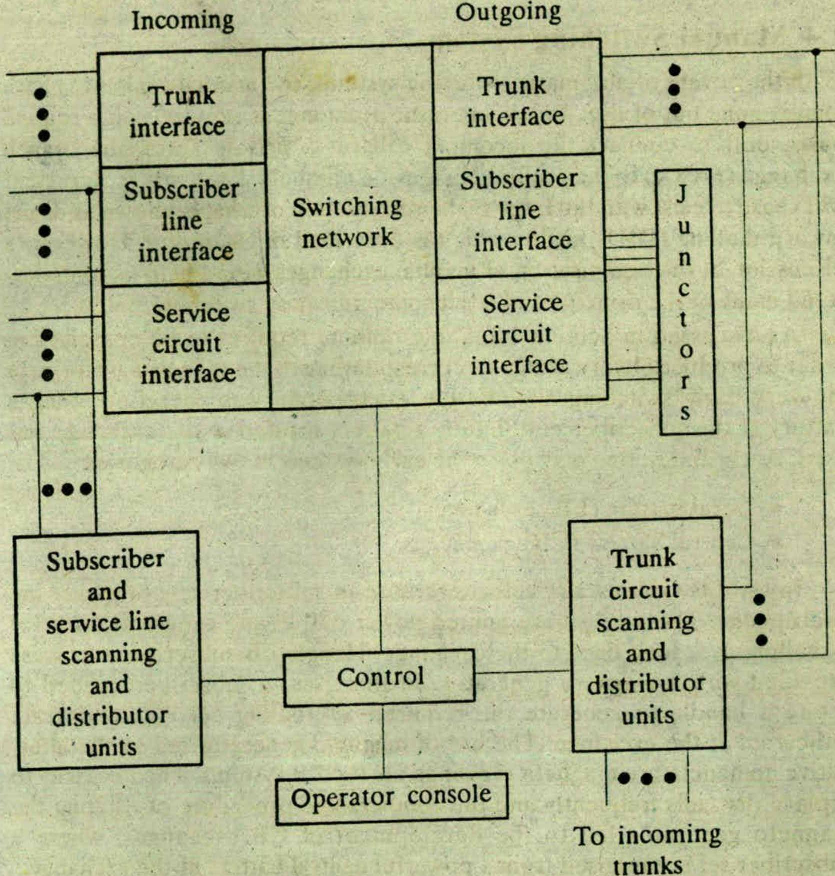


Fig. 1.10 Elements of a switching system.

provide an internal mechanism for local connections without using the junctor circuits. Line scanning units sense and obtain signalling information from the respective lines. Distributor units send out signalling information on the respective lines. Operator console permits interaction with the switching system for maintenance and administrative purposes. In some switching systems, the control subsystem may be an integral part of the switching network itself. Such systems are known as **direct control** switching systems. Those systems in which the control subsystem is outside the switching net work are known as **common control** switching systems. Strowger exchanges are usually direct control systems, whereas crossbar and electronic exchanges are common control systems. All stored program control systems are common control systems. Common control is also known as indirect control or register control.

1.4 Manual Switching System

With the advent of automatic switching systems, the manual exchanges have almost gone out of use. Today, operator assistance is required on a routine basis, only to connect the incoming calls at a private automatic branch exchange (PABX) to the required extension numbers. Even this requirement will cease to exist with the large scale introduction of what is known as direct inward dialling (DID) facility which is described in Chapter 9. However, a discussion of the organisation of manual exchanges would help us to understand many of the principles of a telecommunication switching system.

As discussed in Section 1.2, a microphone requires to be energised in order to produce electrical signals corresponding to the speech waveform. In the very early switching systems the microphone was energised using a battery at the subscriber end. Later, a battery located at the exchange was used. Accordingly, one may place the early systems in two categories:

- Local battery (LB) exchanges
- Central battery (CB) exchanges.

In the LB systems, dry cells were used in subscriber sets to power the microphone. These cells have limited power output and cannot be used for signalling over long lines to the exchange. Hence, LB subscriber sets were provided with a magneto generator. In this case, a subscriber needed to rotate a handle to generate the required alternating current to operate indicators at the exchange. The use of magneto generator led to the alternative nomenclature **magneto exchange** for the LB systems. The necessity to replace dry cells frequently and the cumbersome procedure of rotating the magneto generator led to the development of CB exchanges, where a subscriber set is energised from a powerful central battery at the exchange.

Almost all the present day telephone exchanges are CB systems, although it is not inconceivable that future systems may resort to LB structures if low cost reliable power supplies for the subscriber premises become available. A simple CB system operated by a human being is shown in Fig. 1.11. The system consists of one or more switchboards manned by operators. The subscriber lines are terminated on jacks mounted on the switchboard. There is one jack for every subscriber line. Associated with each jack is a light indicator to draw the attention of the operator. When a subscriber lifts the hand set, the off-hook switch is closed, causing a current to flow through the hand set and the lamp relay coil. The lamp relay operates and the indicator corresponding to the subscriber lights up. The operator establishes contact with the subscriber by connecting the head set to the subscriber line via the headset key and a plug-ended cord pair. A plug-ended cord pair has two cords that are connected internally and terminated with a plug each at the external ends. The plugs mate with the jacks. To establish contact, a cord is plugged into the subscriber jack and the key corresponding to the chosen cord is thrown in position to connect the head set. On being

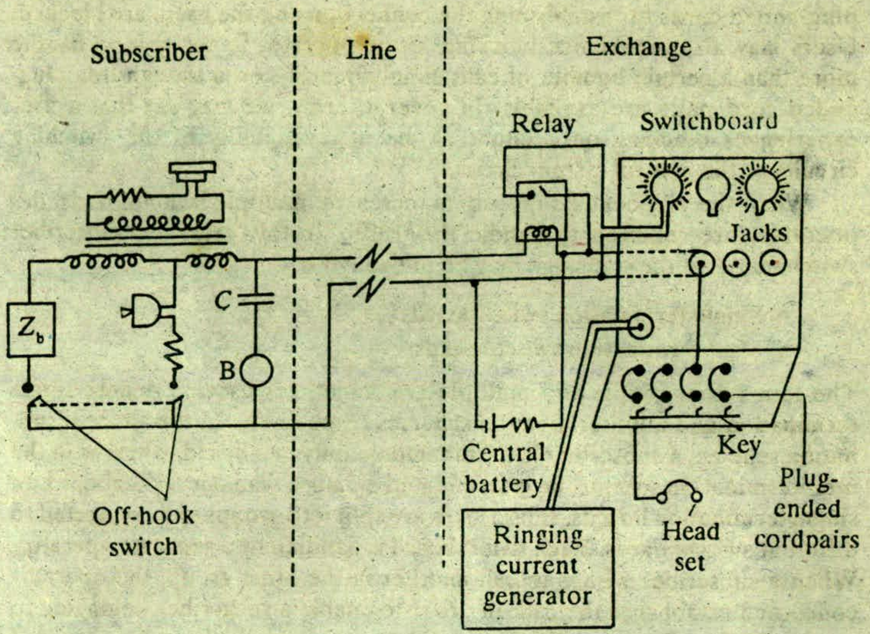


Fig. 1.11 Manned central battery exchange.

told the number required by the subscriber, the operator verifies whether the called party is free, and if so, sends out the ringing current to the called subscriber using a plug-ended cord pair. The ringing circuit at the subscriber end is usually a bell shown as B in Fig. 1.11, with a capacitor C, in series. They remain connected to the circuit always. The capacitor allows the alternating ringing current from the exchange to pass through the bell but prevents the loop direct current. If the called party is busy, the calling subscriber is told about the same. When the called party answers, his indicator lamp lights up. The operator then establishes a connection between the calling and the called party by plugging in the cord pair to the called party jack. In a manual switching system, the operator has full control of a connection. He enables the signalling systems, performs switching and releases a connection after a conversation.

If there are 200 subscribers terminated on a switchboard, there can be a maximum of 100 simultaneous calls. In order to support all these calls, the switchboard must contain 100 plug-ended cord pairs. But a single operator may not be able to handle 100 calls simultaneously. It is, however, rarely that all the subscribers would want to talk simultaneously. Assuming that only 20 subscribers (10 calls) will use the system simultaneously, the switchboard needs to be provided with only 10 plug-ended cord pairs. What happens if more than 20 users want to talk at the same time? The operator will not have plug-ended cords for establishing the connection and the users are blocked. Users may also experience blocking, if the operator is not able to handle more than a certain number of calls simultaneously, even though free plug-ended cord pairs are available. In general terms, we may say that a user experiences blocking on account of the nonavailability of the switching circuits or the control system circuits.

When the number of subscribers increases, multiple switchboards and operators are required to handle the traffic. In this case, the subscriber switchboards at the exchange may be of two types:

- Single termination switchboards
- Multitermination switchboards.

The terms nonmultiple and multiple are sometimes used to denote single termination and multitermination schemes respectively. In the single termination scheme, a subscriber is terminated on only one board, whereas in the multitermination scheme, he is terminated on more than one switchboard. In single termination boards, subscribers are split into groups and connected to different switchboards. Each switchboard is handled by a separate operator. When a subscriber wishes to call another in the same group, the operator concerned establishes the call. In order to enable a subscriber belonging to one group to call a subscriber in another group, transfer lines are provided between the switchboards as shown in Fig. 1.12. The number of transfer lines is determined based on the estimated intergroup traffic. It may be noted that an intergroup call requires the services of two operators manning the two respective groups.

The maximum number of simultaneous calls within a group is limited by the number of plug-ended cord pairs in the group or the number of simulta-

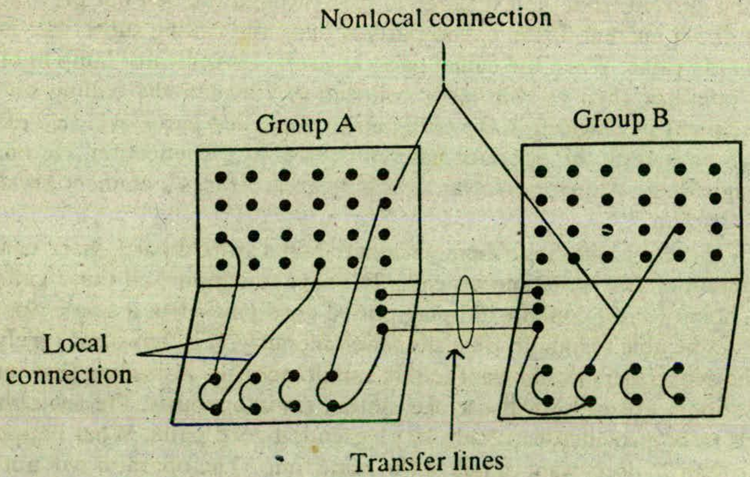


Fig. 1.12 Single termination boards with transfer jacks.

neous calls that can be handled by the operator, whichever is smaller. The number of simultaneous calls between the groups is limited by the number of transfer lines. Single termination systems suffer from the serious disadvantage that as many operators as there are switchboards are always required irrespective of the peak or lean traffic period. During lean traffic period, the average number of simultaneous calls is much less than that during the peak traffic period. Nevertheless, there are likely to be at least a few intergroup calls. Every intergroup call requires two operators to establish the call. Consequently, even a small number of intergroup calls among the switchboards demands that all boards be manned.

The need for two operators per call is avoided in the multitermination switchboard scheme. Here, every subscriber is terminated on all the switchboards as shown in Fig. 1.13. Such an arrangement has the advantage that a

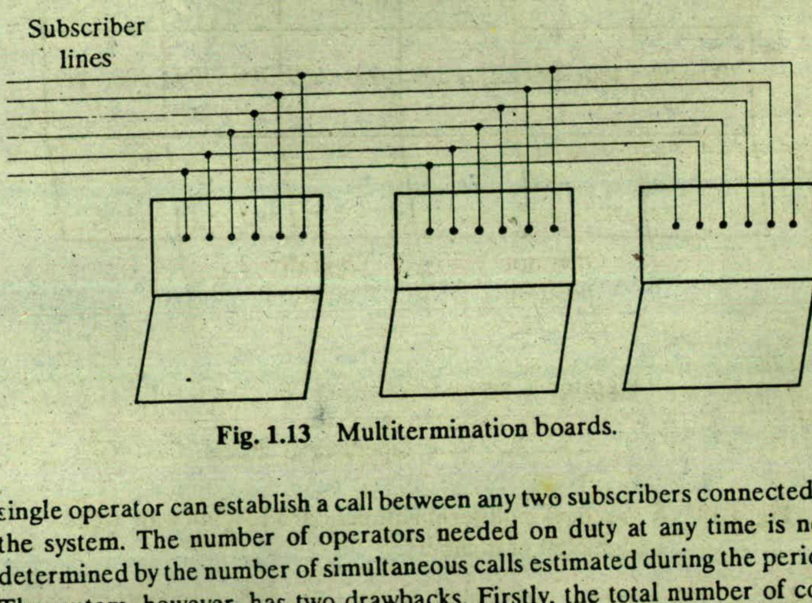


Fig. 1.13 Multitermination boards.

single operator can establish a call between any two subscribers connected to the system. The number of operators needed on duty at any time is now determined by the number of simultaneous calls estimated during the period. The system, however, has two drawbacks. Firstly, the total number of connections in the system increases considerably, thereby reducing the reliability of the system. Secondly, terminating all the subscribers in all the boards, such that the terminations are easily accessible to the operators, poses human engineering problems. The switchboard height becomes large, and the operator does not have easy access to the numbers at the top of the board. The problem is somewhat reduced by terminating half the number of subscribers in alternate switchboards in a cyclic manner and letting an operator have access to one-half of the adjacent boards on the left and right. The scheme is illustrated in Fig.1.14 for two operator positions. Subscribers are terminated on the boards as per the following order:

Subscriber Nos.	Operator board	Right/left Top/bottom
0–99	1	left-top
100–199	1	right-top
200–299	2	left-top
300–399	2	right-top
400–499	1	left-bottom
500–599	1	right-bottom
600–699	2	left-bottom
700–799	2	right-bottom

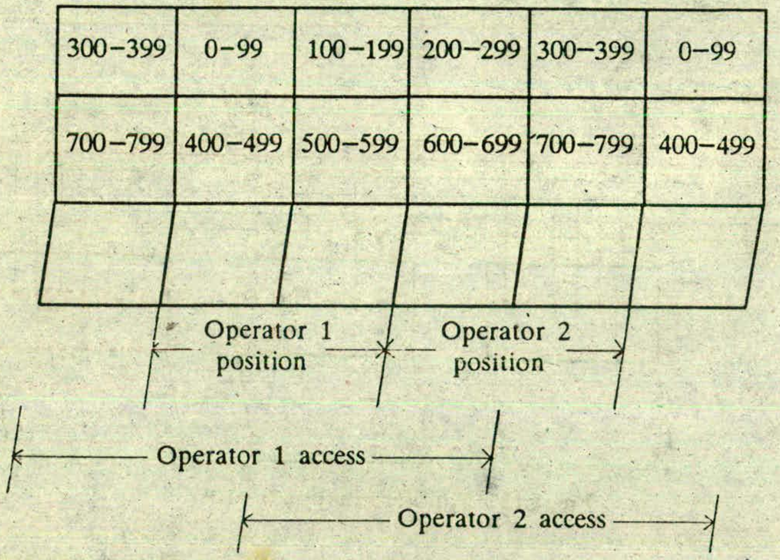


Fig. 1.14 A practical multitermination board scheme with cyclic assignment of numbers.

In addition, the numbers terminated on the right half of operator 2 panel, 300–399 and 700–799, are terminated on the half-panel on the left side of the operator 1. Similarly, the numbers terminated on the left half of operator 1 panel, 0–99 and 400–499, are terminated on the half-panel on the right side of the operator 2. Operator 1 gets access to numbers 300–399 and 700–799 from the left hand side half-panel and to numbers 200–299 and 600–699 from the half-panel of the operator 2. Similarly, operator 2 gets access to numbers 100–199 and 500–599 from the left and to numbers 0–99 and 400–499 from the right. It may be noted that we require one unmanned half-panel at either end of the switchboard row. Thus, in an 8-operator switchboard, there are nine logical switchboards.

As the number of subscribers increases, typically to a thousand or more, manual switching becomes more and more difficult and a method of automatic switching, signalling and control is inevitable.

1.5 Major Telecommunication Networks

Telecommunication networks may be categorised according to their coverage of geographical areas which have distinct telecommunication requirements. Towns and cities have high subscriber densities and relatively heavy traffic per subscriber. They are characterised by many local exchanges and short distances between exchanges as well as subscribers. Networks which are designed taking these factors into account are known as **urban** or **metropolitan** networks. Rural areas are characterised by low subscriber densities, widely dispersed subscribers, lighter traffic per subscriber, just one or two local exchanges usually far apart, long distances between subscribers and exchanges, and less conducive environmental conditions and inadequate infrastructural facilities. These areas are served by **rural networks**. **Long distance** or **toll** or **wide area networks** act as backbone networks interconnecting metropolitan and rural networks. They support intracountry, intercountry and intercontinental communications. Long distances (a few hundred to a few thousand kilometres) involved in such networks call for special consideration in the design of interexchange transmission and signalling systems. In the context of telephone networks, urban and rural networks are sometimes referred to as local networks. But in the context of data networks, the term local network refers to a network within a building or a campus.

The most stupendous telecommunication network in existence is the public switched telephone network (PSTN) or sometimes known as plain old telephone system (POTS). There are over 400 million telephones in the world and the length of wiring in the telephone network is estimated to be over 12 times the distance between the earth and the sun. The growth rate of the telecommunication industry is next only to that of the computer industry. But the growth value in absolute terms far exceeds that of the latter. The present trends seem to indicate that the growth rate may even surpass that of the computer industry in the 90s.

Telecommunication industry is both in the private and government or public sector. It is largely privatised in the United States where there are about 1600 privately owned telephone companies. In most of the other countries, the government has a complete monopoly over all forms of communications including mail, telegraph and telephone. Companies in the United States that provide communication services to the public are known as **common carriers**. In countries where the telecommunication authority is a nationalised company or a department of the government, it is usually known as the **post, telegraph and telephone (PTT)** administration. In India, a single government department known as Posts and Telegraphs (P&T) department dealt with mail, telegraph and telephone communications until the end of 1984. With effect from January 1985, the responsibility was divided between two departments: the **Department of Posts** dealing with mail and the **Department of Telecommunications (DOT)** dealing with telephone, telegraph and data communications.

With a number of different agencies involved in providing telecommunication services, there is clearly a need to ensure compatibility on a worldwide scale to enable internetworking. This coordination is provided by the International Telecommunications Union (ITU), an agency of the United Nations. ITU has three main groups, one of which deals with telephone and data communications. This group is known by the French name: Comité Consultatif Internationale de Télégraphique et Téléphonique (CCITT), i.e. International Consultative Committee for Telegraph and Telephones. Five classes of members, A, B, C, D, and E constitute the CCITT. A description of the members belonging to different classes is presented in Table 1.1. Only class A members have voting rights. Since there is no PTT in the United States, the State Department represents the U.S. government as class A member. In CCITT terminology, a public telephone network is referred to as general switched telephone network (GSTN). In this text, we mostly use the popular term, PSTN; but when discussing standards, we use the CCITT term, GSTN.

Table 1.1 CCITT Members

Class	Members
A	National PTTs
B	Recognised private administrations
C	Scientific and industrial organisations
D	Other international organisations
E	Organisations whose primary function is in fields other than telecommunications, but have an interest in CCITT activities.

Data networks form the second major class of telecommunication networks. They are of recent origin (30–35 years old) and have emerged as a result of coming together of computer and communication technologies. These networks enable sharing of hardware, software and data resources of computer systems. As a result, they have come to be known as computer networks. Early networks interconnected computer systems of the same family. The network project, ARPANET, supported by the Advanced Research Projects Agency of the Department of Defence, United States, is one of the pioneering efforts in interconnecting heterogeneous systems. In some sense, the ARPANET demonstrated the feasibility of data/computer networks. TYMNET is another large scale, general purpose data network introduced in 1970, interconnecting geographically distributed computer systems, users and peripherals.

Prompted by the developments in ARPANET and TYMNET, leading computer vendors announced proprietary architectures for interconnecting their own computer systems. These include IBM's System Network Architecture (SNA) and Digital Equipment Corporation's Digital Network Architecture (DNA). By 1980, the enormous value of computer communication networks became obvious, particularly among research communities and special interest user groups. The developers of Unix operating system quickly realised the advantages of networking and wrote a simple program called **uucp** (Unix-to-Unix Copy) for exchanging files and electronic mail among Unix machines. Two networks, UUNET (Unix Users' Network) and USENET have evolved based on **uucp** communication program. In addition to file transfer and electronic mail, USENET supports a service called **netnews**, which essentially provides a bulletin board on which any user may post a notice or news item to be seen by all other users of the network.

In the academic circle, the computer science community in the United States setup a network with the help of the National Science Foundation to serve all the computer science departments of the U.S. Universities. This network, known as CSNET, uses the transmission facilities of other networks but provides a uniform user level interface. Another academic network started by the City University of New York and Yale University, known as BITNET (Because It's Time Network) aims to serve all the departments of the universities. BITNET has now spread to a large number of sites and spans North America, Europe, Japan and Australia. In Europe, it is called European Academic Research Network (EARN). In the United Kingdom, there is a separate network known as Joint Academic Network (JANET) covering most of the universities and research laboratories.

As in the case of telephone networks, it is apparent that various agencies are involved in setting up and operating data networks and that there is a need for worldwide standards to enable data networks to interwork. Apart from CCITT, significant contributions to data network standards have come from the International Standards Organisation (ISO) which is a voluntary, nontreaty organisation. National standards organisations, like American National Standards Institute (ANSI), British Standards Institution (BSI), Association Française de Normalisation (AFNOR), Deutsches Institut für Normalische (DIN) and Bureau of Indian Standards (BIS) are members of ISO. The Institute of Electrical and Electronic Engineers (IEEE), the largest professional organisation in the world also plays a major role in evolving data network standards. Figure 1.15 presents the organisational structure of the different agencies involved in the coordination of telecommunication network activities.

Integrated services digital network (ISDN) is now emerging as a major telecommunication network. ISDN is envisaged as a single common net-

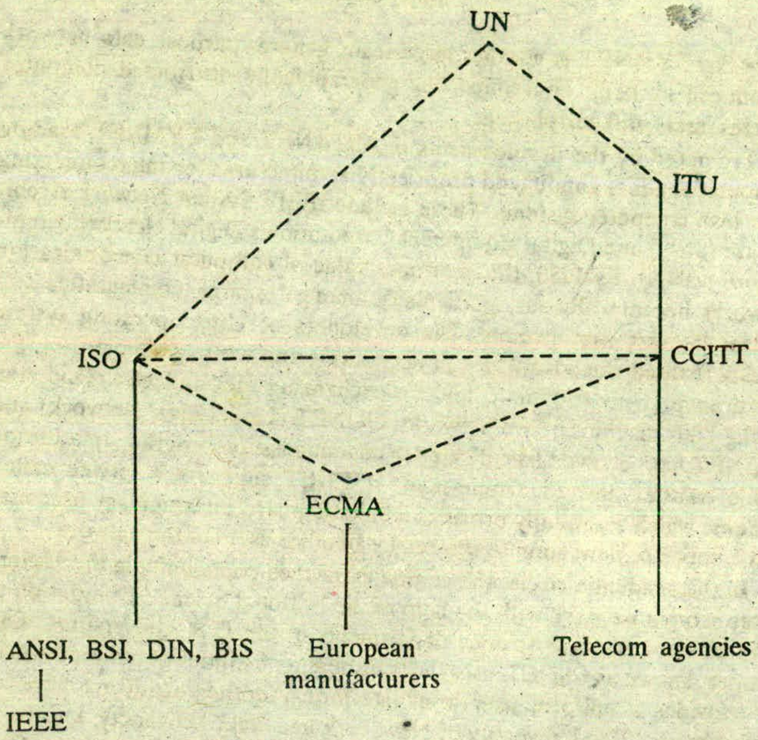


Fig. 1.15 Organisational structure of telecommunication coordination agencies.

work capable of carrying multimedia services like voice, data, video and facsimile. The key to ISDN is the digitalisation of services, transmission, switching and signalling. The digital domain acts as a common substratum for all current and future services. Once digitised, all signals, voice or nonvoice, look alike and a single digital network with adequate speed and signalling capabilities can support a wide range of services. However, meeting a variety of speed and signalling requirements is not easy. This is where the computers come in handy and the ISDN uses them extensively.

Recognising that the telephone network is the primary and extensive international communication infrastructure available today, ISDN is conceived to be a redesign of the existing telephone network to provide end-to-end digital connectivity. Obviously, the redesign cannot take place overnight and the ISDN will have to take an evolutionary path. At present, digital connectivity has been extended to user premises only in some parts of a few countries. Thus, ISDN will have to coexist with the present analog telephone network for some years to come. All these imply that the standards for ISDN must emerge well before its implementation. This aspect has been recognised by CCITT and the first set of key ISDN recommendations were approved in 1984 and further refined in 1988. Unlike telephone and data networks, one may expect a fairly organised and structured growth in the case of ISDN, with CCITT spearheading the coordination.

Perhaps, ISDN is the single most important example of the contribution of computer technology to telecommunications and it may become the most important development as a result of the 'communion' of the computer and communication technologies. The large scale use of computers in ISDN is leading to the concept of **intelligent networks** which are preprogrammed to be adaptive, algorithmic, resourceful, responsive and intelligent. As an example of the possible capabilities of such intelligent networks, one may cite real time machine translation. A telephone conversation originating in Japanese may be heard in English at the receiver end and vice versa. A telex sent in Hindi in Delhi may be delivered in Kannada in Bangalore. Such examples, although somewhat far fetched now, may become a reality in the 21st century.

Telecommunication networks have been evolving in the last 150 years and would continue to evolve to provide wider services in a more convenient form in the coming century. The emerging information society depends heavily on the developments in the field of telecommunications. It is estimated that millions of dollars will be invested by many countries during the next two decades or so in improving the telecommunication facilities. We seem to be entering an era of 'sophisticated' telecommunications.

FURTHER READING

1. Atkinson, J., *Telephony*, Vol. 1, *General Principles and Manual Exchange Systems*, The New Era Publishing Co., London, 1947.
2. Department of Telecommunications, *Forty Years of Telecommunications in Independent India*, New Delhi, 1988.
3. Economic Commission for Europe, The United Nations, *The Telecommunication Industry: Growth and Structural Change*, Paris, 1987.
4. Meyers, R.A. (ed.), *Telecommunications: Encyclopedia of Telecommunications*, Academic Press, San Diego, 1989, pp. 455–476.
5. Schindler, Jr., G.E. (ed.), *A History of Engineering and Science in the Bell System: Switching Technology (1925–1975)*, Bell Telephone Laboratories, 1982.

EXERCISES

1. A fully connected network supports full duplex communication using unidirectional links. Show that the total number of links in such a network with n nodes, is given by $2 \times {}^{n}C_2$.

2. How are switching systems classified? In what way is stored progran control superior to hard-wired control?

3. Estimate the bandwidth requirements of a single satellite that is tc support 20 million telephone conversations simultaneously.

4. An electrical communication system uses a channel that has 20 dB loss. Estimate the received power, if the transmitted power is one watt.

5. If the signal input to an amplifier is 0 dBm, what is the power output in mW if the gain of the amplifier is 20 dB?

6. The channel interfaces in a point-to-point communication system attenuate the signal by 3 dB each. The channel has a loss of 30 dB. If the received signal is to be amplified such that the overall loss is limited to 20 dB, estimate the amplifier gain.

7. If the noise power in a channel is 0.1 dBm and the signal power is 10 mW, what is the (S/N) ratio?

8. What is the significance of (S/N) ratio being –3 dB?

9. For a carbon granule microphone, determine a suitable value for m, if the contribution from each of the higher order terms is to be less than $0.01 I_0$.

10. What is the importance of a steady current flowing through a carbon microphone? Is the harmonic distortion affected by a change in the energising current?

11. Why is it necessary to keep the magnetic diaphragm in an earphone displaced from its unstressed position? How is this achieved?

12. What happens if the ratio ϕ/ϕ_0 is not very small in the case of an earphone?

13. What is the significance of sidetone in a telephone conversation?

14. In the circuit of Fig.1.8, it is desired that 10 per cent of the microphone signal is heard as sidetone. If the number of turns in the coil P is 200, determine the number of turns in the coil Q and the secondary winding in the earphone circuit. Assume that Z_b is exactly matched to the line impedance on the exchange side.

15. In a 100-line folded network, how many switching elements are required for nonblocking operation?

16. A 1000–line exchange is partly folded and partly nonfolded. Forty per cent of the subscribers are active during peak hour. If the ratio of local to external traffic is 4:1, estimate the number of trunk lines required.

17. A central battery exchange is powered with a 48 V battery. The carbon microphone requires a minimum of 24 mA as energising current. The battery has a 400 Ω resistance in series for short circuit protection. The d.c. resistance of the microphone is 50 Ω. If the cable used for subscriber lines offers a resistance of 50 Ω/km, determine the maximum distance at which a subscriber station can be located.

18. A manual switchboard system needs to support 900 subscribers, numbered 100 – 999. Average peak hour traffic is 250 calls, 130 of which are within the number range 400 – 699, 20 of them are between this range and other range of numbers and the remaining are uniformly distributed in the other number ranges. The average lean traffic is 60 calls, of which no call is originated/destined from/to the number range 400 – 699 but uniformly distributed otherwise. An operator is capable of handling 30 simultaneous calls. Suggest a suitable manual switchboard system design that minimises the total number of terminations at the switchboards and employ the minimum number of operators. Estimate the number of terminations in your design.

2

Strowger Switching Systems

Strowger switching system was the first automatic switching system developed by Almon B. Strowger in 1889. The story goes that Strowger was an undertaker whose business seemed to have suffered on account of a telephone operator in a manual exchange. When subscribers rang up the operator and asked for an undertaker, she always connected them to her own husband who was also an undertaker and a competitor to A.B. Strowger. Annoyed at the amount of business he was losing this way, Strowger decided to make a switching system that would replace the human operator. The switch developed by him is named after him. Functionally, the system is classified as step-by-step switching system as the connections are established in a step-by-step manner.

Automatic switching systems have a number of advantages over the manual exchanges. A few important ones are:

- In a manual exchange, the subscriber needs to communicate with the operator and a common language becomes an important factor. In multilingual areas this aspect may pose problems. On the other hand, the operation of an automatic exchange is language independent.
- A greater degree of privacy is obtained in automatic exchanges as no operator is normally involved in setting up and monitoring a call.
- Establishment and release of calls are faster in automatic exchanges. It is not unusual in a manual exchange, for an operator to take quite a few minutes to notice the end of a conversation and release the circuits. This could be very annoying particularly to the business subscribers who may like to make a number of calls in quick succession.
- In an automatic exchange, the time required to establish and release a call remains more or less of the same order irrespective of the load on the system or the time of the day. In a manual system, this may not be true.

2.1 Rotary Dial Telephone

In manual exchanges, a calling subscriber may communicate the identity of the called subscriber in a natural and informal language to the operator. For example, a called subscriber may be identified by his name or profession or designation. In an automatic exchange, informal communication is not possible and a formal numbering plan or addressing scheme is required to identify the subscribers. Numbering plan, in which a subscriber is identified by a number, is more widely used than addressing scheme in which a subscriber is identified by alphanumeric strings. A mechanism to transmit the identity of the called subscriber to the exchange is now required at the telephone set. Two methods are prevalent for this purpose:

- Pulse dialling
- Multifrequency dialling.

Multifrequency dialling is discussed in Section 3.2. Pulse dialling originated in 1895 and is used extensively even today. In this form of dialling, a train of pulses is used to represent a digit in the subscriber number. The number of pulses in a train is equal to the digit value it represents except in the case of zero which is represented by 10 pulses. Successive digits in a number are represented by a series of pulse trains. Two successive trains are distinguished from one another by a pause in between them, known as the **interdigit gap**. The pulses are generated by alternately breaking and making the loop circuit between the subscriber and the exchange. The pulsing pattern is shown in Fig. 2.1 for digits three and two. The pulse rate is usually 10

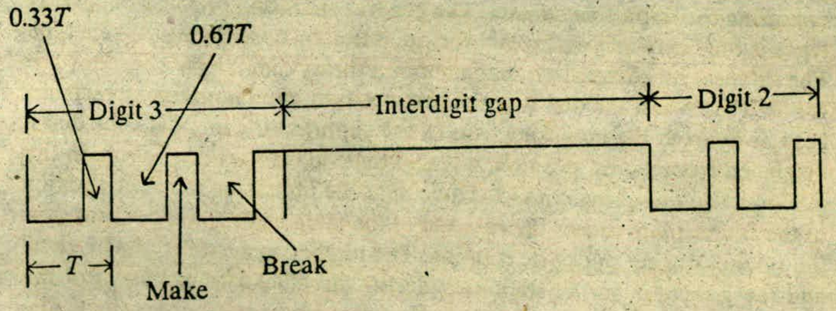


Fig. 2.1 Pulse dialling.

pulses per second with a 10 per cent tolerance. The interdigit gap is at least 200 ms although in some designs the minimum gap requirement may be as much as 400 – 500 ms. In some modern electronic and crossbar exchanges, there exists an upper limit for the interdigit gap (see Section 4.1). The duty ratio of the pulse is 33 per cent nominally.

In introducing dial pulsing mechanism in the telephone set, the following points have to be considered:

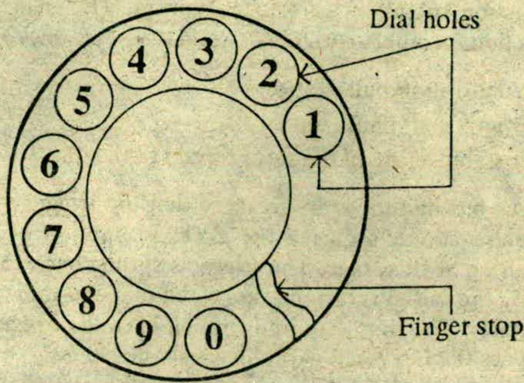
1. Since the pulses are produced by make and break of the subscriber loop, there is likelihood of sparking inside the telephone instrument.
2. The transmitter, receiver and the bell circuits of the telephone set may be damaged if the dialling pulses are passed through them.
3. The dialling habits of the users vary widely and hence all timing aspects should be independent of user action.

A rotary dial telephone uses the following for implementing pulse dialling:

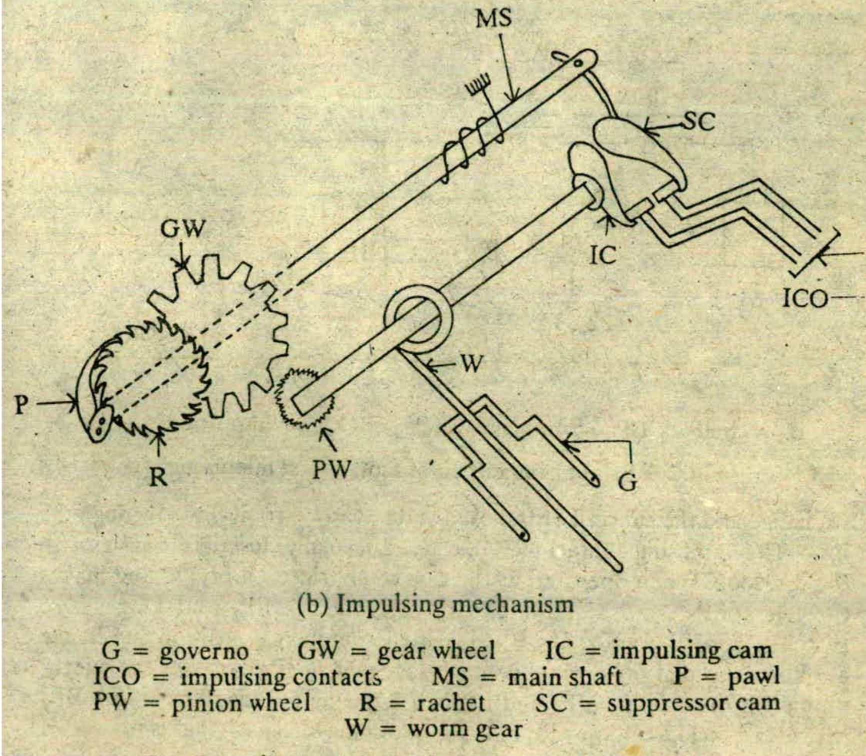
- Finger plate and spring
- Shaft, gear and pinion wheel
- Pawl and ratchet mechanism
- Impulsing cam and suppressor cam or a trigger mechanism
- Impulsing contact
- Centrifugal governor and worm gear
- Transmitter, receiver and bell by-pass circuits.

The arrangement of the finger plate is shown in Fig. 2.2(a). The dial is operated by placing a finger in the hole appropriate to the digit to be dialled, drawing the finger plate round in the clockwise direction to the finger stop position and letting the dial free by withdrawing the finger. The finger plate and the associated mechanism now return to the rest position under the influence of a spring. The dial pulses are produced during the return travel of the finger plate, thus eliminating the human element in pulse timings.

A rotary dial telephone is classified either as cam type or trigger type depending on whether a cam mechanism or a trigger mechanism is used for operating the impulsing contact. The general operating principle of both the types is the same and we explain the operation by considering the cam type. The internal mechanical arrangement of a rotary dial telephone is shown in Fig. 2.2(b). When the dial is in the rest position, the impulsing contacts are kept away from the impulsing cam by the suppressor cam. When the dial is displaced from its rest position, it is said to be in **off-normal** position. In this position, the impulsing contacts come near the impulsing cam. The rotation of the finger plate causes the rotation of the main shaft. The pawl slips over the ratchet during clockwise rotation. The ratchet, gear wheel, pinion wheel and the governor are all stationary during the clockwise movement of the dial. When the dial returns, the pawl engages and rotates the ratchet. The gear wheel, pinion wheel and the governor all rotate. The governor helps to maintain a uniform speed of rotation. The impulsing cam which is attached to a pinion shaft now breaks and makes the impulsing contacts which in turn causes the pulses in the circuit. The shape of the impulsing cam is such that the break and make periods are in the ratio of 2:1. When the dial is about to reach the rest position, the suppressor cam moves the impulsing contacts away from the impulsing cam. This action provides the required interdigit gap timing independent of the pause that may occur between two successive digits, due to human dialling habit. Suppressor cam may also be designed



(a) Finger plate arrangement



(b) Impulsing mechanism

G = governo GW = gear wheel IC = impulsing cam
ICO = impulsing contacts MS = main shaft P = pawl
PW = pinion wheel R = rachet SC = suppressor cam
W = worm gear

Fig. 2.2 Rotary dial telephone parts and mechanism.

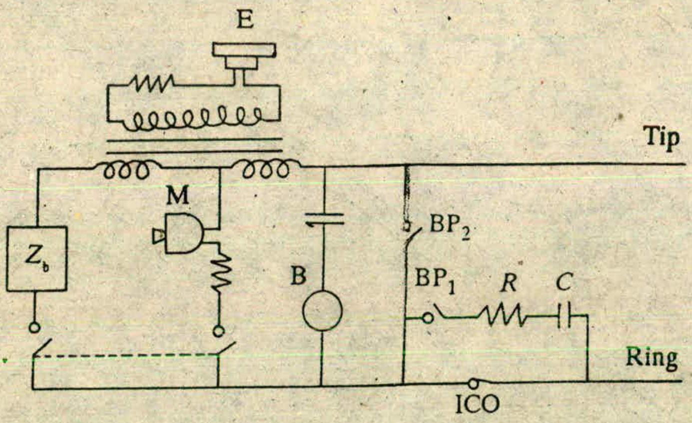
such that the interdigit pause is provided prior to the commencement of the first pulse of a digit.

The trigger dial is an improvement over the cam dial. The precision of operation in the cam dial is affected by the wear and tear of the cam elements and other friction members in the mechanism. The trigger dial design eliminates friction members and helps to achieve

- more uniform impulse ratio,
- larger interdigit pause, and
- better stabilisation of the return speed of the dial.

The trigger mechanism is so arranged that the trigger is sprung away from the impulse contacts during the clockwise motion of the dial, thus preventing pulsing at this stage. The trigger is sprung back to the operative position during the initial return motion of the dial and thereafter operates the pulse contacts. The time required to bring back the trigger to operative position provides the interdigit gap which is about 240 ms.

The impulsing circuit of the rotary dial telephone is shown in Fig. 2.3. When the subscriber lifts his handset (off-hook), the d.c. loop between the



B = bell BP = by-pass switch ICO = impulsing contact

Fig. 2.3 Impulsing circuit of a rotary dial telephone.

exchange and the subscriber is closed and a steady current flows through the loop. The impulsing contact (ICO), which is normally closed, is in series with the d.c. loop. When operated by the cam or the trigger, it breaks and makes the circuit. Figure 2.3 shows two by-pass switches BP_1 and BP_2. These switches close as soon as the dial is moved from its rest position and hence are known as dial-off-normal contacts. The switch BP_2 bypasses the microphone M, the earphone E, and the bell B, during pulsing. The switch BP_1 provides a local *RC* loop with ICO for quenching the spark that is produced when the circuit is broken. In the absence of BP_1, the sparking voltage developed across ICO may affect adversely the other circuits in the telephone set. Once the dialling is complete, the dial is in the rest position, BP_1 and BP_2 are open, and the impulsing contact is closed. Thus the transmitter and the receiver are ready for speech conversation. The two wires connecting the telephone to the exchange are known as **ring and tip**. The central battery voltage of -48 V is connected through a relay to the ring lead and the tip lead is grounded. Ring lead is used to receive signals from the far end and the tip lead is used to transmit the signal.

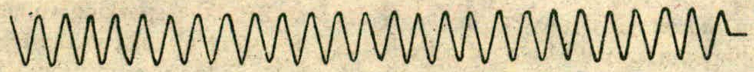
2.2 Signalling Tones

As discussed in Section 1.4, a number of signalling functions are involved in establishing, maintaining and releasing a telephone conversation. These functions are performed by an operator in a manual exchange. In automatic switching systems, the verbal signalling of the operator is replaced by a series of distinctive tones. Five subscriber related signalling functions are performed by the operator:

1. Respond to the calling subscriber to obtain the identification of the called party.
2. Inform the calling subscriber that the call is being established.
3. Ring the bell of the called party.
4. Inform the calling subscriber, if the called party is busy.
5. Inform the calling subscriber, if the called party line is unobtainable for some reason.

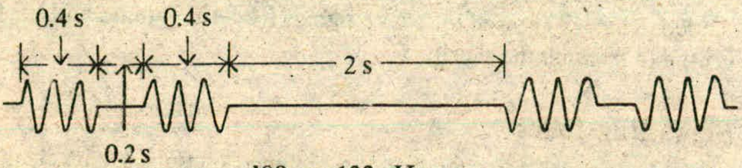
Distinctive signalling tones are provided in all automatic switching systems for functions 1, 3, 4 and 5. A signalling tone for function 2 is usually not available in Strowger exchanges. However, most of the modern exchanges provide a call-in-progress or routing tone for function 2. Although attempts have been made to standardise the tones for various signals, many variations are in vogue in different parts of the world and even in different parts of the same country. Variations are mainly due to different capabilities and technologies of the switching systems used.

The signalling function 1 above is fulfilled by sending a **dial tone** to the calling subscriber. This tone indicates that the exchange is ready to accept dialled digits from the subscriber. The subscriber should start dialling only after hearing the dial tone. Otherwise, initial dial pulses may be missed by the exchange which may result in the call landing on a wrong number. Most often, the dial tone is sent out by the exchange even before the handset is brought near the ear. Sometimes, however, a few seconds may elapse before the dial tone is heard. This happens particularly in common control exchanges which use shared resources for user interfaces. The dial tone is a



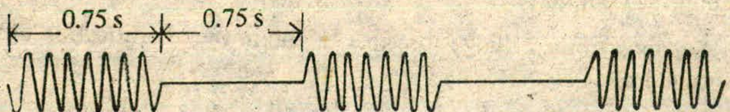
33 or 50 or 400 Hz continuous

(a) Dial tone



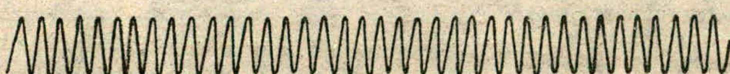
400 or 133 Hz tone

(b) Ringing tone



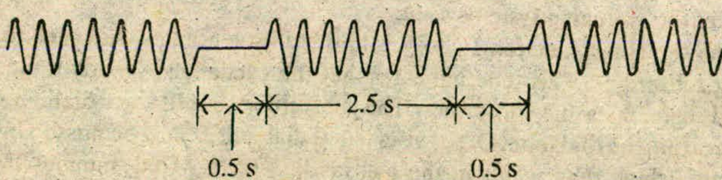
400 Hz

(c) Busy tone



400 Hz continuous

(d) Number unobtainable tone



400 or 800 Hz

(e) Call-in-progress tone

Fig. 2.4 Signalling tones in automatic exchanges.

33 Hz or 50 Hz or 400 Hz continuous tone as shown in Fig. 2.4(a). The 400 Hz signal is usually modulated with 25 Hz or 50 Hz.

When the called party line is obtained, the exchange control equipment sends out the ringing current to the telephone set of the called party. This ringing current has the familiar double-ring pattern. Simultaneously, the control equipment sends out a **ringing tone** to the calling subscriber, which has a pattern similar to that of the ringing current as shown in Fig. 2.4(b). The two rings in the double-ring pattern are separated by a time gap of 0.2 s and two double-ring patterns by a gap of 2 s. The ring burst has a duration of 0.4 s. The frequency of the ringing tone is 133 Hz or 400 Hz, sometimes modulated with 25 Hz or 33 Hz. It may be noted that the ringing current and the ringing tone are two independent quantities. This explains one of the common fault symptoms where a calling subscriber hears the ringing tone whereas no ring is heard at the called subscriber end.

Busy tone pattern is shown in Fig. 2.4(c). It is a bursty 400 Hz signal with silence period in between. The burst and silence durations have the same value of 0.75 s or 0.375 s. A busy tone is sent to the calling subscriber whenever the switching equipment or junction line is not available to put through the call or the called subscriber line is engaged. No distinction is made between these conditions. It is not possible for a calling subscriber to conclude on the basis of the busy tone that the called party was actually engaged in a conversation. While it is technically feasible to introduce different busy tones for different conditions, this would only, perhaps, confuse the subscriber, and not serve any useful purpose.

Figure 2.4(d) shows the **number unobtainable tone** which is a continuous 400 Hz signal. This tone may be sent to the calling subscriber due to a variety of reasons such as the called party line is out of order or disconnected, and an error in dialling leading to the selection of a spare line. In some exchanges the number unobtainable tone is 400 Hz intermittent with 2.5 s *on period* and 0.5 s *off period*.

The **routing tone** or **call-in-progress tone** is a 400 Hz or 800 Hz intermittent pattern. In electromechanical systems, it is usually 800 Hz with 50 per cent duty ratio and 0.5 s *on/off* period. In analog electronic exchanges it is a 400 Hz pattern with 0.5 s *on* period and 2.5 s *off* period. In digital exchanges, it has 0.1 s *on/off* periods at 400 Hz. When a subscriber call is routed through a number of different types of exchanges, one hears different call-in-progress tones as the call progresses through different exchanges. Figure 2.4(e) shows a routing tone pattern.

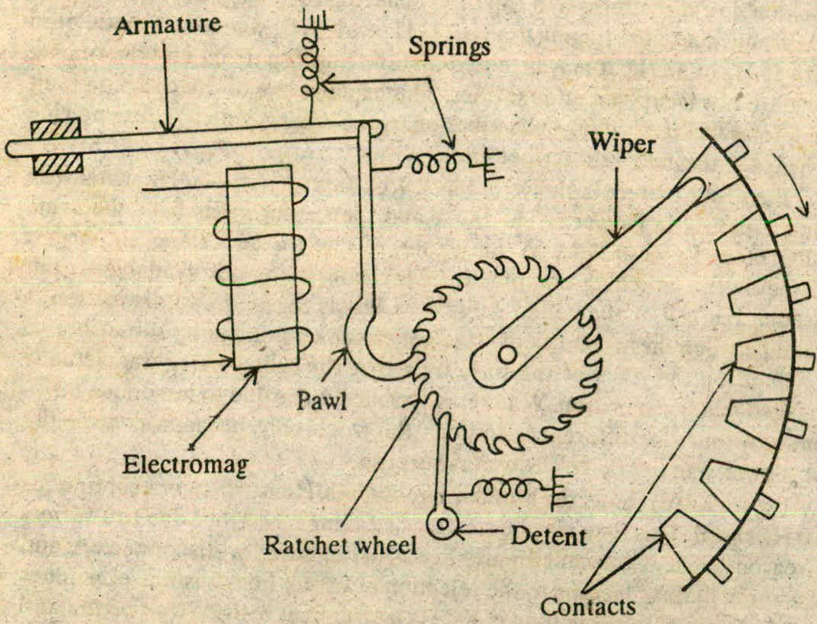
Regular users of telephone in a particular area have little difficulty in recognising signalling tones. It is not unusual that a subscriber in a new area where frequencies or timings of the tones are different from those in his own area, confuses signalling tones. In order to overcome this problem, recorded voices that announce messages like "number engaged" or "busy" are used in some modern exchanges. Voice announcement, however, poses problems in multilingual areas.

2.3 Strowger Switching Components

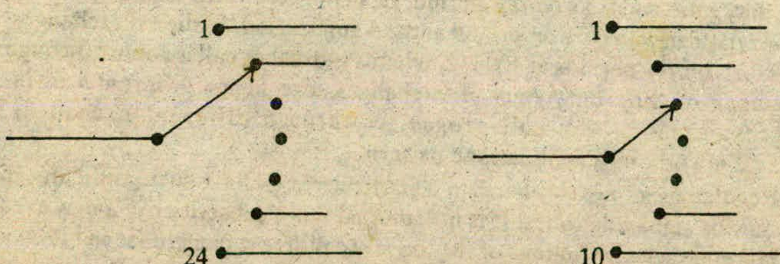
In the Strowger system, there are two types of selectors which form the building blocks for the switching system:

- Uniselector
- Two-motion selector.

These selectors are constructed using electromechanical rotary switches. The drive mechanism of a rotary switch is shown in Fig. 2.5(a).



(a) Drive mechanism of a rotary switch



(b) Schematic representation of uniselectors

Fig. 2.5 Uniselectors.

Whenever the electromagnet is energised, the armature is attracted to it and the pawl falls one position below the present tooth position. The ratchet wheel, however, does not move and is held in position by the detent. When the electromagnet is de-energised, the armature is released and returns to its rest position due to the restoring action of the spring. During this reverse motion of the armature, the pawl moves the ratchet wheel one position up where it is held in position by the detent. The clearance between the armature and the electromagnet is such that during the forward movement of the armature the pawl slips over the ratchet exactly by one position. As the ratchet wheel rotates up by one position, the wiper moves across one contact position in the direction indicated. Thus, if the electromagnet is energised and de-energised five times by applying five pulses, the wiper moves by five contacts. The mechanism shown in Fig. 2.5(a) is known as reverse drive type as the ratchet wheel moves when the armature returns to its rest position. It is possible to arrange the mechanism in such a way that the wheel moves during the forward motion of the armature in which case it is known as forward drive type. Reverse drive type is generally used in uniselectors and the forward drive type in two-motion selectors.

A uniselector is one which has a single rotary switch with a bank of contacts. Typically, there are four banks of which three are used for switching and the fourth one is used for homing. The three switching banks have 25 or 11 contacts each. The first contact in each bank is known as the home contact and the remaining as switching contacts. The homing bank has only two contacts: one at the first position corresponding to the home contacts of the other banks and the other extending as an arc from the second position to the last position. This arc contact is often referred to as the homing arc. Depending upon the number of switching contacts, uniselectors are identified as 10-outlet or 24-outlet uniselectors. Figure 2.5(b) shows a schematic representation of uniselectors.

The wipers associated with the banks of a uniselector, one for each bank, are rigidly mounted to a wiper assembly which moves whenever the ratchet wheel rotates. As a consequence, all the wipers move simultaneously and there is no relative motion amongst them. All wipers lie in the same vertical plane such that each wiper touches the same corresponding bank contact at any instant. There is an **interrupt contact** associated with the uniselector. This contact remains closed normally and opens whenever the armature is close to the end of its forward movement. It breaks the armature energising circuit to enable the armature to return to its rest position. It may be noted that if the drive circuit is permanently energised, the selector will step continuously owing to the constant breaking and making of the interrupter contact.

The proper functioning of a uniselector is dependent on a number of factors:

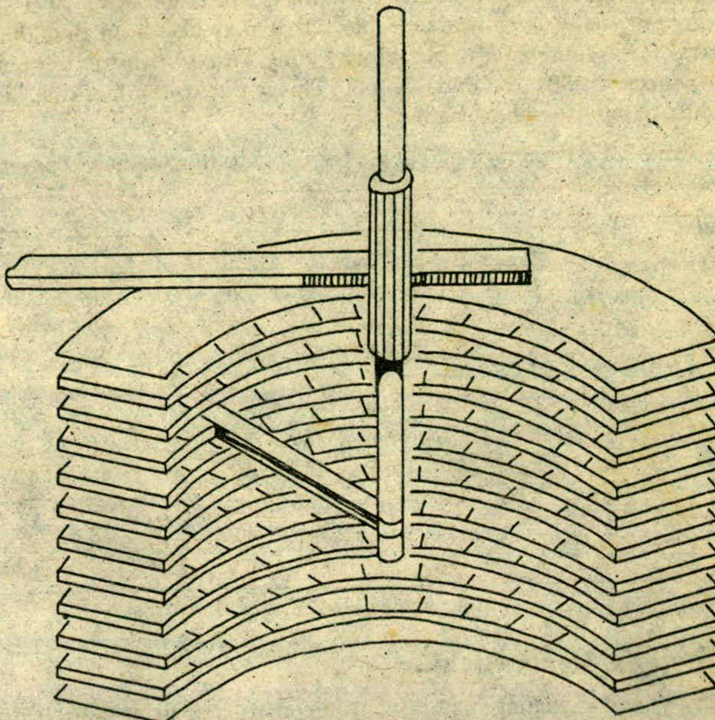
- Energising current level
- Inertia of the moving system

- Friction between wipers and bank contacts
- Friction in drive assembly
- Tension in restoring springs
- Adjustment of interrupter contacts.

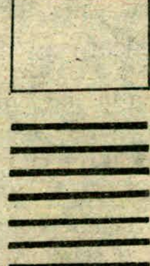
To illustrate the importance of these factors for proper functioning, let us consider the adjustment of interrupter contacts as an example. The interrupter contacts must be adjusted so that they open and close at the correct instants in the stroke of the armature. If they open too soon, the armature may fail to complete its stroke and the pawl may not engage the next ratchet tooth. On the other hand, if they close too soon during the return of the armature, the reverse movement is affected and the stepping of the wiper assembly becomes uncertain. Wear and tear of the selector parts affect the proper functioning adversely and as a result, the selectors require frequent attention for maintenance.

A two-motion selector is capable of horizontal as well as vertical stepping movement. It has two rotary switches, one providing vertical motion for the wiper assembly, and the other providing horizontal movement for the wipers. Many authors use the term 'rotary switch' to mean the switch that causes horizontal movements of the wipers. In this text, the term rotary switch is used in a generic sense to imply a pawl and ratchet arrangement irrespective of whether such an arrangement is being used to cause vertical or horizontal otion. The horizontal movement rotary switch of a two-motion selector has an interrupter contact as in the case of uniselector. Normally, there are 11 vertical positions and 11 horizontal contacts in each vertical position. The lowest vertical position and the first horizontal contact in each vertical level are home positions, and the remaining ones are the actual switching positions. Thus, the wiper in a two-motion selector has access to 100 switching contacts. Access to any particular contact is obtained by moving the wiper assembly vertically to the required level and then rotating the wipers to the desired contact at that level. The arrangement is shown in Fig. 2.6 (a). At each level there are three or four banks of contacts. Depending upon the number of banks, a two-motion selector is sometimes known as a 330-point or 440-point selector. For homing the wiper assembly, it is driven beyond the 11th contact position by the horizontal rotary magnet and its interrupter contact. The wiper assembly then falls vertically to the home level and returns to the horizontal home position under the influence of a restoring spring. In some designs, a third magnet, known as release magnet is used for homing. A set of off-normal contacts are operated by the first vertical and horizontal movements of the wipers and they remain operated until the wiper assembly returns to home position. Figure 2.6(b) shows a schematic representation of a two-motion selector.

The vertical and horizontal motions in a two-motion selector may be effected directly by using two impulse trains from subscriber dialling. The first impulse train corresponding to the first digit operates the vertical mag-



(a) Two-motion selector arrangement



(b) Schematic representation

Fig. 2.6 Two-motion selectors.

net and the second impulse train drives the horizontal rotary switch. In such a case, it follows that the bank contacts are so numbered as to correspond to the digits necessary to reach each contact. The numbering of a standard 100-contact bank is shown in Table 2.1. It may be noted that the lowest vertical level commences with 11 and ends with 10, whilst the tenth level commences with 01 and ends with 00. This is due to the fact that digit zero produces 10 pulses when dialled.

Table 2.1 Numbering Scheme for Two-Motion Selector Contacts

Level	Contacts									
	1	2	3	4	5	6	7	8	9	10
10	01	02	03	04	05	06	07	08	09	00
9	91	92	93	94	95	96	97	98	99	90
8	81	82	83	84	85	86	87	88	89	80
7	71	72	73	74	75	76	77	78	79	70
6	61	62	63	64	65	66	67	68	69	60
5	51	52	53	54	55	56	57	58	59	50
4	41	42	43	44	45	46	47	48	49	40
3	31	32	33	34	35	36	37	38	39	30
2	21	22	23	24	25	26	27	28	29	20
1	11	12	13	14	15	16	17	18	19	10

2.4 Step-by-Step Switching

A step-by-step switching system may be constructed using uniselectors or two-motion selectors or a combination of both. The wiper contacts of these selectors move in direct response to dial pulses or other signals like off-hook from the subscriber telephone. The wiper steps forward by one contact at a time and moves by as many contacts (takes as many steps) as the number of dial pulses received or as required to satisfy certain signalling conditions. Hence the name "step-by-step switching" is given to this method. Most of the necessary control circuits are built in as an integral part of the selectors, thus enabling them to receive and respond to user signalling directly. The relevant signalling tones are sent out to the subscriber by the switching elements (selectors) at the appropriate stages of switching. Thus, a step-by-step switching system is a **direct control** system.

A step-by-step switching system has three major parts as shown in Fig. 2.7. The line equipment part consists of selector hunters or line finders and the other two parts consist of selectors. The selector hunters and line finders represent two fundamental ways in which a subscriber gains access to common switching resources. As the name implies, a selector hunter searches and seizes a selector from the switching matrix part. There is one selector hunter for each subscriber. Usually, 24-outlet uniselectors are used as selector hunters. The selector hunter scheme is sometimes called subscriber uniselector scheme as there is a dedicated uniselector for each subscriber in the system. Line finders are associated with the first set of

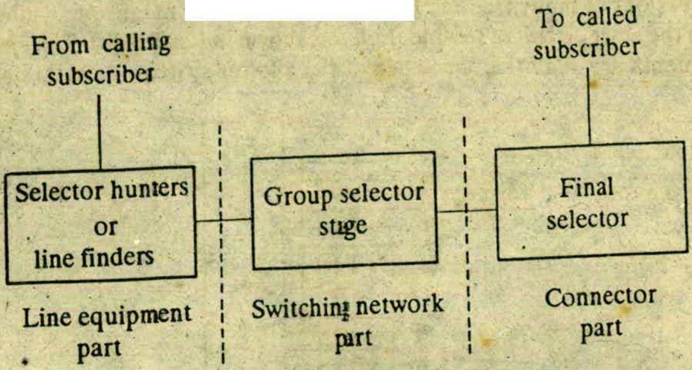


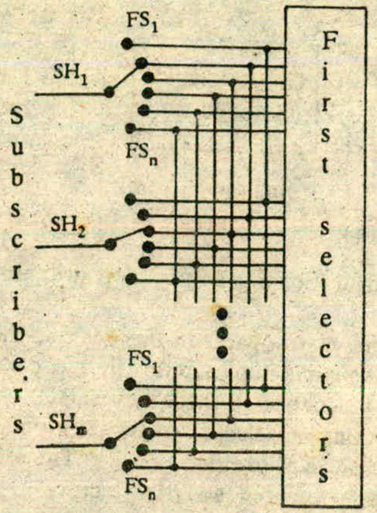
Fig. 2.7 Configuration of a step-by-step switching system.

selectors in the switching matrix part and there is one line finder for each selector in the set. As the name implies, a line finder searches and finds the line of a subscriber to be connected to the first selector associated with it. Line finders are built using uniselectors or two-motion selectors. The line equipment part is also known as preselector stage. The selector hunters and line finders are generically referred to as preselectors.

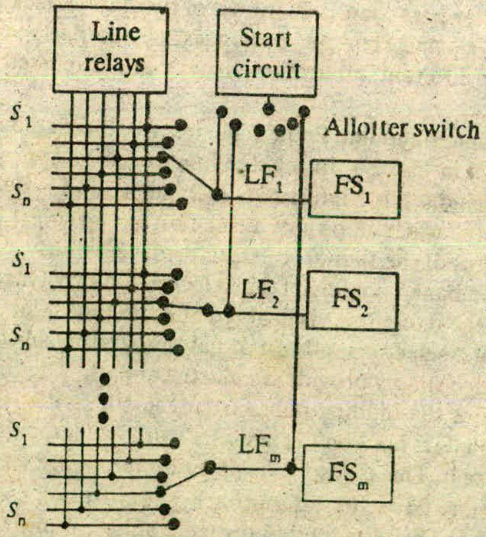
(The switching matrix part consists of one or more sets of two-motion selectors known as first group selector, second group selector, and so on.) The larger the exchange size, the larger is the number of group selector stages. The connector part comprises one set of two-motion selectors known as final selectors. In small Strowger exchanges, all the parts may not exist. Configurations for different capacity exchanges are discussed in Sections 2.6–2.7.

The selector hunter and line finder schemes are illustrated in the trunking diagrams shown in Fig. 2.8. In selector hunter based approach, when a subsciber lifts his hand set, the interrupter mechanism in his selector hunter gets activated and the wiper steps until a free first group selector is found at the outlet. The status of the first group selector, free or busy, is known by a signal in one of the bank contacts of the selector hunter. Once a free first selector is sensed, the interrupter is disabled and the first selector is marked 'busy'. Then, the first selector sends out a dial tone to the subscriber via the selector hunter which simply provides an electrical path. The first selector is now ready to receive the dialling pulses from the subscriber. It is possible that two selector hunters land on the same free first selector simultaneously and attempt to seize it. This is resolved by a suitable seizure circuit.

In the case of line finder based approach, the off-hook signal is sensed by all the line finders. Then the interrupter mechanism of one of the finders, whose associated first selector is free, gets activated and the line finder wiper steps until it reaches the contact on to which the subscriber is terminated. On finding the line, the concerned first selector sends out the dial tone to the subscriber in readiness to receive the dial pulses. The selection of one of the line finders out of many free line finders, is achieved by means of an allotter switch in the start circuit of the line finder as shown in Fig. 2.8 (b). The circuit arrangements are such that the wiper of the allotter switch normally stands



(a) Selector hunter based access



(b) Line finder based access

FS = first selector LF = line finder SH = selector hunter

Fig. 2.8 Subscriber access to Strowger switching system.

on a contact connected to a free line finder and the first selector. When a subscriber lifts his receiver, the start signal from his relay is passed to the particular line finders via the common start circuit and the allotter switch. The line finder then commences to hunt for the calling line. As soon as the calling line is found, the allotter switch steps to the next free line finder. In effect, the line finder and the associated first selector to be used for the next future call is selected in advance by the allotter circuit. In practical designs, several allotter switches are provided in the system to serve calls that may originate in quick succession or simultaneously. Multiple allotter switches also avoid single-point failures, that might lead to complete breakdown of the system.

In designing large exchanges, some practical limitations are encountered in both the above schemes of gaining access to switching resources. Large exchanges are characterised by a large number of subscribers and first group selectors. It is not possible to provide a large number of outlets in the selector hunters or line finders such that any first group selector is accessible by any subscriber. Usually, subscribers are connected in groups of 100 to different sets of line finders which use two-motion selectors. Similarly, sets of selector hunters are connected to different groups of 24 first selectors each. Line finder and selector hunter approaches are advantageous for different sizes of the exchanges. If the exchange is small and the volume of traffic low, line finder approach is economical. For large exchanges with fairly heavy traffic, the selector hunter approach is more cost effective.

When the subscriber starts dialling, the first selector cuts off the dial tone and receives the pulse train corresponding to the first digit dialled by the subscriber. Its wiper assembly steps vertically as many steps as the number of dial pulses. The wipers then move in the horizontal plane across the contacts until they come across a contact to which a free second group selector is connected. This horizontal stepping is completed within the interdigit gap of about 240 ms. Thereafter, the first group selector just provides an electrical path to the second group selector. Each group selector stage functions in a fashion similar to the first group selector by processing one digit of the number dialled by the subscriber and finding a group selector in the subsequent stage. The last two digits of the dialled number are processed by the final selector which steps vertically according to the last but one digit and steps horizontally according to the last digit. Since the final selector responds to digits both in the vertical and horizontal directions unlike the group selectors, it is sometimes referred to as **numerical selector**. If the called subscriber is free, as sensed from a signal at the corresponding bank contact, the final selector sends out a ringing current to the called subscriber and a ringing tone to the calling subscriber. When the called subscriber lifts his handset, the ringing current and tone are cut off and the call metering circuits are enabled by the control circuits associated with the final selectors. If the called subscriber is busy, the final selector sends out a busy tone to the calling subscriber. At any stage of switching, if there is no free selector at the next stage, a busy tone is returned to the calling subscriber.

The control functions in a Strowger system are performed by circuits associated with the selectors. Control and supervisory signals are carried from stage to stage by means of contacts in one of the banks. The wire interconnecting these banks is known as *P*-wire or private wire. Two other bank contacts are used for carrying voice signals and the associated wires are known as negative and positive wires which extend up to the subscriber premises. A selector *X* is said to have seized another selector *Y* in the next stage when the negative, positive and private wires of the selector *X* have been connected to the negative, positive and private wires respectively of the selector *Y*. The complexity and functionality of the control circuits associated with a selector vary depending on the position of the selector in the switching stage.

All the selector control circuits are composed of one or more of the following basic circuits:

1. Guarding circuit
2. Impulsing circuit
3. Homing circuit
4. Metering circuit
5. Ring-trip circuit
6. Alarm circuit.

The guarding circuit is an essential feature of all the selectors. It guards the selector by making it busy as soon as it is seized, lest some other selector involved in setting up another call may also seize it. Once applied, the guarding condition remains set as long as the call is not terminated. The guarding condition is indicated by an earth on the *P*-wire. An earth is supplied to the *P*-wire by the home contact and the homing arc of the home bank. To avoid any unguarded period during the transition of the wiper from the home contact to the homing arc, the wiper is of **bridging** type, i.e. it functions in **make-before-break** fashion; it touches the homing arc before it leaves the home contact.

The impulsing circuit is an essential part of all those selectors which have to respond to dialling pulses. It is used in group and final selectors, but not in line finders or selector hunters. This circuit is usually designed around three relays: one fast acting and the other two slow acting. The fast acting relay faithfully responds to the impulses and passes them on to the *P*-wire circuit. The fast action is achieved by using only one contact spring assembly and an isthmus armature. One of the slow acting relays serves to maintain guarding conditions on the *P*-wire of the incoming circuits and provides for the connection of the selector magnet to the impulsing relay. The third relay is used to recognise the end of a pulse train corresponding to a single digit and prepare the circuit for the next stage in the switching process.

When a selector is searching for a free outlet, the condition on the *P*-wire must be tested to determine whether the outlet is free or not. If an outlet is engaged, the wipers must be allowed to continue the hunting process. If the outlet is free, it must be seized immediately and the incoming positive and negative wires must be switched through to the input of the next stage. At the same time, the hunting process must stop. Once established, the connections must be held until the conversation lasts. All these functions are performed by the testing circuit, and hence this circuit is sometimes referred to as hunting, testing, switching and holding circuit.

There are two methods of indicating the free condition on the *P*-wire: one by means of a simple disconnection and the other by applying a battery to the *P*-wire. As mentioned earlier, the busy condition is indicated by an earth connection. Hence, a testing circuit has to distinguish between an earth and a disconnection in one case and between an earth and a battery in the other. Accordingly, the two methods are referred to as **earth testing** and **battery testing**, respectively. Battery testing is less prone to false connections than earth testing. In any switching process, particularly electromechanical switching, momentary disconnections of lines do occur. Therefore, false switching may take place if the earth testing happens at an instant when a busy outlet is in the course of some switching or release process which temporarily disconnects the guarding earth from the *P*-wire. Such a problem does not occur in the case of battery testing.

At the end of a conversation, all the selectors used for the call must be released and returned to their respective home positions. This operation is performed by homing circuits. The two-motion selectors return to their home position by actuating their self-drive mechanism using interrupt contact. In the case of uniselectors, the necessity of homing arises only for the calling subscriber uniselector. The called subscriber uniselector is already in the home position. Homing operation requires a finite time, and it must be ensured that a hunting selector may not seize a selector which is in the process of homing. Thus, the provision of guarding earth during homing is an integral feature of the homing circuit.

Metering circuit is a special feature of the final selectors. It registers a call against the calling party as soon as the called party answers. The circuit drives a meter containing a simple ratchet-operated counting mechanism with a capacity of 4 to 5 digits. For local calls, the metering is usually independent of the duration of the call and the meter is pulsed only once by the final selector. For long distance calls established using subscriber trunk dialling (STD) facility, the metering is time dependent and the meter is pulsed at an appropriate rate. In this case, the metering pulses are usually received from a remote exchange. Metering is achieved by connecting the meter to the *P*-wire of the subscriber uniselector through a rectifier and applying a positive voltage which makes the rectifier conduct and thereby pulse the meter. The use of the rectifier also ensures that *P*-wire remains guarded during metering.

Ring-trip circuit is a part of the final selectors. The attention of the called subscriber is drawn by ringing the bell of his telephone set. At the same time, a ringing tone is sent out from the final selector to the calling subscriber. Both the ringing current and the ringing tone are cut off by the ring-trip circuit as soon as the called party answers the call. The ringing current in a Strowger system is a 17 Hz alternating current. The ringing tone and the period of interruption of the ringing current are controlled by a relay which is driven by suitable pulsing circuits. To prevent the ringing current from interfering with the speech circuit, the electrical power to the ringing circuit is isolated from the main exchange supply. As soon as the condition of main power being applied to the circuit is sensed, the ringing current is tripped. A common fault of premature tripping of the ringing current occurs when the main supply battery gets connected to the circuit during ringing without the called subscriber actually lifting the handset. If this happens, the bell at the called subscriber telephone set rings only once or twice.

Alarm circuits provide visual and audible indications of any fault or undesirable condition creeping into the selector circuits. Three types of faults are usually detected: **off-hook condition**, **called-subscriber-held**, and **release held**. In the event of a short-circuit in the subscriber line or the subscriber not having replaced his handset properly on the hook, his d.c. loop circuit remains closed and his uniselector hunts and seizes a first selector unnecessarily. To avoid this undesirable use of power and switching elements, every first selector is provided with a permanent glow alarm circuit. This circuit activates an audio and a visual alarm if a selector remains seized for more than six minutes. Called-subscriber-held alarm circuit is necessary in all exchanges where the release of the switching stages is initiated by the calling subscriber replacing his handset. In case the handset is not properly replaced, all the selectors and the called subscriber line remain held, even though the called subscriber has replaced his handset properly. If this happens, neither the called subscriber is able to make any call himself nor can anybody else call him. Thus, the subscriber's instrument remains paralysed. A miscreant can easily create this situation by calling a number and then not replacing his handset on the hook. To prevent this, all final selectors are provided with called-subscriber-held alarm circuit. If the condition of the called subscriber handset having been replaced and the calling subscriber handset not having been replaced lasts for over three minutes, this alarm circuit operates. The third type of alarm circuit, i.e. release-held alarm circuit, senses the failure of a selector to return to home position.

2.5 Design Parameters

When considering the design of a switching system, a number of design alternatives and options may be available. For example, a Strowger switching system may be designed entirely on the basis of uniselectors or two-motion selectors, or a combination of both. It then becomes necessary to compare and evaluate designs to choose from the alternatives. Design parameters assist us in this process. In this section, we define a set of design parameters that characterise the switching systems. These parameters are generic in nature and hence are applicable to all types of switching systems irrespective of the technology or architecture.

The switching network is a major component of any switching system. It is mainly composed of switching elements and the associated circuits. As a result, the cost of the switching network is directly proportional to the number of switching elements in the network. Hence, a good design must attempt to minimise the number of switching elements in the system. When considering the total switching systems, there are other cost elements. For common control systems, the cost of the control subsystem must be taken into account. There is a cost associated with some fixed common hardware elements like ringing current generator, different tone generators and power supplies. A switching network may be realised using one or more stages of switching elements. The higher the number of stages, the longer is the time taken to set up a call as switching is involved in every stage. Every switching system is designed to support a certain maximum number of simultaneous calls, which we call as the **switching capacity**. In most of the designs, the entire switching resources are not utilised even when the switching capacity is fully utilized. Part of the resources remains idle. The fraction of the hardware actually used under full load conditions is an index of the design. Taking these factors into account, we now enumerate the design parameters:

1. Number of subscriber lines, N — 100
2. Total number of switching elements, S — 110
3. Cost of the switching system, C

$$C = S \times C_s + C_c + C_{ch}$$

where

C_s = cost per switching element

C_c = cost of the common control system

C_{ch} = cost of the common hardware

Since the control circuits are associated with switching elements in a Strowger system, C_c is equal to zero. The common hardware is usually a small proportion of the total hardware except for the power supplies and its cost is of the same order in different comparable designs. Hence, we ignore C_{ch} in most of our calculations.

4. Switching capacity, SC
5. Traffic handling capability, TC

$$TC = \frac{\text{switching capacity}}{\text{theoretical maximum load}}$$

$$= \frac{2(SC)}{N}$$

6. Equipment utilisation factor, EUF

$$EUF = \frac{\text{number of switching elements in operation when the } SC \text{ is fully utilised}}{\text{total number of switching elements in the system}}$$

7. Number of switching stages, K
8. Average switching time per stage, T_{st}
9. Call setup time, T_s

$$T_s = T_{st} \times K + T_0$$

where T_0 is the time required for functions other than switching. T_0 is a significant quantity in common control systems where control functions are separated from switching functions. In Strowger (direct control) systems, T_0 may be ignored.

10. Cost capacity index, CCI

$$CCI = \frac{\text{switching capacity}}{\text{cost per subscriber line}} = \frac{N(SC)}{C}$$

The higher the value of CCI, the better is the design. Given the traffic handling capability of a switching system, the stochastic behaviour of the actual traffic and holding time characteristics of a call, it is possible to make reasonable estimates of the blocking probabilities. A detailed treatment of the blocking behaviour of switching systems is presented in Chapter 8. However, simple blocking probability calculations are made when the designs are discussed in the earlier chapters. It may be noted that the blocking probability is more of a performance parameter than a design parameter. However, at the design stage, the traffic handling capability of the switching system must be sized to achieve a low blocking probability in the field. This is done on the basis of estimated traffic.

2.6 100-line Switching System

A 100-line switching system can serve up to 100 subscribers. A 100-line Strowger switching system may be configured in a variety of ways. In this section we discuss five different design alternatives for a 100-line step-by-step switching system. We then compare the designs based on the design parameters discussed in Section 2.5. Simple line diagrams known as **trunking diagrams** are used to represent the configurations of switching systems. For computing the cost of different designs, we assume that the cost of a uniselector is one unit and that of the two-motion selector is two units.

2.6.1 Design 1

An elementary configuration for a 100-line Strowger switching system using 10-outlet uniselectors is shown in Fig. 2.9. The configuration has two stages.

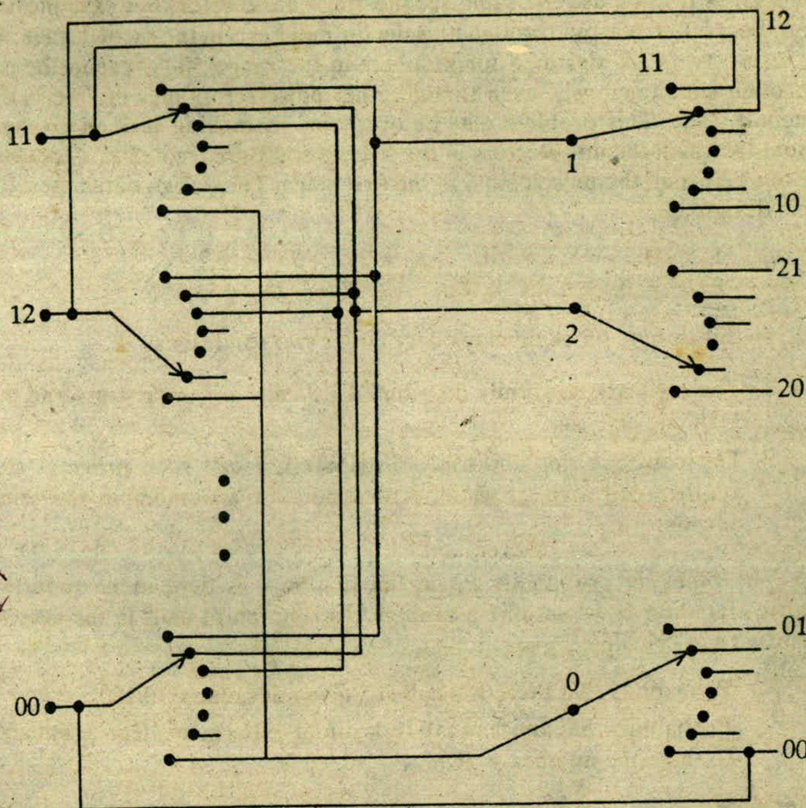


Fig. 2.9 100-line switch using uniselectors

In the first stage, there are 100 uniselectors, one for each subscriber. The second stage has 10 or more uniselectors. The second stage outlets are folded back to the corresponding inlets via suitable control circuitry (not shown in the figure for the sake of simplicity). Usually, each subscriber line is terminated on a relay group at the exchange. The relay group contains all the necessary circuits for the control of the switching mechanism. Functions like testing, switching and return of the tones are done by the relay groups. Similarly, outlets from the first stage are terminated on relay groups at the input of the second stage. The four banks of the uniselectors serve to provide positive, negative, *P*-wire and homing connections. The corresponding outlets of all the first stage uniselectors are commoned or multipled. The first stage responds to the first digit dialled by the user and the second stage to the second digit. Suppose the subscriber 12 dials the number 56, his uniselecto i.e. uniselector 12, steps by five positions and the uniselector 5 in the secor stage steps by six positions. With 10 uniselectors in the second stage, only 1 calls can be established simultaneously. Even this requires that the calls ar uniformly distributed one per decade throughout the number range. All call in a given decade use the same second stage uniselector. For example, the numbers 50–59 are put through uniselector number 5 in the second stage. As a result, two calls destined for numbers in the range 50–59 cannot be put through simultaneously, even though other uniselectors may be free in the second stage. This problem may be overcome by making such an arrangement by which the uniselectors in the second stage are treated as a common resource for all the uniselectors in the first stage. The design parameters for this design are:

$$S = 110, \quad SC = 10, \quad K = 2, \quad TC = 0.2,$$
$$EUF = 0.18, \quad C = 110, \quad CCI = 9.09.$$

In this design, blocking may occur under two conditions:

1. The calls are uniformly distributed, 10 calls are in progress and the 11th one arrives.
2. The calls are not uniformly distributed, a call is in progress and another call arrives, which is destined for a number in the same decade.

The blocking probability P_B in the first case is dependent upon the traffic statistics. If we assume a random distribution of calls in the second case, we can calculate P_B as

Probability that there is a call in a given decade = 10/100

Probability that another call is destined to the same decade but not to the same number = 9/98

Therefore,

$$P_B = (1/10)(9/98) = 0.009$$

2.6.2 Design 2

An alternative scheme which does not involve any logic circuit is to employ 10 uniselectors in the second stage for every one uniselector in the first stage. The total number of uniselectors in the system becomes 1100; 100 in the first stage and 1000 in the second stage. There are 10,000 outlets and 100 inlets. The corresponding outlets associated with every inlet are commoned. For example, all outlets numbered 10 are commoned together. Thus, effectively there are only 100 independent outlets from the switch which are folded back to the corresponding inlets. It may be noted that unlike the previous design, this switching system is nonblocking. The design parameters are:

$$S = 1100, \quad SC = 50, \quad K = 2, \quad TC = 1,$$
$$EUF = 0.09, \quad C = 1100, \quad CCI = 4.54, \quad P_B = 0.$$

Some observations are in order. Apparently, Design 1 appears to have serious limitations. But the values of design parameters, *CCI, EUF* and P_B indicate that it is more cost effective than Design 2. If the traffic statistics indicate that more than 10 calls originate most of the time, the blocking performance of Design 1 becomes unacceptable. In cases where the average number of calls exceeds 10 but still a small fraction (say, less than 20) of the theoretical maximum number of calls, a via media configuration with more than one uniselector per decade in the second stage would be a good solution. But this also calls for a mechanism to choose a free selector out of the many available at the second stage. In step-by-step switching systems, the selection of one out of many selectors in the next subsequent stage is done by deploying a uniselector or the horizontal rotary mechanism of a two motion selector in a self-stepping mode using the interrupter contacts. Designs 4 and 5 discussed later in this section use such arrangements.

2.6.3 Design 3

Another way of realising a 100-line Strowger switching system is to use one two-motion selector for each subscriber. A subscriber is assigned a number in the range 00–99, and the same number is used to identify the two-motion selector assigned to him. The 100 outlets of each two-motion selector are numbered as per the scheme given in Table 2.1. The corresponding outlets in all the 100 two-motion selectors are commoned and folded back to the corresponding inlets. For example, a subscriber with 67 as his number is assigned the two-motion selector 67. The outlet 67 which corresponds to this subscriber is connected to the 7th contact in the 6th vertical position of all the two-motion selectors and folded back to his inlet. The arrangement is shown in Fig. 2.10. If subscriber 23 dials 67, his two motion selector 23 would step vertically 6 times corresponding to the first digit and would step horizontally 7 times to reach the contact to which the subscriber 67 is connected. This switch is nonblocking and uses only one stage of switching elements.

The two-motion selector used to establish a call is dependent upon the initiator of the call. For example, when 23 calls 45, the two-motion selector 23 is used, whereas when 45 calls 23, the two-motion selector 45 is used, although the parties in conversation are the same in both the cases. Since the two-motion selector is activated by the calling party, the call is terminated only when the calling party disconnects the line. If a two-motion selector goes out of order, the subscriber connected to it will not be able to make any outgoing calls but can receive incoming calls. The design parameters of this switch are:

$$S = 100, \quad SC = 50, \quad K = 1, \quad TC = 1,$$
$$EUF = 0.5, \quad C = 200, \quad CCI = 25, \quad P_B = 0.$$

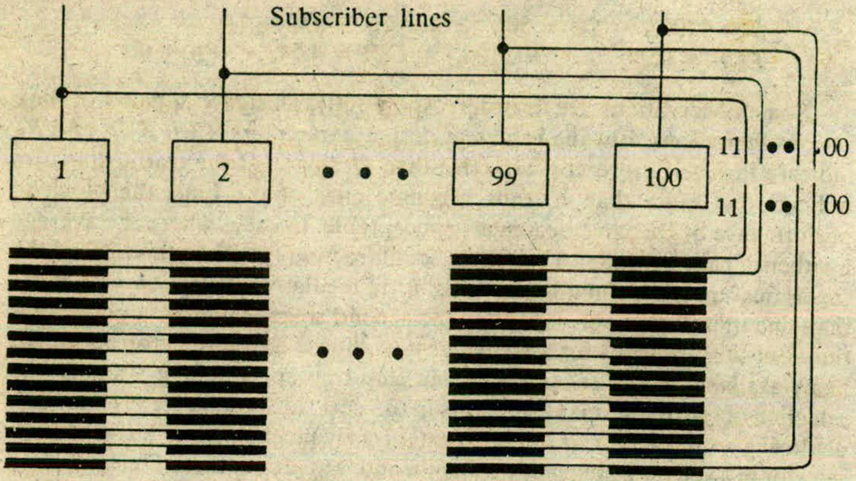


Fig. 2.10 100-line exchange with one two-motion selector per subscriber.

Clearly, Design 3 is superior to Designs 1 and 2. Further improvements to Design 3 are possible if the switching capability is provided to meet only the estimated peak-hour load rather than the theoretical maximum load. Such a design would demand that the switching elements be treated as a common resource accessible by all the subscribers. An elementary requirement of such a design is that the cost savings resulting from placing the switching elements in a common pool should be greater than the cost of the equipment required to associate a subscriber line with a selector. Secondly, the time taken to associate a selector to the subscriber line should not be excessive, and the dial tone must be returned to the subscriber without appreciable delay. Finally, the design must not unduly complicate the maintenance of the equipment and must provide a ready means for tracing connections. Designs 4 and 5 treat switching elements as a common resource.

2.6.4 Design 4

Instead of 100 two-motion selectors as in the case of Design 3, let us consider only 24 two-motion selectors. In this case, 24 simultaneous calls can be put through the switch. The 24 two-motion selectors are shared by all the hundred users. The corresponding outlets of the two-motion selectors are commoned as in the previous case. It is implicitly assumed here that the average peak-hour traffic is 24 simultaneous calls.

We now have to introduce a mechanism by which a user can get hold of a two-motion selector whenever he wants to make a call. Once he seizes a two-motion selector, obtaining the required number follows the same procedure as in the case of Design 3. As discussed in Section 2.4, we may adopt either selector hunter or line finder approach. In this design, we use selector hunters and in Design 5 (see section 2.6.5), we use line finders.

Typically, a 24-outlet uniselector is used as a selector hunter. Each of the 24 outlets is connected to one two-motion selector. Thus, a subscriber has access to all the 24 two-motion selectors in the system. The corresponding outlets of all the selector hunters are commoned and thus, all subscribers have access to all the two-motion selectors. This scheme is shown in Fig. 2.11.

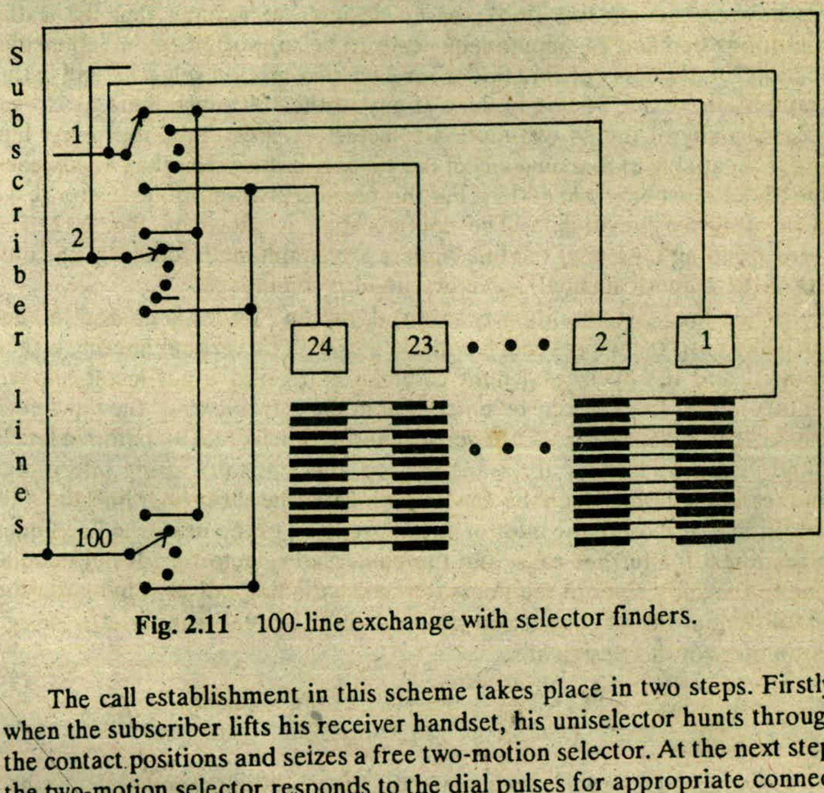


Fig. 2.11 100-line exchange with selector finders.

The call establishment in this scheme takes place in two steps. Firstly, when the subscriber lifts his receiver handset, his uniselector hunts through the contact positions and seizes a free two-motion selector. At the next step, the two-motion selector responds to the dial pulses for appropriate connection. The design parameters of this system are:

S = 100 uniselectors + 24 two-motion selectors

$SC = 24, \quad K = 2, \quad TC = 0.48,$
$EUF = 0.58, \quad C = 148, \quad CCI = 16.2.$

The blocking probability would depend on the traffic characteristics. For an exchange with 100 subscribers, the probability of more than 48 subscribers being active simultaneously is very low. Hence, blocking performance of this design must be satisfactory. This design is clearly superior to Designs 1 and 2. However, the *CCI* of this design is lower than that of Design 3. But the absolute cost is less by 25 per cent. Since the blocking performance of Design 4 is acceptable, it is a better choice than Design 3. In fact, depending upon the traffic estimates, the number of two-motion selectors can be further reduced thereby cutting down the cost. If this number is 10 or less, 10-outlet uniselectors may be used effecting further economy.

2.6.5 Design 5

We now consider a line finder based design. We assume that the traffic conditions demand 24 simultaneous calls to be supported. As mentioned in Section 2.4, there is one line finder for each two-motion selector, and in this example, there are 24 line finders. If any of the 100 subscribers has to get access to any of the 24 two-motion selectors, it is essential that every line finder is capable of reaching any of the 100 subscribers. In other words, each line finder must have 100 outlets. For this purpose, two-motion selectors have to be used as line finders. The configuration is shown in Fig. 2.12. The corresponding outlets of the line finders are commoned. Similarly, the outlets of the numerical (final) selectors are also commoned.

When the start condition is received, the line finder is caused to hunt vertically until the wipers reach a marked level. The vertical hunting is then stopped and the horizontal hunt commences to find a particular marked contact in that level. It may be noted that in the extreme case, the line finder may have to take 20 steps – 10 vertical and 10 horizontal – before a line is found. The line finders are made to step automatically, using interrupter contact mechanism. When the line finder locates the subscriber line, the start condition is removed, the allotter switch steps on to the next free line finder in readiness for further calls, and the numerical selector sends out the dial tone to the subscriber in readiness to receive dialling pulses. Thereafter the establishment of the connection proceeds in the usual manner. The design parameters of this design are:

$$S = 48, \quad SC = 24, \quad K = 1, \quad TC = 0.48,$$
$$EUF = 1, \quad C = 96, \quad CCI = 25.$$

Obviously, Design 5 is by far the best for a 100-line exchange. If we had used uniselectors as line finders, it would have been necessary to divide subscriber lines into small groups of, say 24 each. Such designs involving groupings, function efficiently only under certain specific traffic conditions and generally lead to higher blocking probabilities. However, subscriber grouping or selector grouping cannot be avoided in large exchanges as discussed in Sections 2.7 and 2.8.

2.7 1000-line Blocking Exchange

It is rare that exchanges with more than a few hundred subscribers are designed to be nonblocking. Hence we only consider a blocking design for a

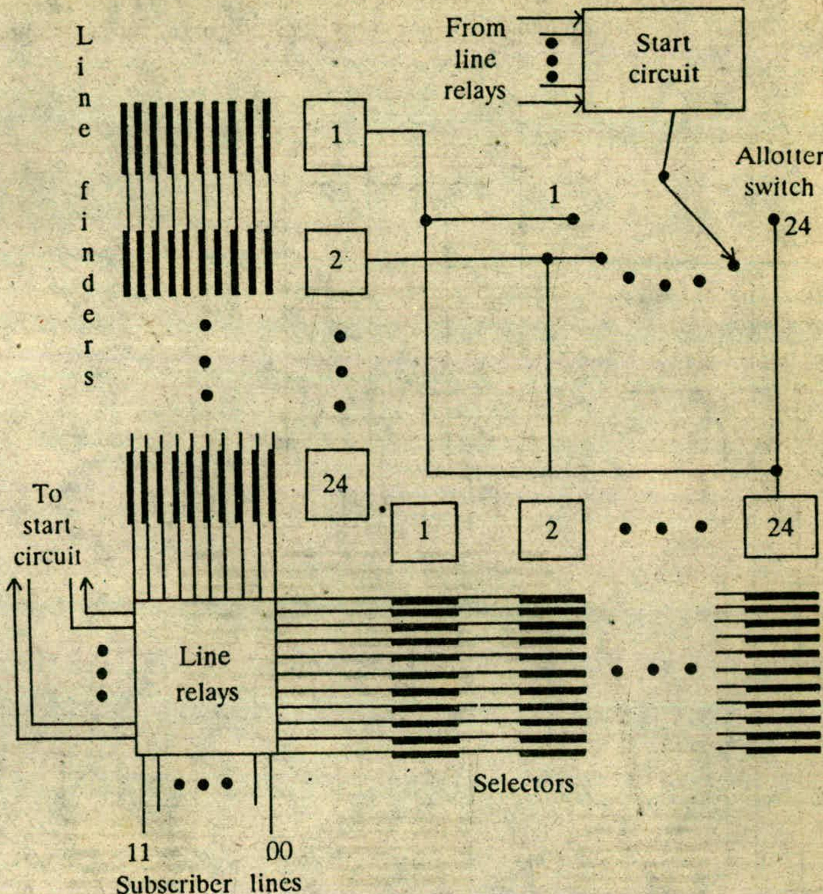
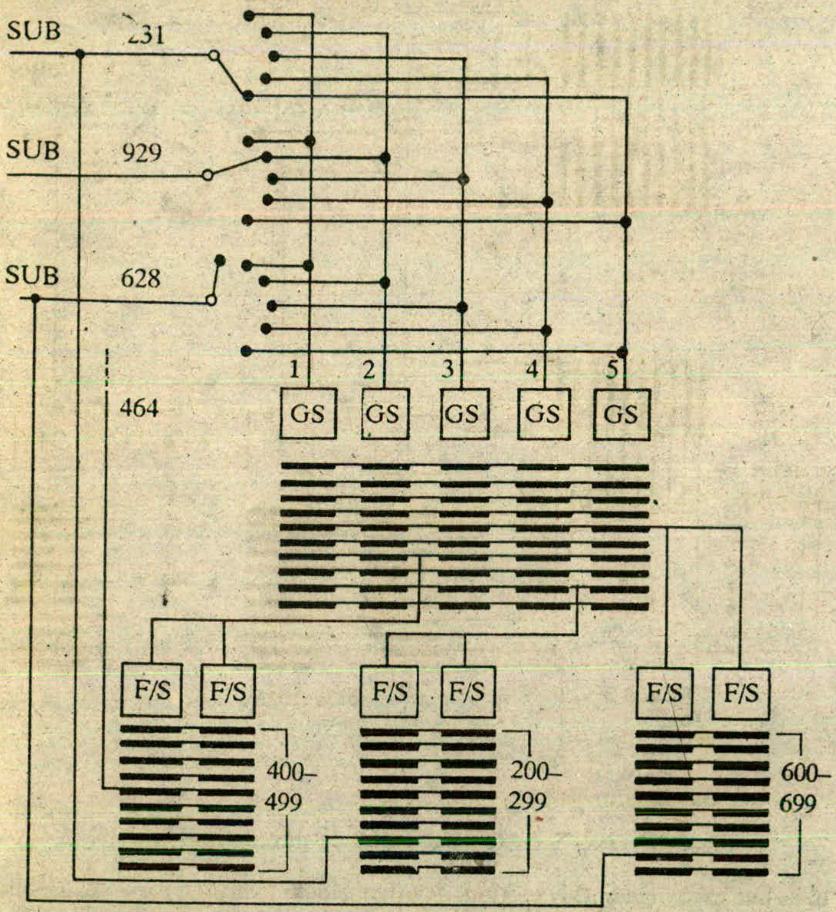


Fig. 2.12 100-line exchange with two-motion line finders.

1000-line exchange in thi, section. Another blocking design for a 10,000-line exchange is described in Section 2.8. In a 1000-line exchange, subscribers are identified by a three-digit number ranging from 000 to 999. As explained in Section 2.4, the final numerical selector in a Strowger system responds to the last two digits dialled by the subscriber. Hence for a 1000-line exchange, we need one more selector stage preceding the final selector stage, which would respond to the first digit of the called subscriber and establish a connection to a final selector. This is a group selector stage which uses two-motion selectors as switching elements. In addition to these two stages, we need either a selector hunter or a line finder stage as a preselector stage. Commercial designs of large exchanges of 1000-lines or more tend to use selector hunters. We, therefore, discuss only the selector hunter based approach here, using one 24-outlet uniselector for each subscriber. Readers are advised to try out a line finder based design.

The trunking diagram for a 1000-line exchange is given in Fig. 2.13. As in the case of 100-line exchange, when a subscriber lifts his receiver, the pre-



F/S = final selector GS = group selector SUB = subscriber

Fig. 2.13 Trunking diagram of a 1000-line exchange.

selector hunts for a free group selector. When a free group selector is obtained, the subscriber is given the dial tone. When the subscriber dials the first digit, the group selector steps up in the vertical direction according to the digit dialled, and hunts for a free final selector in one of its 10 outlets. If a free selector is obtained, it responds to the next two digits and a connection is established, otherwise an engaged tone is sent out to the subscriber.

Each final selector, which is a two-motion selector, provides 100 outlets, and we need a minimum of 10 final selectors to connect 1000 subscribers. With 10 final selectors, only 10 simultaneous calls can be established, which is too small a number for a 1000-line exchange. The design must cater to about 100–200 simultaneous calls by providing as many final selectors. The final selectors are divided into 10 separate groups, each group containing more than one final selector, and giving access to a block of 100 subscriber numbers. The number blocks are 0–99, 100–199, 200–299, and so on. Two hundred final selectors may be uniformly distributed among the 10 groups each having 20 selectors. In such an arrangement, 20 simultaneous calls may be put through in each group. If the number of calls exceeds 20 in a group, the calls are blocked even though final selectors may be free in other groups. In view of this, the selectors may be non-uniformly distributed among the groups depending on the estimated traffic in each group.

EXAMPLE 2.1 In a 1000-line exchange, the number range 000–299 is allotted to business subscribers. Forty per cent of these subscribers in each group of 100 are active during peak hours. The number range 300–999 is allotted to domestic connections. Ten per cent of the domestic subscribers are active in each group at any time. Estimate the total number of final selectors required.

Solution Number of simultaneous calls for business subscriber groups is equal to 20 per group

Number of simultaneous calls for domestic subscriber groups $= 5$ per group

Total number of final selectors required $= 3 \times 20 + 7 \times 5 = 95$

EXAMPLE 2.2 In Example 2.1, if the probability of more than 40 per cent business customers being active is 0.01 and the probability of more than 10 per cent of the domestic customers being active is 0.05, estimate the blocking performance of the exchange. Assume that switching stages other than final selector stage are designed to be nonblocking.

Solution The exchange appears blocking if a business or a domestic customer is blocked. As a result, the blocking probability P_B of the exchange is given by $P_B = 0.01 + 0.05 = 0.06$.

Let us now turn our attention to the group selector stage. The number of selectors in this stage must be at least as many as the number of final selectors to support the required number of simultaneous calls. The problem of distribution of the group selectors is somewhat more complex than that of the final selectors. Firstly, a subscriber uniselector has 24 outlets and hence each user can gain access to a maximum of 24 group selectors. Since the total number of group selectors is less than the number of subscribers, more than one subscriber is terminated on each group selector. With 200 first group selectors in the system and each subscriber connected to 24 of them, 120 users are terminated on each group selector. If a group selector has already been seized by one of the subscribers, the others find the same busy. Every group selector has access to all the 10 groups in the final stage. A level in the group selector stage corresponds to a group in the final selector stage. At each level, there are only 10 outlets and hence only 10 final selectors in a group can be connected to a group selector level. When more than 10 final selectors are provided in a group, special arrangements are required to make all final selectors accessible to a group selector. A technique called **grading group connection** permits more than 10 final selectors to be connected to a group selector. Alternatively, 200 outlets may be made available from each group selector, which permits access to 20 final selectors in each level.

In the trunking diagram of Fig. 2.13, two connections are shown established. Subscriber 231 has seized the group selector 5 and a call is established to subscriber 628. Similarly subscriber 929 is connected to subscriber 464 via group selector 2. Readers may observe the folding connections. During a conversation, the calling subscriber's uniselector, a group selector, a final selector and the called subscriber's uniselector remain busy. At the end of the conversation when both the subscribers replace their handsets, all the switching elements return to their home positions.

Assuming 200 final selectors and 200 group selectors, the design parameters for this exchange are as under:

$$S = 1000 \text{ uniselectors} + 400 \text{ two-motion selectors}$$

$$SC = 200, \quad K = 3, \quad TC = 0.4, \quad EUF = 0.57,$$

$$C = 1800, \quad CCI = 111.$$

The blocking performance would depend upon the traffic characteristics in each group of 100 numbers. We may, however, calculate some empirical value of P_B for a group, assuming 20 final selectors for each group and random distribution of traffic.

$$\text{Probability that one call lands on a given block} = \frac{100}{1000}$$

Probability that two calls are in the same block

$$= \frac{100}{1000} \times \frac{99}{998}$$

$$\text{Probability that there are 20 calls} = \frac{100}{1000} \times \frac{99}{998} \times \cdots \frac{81}{962}$$

In this calculation, it is implied that the calls do not originate from the block under consideration.

2.8 10,000-line Exchange

A 10,000-line exchange has four stages: a preselector stage, two group selector stages and a final selector stage. We consider a selector hunter preselector stage with one 24-outlet uniselector per subscriber. Whenever a subscriber lifts his telephone set, the corresponding pre-selector hunts and obtains a two-motion selector from the first set of group selectors. When the subscriber dials the number, the first group selector responds to the first digit, the second group selector to the second digit and the final selector to the last two digits.

In order to support 10,000 subscribers, we need a minimum of 100 final selectors. Since there are 100 blocks of 100 numbers each (0–99, 100–199, ...) in the number range 0–9999, the final selectors are placed in 100 separate groups. A little reflection reveals that the second group selectors have to be placed in 10 separate groups corresponding to the number ranges 0–999, 1000–1999 and so on. No grouping is required as far as the first group selectors are concerned. Consider a design that can put through 1000 simultaneous calls. We need a minimum of 1000 first group selectors, 1000 second group selectors and 1000 final selectors. For the sake of simplicity, we assume uniform distribution of selectors at every stage. A subscriber has access to 24 first group selectors. Two hundred and forty subscribers are terminated on each first group selector. There are 100 second group selectors belonging to each level of the first group selector. However, a first group selector has access to only 10 second group selectors in a given level. Each second group selector is accessed by 100 first group selectors as there are 10,000 outlets in total at each level of the first group selector stage. There are 10 final group selectors belonging to each level of the second group selector. Each final selector is accessed by 100 second group selectors.

The design parameters are:

S = 10,000 uniselectors + 3000 two-motion selectors
$SC = 1000$, $TC = 0.2$, $K = 4$, $EUF = 0.3$,
$C = 16000$, $CCI = 62.5$.

FURTHER READING

1. Atkinson, J., *Telephony, Vol. 2, Automatic Exchange Systems*, The New Era Publishing Co., London, 1950.
2. Biswas, N. N., *Principles of Telephony*, Radiant Books, Bangalore, 1970.
3. Jolley, E. H., *Introduction to Telephony and Telegraphy*, A. H. Wheeler & Co., Allahabad, 1967.

EXERCISES

1. Draw the pulse dialling waveform for the number 41.
2. Calculate the time required to dial the number 00–91–44–414630 using a rotary dial telephone. Assume that the subscriber takes 600 ms on an average to rotate the dial for a single digit.

3. How can a suppressor cam be designed such that the interdigit gap precedes the dialling pulses?

4. Explain the working of the trigger dial mechanism. How is this superior to cam dial mechanism?

5. A busy tone does not imply that the called party is actually engaged in a conversation. Explain.

6. A regular long-distance caller disconnects one of the call attempts immediately on hearing the ringing tone with a remark that the call has landed on a wrong number. Can he be right? Why?

7. A long-distance dialler hears four different types of call-in-progress signals while establishing a call. What can he conclude?

8. In an English-speaking country, a long distance caller hears the voice announcement 'Lines in this route are busy. Please try after some time'.

   Is it possible for him to determine which segment is busy? Compare the situation in a multilingual country like India.

9. Describe the working of a rotary switch. Differentiate between forward acting and reverse acting types.

10. In a 100-line Strowger exchange using 100 two-motion selectors, show the trunking diagram when the subscriber 85 establishes a connection to subscriber 58. How does the diagram change if the call is initiated by subscriber 58?

11. Give the relative positions of the number pairs (61, 60) and (05, 35) in a two-motion selector.

12. What are the basic approaches to the design of subscriber access to Strowger systems? Describe them.

13. Describe how a uniselector can be used as a selector hunter or line finder.

14. Distinguish between earth testing and battery testing as applied to hunting operations in Strowger exchanges. Discuss the relative merits of each method.

15. A 1000-line exchange has 24 group selectors and 20 final selectors. How many simultaneous calls can be put through this exchange? How many simultaneous calls in the number range 200–299 can be put through if final selectors are uniformly distributed?

16. A 1000-line exchange has 24 group selectors and 50 final selectors uniformly distributed. How many simultaneous calls can be put through the exchange? How many simultaneous calls in the range 200–299 can be put through?

17. Design a 100-line Strowger exchange using line finders to achieve a blocking probability less than 0.05. Calculate the design parameters. Assume that the probability that a subscriber is active is 0.1.

18. In Strowger exchanges, a call may be blocked even though an appropriate path through the switch exists. Explain how this can happen.

19. In a step-by-step switching system it is possible to use a number system other than the decimal system. Given that cost of a two-motion selector is governed by the equation

    $$C_s = 100 + 0.25b^2 \text{ units}$$

    where b is the number base, determine the value of b that would result in the minium cost for a switch that supports N subscribers.

3

Crossbar Switching

The Strowger switching system has been the basis of telephone switching for almost 70 years since its introduction in 1889. However, the major disadvantage of the Strowger system is its dependence on moving parts and contacts that are subject to wear and tear. A two-motion selector moves, on an average, 4 cm vertically and makes a complete rotation horizontally, in establishing and terminating a connection. Such mechanical systems require regular maintenance and adjustment and for this purpose they must be located in places that are easily and speedily accessible by skilled technicians. As the telephone network spread to remote areas, it became necessary to devise switching systems that would require less maintenance and little readjustment after installation. Efforts in this direction led to the invention of crossbar switching systems. The search for a switch in which the contacts operated with only a small mechanical motion using a small number of magnets started in early twentieth century. The first patent for such a device was granted in 1915 to J.N. Reynolds of Western Electric, USA. This was followed by a patent application in 1919 by two Swedish engineers, Betulander and Palmgren for a crossbar switch. Subsequent developments led to the introduction of crossbar switching systems in the field in 1938 by AT&T laboratories in the United States. The first design was christened No. 1 crossbar system. Since then the crossbar systems have been progressively replacing Strowger systems. Apart from the desirable and efficient switch characteristics, crossbar systems differ from Strowger systems in one fundamental respect: they are designed using the common control concept.

3.1 Principles of Common Control

Although common control subsystems were first introduced in crossbar exchanges, the genesis of common control concept can be traced to the Director system used with Strowger exchanges. A Director system facilitates uniform numbering of subscribers in a multiexchange area like a big city and routing of calls from one exchange to another via some intermediate exchanges. Uniform numbering means that to call a particular subscriber, the same number is dialled, no matter from which exchange the call originates a fact to which we are so accustomed, these days. But, it is not possible to implement such a scheme in a direct control switching system without the help of a Director.

Consider the multiexchange network shown in Fig. 3.1. For considerations of economy, it is not a fully connected network. If a subscriber in

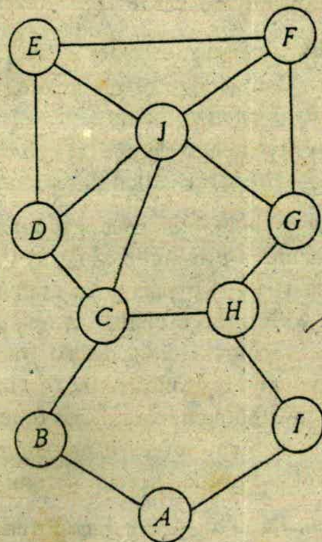


Fig. 3.1 A multiexchange network.

Exchange A wants to call a subscriber in Exchange F, the call has to be routed via at least three exchanges. Two routes are possible: A-B-C-J-F and A-I-H-G-F. In a Strowger system, a call can be sent out of an exchange by reserving a level in the first group selector for outside calls. The 10 outlets at the reserved level can be connected to 10 different exchanges. Let us assume that the reserved level is zero and the outlets are assigned as in the following for the sake of discussions:

From exchange	Outlet	To exchange
A	01	*B*
A	02	*I*
B	04	*C*
C	03	*J*
I	05	*H*
H	01	*G*
G	02	*F*
J	01	*F*

Let 1457 be the subscriber to be called in Exchange *F*. From Exchange *A*, the called subscriber can be reached by dialling either of the following number sequence:

For route *A-B-C-J-F*	01-04-03-01-1457
For route *A-I-H-G-F*	02-05-01-02-1457

The difficulties are now obvious:

- Identification number of a subscriber is route dependent.
- A user must have knowledge of the topology of the network and outlet assignments in each exchange.
- Depending on from which exchange the call originates, the number and its size vary for the same called subscriber.

These difficulties can be overcome if the routing is done by the exchange and a uniform numbering scheme is presented as far as the user is concerned. A number may now consist of two parts: An exchange identifier and a subscriber line identifier within the exchange. An exchange must have the capability of receiving and storing the digits dialled, translating the exchange identifier into routing digits, and transmitting the routing and the subscriber line identifier digits to the switching network. This function is performed by the Director subsystem in a Strowger exchange. Some important observations are in order with regard to the Director system:

- As soon as the translated digits are transmitted, the Director is free to process another call and is not involved in maintaining the circuit for the conversation.
- Call processing takes place independent of the switching network.
- A user is assigned a logical number which is independent of the physical line number used to establish a connection to him. The logical address is translated to actual physical address for connection establishment by an address translation mechanism.

All the above are fundamental features of a common control system. A functional block diagram of a common control switching system is shown in Fig. 3.2. The control functions in a switching system may be placed under four broad categories:

1. Event monitoring
2. Call processing
3. Charging
4. Operation and maintenance.

Events occurring outside the exchange at the line units, trunk junctors and interexchange signalling receiver/sender units are all monitored by the control subsystem. Typical events include call request and call release signals at the line units. The occurrences of the events are signalled by operating

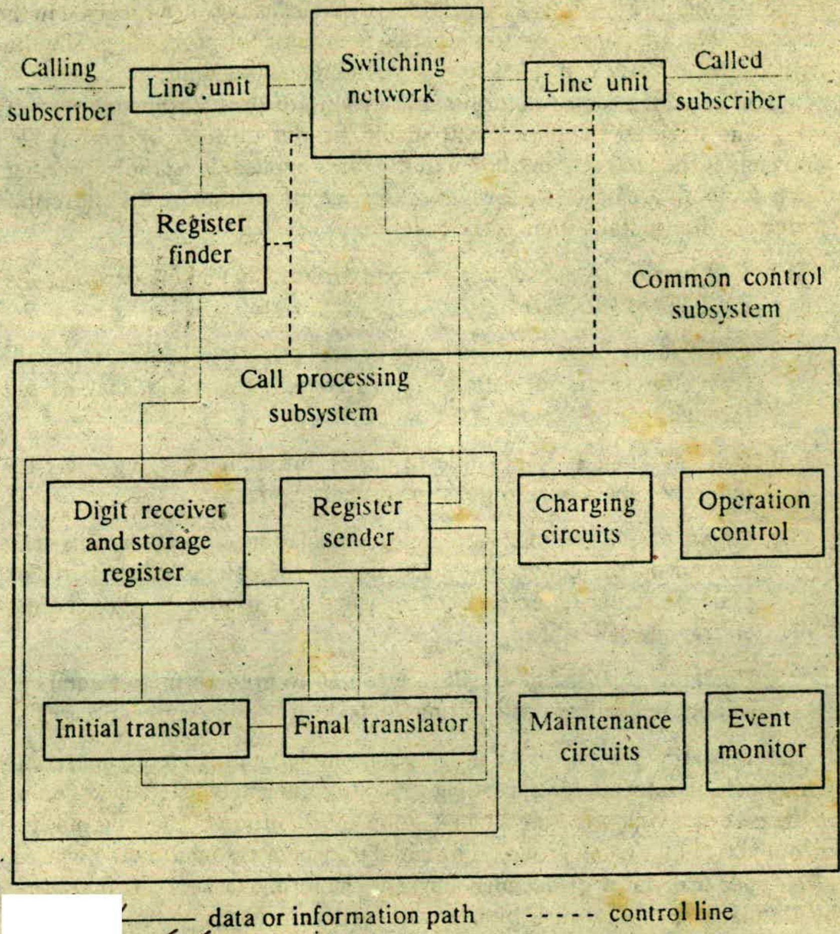


Fig. 3.2 Common control switching system.

relays which initiate control action. The control subsystem may operate relays in the junctors, receivers/senders and the line units, and thus command these units to perform certain functions. Event monitoring may be distributed. For example, the line units themselves may initiate control actions on the occurrence of certain line events.

When a subscriber goes off-hook, the event is sensed, the calling location is determined and marked for dial tone, and the register finder is activated to seize a free register. Identity of the calling line is used to determine line category and the class of service to which the subscriber belongs. As mentioned in Section 2.1, a subscriber telephone may use either pulse dialling or multifrequency dialling, and the line is categorised based on this. A register appropriate to the line category is chosen, which then sends out the dial tone to the subscriber, in readiness to receive the dialling information. As soon as the initial digits (usually 2–5) which identify the exchange are received in the register, they are passed on to the initial translator for processing. Simultaneously, the register continues to receive the remaining digits.

The initial translator determines the route for the call through the network and decides whether a call should be put through or not. It also determines the charging method and the rates applicable to the subscriber. Such decisions are based on the class of service information of the subscriber which specifies details such as the following:

1. Call barring: A subscriber may be barred from making certain calls, e.g. STD or ISD barring.
2. Call priority: When the exchange or network is overloaded, only calls from subscribers identified as priority-call subscribers, may be put through.
3. Call charging: It is possible to define different charging rules for different subscribers in the same exchange.
4. Origin based routing: Routing or destination of certain calls may depend on the geographical location of the calling subscriber. For example, calls to emergency services are routed to the nearest emergency call centre.
5. No dialling calls: These calls are routed to predetermined numbers without the calling party having to dial, e.g. hot line connections.

Initial translation may also take into account instructions from the operating personnel and information regarding the status of the network. For example, in the case of a fault affecting a trunk group, a proportion of the calls may be rerouted via other trunk groups. The initial translator is sometimes known as **office code translator** or **decoder-marker**. The term 'marker' was first used by Betulander, the Swedish pioneer of crossbar technology, to mean controls. The term came into use because the terminals to be interconnected were 'marked' by applying electrical signals. The term is widely used even today when discussing crossbar systems.

If a call is destined to a number in an exchange other than the present one processing the digits, the initial translator generates the required routing digits and passes them on to the register sender. Here, the digits corresponding to the subscriber identification are concatenated and the combined digit pattern is transmitted over the trunks to the external exchange. Register sender uses appropriate signalling technique, depending on the requirements of the destination exchange. If the call is destined to a subscriber within the same exchange, the digits are processed by the final translator. The translation of directory number to equipment number takes place at this stage. The final translator determines the line unit to which a call must be connected and the category of the called line. The category information may influence charging and connection establishment. For example, there may be no charge for emergency lines or fault repair service lines. Some commercial services may offer charge-free or toll-free connection to their numbers. Charging schemes, tariff structures and billing methods are discussed in Chapter 9. In some practical implementations, both initial and final code translator functions are performed by a single translator.

Controlling the operation of the switching network is an important function of the common control subsystem. This is done by marking the switching elements at different stages in accordance with a set of binary data defining the path and then commanding the actual connection of the path. Path finding may be carried out at the level of the common control unit or the switching network. In the former case, the technique is known as **map-in-memory**, and in the latter as **map-in-network**. In the map-in-memory technique, the control unit supplies the complete data defining the path, whereas in the map-in-network technique, the control unit merely marks the inlet and outlet to be connected, and the actual path is determined by the switching network. The former technique is usually present in stored program control subsystems. The latter is more common in crossbar exchanges using markers for control.

Administration of a telephone exchange involves activities such as putting new subscriber lines and trunks into service, modifying subscriber service entitlements and changing routing plans based on the network status. Control subsystems must facilitate such administrative functions. Maintenance activities include supervision of the proper functioning of the exchange equipment, subscriber lines and trunks. It should be possible for the maintenance personnel to access any line or trunk for performing tests and making measurements of different line parameters. The control subsystem should also aid fault tracing without the maintenance personnel having to perform elaborate tests.

3.2 Touch Tone Dial Telephone

In a rotary dial telephone, it takes about 12 seconds to dial a 7-digit number. From the subscriber point of view, a faster dialling rate is desirable. The step-by-step switching elements of Strowger systems cannot respond to rates higher than 10–12 pulses per second. With the introduction of common control in crossbar systems, a higher dialling rate is feasible. It is also of advantage as the common equipment, while not tied up for the duration of a call, is nonetheless unavailable to respond to a new call until it has received and processed all the digits of an earlier call. Pulse dialling is limited to signalling between the exchange and the subscriber and no signalling is possible end-to-end, i.e. between two subscribers. End-to-end signalling is a desirable feature and is possible only if the signalling is in the voice frequency band so that the signalling information can be transmitted to any point in the telephone network to which voice can be transmitted. Rotary dial signalling is limited to 10 distinct signals, whereas a higher number would enhance signalling capability significantly. Finally, a more convenient method of signalling than rotary dialling is preferable from the point of view of human factors. These considerations led to the development of touch tone dial telephones in the 1950s, which were introduced first in 1964 after field trials. They are increasingly replacing rotary dial telephones all over the world.

The touch tone dialling scheme is shown in Fig. 3.3. The rotary dial is replaced by a push button keyboard. 'Touching' a button generates a 'tone'

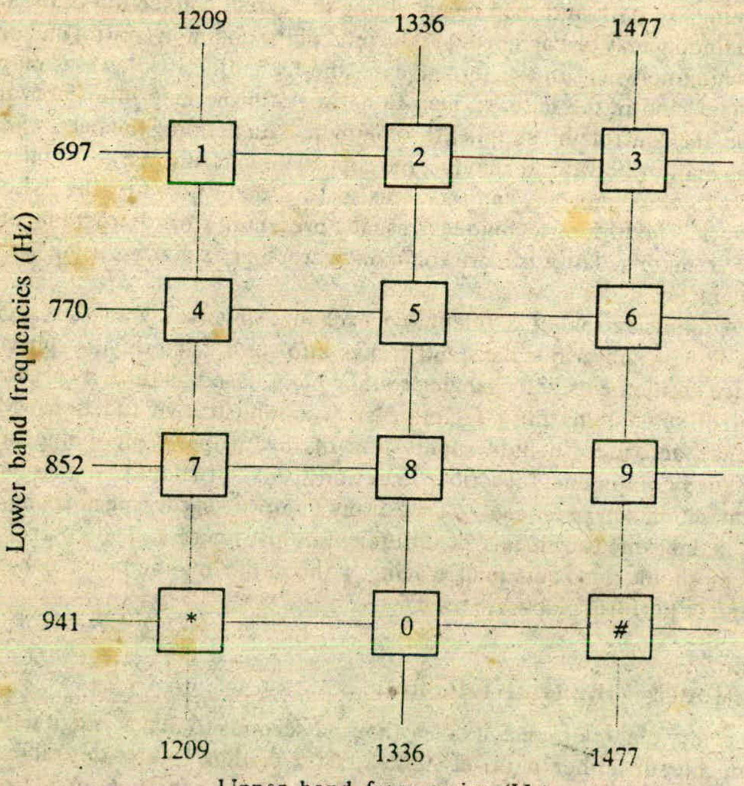


Fig. 3.3 Touch dial arrangement.

which is a combination of two frequencies, one from the lower band and the other from the upper band. For example, pressing the push button 9 transmits 852 Hz and 1477 Hz. An extended design provides for an additional frequency 1633 Hz in the upper band, and can produce 16 distinct signals. This design is used only in military and other special applications. Another design, known as **decadic push button** type, uses a push button dial in place of rotary dial but gives out decadic pulses when a button is pressed as in the case of rotary dial telephone.

3.2.1 Design Considerations

The need for touch tone signalling frequencies to be in the voice band brings with it the problem of vulnerability to 'talk-off' which means that the speech signals may be mistaken for touch tone signals and unwanted control actions such as terminating a call may occur. Another aspect of talk-off is that the speech signal may interfere with the touch tone signalling if the subscriber happens to talk while signalling is being attempted. The main design considerations for touch tone signalling stem from the need for protection against talk-off and include the following factors:

1. Choice of code
2. Band separation
3. Choice of frequencies
4. Choice of power levels
5. Signalling duration.

In addition to these, human factors and mechanical aspects also require consideration.

The choice of code for touch tone signalling should be such that imitation of code signals by speech and music should be difficult. Simple single frequency structures are prone to easy imitation as they occur frequently in speech or music. Hence, some form of multifrequency code is required. Such codes are easily derived by selecting a set of N frequencies and restricting them in a binary fashion to being either present or absent in a code combination. However, some of the 2^N combinations are not useful as they contain only one frequency. Transmitting simultaneously N frequencies involves N-fold sharing of a restricted amplitude range, and hence it is desirable to keep as small as possible the number of frequencies to be transmitted simultaneously. It is also advantageous to keep fixed the number of frequencies to be transmitted for any valid code word. These factors lead to the consideration of P-out-of-N code. Here a combination of P frequencies out of N frequencies constitutes a code word. The code yields $N!/P!(N-P)!$ code words.

Prior to touch tone, P-out-of-N multifrequency signalling, known as multifrequency key pulsing (MFKP), was used between telephone exchanges by the operators. Here, 2-out-of-6 code was used. This code is known to give a talk-off performance of less than 1 in 5000. However, this degree of talk-off performance is inadequate for subscriber level signalling. In order to improve the performance, two measures are adopted. Firstly, while retaining P as two, N is chosen to be seven or eight, depending upon the number of code words desired. Secondly, the chosen frequencies are placed in two separate bands, and a restriction is applied such that one frequency from each band is chosen to form a code word. When multiple frequencies are present in speech signal, they are closely spaced. Band separation of touch tone frequencies reduces the probability of speech being able to produce touch tone combinations. The number of valid combinations is now limited to $N1 \times N2$, where $N1$ and $N2$ represent the number of frequencies in each band. With seven frequencies, four in one band and three in the other, we have 12 distinct signals as represented by the push buttons in Fig. 3.3. With eight frequencies, four in each band, we have 16 possible combinations. Since two frequencies are mixed from a set of seven or eight frequencies, CCITT refers to the touch tone scheme by the name **dual tone multifrequency (DTMF) signalling.**

Band separation of the two frequencies has the following advantages:

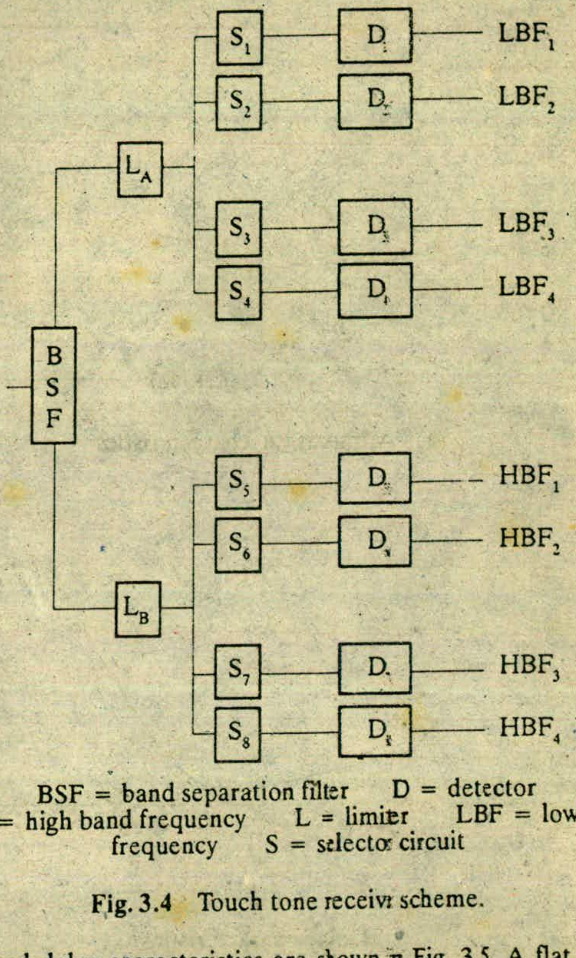
1. Before attempting to determine the two specific frequencies at the receiver end, band filtering can be used to separate the frequency groups. This renders determination of specific frequencies simpler.

2. Each frequency component can be amplitude regulated separately.

3. Extreme instantaneous limiters which are capable of providing substantial guard action can be used for each frequency separately to reduce the probability of false response to speech or other unwanted signals.

Figure 3.4 shows a simplified block diagram of a touch tone receiver. The limiters accentuate differences in levels between the components of an incoming multifrequency signal. For example, if two frequencies reach the limiter with one of them being relatively strong, the output of the limiter peaks with the stronger signal and the weaker signal is further attenuated. If both the signals have similar strengths, the limiter output is much below the full output and neither signal dominates at the output. The selective circuitry is designed to recognise a signal as bonafide when it falls within the specified narrow passband and has an amplitude within about 2.5 dB of the full output of the limiter. The limiter and the selective circuits together reduce the probability of mistaking the speech signal to be a touch tone signal. Speech signals usually have multifrequency components with similar amplitudes and hence, the limiter does not produce a full output. As a result, the selective circuitry rejects the signal as invalid. In order to further improve the talk-off performance, band elimination filters may be used in place of band separation filters at the input of the touch tone receiver: Band elimination filters permit a wider spectrum of speech to be passed to the limiters, thus making it less probable for the limiters to produce a full output at the touch tone signal frequency.

The choice of frequencies for touch tone signalling is dictated by the attenuation and delay distortion characteristics of telephone network circuits for the voice band frequencies (300 Hz–3400 Hz). Typical amplitude



BSF = band separation filter D = detector
HBF = high band frequency L = limiter LBF = low band frequency S = selector circuit

Fig. 3.4 Touch tone receiver scheme.

response and delay characteristics are shown in Fig. 3.5. A flat amplitude response with a very low attenuation and a uniform delay response with a low relative delay value are desirable. Examining the curves shown in Fig. 3.5, frequencies in the range of 700–2200 Hz may be considered. The actual range chosen for touch tone dialling is 700–1700 Hz. Both the lower and the upper frequency bands are defined in this range. The frequency spacing depends in part on the accuracies with which the signal frequencies can be produced. An accuracy of ± 1.5% is easily obtainable at the telephone sets. The selective circuits can be designed to a tolerance of ± 0.5%, leading to a total acceptable variation of ± 2% in the nominal frequency value. Hence, a minimum spacing of 4% is indicated between frequencies. However, a wider spacing has been chosen as is seen from Fig. 3.3, which makes the precise maintenance of the bandwidth less critical.

Having decided on the frequency band and the spacing, the specific values of the frequencies can be so chosen as to avoid simple harmonic

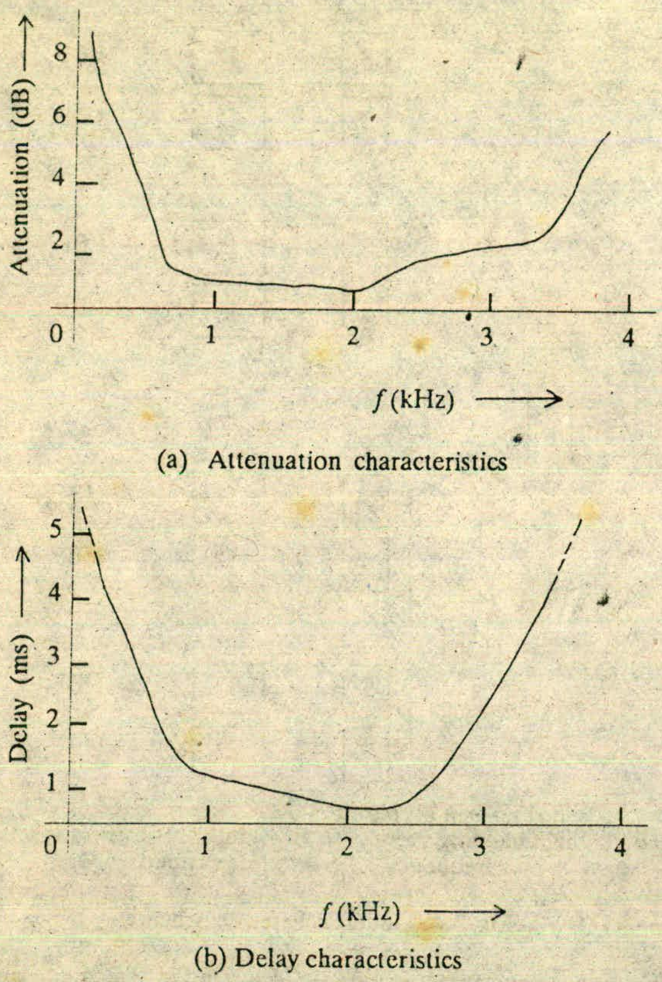


Fig. 3.5 Typical attenuation and delay characteristic of telephone networks.

relationships like 1:2 and 2:3 between adjacent two frequencies in the same band and between pairs of frequencies in the two different bands, respectively. Such a selection improves talk-off performance. As mentioned earlier, sounds composed of a multiplicity of frequencies at comparable levels are not likely to produce talk-off because of the limiter and selector design. Such sounds are produced by consonants. However, vowels are single frequency sounds with a series of harmonic components present in them. Susceptibility to talk-off due to vowels can be reduced by choosing the specific frequencies appropriately. The adjacent frequencies in the same band have a fixed ratio of 21:19, i.e. only the 21st and 19th harmonic components have the same frequency values. Across the bands, the frequencies that lie along the diagonals in Fig. 3.3 have a ratio of 59:34. Thus, the chosen frequency values are such that they almost eliminate talk-off possibility due to harmonics.

Since signalling information does not bear the redundancy of spoken words and sentences, it is desirable that the signal power be as large as possible. A nominal value of 1 dB above 1 mW is provided for at the telephone set for the combined signal power of the two frequencies. As is seen from Fig. 3.5(a), the attenuation increases with the frequency. It has been observed that in the worst case, the increase in attenuation in the subscriber loop between 697 Hz and 1633 Hz could be as much as 4 dB. To compensate for this, the upper band frequencies are transmitted at a level 3 dB higher than that of the lower band frequencies. The nominal output power levels have been chosen as – 3.5 dBm and – 0.5 dBm for the lower and upper band frequencies, respectively.

The probability of talk-off can be reduced by increasing the duration of the test applied to a signal by the receiver before accepting the signal as valid. But, it is clearly unacceptable to expect the user to extend the push button operation for this purpose, beyond an interval that is natural to his dialling habit. Fortunately, this requirement does not arise as even the 'fast' dialler pauses for about 200 ms between digits, and efficient circuits can be designed to accurately determine the signalling frequencies by testing for a much smaller duration. A minimum of 40 ms has been chosen for both signal and intersignal intervals, allowing for a dialling rate of over 10 signals per second. In practice, the median tone duration has been found to be 160 ms and the median interdigit gap to be 350 ms.

Consideration of human factors and mechanical design factors include aspects like button size and spacing, stroke length, strike force, numbering scheme and button arrangement. User preference and performance studies coupled with design considerations have resulted in the following specifications: 3/8-inch square buttons, separated by 1/4-inch, 1/8-inch stroke length; 100 g force at the bottom of the stroke; 4 × 3 array with the digits 1, 2, 3 in the top row and zero in the middle of the last row. The # sign in the third row (see Fig. 3.3) is usually used to redial the last dialled number. The push button is reserved for some special functions. The above specifications correspond to CCITT Q.23 recommendations.

A major advantage of touch tone dialling is the potential for data transmission and remote control. A powerful application of touch tone dialling is the **data in voice answer** (DIVA) system. A customer calling an airline may receive voice announcements like "dial 1 for reservation" and "dial 2 for flight information". Based on the voice announcements, the customer dials further digits which may result in further instruction for additional dialling. Thus, dialling and voice conversation can be interspersed to any level. This is a typical example of end-to-end signalling enabling interaction between a telephone user and a service provider.

3.3 Principles of Crossbar Switching

The basic idea of crossbar switching is to provide a matrix of $n \times m$ sets of contacts with only $n + m$ activators or less to select one of the $n \times m$ sets of contacts. This form of switching is also known as coordinate switching as the switching contacts are arranged in a *xy*-plane. A diagrammatic representation of a crosspoint switching matrix is shown in Fig. 3.6. There is an

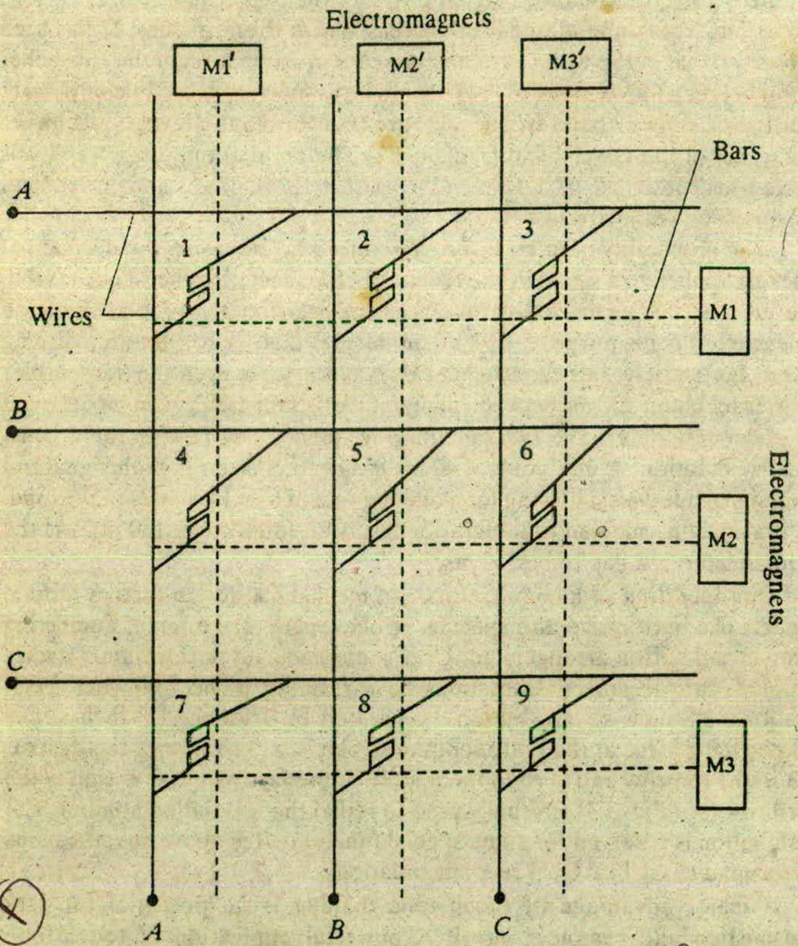


Fig. 3.6 3 × 3 crossbar switching.

array of horizontal and vertical wires shown by solid lines. A set of vertical and horizontal contact points are connected to these wires. The contact points form pairs, each pair consisting of a bank of three or four horizontal and a corresponding bank of vertical contact points. A contact point pair acts as a crosspoint switch and remains separated or open when not in use. The contact points are mechanically mounted (and electrically insulated) on a set of horizontal and vertical bars shown as dotted lines. The bars, in turn, are attached to a set of electromagnets.

When an electromagnet, say in the horizontal direction, is energised, the bar attached to it slightly rotates in such a way that the contact points attached to the bar move closer to its facing contact points but do not actually make any contact. Now, if an electromagnet in the vertical direction is energised, the corresponding bar rotates causing the contact points at the intersection of the two bars to close. This happens because the contact points move towards each other. As an example, if electromagnets M2 and M3′ are energised, a contact is established at the crosspoint 6 such that the subscriber *B* is connected to the subscriber *C*. In order to fully understand the working of the crossbar switching, let us consider a 6 × 6 crossbar schematic shown in Fig. 3.7. The schematic shows six subscribers with the horizantal bars representing the inlets and the vertical bars the outlets. Now consider the

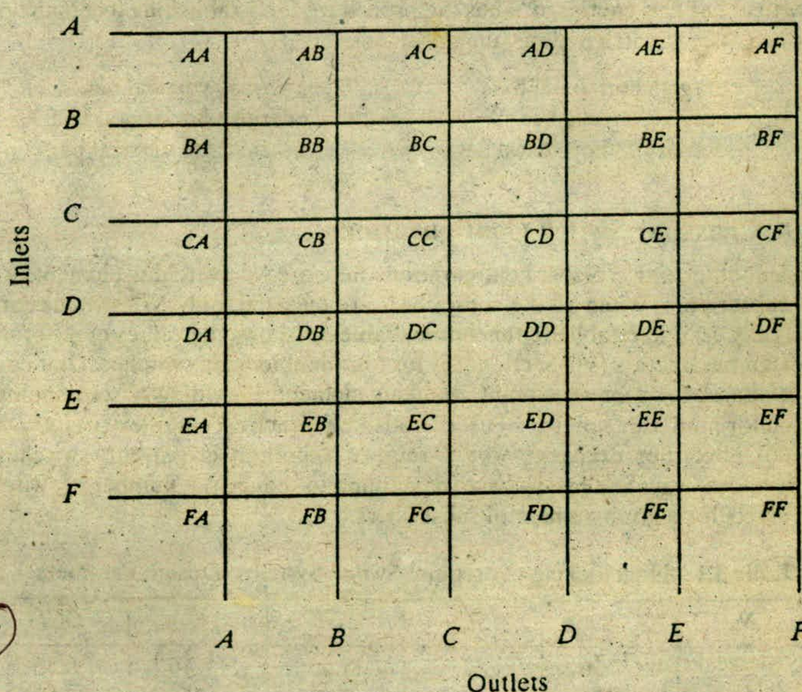


Fig. 3.7 6 × 6 crossbar matrix.

establishment of the following connections in sequence: *A* to *C* and *B* to *E*. First the horizontal bar *A* is energised. Then the vertical bar *C* is energised. The crosspoint *AC* is latched and the conversation between *A* and *C* can now proceed. Suppose we now energise the horizontal bar of *B* to establish the connection B-E, the crosspoint BC may latch and B will be brought into the circuit of A-C. This is prevented by introducing an energising sequence for latching the crosspoints. A crosspoint latches only if the horizontal bar is energised first and then the vertical bar. (The sequence may well be that the vertical bar is energised first and then the horizontal bar). Hence the crosspoint BC will not latch even though the vertical bar C is energised as the proper sequence is not maintained. In order to establish the connection B-E, the vertical bar E needs to be energised after the horizontal bar is energised. In this case, the crosspoint AE may latch as the horizontal bar A has already been energised for establishing the connection A-C. This should also be avoided and is done by de-energising the horizontal bar A after the crosspoint is latched and making a suitable arrangement such that the latch is maintained even though the energisation in the horizontal direction is withdrawn. The crosspoint remains latched as long as the vertical bar E remains energised. As the horizontal bar A is de-energised immediately after the crosspoint AC is latched, the crosspoint AE does not latch when the vertical bar E is energised. Thus the procedure for establishing a connection in a crossbar switch may be summarised as:

energise horizontal bar		energise vertical bar
energise vertical bar	or	energise horizontal bar
de-energise horizontal bar		de-energise vertical bar

3.4 Crossbar Switch Configurations

In a nonblocking crossbar configuration, there are N^2 switching elements for N subscribers. When all the subscribers are engaged, only $N/2$ switches are actually used to establish connections. Table 3.1 shows the values of different design parameters (see Section 2.5) for four nonblocking switches. Unit cost is assumed for each crosspoint switching element. Providing N^2 crosspoints even for moderate number of users leads to impractical complex circuitry. A 1000-subscriber exchange would require 1 million crosspoint switches. Therefore, ways and means have to be found to reduce the number of switch contacts for a given number of subscribers.

Table 3.1 Nonblocking Crosspoint Switch Systems: Design Parameters

N	S	SC	EUF	C	CCI
4	16	2	12.50	16	0.5
16	256	8	3.13	256	0.5
64	4096	32	0.78	4096	0.5
128	16384	64	0.39	16384	0.5

N = No. of subscribers S = No. of switching elements
SC = switching capacity C = total cost
EUF = equipment utilisation factor CCI = cost capacity index

It may be observed in the switch matrix of Fig. 3.7 that different switch points are used to establish a connection between two given subscribers, depending upon who initiates the call. For example, when the subscriber C wishes to call subscriber B, crosspoint CB is energised. On the other hand, when B initiates the call to contact C, the switch BC is used. By designing a suitable control mechanism, only one switch may be used to establish a connection between two subscribers, irrespective of which one of them initiates the call. In this case, the crosspoint matrix reduces to a diagonal matrix with $N^2/2$ switches. A diagonal connection matrix for 4 subscribers is shown in Fig. 3.8. The crosspoints in the diagonal connect the inlets and the

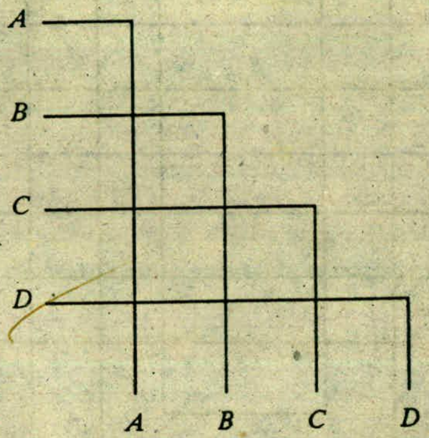


Fig. 3.8 Diagonal crosspoint matrix.

outlet of the same subscriber. This is not relevant. Hence, these are eliminated. The number of crosspoints then reduces to $N(N-1)/2$. It may be recalled that the quantity $N(N-1)/2$ represents the number of links in a fully connected network. So also, the diagonal crosspoint matrix is fully connected. The call establishment procedure here is dependent on the source and destination subscribers. When subscriber D initiates a call, his horizontal bar is energised first and then the appropriate vertical bar. If subscriber A initiates a call, the horizontal bar of the called party is activated first and then the vertical bar of A.

A diagonal crosspoint matrix is a nonblocking configuration. Even $N(N-1)/2$ crosspoint switches can be a very large number to handle practically. The number of crosspoint switches can be reduced significantly by designing blocking configurations. These configurations may be single stage or multistage switching networks. (Single stage configurations are discussed in this section and the multistage networks are dealt with in Chapter 4.) The crossbar hardware may be reduced by connecting two subscribers to a single bar and letting the bar turn both in the clockwise and the anticlockwise directions, and thus closing two different crosspoint contacts. With such an arrangement the number of crossbars reduces, but the number of crosspoint switches remains the same.

In blocking crossbar switches, the number of vertical bars is less than the number of subscribers and determines the number of simultaneous calls that can be put through the switch. Consider the 8 × 3 switch shown in Fig. 3.9.

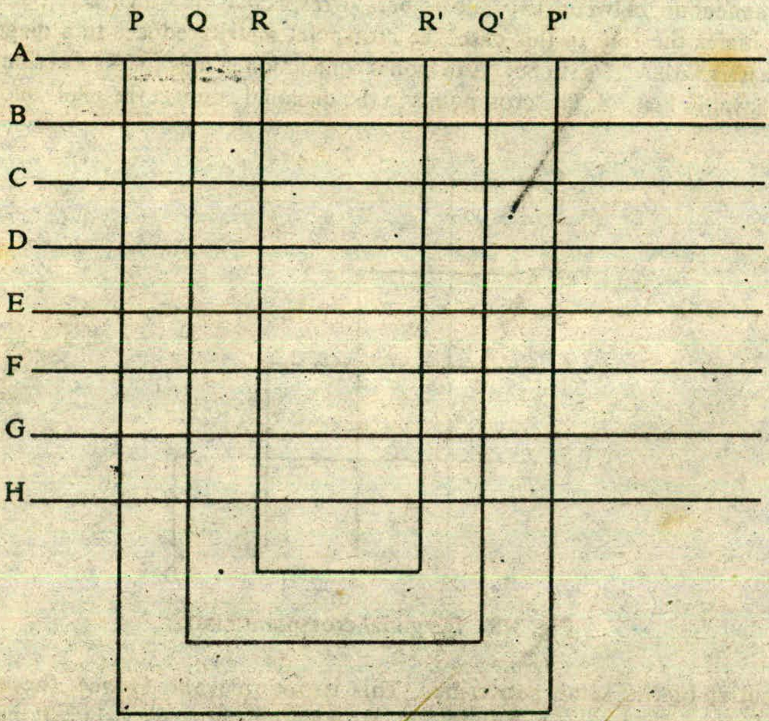


Fig. 3.9 Blocking crossbar switch.

Let a connection be required to be established between the subscribers A and B. First the horizontal bar A is energised. Then one of the free vertical bars, say P, is energised. The crosspoint AP latches. Now if we energise the horizontal bar B, BP will not be latched, as the P vertical is energised before B was energised. In order to be able to connect A to B, we need another vertical crossbar which should electrically correspond to the vertical bar P. In this case, the bar P' is associated with the same electrical wire as the bar P. When P' is energised after B, the crosspoint BP' is latched and a connection between A and B is established. The sequence to be followed in establishing the A-B circuit may be summarised as:

Energise horizontal	A
Energise free vertical	P

De-energise horizontal	A
Energise horizontal	B
Energise vertical	P'
De-energise horizontal	B

We thus see that in blocking crosspoint switches we need to operate four crossbars to establish a connection. The number of switches required is $2NK$, where N is the number of subscribers and K is the number of simultaneous circuits that can be supported. Another alternative to follow is a different sequence of energisation such that a contact is established with the use of only one vertical crossbar instead of two as described above:

Energise horizontal	A and B
Energise vertical	P
De-energise horizontals	A and B

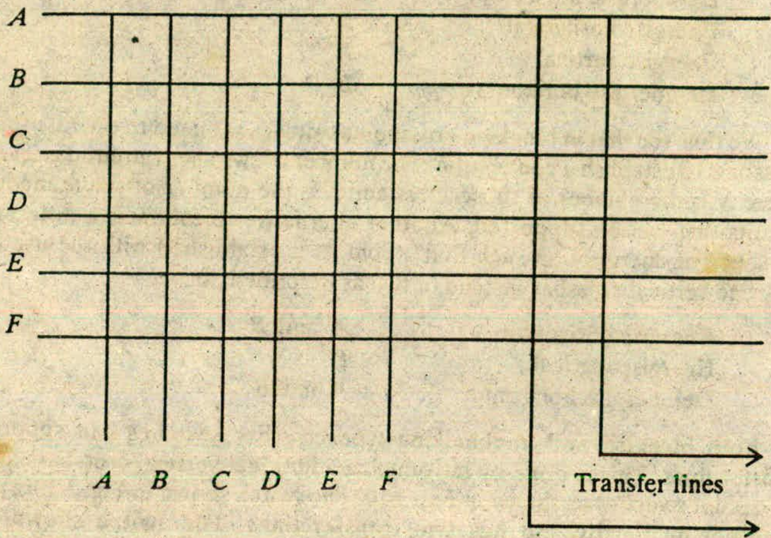
Both blocking and nonblocking type crossbar switches can support transfer lines. This is done by introducing additional vertical crossbars and crosspoint switches as shown in Fig. 3.10. The switch shown in Fig. 3.10(a) is nonblocking locally and has two transfer lines. The switch shown in Fig. 3.10(b) is blocking both locally and externally with two simultaneous local and two simultaneous external calls. The number of crosspoint switches in the first case is $N(N+L)$ and in the second case $N(2K+L)$, where N is the number of subscribers, L the number of transfer lines, and K is the number of simultaneous calls that can be supported locally.

3.5 Crosspoint Technology

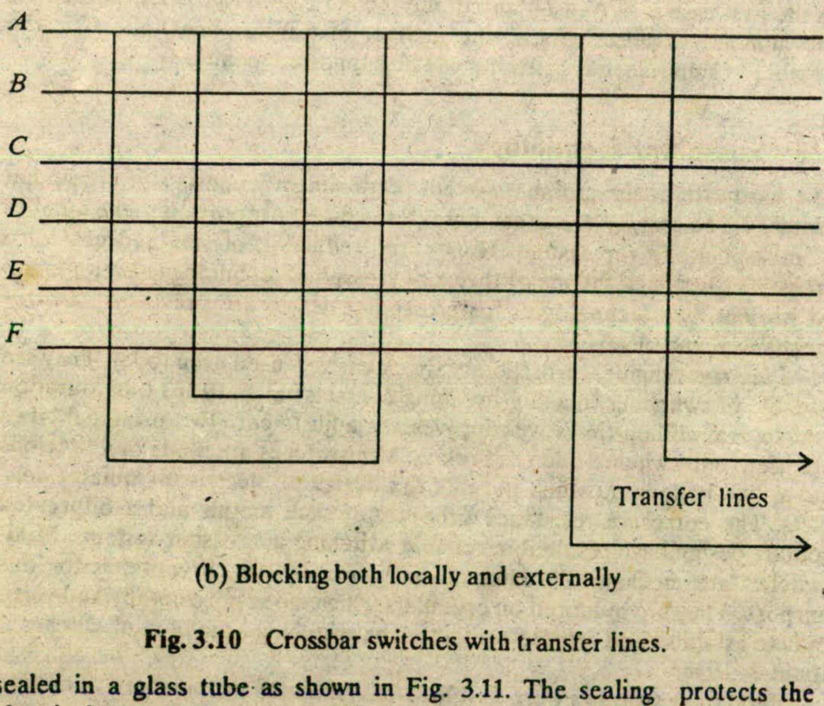
The hardware of the crossbar system predominantly consists of crosspoint switches. The cost of the system increases in direct proportion to the number of crosspoints in the system. Hence, the reduction of size and cost of a crosspoint has been the major thrust of crosspoint technology development. At present, two technologies for crosspoint design are prevalent: electromechanical and electronic.

Electromechanical crosspoints are extensively used even today. They are capable of switching (making/breaking contacts) in 1–10 ms time duration and several million times without wear or adjustment. Two principal types are used: miniswitches and reed relays. Miniswitches are made of a precious metal like palladium which permits the design of electrically quieter contacts. The corrosion resistance property of such metals and a bifurcated contact design have resulted in reliable switching in crossbar systems. Miniswitches are mechanically latched and generally use 'V' notches for this purpose. They are mounted on crossbars which move horizontally and vertically to establish and release contacts. The switching time of miniswitches is about 8–10 ms.

Reed relay switches were developed to eliminate the mechanical motion of bars in a crossbar system, thus increasing the operating life of the system. The reed relay comprises a pair of contacts made of a magnetic material



(a) Locally nonblocking and externally blocking



(b) Blocking both locally and externally

Fig. 3.10 Crossbar switches with transfer lines.

sealed in a glass tube as shown in Fig. 3.11. The sealing protects the electrical contacts from external contamination. The displacement involved in making contacts is about 0.2 mm. and this results in fast switching times

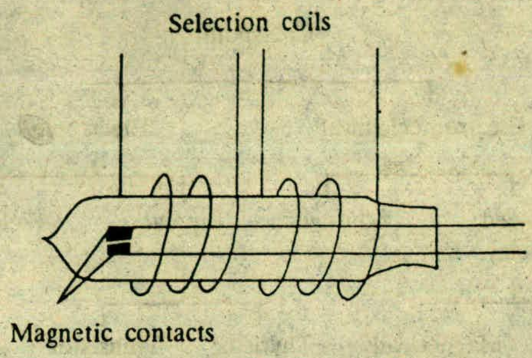


Fig. 3.11 Reed relay crosspoint.

less than 1 ms. A reed relay may be latched electrically or magnetically. The glass tube is surrounded by a pair of coils and when current is passed through both the coils simultaneously, a field is created, which causes the reed contacts to move together. When electrically latched, current is passed continuously through the coil as long as the switched connection is required. Magnetic latching relies on the hysteresis of the magnetic material. The pole pieces required for this purpose may be placed outside the glass tube, or the contacts themselves may be designed to act as poles by choosing an appropriate ferromagnetic material. In the latter case, the reed relay is called **remreed**, signifying remnance property of the contact strips. The residual magnetism in the poles keeps the contacts closed even after the currents are withdrawn from the coils. When a demagnetising current is applied to one or the other of the coils, the contacts open.

A crosspoint matrix is constructed by placing one reed relay at each crosspoint. Crosspoint selection is achieved by connecting one of the coil windings of each relay in series with its vertical neighbour, and the other winding in series with its horizontal neighbour. The required crosspoint is then selected by pulsing the appropriate vertical and horizontal circuits simultaneously. In practice, each reed relay contains a bank of 3 or 4 contacts, as is the case with miniswitches.

One of the first electronic devices to be tried out as a crosspoint is the cold cathode diode. This was soon abandoned because of the practical difficulties in implementation and inadequate transmission characteristics. With the advances in semiconductor technology, transistorised crosspoints were developed in the 1960s. They offered better performance than reed relays at that time but were not economically competitive. With the advent of integrated circuits, many private branch automatic exchanges (PABX) were designed using IC crosspoints. But with the arrival of time division switching technology (see Chapter 6), electronic crosspoints may never find extensive applications, particularly in large public exchanges. Figure 3.12 summarises the crosspoint technologies.

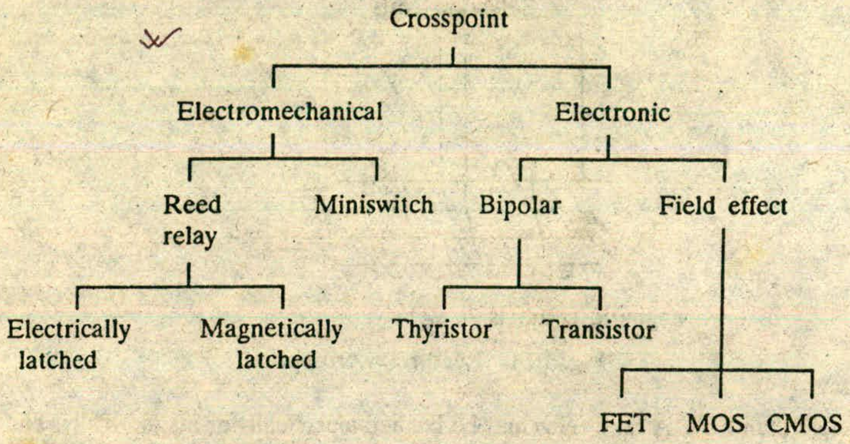


Fig. 3.12 Crosspoint technology.

3.6 Crossbar Exchange Organisation

The basic building blocks of a crossbar exchange are link frames, control markers and registers. Link frames consist of a number of crossbar switches arranged in two stages called primary and secondary with links between them as shown in Fig. 3.13. The two-stage arrangement with links has the effect of

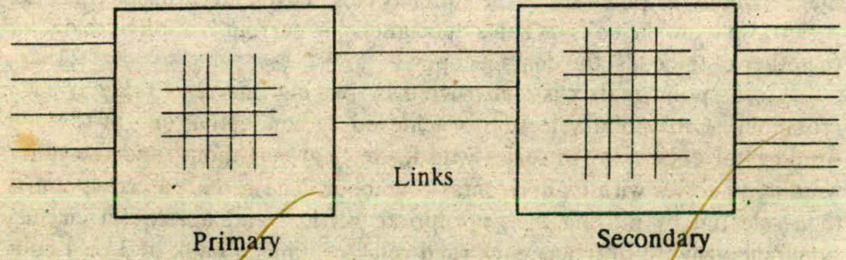


Fig. 3.13 A link frame and its control by a marker.

increasing the number of outlets for a given number of inlets, thereby providing greater selectivity. The switch, in this case, is said to be expanding. Markers control the connections between the inlets and the outlets via the primary section, links and the secondary section. Two-stage networks are discussed in detail in Chapter 4.

A simplified organisation of a crossbar exchange is shown in Fig. 3.14. The line link frames along with the associated markers and registers are known as line unit, and the trunk link frame with its associated hardware as group unit. The trunk link frame may be subdivided into two or three link frames like local office link frame, incoming link frame, etc. Line units are

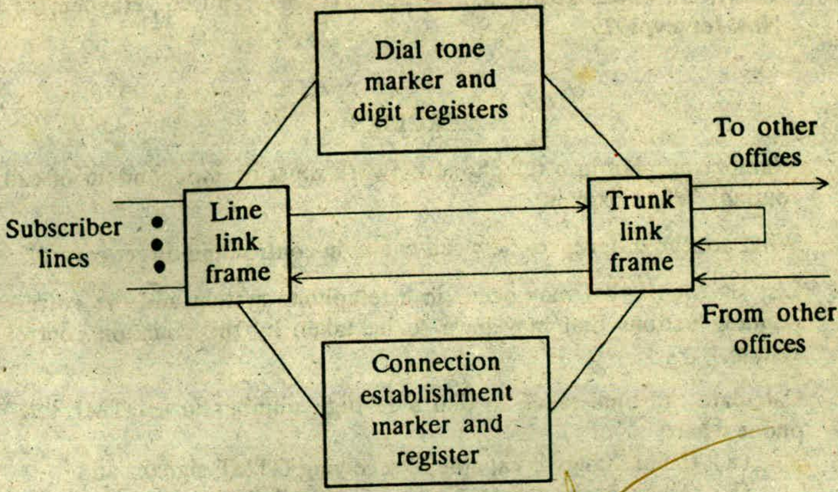


Fig. 3.14 Organisation of a crossbar exchange.

two-way units, that is to say, they can be used for originating as well as terminating calls. It may be noted that this is a significant departure from the Strowger exchange designs where the originating and terminating units are separate and independent. Because of its two-way capability, the secondary section in the line link frame is sometimes called the terminal section. The subscriber lines are terminated on the outlets of the terminal section frames. The group unit is a unidirectional device. It receives the calls from the line unit or from distant exchanges. It routes the calls to the unit of the same exchange or to distant exchanges. It is capable of handling local, outgoing, incoming, terminating and transit calls.

In a crossbar exchange, the call processing progresses in three stages: preselection, group selection, and line selection. Preselection, which is performed by the originating marker, starts from the moment the subscriber lifts the handset of the telephone and ends when the dial tone is sent out to him by a register. In group selection stage, the call is switched through to the desired direction. The direction is decided in accordance with the code given by the translator as described in Section 3.1. In the last stage, the calling subscriber is connected to the called subscriber by the terminating marker. The line of the called party is controlled by the terminating marker which also sets up ringing on the line.

FURTHER READING

1. AT&T, *Switching Systems*, 1961.
2. Chapnis, R. J., *100 Years of Telephone Switching (1878–1978), Part 1: Manual and Electromechanical Switching*, (*North Holland Studies in Telecommunications, Vol. 1*), North Holland, Amsterdam, 1982.

3. Talley, D., *Basic Telephone Switching Systems*, 2nd ed., Hayden Inc., New Jersey, 1979.

EXERCISES

1. 'Numbering plan in a telephone network must be independent of call routing'. Why? Explain.
2. What are the differences between common control and direct control?
3. List six events that may occur in a telephone system and the corresponding actions that may have to be taken by the common control system.
4. Calculate the time taken to dial a 12-digit number in a DTMF telephone when
   (a) the exchange is capable of receiving DTMF signals; and
   (b) the exchange can receive only pulse dialling.
   Compare the result with a rotary telephone dialling.
5. 'Contact bounce' can be a problem in DTMF telephone, i.e. a single press of a push button may be interpreted as more than one press. How does the DTMF dial design take this into account?
6. Show that the harmonic frequencies of any two adjacent base frequencies in DTMF telephone cannot match within the first 15 harmonics.
7. If the transmitted power of the low band frequency signal from a DTMF telephone is 1 mW, what should be the power in mW of the high band frequencies?
8. A telephone exchange supporting 5000 subscribers uses DTMF dialling and a common control subsystem with 100 digit receivers. Each digit receiver is assigned for a duration of five seconds per subscriber for call processing. If 20 per cent of the subscribers attempt to call simultaneously, what is the worst case wait time for a subscriber before he receives the dial tone?
9. A diagonal crosspoint matrix exchange supports 500 users. On an average 1000 calls are put through everyday. If the crosspoint contacts have a mean life of 10000 breaks and makes, estimate as to how often a crosspoint may be replaced in this exchange.
10. Estimate the number of crosspoints required to design an exchange that supports 500 users on a nonblocking basis and 50 transit, outgoing or incoming calls simultaneously.
11. Compare the reliabilities of one transistorised crosspoint switch and a bipolar chip containing 100 crosspoint switches. (Use known reliability data for the two technologies).

12. "The number of crossbars may be reduced by mounting contacts belonging to two subscribers on one bar". Can this be applied to both horizontal and vertical bars simultaneously? Explain how the scheme would work.

13. A blocking crossbar switch is to be designed to support 1000 subscribers. If the estimated peak traffic is 10 erlangs with average holding times of three minutes per call, estimate the number of crosspoints required.

4

Electronic Space Division Switching

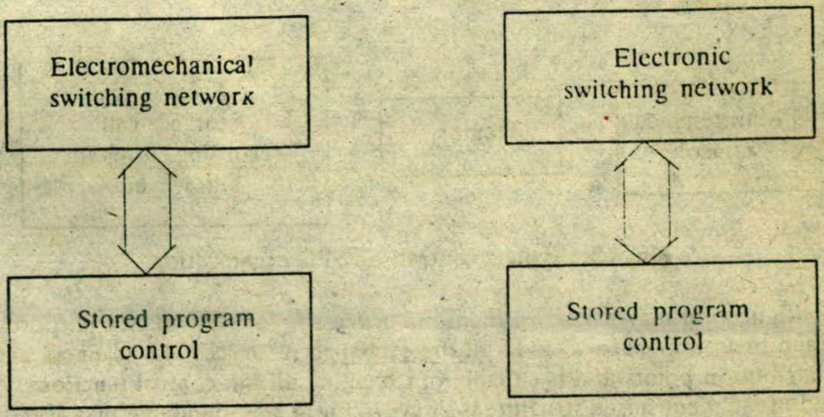
Early crossbar systems were slow in call processing as they used electromechanical components for common control subsystems. Efforts to improve the speed of control and signalling between exchanges led to the application of electronics in the design of control and signalling subsystems. In late 1940s and early 1950s, a number of developmental efforts made use of vacuum tubes, transistors, gas diodes, magnetic drums and cathode ray tubes for realising control functions. Circuits using gas tubes were developed and employed for timing, ring translation and selective ringing of party lines. Vacuum tubes were used in single frequency signalling and transistors in line insulation test circuits. Contemporary to these developments was the arrival of modern electronic digital computers. Switching engineers soon realised that, in principle, the registers and translators of the common control systems could be replaced by a single digital computer.

4.1 Stored Program Control

Modern digital computers use the stored program concept. Here, a program or a set of instructions to the computer is stored in its memory and the instructions are executed automatically one by one by the processor. Carrying out the exchange control functions through programs stored in the memory of a computer led to the nomenclature **stored program control** (SPC). An immediate consequence of program control is the full-scale automation of exchange functions and the introduction of a variety of new services to users. Common channel signalling (CCS), centralised maintenance and automatic fault diagnosis, and interactive human-machine interface are some of the features that have become possible due to the application of SPC to telephone switching.

Introducing a computer to carry out the control functions of a telephone exchange is not as simple as using a computer for scientific or commercial data processing. A telephone exchange must operate without interruption, 24 hours a day, 365 days a year and for say, 30–40 years. This means that the computer controlling the exchange must be highly tolerant to faults. Fault tolerant features were unknown to early commercial computers and the switching engineers were faced with the task of developing fault tolerant hardware and software systems. In fact, major contributions to fault tolerant computing have come from the field of telecommunication switching.

Attempts to introduce electronics and computers in the control subsystem of an exchange were encouraging enough to spur the development of full-fledged electronic switching system, in which the switching network is also electronic. After about 10 years of developmental efforts and field trials, the world's first electronic switching system, known as No.1 ESS, was commissioned by AT&T at Succasunna, New Jersey, in May 1965. Since then, the history of electronic switching system and stored program control has been one of rapid and continuous growth in versatility and range of services. Today, SPC is a standard feature in all the electronic exchanges. However, attempts to replace the space division electromechanical switching matrices by semiconductor crosspoint matrices have not been greatly successful, particularly in large exchanges, and the switching engineers have been forced to return to electromechanical miniature crossbars and reed relays, but with a complete electronic environment. As a result, many space division electronic switching systems use electromechanical switching networks with SPC. Nonetheless, private automatic branch exchanges (PABX) and smaller exchanges do use electronic switching devices. The two types of space division electronic switching systems, one using electromechanical switching network and the other using electronic switching network, are depicted in Fig. 4.1. Both the types qualify as electronic switching systems although only one of them is fully electronic. With the evolution of time division switching, which is done in the electronic domain, modern exchanges are fully electronic. Principles of time division switched electronic exchanges are discussed in Chapter 6.



(a) Electromechanical switching (b) Electronic switching

Fig. 4.1 Electronic space division switching systems.

There are basically two approaches to organising stored program control: centralised and distributed. Early electronic switching systems (ESS) developed during the period 1970–75 almost invariably used centralised control. Although many present day exchange designs continue to use centralised SPC, with the advent of low cost powerful microprocessors and very large scale integration (VLSI) chips such as programmable logic arrays (PLA) and programmable logic controllers (PLC), distributed SPC is gaining popularity.

4.2 Centralised SPC

In centralised control, all the control equipment is replaced by a single processor which must be quite powerful. It must be capable of processing 10 to 100 calls per second, depending on the load on the system, and simultaneously performing many other ancillary tasks. A typical control configuration of an ESS using centralised SPC is shown in Fig. 4.2. A centralised SPC

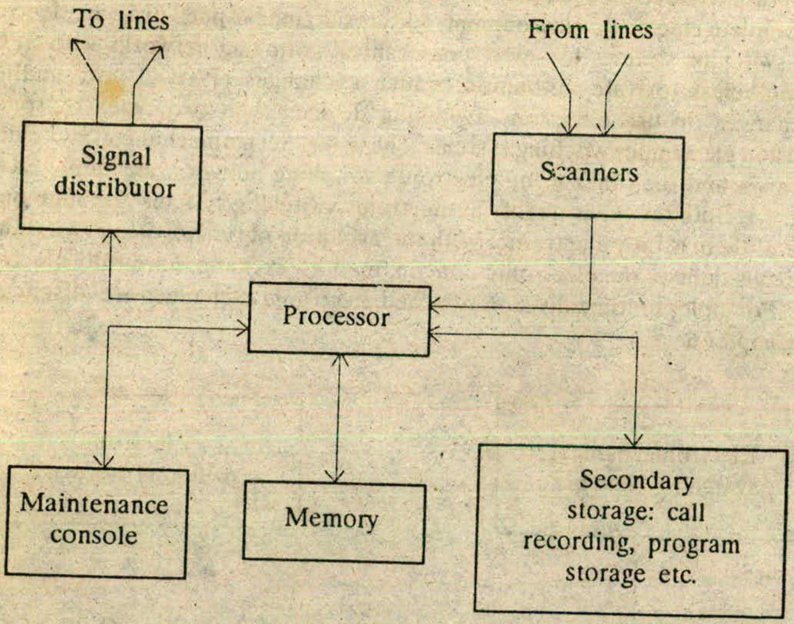


Fig. 4.2 Typical centralised SPC organisation.

configuration may use more than one processor for redundancy purposes. Each processor has access to all the exchange resources like scanners and distribution points and is capable of executing all the control functions. A redundant centralised structure is shown in Fig. 4.3. Redundancy may also be provided at the level of exchange resources and function programs. In actual implementation, the exchange resources and the memory modules contain-

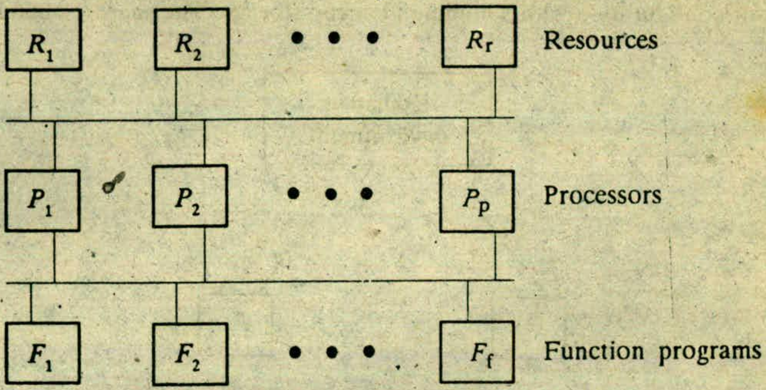


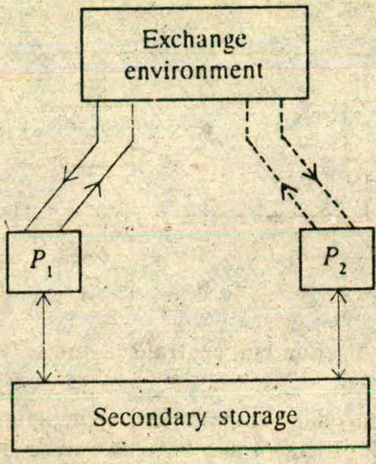
Fig. 4.3 A redundant centralised control structure.

ing the programs for carrying out the various control functions may be shared by processors, or each processor may have its own dedicated access paths to exchange resources and its own copy of programs and data in dedicated memory modules.

In almost all the present day electronic switching systems using centralised control, only a two-processor configuration is used. A dual processor architecture may be configured to operate in one of three modes:

1. Standby mode
2. Synchronous duplex mode
3. Load sharing mode.

Standby mode of operation is the simplest of dual processor configuration operations. Normally, one processor is active and the other is on standby, both hardware and software wise. The standby processor is brought online only when the active processor fails. An important requirement of this configuration is the ability of the standby processor to reconstitute the state of the exchange system when it takes over the control, i.e. to determine which of the subscribers and trunks are busy or free, which of the paths are connected through the switching network etc. In small exchanges, this may be possible by scanning all the status signals as soon as the standby processor is brought into operation. In such a case, only the calls which are being established at the time of failure of the active processor are disturbed. In large exchanges, it is not possible to scan all the status signals within a reasonable time. Here, the active processor copies the status of the system periodically, say every five seconds, into a secondary storage. When a switchover occurs, the online processor loads the most recent update of the system status from the secondary storage and continues the operation. In this case, only the calls which changed status between the last update and the failure of the active processor are disturbed. Figure 4.4 shows a standby dual processor configuration with a common backup storage. The shared secondary



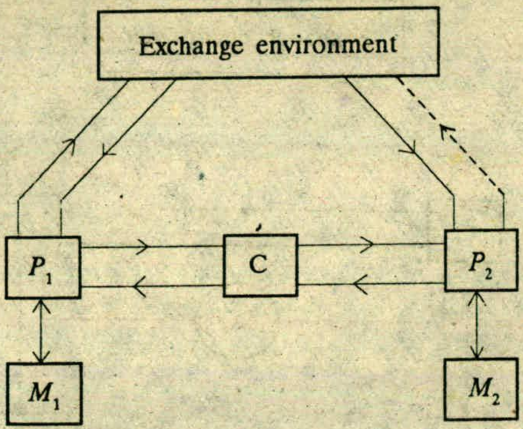
P_1 = active processor P_2 = standby processor

Fig. 4.4 Standby dual processor configuration.

storage need not be duplicated and simple unit-level redundancy would suffice.

In synchronous duplex mode of operation, hardware coupling is provided between the two processors which execute the same set of instructions and compare the results continuously. If a mismatch occurs, the faulty processor is identified and taken out of service within a few milliseconds. When the system is operating normally, the two processors have the same data in their memories at all times and simultaneously receive all information from the exchange environment. One of the processors actually controls the exchange, whereas the other is synchronised with the former but does not participate in the exchange control. The synchronously operating configuration is shown in Fig. 4.5. If a fault is detected by the comparator, the two processors P_1 and P_2 are decoupled and a check-out program is run independently on each of the machines to determine which one is faulty. The check-out program runs without disturbing the call processing which is suspended temporarily. When a processor is taken out of service on account of a failure or for maintenance, the other processor operates independently. When a faulty processor is repaired and brought into service, the memory contents of the active processor are copied into its memory, it is brought into synchronous operation with the active processor and then the comparator is enabled.

It is possible that a comparator fault occurs on account of a transient failure which does not show up when the check-out program is run. In such



C = comparator M = memory P = processor

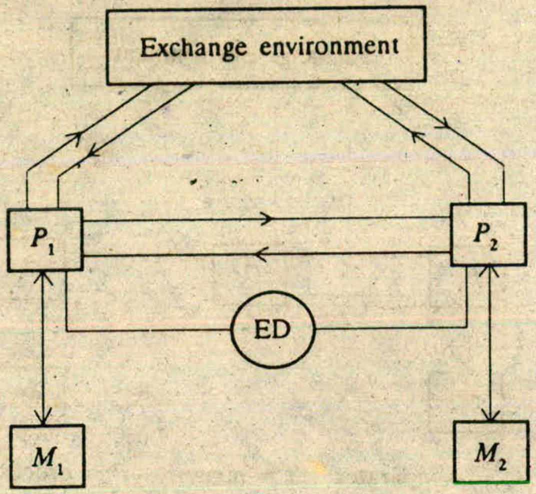
Fig. 4.5 Synchronous duplex operation.

cases, the decision as to how to continue the operation is arbitrary and three possibilities exist:

1. Continue with both the processors.
2. Take out the active processor and continue with the other processor.
3. Continue with the active processor but remove the other processor from service.

Strategy 1 is based on the assumption that the fault is a transient one and may not reappear. Many times the transient faults are the forerunners of an impending permanent fault which can be detected by an exhaustive diagnostic test of the processor under marginal voltage, current and temperature conditions. Strategies 2 and 3 are based on this hypothesis. The processor that is taken out of service is subjected to extensive testing to identify a marginal failure in these cases. A decision to use strategy 2 or 3 is somewhat arbitrary.

In **load sharing** operation, an incoming call is assigned randomly or in a predetermined order to one of the processors which then handles the call right through completion. Thus, both the processors are active simultaneously and share the load and the resources dynamically. The configuration is shown in Fig. 4.6. Both the processors have access to the entire exchange environment which is sensed as well as controlled by these processors. Since the calls are handled independently by the processors, they have separate memories for storing temporary call data. Although programs and semipermanent data can be shared, they are kept in separate memories for redundancy purposes. There is an interprocessor link through which the



ED = exclusion device

Fig. 4.6 Load sharing configuration.

processors exchange information needed for mutual coordination and verifying the 'state of health' of the other. If the exchange of information fails, one of the processors which detects the same takes over the entire load including the calls that are already set up by the failing processor. However, the calls that were being established by the failing processor are usually lost. Sharing of resources calls for an exclusion mechanism so that both the processors do not seek the same resource at the same time. The mechanism may be implemented in software or hardware or both. Figure 4.6 shows a hardware exclusion device which, when set by one of the processors, prohibits access to a particular resource by the other processor until it is reset by the first processor. Software exclusion mechanism is discussed in detail in Section 4.4.

Under normal operation, each processor handles one-half of the calls on a statistical basis. The exchange operators can, however, send commands to split the traffic unevenly between the two processors. This may be done, for example, to test a software modification on one processor at low traffic, while the other handles majority of the calls. Load sharing configuration gives much better performance in the presence of traffic overloads as compared to other operating modes, since the capacities of both the processors are available to handle overloads. Load sharing configuration increases the effective traffic capacity by about 30 per cent when compared to synchronous duplex operation. Load sharing is a step towards distributed control.

One of the main purposes of redundant configuration is to increase the overall availability of the system. A telephone exchange must show more or less a continuous availability over a period of perhaps 30 or 40 years. We now compare the availability figures of a single processor and a dual processor system. The availability of a single processor system is given by

$$A = \frac{MTBF}{MTBF + MTTR} \tag{4.1}$$

where

$MTBF$ = mean time between failure

$MTTR$ = mean time to repair

The unavailability of the system is given by

$$U = 1 - A$$

$$= 1 - \frac{MTBF}{MTBF + MTTR}$$

$$= \frac{MTTR}{MTBF + MTTR} \tag{4.2}$$

If $MTBF \gg MTTR$, then

$$U = \frac{MTTR}{MTBF} \tag{4.3}$$

For a dual processor system, the mean time between failures, $MTBF_D$, can be computed from the $MTBF$ and $MTTR$ values of the individual processors. A dual processor system is said to have failed only when both the processors fail and the system is totally unavailable. Such a situation arises only when one of the processors has failed and the second processor also fails when the first one is being repaired. In other words, this is related to the conditional probability that the second processor fails during the $MTTR$ period of the first processor when the first processor has already failed. Without going into the detailed derivations, we just state the result for $MTBF_D$ *as*

$$MTBF_D = \frac{(MTBF)^2}{2\,MTTR} \tag{4.4}$$

Therefore, the availability of the dual processor system, A_D, is given by

$$A_D = \frac{MTBF_D}{MTBF_D + MTTR}$$

$$= \frac{(MTBF)^2/2\,MTTR}{[(MTBF)^2/2\,MTTR] + MTTR}$$

$$= \frac{(MTBF)^2}{(MTBF)^2 + 2(MTTR)^2} \tag{4.5}$$

Therefore, the unavailability U_D is

$$U_D = 1 - A_D$$

$$= 1 - \frac{(MTBF)^2}{(MTBF)^2 + 2(MTTR)^2}$$

$$= \frac{2(MTTR)^2}{(MTBF)^2 + 2(MTTR)^2}$$

If $MTBF \gg MTTR$, then we have

$$U_D = \frac{2(MTTR)^2}{(MTBF)^2} \tag{4.6}$$

EXAMPLE 4.1 Given that $MTBF$ = 2000 hours and $MTTR$ = 4 hours, calculate the unavailability for single and dual processor systems.

Solution

$$U = 4/2000 = 2 \times 10^{-3}$$

i.e. 525 hours in 30 years.

$$U_D = 2 \times 16/2000 \times 2000 = 8 \times 10^{-6}$$

i.e. 2.1 hours in 30 years.

As discussed in Section 3.1, event monitoring, call processing, charging and operation and maintenance (O&M) are the four important functions of a control subsystem in an exchange. Considering the real time response requirements, these functions may be grouped under three levels as shown in Fig. 4.7. Event monitoring has the highest real time constraint and the O&M

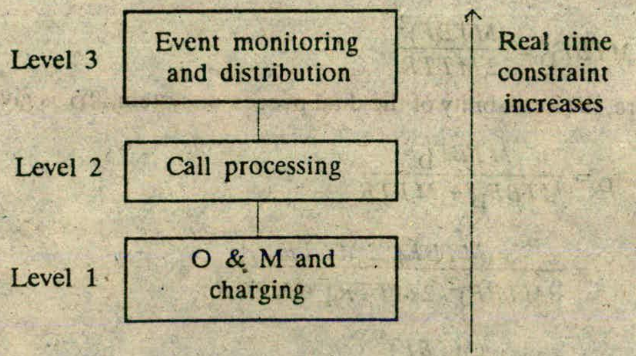


Fig. 4.7 Levels of control functions.

and charging the least. The real time constraint necessitates a priority interrupt facility for processing in centralised control. If an event occurs when O&M function is being carried out by the control processor, the O&M processing has to be interrupted, the event processing taken up and completed, and then the O&M function processing resumed. Nesting of interrupts is necessary to suspend any low level function and to take up the processing of higher level functions as shown in Fig. 4.8. When an interrupt

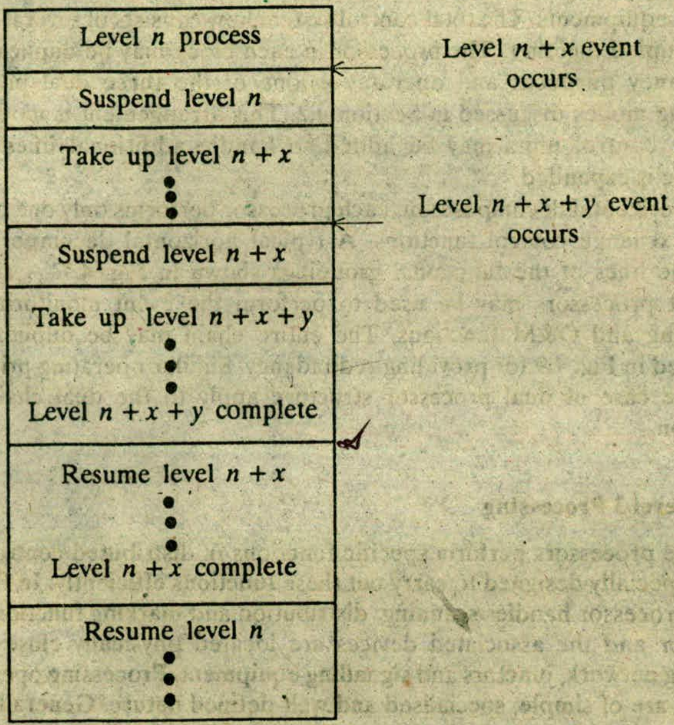


Fig. 4.8 Interrupt processing.

occurs, program execution is shifted to an appropriate service routine address in the memory through a branch operation. There are two methods of accomplishing this. One is called **vectored** interrupt and the other **non-vectored** interrupt. In nonvectored interrupt, the branch address is fixed and a main interrupt service routine scans the interrupt signals and decides on the appropriate routine to service the specific interrupt. In vectored interrupt, the interrupting source supplies the branch address information to the processor. The set of addresses supplied by different interrupting sources is known as **interrupt vector**. Obviously, vectored interrupt is faster in response than the nonvectored interrupt.

4.3 Distributed SPC

In distributed control, the control functions are shared by many processors within the exchange itself. This type of structure owes its existence to the low cost microprocessors. This structure offers better availability and reliability than the centralised SPC.

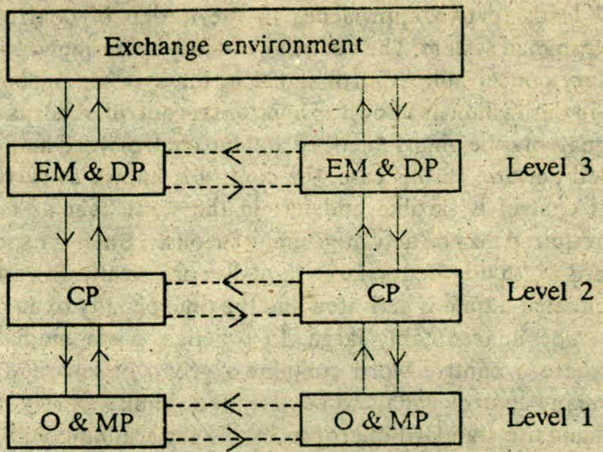
Exchange control functions may be decomposed either 'horizontally' or 'vertically' for distributed processing. In vertical decomposition, the exchange environment is divided into several blocks and each block is assigned to a processor that performs all control functions related to that block of equipments. The total control system now consists of several control units coupled together. The processor in each block may be duplicated for redundancy purposes and operates in one of the three dual processor operating modes discussed in Section 4.2. This arrangement is modular so that the control units may be added to handle additional lines as the exchange is expanded.

In horizontal decomposition, each processor performs only one or some of the exchange control functions. A typical horizontal decomposition is along the lines of the functional groupings shown in Fig. 4.7. A chain of different processors may be used to perform the event monitoring, call processing and O&M functions. The entire chain may be duplicated as illustrated in Fig. 4.9 for providing redundancy. Similar operating principles as in the case of dual processor structure apply to the dual chain configuration.

4.3.1 Level 3 Processing

Since the processors perform specific functions in distributed control, they can be specially designed to carry out these functions efficiently. In Fig. 4.9, level 3 processor handles scanning, distribution and marking functions. The processor and the associated devices are located physically close to the switching network, junctors and signalling equipment. Processing operations involved are of simple, specialised and well-defined nature. Generally, processing at this level results in the setting or sensing of one or more binary conditions in flipflops or registers. It may be necessary to sense and alter a set of binary conditions in a predefined sequence to accomplish a control function. Such simple operations are efficiently performed either by wired logic or microprogrammed devices.

A control unit, designed as a collection of logic circuits using logic elements, electronic or otherwise, is called a 'hard-wired' control unit. A hard-wired unit can be exactly tailored to the job in hand, both in terms of the function and the necessary processing capacity. But it lacks flexibility and cannot be easily adapted to new requirements. A microprogrammed unit is more universal and can be put to many different uses by simply modifying the microprogram and the associated data. With the same technology, the microprogrammed units tend to be more expensive and slower than hard-wired



CP = call processor EM&DP = event monitoring and distribution processor O&MP = operation and maintenance processor

Fig. 4.9 Dual chain distributed control.

units for an equivalent processing capacity. When the processing is complex, microprogramming implementation is easier. Table 4.1 summarises the characteristics of microprogrammed and hard-wired control. With the advent of low cost microprocessors and VLSI programmable logic arrays and controllers, microprogramming is the favoured choice for level 3 processing.

Table 4.1 Characteristics of Electronic Control Schemes

Microprogrammed control	Hard-wired control
Flexible	Not flexible
Slower	Faster
More expensive for moderate processing functions	Less expensive for moderate simple and fixed processing
Easier to implement complex processing functions	Difficult to implement complex functions
Introducing new services is easy	Not easily possible
Easier to maintain	Difficult to maintain

In microprogramming, the binary conditions required for control functions are altered through a control word which contains a bit pattern that activates the appropriate control signals. By storing a set of control words in a memory and reading them out one after another, control signals may be activated in the required sequence. Recognition of this fundamental aspect of control leads to two approaches to the design of control word in a microprogrammed system. The control word may be designed to contain one bit per every conceivable control signal in the system. A control scheme organised in this fashion is known as **horizontal control**. Alternatively, all the control signals may be binary encoded and the control word may contain only the encoded pattern. In this case, the control is known as **vertical control**. Horizontal control is flexible and fast in the sense that as many control signals as required may be activated simultaneously. But it is expensive as the control word width may be too large to realise practically. In vertical control only one signal at a time is activated and the time penalty to activate a set of signals may be unacceptably large. In practice, a via media solution is adopted where a control word contains a group of encoded words that permit as many control signals to be activated simultaneously. Some of the recent designs use standard microprocessors for scanning and distribution functions instead of designing a microprogrammed unit. The microprocessor based design is somewhat slower than the microprogrammed unit, and the latter is likely to dominate until low cost custom ICs for these functions become available.

4.3.2 Level 2 Processing

The processors employed for call processing in level 2 of Fig. 4.9 have, in most cases, been specially developed for this purpose in the past. Level 2 processor is usually termed as **switching processor**. Early general purpose computers were ill suited to real time applications and were large in size and expensive. With the arrival of minicomputers and then microprocessors, a number of real time applications outside the field of telecommunications have sprung up. This, in turn, has led to the appearance of standard processors suitable for real time applications in the market. Nonetheless, the exchange manufacturers have continued to prefer house-developed switching processors for some time in order to maintain full control over the products and to contain the costs. Of late, however, the trend is to employ commercially available standard microprocessors for the switching processor functions.

Switching processors are not fundamentally different from general purpose digital computers. There are, however, certain characteristics that are specific to switching processors, as in the case of processors employed in process control or other industrial real time applications. Processor instructions, for instance, are designed to allow data to be packed more tightly in memory without unduly increasing the access time. Single bit and half-byte manipulation instructions are used extensively in switching applications. Special instructions for task and event queue management, which would enable optimal run times for certain scheduler functions, are desirable. The architecture of switching processors is designed to ensure over 99.9% availability, fault tolerance and security of operation. In the input/output (I/O) area, the switching processors differ from general purpose computers, mainly on account of the existence of telephone peripherals such as scanners, distributors and markers along with the conventional data processing type peripherals like teleprinters, magnetic tapes etc. The total I/O data transfer is not very high in switching processors and is of the order of 100 kilobytes per second for large systems. Both program controlled data transfer and direct memory access (DMA) techniques are used for I/O data transfer. Sometimes, the exchange peripherals are located far away from the switching processor; consequently, special communication links are required to connect them to the I/O controller.

The traffic handling capacity of the control equipment is usually limited by the capacity of the switching processor. The load on the switching processor is measured by its occupancy t, estimated by the simple formula

$$t = a + bN \tag{4.7}$$

where

a = fixed overhead depending upon the exchange capacity and configuration
b = average time to process one call
N = number of calls per unit time

The occupancy t is expressed as a fraction of the unit time for which the processor is occupied. The parameter a depends to a large extent on the scanning workload which, in turn, depends usually on the number of subscriber lines, trunks and service circuits in the exchange. The parameter value may be estimated by knowing the total number of lines, the number of instructions required to scan one line, and the average execution time per instruction. The estimation of the value of the parameter b requires the definition of call mix, comprising incoming, outgoing, local and transit calls. This is because the number of instructions required to process each type of call varies considerably. For example, the number of instructions required to process an incoming call where there is no need to retransmit the address digits is much less than the number required to process a transit call. The result of a call attempt such as call put through, called party busy or no answer also affects the number of instructions to be executed. The number of subscribers with DTMF and rotary dial telephones and the percentage of calls to grouped (PBX) lines are also important factors. Taking these factors into account, a call mix may be worked out and the mean processing time per call attempt calculated, by taking the weighted average of the processing times for various types of calls.

Usually, the switching processor is designed to handle a traffic load which is 40% higher than the nominal load. When this overload occurs, the processor may be loaded only to 95% of its capacity so that traffic fluctuations can be absorbed. Such a consideration renders Eq. (4.7) as

$$0.95 = a + 1.4bN_N$$

where

N_N = nominal load in terms of number of calls per unit time

or

$$N_N = \frac{0.95 - a}{1.4b} \tag{4.8}$$

The average instruction execution time is dependent on the instruction mix as different instructions take different times. The best way to evaluate a switching processor is to prepare a benchmark comprising representative call mix and measure the actual processing time under this load.

4.3.3 Level 1 Processing

The level 1 control handles operations and maintenance (O&M) functions which involves the following steps:

- Administer the exchange hardware and software.
- Add, modify or delete information in translation tables.
- Change subscriber class of service.
- Put a new line or trunk into operation.
- Supervise operation of the exchange.
- Monitor traffic.
- Detect and locate faults and errors.
- Run diagnostic and test programs.
- Man-machine interaction.

The complex nature of the functions demands a large configuration for the level 1 computer involving large disk or tape storage. As a result, O&M processor in many cases is a standard general purpose computer, usually a mainframe. The complexity and volume of the software are also the highest when compared to level 2 and 3 processing. The O&M functions are less subject to real time constraints and have less need for concurrent processing. Hence, it is a common practice that a single O&M computer is shared among several exchanges located remotely as shown in Fig. 4.10. In such an arrangement, the exchanges contain only the level 2 and level 3 processing modules. Remote diagnosis and maintenance permit expert maintenance personnel to attend to several exchanges from one central location.

The three-level distributed control discussed in this section, though typical and popular, is by no means the only method of distributing control functions. Many exchange designs use a single computer located physically at

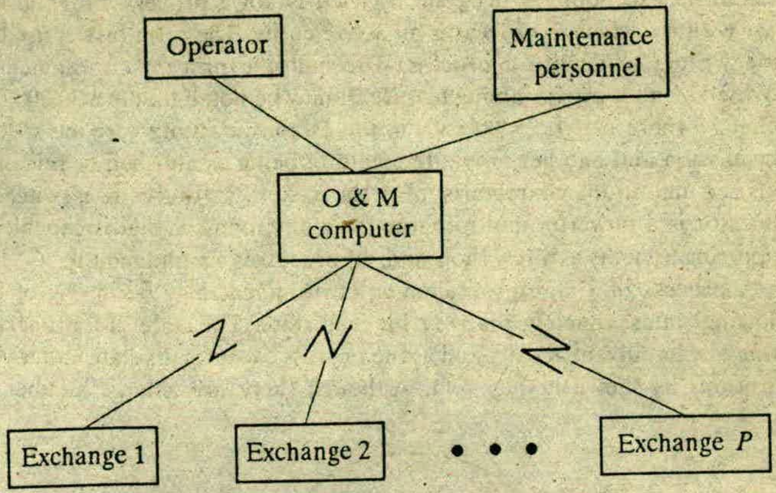


Fig. 4.10 Remote operation and maintenance.

the exchange site, to perform both the O&M and call processing functions. Some designs may use two different processors for O&M and call processing, but may not resort to remote O&M. Instead, they may use one dedicated O&M processor for each exchange.

4.4 Software Architecture

As is the case with general purpose computers, the software of the SPC systems may be placed under two broad categories: system software and application software. Software architecture deals with the system software environment, including the language processors. As stated in Section 4.3.3, O&M functions are usually carried out by a general purpose computer, and the software architecture required for this purpose is similar to that of a general data processing system. Call processing, being specific to switching systems and demanding real time responses, requires a software system with special features. In this book, we discuss the SPC software architecture only in the context of call processing functions. But many of the features discussed are also a part of the operating system under which O&M functions are carried out. Application software details are covered in Section 4.5.

Call processing is an event oriented processing function. It is triggered by an event occurring at a subscriber line or trunk. Call setup is not done in one continuous processing sequence in the exchange. Instead, it involves several elementary processing actions, each lasting a few tens or hundreds of milliseconds, separated by periods of waiting for external events. Sometimes, these waiting periods may be as long as 20 seconds. Hence, as far as the processor is concerned, many calls are processed simultaneously, with each call being handled by a separate **process**. A process, in this context, is a program in execution. A program by itself is not a process. Program is a passive entity, whereas process is an active entity. The term **task** is used by some writers to denote a process. The multiple process environment is provided by multiprogramming feature. It may be noted that in a 30,000 line exchange, there may be 3000 calls in the established state carrying speech transmission and another 500 in the state of being established or released. Thus, an important characteristic of the system software of a switching processor is a powerful multiprogramming environment that is capable of supporting as many as a few thousands of processes simultaneously.

A process in a multiprogramming environment may be in one of the following states: running, ready or blocked state. The state of a process is defined by its current activity and as the process executes, its state undergoes transitions as shown in Fig. 4.11. Although there are a large number of

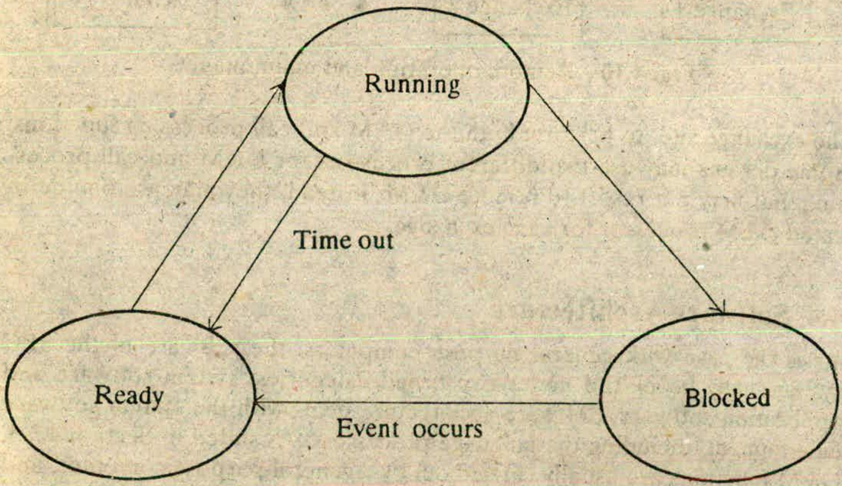


Fig. 4.11 Process states and transitions.

processes present in the system, the central processing unit (CPU) can be allocated to only one process at any instant of time. A process is said to be **running** if it currently has the CPU allocated to it. A process is said to be **ready** if it could use a CPU if one were available. A process is said to be **blocked** if it is waiting for some event to occur before it can proceed. While only one process may be running at any time, several processes may be ready and several blocked. The ready processes are ordered according to some priority so that the next process to receive the CPU is the first ready process in the ordered list. There may be several ready lists for each level of priority. The blocked processes are unordered and they unblock in the order in which the events they are awaiting occur. To prevent any one process from monopolising the CPU, either accidentally or maliciously, a timer is set, and if it runs out, the process is forced to the ready state to join at the end of the appropriate ready list as determined by its priority.

Each process is represented in the operating system by a **process control block** (PCB) which is a data structure containing, *inter alia*, the following information about the process:

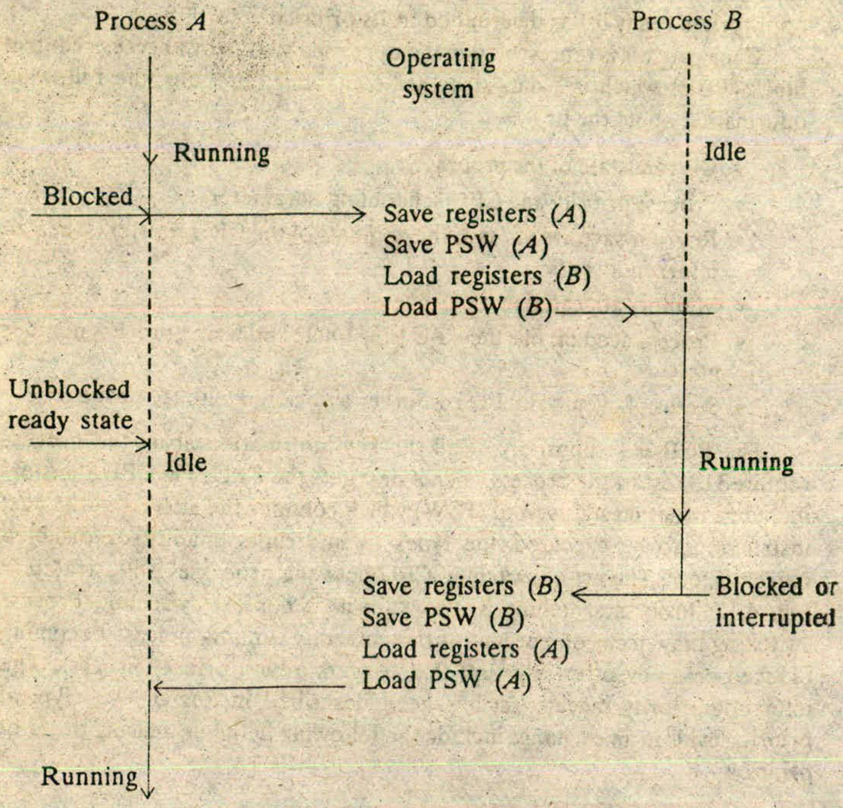
- Current state of the process
- Process priority and CPU scheduling parameters
- Register save area to save the contents of the CPU registers when an interrupt occurs
- Memory allocated to the process
- Process account like the CPU time limits and usage, process number etc.
- Status of events and I/O resources associated with the process.

The PCB is a repository of all the key information about the process required to restart the process when it next gets the CPU. The CPU registers include a program status word (PSW) which contains the address of the next instruction to be executed, the types of interrupts enabled or disabled currently, etc. The act of switching CPU between processes is illustrated in Fig. 4.12. Process switching is also known as context switching. Process switching may occur on account of the currently running process becoming blocked or an event or interrupt that triggers a high priority process. The interrupt priority mechanism has been described in Section 4.2. Typical priority levels in an exchange include the following in the decreasing order of priority:

- Fault alarms
- Interprocessor communication, high speed peripherals (disk, etc.)
- High speed clock interrupt for periodic tasks
- Exchange peripherals
- Low speed clock interrupt for periodic tasks
- Call processing
- Slow speed peripherals (terminals, etc.)
- O&M tasks.

As is now obvious, context switching, which implies PCB saving and loading, occurs very often in a switching processor. In order to speed up this activity, special hardware instructions that save and load PCBs are provided in the switching processors.

Processes in a switching system cooperate in the sense that they share common variables, update common tables, write into common files, and so on. Information about the resources of the exchange (trunks, registers etc.) and their current utilisation is kept in the form of tables. For example, a subscriber line state table contains information about each subscriber line,



PSW = program status word

Fig. 4.12 Process switching.

whether it is free or busy. Now, consider some process *A* which scans the subscriber line state table and finds that a particular subscriber is free. At this instant, process *B* which has a higher priority interrupts and seeks the same subscriber line and sets it busy for its own purpose. When the control is returned to process *A*, it is blissfully unaware of the action of process *B* and sets the subscriber line busy once again and allocates the same for its own purpose. Thus, a single subscriber line is now allocated to two different processes (calls) at the same time, which is incorrect. A similar problem can arise with other shared tables and files also. The problem can be solved by giving each process exclusive access to a shared table. When one process accesses a shared table, all others wanting to access the same table are kept waiting. When the first process finishes accessing, one of the waiting processes is given access. Thus, there is **mutual exclusion** of processes in accessing shared data.

When a process is accessing shared data, it is said to be in its **critical section** or **critical region**. Mutual exclusion implies that only one process can be in critical region at any instant for a given shared data. It may, however, be noted that when one process is in its critical section, others may continue execution outside their critical sections. Since a process being in its critical section forbids others from entering their critical regions, the critical section of any process must execute as quickly as possible. A process must not block within its critical region. The coding of the critical sections must be done carefully to avoid infinite loops etc.

Many hardware and software solutions have been devised to enforce mutual exclusion. We present here one of the more generalised software solutions; the use of a synchronisation tool called a **semaphore**. A semaphore is a protected variable that can assume non-negative integer values. There is one semaphore for each shared resource in the system. The initial value of a semaphore is set equal to the number of identical resources in a pool, say the number of trunk lines. Two standard individual operations test (P) and increment (V) are possible on a semaphore (S). Each P operation tests S to see if it is nonzero, and if so, decrements the value of S by 1, indicating that a resource has been removed from the common pool. If S is zero, the process is blocked to be released by a V operation which increments the value of S by one signifying the return of a resource to the pool. A blocked process is placed in a queue of processes, all of which are blocked for want of the same resource. In the case of a shared table, S can assume only binary values, zero or one.

The implementation of a semaphore with a blocked queue may result in a situation where two or more processes are waiting indefinitely for a V operation that can be caused only by one of the blocked processes. If this happens, the processes are said to be **deadlocked**. As an illustration, consider a system consisting of two processes P_0 and P_1, each accessing two binary semaphores S_0 and S_1 initialised to the value 1. The operations performed by P_0 and P_1 are:

P_0	P_1
$P(S_0)$; $S_0 = S_0 - 1$	$P(S_1)$; $S_1 = S_1 - 1$
$P(S_1)$; blocked (S_1)	$P(S_0)$: blocked (S_0)

The process P_0 having seized the semaphore S_0 gets blocked on S_1 and the process P_1 having seized the semaphore S_1 gets blocked on S_0. Both the processes are blocked for a V operation to be performed by the other, and there is a deadlock. Techniques for deadlock prevention, avoidance, detection and recovery have been developed but are not discussed in this text. Readers are referred to Further Reading [5, 9].

Having discussed the salient features of the operating system for switching processors, we now turn our attention to software production and language processors. The important place that software production occupies in the switching industry is due to five basic factors associated with switching software:

1. Complexity and size of the software
2. Long working life required
3. Real time operation
4. Stringent reliability and availability
5. Software portability.

The complexity of switching software stems from the complexity of switching systems. SPC system software runs into hundreds of thousands of lines of code. Thousands of programmer-years are involved in developing software for switching systems. The life expectancy of SPC software, which is about 40 years, is quite exceptional compared with that of 10 or 15 years for software of other major systems. Attendant upon the long life requirement of switching software is the need to develop and expand it during its lifetime for the following purposes:

- Supporting new services
- Meeting new requirements that may arise in network management
- Coping with applications that may differ from country to country
- Permitting technological enhancements in system hardware.

The need for real time operation and the requirement for high reliability and availability have been discussed already. The high investment of cost and manpower required to develop switching software has led to the adaptation of software systems to successive generations of hardware. This is known as software portability.

The study of problems encountered in the production and maintenance of large scale software for complex systems has led to the emergence of a special branch of engineering, known as **software engineering**. The practice of software engineering techniques calls for four stages in the production of software systems:

1. Functional specifications
2. Formal description and detailed specifications
3. Coding and verification
4. Testing and debugging.

The recommended approach to software design is top-down, i.e. proceeding from the general to the particular, with an increasing level of complexity and abstractions. To design a system, its functional specifications are usually described first in natural language. To avoid the ambiguities of a natural language, the concepts defined in the first writing are described and detailed specifications spelt out using a formal language. The formal description and specification language, usually based on diagrams, clearly shows all possible imbrications and bifurcations that can occur in the program. The next stage in software production is the programming itself.

This is a translation process, converting the computer related part of the formal specification into programming language. Programs are partitioned into modules which are developed using a stepwise decomposition or refinement process. Intermodule interfaces are carefully worked out so that the modules are relatively independent, making them easier to code, test and later modify. The last stage in software production consists of program testing and validation. During this stage, errors in the program are eliminated to a large extent by carefully selecting test patterns. This phase of activity is usually carried out in a bottom-up fashion.

The subject of software for SPC systems was entrusted to the CCITT study group on 'Telephone Switching', SG XI, by the 1968 CCITT Plenary Assembly. More or less at the same time, the software engineering concepts were beginning to crystallise in the field of computers. In the early 1970s, SG XI identified three major areas for software standardisation. Two of them correspond to two stages in software production, viz. formal specification and coding. The third is the language for man-machine interaction required for performing O&M functions in an exchange. Work in these areas has resulted in three CCITT standards in the 'Z' series:

- Z.100 – Specification description language (SDL)
- Z.200 – CCITT high level language (CHILL)
- Z.3xx – Man-machine language (MML).

SDL is a formalised method of describing the functional specifications and the internal processes necessary to implement the specifications. It is both a software development tool and a high level documentation technique. SDL is based primarily on **state transition diagrams.** It provides traditional flowcharting features to represent sequential and conditional computations. It also has features termed **signals,** to represent communication and synchronisation between concurrent processes. Each SDL diagram describes a single concurrent process. The standard symbols of SDL are summarised in Fig. 4.13.

SDL offers several advantages to both switching equipment manufacturer and operating and maintenance agency of the system:

1. It is a good top-down design technique.
2. It is oriented towards process interactions in a switching system.
3. Many tools are available for storage, updating and logical verification of SDL specifications. The SDL diagrams are easily transformed into **Petri Nets,** which is an excellent mathematical tool for analysis and study of concurrent processes.
4. Tools are available for translating SDL specifications into CHILL code and vice versa.
5. It is easy to learn, interpret and use.

6. It provides unambiguous specifications and descriptions for tendering and evaluation of offers.
7. It provides a basis for meaningful comparison of the capabilities of different SPC systems.

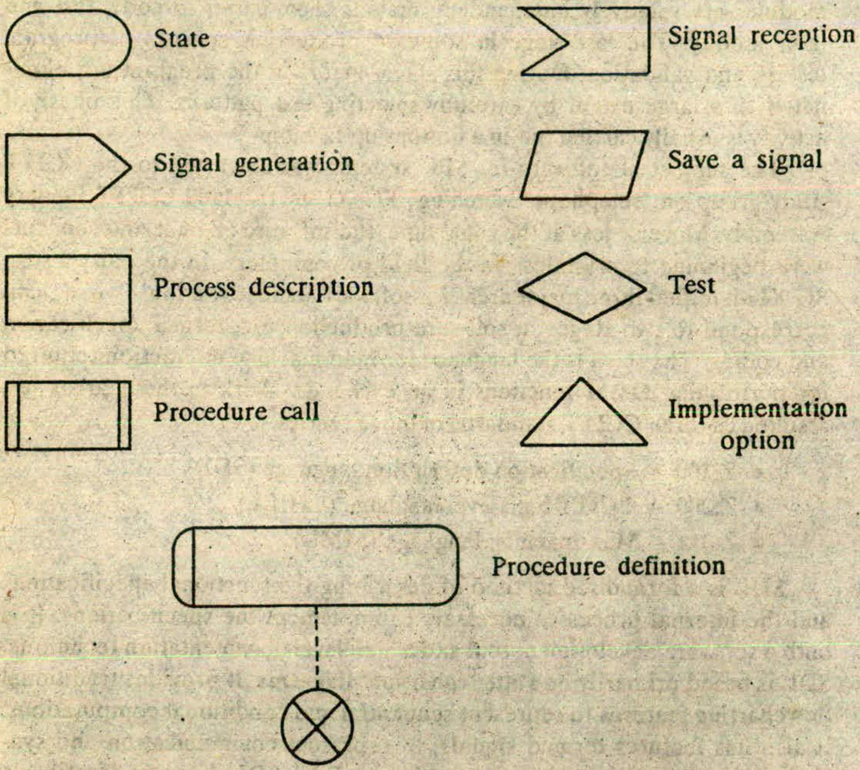


Fig. 4.13 Standard symbols in SDL.

Switching systems basically belong to the class of finite state machines (FSM) which are asynchronous in nature and follow a sequential logic for their operation. They can be modelled by using a combinational part and a memory part as shown in Fig. 4.14. In FSM, the status of the output circuits not only depends upon the inputs but also upon the current state of the machine. Asynchronous sequential operation gives rise to many problems due to transient variations that may occur in the logic circuits and memory elements. Clocked synchronous operation shown in Fig. 4.15 overcomes such problems. The theory of the operating principles of synchronous finite state machines forms the basis of design of SDL.

Both assembly and high level languages are used in producing switching software. Early electronic switching systems used assembly language programming extensively. The present trend is to use more and more of high

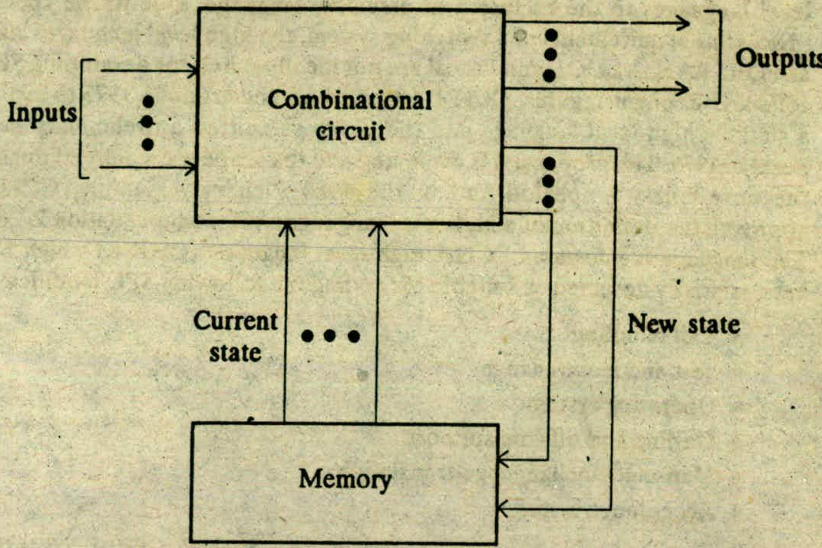


Fig. 4.14 Finite state machine model.

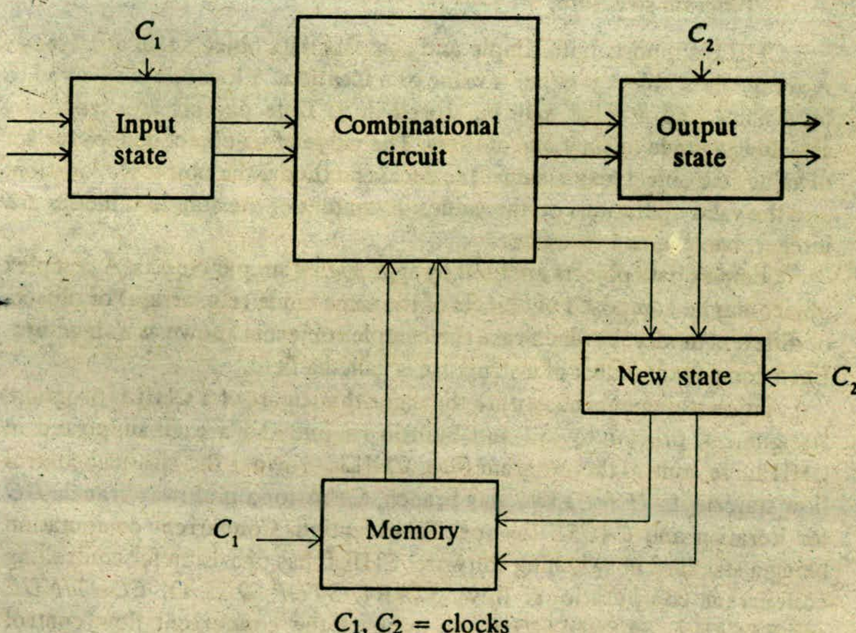


Fig. 4.15 Synchronous FSM.

level languages to the exclusion of assembly language. Due to the special rocessing requirements of a switching system, the high level languages such as FORTRAN, ALGOL and Pascal are not ideally suited for developing SPC software. Recognising this, CCITT set up an ad hoc group in 1975 to evolve a suitable high level language and the group submitted a preliminary proposal in 1976. The proposal was evaluated and the scope was enlarged during the next four-y period and in the 1980 Plenary Assembly, CCITT approved the definition of a high level language as Recommendation Z.200. The language is known as CCITT high level language (CHILL) which has been specially designed as suitable for coding the following SPC modules:

- Call handling
- Test and maintenance
- Operating system
- On-line and off-line support
- Man-machine language translation
- Acceptance testing.

Like many other programming languages (e.g. Pascal), CHILL has three major features:

1. Data structure
2. Action statements
3. Program structure.

CHILL supports both simple and complex data objects that are 'typed'. A simple data object is either a **value** or a **location**. A location is referred to by a name and a value may be stored in it. Data objects are 'typed' by attaching a **mode** to each one of them. The mode of an object defines the set of values the object may assume, the access method if the object is a location, and the valid operations on the values. Examples of the standard modes are integer, boolean and character.

Complex data objects are built by aggregating simple objects. A complex object may be composed of objects of the same mode (e.g. arrays) or objects of different modes, in which case the complex object is known as a **structure**. Each component object of a structure is called a field.

Action statements constitute the algorithmic part of a CHILL program. Assignment, procedure call and built-in routine call are all supported in CHILL. To control the program flow, CHILL provides the classical control flow statements: *IF* for a two-way branch, *CASE* for a multiway branch, *DO* for iteration and *CAUSE* for specific exception. Concurrent computation being a standard in switching software, CHILL has provision for controlling concurrent computational flow. *START, STOP, DELAY, CONTINUE* and *RECEIVE EXPRESSION* are some of the concurrent flow control statements.

Program structuring consists of grouping together actions and objects that are related. Program structuring controls the use of names, e.g. local or global, and the lifetime of objects. To enable concurrent computation, CHILL program structure supports the process concept discussed at the beginning of this section. Operations on processes include: create, destroy, suspend, resume, block, name and set priority.

No switching system is conceivable without arrangements for operation and maintenance of the same. Electronic switching systems are no exception. Recognising this, the work on CCITT man-machine language (MML) was initiated in early 1970s, and the basics of MML were approved as Recommendations Z.3xx series in the 1976 Plenary Assembly. The four main functions covered by MML are:

1. Operation
2. Maintenance
3. Installation
4. Acceptance testing.

MML is designed to be used by novices as well as experts, to be adaptable to different national languages and organisations and to be flexible to allow incorporation of new technology. Today, MML is used not only at exchange control desks but also at operation and maintenance centres administering subscriber connections, routing, and performing traffic measurement and network management.

MML is defined by its syntax, semantics and information interexchange procedures. Syntax defines the character set and the use of symbols, keywords and special codes to construct grammatically correct language sequences. In MML, the syntax definitions are conveyed through syntax diagrams that are similar to the ones used in Pascal. Semantics define the interpretation of each language element or statement. The information interchange procedures deal with the conversational or interactive aspects of the man-machine dialogue required in the switching system operation and maintenance. An operator error can sometimes have serious consequences for the exchange, and MML information interchange procedures are defined to prevent such occurrences. Operators are provided access to the system only through functional commands like **place a line in service, take a line out of service, modify service entitlement** etc. No direct access to data tables in the memory is permitted. For commands that have potentially serious consequences, MML procedure demands a reconfirmation from the operator before they are executed, thereby drawing the attention of the operator to the risks attendant on the requested action. MML procedures also have a **roll back** mechanism to resume operation from an earlier state of the system in case the present command is not properly executed due to a momentary failure in the system.

4.5 Application Software

The application software of a switching system may be divided into three main classes:

1. Call processing software
2. Administrative software
3. Maintenance software.

A software package is described by its organisation, the data structures it uses and the processing functions it performs. Application software packages of a switching system use a modular organisation. The software packages are divided into program modules, each dealing with a specific task. The size of a module varies depending on the task. Generally speaking, the modules are not self-contained. They exchange data with other modules, either directly through interfaces or indirectly through data tables. Several modules are grouped together to constitute functional units corresponding to independent functions. A module may be a part of more than one function unit. Usually, a functional unit runs as a separate process in the system. The modules of a process are strung together through special programs or chaining tables. Module chaining through tables is illustrated in Fig. 4.16. Associated with every module is a pointer to a set of entries in the chaining table pertaining to that module. Each entry in the chaining table consists of a key and a module number. Whenever a module completes execution, its corresponding entries in the chaining table are scanned and the keys are compared to a function status key. If a match occurs, the corresponding module in the chaining entry is executed next. This approach provides flexibility for adding new modules to a function or deleting old modules by simply modifying the chaining data.

Application software accounts for about 80 per cent of the total volume of the software in a switching system. Administration and maintenance programs together constitute about 65 per cent of the total volume. The total software typically comprises between 400,000 and 500,000 machine instructions. The entire software need not be core resident. Considering the real time constraints, the system software and the call processing application software are usually core resident. The administration and maintenance modules reside on a back up storage and are brought into the main memory as and when required. Depending on the architectural support available from the switching processor, the operating system may use overlay or virtual memory technique for this purpose.

Switching system software almost always uses a parameterised design. This enables the same package to be used over a wide range of exchanges by adapting the package to specific exchange characteristics. The parameters may be divided into system parameters and office parameters. The system parameters afford flexibility at the overall system level while the office parameters define program execution limits at specific exchanges. The system

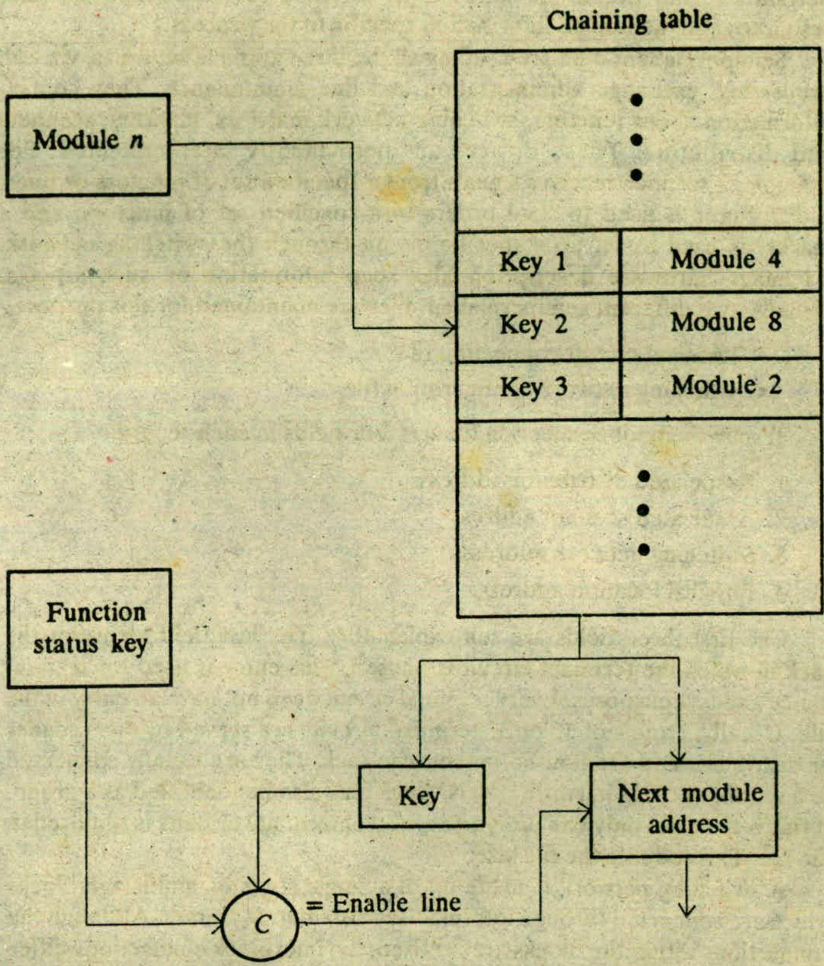


Fig. 4.16 Module chaining through table.

parameters are the same at all exchanges of a given type. Examples include signalling time delays, fault thresholds etc. Examples of office parameters are the number of subscribers, the maximum number of simultaneous calls etc. There are usually more than 100 office parameters which define the characteristics of an exchange.

Parametric data are stored in the form of tables or files in the system. The nature of the parametric data may be classified as either **semipermanent** or **temporary**. Semipermanent data consist of parameters that describe the hardware characteristics of the exchange and its environment. These data are updated rather infrequently, say when there is an expansion of the exchange. Temporary data have a lifetime equal to that of the process they pertain to. They define the state of system resources and temporary links resources and include all information specific to the process.

Semipermanent data are used by all the three application areas, viz. call processing, exchange administration and line maintenance. They contain information about junctors, switching network matrices, markers, scanners and distributors. These devices are permanently interconnected. For example, a scanner receives signals from a specified set of junctors or lines; a distributor is used to send orders to a specified set of junctors; and a marker is used to set up a specific circuit through the switching network. Exchange hardware description files keep information on such aspects. Usually, two different semipermanent files are maintained for this purpose:

- Terminal circuit connection file
- Switching network configuration file.

Terminal circuit connection file has four fields in each record:

1. Associated distributor address
2. Associated scanner address
3. Switching network address
4. Physical location address.

The first three fields are self-explanatory. The last field identifies the rack in which the terminal circuit is housed. This entry is used for maintenance and test purposes. Every terminal circuit need not have an entry in the file. Usually, groups of 16 or 32 terminal circuits are served by one scanner or distributor and are housed in the same rack. They are usually connected to a set of switching network points which may also be identified as a group. If this is the case, only one entry per group of terminal circuits is required in the file. This reduces the file size.

A switching network is made up of a single stage or multistage blocks which are connected through intermediary distribution frames. Although the connections within the blocks are prefixed, the interblock connections differ from exchange to exchange. The switching network configuration file contains information about the interblock connections.

The description of the exchange environment involves maintaining the following information:

- Connected subscriber lines
- Trunks towards other exchanges
- Rules for digit translation and routing.

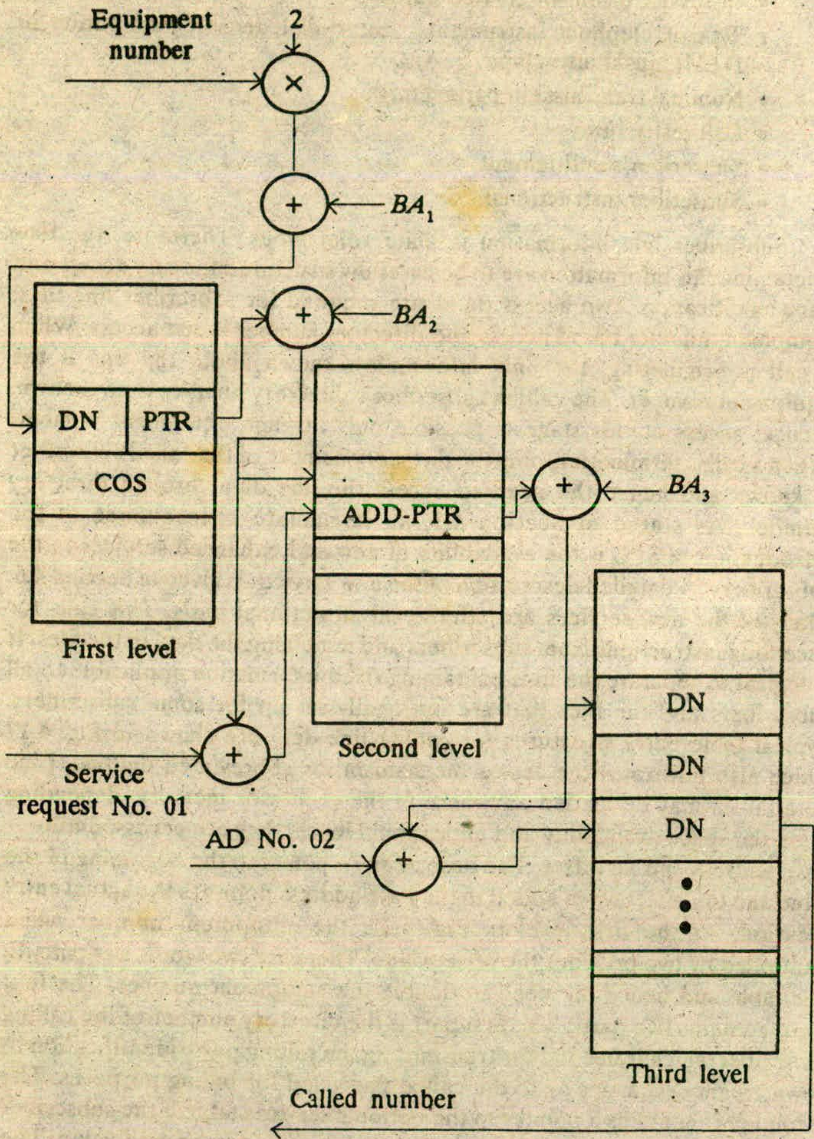
An extensive characterisation of subscriber lines is possible in the SPC systems when compared to electromechanical switching systems. The information maintained includes:

- Correspondence between line equipment number and directory number

- In-service or out-of-service status
- Type of telephone instrument – rotary dial, decadic push button or DTMF pushbutton type
- Nominal transmission parameters
- Call restrictions
- Subscriber's entitlement
- Subscriber instructions.

Subscriber line information is quite voluminous. Therefore, the files containing this information are to be carefully structured for easy access and easy modification. Two access paths are required for subscriber line files: equipment number based access and directory number based access. When a call is originating, the only information known about the line is the equipment number. The calling subscriber's directory number is not known. Hence, access at this stage is possible only through equipment number. When a call is terminating, only the directory number of the called subscriber is known and hence the need to access the line data through directory number. As stated in Section 4.1, an immediate consequence of the introduction of SPC is the availability of new and enhanced services to the subscribers. A detailed description of the new services is given in Section 4.6. Many of the new services are offered on an optional basis. This calls for accepting instructions from subscribers and modifying the data in the files. It is logical to separate the files containing fixed information applicable to all subscribers and the files that are optionally set up for some subscribers. Typical table entry structures for calling line data are shown in Fig. 4.17 which also illustrates the access mechanism for abbreviated dialling. Line data tables may be loaded anywhere in the processor memory, depending upon the space availability and allocation. Hence, they are accessed usin base address and an offset. The base addr points to the beginning of the table and the offset when added to the base address points to the actual entry location. At the first level in Fig. 4.17, the equipment number when multiplied by two provides the offset value. There are two words per entry in this table and hence the need to double the equipment number. The first word contains two parts. The first part is the directory number of the calling subscriber which is needed for transmitting the calling party identification to downstream exchanges or to the called party and for billing purposes. The second part contains a pointer to the optional services table if the subscriber has availed any of the optional services. Otherwise, it has a null value. The second word cont class of service information which includes the type of instrument, the type of line (individual telephone or public coin box), etc.

At the second level, each entry takes as many words as the number of optional services offered by the system if we assume that each word in this entry stores information regarding one optional service. The value at the first level is suitably adjusted taking this into account. When the pointer value is added to the base address 2, the starting address of the entry corresponding



AD = abbreviated dialling ADD = abbreviated dial directory
BA = base address COS = class of service
DN = directory number PTR = pointer

Fig. 4.17 Access to calling line data

to the calling subscriber is obtained. The user request for a service is converted to offset which is added to the starting address of the entry to obtain the word corresponding to that service. In our example in Fig. 4.17, the calling subscriber requests abbreviated dialling facility which is given the service request number 01. The word 1 of the entry corresponding to the abbreviated dialling service contains a pointer, ADD-PTR, to a third level table which is the abbreviated dialling directory for the subscribers. An access mechanism similar to the one in the second level gives out the directory number of the called subscriber corresponding to the abbreviated dialling number two.

As far as the called subscriber line data is concerned, the main purpose here is to identify the equipment number corresponding to the directory line number and class of service information with regard to reception. Assuming a 4-digit directory line number (exchange code excluded) a 10,000-entry table would enable one-level translation. But in exchanges where the actual number of connected subscribers is small (say 2000–5000), many of the thousandth position digits in the directory number are not used. In such cases, a two-level translation would result in saving storage space. Readers are urged to work out such a scheme.

Semipermanent data associated with the trunks gives the assignment of trunk groups to different exchanges and the signalling method to be used for each group. If the trunk circuits in sequence are grouped together, the organisation of trunk circuits data becomes simple. There is one entry per trunk group in the table which is accessed using the group address as offset. When a signal arrives on an incoming trunk, the trunk group is easily identified by discarding the lower order bits of the binary address of the trunk circuits. If there are eight circuits per group, the lower three bits are discarded. The table is then accessed using the higher order bits which represent the group address. For outgoing trunks, the trunk group is determined based on the translation of dialled digits corresponding to the office code. In the case of transit calls, the office code address received on the incoming trunk is translated to determine the outgoing trunk group.

Office code address translation may be organised by using either a linear or pyramidal structure for the data tables. Linear structure is suitable if the number of digits to be translated is fixed. For example, a 2-digit office code is translated by using a 100-entry table which is accessed using the 2-digit address as offset.

In telephone environment, the number of address digits to be translated is rarely fixed. Exchange codes are either two or three digits. An address that starts with digit zero implies an intercity or international call requiring translation action immediately after the first digit is received. In view of this variable number of digits, pyramidal table structures are preferred in electronic exchanges. Here, the digits are decoded one by one as they arrive. At each decoding level, the received digit indexes a table and the entry in the table determines the next step. The scheme is illustrated in Fig. 4.18. All tables in this scheme have 10 entries each. There is one table in the first level,

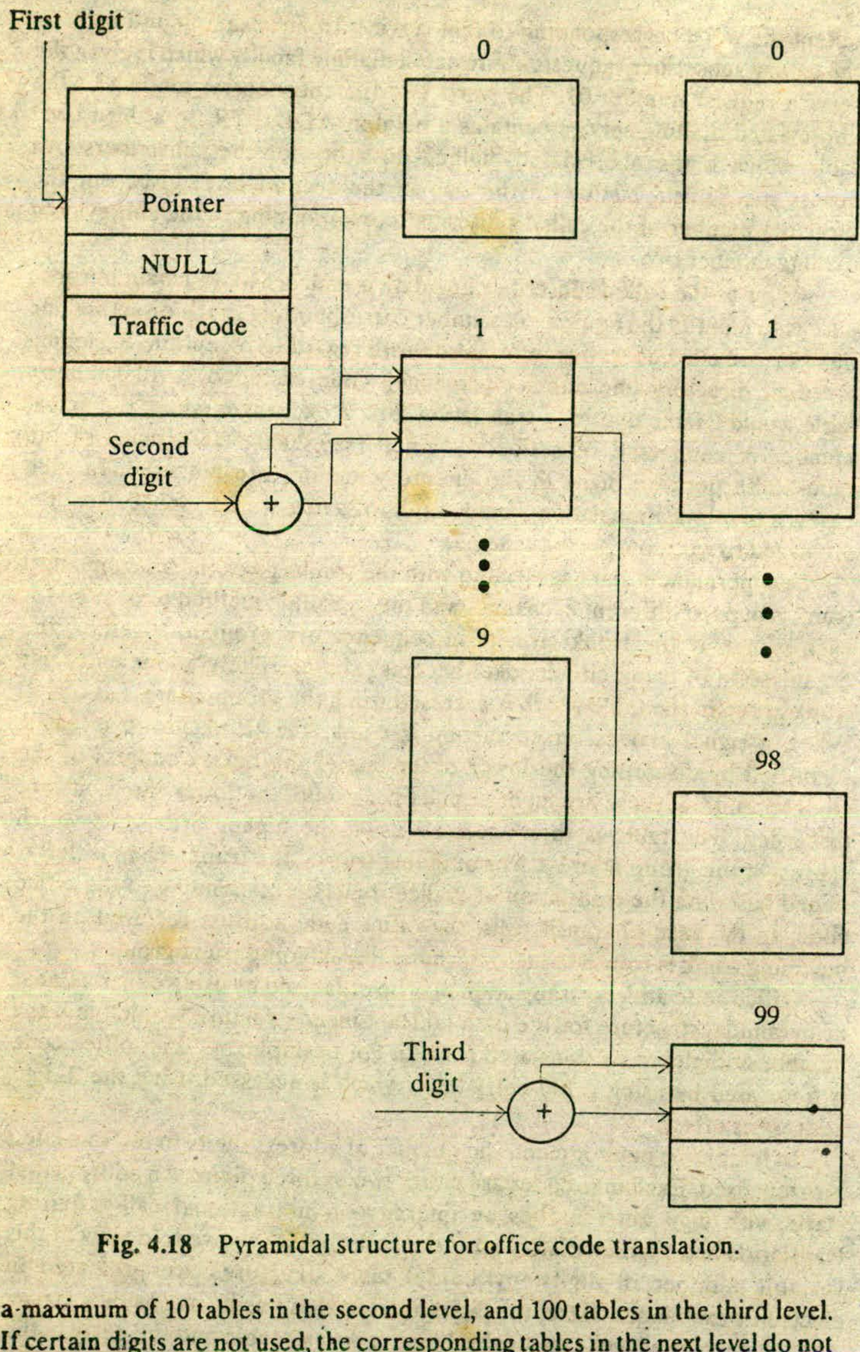


Fig. 4.18 Pyramidal structure for office code translation.

a maximum of 10 tables in the second level, and 100 tables in the third level. If certain digits are not used, the corresponding tables in the next level do not exist. An entry in the table may contain a traffic code, or a pointer to the beginning of a block in the next level or a null value. The null value means that the particular digit combination is not in use. The traffic code supplies the outgoing trunk group number and the associated signalling characteristics. It may also provide an alternate trunk group number to be used in case of overflow in the primary trunk group.

Temporary data are created and modified during call processing. They define the dynamic state of resources and links. They constitute the call related data stored in the working area of a process. The main state description tables are:

1. Subscriber line state table
2. Terminal circuit state table
3. Switching network link state table
4. Working area of a process.

The working area of a process holds call progress information, dialled digits, translation data read from the routing tables, class of service information for both calling and called subscriber etc. The working area for a process is created when the process is created and destroyed when the process is terminated.

In this section, we have so far discussed the organisation of the application software and the data structures used by it. We conclude the discussions regarding application software by giving a brief account of the process functions of each of the three application areas.

Call processing in electronic exchanges is somewhat similar to the one in a common control crossbar system except that most of the functions are performed by software. Call processing is usually handled by a number of different software processes that are created and terminated for every call by the main system control process. In addition, there are other system processes that run periodically and perform certain functions related to call processing. Among the periodically scheduled processes is a system process that scans all the subscriber lines every few hundred milliseconds looking for off-hook conditions. When an off-hook condition is detected, the matter is reported to the main control process which activates another process to handle all functions associated with the calling line. The functions include looking up the subscriber line data table to determine whether the subscriber has a rotary dial, decadic push button or DTMF push button telephone, allocate an appropriate digit receiver and monitor the receipt of dialled digits. The collected digits are passed on to another process which performs digit translation and routing. Another process obtains the routing information and sets up switching network links. In the meantime a process is created to perform all the functions related to the called line.

In addition to the usual administrative functions, administrative programs in electronic exchanges generate traffic reports, monitor traffic flow, uncover traffic sensitive network or terminal problems and gather information for billing. If the traffic load exceeds the capacity of the system, an overload control process is initiated which reschedules priorities and frequencies of activities to ensure that the system continues to process as many calls as practicable. One way of overload control is to restrict the number of call originations per unit time. This is done by delaying the sending of dial tone for a few seconds to a subscriber who goes off-hook.

Maintenance programs are run for performing either diagnostic function or preventive maintenance. During periods of normal traffic, there are preventive maintenance programs that take advantage of unused real time to run test programs of hardware and to audit system memory contents for correctness and consistency. In periods of high traffic, these programs are deferred. If a fault occurs in the system, the operating system activates unscheduled maintenance programs to recover the system from the fault with minimal mutilation of calls in progress. Sections of the exchange hardware may be isolated and diagnostic program run to enable maintenance personnel to fix faults.

4.6 Enhanced Services

One of the immediate benefits of stored program control is that a host of new or improved services can be made available to the subscribers. Over a hundred new services have already been listed by different agencies like CCITT, and the list is growing day by day. In fact, the only limitations in introducing new services seem to be the imagination of the designers and the price the market is prepared to pay for the services. Although there are a large number of services, they may be grouped under four broad categories:

1. Services associated with the calling subscriber and designed to reduce the time spent on dialling and the number of dialling errors
2. Services associated with the called subscriber and designed to increase the call completion rate
3. Services involving more than two parties
4. Miscellaneous services.

These new services are known as supplementary services and some of the prominent ones are as follows:

Category 1:

- Abbreviated dialling
- Recorded number calls or no dialling calls
- Call back when free.

Category 2:

- Call forwarding

- Operator answer.

Category 3:

- Calling number record
- Call waiting
- Consultation hold
- Conference calls.

Category 4:

- Automatic alarm
- STD barring
- Malicious call tracing.

Before discussing some of the important services in each of these categories, we describe the general procedure used for communicating subscriber commands to the exchange for obtaining these services. A subscriber issues commands to an exchange to activate or deactivate a service, record or clear data in the subscriber line data area or solicit an acknowledgement from the exchange. As an example, a user may enable or disable STD facility on his line by using a command. A command may or may not have data associated with it. The number of digits in the data, when present, may vary depending upon the command. As a result, subscriber commands are designed to be of variable length necessitating the use of an end-of-command symbol. The general command syntax is shown in Fig. 4.19.

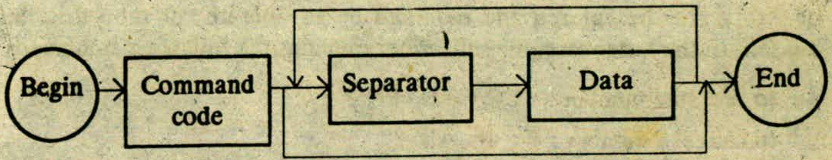


Fig. 4.19 Syntax of user commands.

The command code is usually a 2-digit number. Enhanced services are, in general, available only to subscribers with DTMF push button telephone. The two push buttons with symbols * and # are used extensively for communicating subscriber commands to the exchange. For easy handling of the commands at the exchange end, the commands are placed under four groups as shown in Table 4.2 which also gives the most popularly used command formats with and without data for each group. The symbol * is used as a separator and the symbol # as the end-of-command symbol. The beginning symbols of a command depends on the group of the command. To make abbreviated dialling as simple as possible, neither a symbol is used to indicate the beginning of the command nor is there any command code used in this case.

Table 4.2 User Command Formats

Command group	Format
Service activation and data recording	* CC #
	* CC * NNN #
Service deactivation and data clearing	# CC #
	# CC * NNN #
Interrogation	*# CC #
	*# CC * NNN #
Special abbreviated dialling	AN #

CC = command code AN = abbreviated number
NNN = data associated with the command

Abbreviated dialling (AD) facility allows an entitled subscriber to call any of a predefined list of other subscribers by dialling just one or two digits. Abbreviated dialling may be implemented through 'repertory diallers' or similar equipment attached to telephone sets, although we are here concerned with the service provided by the exchange with the subscribers using simple DTMF instruments. As illustrated in Section 4.5, memory area, called abbreviated dial directory, is reserved for each user availing AD facility, which contains the abbreviated number (AN) and the corresponding full number (FN) of the subscriber to be called. Call processing program translates the AN to FN by consulting this AD directory. The data in the AD directory may be entered and modified by an operator or the subscriber himself. In the latter case, the subscriber executes the following commands:

to record a number : * CC * AN *FN #

to cancel a number : # CC * AN #

to dial a number : AN #.

The use of some single digit and two-digit values as abbreviated numbers may be prohibited. For example, in some countries, no AN may start with a zero and no 2-digit AN may start with a one, thus restricting the AN range to 1–9 and 20–99 providing a maximum of 89 abbreviated numbers. Permitting a user to have AD facility implies reservation of memory space in the SPC processor. Hence, the charges for the service may be proportional to the number of ANs a subscriber wishes to have.

The facility of recorded number calls or no-dialling calls permits a subscriber to call a predetermined number by simply lifting the handset without dialling any digit whatsoever. Unlike a hot line facility where a dedicated line between the calling and called subscribers exists and no other calls are permitted using this line and instrument, recorded number call service is a programmable one. Here, the subscriber may use his telephone in the normal way and at the same time have recorded number call facility. If the subscriber goes off-hook and does not dial any digit for a few seconds (predetermined delay), the exchange automatically starts setting up the call to the previously recorded number. If he dials a digit within the predetermined delay, a normal call is assumed. The subscriber may record or cancel a number to be dialled automatically by using the appropriate subscriber commands.

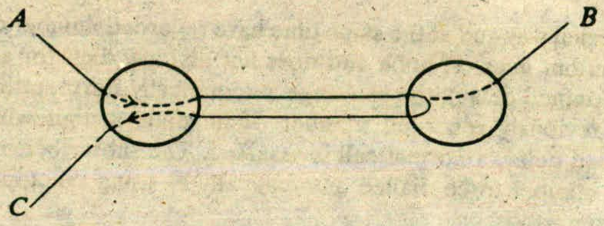
Whenever a call does not materialise, a subscriber would want to attempt the call again. In this case, he can request for an **automatic redialling or repeat dialling** feature in which the most recently dialled number is automatically redialled by the exchange. Continuous repeat dialling has the effect of increasing the call noncompletion rate. Hence, the automatic repeat dialling is usually limited to a few trials.

Call back when free feature permits the calling subscriber to instruct the exchange, when the called party is busy, to monitor the called line and ring him back when it becomes free. It is fairly easy to implement this feature within a local exchange. Monitoring distant calls requires extensive signalling between exchanges.

Call forwarding enables a subscriber to instruct the exchange to forward all his calls to another number. It is relatively straightforward to implement this facility in a PABX or a local exchange. If call forwarding is to be done across exchanges, a number of difficulties arise. These concern routing, charging and trunk utilisation. Suppose, in Fig. 4.20(a), *A* calls *B*, who has given instructions to forward this call to *C* who is a subscriber in the originating exchange. An interexchange call is now set up instead of a local call. Consider the situation depicted in Fig. 4.20(b), where *A* calls *B* who has given instructions to forward his calls to *C* who in turn has given instructions to forward his calls to *B*. There is a ping-pong effect between *B* and *C* and soon all the trunks are used up or captured in attempting to establish the call. In Fig. 4.20(c), *A* calls *B* who is in the same city and has given instructions to forward his calls to *C* who is in another city. If *A*'s call is forwarded to *C*, the question arises as to who, *A* or *B*, should bear the cost of the intercity call. If *A* has to pay, he must know that his local call is being changed to an intercity call so that he has the option not to go ahead with the call.

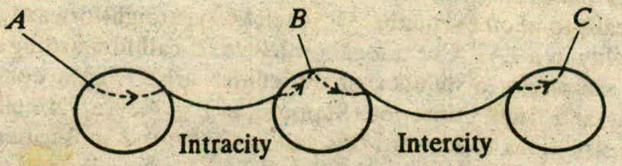
Operator answer service diverts all the calls of a subscriber to an operator, who answers the calls, takes down messages which are communicated to the subscriber whenever he calls the operator. With the availability of efficient and relatively cheap telephone answering machines, the usefulness of operator answering service has significantly diminished in the recent years.

Calling number record feature keeps a record of the numbers calling the subscriber when he is unable to attend to the calls for same reason or the other. The number of calling numbers recorded is usually limited to a few, say up to five most recently dialled numbers. On return, the user may request the exchange to dial these numbers and return the calls.



(a) Inefficient routing

(b) Trunk capturing



(c) Charging difficulty

Fig. 4.20 Difficulties in call forwarding across exchanges.

Call waiting feature provides an indication to a busy subscriber that another party is trying to reach him. The indication is given through a short audible tone, lasting typically about three seconds. The subscriber may then

- ignore the incoming call and continue with the present one,
- place the incoming call on hold and continue with the first call,
- place the first call on hold and answer the new call, or
- release the first call and accept the new one.

Call-waiting feature requires two switching paths to be set up simultaneously. Both the paths must use the same signalling scheme.

Consultation hold is a facility that enables a subscriber in conversation to place the other subscriber on hold and contact a third subscriber for consultation. This is like the telephone extension service used in offices where a secretary may consult the executive while holding an incoming call except that any subscriber number can be dialled for consultation. It may be possible for a subscriber to switch back and forth between the original party and the consulting party, alternately placing one of them on hold.

Conference call facility is an extension of the consultation hold feature. After the third party is brought in, a conference connection is set up among all the three. Each of the parties then receives the speech signals of the others and can proceed with the conversation in a conferencing mode. Setting up a three-party conference connection calls for a special equipment at the exchange to sum the speech signals of two parties and provide the same on the ring line of the third party. The arrangement is shown in Fig. 4.21. For a

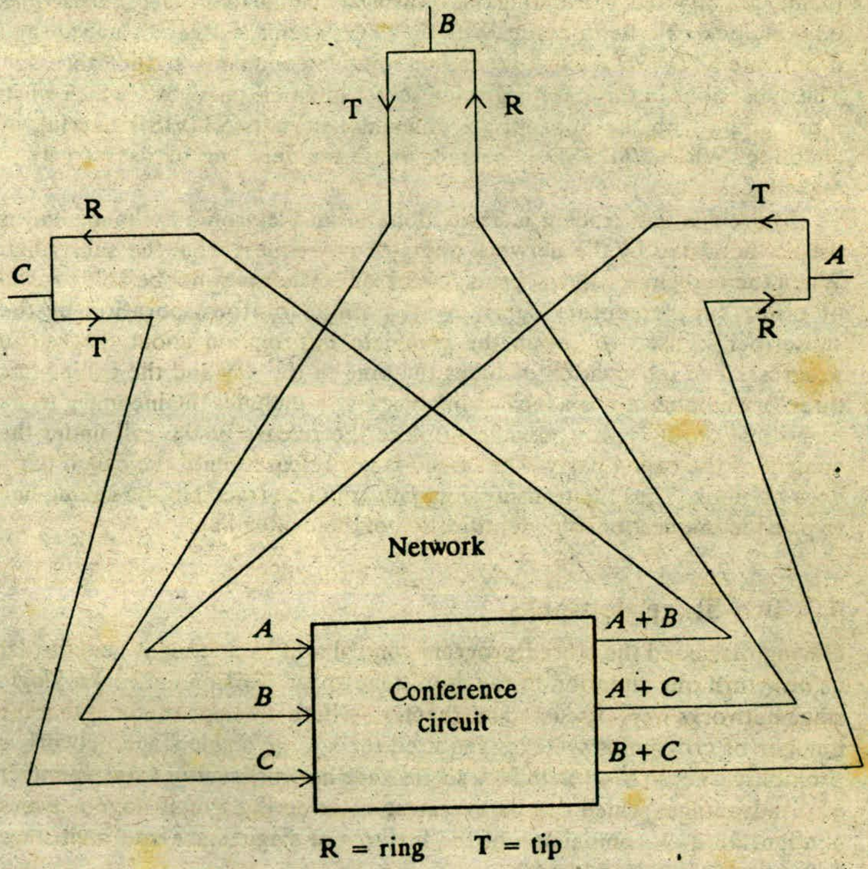


Fig. 4.21 Three-party conferencing connection.

conference involving N persons, N separate summations must be performed; one for each person containing all signals but his own. Another conferencing technique involves monitoring the activity of all the persons and switching the signals of the loudest talker to all others.

Automatic alarm call facility allows a user to record the time at which the alarm call is to be given. The SPC processor rings back the subscriber at the appropriate time. Usually, a separate synchronous process that runs periodically, say every five minutes, scans a table that holds the alarm times and the corresponding subscriber numbers. When the alarm time matches the time of the day, a ringing tone is sent to the subscriber's instrument, usually for one minute or until he answers, whichever is earlier. When the subscriber answers, a recorded greeting message is played to inform him that it is a wake-up call.

Many a time, particularly in office premises, telephones with STD/ISD facility are misused. Preventing this by using a manual lock places restrictions on even local calls being made. With the SPC systems, a user can activate and deactivate STD/ISD facility by sending suitable commands to the processor. The subscriber in this case is given a secret number, password, which when input along with the appropriate command permits **STD/ISD barring or enabling**. When STD/ISD is barred, the instrument can be used freely to make local calls.

Malicious call tracing is easily done in an electronic exchange and is usually activated by the network operator on request from the subscriber. When the malicious call tracing is on, the subscriber may not be able to avail of other supplementary services as a single button operation by the subscriber is used to obtain the complete information about the call in progress. The information includes the time of the day and the calling line directory number if it is a local call; otherwise it includes the incoming trunk identification. It is also possible to place the release of the call under the control of the called party. The circuit is not released until the called party goes on-hook. Thus the transmission path could be traced to the calling line to provide unquestionable identification of the calling line.

4.7 Two-Stage Networks

Having discussed the stored program control and its attendant benefits, let us now turn our attention to the multistage space division networks. Single stage networks were discussed in Chapter 3 where it was pointed out that the number of crosspoint switches required for a large single stage network is prohibitive, i.e. $N(N-1)/2$. In fact, single stage networks suffer from a number of disadvantages which can be overcome by adopting a multistage network configuration. A comparison of the features of single stage and multistage networks is given in Table 4.3.

Without Proof, we state the theorem that for any single stage network there exists an equivalent multistage network. A $N \times N$ single stage network with a switching capacity of K connections can be realised by a two-stage network of $N \times K$ and $K \times N$ stages as shown in Fig. 4.22. A connection needs two switching elements. Any of the N inlets can be connected to any of the K outputs of the first stage. Similarly, any of the

Table 4.3 Single Stage vs. Multistage Networks

S.No.	Single stage	Multistage
1.	Inlet to outlet connection is through a single crosspoint.	Inlet to outlet connection is through multiple crosspoints.
2.	Use of a single crosspoint per connection results in better quality link.	Use of multiple crosspoints may degrade the quality of a connection.
3.	Each individual crosspoint can be used for only one inlet/outlet pair connection.	Same crosspoint can be used to establish connection between a number of inlet/outlet pairs.
4.	A specific crosspoint is needed for each specific connection.	A specific connection may be established by using different sets of crosspoints.
5.	If a crosspoint fails, associated connection cannot be established. There is no redundancy.	Alternative cross-points and paths are available.
6.	Crosspoints are inefficiently used. Only one crosspoint in each row or column of a square or triangular switch matrix is ever in use, even if all the lines are active.	Crosspoints are used efficiently.
7.	Number of crosspoints is prohibitive.	Number of crosspoints is reduced significantly.
8.	A large number of crosspoints in each inlet/outlet leads to to capacitive loading.	There is no capacitive loading problem.
9.	The network is nonblocking in character.	The network is blocking in character.
10.	Time for establishing a call is less.	Time for establishing a call is more.

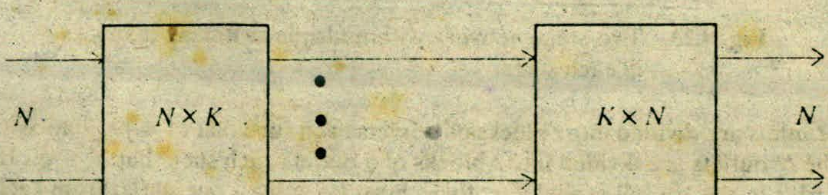


Fig. 4.22 A two-stage representation of an $N \times N$ network.

K inputs of the second stage can be connected to any of the N outlets. As a result, there are K alternative paths for any inlet/outlet pair connection. The network is said to provide **full connectivity** or **full availability**, in the sense that any of the N inlets can be connected to any of the N outlets in the network. The term *full connectivity* must be distinguished from the term *fully connected network* defined in Section 1.1. Each stage of the network has NK switching elements. Assuming about 10 per cent of the subscribers to be active on an average, K may be set equal to $(N/16)$. In this case, the number of switching elements, S, in the network is $(N^2/8)$. For $N = 1024$, we have $K = 64, S = 131{,}072$.

For large N, the switching matrix $N \times K$ may still be difficult to realise practically. It is necessary to consider architectures that use smaller sized switching matrices. Let us consider the two-stage realisation of an $M \times N$ switch using a number of smaller switching matrices as shown in Fig. 4.23.

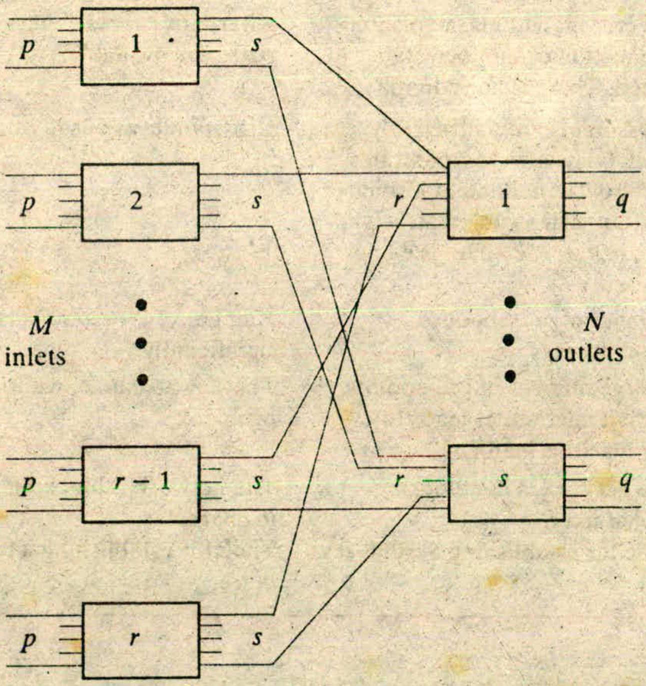


Fig. 4.23 Two-stage network with multiple switching matrices in each stage.

M inlets are divided into r blocks of p inlets each such that $M = pr$. Similarly, the N outlets are divided into s blocks of q outlets each such that $N = qs$. In order to ensure full availability, there must be at least one outlet from each block in the first stage terminating as inlet on every block of the second stage.

This determines the block sizes as $p \times s$ and $r \times q$ for the first and second stages respectively. Total number of switching elements S is given by

$$S = psr + qrs$$

Substituting for p and q in terms of M, N, r and s, we get

$$S = Ms + Nr \qquad (4.9)$$

The number of simultaneous calls that can be supported by the network, i.e. the switching capacity, SC, is equal to the number of links between the first and the second stage. Hence,

$$SC = rs \qquad (4.10)$$

For rs connections to be simultaneously active, the active inlets and outlets must be uniformly distributed. In other words, there must be s active inlets in each of the r blocks in the first stage and r active outlets in each of the s blocks in the second stage. Further, the s active inputs in one block of the first stage must be uniformly distributed across all the s blocks in the second stage at the rate of one per block.

This two-stage network is blocking in nature and the blocking may occur under two conditions:

1. The calls are uniformly distributed as described above: there are $r \times s$ calls in progress and the $(rs + 1)$-th call arrives.
2. The calls are not uniformly distributed; there is a call in progress from I-th block in the first stage to the J-th block in the second stage and another call originates in the I-th block destined to the J-th block.

   In the first case, the blocking probability P_B is dependent upon the traffic statistics. In the second case, we may calculate P_B as follows:

Let α be the probability that a given inlet is active. Then, the probability that an outlet at the I-th block is active is

$$\beta = (p\alpha)/s$$

The probability that another inlet becomes active and seeks an outlet other than the one which is already active is given by

$$(p - 1)\alpha/(s - 1)$$

The probability that the already active outlet is sought is, therefore,

$$P_B = \frac{p\alpha}{s}\left[1 - \frac{(p-1)\alpha}{s-1}\right]$$

Substituting $p = M/r$, we have

$$P_B = \frac{M\alpha(s-1) - ((M/r) - 1)\alpha}{rs(s-1)} \qquad (4.11)$$

Readers are advised to compare this case with the two-stage Strowger design described in Section 2.6.1.

From Eq. (4.9), we see that the number of switching elements can be minimised if r and s are as small as possible. On the other hand, if r and s are reduced, the blocking probability P_B goes high as seen from Eq. (4.11). We, therefore, have to choose values of r and s which are as small as possible but give sufficient links to provide a reasonable grade of service to subscribers. It may be noted that if $N > M$, the network is expanding the traffic; if $N < M$, the network is concentrating the traffic. The case when $M = N$ occurs often deserves attention. In this case, it is reasonable to assume that a uniform matrix size is used for both the stages, i.e. $r = s$ and $p = q$. The total number of switching elements, S, works out to be $2Nr$ and the switching capacity $SC = r^2$.

Often square switching matrices are available as standard IC chips which can be used as building blocks for switching networks. In such a case, $p = r = s = q = \sqrt{N}$. Thus the network has $\sqrt{N}$ blocks each in the first and second stages and each block is a square matrix of $\sqrt{N} \times \sqrt{N}$ inlets and outlets. If N is not a perfect square, the switching matrices are chosen to have a size of $\lceil \sqrt{N} \rceil \times \lceil \sqrt{N} \rceil$, where the symbol $\lceil\ \rceil$ denotes a ceiling function that gives the smallest integer equal to or greater than N. By substituting values in Eq. (4.9), we get

$$S = N\sqrt{N} + N\sqrt{N} = 2N\sqrt{N} \tag{4.12}$$

and from Eq. (4.10) we have

$$SC = \sqrt{N} \times \sqrt{N} = N \tag{4.12a}$$

As explained earlier in this section, N simultaneous calls can be supported on this network only if the traffic is uniformly distributed. Networks that support N simultaneous connections but under restricted traffic distribution conditions are known as **baseline networks.**

In the two-stage structure we have discussed so far, there is only one link between a block in the first stage and a block in the second stage. As a result, a link failure would cut off connection between p inlets and q outlets. This one-link structure gives rise to severe blocking in the network. The blocking performance can be improved by increasing the number of links between the blocks of the stages. Consider k links being introduced between every first and second stage block pair. Then, the design parameters for $M = N$ are $p = q = \sqrt{N}$, $s = r = K\sqrt{N}$ and $S = 2Nk\sqrt{N}$ and $SC = N$. To make the network nonblocking, we must have $K = \sqrt{N}$, we then get

$$S = 2N^2; \quad SC = N \tag{4.13}$$

Thus, a two-stage nonblocking network requires twice the number of switching elements as the single stage nonblocking network. In fact, for a nonblocking configuration, a two-stage network offers no distinct advantage over a single stage network except that it provides N alternative paths for establishing a connection. However, a standard way of designing blocking networks with full availability is to use two or more stages. Blocking networks require both concentrating and expanding network structures which are easily implemented as two separate parts. The purpose of using a two-stage arrangement in Section 3.6 must be clear now. The real advantages of multistage networks become evident when we consider networks of three or more stages.

4.8 Three-Stage Networks

The blocking probability and the number of switching elements can be reduced significantly by adopting a three-stage structure in place of two-stage networks. The general structure of an $N \times N$ three-stage blocking network is shown in Fig. 4.24. The N inlets and N outlets are divided into r blocks of p inlets and p outlets each respectively. The network is realised by using switching matrices of size $p \times s$ in stage 1, $r \times r$ in stage 2, and $s \times p$ in

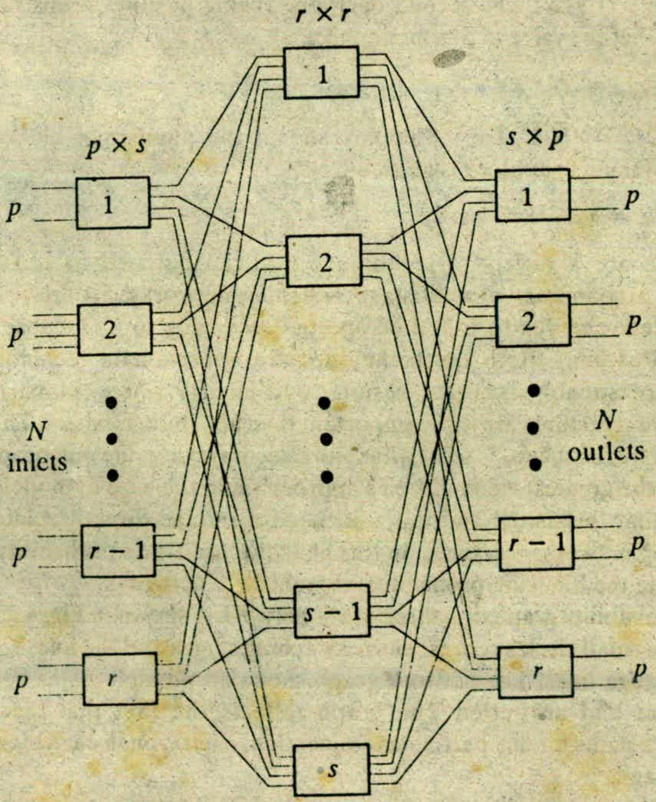


Fig. 4.24 $N \times N$ three-stage switching network.

stage 3. Unlike the two-stage network discussed in Section 4.7, here any arbitrary inlet in the first stage has s alternative paths to reach any arbitrary outlet in the third stage. The total number of switching elements is given by

$$S = rps + sr^2 + spr = 2Ns + sr^2 = s(2N + r^2) \tag{4.14}$$

If we use square matrices in the first and third stages, we have $p = s = (N/r)$ and, therefore,

$$S = \frac{2N^2}{r} + Nr \tag{4.15}$$

Equation (4.15) indicates that there is an optimum value for r that would minimise the value of S. To obtain this value of r, we differentiate Eq. (4.15), set it equal to zero and determine the value of r:

$$\frac{dS}{dr} = \frac{-2N^2}{r^2} + N = 0$$

Therefore, $r = \sqrt{2N}$. The second derivative, being positive at this value of r, indicates that the value of S is minimum, i.e.

$$S_{\min} = 2N\sqrt{2N} \tag{4.16}$$

and $p = N/r = \sqrt{N/2}$. The optimum ratio of the number of blocks to the number of inputs per block is given by

$$r/p = \sqrt{2N}/\sqrt{N/2} = 2 \tag{4.17}$$

There are a variety of techniques that can be used to evaluate the blocking probabilities of multistage switching networks. Of these, two are widely used: one due to C.Y. Lee and the other due to C. Jacobaeus (see Further Reading). Both the techniques are approximate techniques and provide reasonably accurate results, particularly when comparisons of alternative structures are more important than absolute numbers. The model proposed by Jacobaeus is somewhat more accurate than the one proposed by Lee. But the greatest value of Lee's approach is in the ease of modelling and the fact that the model and the associated formulae directly relate to the underlying network structures. In this book, we use Lee's probability graphs to estimate the blocking probability of multistage networks.

A probability graph of a three-stage network is shown in Fig. 4.25. In the graph, the small circles represent the switching stages and the lines represent the interstage links. The network graph shows all possible paths between a given inlet and an outlet. The graph reflects the fact that there are s alternative paths for any particular connection, one through each block in the second stage.

Blocking probabilities in the network may be estimated by breaking down a graph into serial and parallel paths. Let

β = probability that a link is busy

β' = probability that a link is free

Then,

$$\beta = 1 - \beta'$$

If there are s parallel links, the blocking probability is the probability that all

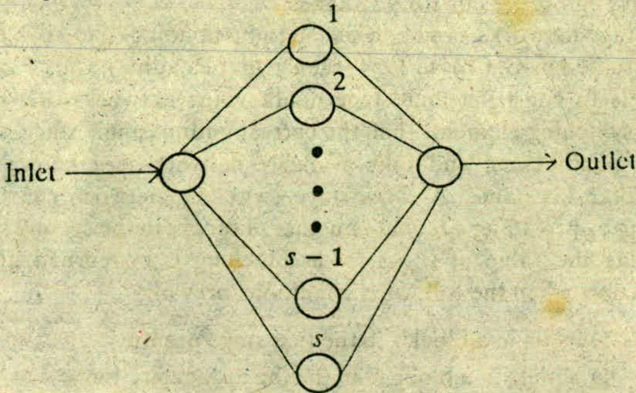


Fig. 4.25 Lee's graph for a three-stage network.

the links are busy:

$$P_B = \beta^s; \qquad Q_B = 1 - P_B = 1 - \beta^s$$

When a series of s links are needed to complete a connection, the blocking probability is easily determined as one minus the probability that they are available:

$$P_B = 1 - (\beta')^s = 1 - (1 - \beta)^s$$

For a three-stage network, there are two links in series for every path and there are s parallel paths. Therefore,

$$P_B = [1 - (\beta')^2]^s = [1 - (1 - \beta)^2]^s \tag{4.18}$$

If α is the probability that an inlet at the first stage is busy, then

$$\beta = p\alpha/s = \alpha/k \tag{4.19}$$

Substituting the value of β from Eq. (4.19) into Eq. (4.18), we get the complete expression for the blocking probability of a three-stage switch in terms of its inlet utilisation α as

$$P_B = [1 - (1 - \alpha/k)^2]^s \tag{4.20}$$

The factor k in Eq. (4.19) represents either space expansion or concentration. If s is greater than p, the first stage provides expansion; otherwise there is concentration. From Eq. (4.20) we see that in order to have a low value for the blocking probability the factor (α/k) must be small. If α is large, k must be large, i.e. if the inlets are well loaded, we need an expanding first stage. This is usually the case with transit exchanges where the incoming trunks are heavily loaded and expansion is needed to provide adequately low blocking probabilities. On the other hand, if a is small, k may be small, i.e. if the inputs are lightly loaded, the first stage may be a concentrating one. This is usually the case with end offices or PBX switches.

Multistage networks can be designed to be nonblocking. Such networks were first studied by C. Clos in 1954 (see Further Reading). Clos showed that it is possible to construct multistage nonblocking networks that have less number of switching elements than the corresponding single stage networks. Multistage nonblocking and fully available networks are known as Clos networks after his name. A three-stage switching network can be made nonblocking by providing adequate number of blocks in the second stage, i.e. by increasing the value of s. As far as blocking is concerned, the worst situation occurs when the following conditions prevail:

1. $(p - 1)$ inlets in a block I in the first stage are busy.
2. $(p - 1)$ outlets in a block O in the third stage are busy.
3. The $(p - 1)$ second-stage blocks, on which the $(p - 1)$ outlets from block I are terminated, are different from the $(p - 1)$ second-stage blocks from which the links are established to the block O.
4. The free inlet of block I needs to be terminated on the free outlet of block O.

The conditions are illustrated in Fig. 4.26. Under these circumstances, we require an additional block in the second stage. Thus, the number of blocks required in the second stage for nonblocking operation is given by

$$s = 2(p - 1) + 1 = 2p - 1 \tag{4.21}$$

The number of switching elements in the nonblocking configuration is given by

$$\begin{aligned} S &= p(2p - 1)r + (2p - 1)r^2 + p(2p - 1)r \\ &= 2N\left(\frac{2N}{r} - 1\right) + r^2\left(\frac{2N}{r} - 1\right) \\ &= \frac{4N^2 - 2Nr}{r} + 2Nr - r^2 \end{aligned} \tag{4.22}$$

The switching matrix sizes in stages one, two and three are $p \times (2p - 1)$, $r \times r$ and $(2p - 1) \times p$, respectively. The optimum value of r for minimising the number of switching elements may be computed as

$$\frac{dS}{dr} = -\frac{4N^2}{r^2} + 2N - 2r = 0$$

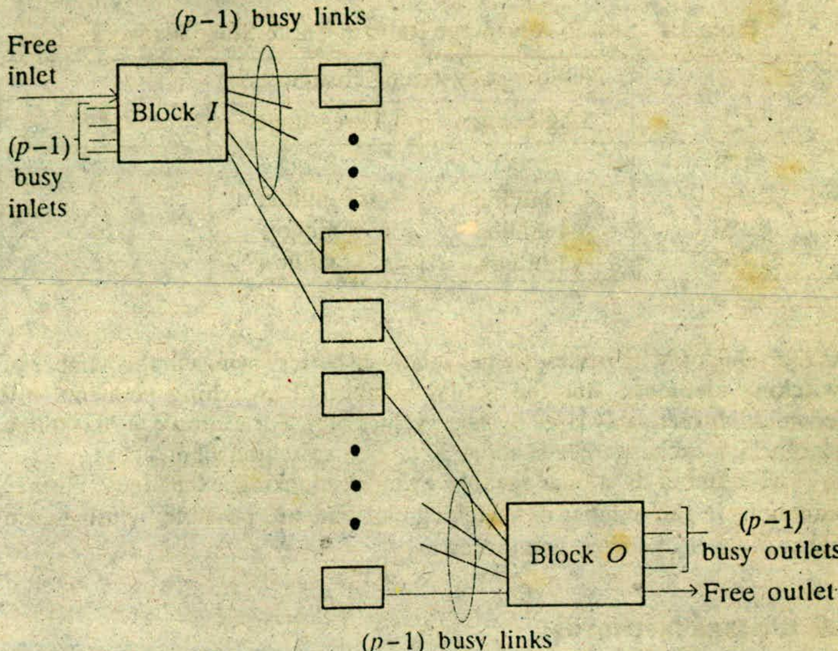


Fig. 4.26 Three-stage nonblocking configuration.

i.e. $-2N^2 + Nr^2 - r^3 = 0$, or

$$r^2(N - r) = 2N^2 \tag{4.23}$$

For large values of N, we have $N - r = N$, and hence

$$r = \sqrt{2N}, \quad p = \sqrt{N/2} \tag{4.24}$$

Substituting the value of r in Eq. (4.22), we have

$$S_{\min} = \frac{4N^2}{\sqrt{2N}} - 2N + 2N\sqrt{2N} - 2N$$

$$= 4N\sqrt{2N} - 4N = 4N(\sqrt{2N} - 1) \tag{4.25}$$

$$= 4N\sqrt{2N} \tag{4.26}$$

We may now define a factor called switching elements advantage ratio λ as

$$\lambda = \frac{\text{number of switching elements in a nonblocking single-stage network}}{\text{number of switching elements in a nonblocking three-stage network}}$$

Table 4.4 shows the values of S and λ for some values of N.

Table 4.4 Switch Advantage Ratio in Three-Stage Networks

N	Number of switching elements, S		λ
	Single stage	Three stages	
128	16,384	8,192	2
2,048	4 million	0.5 million	8
8,192	64 million	4 million	16
32,768	1 billion	32 million	32

As the value of N increases, we get relatively better savings in the number of switching elements. But the actual number of switching elements still becomes impracticably large for large values of N. For example, a 30,000-line nonblocking exchange needs about 30 million switching elements.

The number is unmanageable even in blocking exchanges. Further reductions in the number of switching elements are possible by using even higher number of stages than three.

4.9 *n*-Stage Networks

A variety of ways exist in which switching networks with four or more stages can be constructed. A description of all such networks is beyond the scope of this book. As an illustrative example, we discuss a five-stage network shown in Fig. 4.27. This network is formed by replacing each block of the centre stage of the network shown in Fig. 4.24 with a three-stage network. There are r inlets to a block in the centre stage of the network in Fig. 4.24. These are now terminated on the three-stage network in Fig. 4.26 that replaces the block in Fig. 4.24. The r inlets are distributed among the r_1 blocks shown in Fig. 4.26 with (r/r_1) inlets per block.

In order to compare the requirements of the switching elements in the case of three-stage and five-stage networks, let us assume that the three-stage network is realised with optimum number of square blocks in each stage so that the minimum number of switching elements are used. Taking a specific example of 2^{15} subscribers, for the three-stage network, from Eq.(4.16) we have the relations

$$S = 16 \times 2^{20}, \qquad p = 128, \qquad r = 256$$

In order to maintain the same level of blocking performance for the five-stage network as the three-stage network, let us assume that the centre three stages of the five-stage network are designed to be nonblocking and estimate the number of switching elements:

$$\text{Switching element in the first stage} = 2^8 \times 2^7 \times 2^7 = 2^{22}$$
$$\text{Switching elements in the last stage} = 2^{22}$$

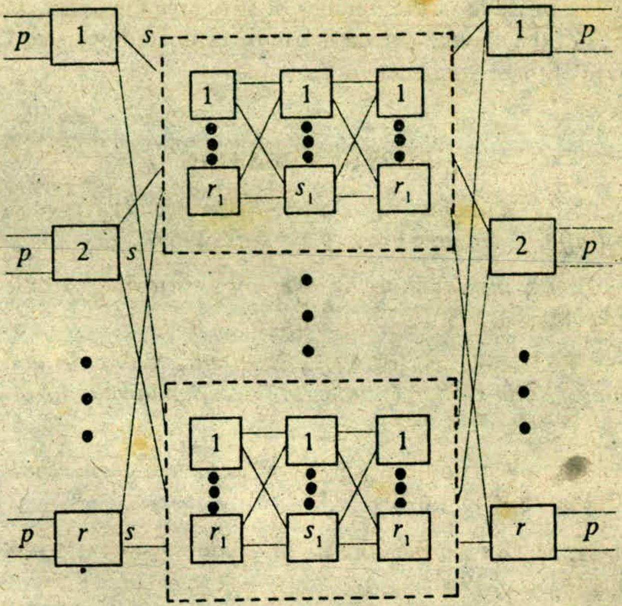


Fig. 4.27 Five-Stage switching network.

Switching elements in each three-stage block of the middle stages from Eq.(4.26) $= 2^{14} \times \sqrt{2}$

Number of switching elements in all the three-stage blocks $= 2^{14} \times \sqrt{2} \times 2^{8} = \sqrt{2} \times 2^{22}$

Total number of switching elements $= 3.4 \times 2^{22} \approx 13.6$ million

If we were to accept certain amount of blocking in the middle stages and use square matrices and optimum number of blocks, we obtain the following values from Eqs.(4.16) and (4.17):

$p \approx 12, \quad r_1 = 2 \times p = 24$

$S = \sqrt{2} \times 2^{13}$ per block of three-stage network

$S = \sqrt{2} \times 2^{20}$ for all the middle three-stage blocks

$S = 9.4$ million for the entire five-stage network.

The process of replacing the middle blocks with three-stage network blocks can be continued to obtain any number of stages (odd) in a network. If the replacement network is designed to be strictly nonblocking, the resulting $(N + 2)$ stage network has the same blocking performance as the n-stage network, but with reduced number of switching elements. By accepting a small amount of blocking in the middle stages, even further reductions are possible in the number of switching elements. The blocking performance of the n-stage networks can be estimated by using Lee's graph as in the case of three-stage network.

FURTHER READING

1. Benes, V. E.; *Mathematical Theory of Connecting Networks and Telephone Traffic*, Academic Press, New York, 1965.
2. Briley, B. E., *Introduction to Telephone Switching*, Addison-Wesley, Reading (Mass.), 1983.
3. Chapuis, R. J. and Joel, Jr., A. E., *Electronics, Computer and Telephone Switching, (North-Holland Studies in Telecommunication, Vol. 13)*, North Holland, Amsterdam, 1990.
4. Clos, C., A study of non-blocking network, *Bell System Technical Journal*, Vol. 32, March 1953, pp 406–424.
5. Deitel, H. M., *An Introduction to Operating Systems*, Addison-Wesley, Reading (Mass.), 1984.
6. GRINSEC, *Electronic Switching, (North Holland Studies in Telecommunications, Vol. 2)*, North Holland, Amsterdam, 1983.
7. Jacobaeus, C., *A Study of Congestion in Link Systems*, Ericsson Techniques, No. 48, Stockholm, 1950, pp 1–70.
8. Lee, C.Y., Analysis of switching networks, *Bell System Technical Journal*, Vol. 34, November 1955, pp 1287–1315.
9. Silberschatz, A. and Peterson, J. L., *Operating System Concepts*, Addison-Wesley, Reading (Mass.), 1988.

EXERCISES

1. Define each of the following terms: program, procedure, process, user, task, job and subroutine.
2. Blocked processes list is not maintained in priority order. Why?
3. In a switching system running thousands of processes, it cannot easily be determined that a process is in infinite loop. What safeguards can be built into the operating system to prevent processes running indefinitely?
4. Show that if the P and V operations are not executed atomically (indivisibly), the mutual exclusion may be violated.
5. Show how a general semaphore can be implemented using binary semaphores.

6. Deadlock may occur in a road traffic junction. Illustrate this with the help of a diagram.

7. A transit exchange may be likened to a road traffic junction. Show the *P* semaphore operations that can lead to a deadlock in a transit exchange that switches traffic among four exchanges. Assume one incoming and one outgoing trunk for each exchange.

8. Semaphore waiting lists are usually implemented in first-in-first-out (FIFO) order. What problems do you foresee if they are implemented in last-in-first-out (LIFO) order?

9. A local switching system has a capacity of *SC*. When there are no calls the SPC processor is idle and is activated when a call arises. If a call arrives when the processor is busy, it is put on a wait provided there is spare capacity available in the system. Otherwise, the call is considered lost. Write a program to coordinate the SPC processor function and the call arrivals.

10. Write a program in Pascal or CHILL that implements an alarm service that sends out two reminder rings after the first wake-up call at an interval of five minutes each.

11. A three-stage switching structure supports 128 inlets and 128 outlets. It is proposed to use 16 first stage and third stage matrices.
    (a) What is the number of switching elements in the network if it is nonblocking?
    (b) At peak periods, the occupancy rate of an inlet is 10%. If the number of switching elements required for nonblocking operation is reduced by a factor of 3, what is the blocking probability of the network?

12. Determine the switch advantage ratio of a three-stage network with N inlets and N outlets for the cases when (a) $N = 128$ and (b) $N = 32,768$.

13. A three-stage network is designed with the following parameters: $M = N = 512$, $p = q = 16$ and $\alpha = 0.7$. Calculate the blocking probability of the network if (a) $s = 16$, (b) $s = 24$, and (c) $s = 31$ using the Lee equation. Determine the inaccuracy of the result in case (c).

14. The Jacobaeus equation for the blocking probability in a three-stage network is given by

$$P_B = \frac{(p\,!)^2}{s\,!\,(2p - s)\,!}\,\alpha^s(2 - \alpha)^{2p - s}$$

where the symbols have the same meaning as in Eqs. (4.18) and (4.19). For the three-stage parameters and the cases given in Exercise 13, calculate the blocking probabilities. Compare the results with those obtained in Exercise 13.

15. Determine the design parameters of a three-stage switch with inlet utilisation of 0.1 to achieve a $P_B = 0.002$ for (a) $N = 128$, (b) $N = 2048$, and (c) $N = 8192$.

16. Using the Lee graph, show that the blocking probability of a five-stage network is given by

$$P_B = [1 - (1 - \alpha_1)^2 [1 - \{1 - (1 - \alpha_2^2)^{s_1}\}]^s$$

where $\alpha_1 = \alpha(p/s)$, $\alpha_2 = \alpha_1 \dfrac{r}{r_1 s_1}$, α is the probability that an input line is active and r, r_1 and s_1 have the same significance as in Fig.4.27.