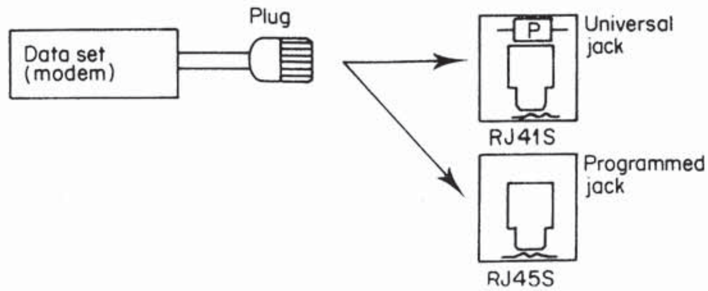

Figure 1.69 Programmed arrangement jack

Table 1.29 Telephone ordering options

Decision	Description	
A	1	Telephone set controls line
	2	Data set controls line
B	3	No aural monitoring
	4	Aural monitoring provided
C	5	Touchtone dialing
	6	Rotary dialing
D	7	Switchhook indicator
	8	Mode indicator

When the telephone set is optioned for telephone set controls the line, calls are originated or answered with the telephone by lifting the handset off-hook. To enable control of the line to be passed to a modem or data set an 'exclusion key' is required.

The exclusion key telephone permits calls to be manually answered and then transferred to the modem using the exclusion key. The exclusion key telephone is wired for either 'telephone set controls line' or 'data set controls line'. Data set control is normally selected if you have an automatic call or automatic answer modem since this permits calls to be originated or answered without taking the telephone handset off-hook. To use the telephone for voice communications the handset must be raised and the exclusion key placed in an upward location.

The telephone set control of the line option is used with manual answer or manual originate modems or automatic answer or originate modems that will be operated manually. To connect the modem to the line the telephone must be off-hook and the exclusion key placed in an upward position. To use telephone for voice communications the telephone must be off-hook while the exclusion key is placed in the downward position.

When the dat set controls the line option is selected, calls can be automatically originated or answered by the data equipment without lifting the telephone handset.

Aural monitoring enables the telephone set to monitor call progress tones as well as voice answer back messages without requiring the user to switch from data to voice.

You can select option B3 if aural monitoring is not required, while option B4 should be selected if it is required. Option C5 should be selected if touchtone dialing is to be used, while option C6 should be specified for rotary dial telephones. Under option D7, the exclusion key will be bypassed, resulting in the lifting of the telephone handset causing the closure of the switchhook contact in the telephone. In comparison, option D8 results in the exclusion key contacts being wired in series with the switchhook contacts, indicating to the user whether he or she is in a voice or data mode.

Ordering the business line

Ordering a business line to transmit data over the switched telephone network currently requires you to provide the telephone company with four items of information. First, you must supply the telephone company with the Federal Communications Commission (FCC) registration number of the device to be connected to the switched telephone network. This 14-character number can be obtained from the vendor who must first register their device for operation on the switched network prior to making it available for use on that network.

Next, you must provide the ringer equivalence number of the data set to be connected to the switched network. This is a three-character number, such as 0.4A, and represents a unitless quotient formed in accordance with certain circuit parameters. Finally, you must provide the jack numbers and arrangement to be used as well as the telephone options if you intend to use a handset.

1.15 THE DATA LINK LAYER

In the ISO model, the data link layer is responsible for the establishment, control, and termination of connections among network devices. To accomplish these tasks the data link layer assumes responsibility for the flow of user data as well as for detecting and providing a mechanism for recovery from errors and other abnormal conditions, such as a station failing to receive a response during a predefined time interval.

In this section, we will first examine the key element that defines the data link layer—its protocol. In this examination, we will differentiate between terminal protocols and data link protocols to eliminate this terminology as a potential area of confusion. Next, we will focus our attention upon several specific wide area networking protocols, starting with simple asynchronous line-by-line protocols. Protocols examined in the second portion of this section include an asynchronous teletype protocol, several popular asynchronous file transfer protocols used to transfer data to and from personal computers, IBM's character-oriented binary synchronous communication (commonly referred to as BSC or bisync), Digital Equipment Corporation's Digital Data Communications Message Protocol (DDCMP), and the bit-oriented Higher Level Data Link Control (HDLC) as well as its IBM near-equivalent, Synchronous Data Link Control (SDLC).

Terminal and data link protocols

Two types of protocols should be considered in a data communications environment: terminal protocols and data link protocols.

The data link protocol defines the control characteristics of the network and is a set of conventions that are followed which govern the transmission of data and control information. A terminal or a personal computer can have a predefined control character or set of control characters which are unique to the terminal and are not interpreted by the line protocol. This internal protocol can include such control characters as the bell, line feed and carriage return for conventional teletype terminals, blink and cursor positioning characters for a display terminal and form control characters for a line printer.

For experimenting with members of the IBM PC series and compatible computers, you can execute the one line BASIC program PRINT CHR$(X) "DEMO", substituting different ASCII values for the value of *X* to see the effect of different PC terminal control characters. As an example, using the value 7 for X, the IBM PC will beep prior to displaying the message DEMO, since ASCII 7 is interpreted by the PC as a request to beep the speaker. Using the value 9 for X will cause the message DEMO to be printed commencing in position 9, since ASCII 9 is a tab character which causes the cursor to move on the screen 8 character positions to the right. Another example of a terminal control character is ASCII 11, which is the home character. Using the value 11 for X will cause the message DEMO to be printed in the upper left-hand corner of the screen since the cursor is first placed at that location by the home character.

Although poll and select is normally thought of as a type of line discipline or control, it is also a data link protocol. In general, the data link protocol enables the exchange of information according to an order or sequence by establishing a series of rules for the interpretation of control signals which will govern the exchange of information. The control signals govern the execution of a number of tasks which are essential in controlling the exchange of information via a communications facility. Some of these information control tasks are listed in Table 1.30.

Table 1.30 Information control tasks

Connection establishment	Transmission sequences
Connection verification	Data sequence
Connection disengagement	Error control procedures

Connection establishment and verification

Although all of the tasks listed in Table 1.30 are important, not all are required for the transmission of data, since the series of tasks required is a function of the total data communications environment. As an example, a single terminal or personal computer connected directly to a mainframe or another terminal device by a leased line may not require the establishment and verification of the connection. Several devices connected to a mainframe computer on a multidrop or multipoint line would, however, require the verification of the identification of each terminal device on the line to insure that data transmitted from the computer would be received by the proper device. Similarly, when a device's session is completed, this fact must be recognized so that the mainframe computer's resources can be made available to other users. Thus, connection disengagement on devices other than those connected on a point-to-point leased line permits a port on the front-end processor to become available to service other users.

Transmission sequence

Another important task is the transmission sequence which is used to establish the precedence and order of transmission, to include both data and control information. As an example, this task defines the rules for when devices on a multipoint circuit may transmit and receive information. In addition to the transmission of information following a sequence, the data itself may be sequenced. Data sequencing is normally employed in synchronous transmission and in asynchronous file transfer operations where a long block is broken into smaller blocks for transmission, with the size of the blocks being a function of the personal computer's or terminal's buffer area and the error control procedure employed. By dividing a block into smaller blocks for transmission, the amount of data that must be retransmitted, in the event that an error in transmission is detected, is reduced.

Although many error-checking techniques are more efficient when short blocks of information are transmitted, the efficiency of transmission correspondingly decreases since an acknowledgement (negative or positive) is returned to the device transmitting after each block has been received and checked. For communications between remote job entry terminals and computers, blocks of up to several thousand characters are typically used. Block lengths from 80 to 1024 characters are, however, the most common sizes. Although some protocols specify block length, most protocols permit the user to set the size of the block, while other protocols automatically vary the block size based upon the error rate experienced by the transmission progress.

Error control

The simplest method of error control does not actually ensure errors are corrected. This method of error control is known as echoplex and results in each character transmitted to a receiving device being sent back or echoed from the receiver to the transmitting device, hence, the term 'echoplex'.

Under the echoplex method of error correction, the transmitting device examines the echoed data. If the echoed data differs from the transmitted data an error is assumed to have occurred and the data must then be retransmitted. Since a transmission error can occur in either direction, it is possible for a character corrupted to another character during transmission in one direction to be corrupted back into its original bit form when echoed to the transmitting device. In addition, a character received correctly may be corrupted during its echo, resulting in the false impression that an error occurred.

Echoplex was a popular method for detecting transmission errors that was used with teletype terminals. This method of error detection was also used in such message switching systems as TWX and is currently used with many types of asynchronous transmission, including personal computers. Concerning the latter, a PC communicating in an asynchronous full-duplex mode to another full-duplex computer will have each character transmitted echoed back. Since the detection of erroneous characters, however, depends upon the visual accuity of the operator more modern methods of error detection and correction have replaced the use of echoplex in applications where we cannot rely upon an operator to correct errors. Thus, a large number of file transfer protocols that group characters into data blocks and append a checking mechanism were developed to automatically detect and correct transmission errors.

Today, the most commonly employed method to correct transmitted errors is to inform the transmitting device simply to retransmit a block. This procedure requires coordination between the sending and receiving devices, with the receiving device either continuously informing the sending device of the status of each previously transmitted block or transmitting a negative acknowledgement only when a block is received in error.

If the protocol used requires a response to each block and the block previously transmitted contained no detected errors, the receiver will transmit a positive acknowledgement and the sender will transmit the next block. If the receiver detects an error, it will transmit a negative acknowledgement and discard the block containing an error. The transmitting station will then retransmit the previously sent block. Depending upon the protocol employed, a number of retransmissions may be attempted. However, if a default limit is reached owing to a bad circuit or other problems, then the computer or terminal device acting as the master station may terminate the session, and the operator will have to re-establish the connection.

If the protocol supports transmission of a negative acknowledgement only when a block is received in error, additional rules are required to govern transmission. As an example, the sending device could transmit several blocks and, in fact, could be transmitting block $n+4$ prior to receiving a negative acknowledgement concerning block n. Depending upon the protocol's rules, the transmitting device could retransmit block n and all blocks after that block or finish transmitting block $n+4$, then transmit block n and resume transmission with block $n+5$.

Types of protocols

Now that we have examined protocol tasks, let us focus our attention upon the characteristics, operation and utilization of several types of protocols that provide a predefined agreement for the orderly exchange of information. To facilitate this examination we will start with an overview of one of the simplest protocols in use and structure our overview of protocols with respect to their complexity.

Teletype protocols

Teletype and teletype compatible terminals support relatively simple protocols used for conveying information. In general, a teletype protocol is a line-by-line protocol that requires no acknowledgement of line receipt. Thus, the key elements of this protocol define how characters are displayed and when a line is terminated and the next line is to be displayed. Some additional elements included in line by line teletype protocols actually are part of the terminal protocol, since they define how the terminal should respond to specific control characters.

Teletype Model 33

One commonly used teletype protocol is the Teletype® Model 33 data terminal. This terminal transmits and receives data asynchronously on a line by line basis using a modified ASCII code in which lower-case character received by a Model 33 are actually printed as their upper-case equivalent, a term known as 'fold-over' printing. Although the ASCII code defines the operation of 32 control characters, only 11 control characters can be used for communications control purposes. Prior to examining the use of communications control characters in the teletype protocol, let us first review the operational function and typical use of each control character. These characters were previously listed in Table 1.11 with the two-character designator CC following their meaning and will be reviewed in the order of their appearance in the referenced table.

Communications control characters

NUL

The null (NUL) character is a non-printable time delay or filler character. This character is primarily used for communicating with printing devices that require a defined period of time after each carriage return in which to reposition the printhead to the beginning of the next line. In the early days of PC communications many mainframe computers would be programmed to prompt users to 'Enter the number of nulls'; this is a mechanism to permit electromechanical terminal devices that require a delay to return the print head to the first position on the next line without obtaining garbled output.

SOH

The start of heading (SOH) is a communications control character used in several character-oriented protocols to define the beginning of a message heading data block. In synchronous transmission on a multipoint or multidrop line structure, the SOH is followed by an address which is checked by all devices on the common line to ascertain if they are the recipient of the data. In asynchronous transmission, the SOH character can be used to signal the beginning of a filename during multiple file transfers, permitting the transfer to occur without treating each file transfer as a separate communications session. Since asynchronous communications typically involve point-to-point communications, no address is required after the SOH character; however, both devices must have the same communications software program that permits multiple file transfers in this manner.

STX

The start of text (STX) character signifies the end of heading data and the beginning of the actual information contained within the block. This communications control character is used in the bisynchronous protocol that will be examined later in this section.

ETX

The end of text (ETX) character is used to inform the receiver that all the information within the block has been transmitted and normally terminates a block of data started with an STX or SOH. This character is also used to denote the beginning of the block check characters appended to a transmission block as an error detection mechanism. This communications control character is primarily used in the bisynchronous protocol and its receipt requires a status acknowledgement, such as an ACK or NAK.

EOT

The end of transmission (EOT) character defines the end of transmission of all data associated with a message transmitted to a device. If transmission occurs on a multidrop circuit the EOT also informs other devices on the line to check later transmissions for the occurrence of messages that could be addressed to them. The EOT is also used as a response to a poll when the polled station has nothing to send and as an abort signal when the sender cannot continue transmission. In the XMODEM protocol the EOT is used to indicate the end of a file transfer operation.

ENQ

The enquiry (ENQ) communications control character is used in the bisynchronous protocol to request a response or status from the other station on a point to point line or to a specifically addressed station on a multidrop line. In response to the ENQ character, the receiving station may respond with the number of the last block of data it successfully received. In a multidrop environment, the mainframe computer would poll each device on the line by addressing the ENQ to one particular station at a time. Each station would respond to the poll positively or negatively, depending upon whether or not they had information to send to the mainframe computer at that point in time.

ACK and DLE

The acknowledgement (ACK) character is used to verify that a block of data was received correctly. After the receiver computes its own 'internal' checksum or cyclic code and compares it to the one appended to the transmitted block, it will transmit the ACK character if the two checksums match. In the XMODEM protocol the ACK character is used to inform the transmitter that the next block of data can be transmitted. In the bisynchronous protocol the data link escape (DLE) character is normally used in conjunction with the 0 and 1 characters in place of the ACK character. Alternating DLE0 and DLE1 as positive acknowledgement to each correctly received block of data eliminates the potential of a lost or garbled acknowledgement resulting in the loss of data. In some literature, DLE0 and DLE1 are referred to as ACK0 and ACK1.

NAK

The negative acknowledgement (NAK) communications control character is transmitted by a receiving device to request the transmitting device to retransmit the previously sent data block. This character is transmitted when the receiver's internally generated checksum or cyclic code does not match the one transmitted, indicating that a transmission error has occurred. In the XMODEM protocol this character is used to inform the transmitting device that the receiver is ready to commence a file transfer operation as well as to inform the transmitter of any blocks of data received in error. In the bisynchronous protocol, the NAK is also used as a station-not-ready reply to an ENQ line bid or a station selection.

SYN

The synchronous idle (SYN) character is employed in the bisynchronous protocol to establish and then maintain line synchronization between the transmitter and receiver during periods when no data is transmitted on the line. When a series of SYN characters is interrupted, this indicates to the receiver that a block of data is being transmitted.

ETB

The end of transmission block (ETB) character is used in the bisynchronous protocol in place of an ETX character when data is transmitted in multiple blocks. This character then indicates the end of a particular block of transmitted data that commenced with an SOH or STX character. A block check character (BCC) is sent after an ETB. The receipt of an ETB is followed by an acknowledgement by the receiving device, such as an ACK or NAK.

Information flow

Figure 1.70 illustrates in a time chart format the possible flow of information between a teletype compatible terminal and a computer system employing a basic

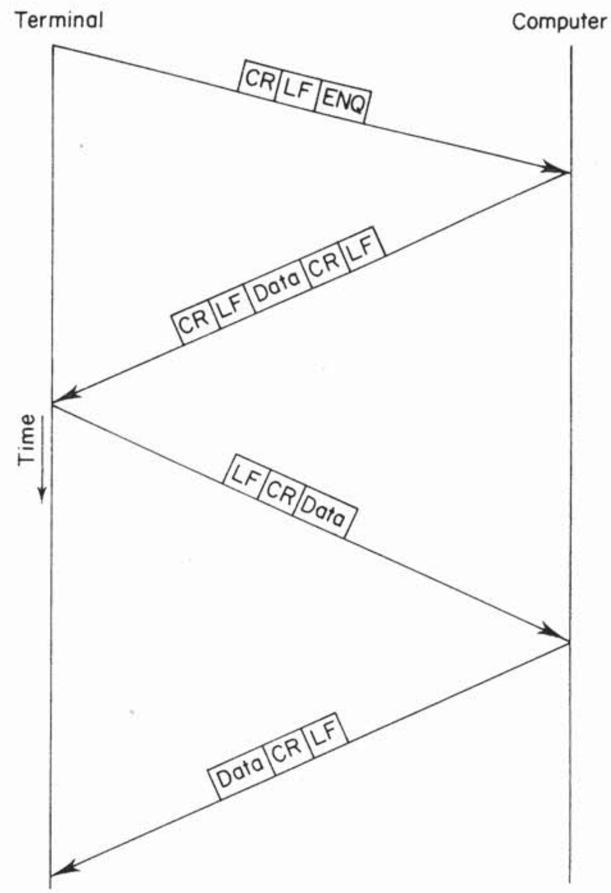

Figure 1.70 Basic teletype protocol

teletype protocol. In this protocol the terminal operator might first transmit the ENQ character, which is formed by pressing the shift and E keys simultaneously. If the call originated over the PSTN, the ENQ character, in effect, tells the computer to respond with its status. Since the computer is beginning its servicing of a new connection request, it normally responds with a log-on message. This log-on message can contain one or more lines of data.

The first line of the log-on message shown in Figure 1.70 is prefixed with a carriage return (CR) line feed (LF) sequence, which positions the printhead to the first column on a new line prior to printing the data in the received log-on message line. The log-on message line, as well as all following lines transmitted by the computer, will have a CR LF suffix, in effect, preparing the terminal for the next line of data. Upon receipt of the log-on message the terminal operator keys in his or her log-on code, which is transmitted to the computer as data followed by the CR LF suffix which terminates the line entry.

Variations

There are numerous variations to the previously discussed teletypewriter protocol, of which space permits mentioning only two.

Some computers will not recognize an ENQ character on an asynchronous ASCII port. Those computers are normally programmed to respond to a sequence of two or more carriage returns. Thus, the sequence ENQ LF CR would be replaced by CR CR or CR CR CR.

With the growth in popularity and use of personal computers as terminals, it was found that the time delay transmitted by computers in the form of null characters to separate multiple lines of output from one another by time was not necessary. Originally, the transmission of one line was separated through the use of NUL characters by several character intervals from the next line. This separation was required to provide the electromechanical printer used on teletype terminals with a sufficient amount of time to reposition its printhead from the end of one line to the beginning of the next line prior to receiving the first character to be printed on the next line. Since a cursor on a video display can be repositioned almost instantly, the growth in the use of personal computers and video display terminals resulted in the removal of time delays between computer transmitted lines.

Today, some computer software designed to service asynchronous terminals, as well as personal computers, will prompt the terminal operator with a message similar to: "ENTER NULLS (0 TO 5)-". This message provides the terminal operator with the ability to inform the computer whether he or she is using an electromechanical terminal. If 0 is entered, the computer assumes the terminal has a CRT display and does not separate multiple lines transmitted from the computer by anything more than the standard CR LF sequence. If a number greater than zero is entered, the computer separates multiple lines by the use of the indicated number of null characters. The NUL character, also called a PAD character, is considered to be a blank character which is discarded by the receiver. Thus, transmitting one or more NUL characters between lines only serves to provide time for the terminal's printhead to be repositioned to column 1 and has no effect upon the received data. If you are accessing a computer system that assumes all users have CRT terminals or personal computers, more than likely the null message will not be displayed. Such systems assume all users do not require a timing delay between transmitted lines and do not insert NUL characters between lines.

Error control

What happens if a line hit occurs during the transmission of data when a teletype protocol is used? Unfortunately, the only error detection mechanisms employed by teletype terminals and computer ports that supports this protocol are parity checking and echoplex.

If parity checking is supported by the terminal, it may simply substitute and display a special error character received with a parity error. This places the responsibility for error detection and correction upon the terminal operator, who must first visually observe the error and then request the computer to retransmit the line containing the parity error. Similarly, echoplex requires the operator to visually note that the echoed character differs from the character key just pressed.

As previously discussed in Section 1.11, the response of a computer to a parity error can range from no action to the generation of a special symbol to denote the occurrence of a parity error. In fact, most asynchronous line by line protocols do not check for parity errors. These protocols use echoplex, which as previously explained can result in a false indication of a transmission error or the appearance that all is well even though a transmitted character was received in error.

XMODEM protocol

The XMODEM protocol which was originally developed by Ward Christensen has been implemented into many asynchronous personal computer communications software programs and is supported by a large number of bulletin boards. Figure 1.71 illustrates in a time chart format the use of the XMODEM protocol for a file transfer consisting of two blocks of data. As illustrated, under the XMODEM protocol the receiving device transmits a negative acknowledgement (NAK) character to signal the transmitter that it is ready to receive data.

The XMODEM protocol is a 'receiver-driven' protocol in that the receiver transmits a character as a signal for the transmitter to start its data transfer operation. Under the XMODEM protocol the receiver has a 10 s timeout. It transmits a NAK each time it times out; hence, if the software on the personal computer that is to transmit a file is not set up to do so a period of 10 s can transpire until transmission actually starts. In response to the NAK the transmitter sends a start of header (SOH) communications control character followed by two characters that represent the block number and the one's complement of the block number.

The block number used in the XMODEM protocol starts at 01, increments by 1, and wraps from a maximum value of 0FFH to 00H and not to 01H. The one's complement is obtained by subtracting the block number from 255. Next a 128-character data block is transmitted which in turn is followed by the checksum character. As previously discussed in Section 1.11, the checksum is computed by first adding the ASCII values of each of the characters in the 128-character block and dividing the sum by 255. Next, the quotient is discarded and the remainder is retained as the checksum.

If the data blocks are damaged during transmission, the receiver can detect the occurrence of an error in one of three ways. If the start of header is damaged, it will be detected by the receiver and the data block will be negatively acknowledged. If either the block count or the one's complement field are damaged, they will not be the one's complement of each other. Finally, the receiver will compute its own checksum and compare it to the transmitted checksum. If the checksums do not match this is also an indicator that the transmitted block was received in error.

Each of the preceding situations results in the block being considered to have been received in error. Then the receiving station will transmit a NAK character which serves as a request to the transmitting station to retransmit the previously transmitted block. As illustrated in Figure 1.71, a line hit occurring during the transmission of the second block resulted in the receiver transmitting a NAK and

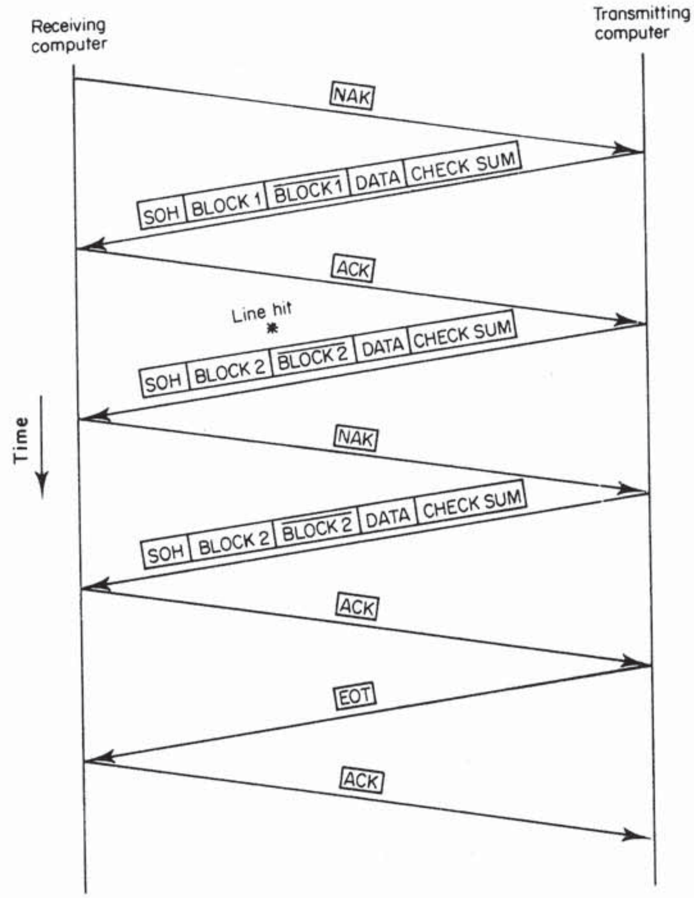

Figure 1.71 XMODEM protocol file transfer operation

the transmitting device resending the second block. Suppose more line hits occur which affects the retransmission of the second block. Under the XMODEM protocol the retransmissions process will be repeated until the block is correctly received or until ten retransmission attempts occur. If, owing to a thunderstorm or other disturbance, line noise is a problem, after 10 attempts to retransmit a block the file transfer process will be aborted. This will require a manual operator intervention to restart the file transfer at the beginning and is one of many deficiencies of the XMODEM protocol. Other deficiencies of the XMODEM protocol include its relatively small block size, its half-duplex transmission scheme, and its use of the checksum that provides a less reliable error detection capability in comparison to the use of a CRC.

In spite of the limitations of the XMODEM protocol, it is one of the most popular protocols employed by personal computer users for asynchronous data transfer because of several factors. First, the XMODEM protocol is in the public domain which means it is readily available at no cost for software developers to incorporate into their communications programs. Secondly, the algorithm employed to generate the checksum is easy to implement using a higher level language such as BASIC or Pascal. In comparison, a CRC-16 block-check character is normally generated using assembly language.

As a result of the previously mentioned limitations associated with the XMODEM protocol, several extensions to that protocol were developed. In addition, many commercial software developers designed proprietary file transfer protocols that were also structured to overcome one or more of the limitations associated with the XMODEM protocol. Five of the more popular extensions of the XMODEM protocol are XMODEM/CRC, YMODEM, YMODEM-G, XMODEM-1K and ZMODEM. Some of those protocols also have what are known as batch extensions which support the transfer of multiple (batched) files. Concerning commercial proprietary software protocols, two of the more popular are the BLAST and CrossTalk protocols. To provide readers with an indication of the advantages and disadvantages of each protocol we will compare and contrast several of those protocols to the original XMODEM protocol.

XMODEM/CRC protocol

The XMODEM/CRC is very similar to the XMODEM protocol, except that a two-byte CRC-16 is used in place of the one-character arithmetic checksum used with the original protocol. To differentiate the use of a CRC-16 from the use of a checksum, the receiver specifies the CRC-16 by transmitting the character C (Hex 43) instead of a NAK when requesting the first packet.

Figure 1.72 illustrates the block format of the XMODEM/CR protocol. In comparing the format of the XMODEM/CRC protocol to the XMODEM protocol, you will note the similarity between the two protocols, since only the error detection mechanism has changed.

SOH	Block Number	1's Complement of Block Number	128 Data Characters	CRC High Byte	CRC Low Byte

Figure 1.72 XMODEM/CRC block format

XMODEM/CRC

Through the use of a CRC, the probability of an undetected error is significantly reduced in comparison to the use of the XMODEM checksum. The CRC will detect all single- and double-bit errors, all errors with an odd number of bits, all bursts of errors up to 16 bits in length, 99.997% of 17-bit error bursts, and 99.998% of 18-bit and longer bursts.

Although the XMODEM/CRC protocol significantly reduces the probability of an undetected error, it is a half-duplex protocol similar to XMODEM and uses the same size data block. Thus, it removes only one of the three constraints associated with the XMODEM protocol.

YMODEM and YMODEM batch protocols

The YMODEM protocol was developed as an extension to XMODEM to overcome several constraints of the latter as well as to provide additional capabilities beyond those provided by both the XMODEM and XMODEM/CRC protocols. Under the YMODEM protocol, a header block was added to relay the filename and other information and multiple file transfers are supported in a batch mode. In addition, data is normally transferred in 1024-byte blocks, which results in more time being spent actually transferring data and less time spent computing checksums or CRCs and sending acknowledgements.

The original development of the YMODEM protocol was limited to transferring one file at a time using 1024-byte (1 K) blocks. Although many communications software programmes implemented YMODEM correctly as it was designed—as a single file protocol, other programs implemented it as a multiple file protocol. In actuality, the multiple file protocol version of YMODEM is normally and correctly referenced as YMODEM BATCH. Since YMODEM BATCH is the same as YMODEM except that the former allows multiple file (batch) transfers, we will examine both protocols and collectively refer to them as YMODEM, although this is not absolutely correct.

The format of the YMODEM protocol is illustrated in Figure 1.73. Under this protocol, the start of text (STX) character whose ASCII value is 02H replaces the SOH character used by the XMODEM and XMODEM/CRC protocols. The use of the STX character informs the receiver that the block contains 1024 data characters; however, the receiver can also accept 128 data character blocks. When 128 data character blocks are sent, the SOH character replaces the STX character.

STX	Block Number	1's Complement of Block Number	1024 Data Characters	CRC High Byte	CRC Low Byte

Figure 1.73 YMODEM block format

Under the YMODEM protocol multiple files can be transmitted through the use of a single command which, in some implementations, accept global characters. For example, specifying *.DAT would result in an attempt to transmit all files with the extension DAT. If no files with that extension are found, then zero files are transmitted.

When a data transfer is initiated, the receiver transmits the ASCII C character to the sender to synchronize transmission startup as well as to indicate that CRC checking is to be employed. The sender then opens the first file and transmits block number 0 instead of block number 1 used with the XMODEM and XMODEM/CRC protocols. Block number 0 will contain the filename of the file being transmitted and may optionally contain the file length and file creation date. Due to the manner in which most personal computer operating systems work, the creation or modification date of a file being downloaded will be modified to the current date when the file is received. For example, if the file being downloaded was created on JUN 15 1998 and today's date is JLY 19 1998, the file data would be changed to JLY 19 1998 when the file is downloaded onto your computer. The remaining data characters in block 0 are then set to nulls.

Once block 0 is correctly received, it will be ACKed if the receiver can perform a 'write open' operation. Otherwise, the receiver will transmit the CAN character to cancel the file transfer operation. After block 0 is acknowledged, the sender will commence transferring the contents of the file similar to the manner in which data is transmitted using the XMODEM/CRC protocol. During actual data transfer, the sender can switch between 128 and 1024 data character blocks by prefixing 128 character blocks with the SOH character and 1024 data character blocks with the STX character. After the contents of a file are successfully transmitted, the receiver will transmit an ASCII C which serves as a request for the next file. If no additional files are to be transmitted, the sender will transmit a 128 character data block, with the value of each character set to an ASCII 00H or null character.

Figure 1.74 illustrates the transmission of the file named STOCK.DAT which was last modified on JLY 19 1998 at 20:30 hours and which contains 2276 characters of information.

To initiate the file transfer, the receiver transmits an ASCII C to the sender. Upon its receipt the sender transmits a 128 character data block numbered as block 0. This block is prefixed with the SOH character to differentiate it from a 1024 data character block. The 'file info' field in block number 0 contains the filename (STOCK.DAT) followed by the time the file was created or last modified (20:30), the date the file was created or last modified (JLY 19 1998), and the file size in bytes (2276). A single space is used to separate the date from the file size, resulting in a total of 30 characters used to convey file information. Since the smallest data block contains 128 characters, 98 nulls are added to complete this block. After this block is acknowledged, the sender then transmits the first 1024 data characters through the use of a 1024 character data block, prefixing the block with the STX character.

In examining Figure 1.74, note that blocks 01 and 02 are 1024 characters in length, while blocks 03 and 04 are 128 characters in length. Since the file size was 2276 characters, the YMODEM protocol attempts to use as many 1024 data character blocks as possible and then transmits 128 character data blocks to complete the transmission. In doing so, the last 28 characters in block 4 are set to NULs.

Once the last block is successfully transferred, the sender transmits the EOT character to denote the completion of the file transfer. The receiver then transmits the ASII C as an indicator to the sender to initiate the transfer of the next file. Since only one file was to be transmitted, the sender transmits a new block number 00H that contains 128 NUL characters, signifying that no more files remain to be transmitted. Once this block is acknowledged, the transmission session is completed.

In addition to providing an increase in throughput over XMODEM and XMODEM/CRC, when transmission occurs over relatively noise-free lines the header information carried by YMODEM enables communications programs to compute the expected duration of the file transfer operation. This explains why

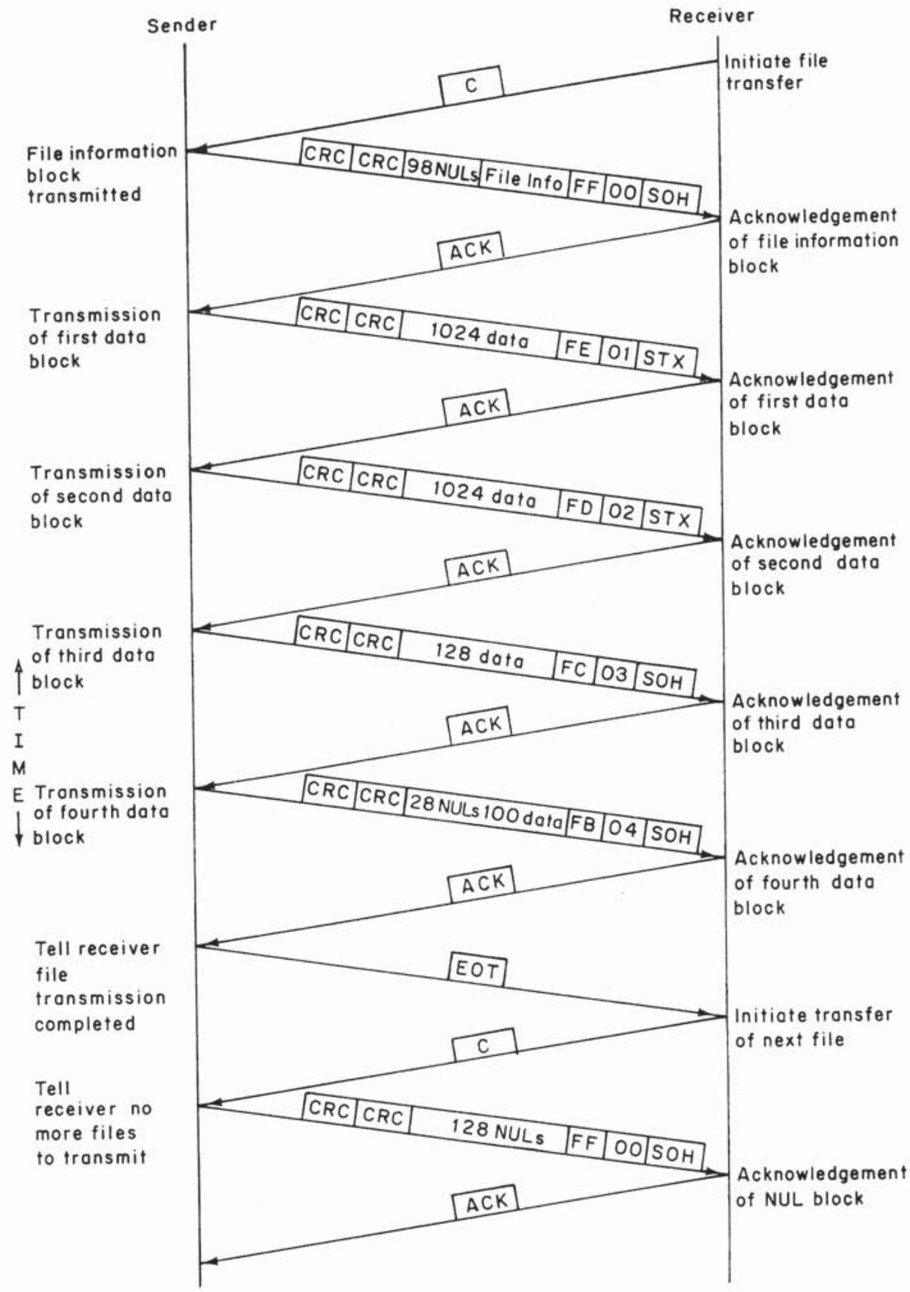

Figure 1.74 YMODEM protocol file transfer example

most communications programs will visually display the file transfer time and some programs will provide an updated bar chart of the progress of a YMODEM file transfer, while they cannot do the same when the XMODEM or XMODEM/CRC protocols are used.

XMODEM-1K protocol

The XMODEM-1K protocol is a derivative of the XMODEM standard. The XMODEM-1K protocol follows the previously described XMODEM protocol, substituting 1024-byte blocks in place of byte data blocks. The XMODEM-1K is not compatible with the YMODEM nor the YMODEM BATCH protocols, as the former does not send or accept a block 0, which contains file information. Since the block size of this protocol is significantly longer than that of the XMODEM protocol, you can expect a higher level of throughput when transmitting on good quality circuits using XMODEM-1K.

YMODEM-G and YMODEM-G BATCH protocols

The development of error correction and detection modems essentially made the use of CRC checking within a protocol redundant. In recognition of this, a 'G' option was originally added to the YMODEM protocol which changed it into a 'streaming' protocol in which all data blocks are transmitted one after another, with the receiver then acknowledging the entire transmission. This acknowledgement simply acknowledges the entire transmission without the use of error detection and correction. In fact, the two-byte CRC field is set to zero during a YMODEM-G transmission. Thus, this protocol should only be used with error correcting modems that provide data integrity. The use of error correcting modems is described in detail in Chapter 5.

Although some software programs enable users to initiate YMODEM-G by entering the character G as an optional parameter, most programs consider YMODEM-G as a separate protocol selected from a pull-down menu or via a command line entry.

Like the YMODEM BATCH protocol, YMODEM-G BATCH protocol permits multiple files to be transmitted and sends the first 128 data character block with file information in the same manner as carried by the YMODEM BATCH protocol. Typically, the multiple file transfer capability is selected by the use of a YMODEM-G BATCH option available with many communications programs. To differentiate YMODEM-G BATCH from YMODEM-G, the receiver will initiate the batch transfer by sending the ASCII G instead of the ASCII C. When the sender recognizes the ASCII G, it bypasses the wait for an ACK to each transmitted block and sends succeeding blocks one after another, subject to any flow control signals issued by an attached modem or by a packet network if that network is used to obtain a transmission path. When the transmission is completed, the sender transmits an EOT character and the receiver returns an ACK which serves to acknowledge the entire file transmission. The ACK is then followed by the receiver transmitting another ASCII G to initiate the transmission of the next file. If no additional files are to be transmitted, the sender then transmits a block of 128 characters with each character set to an ASCII 00H or NUL character.

Figure 1.75 illustrates the transmission of the previously described STOCK.-DAT file using the YMODEM-G protocol. In comparing Figure 1.75 with Figure

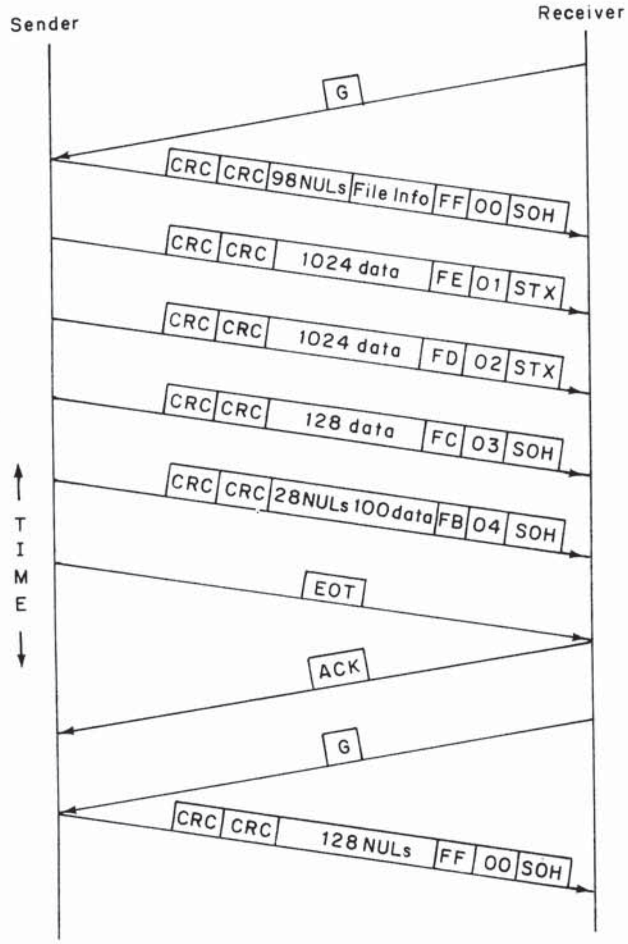

Figure 1.75 YMODEM-G protocol file transfer example

1.74, it becomes obvious that the streaming nature of YMODEM-G and YMODEM-G BATCH increases transmission throughout, resulting in a decrease in the time required to transmit a file or group of files.

ZMODEM

The development of the ZMODEM protocol was funded by Telenet which is now operated by Sprint as SprintNet. This packet switching vendor turned to Mr Chuck Forsberg, the author of the original YMODEM protocol, to develop a file transfer protocol that would provide a more suitable mechanism for transferring information via packet networks. The resulting file transfer protocol, which was called ZMODEM, corrected many of the previously described constraints associated with the use of the XMODEM and YMODEM protocols. Significant features of the ZMODEM protocol include its streaming file transfer operation, an extended error detection capability, automatic file transfer capability, the use of data compression, and downward compatibility with the XMODEM-1K and YMODEM protocols.

The streaming file transfer capability of ZMODEM is similar to that incorporated into YMODEM-G—that is, the sender will not receive an acknowledgement until the file transfer operation is completed. In addition to the streaming capability, ZMODEM supports the transmission of conventional 128- and 1024-byte block lengths of XMODEM-based protocols. In fact, ZMODEM is backward-compatible with XMODEM-1K and YMODEM.

The extended error detection capability of ZMODEM is based upon the ability of the protocol to support both 16- and 32-bit CRCs. According to the protocol developer, the use of a 32-bit CRC reduces the probability of an undetected error by at least five orders of magnitude below that obtainable from the use of a 16-bit CRC. In fact, 32-bit CRCs are commonly used with local area network protocols to reduce the probability of undetected errors occurring on LANs.

The automatic file transfer capability of ZMODEM enables a sending or receiving computer to trigger file transfer operations. In comparison, XMODEM and YMODEM protocols and their derivatives are receiver driven. Concerning the file transfer startup process, a file transfer begins immediately under ZMODEM while XMODEM and YMODEM protocols and their derivatives have a 10-second delay as the receiver transmits NAKs or another character during protocol startup operations.

An additional significant feature associated with the ZMODEM protocol is its support of data compression. When transmitting data between Unix systems, ZMODEM compresses data using a 12-bit modified Lempel–Ziv compression technique, similar to the modified Lempel–Ziv technique incorporated into the ITU-T V.42 bis modem recommendation. When ZMODEM is used between non-Unix systems, compression occurs through the use of Run Length Encoding similar to MNP Class 5.

Kermit

Kermit was developed at Columbia University in New York City primarily as a mechanism for downloading files from mainframes to microcomputers. Since its original development this protocol has evolved into a comprehensive communications system which can be employed for transferring data between most types of intelligent devices. Although the name might imply some type of acronym, in actuality, this protocol was named after Kermit the Frog, the star of the well-known Muppet television show.

Kermit is a half-duplex communications protocol which transfers data in variable sized packets, with a maximum packet size of 96 characters. Packets are transmitted in alternate directions since each packet must be acknowledged in a manner similar to the XMODEM protocol.

In comparison to the XMODEM protocol and its derivatives which permit 7- and 8-level ASCII as well as binary data transfers in their original data composition, all Kermit transmissions occur in 7-level ASCII. The reason for this restriction is the fact that Kermit was originally designed to support file transfers to 7-level ASCII mainframes. Binary file transfers are supported by the protocol prefixing each byte whose eighth bit is set by the ampersand (&) character. In addition, all characters transmitted to include 7-level ASCII must be printable, resulting in Kermit transforming each ASCII control character with the pound (£) character. This transformation is accomplished through the complementation of the seventh bit of the control character. Thus, 64 modulo 64 is added or subtracted from each control character encountered in the input data stream. When an 8-bit byte is encountered whose low order 7 bits represent a control character, Kermit appends a double prefix to the character. Thus, the byte 100000001 would be transmitted as &£A.

Although character prefixing adds a considerable amount of overhead to the protocol, Kermit includes a run length compression facility which may partially reduce the extra overhead associated with control character and binary data transmission. Here, the tilde (~) character is used as a prefix character to indicate run length compression. The character following the tilde is a repeat count, while the third character in the sequence is the character to be repeated. Thus, the sequence XA is used to indicate a series of 88 As, since the value of X is 1011000 binary or decimal 88. Through the use of run length compression the requirement to transmit printable characters results in an approximate 25% overhead increase in comparison to the XMODEM protocol for users transmitting binary files. If ASCII data is transmitted, Kermit's efficiency can range from more efficient to less efficient in comparison to the XMODEM protocol, with the number of control characters in the file to be transferred and the susceptibility of the data to run length compression the governing factors in comparing the two protocols.

Figure 1.76 illustrates the format of a Kermit packet. The header field is the ASCII start of header (SOH) character. The length field is a single character whose value ranges between 0 and 94. This one-character field defines the packet length in characters less two, since it indicates the number of characters to include the checksum that follow this field.

Header	Length	Sequence	Type	Data	Check

Figure 1.76 The Kermit packet format. The first three fields in the Kermit packet are one character in length and the maximum total packet length is 96 or fewer characters

The sequence field is another one-character field whose value varies between 0 and 63. The value of this field wraps around to 0 after each group of 64 packets is transmitted.

The type field is a single printable character which defines the activity the packet initiates. Packet types include D (data), Y (acknowledgement), N (negative acknowledgement), B (end of transmission or break), F (file header), Z (end of file) and E (error).

The information contents of the packet are included in the data field. As previously mentioned, control characters and binary data are prefixed prior to their placement in this field.

The check field can be one, two or three characters in length depending upon which error detection method is used since the protocol supports three options. A single character is used when a checksum method is used for error detection. When this occurs, the checksum is formed by the addition of the ASCII values of all characters after the Header character through the last data character and the low order 7 bits are then used as the checksum. The other two error detection methods supported by Kermit include a two-character checksum and a three-character 16-bit CRC. The two-character checksum is formed similar to the one-character checksum; however, the low order 12 bits of the arithmetic sums are used and broken into two 7-bit printable characters. The 16-bit CRC is formed using the CCITT standard polynomial, with the high order 4 bits going into the first character while the middle 6 and low order 6 bits are placed into the second and third characters, respectively.

By providing the capability to transfer both the filename and contents of files, Kermit provides a more comprehensive capability for file transfers than XMODEM. In addition, Kermit permits multiple files to be transferred in comparison to XMODEM, which requires the user to initiate file transfers on an individual basis.

Bisynchronous protocols

During the 1970s IBM's BISYNC (binary synchronous communications) protocol was one of the most frequently used for synchronous transmission. This particular protocol is actually a set of very similar protocols that provides a set of rules which effect the synchronous transmission of binary-coded data.

Although there are numerous versions of the bisynchronous protocol in existence, three versions account for the vast majority of devices operating in a bisynchronous environment. These three versions of the bisynchronous protocol are known as 2780, 3780 and 3270. The 2780 and 3780 bisynchronous protocols are used for remote job entry communications into a mainframe computer, with the major difference between these versions the fact that the 3780 version performs space compression while the 2780 version does not incorporate this feature. In comparison to the 2780 and 3780 protocols that are designed for point to point communications, the 3270 protocol is designed for operation with devices connected to a mainframe on a multidrop circuit or devices connected to a cluster controller which, in turn, is connected to the mainframe. Thus, 3270 is a poll and select software protocol.

Originally, 2780 and 3780 workstations were large devices that controlled such peripherals as card readers and line printers. Today, an IBM PC or compatible computer can obtain a bisynchronous communications capability through the installation of a bisynchronous communications adapter card into the PC's system unit. This card is designed to operate in conjunction with a bisynchronous communications software program which with the adapter card enables the PC to operate as an IBM 2780 or 3780 workstation or as an IBM 3270 type of interactive terminal.

The bisynchronous transmission protocol can be used in a variety of transmission codes on a large number of medium- to high-speed equipment. Some of the constraints of this protocol are that it is limited to half-duplex transmission and that it requires the acknowledgement of the receipt of every block of data transmitted. A large number of protocols have been developed owing to the success of the BISYNC protocol. Some of these protocols are bit-oriented, whereas BISYNC is a character-oriented protocol, and some permit full-duplex transmission, whereas BISYNC is limited to half-duplex transmission.

Data code use

Most bisynchronous protocols support several data codes including the 6-bit transcode (SBT), 7-bit ASCII and 8-bit EBCDIC. Normally, error control is obtained by using a two-dimensional parity check (LRC/VRC) when transmission is in ASCII. When transmission is in EBCDIC the CRC-16 polynomial is used to generate a block-check character, while the use of the SBT code is accompanied by the use of a CRC-12 polynomial.

Figure 1.77 illustrates the generalized bisynchronous block structure. For synchronization, most BISYNC protocols require the transmission and detection of two successive synchronization (SYN) characters. The start of message control code is normally the STX communications control character. The end of message control code can be either the end of text (ETX), end of transmission block (ETB), or the end of transmission (EOT) character; the actual character, however, depends upon whether the block is one of many blocks, the end of the transmission block, or the end of the transmission session.

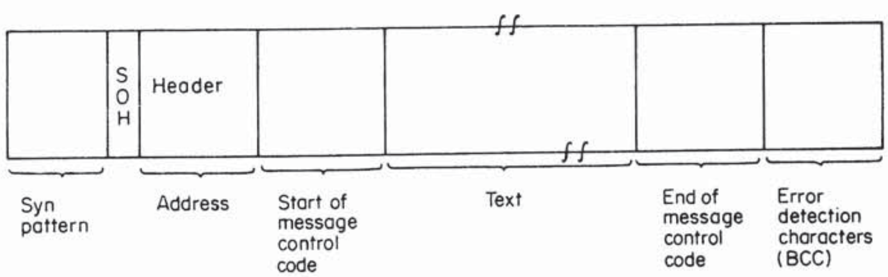

Figure 1.77 Generalized BSC block structure

The ETX character is used to terminate a block of data started with a SOH or STX character which was transmitted as an entity. SOH identifies the beginning of a block of control information, such as a destination address, priority and message sequence number. The STX character denotes both the end of the message header and the beginning of the actual content of the message. A BCC character always follows an ETX character. Since the ETX only signifies the end of a message, it

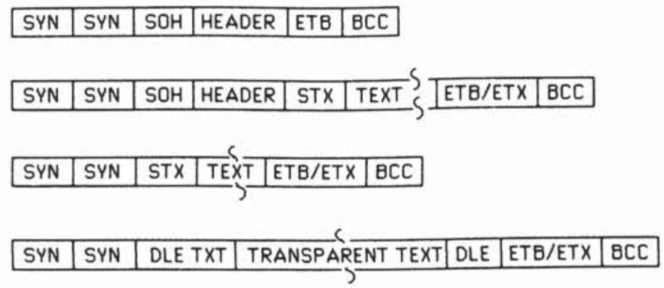

Figure 1.78 Common BISYNC data block formats

requires a status reply from the receiving station prior to subsequent communications occurring. A status reply can be an DLE0, DLE1, NAK, WACK or RVI character, with the meaning of the last three characters discussed later in this section.

The ETB character identifies the end of a block that was started with a SOH or STX. Similar to ETX, a BCC is sent immediately after the ETB and the receiving station is required to furnish a status reply.

The EOT code defines the end of message transmission for a single or multiple block message. The effect of the EOT is to reset all receiving stations. In a multidrop environment, the EOT is used as a response to a poll when an addressed station has no data to transmit.

The generalized block structure illustrated in Figure 1.77 can vary considerably based upon many factors, including the link topology (point-to-point or multipoint), operational mode (contention or polled), and the type of data to be transported by the protocol (alphanumeric or binary). Figure 1.78 illustrates four specific BISYNC data block formats commonly used.

The use of a header field is optional and the format and contents of that field are specified by the user. Typically, the header field is used for device selection purposes or for other routing information. The first block format shown in Figure 1.78 will normally precede the third block format, while the second block format can be viewed as a combination of the first and third formats. The fourth block format allows any bit pattern to be carried into the text field of the block while avoiding the possibility of the pattern being misinterpreted as a control character. This format, which permits text to be treated as transparent data, will be described later in this section.

Figure 1.79 illustrates the error control mechanism employed in a bisynchronous protocol to handle the situation where a line hit occurs during transmission or if an acknowledgement to a previously transmitted data block becomes lost or garbled.

In the example on the left portion of Figure 1.79, a line hit occurs during the transmission of the second block of data from the mainframe computer to a terminal or a personal computer. Note that, although Figure 1.79 is an abbreviated illustration of the actual bisynchronous block structure and does not show the actual block-check characters in each block, in actuality they are contained in each block. Thus, the line hit which occurs during the transmission of the second block

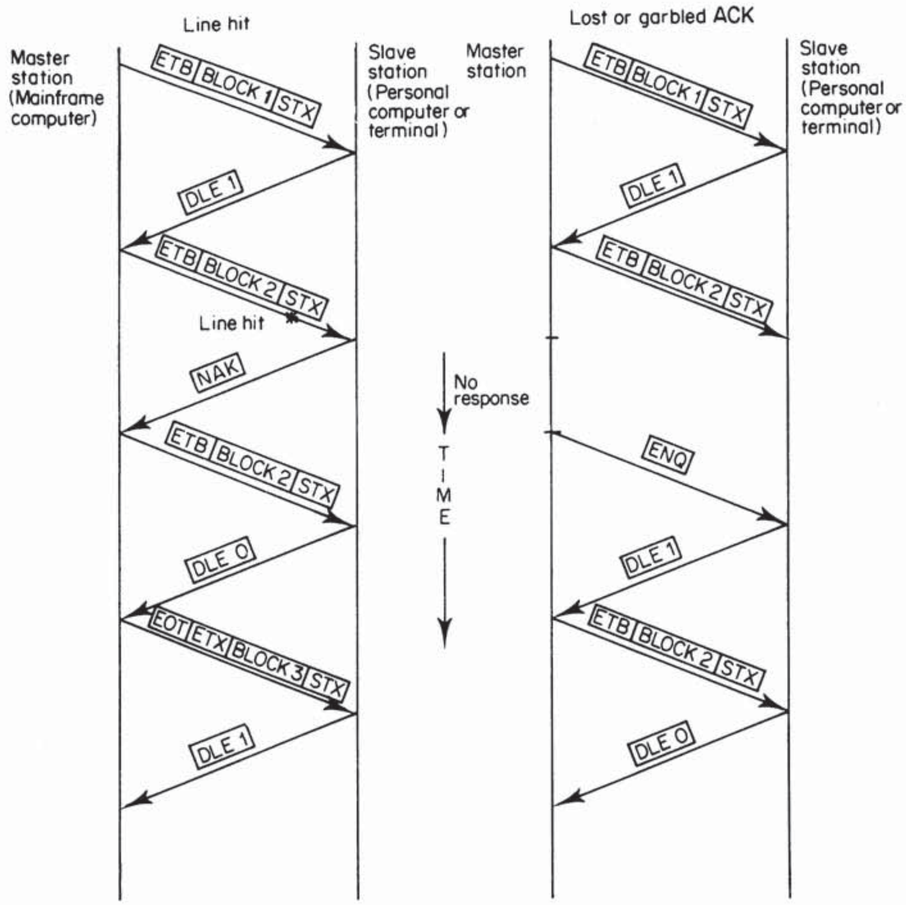

Figure 1.79 BSC, error control methods

results in the 'internally' generated BCC being different from the BCC that was transmitted with the second block. This causes the terminal device to transmit a NAK to the mainframe, which results in the retransmissions of the second block.

In the example on the right-hand part of Figure 1.79, let us assume that the terminal received block 2 and sent an acknowledgement which was lost or garbled. After a predefined timeout period occurs, the master station transmits an ENQ communications control character to check the status of the terminal. Upon receipt of the ENQ, the terminal will transmit the alternating acknowledgement, currently DLE1; however, the mainframe was expecting DLEO. Thus, the mainframe is informed by this that block 2 was never acknowledged and as a result retransmits that block.

Other control codes

Three additional transmission codes commonly used in a bisynchronous protocol are the WACK, RVI and TTD characters.

The wait-before-transmit affirmative acknowledgement (WACK) code is used by a receiving station to inform a transmitting station that the former is a temporary not-ready-to-receive condition. In addition to denoting the previously described condition, the WACK also functions as an affirmative acknowledgement (ACK) of the previously received data block. Once a WACK is received by the sending station, that station will normally transmit ENQs at periodic intervals to the receiving station. The receiving station will continue to respond to each ENQ with a WACK until it is ready to receive data.

The reverse interrupt (RVI) code is used by a receiving station to request the termination of a current session to enable a higher priority message to be sent. Similar to a WACK, the RVI also functions as a positive acknowledgement to the most recently received block.

The temporary text delay (TTD) is used by a sending station to keep control of a line. TTD is normally transmitted within 2 s of a previously transmitted block and indicates that the sender cannot transmit the next block within a predefined timeout period.

Although BISYNC usage has considerably declined during the past few years, it continues to retain a degree of popularity. Unfortunately, there are actually three versions of BISYNC you must consider, with each version slightly different with respect to the character composition supported, use of control codes, and the method error control. Table 1.31 summarizes the major differences between the three versions of the BISYNC protocol.

Table 1.31 Comparing BISYNC versions

	Character codes		
Function/control code;	SBT	ASCII	EBCDIC
Composition			
Number of bits per character	6	7	8
Number of defined characters	64	128	144 (256 possible)
Control code			
DLEO code	DLE-	DLEO	DLE70
DLEI code	DLET	DLE1	DLE/
WACK	DLEW	DLE;	DLE,
RVI	DLE2	DLE	DLE@
TTD	STX ENQ	STX ENQ	STX ENQ
Error control			
No transparency	CRC-12	VRC/LRC	CRC-16
Tranparent mode installed and operating	CRC-12	CRC-16	CRC-16
Transparent mode installed but not operating	CRC-12	VRC/CRC-16	CRC-16

Timeouts

Timeouts are incorporated into most communications protocols to preclude the infinite seizure of a facility due to an undetected or detected but not corrected error condition. The bisynchronous protocol defines four types of timeouts–transmit, receive, disconnect and continue.

The transmit timeout defines the rate of insertion of synchronous idle character sequences used to maintain synchronization between a transmitting and receiving station. Normally, the transmitting station will insert SYN SYN or DLE SYN sequences between blocks to maintain synchronization. Transmit timeout is normally set for 1 s.

The receive timeout can be used to limit the time a transmitting station will wait for a reply, signal a receiving station to check the line for synchronous idle characters or to set a limit on the time a station on a multidrop line can control the line. The typical default setting of the receive timeout is 3 s.

The disconnect timeout causes a station communicating on the PSTN to disconnect from the circuit after a predefined period of inactivity. The default setting for a disconnect timeout is normally 20 s of inactivity.

The fourth timeout supported by bisynchronous protocols is the continue timeout. This timeout causes a sending station transmitting a TTD to send another TTD character if it is unable to send text. A receiving station must transmit a WACK within two seconds of receiving the TTD if it is unable to receive.

Although the default timeout values are sufficient for most applications, there is one area where they almost always result in unnecessary problems—the situation where satellite communications facilities are used. Satellite communications add at least a 52 000-mile round trip delay to signal propagation, resulting in a built-in round trip delay of approximately 0.5 s. Due to this, you may always experience transmit and continue timeouts and can even experience many receive timeouts that are unwarranted if default timeout values are used. To eliminate the occurrence of unwarranted timeouts you should add 1 s to the default timeout values for each satellite 'hop' in a communications path, where a 'hop' can be defined as the transmission from one earth station to another earth station via the use of a satellite.

To illustrate the deterioration in a bisynchronous protocol when transmission occurs on a satellite circuit, assume you wish to transmit 80-character data blocks at 9.6 kbps and use modems whose internal delay time is 5 ms. Let us further assume there is a single satellite hop transmission will flow over, resulting in a one-way propagation delay of 250 ms. Since each message block must be acknowledged prior to the transmission of the next block, let us assume there are eight characters in each acknowledgement message. Based upon those assumptions, Table 1.32 lists each of the delay times associated with the transmission of one message block until an acknowledgement is received as well as the computation of the protocol efficiency.

There are three methods you can consider to improve throughput efficiency. You can increase the size of the message block, use high-speed modems or employ a full-duplex protocol. The first two methods have distinct limitations. As the size of the data block increases a point will be reached where the error rate on the data link results in the retransmission of the larger size message every so often, negating the efficiency increase from an increased block size. Since the data rate obtainable is a function of the bandwidth of a channel, it may not be practical to increase the data transmission rate, resulting in a switch to a full-duplex protocol being the

Table 1.32 Bisynchronous protocol efficiency example

		Time (ms)
Message transmission time	$\frac{80 \text{ characters} \times 8 \text{ bits/character}}{9600 \text{ bps}}$	67
Propagation delay		250
Modem delay time		10
Acknowledgement delay time	$\frac{8 \text{ characters} \times 8 \text{ bits/character}}{9600 \text{ bps}}$	7
Propagation delay		250
Modem dalay		10
		594

$$\text{Efficiency} = \frac{\text{Time spent transmitting data}}{\text{Total time to transmit and acknowledge}} = \frac{67}{594} = 11.3\%$$

method used by most organizations to increase efficiency when transmitting via a satellite link.

Data transparency

In transmitting data between two devices there is always a probability that the composition of an 8-bit byte will have the same bit pattern as a bisynchronous control character. This probability significantly increases if, as an example, you are transmitting the binary representation of a compiled computer program.

Since 8-bit groupings are examined to determine if a specific control character has occurred, a bisynchronous protocol would normally be excluded from use if you wished to transmit binary data. To overcome this limitation, protocols have what is known as a transparent mode of operation.

The control character pair DLE STX is employed to initiate transparent mode operations while the control character pairs DLE ETB or DLE ETX are used to terminate this mode of operation. Any control characters formed by data when the transparent mode is in operation are ignored. In fact, if a DLE character should occur in the data during transparent mode operations, a second DLE character will be inserted into the data by the transmitter. Similarly, if a receiver recognizes two DLE character in sequence, it will delete one and treat the second one as data, eliminating the potential of the composition of the bit patterns of the data causing a false ending to the transparent mode of operation.

DDCMP

Digital Equipment Corporation's Digital Data Communications Message Protocol (DDCMP) is a character-oriented data link protocol similar to IBM's bisynchronous protocol. Unlike IBM's protocol that is restricted to synchronous transmission, DDCMP can operate either asynchronously or synchronously over switched or non-switched facilities in a full- or half-duplex transmission mode.

Figure 1.80 illustrates the DDCMP protocol format, in which the header contains 56 bits partitioned into six distinct fields.

SYN	SYN	Type	Header						Information (any number of 8-bit characters)	CRC-16 (16 bits)
			Count (14 bits)	Flags (2 bits)	Response (8 bits)	Sequence (8 bits)	Address (8 bits)	CRC-16 (16 bits)		

Figure 1.80 Digital data communications message protocol (DDCMP) format

Structure

Like IBM's bisynchronous protocol, DDCMP uses two SYNC characters for synchronization. The type field is a one-character field which defines the type of message being transmitted. Data messages are indicated by a SOH character, while control messages which in DDCMP are either an ACK or NAK, are indicated by the ENQ character. A third type of message, maintenance, is denoted by the use of the DLE character in the type field.

When data is transferred the count field defines the number of bytes in the information field to include CRC bytes. One advantage of this structure is its inherent transparency, since the count field defines the number of bytes in the information field, the composition of the bytes will not be misinterpreted as they are not examined as part of the protocol. If the message is a control message, the count field is used to clarify the type of NAK indicated in the type field. Although there is only one type of ACK, DDCMP supports several types of NAKs, such as buffer overrun or the occurrence of a block-check error on a preceding message.

Table 1.33 lists the composition of the rightmost six bits of the count field which are used to define the reason for a NAK. Both ACK and NAK are denoted by an ENQ in the type field (00000101), so the count field is also used to distinguish between an ACK and NAK. If the first eight bits in the count field are binary 00000001, an ACK is defined, whereas, if the first eight bits in the count field are binary 00000010, this bit composition defines a NAK. Thus, the bit compositions listed in Table 1.33 are always prefixed by binary 00000010 which defines a NAK.

The high order bit of the flag field denotes the occurrence of a SYNC character at the end of the current message. This allows the receiver to reinitialize its synchronization detection logic. The low order bit of the FLAG field indicates the current message to be the last of a series the transmitter intends to send. This allows the addressed station to begin transmission at the end of the current message.

Table 1.33 Count field NAK definitions

Count field value (rightmost 6 bits)	NAK definition
000001	CRC header error
000010	CRC data error
000011	Reply response
001000	Buffer unavailable
001001	Receiver overrun
010000	Message too long
010001	Header format error

Both the response and sequence fields are used to transmit message numbers. DDCMP stations assign a sequence number to each message they transmit, placing the number in the sequence field. If message sequencing is lost, the control station can request the number of the last message previously transmitted by another station. When this request is received, the answering station will place the last accepted sequence number in the response field of the message it transmits back to the control station.

The address field is used in a multipoint line configuration to denote stations destined to receive a specific message. The following CRC1 field provides a mechanism for the detection of errors in the header portion of the message. This CRC field is required since error-free transmission depends upon the count field being detected correctly. The actual data is placed in the information field and, as previously mentioned, can include special control characters. Finally the CRC2 field provides an error detection and correction mechanism for the data in the information field.

Operation

Unlike IBM's bisynchronous protocol, DDCMP does not require the transmission of an acknowledgement to each received message. Only when a transmission occurs or if traffic is light in the opposite direction, a condition where no data messages are to be sent, is it necessary to transmit a special NAK or ACK.

The number in the response field of a normal header or in either a special NAK or ACK message is used to specify the sequence number of the last good message received. To illustrate this, assume messages 3, 4, 5 and 6 were received since the last time an acknowledgement was sent and message 7 contains an error. Then, the header in the NAK message would have a response field value of 6, indicating that messages 3, 4, 5 and 6 were received correctly and message 7 was received incorrectly. Under the DDCMP protocol up to 255 messages can be outstanding due to the use of an 8-bit response field.

Another advantage of DDCMP over IBM's bisynchronous protocols is the ability of DDCMP to operate in a full-duplex mode. This eliminates the necessity of line turnarounds and results in an improved level of throughput. Another function of the response field is to inform a transmitting station of the occurrence of a sequence error. This is accomplished by the transmitting station examining the contents of the response field. For example, if the next message the receiver expects is 4 and it receives 5, it will not change of the response field of its data messages which contains a 3. In effect, this tells the transmitting station that the receiving station has accepted all messages up through message 3 and is still awaiting message 4.

Bit-oriented protocols

A number of bit-oriented line control procedures were implemented by computer vendors that are based upon the International Organization for Standardization (ISO) procedure known as high-level data link control (HDLC). Various names for line control procedures similar to HDLC include IBM's synchronous data link control (SDLC) and Unisys' data link control (UDLC). We refer to each of these protocols as being bit oriented, as a receiver continuously monitors data bit by bit.

The advantages of bit-oriented protocols are threefold. First, their full-duplex capability supports the simultaneous transmission of data in two directions, resulting in a higher throughput than is obtainable in BISYNC. Secondly, bit-oriented protocols are naturally transparent to data, enabling the transmission of pure binary data without requiring special sequences of control characters to enable and disable a transparency transmission mode of operation as required with BISYNC. Lastly, most bit-oriented protocols permit multiple blocks of data to be transmitted one after another prior to requiring an acknowledgement. Then, if an error affects a particular block, only that block has to be retransmitted.

HDLC link structure

Under the HDLC transmission protocol one station on the line is given the primary status to control the data link and supervise the flow of data on the link. All other stations on the link are secondary stations and respond to commands issued by the primary station.

The vehicle for transporting messages on an HDLC link is called a frame and is illustrated in Figure 1.81. The frame provides a common format for all supervisory and information transfers. In addition, it provides a structure which contains fields that have a predefined general interpretation.

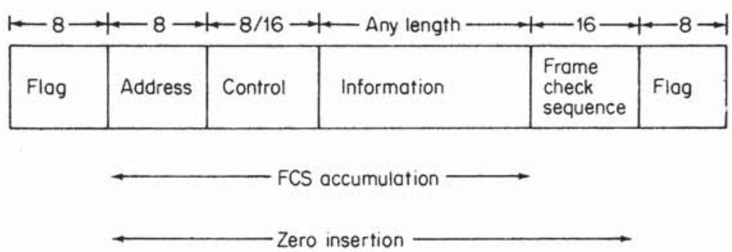

Figure 1.81 HDLC frame format. HDLC flag is 01111110 which is used to delimit an HDLC frame. To protect the flag and assure transparency the transmitter will insert a zero bit after a fifth 1 bit to prevent data from being mistaken as a flag. The receiver always deletes a zero after receiving five 1s

The HDLC frame contains six fields, wherein two fields serve as frame delimiters and are known as the HDLC flag. The HDLC flag has the unique bit combination of 01111110 (7EH), which defines the beginning and end of the frame. To protect the flag and assure transparency the transmission device will always insert a zero bit after a sequence of five 1-bits occurs to prevent data from being mistaken as a flag. This technique is known as zero insertion. The receiver will always delete a zero after receiving five ones to ensure data integrity.

The zero bit insertion technique insures that any string of more than five 1-bits will be interpreted as either a flag, a transmission error, or a deliberately

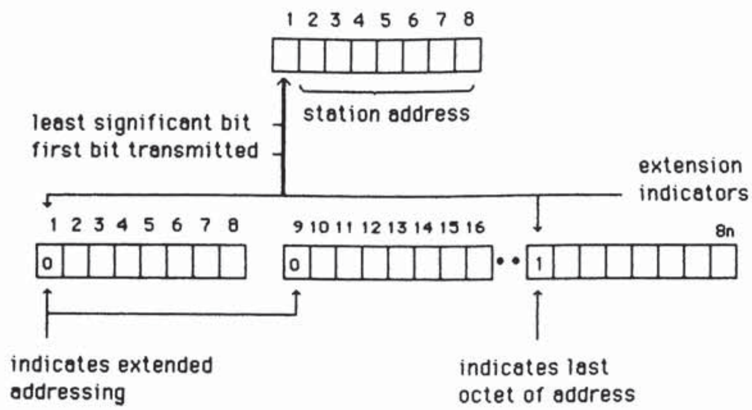

Figure 1.82 HDLC address field formats

transmitted fill pattern. Once a frame delimiter is recognized, the receiving device knows that the beginning of a frame has occurred when it receives an 8-bit non-flag address field. Upon detecting a subsequent frame delimiter, the receiving device knows that the frame has ended.

The address field is normally an 8-bit pattern that identifies the secondary station involved in the data transfer. If the least significant bit position of the address field is a zero, this indicates that extended addressing is used. The resulting extended address can contain any number of bytes and is terminated by a binary 1 in the least significant bit position of the last byte that constitutes the extended address. Figure 1.82 illustrates the HDLC address field formats.

The control field can be either 8 or 16 bits in length. This field identifies the type of frame transmitted as either an information frame or a command/response frame. Command frames are those transmitted by a primary station, while response frames are those transmitted by secondary stations. In the command frame, the address identifies the destination station for the command issued by the primary station. Similarly, in a response frame, the address field identifies the station transmitting the response. The information field can be any length and is treated as pure binary information, while the frame check sequence (FCS) contains a 16-bit value generated using a cyclic redundancy check (CRC) algorithm.

Control field formats

The 8-bit control field formats are illustrated in Figure 1.83. N(S) and N(R) are the send and receive sequence counts. They are maintained by each station for Information (I-frames) defined by the least significant bit having a value of 0 that are sent and received by that station. Each station increments its N(S) count by one each time it sends a new frame. The N(R) count indicates the expected number of the next frame to be received.

Using an 8-bit control field, the N(S)/N(R) count ranges from 0 to 7. Using a 16-bit control field the count can range from 0 to 127. There P/F bit is a poll/final bit. It is used as a poll by the primary (set to 1) to obtain a response from a

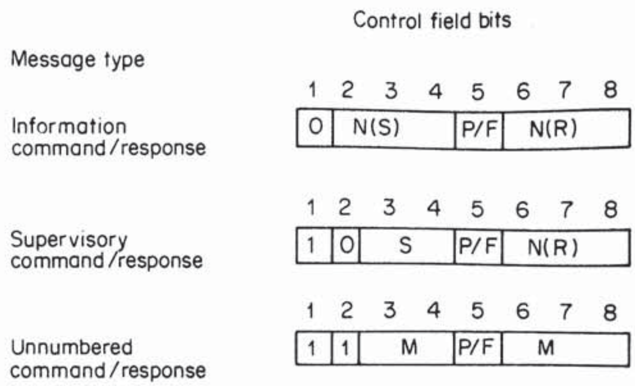

Figure 1.83 HDLC control field formats. N(S) = send sequence cont; N(R) = receive sequence count; S = supervisory function bits; M = modifier function bits; P/F = poll/final bit

secondary station. It is set to 1 as a final bit by a secondary station to indicate the last frame of a sequence of frames.

The supervisory command/response frame is used in HDLC to control the flow of data on the line. Figure 1.84 illustrates the composition of the supervisory control field: supervisory frames (S-frames) contain an N(R) count and are used to acknowledge I-frames, request retransmission of I-frames, request temporary suspension of I-frames, and perform similar functions.

As indicated in Figure 1.84, a supervisory frame is identified when the two least significant bits in the control field have a value of 01. Then, the value of the two following supervisory function bits indicate which of four functions are being invoked. If the first two bits in the control field are set to 11, a management frame referred to as an unnumbered command/response occurs. The latter reference results from the absence of N(S) and N(R) numbering fields. Five modifier (M) bits are used in the control field to define up to 32 general link control functions. Table 1.34 lists the unnumbered management functions presently defined for HDLC.

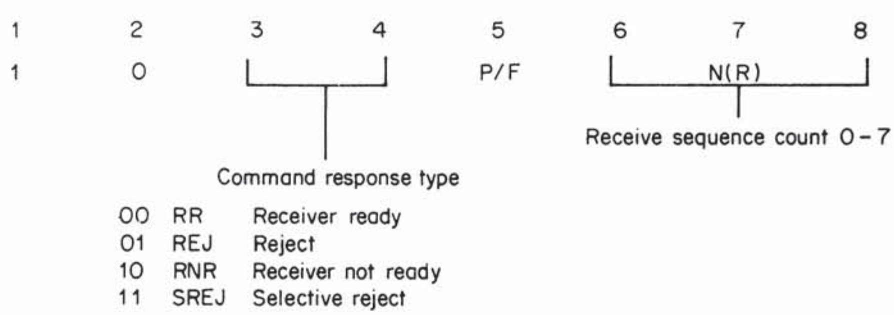

Figure 1.84 Supervisory control field

Table 1.34 Unnumbered command/response

Designation	Abbreviation	Command	Response
Disconnect	DISC	×	
Disconnect mode	DM		
Frame reject	FRMR		
Request initialization	RIM		×
Reset	RSET	×	×
Set asynchronous balanced mode	SABM	×	×
Set asynchronous balanced mode extended	SABME	×	
Set asynchronous response mode	SARM	×	
Set asynchronous response mode extended	SARME	×	
Set initialization mode	SIM		
Set normal response mode	SNRM	×	
Set normal response mode extended	SNRME	×	
Test	TEST	×	×
Unnumbered acknowledgement	UA		×
Unnumbered information	UI	×	×
Unnumbered poll	UP	×	
Exchange identification	XID	×	×

Operational modes

HDLC support three operational modes—normal response, asynchronous response, and asynchronous response balanced. In a normal response mode, a secondary station can only initiate transmission after receiving explicit permission from a controlling primary station. This operational mode is best suited for multipoint operations.

The asynchronous response mode is only applicable when there is one secondary station under primary control. In this operational mode, the secondary station can initiate transmission without having to receive explicit permission from a primary station.

The third operational mode supported by HDLC is asynchronous balanced. This operational mode enables the symmetrical transfer of data between two 'combined' stations on a point-to-point circuit. Here, each station has the ability to initialize and disconnect the circuit and is responsible for both controlling its own data flow and for recovering from error conditions. This operational mode is commonly used in packet switching. Figure 1.85 illustrates the difference between balanced and unbalanced operational modes.

To illustrate the advantages of HDLC over BISYNC transmission, consider the full-duplex data transfer illustrated in Figure 1.86. For each frame transmitted, this figure shows the type of frame, $N(S)$, $N(R)$ and poll/final (P/F) bit status.

In the transmission sequence illustrated in the left part of Figure 1.86, the primary station has transmitted five frames, numbered zero through four, when its poll bit is set in frame four. This poll bit is interpreted by the secondary station as a request for it to transmit its status and it responds by transmitting a receiver ready (RR) response, indicating that it expects to receive frame five next. This serves as an indicator to the primary station that frames zero through four were received correctly. The secondary station sets its poll/final bit as a final bit to indicate to the primary station that its transmission is completed.

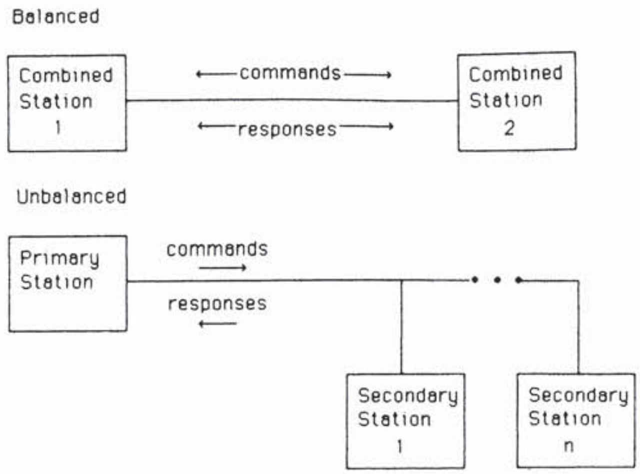

Figure 1.85 Balanced as opposed to unbalanced operations

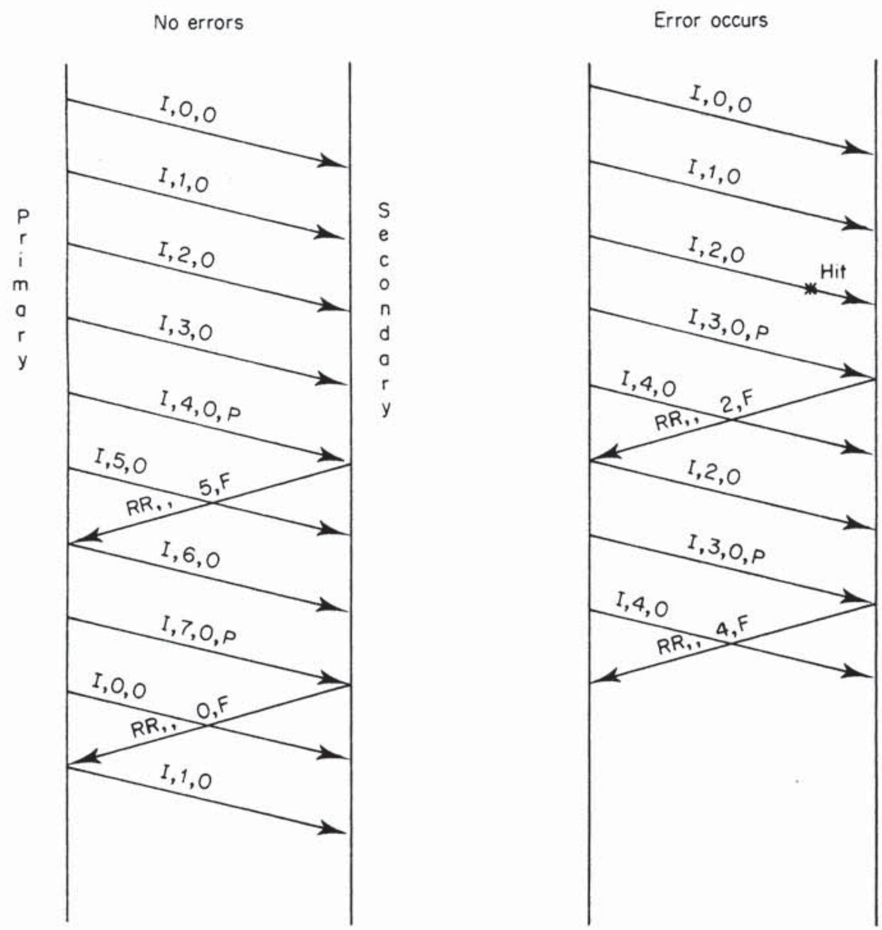

Figure 1.86 HDLC full-duplex data transfer. Format: Type, N(S), N(R), p/f

Note that since full-duplex transmission is permissible under HDLC, the primary station continues to transmit information (I) frames while the secondary station is responding to the primary's polls. If an 8-bit control field is used, the maximum frame number that can be outstanding is limited to seven since 3 bit positions are used for $N(S)$ frame numbering. Thus, after frame number seven has been transmitted, the primary station then begins frame numbering again at $N(S)$ equal to zero. Notice that when the primary station sets its poll bit when transmitting frame seven the secondary station responds, indicating that it expects to receive frame zero. This indicates to the primary station that frames five to seven were received correctly, since the previous secondary response acknowledge frames zero to four.

In the transmission sequence indicated on the right-hand side of Figure 1.86, assume a line hit occurs during the transmission of frame two. Note that in comparison to BISYNC, under HDLC the transmitting station does not have to wait for an acknowledgement of the previously transmitted data block; and it can

Table 1.35 Protocol characteristics comparison

Feature	BISYNC	DDCMP	SDLC	HDLC
Full duplex	No	Yes	Yes	Yes
Half duplex	Yes	Yes	Yes	Yes
Message format	Variable	Fixed	Fixed	Fixed
Link control	Control character, character sequences, optional header	Header (fixed)	Control field (8 bits)	Control field (8/16 bits)
Station addressing	Header	Header	Address field	Address field
Error checking	Information field only	Header information field	Entire frame	Entire frame
Error detection	VRC/LRC-8 VRC/CRC-16	CRC-16	CRC-ITU	CRC-ITU
Request for retransmission	Stop and wait	Go back N	Go back N	Go back N, selected reject
Maximum frames outstanding	1	255	7	127
Framing—start	2 SYNs	2SYNs	Flag	Flag
—end	Terminating characters	Count	Flag	Flag
Information transparency	Transparent mode	Inherent (count)	Inherent (zero insertion/ deletion)	Inherent (zero insertion/deletion
Control character	Numerous	SOH, DLE, ENQ	None	None
Character codes	ASCII EBCDIC Transcode (SBT)	ASCII (control character only)	Any	Any

continue to transmit frames until the maximum number of frames outstanding is reached; or, it can issue a poll to the secondary station to query the status of its previously transmitted frames while it continues to transmit frames up until the maximum number of outstanding frames is reached.

The primary station polled the secondary in frame three and then sent frame four while it waited for the secondary's response. When the secondary's response was received, it indicated that the next frame the secondary expected to receive $N(\mathrm{R})$ was two. This informed the primary station that all frames after frame one would have to be retransmitted. Thus, after transmitting frame four the primary station then retransmitted frames two and three prior to retransmitting frame four.

It should be noted that if selective rejection is implemented, the secondary could have issued a selective reject (SREJ) of frame two. Then, upon its receipt, the primary station would retransmit frame two and have then continued its transmission with frame five. Although selective rejection can considerably increase the throughput of HDLC, even without its use this protocol will provide the user with a considerable throughput increase in comparison to BISYNC.

For comparison purposes Table 1.35 compares the major features of BISYNC, DDCMP, IBM's SDLC and the ITU HDLC protocols.

Other protocols

Most of the previously mentioned protocols are restricted to use on wide area networks. Other protocols that are considerably more popular than those previously discussed operate on both LANs and WANs and will be covered in detail in Chapter 2 and 3. These protocols include Novell's NetWare IPX/SPX and TCP/IP, the latter being the only protocol that can be used on the Internet.

1.16 INTEGRATED SERVICES DIGITAL NETWORK

In this section we will examine both data and voice communications in the form of the Integrated Services Digital Network (ISDN), which at one time was expected to replace most, if not all, existing analog networks. Although ISDN has not lived up to the hype that surrounded its introduction, its service availability has considerably expanded since its introduction during the 1980s. One of the main limitations of ISDN availability was the requirement for local telephone companies to upgrade their switching infrastructure to support the technology. Many telephone companies postponed the upgrade of their switches through the mid-1990s as an economy measure while waiting for consumer demand for the service to develop. Since 1995 the upgrade of telephone switches has proceeded at a very high rate, resulting in ISDN service becoming available to approximately 80% of all telephone users in the United States by 1998. Thus, the limited availability of ISDN which acted as a constraint on its usage has considerably diminished over the past few years.

ISDN offers the potential for the development of a universal international digital network, with a series of standard interfaces that will facilitate the connection of a wide variety of telecommunications equipment to the network. Although the full transition to ISDN may require several decades and some ISDN functions may never be offered in certain locations, its potential cannot be overlooked. Since many ISDN features offer a radical departure from existing services and current methods of communications, we will review the concept behind ISDN, its projected features and services that can result from its implementation in this section.

Concept behind ISDN

The original requirement to transmit human speech over long distances resulted in the development of telephone systems designed for the transmission of analog data. Although such systems satisfied the basic requirement to transport human speech, the development of computer systems and the introduction of remote processing required a conversion of digital signals into an analog format. This conversion was required to enable computers and business machines to use existing telephone company facilities for the transmission of digital data. Not only was this conversion awkward and expensive due to the requirement to design high speed modems by developing and incorporating advanced encoding and error correcting techniques, but, in addition, analog facilities of telephone systems limit the data transmission rate obtainable when such facilities are used.

The evolution of digital processing and the rapid decrease in the cost of semiconductors resulted in the application of digital technology to telephone systems. By the late 1960s, telephone companies began to replace their electromechanical switches in their central offices with digital switches, while by the early 1970s, several communications carriers were offering end-to-end digital transmission services. By the mid-1980s, a significant portion of the transmission facilities of most telephone systems were digital, with over 99.9% of long distance transmission converted to digital by 1998. On such systems, human speech is encoded into digital format for transmission over the backbone network of the telephone system. At the local loop of the network, digitized speech is reconverted into its original analog format and then transmitted to the subscriber's telephone.

Based upon the preceding, ISDN can be viewed as an evolutionary progression in the conversion of analog telephone systems into an eventual all-digital network, with both voice and data to be carried end-to-end in digital form.

ISDN architecture

Under ISDN, network access functions that govern the methods by which user data flows into the network were separated from actual network functions, such as the manner by which signaling information is conveyed. In fact, ISDN's architecture resulted in a separate signaling network referred to as Common Channel Signaling Number 7 (CCS7) being used to convey signaling information. Although separate data and signaling networks govern the operation of ISDN, a single user–network interface provides a ubiquitous interface to ISDN network users.

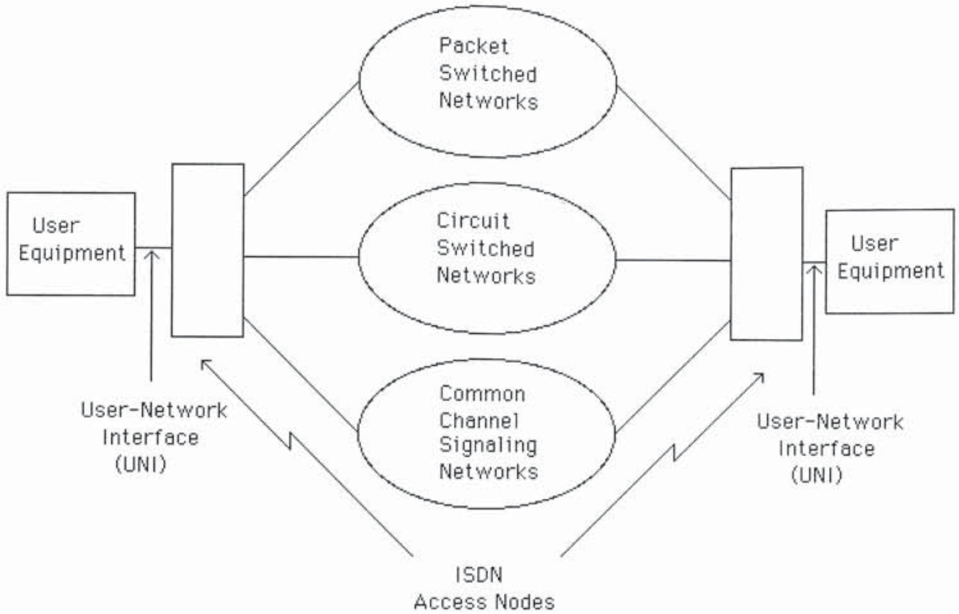

Figure 1.87 Basic ISDN architecture

Figure 1.87 illustrates the basic ISDN architecture to and between user–network interfaces. Note that under ISDN the user–network consists of both circuit switched and packet switched services. The circuit switched service provides for the routing of a call between access nodes to obtain an end-to-end connection similar to the manner by which calls are routed via the PSTN. The packet switched service enables packetized X.25 data transfer between access nodes as well as between access nodes and a device located on an X.25 packet network.

Types of service

Two types of ISDN are now standardized—Narrowband ISDN (N-ISDN) and broadband ISDN (B-ISDN). Broadband ISDN involves the logical grouping of N-ISDN facilities into a higher operating rate facility to obtain a high speed data transmission capability. Asynchronous Transfer Mode (ATM) evolved from the development work of B-ISDN. Since ATM is covered as a separate section in Chapter 2, we will focus our attention primarily upon N-ISDN in this section. In doing so we will examine the two major narrowband ISDN connection methods: basic access and primary access.

Basic access

Basic access defines a multiple channel connection derived by multiplexing data on twisted-pair wiring. This multiple channel connection is between an end-user terminal device and a telephone company office or a local Private Automated

F	B1 channel	D	B2 channel

Figure 1.88 ISDN basic access channel format

Branch Exchange (PABX). The ISDN basic access channel format is illustrated in Figure 1.88.

As indicated in Figure 1.88, basic access consists of framing (F), two bearer (B) channels and a data (D) channel that are multiplexed by time onto a common twisted-pair wiring media. Each bearer channel can carry one pulse code modulation (PCM) voice conversation or data at a transmission rate of 64 kbps.

PCM is a voice digitization technique which results in a data rate of 64 kbps being used to represent a voice conversation. Since there are two B channels this enables basic access to provide the end-user with the capability to simultaneously transmit data and conduct a voice conversation on one telephone line or to be in conversation with one person and receive a second telephone call. In the case of the latter situation, assuming the end-user has an appropriate telephone instrument, he or she could place one person on hold and answer the second call.

The basic access frame

The actual manner by which basic access framing is accomplished is significantly more complex than previously illustrated in Figure 1.88. In actuality, a variety of framing, echo, activate and DC balancing bits are spread out within a 48-bit basic rate interface frame along with B and D channel bits that transport data. The 48-bit frame actually carries only 36 data bits, with the remaining 12 bits representing line overhead. Since the frame is repeated 4000 times per second this results in a line operating rate of 48 bits/frame $\times$ 4000 frames/s, or 192 kbps, with the actual data transfer rate becoming 36 bits/frame $\times$ 4000 frames/s, or 144 kbps.

There are two frame formats used for ISDN basic access. One format is used when frames are transmitted from the network to ISDN terminal equipment. Since the frame flows from the Network Termination (NT) to the terminal equipment (TE) it is referred to as a basic access NT frame. The second type of ISDN basic access frame flows from the TE to the NT and, as you might expect, is referred to as a basic access TE frame. Figure 1.89 illustrates the format of the two basic-access frames.

In examining Figure 1.89 note that both NT and TE frames commence with a framing bit for frame alignment or synchronization. In an NT frame the L bits are used to electrically balance the entire frame. In comparison, in the TE frame the L bits are used to balance each octet of B channel information and each individual D channel bit. Through electrical balancing a situation where an excess number of binary zeros could occur that would prevent receiver synchronization with data is avoided. In addition, frame balancing limits DC voltage buildup and enables devices to communicate at greater distances from one another.

The A bit in the NT frame is used to activate or deactivate terminal equipment (TE), enabling the TE to be placed in a low power consumption mode when there

Basic Access NT Frame

F	L	Eight B1 bits	E	D	A	Fa	N	Eight B2 bits	E	D	S	Eight B1 bits	E	D	S	Eight B2 bits	E	D	L

Basic Access TE Frame

F	L	Eight B1 bits	L	D	L	Fa	L	Eight B2 bits	L	D	L	Eight B1 bits	L	D	L	Eight B2 bits	L	D	L

where: A = Activate/deactivate bit
B1 = B1 channel bits
B2 = B2 channel bits
D = D channel bit
E = D channel echo bit
F = Framing bit
S = Reserved for future standardization
Fa = Auxiliary framing bit
L = DC balancing bit
NT = Network termination
TE = Terminal equipment

Figure 1.89 ISDN basic-access frame formats

is no activity or to come on-line. The S bits are reserved for future standardization, while the E bits represent the echoes of previously transmitted D channel bits in TE frames. That is, when the NT receives a D channel bit from a TE the NT echoes the bit in the next E bit position in the basic access NT frame flowing to the TE.

The D channel

The D channel was designed for both controlling the B channels through the sharing of network signaling functions on this channel as well as for the transmission of packet switched data. Concerning the transmission of packet switched data, the D channel provides the capability for a number of applications to include monitoring home alarm systems and the reading of utility meters upon demand. Since these types of applications have minimum data transmission requirements, the D channel can be expected to be used for a variety of applications in addition to providing the signaling required to set up calls on the B channels.

When terminal equipment has D channel information to transmit it monitors the flow of E bits in the NT frame for a specified number of set bits that indicates the D channel is not in use. When that number is reached the TE will transmit its

Country Code	National Destination Code	Subscriber Number	ISDN Sub-Address
≤17 digits			≤40 digits

Figure 1.90 Basic ISDN address structure

D channel data to the NT. The number of set E bits the TE must receive depends upon the type of information to be transmitted. Signaling information has a higher priority than non-signaling information, resulting in eight set E bits that must be received for the TE to transmit signaling information, and 10 set E bits for transmitting non-signaling information.

Data carried by an ISDN D channel is encoded using the link access protocol D-channel (LAPD) format. LAPD is a layer 2 protocol defined by the ITU-T Q.921 recommendation whose frame format is identical to HDLC. The composition of the address field of Q.921, however, significantly differs from HDLC. ISDN addressing is standardized by the ITU I.331 Recommendation. That recommendation specifies a primary address up to 17 digits in length and an optional sub-address up to 40 digits in length. Figure 1.90 illustrates the basic structure of an ISDN address.

The country code, national destination code, and subscriber number uniquely identify each ISDN subscriber. This variable length number can be up to 17 digits in length, with the country code standardized in the ITU E.163 Recommendation. The optional sub-address field shown in Figure 1.90 enables organizations to extend their addressing within a private network which is accessible via ISDN. To provide interoperability between ISDN and other networks, such as the PSTN and public packet networks, the ITU developed a series of additional recommendations. Its I.330 series of recommendations define internetworking between ISDN and the North American Numbering Plan (NANP) for switched network telephone calls as well as for D channel transmission to and from X.25 public packet networks which use the ITU X.121 Recommendation for network addresses.

One of the more publicized features of ISDN is its calling line identification (CLID) capability. Essentially, CLID results in the transmission of the caller's telephone number via the D channel where it will be displayed on a liquid crystal display (LCD) built into most ISDN telephones, which are commonly referred to as digital telephones. Since the calling line identification is carried as a series of binary numbers, it becomes possible to integrate incoming numbers into a computer system. This makes it possible for a business to use the incoming number as a database search element. For example, an insurance company could route the calling number to their mainframe computer. As the telephone operator answers the call, the computer could simultaneously search a database and retrieve and display policy information on the operator's terminal. This capability not only enhances customer service, but, in addition, increases the productivity of organization employees.

Although calling line information can be received via a PSTN connection, that information is transferred between rings as a sequence of modem modulated symbols. This normally requires the use of a separate caller ID box to display the called number, although a few analog telephone sets now have a built-in caller ID display. While the displayed digits can be used in a manner similar to the previously described ISDN CLID capability, a specialized device would be required to frame the digits in a manner suitable for recognition by the computer. In comparison, the X.25 frame format used by the ISDN D channel for transporting CLID information represents a standardized method for conveying information.

Primary access

Primary access can be considered as a multiplexing arrangement whereby a grouping of basic access users shares a common line facility. Typically, primary access will be employed to directly connect a Private Automated Branch Exchange (PABX) to the ISDN network. This access method is designed to eliminate the necessity of providing individual basic access lines when a group of terminal devices shares a common PABX which could be directly connected to an ISDN network via a single high-speed line. Due to the different types of T1 network facilities in North America and Europe, two primary access standards have been developed.

In North America, primary access consists of a grouping of 23 B channels and one D channel to produce a 1.544 Mbps composite data rate, which is the standard T1 carrier data rate. In Europe, primary access consists of a grouping of 30 B channels plus one D channel to produce a 2.048 Mbps data rate, which is the T1 carrier transmission rate in Europe.

Other channels

In addition to the previously mentioned B and D channels, ISDN standards define a number of additional channels. These channels include the A, C and H series of channels.

The A channel is a 56 kbps wideband analog channel. The C channel is a digital channel that is used with the A channel during the transition to ISDN. The C channel operates at 8 or 16 kbps and carries signaling information similar to the manner in which the D channel controls B channels.

H channels are also known as broadband ISDN (B-ISDN) and are formed out of multiple B channels. For example, the H0 channel operates at 384 kbps and represents 6 B channels. Other H channels include the H11 channel which operates at 1.536 Mbps, the H12 channel which operates at 1.92 Mbps, and the H4 channel which operates at approximately 135 Mbps. The use of ISDN H channels provides a mechanism to interconnect local and wide area networks as well as to support such applications as full motion video teleconferencing and high-speed packet switching, the latter commonly referred to as frame relay. Both packet switching and frame relay are discussed in Chapter 2.

Network characteristics

Four of the major characteristics of an ISDN network are listed in Table 1.36. These characteristics can also be considered as driving forces for the implementation of the network by communications carriers.

Table 1.36 ISDN characteristics

Integrates voice, data and video services
Digital end-to-end connection resulting in high transmission quality
Improved and expanded services due to B and D channel data rates
Greater efficiency and productivity resulting from the ability to have several simultaneous calls occur on one line

Due to the digital nature of ISDN, voice, data, and video services can be integrated, alleviating the necessity of end-users obtaining separate facilities for each service. Since the network is designed to provide end-to-end digital transmission, pulses can be easily regenerated throughout the network, resulting in the generation of new pulses to replace distorted pulses. In comparison, analog transmission facilities employ amplifiers to boost the strength of transmission signals, which also increases any impairments in the signal. As a result of regeneration being superior to amplification, digital transmission has a lower error rate and provides a higher transmission signal quality than an equivalent analog transmission facility.

Due to basic access in effect providing three signal paths on a common line, ISDN offers the possibilities of both improvements to existing services and an expansion of services to the end-user. Concerning existing services, current analog telephone line bandwidth limitations normally preclude bidirectional data transmission rates over 33.6 kbps occurring on the switched telephone network. In comparison, under ISDN each B channel can support a 64 kbps transmission rate while the D channel will operate at 16 kbps. In fact, if both B channels and the D channel were in simultaneous operation a data rate of 144 kbps would be obtainable on a basic access ISDN circuit, which would extend current analog circuit data rates by a factor of 4.

Since each basic access channel in effect consists of three multiplexed channels, different operations can occur simultaneously without requiring an end-user to acquire separate multiplexing equipment. Thus, an end-user could receive a call from one person, transmit data to a computer and have a utility company read their electric meter at a particular point in time. Here, the ability to conduct simultaneous operations on one ISDN line should result in both greater efficiency and productivity. Efficiency should increase since one line can now support several simultaneous operations, while the productivity of the end-user can increase due to the ability to receive telephone calls and then conduct a conversation while transmitting data.

Terminal equipment and network interfaces

One of the key elements of ISDN is a small set of compatible multipurpose user–network interfaces that were developed to support a wide range of applications. These network interfaces are based upon the concept of a series of reference points for different user terminal arrangements which is then used to define these interfaces. Figure 1.91 illustrates the relationship between ISDN reference points and network interfaces.

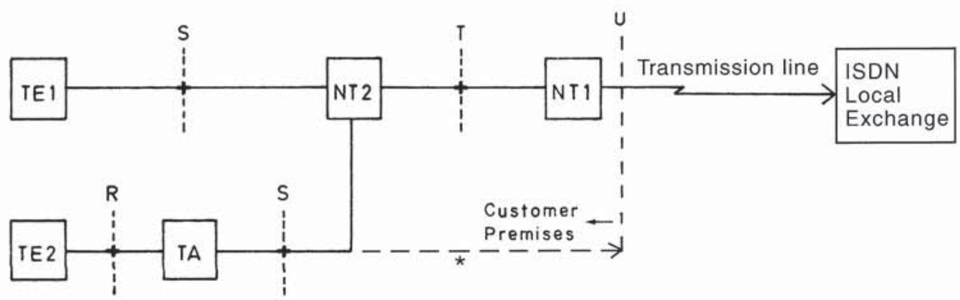

Figure 1.91 ISDN reference points and network interfaces. TE1 (Terminal Equipment 1) type devices comply with the ISDN network interface. TE2 (Terminal Equipment 2) type devices do not have an ISDN interface and must be connected through a TA (Terminal Adapter) functional grouping. NT2 (Network Termination 2) includes switching and concentration equipment which performs functions equivalent to layers 1 through 3 of the OSI Reference Model. NT1 (Network Termination 1) includes functions equivalent to layer 1 of the OSI Reference Model. A terminal adapter with a built-in NT1 can be directly connected to the U interface, eliminating the need for a separate NT1. (Reprinted with permission from *Data Communications Management,* © 1987 Auerbach Publishers, New York, NY.)

The ISDN reference configuration consists of functional groupings and reference points at which physical interfaces may exist. The functional groupings are sets of functions that may be required at an interface, while reference points are employed to divide the functional groups into distinct entities.

The TE (terminal equipment) functional grouping is comprised of TE1 and TE2 type equipment. Examples of TE equipment include digital telephones, conventional data terminals, and integrated voice/data workstations.

TE1

TE1 type equipment complies with the ISDN user–network interface and permits such equipment to be directly connected to an ISDN 'S' type interface which supports multiple B and D channels. TE1 equipment connects to ISDN via a twisted-pair four-wire circuit. Transmission is full-duplex and occurs at 192 kbps for basic access and at 1.544 or 2.048 Mbps for primary access.

TE2

TE2 type equipment are devices with non-ISDN interfaces, such as RS-232 or the ITU X or V-series interfaces. This type of equipment must be connected through a TA (terminal adapter) functional grouping, which in effect converts a non-ISDN interface (R) into an ISDN Sending interface (S), performing both a physical interface conversion and protocol conversion to permit a TE2 terminal to operate on ISDN.

Terminal adapters

Due to the large base of non-ISDN equipment currently in operation, the terminal adapter can be expected to play an important role as the use of this digital network expands. The terminal adapter (TA) performs a series of functions to convert non-ISDN equipment for use on the ISDN network. First, it must adapt the data rate of the non-ISDN device to either a 64 kbps B channel or a 16 kbps D channel operating rate. Next, it must perform the conversion of data from the non-ISDN device to a format acceptable to ISDN. For example, a non-ISDN device, such as an intelligent modem, might have its AT commands converted into ISDN D-channel signaling information. Other functions performed by TAs include the conversion of electrical, mechanical, functional, and procedural characteristics of non-ISDN equipment interfaces to those required by ISDN and the mapping of network layer data to enable a signaling terminal to be 'understood' by ISDN equipment.

Since a basic access channel operates at a multiple of most non-ISDN equipment rates, most terminal adapters include a multiple number of R interface ports. This allows, for example, an asynchronous modem connected to a personal computer, a facsimile machine, and a telephone to be connected to a basic access line via the R interface. In fact, most commercially available terminal adapters have three or four R interface ports.

Rate adaption

Rate adaption is the process during which the data rate of slow-speed devices is increased to the 64 kbps synchronous data rate of an ISDN B channel. During the rate adaption process, the data stream produced by a non-ISDN device is padded with dummy bits by the terminal adapter and clocked at a 64 kbps data rate.

In 1984, the CCITT approved its rate adaption standard known as the V.110 recommendation. Originally, this recommendation was strictly for synchronously operated devices and was modified in 1988 to support asynchronous devices. The framing specified by the V.110 recommendation is complex, with each 80-bit frame containing a 17-bit frame alignment pattern, while the actual rate adaption process can involve between one and three steps, with the actual number of steps dependent upon the operating rate of the terminal and its operating mode. Asynchronous devices require a three-stage process, while synchronous terminals operating below 64 kbps require a two-stage process as indicated in Figure 1.92.

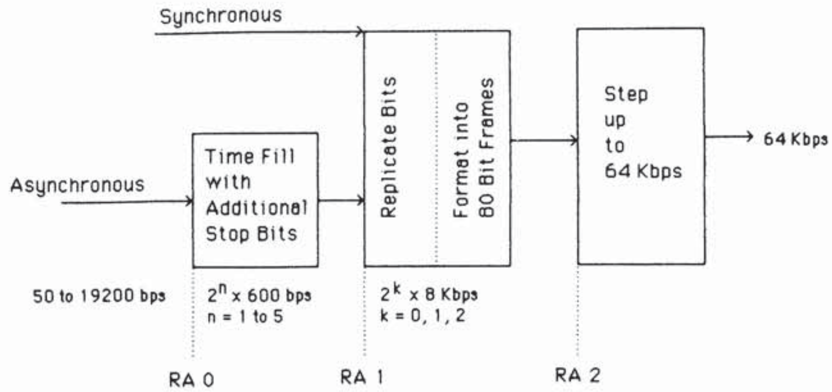

Figure 1.92 V.110 rate adaption

As indicated in Figure 1.92, asynchronous devices require a three-stage rate adaption process. In the first stage, extra stop bits are appended to each character to make the operating rate a multiple of 600 bps. The second stage of the V.110 rate adaption process services both asynchronous and synchronous data. In this stage, bits are replicated to create an 80-bit frame operating at an intermediate data rate of 8, 16 or 32 kbps. In the third stage, a process called bit positioning occurs in which one, two or four bits are added for each bit to bring the data rate up to 64 kbps. In addition to defining a complex rate adaption process, V.110 does not completely define flow control, fails to define a mechanism for error detection on its 80-bit frame, and is often difficult for the procedure defined by the recommendation to detect and adjust to a change in the data rate of a non-ISDN device. Due to these problems, a second rate adaption standard, which was originated in the United States by the ANSI T1E1 standards committee, has gained widespread acceptance. Known as V.120, this standard is based upon HDLC which makes flow control, error detection and correction, and other functions very easy to perform although they are either not included in or difficult to perform under the V.110 recommendation. The V.120 rate adaption is based upon the use of the LAPD protocol, with flag stuffing used to adapt the data rate to 64 kbps. Here the term 'flag stuffing' refers to the addition of a sufficient number of flag (01111110) bytes to bring the operating rate to 64 kbps. Today, V.120 is primarily used in the United States, while V.110 is primarily used in Europe and Japan. Both procedures provide support for V.24 and V.35 R interface.

Three additional rate adaption schemes that warrant mention include AT&T's Digital Multiplexed Interface (DMI), the ITU X.32 Recommendation, and Northern Telecom's T-Link. DMI represents AT&T's computer to PBX interface and is supported by their large 5ESS central office switches. In actuality there are three types of DMI rate adaption methods. DMI-1 supports 56 kbps data service and is compatible with the V.110 Recommendation at that data rate. DMI-2 supports rate adaption from devices operating below 20 kbps, such as regular RS-232 interfaces. DMI-3 is based on the ITU LAP-D protocol and is similar to the V.120 rate adaption scheme.

Table 1.37 ITU terminal adapter standards

Feature	V.110	V.120	X.31
ISDN Beaver service	Circuit	Circuit	Circuit/packet
B-channel multiplexing	Q.931	LinkID	Channel number
Error detection	None	CRC and V.41	CRC and V. 41
Error correction	None	Retransmission	Retransmission
Flow control	Unidirectional	Yes	Yes
HDLC-based	No	Yes	Yes
Multiple destinations	No	No	Yes
Rate adaption method	Multistep	Brit stuffing	Bit stuffing
Type of DTE/DCE at R-interface	Asynchronous and synchronous	Asynchronous, HDLC, transparent	X.25 synchronous

The X.31 Recommendation governs the rate adaption of packet mode X.25 equipment to an ISDN channel. Under the X.31 Recommendation call control procedures between X.25 and Q.931 are defined, enabling X.25 signaling to be converted to ISDN's and vice versa. The Northern Telecom T-Link rate adaption scheme is a circuit-mode terminal adapter protocol used in that vendor's ISDN terminal products. T-Link adapts the user data rate to an ISDN 64 kbps channel for transmission through Northern Telecom DMS-100 central office switches.

Until 1991 differences in rate adaption methods and ISDN line provisioning made the installation and configuration of an ISDN circuit a most challenging process. In 1991 Bellcore defined a National ISDN-1 standard which defines the method by which ISDN capable devices signal their status, such as busy, available, or no answer to carrier switches. This was a significant step in enabling different rate adaption methods to correctly pass required signaling information, for example to AT&T 5ESS and Northern Telecom DMS-100 switches.

A second significant event occurred a few years later with the development of a standard set of ISDN ordering codes. This enables telephone company installers to easily configure their ISDN line to work with user equipment manufactured to operate in a predefined manner. Table 1.37 provides a comparison of the nine features between the V.110, V.120, and X.31 standardized rate adaption Recommendations.

NT1

The NT1 (network termination 1) functional group is the ISDN digital interface point and is equivalent to layer 1 of the OSI reference model. Functions of NT1 include the physical and electrical termination of the loop, line monitoring, timing, and bit multiplexing. In Europe, where most communications carriers are government owned monopolies, NT1 and NT2 functions may be combined into a common device, such as a PABX. In such situations, the equipment serves as an NT12 functional group. In comparison, in the United States the communications carrier may provide only the NT1, while third-party equipment would connect to the communications carrier equipment at the T interface.

NT2

The NT2 (network termination 2) functional group includes devices that perform switching and data concentration functions equivalent to the first three layers of the OSI Reference Model. Typical NT2 equipment can include PABXs, terminal controllers, concentrators, and multiplexers.

Interfaces

As previously explained, the R interface is the point of connection between non-ISDN equipment and a terminal adapter. Although the R interface can consist of any common DTE interface, most terminal adapters support RS-232 and V.35.

The S interface is the standard interface between the TE and NT1 or between the TA and NT1. This is the interface to a 192 kbps, 2B + D, four-wire circuit. The S/T interface can operate up to distances of 1000 m using a pseudo-ternary coding technique. In this coding technique, a binary one is encoded by the transmission of an electrical zero, while binary zeros are encoded by transmitting alternating positive and negative pulses. Since there are three signal states (+, 0, − voltage) used to encode two symbols, the coding method is called pseudo-ternary. An example of this coding technique is illustrated in Figure 1.93.

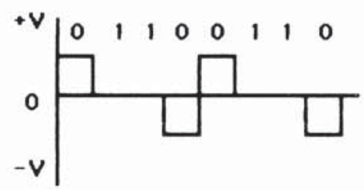

Figure 1.93 Pseudo-ternary coding example

The T interface is the customer end of an NT1 onto which you connect an NT2. For basic access, the T interface is a 192 kbps four-wire, 2B + D interface. For primary access, the T interface is a 1.544 Mbps, 23B + D, four-wire circuit or a 2.048 Mbps, 30B + D, four-wire circuit.

The U interface is the ISDN reference point that occurs between the NT1 and the network and is the first reference point at the customer premises. The coding scheme for information on the U interface is known as 2B1Q which is an acronym for 'two binary, one quaternary'. Under this coding scheme every two bits are encoded into one of four distinct states that are known as quats. The top portion of Figure 1.94 illustrates an example of 2B1Q encoding, while the lower portion of that illustration indicates the relationship between each dibit value and its quats code.

From the U interface, transmission occurs at 160 kbps to the telephone company central office. Due to the use of 2B1Q coding a maximum transmission distance of 18 000 feet is supported at a data rate of 160 kbps. This data rate represents 144 kbps used for the 2B + D channels and 16 kbps used for synchronization. In

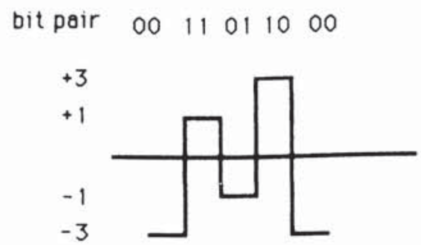

Figure 1.94 2BIQ coding example

Dibit	Quats
00	−3
01	−1
10	+3
11	+1

comparison, from the U interface to the S/T interface the data rate is 192 kbps, with 48 kbps used for synchronization and line balancing.

The future of ISDN

Although the use of ISDN has many distinct advantages, especially in the areas of dial Internet access and the use of multiple calls grouped together to obtain sufficient bandwidth required for videoconferencing, the use of this digital service has significantly fallen below expectations. Among the reasons provided for the slower than expected growth in the use of ISDN are its cost and pending competition in the form of digital subscriber lines (DSLs). Concerning its cost, installation of an ISDN line can easily exceed $300 and usage is typically billed at 5 or 10 cents/minute. When compared to the cost of a conventional analog line that can support bidirectional modem transfer at up to 33.6 kbps without incurring a $3 to $6 per hour usage surcharge, the price/performance ratio associated with the use of ISDN leaves much to be desired. Concerning DSL competition, although in its infancy, several types of digital subscriber lines have been developed that enable transmission rates up to 8 Mbps on the local analog loop between a subscriber and the serving telephone company office. Since DSL represents a modulation technology, it will be covered in detail in Chapter 5.

At the time this book revision was prepared, several field trials of DSL technology were being conducted and a few communications carriers and Internet Service Providers were offering commercial services using the technology. Although it is probably premature to predict the ultimate effect of DSL upon ISDN, it should be noted that ISDN represents a circuit switched and packet switched technology that enables calls to be routed. In comparison, DSL technology is limited to providing a high speed point-to-point access from a subscriber to a telephone company central office. Thus, the primary competition between the two will probably occur in the market for obtaining fast Internet access. This is because a subscriber requiring only fast Internet access would have previously used ISDN's basic access to group two B channels to obtain a 128 kbps data transfer capability to a single location. In those circumstances the use of DSL technology would provide a substitute that could be competitive depending upon its cost structure.

REVIEW QUESTIONS

To facilitate the reference of material in this book to review questions, each question has a three-part number. The first digit references the chapter related to the question. The second part of the number bounded by two decimal points references the section in the chapter upon which the review question is based. The third part of the question number references the question related to a specific section. For example, question 1.12.3 is the third question that references the material in Section 12 of Chapter 1.

1.1.1 Discuss the function of each of the major elements of a transmission system.

1.2.1 What is the relationship between each of the three basic types of line connections and the use of that line connection for short or long duration data transmission sessions?

1.3.1 Discuss the relationship of modems and digital service units to digital and analog transmission systems. Why are these devices required and what general functions do they perform?

1.3.2 Name four types of analog facilities offered by communications carriers and discuss the utilization of each facility for the transmission of data between terminal devices and computer systems.

1.3.3 Why is unipolar non-return to zero signaling unsuitable for use on a wide area network?

1.3.4 Why is a signaling technique that does not produce residual dc important on a wide area network?

1.3.5 What is a key advantage associated with the use of bipolar return to zero signaling?

1.3.6 What is the effect of 'bit robbing' on the ability to transmit data on a 64 kbps DS0 channel?

1.3.7 What is the function of an analog extension?

1.3.8 Name four types of digital transmission facilities offered by communications carriers and discuss the possible use of each facility by a large organization.

1.4.1 What is the difference between simplex, half-duplex, and full-duplex transmission?

1.4.2 Discuss the relationship between the modes of operation of terminals and computers with respect to the printing and display of characters on a terminal in response to pressing a key on the terminal's keyboard.

1.4.3 Assume your terminal is placed into a full-duplex mode of operation and you are accessing a similar operating computer. As you press keys on your keyboard, what do you see on your display?

1.5.1 In asynchronous transmission how does a receiving device determine the presence of a start bit?

1.5.2 What is the difference between asynchronous and synchronous transmission with respect to the timing of the data flow?

1.6.1 What is the difference between serial and parallel transmission? Why do most communications systems use serial transmission?

1.7.1 Discuss the difference in terminal requirements with respect to point-to-point and multidrop line usage.

1.8.1 What is a line discipline? Why is it required?

1.9.1 Discuss the use of three network topologies.

1.9.2 Why does a WAN designer normally attempt to structure their network to minimize the distance of circuits connecting geographically separated locations?

1.10.1 Discuss the difference between analog and digital transmission with respect to currently available operating rates.

1.11.1 Why is the Morse code basically unsuitable for transmission by terminal devices?

1.11.2 How does Baudot code, which is a 5-level code, permit the representation of more than 32 unique characters?

1.11.3 What is the bit composition of the ASCII characters A and a?

1.12.1 Assuming even parity checking is employed, what are the parity bits assigned to the ASCII characters A, E, I, O, and U? What are the parity bits if odd parity checking is employed?

1.12.2 What are the major limitations of parity checking?

1.12.3 Assume a file on your personal computer contains 3000 lines of data, with an average of 60 characters per line. If you transmit the file using 8-bit character transmission and the probability of an error occurring is 1.5 per 100 000 bits; how many characters can be expected to be received in error if the bit errors occur randomly and are singular in occurrence per transmitted character?

1.12.4 Under the XMODEM protocol, what would be the value of the checksum if the data contained in a block consisted of all ASCII X characters?

1.12.5 Discuss the relationship between a transmitted cyclic redundancy check character and an internally generated cyclic redundancy check character with respect to the data integrity of the block containing the transmitted cyclic redundancy check character.

1.13.1 Discuss the importance of having standards.

1.13.2 Discuss the difference between national, international, and *de facto* standards. Cite an example of each.

1.13.3 Why would it be in the best interest of a manufacturer to build a product compatible with appropriate standards, such as the RS-232/V.24 standard?

1.13.4 Discuss the applicability of FIPS and ANSI standards with respect to federal agencies and private sector firms.

1.13.5 Why can it take up to four years or more for the ITU to adopt a recommendation?

1.13.6 Name two sources of *de facto* communications standards.

1.13.7 What is a Request For Comment (RFC)?

1.13.8 What is the purpose of layer isolation in the OSI reference model?

1.13.9 What are the functions of nodes and paths in a network?

1.13.10 Discuss the seven OSI layers and the functions performed by each layer.

1.14.1 What is the primary difference between RS-232-C and RS-232-D with respect to interchange circuits and connectors?

1.14.2 What is the difference in connector requirements between RS-232, ITU V.24, V.35 and X.20 standards?

1.14.3 Discuss the relationship between the voltage to represent a binary one in a terminal and the RS-232 signal characteristics that represent a binary one.

1.14.4 What are three methods commonly used to refer to RS-232 circuits? Which method do you feel is most popular in industry? Why?

1.14.5 What is the purpose of the ring indicator signal? Why do some modems require two rings prior to answering a call?

1.14.6 What is the difference between internal and external timing?

1.14.7 What are two key limitations associated with RS-232? Describe how differential signaling associated with RS-449 and RS-530 alleviate a considerable portion of those limitations.

1.14.8 What is balanced signaling?

1.14.9 What is the primary application for using a V.35 interface?

1.14.10 What is the purpose of the RS-366-A interface?

1.14.11 Why is the X.21 interface more costly than an RS-232/V.24 interface?

1.14.12 What is the purpose of the X.21 bis interface?

1.14.13 What are the operating rate differences between the V.35 and HSSI interfaces?

1.14.14 What are the key differences between the use of the HSSI and HIPPI interfaces with respect to transmission distance?

1.14.15 What is a null modem? Why are pins 2 and 3 reversed on that cable?

1.15.1 What is the difference between a terminal protocol and a data link protocol?

1.15.2 What is the purpose of data sequencing in which a large block is broken into smaller blocks for transmission?

1.15.3 Define the characteristics of a teletype protocol.

1.15.4 What is the purpose in using one or more null character after a carriage return line feed sequence?

1.15.5 What error detection method is used in the teletype protocol? How are errors corrected when they are detected?

1.15.6 What is echoplex? When echoplex is used who is responsible for examining locally printed characters?

1.15.7 How can a receiver detect the occurrence of an error when the XMODEM protocol is used?

1.15.8 Discuss two limitations of the XMODEM protocol.

1.15.9 What are two advantages of the ZMODEM protocol in comparison to the XMODEM protocol?

1.15.10 What are the major differences between the 2780, 3780 and 3270 protocols?

1.15.11 What is the purpose of bisynchronous protocol transmitting alternating acknowledgements (DLE1 and DLE0)?

1.15.12 Why are alternating DLE0 and DLE1 characters transmitted as positive acknowledgements in bisynchronous transmission?

1.15.13 What procedure is used to prevent a stream of binary date from being misinterpreted as an HDLC flag? Explain the operation of this procedure.

1.15.14 What are the advantages of a bit-oriented protocol in comparison to a character-oriented protocol?

1.15.15 If a secondary station responds to the poll of a primary station by settin $N(\mathrm{R})$ equal to five in its response, what does this signify to the primary station?

1.15.16 How does the DDCMP protocol provide data transparency?

1.15.17 Assume the response field of a NAK message in a DDCMP protocol has a value of 14 and messages 12, 13, 14, 15 and 16 are outstanding. What does this indicate?

1.15.18 What is the advantage of a selective reject command?

1.16.1 Why can you expect transmission quality on ISDN facilities to be superior to existing analog facilities?

1.16.2 Discuss the data transmission rate differences between a basic access ISDN circuit and that obtainable on the switched telephone network.

1.16.3 What is the actual data transfer rate obtainable on an ISDN basic access line?

1.16.4 What is the purpose of the A bit in an NT frame?

1.16.5 What function does a terminal adapter perform?

1.16.6 What is pseudo-ternary coding? How would the bit sequence 1010 be encoded using this coding technique?

1.16.7 What is rate adaption?

1.16.8 Discuss the difference between the V.110 and V.120 rate adaption standards.

1.16.9 Discuss the use of the R, S, T and U interfaces.

1.16.10 How would the bit sequence 00101001 be encoded using 2B1Q coding?

1.16.11 In what application would digital subscriber line technology appear to compete with ISDN?

4

WIDE AREA NETWORK TRANSMISSION EQUIPMENT

One method of categorizing data communications components is by the function or group of functions they are designed to perform. In this chapter, the operation and utilization of networking devices designed primarily to provide data transmission via a wide area network communications medium will be covered. Specific devices which will be explored in this chapter include a variety of analog and digital transmission products that can be used to transmit data over analog or digital wide area network transmission facilities. Analog transmission devices covered in this chapter range in scope from the nearly obsolete acoustic coupler to a half-dozen types of modems, including state-of-the-art cable and digital subscriber line modems. Digital transmission devices covered in this chapter include different types of data and channel service units and a device that permits the extension of parallel transmission from a computer to peripheral units located at a distance from a computational facility.

4.1 ACOUSTIC COUPLERS

An acoustic coupler is in essence a modem which permits data transmission through the utilization of the handset of an ordinary telephone. Similar in functioning to a modem, an acoustic coupler is a device which will accept a serial asynchronous data stream from terminal devices, modulates that data stream into the audio spectrum, and then transmit the audio tones over a switched or dial-up telephone connection.

Acoustic couplers are equipped with built-in cradles or fittings into which a conventional telephone headset is placed. Through the process of acoustic coupling, the modulated tones produced by the acoustic coupler are directly picked up by the attached telephone headset. Likewise, the audible tones transmitted over a telephone line are picked up by the telephone earpiece and demodulated by the acoustic coupler into a serial data stream which is acceptable to the attached terminal.

Acoustic couplers normally use two distinct frequencies to transmit information, while two other frequencies are employed for data reception. A frequency from each pair is used to create a mark tone which represents an encoded binary one from the digital data stream, while another pair of frequencies generates a space tone which represents a binary zero. This utilization of two pairs of frequencies permits full-duplex transmission to occur over the two-wire switched telephone network.

Since acoustic couplers enable any conventional telephone to be used for data transmission purposes, the coupler does not have to be physically wired to the line. This permits considerable flexibility in selecting a transmission location, which can include pay telephones in airports and hard-wired telephones in hotel rooms. Acoustic couplers are manufactured both as separate units and as built-in units to data terminals, as shown in Figure 4.1. Due to the significant replacement of portable terminals by laptop computers with built-in modems with modular jacks, the market for both portable terminals and acoustic couplers has greatly diminished. However, if you read some mobile computing trade magazines you will still note advertisements for battery-powered acoustic couplers as they permit the transmission mobility many travelers require.

Figure 4.1 Varying coupler connections: (a) terminal with built-in coupler; (b) terminal connected to coupler

US and European compatibility

Since acoustic couplers are normally employed to permit portable data processing devices to communicate with data-processing facilities, and since a large portion of low-speed modems at such facilities in the United States were originally furnished by AT&T and its operating companies prior to its break-up into independent organizations, most manufacturers of acoustic couplers designed them to be compatible with low-speed 'Bell System' modems. Here the term 'Bell System' refers to the operating characteristics of modems that were manufactured by Western Electric for use by AT&T operating companies prior to those operating companies becoming independent organizations.

In Europe, most acoustic couplers are designed to be compatible with ITU recommendations that govern the operation of low-speed modems. To understand the differences between low speed Bell System and ITU modems, we will examine acoustic couplers that operate at data rates between 0 and 450 bps. In the United States, such couplers are compatible with Bell System 103 and 113 type modems while in Europe such couplers are compatible with the ITU V.21 recommendation. Table 4.1 lists the operating frequencies of acoustic couplers designed to operate with Bell System 103/113 type modems and modems that follow the ITU V.21 recommendation.

Table 4.1 Acoustic coupler modem compatibility (operating frequencies in Hz)

	Bell System 103/113 type		ITU V.21	
	Originate	Answer	Originate	Answer
Transmit				
Mark	1270	2225	980	1650
Space	1070	2025	1180	1850
Receive				
Mark	2225	1270	1650	980
Space	2025	1070	1850	1180

Couplers like low-speed modems must operate in one of two modes—originate or answer. This operational mode should not be confused with a transmission mode of simplex, half-duplex, or full-duplex. What the operational mode refers to is the frequency assignments for transmitting marks and spaces. Thus, from Table 4.1, an acoustic coupler compatible with a Bell System 103/113 type modem would transmit a tone at 1270 Hz to represent a mark and a tone at 1070 Hz to represent a space when it is in the originate mode of operation. To communicate effectively, the device (modem or coupler) at the other end of the line must be in the answer mode of operation. If so, then it would receive a mark at 1270 Hz and a space at 1070 Hz, ensuring that the tones transmitted by the originate mode device would be heard by the receive mode device. This explains why two terminal operators, each with an originate mode coupler, could not communicate with one another. This communications incompatibility results from the fact that one coupler would transmit a mark at 1270 Hz, while the other coupler would be set to receive the mark at 2225 Hz. Thus, the second coupler would never hear the tone originated by the first coupler.

By convention, originate mode couplers and modems are connected to terminals while answer mode devices are connected to computer ports, since terminals original calls and computers typically answer such calls. Some couplers can be obtained with an originate/answer mode switch. By changing the position of the switch, you change the coupler's operating frequency assignments.

In Figure 4.2 the frequency assignments of couplers designed to be compatible with Bell System 103/113 type modems is graphically illustrated. Note that 1170 Hz and 2125 Hz are the channel center frequencies and two independent data channels are derived by frequency, permitting full-duplex transmission to occur over the two-wire public switched telephone network.

Returning to Table 4.1, note that the operating frequencies of Bell System 103/113 type modems are completely different from modems designed to operate according to the V.21 recommendation. This frequency incompatibility explains why, for example, an American traveling in Europe will often be unable to use his or her portable personal computer to communicate with either a public packet network or his or her company's mainframe computer located in Europe.

Originally acoustic couplers were developed to transmit and receive data at 300 bps. Today, a few vendors market devices that operate at 1200 bps. Such couplers

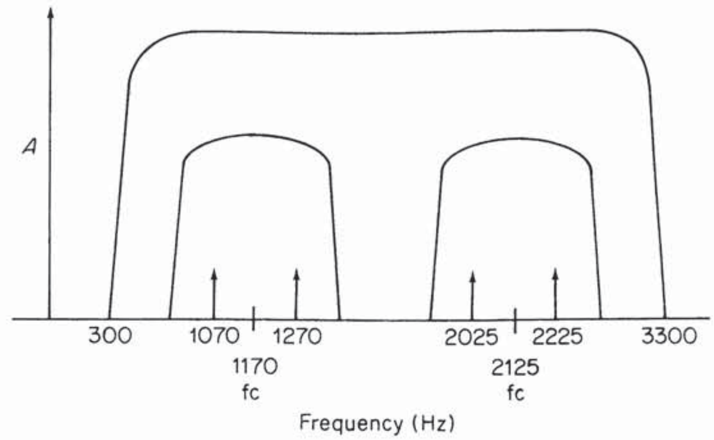

Figure 4.2 Bell System 103/113 frequency spectrum

are compatible with either Bell System 202 or 212A or ITU V.22 modems and their method of modulation will be described in Section 4.2 where modems are covered.

Operation

The operation of an acoustic coupler is a relatively simple process. An operator wishing to transmit data dials a telephone access number and upon establishing the proper connection by hearing a high-pitched tone, places the telephone headset into the coupler. Although usage varies by numerous applications, the prevalent utilization of acoustic couplers is in obtaining access to a packet network.

In a packet network, a group of dial-in computer telephone access numbers are interfaced to rotary which enables users to dial the low telephone number of the group and automatically 'step' or bypass currently busy numbers. Each telephone line is then connected to a modem on a permanent basis, and the modem in turn is connected to a computer port or channel. An automatic answering device in each modem automatically answers the incoming call and in effect establishes a connection from the user who dialed the number to the computer port, as shown in Figure 4.3. Once access is established to a packet network, the operator enters a routing code to establish a virtual path through the network to a corporate computer connected to the packet network.

Problems in Usage

A disadvantage associated with the use of acoustic couplers is a reduction of transmission rates when compared to rates which can be obtained by using modems. Owing to the properties of carbon microphones in telephone headsets, the frequency band that can be passed is not as wide as the band modems can pass. Although typical data rates of acoustic couplers vary between 110 and 300 bps,

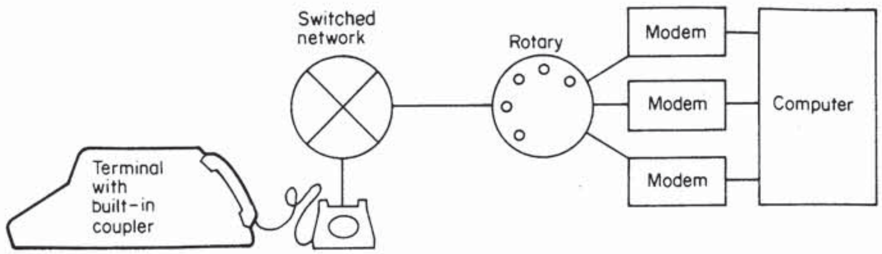

Figure 4.3 Network access in a time-sharing environment. After dialing the computer access number the terminal user places the telephone headset into the cradle of the acoustic coupler

some units manufactured permit transmission at 450, 600 and 1200 bps. For usage with slow-speed terminals, the acoustic coupler can be viewed as a low-cost alternative to a modem while increasing user transmission location flexibility.

One item which may warrant user attention is the placement of a piece of cotton inside the earpiece behind the receiver of the telephone. Although the placement of cotton at this location is normally done by most telephone companies, this should be checked, since the cotton keeps speaker and receiver noise from interfering with each other and acts to prevent transmitted data from interfering with received data.

An easily resolved problem is the placement of the telephone headset into the coupler. On many occasions users have hastily placed the handset only partially into the coupler, and this will act to reduce the level of signal strength necessary for error-free transmission.

4.2 MODEMS

Today, despite the introduction of a number of all-digital transmission facilities by most communications carriers, the analog telephone system remains the primary facility utilized for data communications. Since terminals and computers produce digital pulses, whereas telephone circuits are designed to transmit analog signals which fall within the audio spectrum used in human speech, a device to convert the digital data pulses of terminals and computers to analog tones that are carried on telephone circuits becomes necessary to transmit data over such circuits. Such a device is called a modem, which derives its meaning from a contraction of the two main functions of such a unit—modulation and demodulation. Although modem is the term most frequently used for such a device that performs modulation and demodulation, 'data set' is another common term whose use is synonymous in meaning.

In its most basic form a modem consists of a power supply, a transmitter, and a receiver. The power supply provides the voltage necessary to operate the modem's circuitry. In the transmitter a modulator and amplifier, as well as filtering, wave-shaping, and signal control circuitry convert digital direct current pulses; these

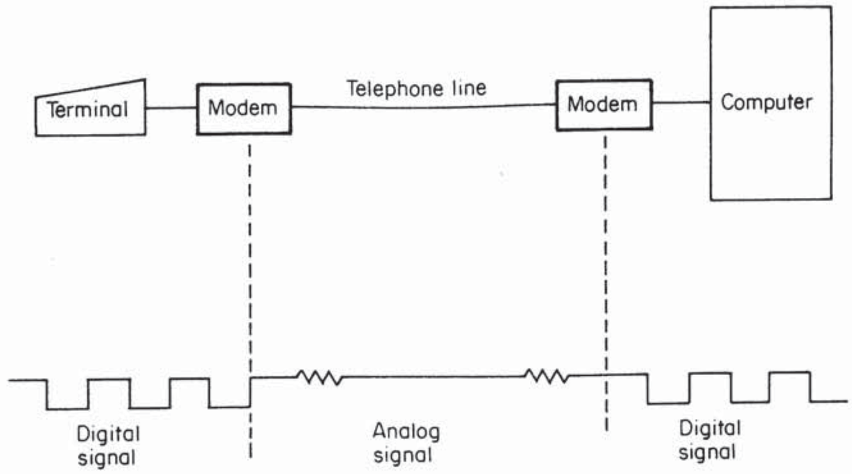

Figure 4.4 Signal conversion performed by modems. A modem converts a digital signal to an analog tone (modulation) and reconverts the analog tone (demodulation) into its original digital signal

pulses, originated by a computer or terminal, are converted into an analog, wave-shaped signal which can be transmitted over a telephone line. The receiver contains a demodulator and associated circuitry which reverse the process by converting the analog telephone signal back into a series of digital pulses that is acceptable to the computer or terminal device. This signal conversion is illustrated in Figure 4.4.

Basic components

In Figure 4.5 the basic components of a modem are indicated in a block diagram format, with those components associated with the transmitter at the top portion of the illustration, while those components associated with the receiver are located in the lower portion of the figure. Prior to examining each of the components, readers should note that Figure 4.5 represents a 'general' modem and those components indicated by dashed lines are only applicable to synchronous devices. In addition, other components, including a microprocessor, ROM and RAM, which can be included to provide a modem with intelligence have been purposely omitted from Figure 4.5 to enable us to focus our attention upon the modulation and demodulation of data by the transmitter and receiver in each device.

Modem transmitter section

The key components of a modem's transmitter section include a data encoder, scrambler, modulator, amplifier, filter, timing source and transmit control circuits. Of these components, the scrambler and transmit clock provided by the timing source are used only in synchronous modems.

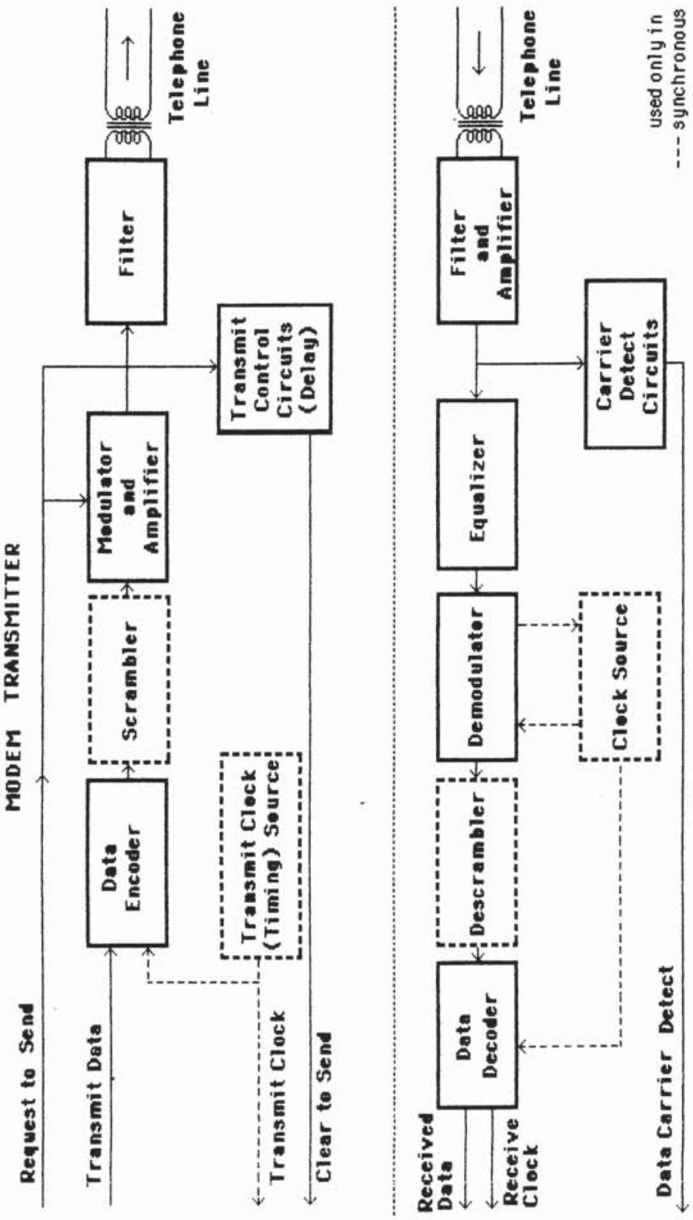

Figure 4.5 Basic components of a modem

The data encoder is an option built into many modems. The encoder is used in conjunction with some modulation schemes, enabling each signal change to represent more than one bit of information. We will discuss the use of data encoders and decoders later in this chapter when we examine the difference between a bit and a baud.

Scramblers

As previously discussed in Chapter 1, synchronous modems provide clocking signals on pins 15 and 17 of the RS-232 interface. When a modem receives a modulated synchronous data stream and passes the demodulated data to an attached terminal device, it also provides a clocking signal to the data terminal. This clocking signal tells the terminal device when to sample pin 3, the received data circuit, and is produced by the modem from the received data. Thus, the received clocking signal is commonly referred to as a derived clocking signal as it is derived from the received data.

For a synchronous modem's received clock to function correctly it must remain in synchronization with the data being received. This requires a sufficient number of changes in the composition of the data, e.g., 0 to 1 and 1 to 0, to permit the receiving modem's circuitry to derive timing from the received data. Since the data stream can consist of any arbitrary bit pattern, it is quite possible that the data will randomly contain long sequences of 0s or 1s. When these sequences occur the data will not provide the modem's receiver with a sufficient number of transitions for clock recovery, a condition which resulted in the incorporation of scramblers into synchronous modems.

A scrambler modifies the data to be modulated based upon a predefined algorithm. This algorithm is normally implemented through the use of a feedback shift register, which examines a defined sequence of bits and modifies their composition to ensure that every possible bit combination is equally likely to occur. At the receiving modem a descrambler employs the inverse of the predefined algorithm to restore the data into its original serial data stream.

Modulator, amplifier and filter

The modulator acts upon a serial data stream by using the composition of the data to alter the carrier tone which the modem places on the communications line. When a connection between two modems is established one modem will 'raise' a carrier tone that is heard by the distant modem. The reader is referred to 'The modulation process' in this section for information concerning the operation of different methods used to modulate data.

The amplifier boosts the level of the modulated signal for transmission onto the telephone line while the filter limits the frequencies of the tones placed on the line to comply with federal regulations. At the receiver, modulated tones received from the telephone line are filtered to remove extraneous tones caused by noise and then amplified to boost the received signal level.

Equalizer

The equalizer illustrated in the modem receiver section in Figure 4.5 is designed to measure the characteristics of a received analog signal and to adjust itself to that signal. In doing so the equalizer minimizes the effect of attenuation and delay upon the various components of a transmitted signal. To accomplish this task the modem's transmitter will prefix each transmission with a short 'training' signal whenever the direction of transmission changes. This training signal represents a predefined modulation of the carrier whose ideal reception characteristics are known by the equalizer in the receiver of the distant modem. Thus, the equalizer will be adjusted by the receiving modem until the best possible signal is received.

To obtain an appreciation for the operation and utilization of equalizers, let us first focus our attention upon several basic data channel parameters and the method by which communications carriers create a telephone channel. This will provide us with the foundation for discussing several analog signal impairments associated with transmission on the switched telephone network and the method by which equalizers can be used to compensate for those impairments.

Bandwidth

Bandwidth is a measurement of the width of a range of frequencies such that

$$B = f_2 - f_1$$

where B is the bandwidth, f_2 the highest frequency and f_1 the lowest frequency in a range of frequencies. Figure 4.6 illustrates the bandwidth of a telephone channel in comparison to the audio spectrum heard by the human ear. Here the unit hertz (Hz) is used to represent a cycle per second.

The 3000 Hz bandwidth which forms a telephone channel is commonly referred to as the passband of the channel. The term passband references a contiguous portion of an area in the frequency spectrum which permits a predefined range of frequencies to pass. Thus, the passband of a telephone channel permits frequencies between 300 and 3300 Hz to pass.

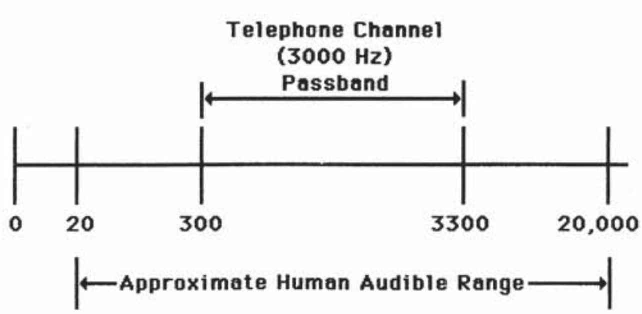

Figure 4.6 Bandwidth of a telephone channel

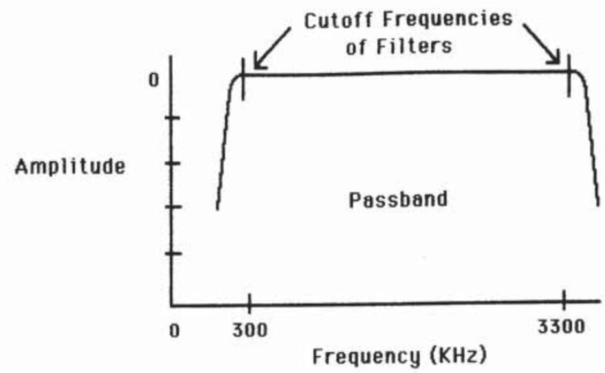

Figure 4.7 Telephone channel passband creation

The rationale for the telephone channel passband is economics. Frequencies under 300 Hz and above 3300 Hz are essentially not required to understand a telephone conversation even though it precludes a soprano from being fully appreciated at the other end of the telephone connection. By transmitting only 3000 Hz instead of the 20 000 Hz that the human ear can hear the bandwidth required for each call is reduced by a factor of approximately six. This bandwidth reduction enabled telephone companies to more efficiently employ frequency division multiplexing, a technique which allows many voice calls to be simultaneously carried on a common circuit routed between telephone company offices.

To construct a telephone channel passband the telephone company uses low and high-pass filters which are designed to permit either all signals up to a predefined frequency or all signals under a predefined frequency to pass through the channel. As a result of the use of filters the amplitude-frequency response becomes rounded at the cut-off frequencies at which the filters operate and then begin to approach large negative values as the filters' attenuation becomes more pronounced. Figure 4.7 illustrates how the use of filters results in the creation of a passband on a telephone channel.

Ideally, all frequencies across the passband of a telephone channel should undergo the same amount of attenuation as illustrated by the straight line between the cut-off frequencies shown in Figure 4.7. Unfortunately, high frequencies lose their strength more rapidly than low frequencies, which results in attenuation increasing as frequencies increase towards the end of the passband. In addition, attenuation increases as the edges of the operating frequencies of bandpass filters on a channel are approached. As a result of the two previously mentioned factors the amplitude-frequency response of a telephone channel which indicates the attenuation distortion that signals experience will resemble that illustrated in Figure 4.8.

To minimize the effect of attenuation distortion some modems include an attenuation equalizer. This type of equalizer introduces a variable gain at frequencies within the passband which compensate for the differences in attenuation between high and low frequencies as well as the increased attenuation at the edges

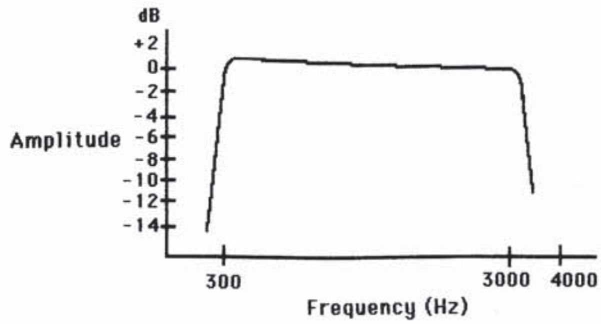

Figure 4.8 Typical amplitude–frequency response across a voice channel

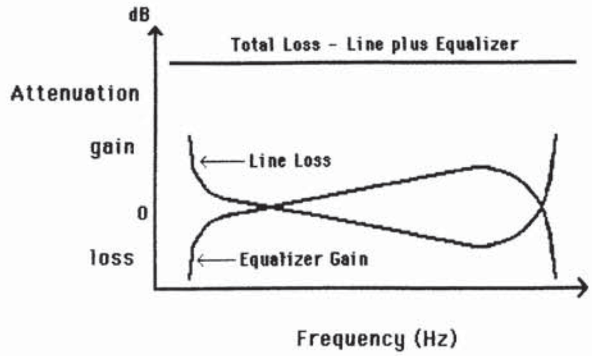

Figure 4.9 Using an equalizer to correct attenuation distortion

of the passband. Figure 4.9 illustrates the operation of an attenuation equalizer whose use results in a near uniform signal level across the passband.

Delay distortion

A second type of distortion which affects the recovery of information from a received signal is delay distortion. In a distortionless channel, all frequencies pass through the channel at the same speed. This results in the frequency and phase of the signal having a constant linear relationship with respect to time and ensures that the transmission of one signal will not interfere with the reception of a previously transmitted signal. Unfortunately, all channels except perhaps those in the laboratory have a degree of distortion. When distortion occurs, the relationship between the phase and frequency of a signal becomes non-linear. As the level of distortion increases, the relationship between the phase and frequency of a signal degenerates further. This degeneration, which is called phase delay, is measured at a particular point on the frequency spectrum by dividing the phase of the signal by

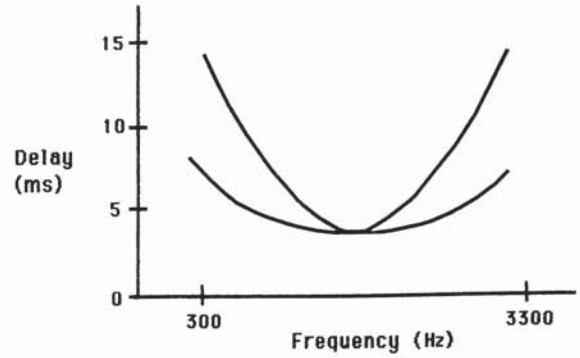

Figure 4.10 Typical envelope delay curves

its frequency. The direct measurement of phase delay is not practical due to the requirement to have an absolute phase reference and keep track of phase changes over multiples of 360°. This resulted in the use of the slope of the phase plotted against frequency, which is known as envelope delay.

Mathematically, envelope delay is the first derivative of phase delay. The shape of the envelope delay curve obtained by measuring delays at different frequencies reflects the degree of change in the slope of the phase versus frequency curve. This delay change varies based upon the transmission distance. Figure 4.10 illustrates two typical delay curves for signals transmitted on a telephone channel, with the steeper curve representing the envelope delay on a longer distance circuit than the flatter curve.

To illustrate the potential effect of envelope delay upon communications, assume a modem transmits one of two tones (f_1) to represent a binary zero and the second tone (f_2) to represent a binary one. This is one of the earliest methods used for modulation, a technique referred to as frequency shift keying or FSK. If the envelope delay curve is not symmetrical, consider what can happen if tone f_1 is transmitted, followed by tone f_2. Due to different delays associated with different frequencies there now exists the possibility that the delay in f_1 being received could result in that tone reaching the receiving modem at the same time that tone f_2, which represents a different binary value, is received. This could cause one received signal to be superimposed upon the second signal by time, resulting in one tone distorting the other tone.

Although all communications circuits will exhibit a degree of delay, it is important to flatten the delay time across the passband to minimize the potential for one tone of a signal to be superimposed on another tone by time. Some modems are designed with delay equalizers which introduce a delay approximately inverse to that exhibited by the telephone channel. Through the use of a delay equalizer the delay time associated with frequencies within the passband can be made relatively flat as illustrated in Figure 4.11. Doing so reduces the potential of one tone interfering with another, a condition formally referred to as intersymbol interference.

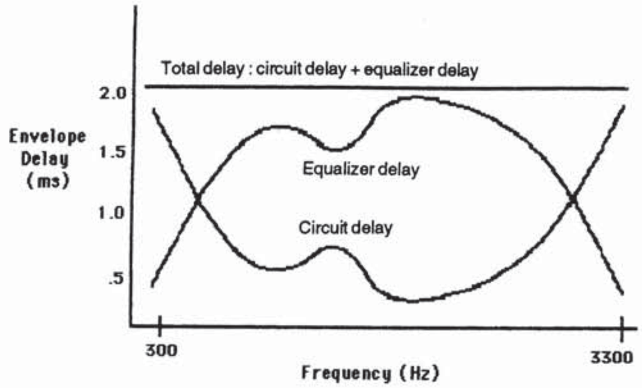

Figure 4.11 Using a delay equalizer

The modulation process

The modulation process alters the characteristics of a carrier signal. By itself, a carrier is a repeating signal that conveys no information. However, when the carrier is changed by the modulation process information is impressed upon the signal. For analog signals, the carrier is a sine wave, represented by

$$a = A \sin(2\pi ft + \phi)$$

where a is the instantaneous value of voltage at time t, A the maximum amplitude, f the frequency and ϕ the phase.

The carrier's characteristics that can be altered are thus the carrier's amplitude for amplitude modulation (AM), the carrier's frequency for frequency modulation (FM), and the carrier's phase angle for phase modulation (ϕM).

Amplitude modulation

The simplest method of employing amplitude modulation is to vary the magnitude of the signal from a zero level to represent a binary zero to a fixed peak-to-peak voltage to represent a binary one. Figure 4.12 illustrates the use of amplitude modulation to encode a digital data stream into an appropriate series of analog

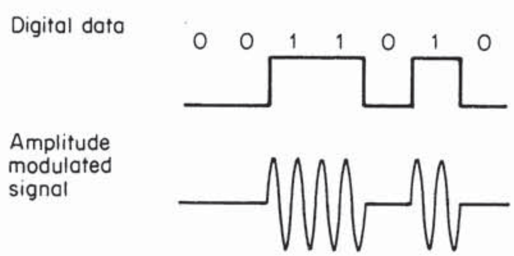

Figure 4.12 Amplitude modulation

signals. Although pure amplitude modulation is normally used for very low data rates, it is also employed in conjunction with phase modulation to obtain a method of modulating high-speed digital data sources.

Frequency modulation

Frequency modulation refers to how frequently a signal repeats itself at a given amplitude. One of the earliest uses of frequency modulation was in the design of low-speed acoustic couplers and modems where the transmitter shifted from one frequency to another as the input digital data changed from a binary one to a binary zero or from a zero to a one. This shifting in frequency is known as frequency shift keying (FSK) and is primarily used by modems operating at data rates up to 300 bps in a full-duplex mode of operation and up to 1200 bps in a half-duplex mode of operation. Figure 4.13 illustrates frequency shift keying frequency modulation.

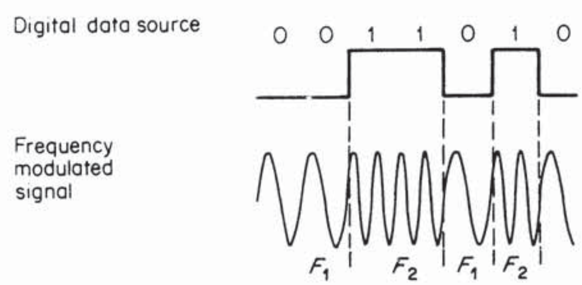

Figure 4.13 Frequency modulation

One of the earliest modems to use FSK modulation was the Bell System 103/113 type modem. Here the term Bell System refers to the operating characteristics of modems that were manufactured by Western Electric for use by AT&T operating companies prior to those operating companies becoming independent organizations. A large portion of low-speed modems in the United States were furnished by AT&T and its operating companies prior to the court-ordered divestiture of those companies. Due to this large installed base of older devices, many modern modems marketed today are designed to be compatible with earlier modems whose operating characteristics can be considered to be *de facto* standards.

The Bell System 103/113 type modem was designed to operate in one of two modes—originate or answer. This operational mode refers to the frequency assignment used by the modem for transmitting marks and spaces and should not be confused with the half- or full-duplex transmission modes that reference the modem's ability to transmit and receive data alternately or simultaneously. By splitting a two-wire circuit into two separate transmission paths by frequency, full-duplex transmission becomes possible on a half-duplex line facility.

Table 4.1 previously listed the operating frequencies of the Bell System 103/113 type modem. The originate-mode modem is normally connected to a terminal device that originates calls, whereas the answer-mode modem is normally connected to computer ports that answer calls occurring over the switched telephone network (PSTN).

In Figure 4.2 the frequency assignments of modems designed to be compatible with Bell System 103/113 type modems is graphically illustrated. Note that 1170 Hz and 2125 Hz are the channel center frequencies and two independent data channels are derived by frequency, permitting full-duplex transmission to occur over the two-wire public switched telephone network.

Phase modulation

Phase modulation is the process of varying the carrier signal with respect to the origination of its cycle as illustrated in Figure 4.14. Several forms of phase modulation are used in modems to include single and multiple-bit phase-shift keying (PSK) and the combination of amplitude and multiple-bit phase-shift keying.

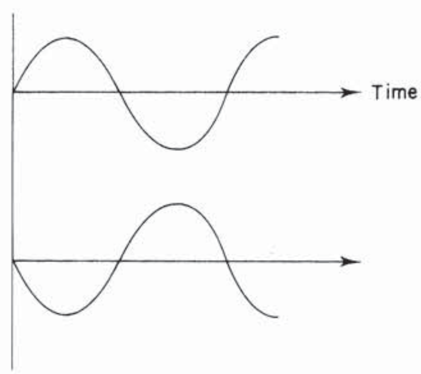

Figure 4.14 Phase modulation. Phase is the position of the wave form of a signal with respect to the origination of the carrier cycle. In this illustration, the bottom wave is 180 degrees out of phase with a normal sine wave illustrated at the top

In single-bit, phase-shift keying, the transmitter simply shifts the phase of the signal to represent each bit entering the modem. Thus, a binary one might be represented by a 90° phase change while a zero bit could be represented by a 270° phase change. Due to the variance of phase between two-phase values to represent binary ones and zeros this technique is known as two-phase modulation.

Prior to discussing multiple-bit, phase-shift keying let us examine the difference between the data rate and signaling speed. This will enable us to understand the rational for the utilization of multiple-bit, phase-shift keying, where two or more bits are grouped together and represented by one phase shift in a signal.

Bps versus baud

Bits per second is the number of binary digits transferred per second and represents the data transmission rate of a device. Baud is the signaling rate of a device such as a modem. If the signal of the modem changes with respect to each bit entering the device, then 1 bps = 1 baud. Suppose a modem is constructed such that one signal change is used to represent two bits. Then the baud rate would be one-half the bps rate.

When one baud is used to represent two bits the encoding technique is known as dibit encoding. Similarly, the process of using one baud to represent three bits is known as tribit encoding and the bit rate is then one-third of the baud rate. Both dibit and tribit encoding are known as multilevel coding techniques and are commonly implemented using phase modulation.

Voice circuit parameters

Bandwidth is a measurement of the width of a range of frequencies. As previously explained, a voice-grade telephone channel has a passband, which defines its slot in the frequency spectrum, which ranges from 300 to 3300 Hz. The bandwidth of a voice-grade telephone channel is thus 3300 − 300 or 3000 Hz.

As data enters a modem it is converted into a series of analog signals, with the signal change rate of the modem known as its baud rate. In 1928, Nyquist developed the relationship between the bandwidth and the baud rate on a circuit as

$$B = 2W$$

where B is the baud rate and W the bandwidth in Hz.

For a voice-grade circuit with a bandwidth of 3000 Hz, this relationship means that data transmission can only be supported at baud rates lower than 6000 symbols or signaling elements per second, prior to one signal interfering with another and causing intersymbol interference.

Since any oscillating modulation technique immediately halves the signaling rate, this means that most modems are limited to operating at one-half of the Nyquist limit. Thus, in a single-bit, phase-shift keying modulation technique, where each bit entering the modem results in a phase shift, the maximum data rate obtainable would be limited to approximately 3000 bps. In such a situation the bit rate would equal the baud rate, since there would be one signal change for each bit.

To overcome the Nyquist limit required engineers to design modems that first group a sequence of bits together, examined the composition of the bits, and then implemented a phase shift based upon the value of the grouped bits. This technique is known as multiple bit, phase-shift keying or multilevel, phase-shift keying. Two-bit codes called dibits and three-bit codes known as tribits are formed and transmitted by a single phase shift from a group of four or eight possible phase states.

Most modems operating at 600 to 4800 bps employ multilevel, phase-shift keying modulation. Some of the more commonly used phase patterns employed by modems using dibit and tribit encoding are listed in Table 4.2.

Table 4.2 Common phase-angle values used in multilevel, phase-level keying

Bits transmitted	Possible phase-angle values (degrees)		
dibit			
00	0	45	90
01	90	135	0
10	180	225	270
11	270	315	180
tribit			
000	0	22.5	45
001	45	67.5	0
010	90	112.5	90
011	135	157.5	135
100	180	202.5	180
101	225	247.5	225
110	270	292.5	270
111	315	337.5	315

Combined modulation techniques

Since the most practical method to overcome the Nyquist limit is obtained by placing additional bits into each signal change, modem designers have combined modulation techniques to obtain very high-speed data transmission over voice-grade circuits. One combined modulation technique commonly used involves both amplitude and phase modulation. This technique is known as quadrature amplitude modulation (QAM) and results in four bits being placed into each signal change, with the signal operating at 2400 (baud), causing the data rate to become 9600 bps.

The first implementation of QAM involved a combination of phase and amplitude modulation, in which 12 values of phase and three values of amplitude are employed to produce 16 possible signal states as illustrated in Figure 4.15. One of the earliest modems to use QAM in the United States was the Bell System 209, which modulated a 1650 Hz carrier at a 2400 baud rate to effect data transmission at 9600 bps. Today, most 9600 bps modems manufactured adhere to the ITU V.29 standard. The V.29 modem uses a carrier of 1700 Hz which is varied in both phase and amplitude, resulting in 16 combinations of eight phase angles and four amplitudes. Under the V.29 standard, fallback data rates of 7200 and 4800 bps are specified.

In addition to combining two modulation techniques, QAM also differs from the previously discussed modulation methods by its use of two carrier signals. Figure 4.16 illustrates a simplified block diagram of a modem's transmitter employing QAM. The encoder operates upon four bits from the serial data stream and causes both an in-phase (IP) cosine carrier and a sine wave that serves as the quadrature component (QC) of the signal to be modulated. The IP and QC signals are then summed and result in the transmitted signal being changed in both amplitude and

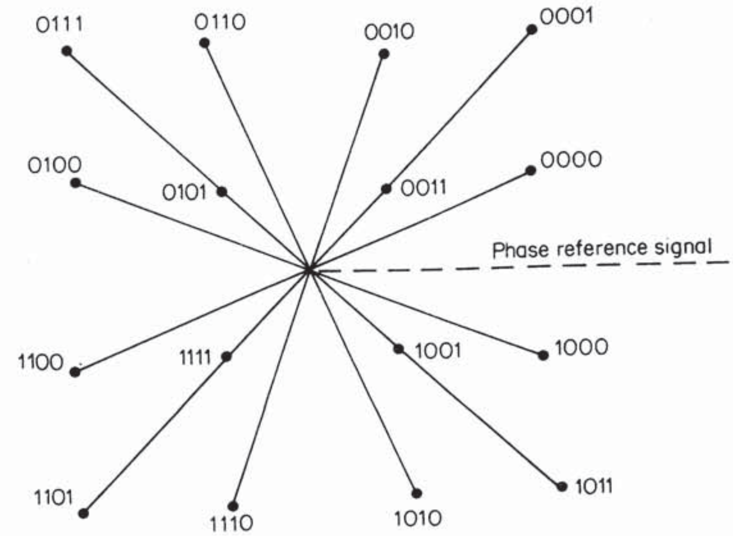

Figure 4.15 Quadrature amplitude modulation produces 16 signal states from a combination of 12 angles and three amplitude levels

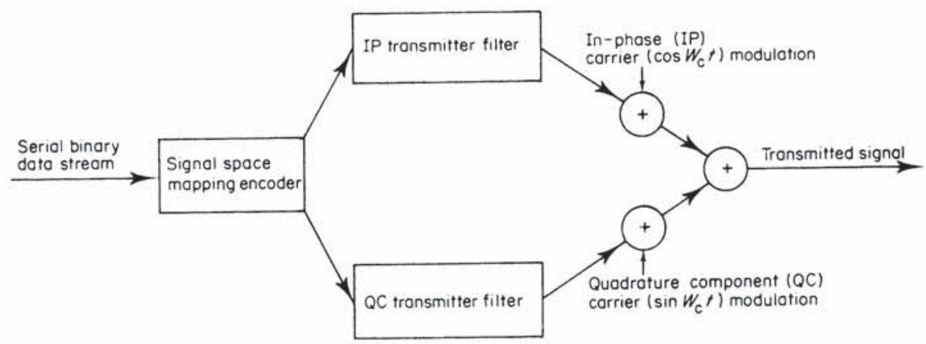

Figure 4.16 QAM modem transmitter

phase, with each point placed at the *x-y* coordinates representing the modulation levels of the cosine carrier and the sine carrier.

If you plot the signal points previously illustrated in Figure 4.15 which represent all of the data samples possible in that particular method of QAM, the series of points can be considered to be the signal structure of the modulation technique. Another popular term used to describe these points is the constellation pattern. By an examination of the constellation pattern of a modem, it becomes possible to predetermine its susceptibility to certain transmission impairments. As an example, phase jitter which causes signal points to rotate about the origin can result in one signal being misinterpreted for another, which would cause four bits to be received in error. Since there are 12 angles in the QAM method illustrated in Figure 4.15, the minimum rotation angle is 30°, which provides a reasonable immunity to phase jitter.

Other modulation techniques

By the late 1980s several vendors were offering modems that operated at data rates up to 19 200 bps over leased voice-grade circuits. Originally, modems that operated at 14 400 bps employed a quadrature amplitude modulation technique, collecting data bits into a 6-bit symbol 2400 times per second, resulting in the transmission of a signal point selected from a 64-point signal constellation. The signal pattern of one vendor's 14 400 bps modem is illustrated in Figure 4.17. Note that this particular signal pattern appears to form a hexagon and according to the vendor was used since it provides a better performance level with respect to signal-to-noise (S/N) ratio and phase jitter than conventional rectangular grid signal structures. However, in spite of hexagonal packed signal structures, it should be obvious that the distance between signal points for a 14 400 bps modem are closer than the resulting points for a 9600 bps modem. This means that a 14 400 bps conventional QAM modem is more susceptible to transmission impairments and the overall data throughput under certain situations can be less than that obtainable with 9600 bps modems. Figure 4.18 illustrates the typical throughput variance of 9600 and 14 400

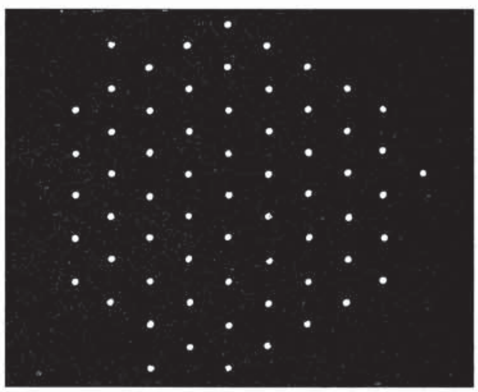

Figure 4.17 14 400 bps hexagonal signal constellation pattern

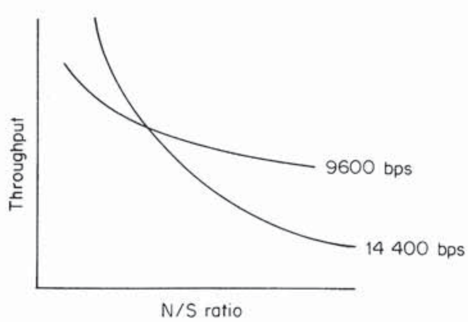

Figure 4.18 Throughput variance. Under certain conditions the throughput obtained by using 9600 bps modems can exceed the throughput obtained when using 14 400 bps devices

bps modems with respect to the ratio of noise to the strength of the signal (N/S) on the circuit. From this illustration, it should be apparent that 14 400 bps modems using conventional quadrature amplitude modulation should only be used on high-quality circuits.

Modems that transmit data at 16 000 bps are very similar to 14 400 bps devices, with the major difference being in the baud rate. Thus, most 16 000 bps modems encode data into 6-bit symbols and transmit the signals 2667 times per second. This method also employs a total of 64 signal points; however, the baud rate is increased from 2400 to 2667 to obtain the higher data transfer rate.

Trellis coded modulation

Due to the susceptibility of conventional QAM modems to transmission impairments, a new generation of modems based upon Trellis coded modulation (TCM) was developed. Such modems tolerate more than twice as much noise power as conventional QAM modems, permitting bidirectional data rates up to 33.6 kbps to be obtained over the switched telephone network.

To understand how TCM provides a higher tolerance to noise and other line impairments, to include phase jitter and distortion, let us consider what happens when a line impairment occurs when conventional QAM modems are used. Here the impairment causes the received signal point to be displaced from its appropriate location in the signal constellation. The receiver then selects the signal point in the constellation that is closest to what it received. Obviously, when line impairments are large enough to cause the received point to be closer to a signal point that is different from the one transmitted, an error occurs. To minimize the possibility of such errors, TCM employs an encoder that adds a redundant code bit to each symbol interval.

In actuality, at 14 400 bps the transmitter converts the serial data stream into 6-bit symbols and encodes two of the six bits employing a binary convolutional encoding scheme as illustrated in Figure 4.19. The encoder adds a code bit to the two input bits, forming three encoded bits in each symbol interval. As a result of this encoding operation, three encoded bits and four remaining data bits are then

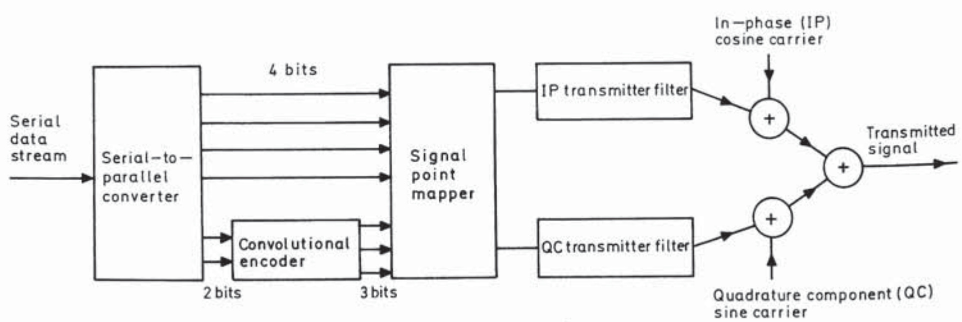

Figure 4.19 Trellis coded modulation

mapped into a signal point which is selected from a 128-point (2^7) signal constellation.

The key to the ability of TCM to minimize errors at high data rates is the employment of forward error correcting (FEC) in the form of convolutional coding. With convolutional coding, each bit in the data stream is compared with one or more bits transmitted prior to that bit. The value of each bit, which can be changed by the convolutional encoder, is therefore dependent upon the value of other bits. In addition, a redundant bit is added for every group of bits compared in this manner. The following examination of the formation of a simple convolutional code clarifies how the convolutional encoder operates.

Convolutional encoder operation

Assuming that a simple convolutional code is formed by the modulo 2 sum of the two most recent data bits, then two output bits will be produced for each data bit—a data bit and a parity bit. If we also assume that the first output bit from the encoder is the current data bit then the second output bit is the modulo 2 sum of the current bit and its immediate predecessor. Figure 4.20 illustrates the generation of this simple convolutional code.

Because each parity bit is the modulo 2 sum of the two most recent data bits, the relationship between the parity bits and the data bits becomes:

$$P_i = b_i + b_{i-1} \qquad i = 1, 2, 3 \ldots$$

If the composition of the first four data bits entering the encoder was 1101 ($b_4b_3b_2b_1$), the four parity bits are developed as follows:

$$\begin{aligned} P_1 &= b_1 + b_0 = 1 + 0 = 1 \\ P_2 &= b_2 + b_1 = 0 + 1 = 1 \\ P_3 &= b_3 + b_2 = 1 + 0 = 1 \\ P_4 &= b_4 + b_3 = 1 + 1 = 0 \end{aligned}$$

Thus, the four-bit sequence 1101 is encoded as 01111011.

The preceding example also illustrates how dependencies can be constructed. In actuality, there are several trade-offs in developing a forward error correction scheme based upon convolutional coding. When a bit is only compared with a previously transmitted bit the number of redundant bits required for decoding at the receiver is very high. The complexity of the decoding process is, however, minimized. When the bit to be transmitted is compared with a large number of previously transmitted bits the number of redundant bits required is minimized. The processing required at both ends, however, increases in complexity.

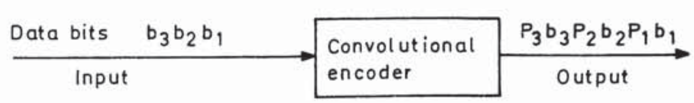

Figure 4.20 Simple convolutional code generation

In a 14 400 bps TCM modem, the signal point mapper uses the three encoded bits to select one of eight (2^3) subsets consisting of 16 points developed from the four data bits. This encoding process ensures that only certain points are valid. At the receiving modem, the decoder compares the observed sequence of signal points and selects the valid point closest to the observed sequence. The encoder makes this selection process possible by generating redundant information that establishes dependencies between successive points in the signal constellation. At the receiving modem, the decoder uses an alogorithm that compares previously received data with currently received data. The convolutional decoding algorithm then enables the modem to select the optimum signal point. Because of this technique, a TCM modem is twice as immune to noise as a conventional QAM modem. In addition, the probability of an error occurring when a TCM modem is used is substantially lower than when an uncoded QAM modem is used.

TCM modem developments

The first TCM modems marketed in 1984 operated at 14.4 kbps. Since then, a number of modems that use more complex TCM techniques have been introduced. These modems operate synchronously at 19.2 kbps, 24.4 kbps, 28.8 kbps and 33.6 kbps. By 1998, the application of TCM to modems enabled network users to more than double the transmission speed on analog circuits from that achievable only a few years before.

Echo cancellation

Prior to 1984, full-duplex transmission was achieved on the switched telephone network by splitting the available bandwidth on the line into two separate channels, each of which was used to transmit and receive data simultaneously. While this method of frequency division of the channel enabled full-duplex transmission at data rates up to 2400 bps, higher data rates using this modulation technique required more bandwidth than was available on the line. This constraint resulted in modem engineers developing a new modulation technique to achieve 9600 bps full-duplex transmission on the switched telephone network that was based upon the use of echo cancellation technology. Two other techniques used by modem designers that simulate full-duplex transmission are asymmetric transmission and 'ping-pong' transmission. Both of these transmission techniques are discussed later in this section.

Using echo cancellation, both the sending and receiving modem use the same frequency, which would normally result in the occurrence of interference between transmitted and received signals. By the use of echo cancellation technology it becomes possible for the modem's receiver to cancel out the effect of its own transmitted signal, enabling the modem to obtain the ability to distinguish the received signal.

Echo cancellation is used in the ITU V.32, V.32 bis, V.33 and V.34 modems that are described later in this section. The V.32 modem uses QAM to encode 4 bits into one signal change or baud, operating at 2400 baud to provide 9600 bps full-duplex transmission through the use of echo cancellation. In addition, trellis coding of data is an optional mode of operation for V.32 modems which, when in effect, results in the probability of a bit error occurring less than that of most modems operating at significantly lower data rates than a V.32 modem.

Types of modems and features

Modems can be categorized based upon a large number of features, their physical construction, the type of data they can modulate, and the type of telephone facility they can operate upon. To become familiar with the basic types of modems, we will examine many of the more popular features that can be used to define different types of this communications device.

Mode of transmission

If the transmitter or the receiver of the modem is such that the modem can send or receive data in one direction only, the modem will function as a simplex modem. If the operations of the transmitter and receiver are combined so that the modem may transmit and receive data alternately, the modem will function as a half-duplex modem. In the half-duplex mode of operation, the transmitter must be turned off at one location, and the transmitter of the modem at the other end of the line must be turned on before each change in transmission direction. The time interval required for this operation is referred to as turnaround time. If the transmitter and receiver operate simultaneously, the modem will function as a full-duplex modem. This simultaneous transmission in both directions can be accomplished by the use of echo cancellation, splitting the telephone line's bandwidth into two channels on a two-wire circuit, or by the utilization of two two-wire circuits, such as are obtained on a four-wire leased line.

Transmission technique

Modems are designed for asynchronous or synchronous data transmission. Asynchronous transmission is also referred to as start-stop transmission and is usually employed by unbuffered terminal devices where the time between character transmission occurs randomly.

In asynchronous transmission, the character being transmitted is initialized by the character's start bit as a mark-to-space transition on the line and terminated by the character's stop bit which is converted to a 'space/marking' signal on the line. The digital pulses between the start and stop bits are the encoded bits which determine the type of character which was transmitted. Between the stop bit of one character and the start bit of the next character, the asynchronous modem places the line in the 'marking' condition. Upon receipt of the start bit of the next character the line is switched to a mark-to-space transition, and the modem at the other end of the line starts to sample the data.

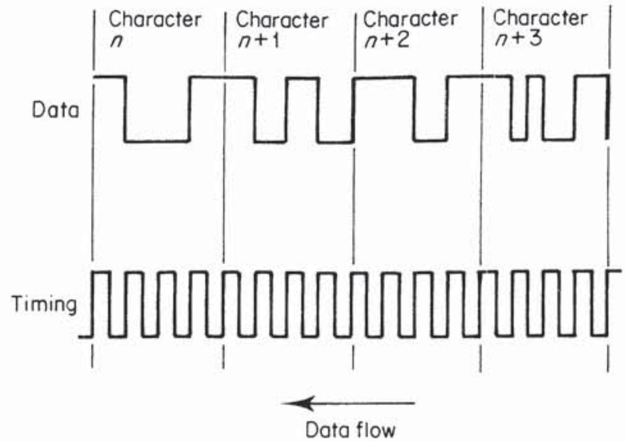

Figure 4.21 Synchronous timing signals. The timing signal is used to place the bits that form each character into a unique time period

The marking and spacing conditions are audio tones produced by the modulator of the modem to denote the binary data levels. These tones are produced at predefined frequencies, and their transition between the two states as each bit of the character is transmitted defines the character.

Synchronous transmission permits more efficient line utilization since the bits of one character are immediately followed by the bits of the next character, with no start and stop bits required to delimit individual characters. In synchronous transmission, groups of characters are formed into data blocks, with the length of the block varying from a few characters to a thousand or more. In synchronous transmission, the individual bits of each of the characters within each block are identified based upon a transmitted timing signal which is usually provided by the modem and which places each bit into a unique time period. This timing or clock signal is transmitted simultaneously with the serial bit stream as shown in Figure 4.21.

Line use classification

Modems can be classified into many categories to include the mode of transmission and transmission technique as well as by the application features they contain and the type of lines they are built to service. Generaly, modems can be classified into four line-servicing groups: subvoice or narrowband lines, voice-grade lines, wideband lines, and dedicated lines. Subvoice-band modems require only a portion of the voice-grade channel's available bandwidth and are commonly used with equipment operating at speeds up to 300 bps. On narrow-band facilities, modems can operate in the full-duplex mode by using one-half of the available bandwidth for transmission in each direction and use an asynchronous transmission technique.

Modems designed to operate on voice-grade facilities may be asynchronous or synchronous, half-duplex or full-duplex. Asynchronous transmission is normally employed at speeds up to and including 33.6 kbps. Although a leased, four-wire line will permit full-duplex transmission at high speeds without requiring the use of echo cancellation, transmission via the switched telephone network at high data rates requires the use of this technology to obtain a full-duplex operational capability.

Wideband modems, which are also referred to as group-band modems since a wideband circuit is a grouping of lower-speed lines, permit users to transmit synchronous data at speeds in excess of 48 800 bps. Although wideband modems are primarily used for computer-to-computer transmission applications, they are also used to service multiplexers which combine the transmission of many low- or medium-speed terminals to produce a composite higher transmission speed. Dedicated or limited-distance modems, which are also known by such names as shorthaul modems and modem bypass units, operate on dedicated solid conductor twisted-pair or coaxial cables, permitting data transmission at distances ranging up to 15 to 20 miles, depending upon the modem's operating speed and the resistance of the conductor.

Intelligence

Until the 1970s, modems could be categorized as dumb or non-intelligent devices. This categorization referenced their inability to perform different functions based upon a request initiated from a computer or terminal operator whose equipment was connected to the modem.

The incorporation of microprocessor technology and addition of random access memory (RAM), read only memory (ROM), and erasable programmable read only memory (EPROM) to modem circuit boards allowed manufacturers to add intelligence to their products. Through built-in routines in ROM and the ability to recognize and act upon commands and events, modems can perform such functions as automatically performing dialing operations, negotiating the method of modulation to be used to communicate with a distant modem, performing error detection and correction operations to insure data integrity, responding to status requests to report the 'state of its health', and many other functions.

Method of fabrication

Modems are manufactured in three different configurations—standalone, fabricated on adapter cards, and fabricated on rack-mount cards. The standalone modem is a self-contained device that includes one or more circuit boards contained in a common housing.

Modems fabricated on adapter cards are designed for insertion into the system unit of specific personal computers, such as an IBM PC or Apple Macintosh. Unlike an external modem that requires a separate power supply, the internal modem obtains its power from the personal computer. This eliminates the necessity of having an additional outlet for powering the modem.

Two additional differences between internal and external modems concern their cabling and desk space utilization. External modems must be cabled to a personal computer, whereas, an internal modem can use the bus in the system unit of the PC for data transfer between the modem and the computer. Concerning desk space or what many persons call footprint, since an internal modem fits into the system unit of a personal computer it requires no additional footprint or desk space.

Both internal and external modems are designed to operate as singular devices. In comparison, rack-mount modem cards are designed for insertion into a rack which is designed to provide both power and control to a common group of modems. Rack-mount modem cards are normally installed at data processing centers or at locations where communications equipment is located to support the transmission requirements of many users to a remote location, such as between dial-in lines and the ports on a multiplexer.

Reverse and secondary channels

To eliminate turnaround time when transmission is over the two-wire switched network or to relieve the primary channel of the burden of carrying acknowledgement signals on four-wire dedicated lines, modem manufacturers developed a reverse channel which is used to provide a path for the acknowledgement of transmitted data at a slower speed than the primary channel. This reverse channel can be used to provide a simultaneous transmission path for the acknowledgement of data blocks transmitted over the higher speed primary channel at up to 150 bps.

A secondary channel is similar to a reverse channel. It can, however, be used in a variety of applications which include providing a path for a high-speed terminal and a low-speed terminal. When a secondary channel is used as a reverse channel, it is held at one state until an error is detected in the high-speed data transmission and is then shifted to the other state as a signal for retransmission. Another application where a secondary channel can be utilized is when a location contains a high-speed, synchronous terminal and a slow-speed, asynchronous terminal such as a Teletype. If both devices are required to communicate with a similar distant location, one way to alleviate dual line requirements as well as the cost of extra modems to service both devices is by using a pair of modems that have secondary channel capacity, as shown in Figure 4.22. Although a reverse channel is usable on both two-wire and

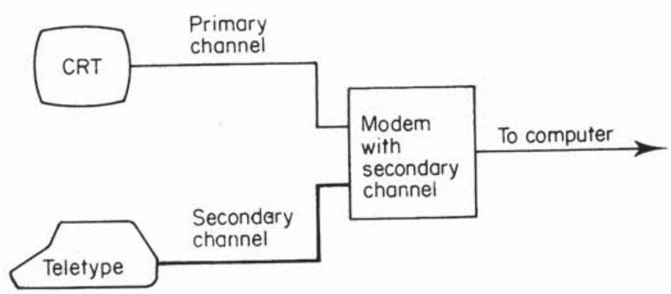

Figure 4.22 Secondary channel operation. Two terminals can communicate with a distant location by sharing a common line through the use of a modem with a secondary channel

four-wire telephone lines, the secondary channel technique is usable only on a four-wire circuit. A secondary channel modem derives two channels from the same line, a wide one to carry synchronous data and a narrow channel to carry asynchronous teletype-like data. Some modems with the secondary channel option can actually provide two slow-speed channels as well as one high-speed channel, with the two slow-speed channels being capable of transmitting asynchronous data up to a composite speed of 150 bps.

Equalization

Inconsistencies inherent in a transmission medium designed for voice rather than data transmission required modem manufacturers to build equalizers into their products to compensate for those inconsistencies produced by the telephone circuit, amplifiers, switches, and relays, as well as other equipment that data may be transmitted across in establishing a connection between two or more points. An equalizer is basically an inverse filter which is used to correct amplitude and delay distortion which, if uncorrected, could lead to intersymbol interference during transmission. A well designed equalizer matches line conditions by maintaining certain of the modem's electrical parameters at the widest range of marginal limits in order to take advantage of the data rate capability of the line while eliminating intersymbol interference. The design of the equalizer is critical, since if the modem operates too near or outside these marginal limits, the transmission error rate will increase. There are three basic methods for achieving equalization. The first method, the utilization of fixed equalizers, is typically accomplished by using marginally adjustable high-Q filter sections. Modems with transversal filters use a tapped delay line with manually adjustable variable tap gains, while automatic equalization is usually accomplished by a digital transversal filter with automatic tap gain adjustments. The faster the modem's speed, the greater the need for equalization and the more complex the equalizer. Most modems with rated speeds up to 4800 bps incorporate non-adjustable, fixed equalizers which have been designed to match the average line conditions that have been found to occur on the dial-up network. Most modems with fixed or non-adjustable equalizers are thus designed for a normal, randomly routed call between two locations over the dial-up network. If the modem is equipped with a signal-quality light which indicates an error rate that is unacceptable, or if the operator encounters difficulty with the connection, the problem can be alleviated by simply disconnecting the call and dialing a new call, which should reroute the connection through different points on the dial-up network.

Manual equalization

Manually adjusted equalization was originally employed on some 4800 bps modems used for transmission over leased lines, with the parameters being tuned or preset at installation time, and re-equalization usually not required unless the lines are reconfigured. Primarily designed to operate over unconditioned leased telephone lines, manually equalized modems allow the user to eliminate the monthly expense associated with line conditioning. Due to the incorporation of microprocessors into modems for signal processing they were soon employed to perform automatic equalization. This resulted in most modern modems incorporating automatic equalization.

Automatic equalization

Automatic equalization is used on most 4800 bps and above modems designed for operation over the switched telephone network and on all 7200 bps and above modems which are primarily designed to operate over dedicated lines but which can operate over the switched network in a fallback operational mode. With automatic equalization, a certain initialization time is required to adapt the modem to existing line conditions. This initialization time becomes important during and after line outages, since long initial equalization times can extend otherwise short dropouts unnecessarily. Recent modem developments have shortened the initial equalization time to between 15 and 25 ms, whereas only a few years ago a much longer time was commonly required. After the initial equalization, the modem continuously monitors and compensates for changing line conditions by an adaptive process. This process allows the equalizer to 'track' the frequently occurring line variations that occur during data transmission without interrupting the traffic flow. On one 9600 bps modem, this adaptive process occurs 2400 times a second, permitting the recognition of variations as they occur.

Synchronization

For synchronous communications, generally in speeds exceeding 1800 bps, the start–stop bits characteristic of asynchronous communications can be eliminated. Bit synchronization is necessary so that the receiving modem samples the link at the exact moment that a bit occurs. The receiver clock is supplied by the modem in phase coherence with the incoming data bit stream, or more simply stated, tuned to the exact speed of the transmitting clock. The transmitting clock can be supplied by either the modem (internal) or the terminal (external).

The transmission of synchronous data is generally under the control of a master clock which is the fastest clock in the system. Any slower data clock rates required are derived from the master clock by digital division logic, and those clocks are referred to as slave clocks. For instance, a master clock oscillating at a frequency of 96 kHz could be used to derive 9.6 kpbs (1/10), 4.8 kbps (1/20), and 2.4 kbps (1/40) clock speeds.

Multiport capability

Modems with a multiport capability offer a function similar to that provided by a multiplexer. In fact, multiport modems contain a limited function time division multiplexer (TDM) which provides the user with the capability of transmitting

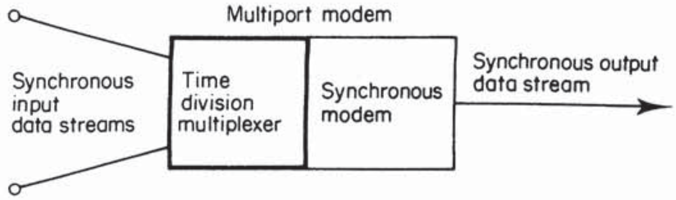

Figure 4.23 Multiport modem. Containing a limited function time division multiplexer, a multiport modem combines the input of a few synchronous input data streams for transmission

more than one synchronous data stream over a single transmission line, as shown in Figure 4.23.

In contrast with typical multiplexers, the limited function multiplexer used in a multiport modem combines only a few high-speed synchronous data streams, whereas multiplexers can normally concentrate a mixture of asynchronous and synchronous, high- and low-speed data streams. A further description, as well as application examples, will be found in Section 4.4.

Security capability

To provide an additional level of network protection for calls originated over the PSTN several vendors market 'security' modems. Some of these modems contain a buffer area into which a network administrator enters authorized passwords and associated telephone numbers. When a potential network user dials the telephone number assigned to the security modem that device prompts the person to enter his password. If a valid match between the entered and previously stored password occurs the security modem disconnects the line and then dials the telephone number associated with the password. A modem with a security capability thus provides a mechanism to verify the originator of calls over the PSTN by his telephone number. A second type of security modem verifies the location of the call originator. To do so this type of modem is first configured by associating callback telephone numbers to passwords. Once a caller establishes a connection with this type of security modem and enters his or her password, the distant modem breaks the connection and initiates a callback using the telephone number associated with the password. Based upon this type of operation, this feature is referred to as callback security. A further description of security modems is contained in Section 4.6.

Multiple speed selection capability

For data communication systems which require the full-time service of dedicated lines but need to access the switched network if the dedicated line should fail or degrade to the point where it cannot be used, dial backup capability for the modems used is necessary. Since transmission over dedicated lines usually occurs at a higher speed than one can obtain over the switched network, one method of facilitating dial backup is through switching down the speed of the modem. Thus, a multiple speed modem which is designed to operate at 9600 bps over dedicated lines may be switched down to 7200 or 4800 bps for operation over the dial-up network until the dedicated lines are restored.

Voice/data capability

Many modems can now be obtained with a voice/data option which provides the user with a voice communication capability over the same line which is used for data transmission. Depending upon the modem, this voice capability can be either alternate voice/data or simultaneous voice/data. The user may thus communicate with a distant location at the same time as data transmission is occurring, or the user may transmit data during certain times of the day and use the line for voice communications at other times. Voice/data capability can also be used to minimize normal telephone charges when data transmission sequences require voice coordination.

Modem operations and compatibility

Many modem manufacturers describe their product offerings in terms of compatibility or equivalency with modems manufactured by Western Electric for the Bell System, prior to its breakup into independent telephone companies, or with International Telecommunications Union (ITU) recommendations. The International Telecommunications Union, which is based in Geneva, is an agency of the United Nations which developed a series of modem standards for recommended use. These recommendations are primarily adapted by the Post, Telephone and Telegraph (PTT) organizations that operate the telephone networks of many countries outside the United States; however, owing to the popularity of certain ITU recommendations, they have also been followed in designing certain modems for operation on communications facilities within the USA. The following examination of the operation and compatibility of the major types of Bell System and ITU modems is based upon their operating rate.

300 bps

Modems operating at 300 bps use a frequency shift keying (FSK) modulation technique as previously described during the discussion of acoustic couplers in Section 4.1. In this technique the frequency of the carrier is alternated to one of two frequencies, one frequency representing a space or zero bit while the other frequency represents a mark or a one bit. Table 4.3 lists the frequency assignment for Bell System 103/113 and ITU V.21 modems which represent the two major types of modems that operate at 300 bps.

Table 4.3 Frequency assignments (Hz) for 300 bps modems

Major modem types	Originate	Answer
Bell System	Mark 1270	2225
(103/113 type)	Space 1070	2025
ITU V.21	Mark 980	1650
	Space 1180	1850

Bell System 103 and 113 series modems are designed so that one channel is assigned to the 1070–1270 Hz frequency band while the second channel is assigned to the 2025–2225 Hz frequency band. Modems that transmit in the 1070–1270 Hz band but receive in the 2025–2225 Hz band are designated as originate modems, while a modem which transmits in the 2025–2225 Hz band but receives in the 1070-1270 Hz band is designated as an answer modem. When using such modems, their correct pairing is important, since two originate modems cannot communicate with each other.

Bell System 113A modems are originate only devices that should normally be used when calls are to be placed in one direction. This type of modem is mainly used to enable teletype-compatible terminals to communicate with time-sharing systems where such terminals only originate calls. Bell System 113B modems are answer only and are primarily used at computer sites where users dial in to establish communications.

Modems in the 103 series, which includes the 103A, E, F, G and J modems, can transmit and receive in either the low or the high band. This ability to switch modes is denoted as 'originate and answer', in comparison to the Bell 113A which operates only in the originate mode and the Bell 113B which operates only in the answer mode.

As indicated in Table 4.3, modems operating in accordance with the ITU V.21 recommendation employ a different set of frequencies for the transmission and reception of marks and spaces. Thus, Bell System 103/113 type modems and V.21 devices can never communicate with one another.

The Bell System 103/113 type modem was used extensively in North America, and V.21 devices are primarily used in Europe. This incompatibility explains many problems that transportable and lap-top computer users experience when traveling internationally. An American traveling to Europe with a Bell 103/113 type modem built into the computer is thus precluded from communicating at 300 bps with modems in Europe. Similarly, Europeans traveling to North America with a V.21 type modem in their computers cannot communicate at 300 bps, because modems in North America normally are Bell System 103/113 compatible at that data rate.

The two pairs of frequencies used by the modems listed in Table 4.3 permit the bandwidth of a communications channel to be split into two subchannels by frequency. This technique was illustrated in Figure 4.2 for Bell System 103/113 modems. Since each subchannel can permit data to be transmitted in a direction opposite that transmitted on the other subchannel, this technique permits full-duplex transmission to occur on the switched telephone network which is a two-wire circuit that normally can only support half-duplex transmission.

Echo suppression

One of the problems associated with designing full-duplex modems for operation on the switched telephone network is the effect of echo suppressors upon modem operations. To understand how modem designers overcome the effect of echo suppressors, let us first examine why they are required. Then we can note how modem designers can disable the echo suppressors in the path established by a PSTN call as well as the effect of echo suppression upon both full- and half-duplex modems.

Although the switched telephone network is considered to be a two-wire network, in actuality, the routing of each call that goes through more than one telephone office will be via both two-wire and four-wire paths. Each local loop from a subscriber to the local telephone company serving office is a two-wire path; however, each connection between telephone offices is a four-wire path as illustrated in Figure 4.24.

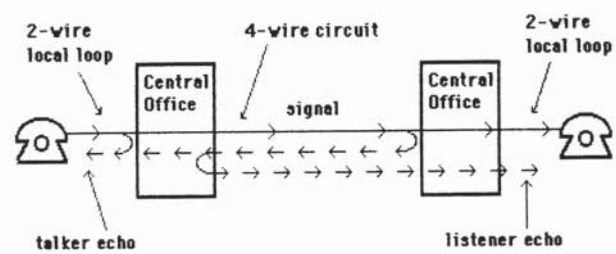

Figure 4.24 Talker and listener echoes

The actual conversion of signals from a two-wire circuit to a four-wire circuit results in an impedance mismatch. This impedance mismatch causes a portion of signal energy to be reflected back toward the originator and is referred to as 'talker', 'local', or 'near-end' echo. If this echo encounters another impedance mismatch as it flows back to the originator, another echo will be produced which now flows in the same direction as the desired signal. This doubly reflected echo, which would be heard by the listener in a telephone conversation, is called a 'listener' echo.

To minimize the effect of echoes, communications carriers added echo suppressors to their network. Echo suppressors are signal-activated devices which, unless disabled, insert a high degree of attenuation in the return echo path during the time a signal flows in the opposite path. While the use of echo suppressors provides a better quality voice circuit, their operation normally limits data transmission to one direction at a time. In addition, since they require a small amount of time to disable every time the direction of transmission is reversed, their suppression in a half-duplex environment will adversely affect data transmission throughput.

Disabling echo suppressors

One method to obtain full-duplex transmission on the PSTN is to use different frequencies for each transmission direction, similar to that described for FSK 300 bps modems. To accomplish simultaneous transmission in both directions requires the disabling of echo suppressors that lie in the switched network path established for the call. Echo suppressor disabling is accomplished by modems generating a signal in the 2010-2240 Hz band for at least 400 ms. Once echo suppressors have been disabled by an echo suppressor tone (usually 2100 Hz in North America), the signal energy of the modulated carrier in the 300-3300 Hz band in either direction of transmission is sufficient to keep the suppressors disabled as long as power is not interrupted for more than a period of 100 ms.

300 to 1800 bps

There are several Bell System and ITU V series modems that operate in the range between 300 to 1800 bps. Some of these modems such as the Bell System 212A and ITU V.22 devices can operate at either of two speeds; and other modems such as the Bell System 202 and the ITU V.23 only operate at one data rate. We will examine these modems in pairs, enabling their similarities and differences to be compared.

Bell System 212A and V.22 modems

In the late 1970s and early 1980s, the Bell System 212A and ITU V.22 modems represented the largest base of installed devices used for data transmission on the switched telephone network. Although both types of modems were rapidly replaced by more modern and higher operating rate devices, they can be expected to remain in use at many locations through the 1990s. Most higher-speed modems manufactured today are thus downward compatible with Bell System 212A and ITU V.22 devices.

The Bell System 212A modem permits either asynchronous or synchronous transmission over the public switched telephone network. The 212A contains a 103-type modem for asynchronous transmission at speeds up to 300 bps. At this data rate FSK modulation is employed, using the frequency assignments previously indicated in Table 4.3. At 1200 bps, differential phase shift keyed (DPSK) modulation is used which permits the modem to operate either asynchronously or synchronously. The phase shift encoding of the 212A type modem is illustrated in Table 4.4.

The actual phase shifts listed in Table 4.4 occur with respect to the phase angle of the previous phase shift. Thus, the name for this modulation technique is differential phase-shift keying or DPSK. In comparison, since no phase shift occurs at

Table 4.4 212A type modem phase shift encoding

Dibit	Phase shift (degrees)
00	90
01	0
10	180
11	270

300 bps the FSK modulation scheme used at that data rate is sometimes referred to as continuous phase FSK modulation.

Although a 212A type modem is designed to accept asynchronous data, transmission between two 212As at 1200 bps is actually synchronous. For this reason a 212A modem is not transparent to the binary data stream and must therefore be able to be set to support the correct character length.

All 212A type modems operating at 1200 bps support 9- and 10-bit characters, while some vendor modems also support 8- and 11-bit characters. Since the character bit length includes start and stop bits, the universal support of 10-bit characters permits the transmission of 7-bit ASCII plus parity and 8-bit extended ASCII and EBCDIC characters by all 212A modems.

When operating at 1200 bps the 212A modem uses 1200 Hz originate and 2400 Hz answer carrier frequencies. This enables the 3000 Hz bandwidth of the telephone channel to be subdivided into two channels by frequency, enabling full-duplex transmission to occur on the PSTN.

One advantage in the use of this modem is that it was the first device developed to permit the reception of transmission at two different transmission speeds. Before the operator initiates a call, he or she selects the operating speed at the originating set. The manner in which the operating speed is selected depends upon the type of 212A modem used. If the modem is what is now commonly referred to as a 'dumb' modem the operator selects the higher operating speed by pressing a 'HS' (high speed) button on the front panel of the modem. If the modem is an intelligent modem built to respond to software commands the operator can either use a communications program or send a series of commands through the serial port of a personal computer or terminal connected to the modem to set its operating speed. Due to the substantial use of intelligent modems with personal computers these modems will be reviewed as a separate entity later in this chapter. When the call is made, the answering 212A modem automatically switches to that operating speed. During data transmission, both modems remain in the same speed mode until the call is terminated, when the answering 212A can be set to the other speed by a new call. The dual-speed 212A permits both terminals connected to Bell System 100 series data sets operating at up to 300 bps or terminals connected to other 212A modems operating at 1200 bps to share the use of one modem at a computer site and thus can reduce central computer site equipment requirements.

The V.22 standard is for modems that operate at 1200 bps on the PSTN or leased circuits and has a fallback data rate of 600 bps. The modulation technique

Table 4.5 V.22 modulation phase shift as opposed to bit patterns

Dibit values 1200 bps)	Bit values (600 bps)	Phase change modes 1,2,3,4	Phase change mode 5
00	0	90	270
01	—	0	180
11	1	270	90
10	—	180	0

employed is 4-phase PSK at 1200 bps and 2-phase PSK at 600 bps, with five possible operational modes specified for the modem at 1200 bps. Table 4.5 lists the V.22 modulation phase shifts with respect to the bit patterns entering the modem's transmitter. Modes 1 and 2 are for synchronous and asynchronous data transmission at 1200 bps respectively, while mode 3 is for synchronous transmission at 600 bps. Mode 4 is for asynchronous transmission at 600 bps while mode 5 represents an alternate phase change set for 1200 bps asynchronous transmission.

In comparing V.22 modems to the Bell System 212A devices it should be apparent that they are totally incompatible at the lower data rate, since both the operating speed and modulation techniques differ. At 1200 bps the modulation techniques used by a V.22 modem in modes one through four are exactly the same as that used by a Bell System 212A device. Unfortunately, a Bell 212A modem that answers a call sends a tone of 2225 Hz on the line that the originating modem is supposed to recognize. This frequency is used because of the construction of the switched telephone network in the United States and other parts of North America. Under V.22, the answering modem first sends a tone of 2100 Hz since this frequency is more compatible with the design of European switched telephone networks. Then, the V.22 modem sends a 2400 Hz tone that would not be any better except that the V.22 modem also sends a burst of data whose primary frequency is about 2250 Hz, which is close enough to the Bell standard of 2225 Hz that many Bell 212A-type modems will respond. Thus, some Bell 212A modems can communicate with V.22 modems at 1200 bps while other 212-type modems may not be able to communicate with V.22 devices, with the ability to successfully communicate being based upon the tolerance of the 212 type modem to recognise the V.22 modem's data burst at 2250 Hz.

Bell System 202 series modems

Bell System 202 series modems were introduced during the early 1960s and provided users with high-speed transmission for their time. Today most 202 type modems have been replaced by the use of higher operating devices, however, some third-party vendors continue to manufacture 202 compatible products.

Bell System 202 series modems are designed for speeds up to 1200 or 1800 bps. One model in the 202 series, the 202C modem, can operate on either the switched network or on leased lines, in the half-duplex mode on the former and the full-duplex mode on the latter. The 202C modem can operate half-duplex or full-duplex on leased lines. This series of modems uses frequency shift keyed (FSK) modulation, and the frequency assignments are such that a mark is at 1200 Hz and a space at 2200 Hz. When either modem is used for transmission over a leased four-wire circuit in the full-duplex mode, modem control is identical to the 103 series modem in that both transmitters can be strapped on continuously which alleviates the necessity of line turnarounds.

Since the 202 series modems do not have separate bands, on switched network utilization half-duplex operation is required. This means that both transmitters (one in each modem) must be alternately turned on and off to provide two-way communication.

The time required to turn off the modem's transmitter and turn on its receiver is referred to as its turnaround delay time. While not significant on an individual basis, when transmitting and receiving small amounts of data the cumulative turnaround delay time will adversely affect performance. The use of 202 series modems on the switched network has thus been essentially replaced by 212A, V.22 bis, and other higher operating rate full-duplex modems whose use eliminates adverse performance due to turnaround delays.

The Bell 202 series modems have a 5 bps reverse channel for switched network use, which employs amplitude modulation for the transmission of information. The channel assignments used by a Bell System 202 type modem are illustrated in Figure 4.25, where the 387 Hz signal represents the optional 5 bps AM reverse channel. Owing to the slowness of this reverse channel, its use is limited to status and control function transmission. Status information such as 'ready to receive data' or 'device out of paper' can be transmitted on this channel. Also owing to the slow transmission rate, error detection of received messages and an associated NAK and request for retransmission is normally accomplished on the primary channel since even with the turnaround time, it can be completed at almost the same rate one obtains in using the reverse channel for that purpose. Non-Bell 202-equivalent modems produced by many manufacturers provide reverse channels of 75 to 150 bps which can be utilized to enhance overall system performance. Reverse keyboard-entered data as well as error detection information can be practically transmitted over such a channel.

While a data rate of up to 1800 bps can be obtained with the 202D modem, transmission at this speed requires that the leased line be conditioned for transmission by the telephone company. The 202S and 202T modems are additions to

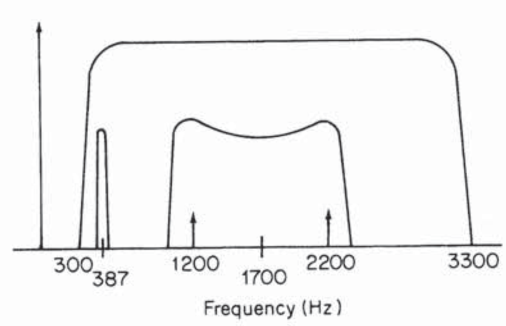

Figure 4.25 Bell System 202-type modem channel assignments

the 202 series and are designed for transmission at 1200 and 1800 bps over the switched network and leased lines, respectively. At speeds in excess of 1400 bps, the 202T requires line conditioning when interfaced to either two- or four-wire circuits, whereas for a two-wire circuit, conditioning is required at speeds in excess of 1200 bps when an optional reverse channel is used.

V.23 modems

The V.23 standard is for modems that transmit at 600 or 1200 bps over the PSTN. Both asynchronous and synchronous transmission is supported by using FSK modulation; and, an optional 75 bps backward or reverse channel can be used for error control. Figure 4.26 illustrates the channel assignments for a V.23 modem. In comparing Figure 4.26 with Figure 4.25, it is obvious that Bell System 202 and V.23 modems are incompatible with each other.

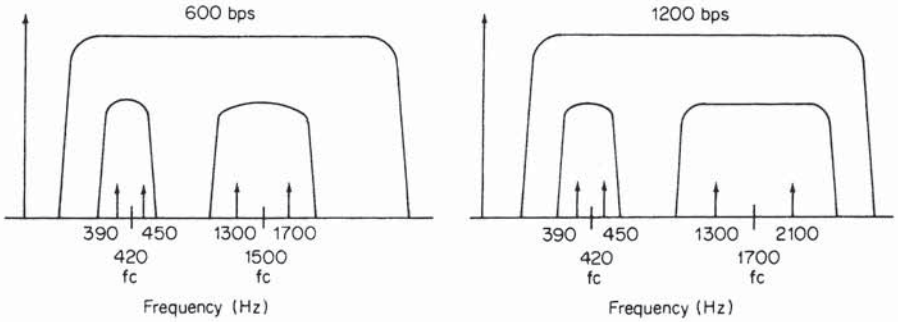

Figure 4.26 V.23 channel assignments

The V.23 reverse channel uses a frequency of 390 Hz to represent a binary 1 and a frequency of 450 Hz to represent a binary 0. In this modem's first mode of operation, which governs transmission up to 600 bps, a binary 1 is represented by a 1300 Hz tone while a binary 0 is represented by a 1700 Hz tone. In the modem's second mode of operation, which governs data transmission at 1200 bps, a binary 1 is represented by a frequency of 1300 Hz while a binary 0 is represented by a frequency of 2100 Hz.

Unlike the Bell System 202 that has been essentially replaced by more modern modems, the 75 bps reverse channel of the V.23 modem has prolonged its life as it makes the modem well suited to support many applications, including Videotext transmission requirements in Europe. When used in this manner, a V.23 modem is built into television sets and special terminal devices. The reverse channel which operates at 75 bps is used to transmit screen selection information to the computer controlling the Videotext system. Since the screen selection information may consist of a few alphanumeric characters, the 75 bps data rate is normally more than sufficient to support the transmission of data to the computer. The primary channel which operates at 1200 bps is then used as a reverse channel with respect to the requester providing the transmission path for data sent from the computer to the user's screen.

2400 bps

Examples of modems that operate at 2400 bps include the Bell System 201, the ITU V.26 series, and the V.22 bis modem. The Bell System 201 and ITU V.26 series modems are designed for synchronous bit serial transmission at a data rate of 2400 bps, while the V.22 bis standard governs 2400 bps asynchronous transmission.

Bell System 201B/C

The Bell System 201 series was originally introduced during the mid-1960s to support what was then considered to be high-speed synchronous data transmission. Although 201 modems are no longer manufactured by Western Electric, compatible modems are manufactured by that subsidiary of AT&T as well as by several third-party vendors.

Members of the 201 series include the 201B and 201C models. Both of these modems use dibit phase shift keying modulation, with the phase shifts based upon the dibit values listed in Table 4.6. The 201B modem is designed for half- or full-duplex synchronous transmission at 2400 bps over leased lines. In comparison, the 201C is designed for half-duplex, synchronous transmission over the PSTN. A more modern version of the 201C is AT&T's 2024A modem, which is compatible with the 201C.

Table 4.6 Bell System 201 B/C phase as opposed to bit pattern

Dibit values	Phase shift (degrees)
00	225
10	315
11	45
10	135

V.26 modem

The V.26 standard specifies the characteristics for a 2400 bps synchronous modem for use on a four-wire leased line. Modems operating according to the V.26 standard employ differential phase shift keying, using one of two recommended coding schemes. The phase change based upon the dibit values for each of the V.26 coding schemes is listed in Table 4.7. The use of each phase change pattern results in a modulation rate of 1200 baud providing a data signaling rate of 2400 bps. Under the V.26 standard a reverse channel can be used. This channel has the same specifications as the V.23 reverse channel.

Table 4.7 V.26 modulation phase as opposed to bit pattern

	Phase change	
Dibit values	Pattern A	Pattern B
00	0	45
01	90	135
11	180	225
10	270	315

Two similar recommendations to V.26 are V.26 bis and V.26 ter. The V.26 bis recommendation defines a dual speed 2400/1200 bps modem for use on the PSTN. At 2400 bps the modulation and coding method is the same as the V.26 recommendation for pattern B listed in Table 4.7. At the reduced data rate of 1200 bps a two-phase shift modulation scheme is employed, with a binary zero represented by a 90° phase shift while binary one is represented by a 270° phase shift. The V.26 bis recommendation also includes an optional reverse or backward channel that can be used for data transfer up to 75 bps. When employed, frequency shift keying is used to obtain this channel capacity, with a mark or one bit represented by a 390 Hz signal and a space or zero bit represented by a 450 Hz signal.

The V.26 ter recommendation uses the same phase shift scheme as the V.26 modem, but incorporates an echo-canceling technique that allows transmitted and received signals to occupy the same bandwidth. The V.26 ter modem is thus capable of operating in full duplex at 2400 bps on the PSTN. Echo canceling will be described later in this chapter when the V.32 modem is examined. Although the V.26 ter modem is popular in France, its use has been superseded by the V.22 bis modem in most countries.

V.22 bis

The ITU V.22 bis recommendation governs modems designed for 2400 bps full-duplex transmission on the PSTN and two-wire leased lines. When operating at 2400 bps, a V.22 bis modem can accept either asynchronous or synchronous data; however, transmission between modems occurs synchronously using quadrature amplitude modulation at 600 baud. Since each baud represents four bits, this results in a 2400 bps data rate.

Similar to 212A and V.22 modems, a V.22 bis modem uses carrier frequencies of 1200 and 2400 Hz for each channel obtained by frequency division. Unlike those modems that use DPSK modulation, the V.22 bis when operating at 2400 bps uses QAM. In this modulation technique the data to be transmitted is divided into groups of four consecutive bits known as quadbits. The first two bits are encoded as a phase change relative to the quadrant occupied by the preceding signal element. Table 4.8 indicates these phase quadrant changes. The last two bits of each quadbit define one of four signaling elements associated with each new quadrant. Figure 4.27 illustrates the signal constellation of a V.22 bis modem in which all possible

Table 4.8 V.22 bis dibit to phase quadrant change encoding

First two bits in quadbit (2400 bps) or dibit values (1200 bps)	Phase quadrant change (degrees)	
00	1 → 2	90
	2 → 3	
	3 → 4	
	4 → 1	
01	1 → 1	0
	2 → 2	
	3 → 3	
	4 → 4	
11	1 → 4	270
	2 → 1	
	3 → 2	
	4 → 3	
10	1 → 3	180
	2 → 4	
	3 → 1	
	4 → 2	

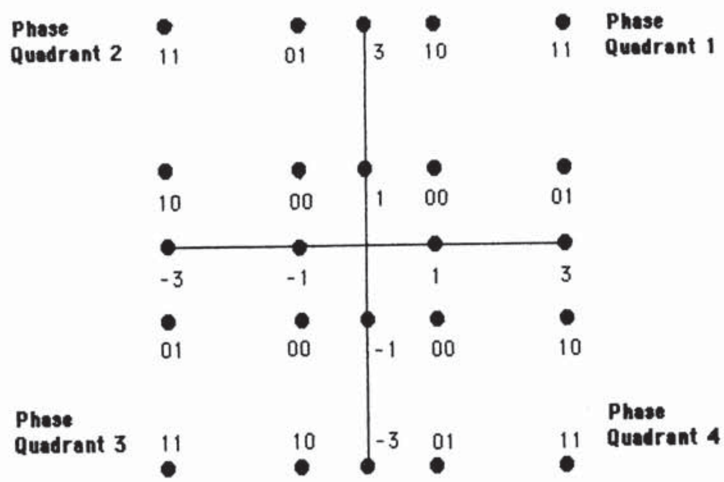

Figure 4.27 V.22 bis signal constellation

signal points are shown. Note that the dibit values in each phase quadrant represent the last two bits in a quadbit.

When operating at 1200 bps a V.22 bis modem follows the modulation scheme used by a V.22 device. That is, it uses DPSK modulation in which the value of each dibit is used to generate a phase change relative to the previous phase change as indicated in Table 4.5.

Since V.22 bis defines operations at 1200 bps to follow the V.22 format, when using a V.22 bis modem manufactured in Europe communications capability with

Bell System 212A type modems at that data rate may not always be possible due to the answer tone incompatibility usually encountered between modems following Bell System specifications and ITU recommendations. In addition, V.22 bis modems manufactured in the United States are usually not compatible with such modems manufactured in Europe at fallback data rates. This is because V.22 bis modems manufactured in Europe that are fully compliant with the recommendation only support four modes of operation to include asynchronous and synchronous transmission at 2400 and 1200 bps. This precludes the operation of those modems at 300 bps. At 1200 bps, the incompatibility between most European telephone networks (which are designed to accept only 2100 Hz answer tones) and the US telephone network (which usually accepts an answer tone between 2100 and 2225 Hz) may preclude communications between a US and a European manufactured V.22 bis modem at 1200 bps.

Since most US manufactured V.22 bis modems are fully compatible with Bell System 212A devices, this means that such modems can also operate at 300 bps using FSK modulation. At that data rate, a US manufactured V.22 bis modem would be incompatible with a European manufactured V.22 bis modem.

In spite of the previously mentioned problems, V.22 bis modems became a *de facto* standard for use with personal computers communicating over the PSTN during the early 1990s. This is due to several factors, to include the manufacture in the United States of V.22 bis modems that are Bell System 212A compatible, permitting PC users with such modems to be able to communicate with other PCs and mainframe computers connected to either 212A or 103/113 type modems. In addition, at 2400 bps, US V.22 bis modems can communicate with European V.22 bis modems, in effect, providing worldwide communications capability over the PSTN.

4800 bps

Modems that operate at 4800 bps were originally considered high-speed devices. In the late 1990s this operating rate is more representative of a low to medium speed.

The two modems examined in this section represent devices that were originally marketed during the early 1970s. Although the modulation method employed by both modems covered in this section can be considered as dated, the use of each modem can be expected to continue in a limited manner for the foreseeable future based upon reasons explained later in this section.

Bell System 208 and V.27 modems

The Bell System 208 series and ITU V.27 modems represent the most common types of modems designed for synchronous data transmission at 4800 bps. The Bell System 208 Series modems use a quadrature amplitude modulation technique. The 208A modem is designed for either half-duplex or full-duplex operation at 4800 bps over leased lines. The 208B modem is designed for half-duplex operation at 4800 bps on the switched network. The V.27 modem is designed for full- or half-duplex operations on leased lines. Later versions of the 208A were offered by

AT&T as the 2048A and 2048C models, which were designed for four-wire leased line operation. The 2048C has a start-up time less than one half of the 2048A, which makes it more suitable for operations on multidrop lines. This is because a requirement by any remote modem in a multidrop network for retraining can be satisfied quicker than by the use of a 2048A modem. Other third-party vendors introduced 208B type modems with rapid equalization times that provide enhanced throughput capability for switched network transmission. This primarily occurs during interactive transmission since the half-duplex operation of the 208B on the switched network results in numerous line turnarounds. By shortening the time required for the modems to re-equalize each time the direction of transmission changes the modem can begin to modulate data quicker after a line turnaround, in effect, boosting data throughput.

Both Bell 208 type modems and ITU V.27 modems pack data three bits at a time, encoding them for transmission as one of eight phase angles. Unfortunately, since each type of modem uses different phase angles to represent a tribit value, they cannot talk to each other. Table 4.9 lists the V.27 modulation phase shifts with respect to each of the eight possible tribit values.

Table 4.9 V.27 modulation phase drift versus bit pattern

Tribit values	Phase change (degrees)
001	0
000	45
010	90
011	135
111	180
110	225
100	270
101	315

During the 1970s to the mid-1980s, the primary use of 208 and V.27 modems was to support the transmission requirements of synchronous remote batch terminals (RBTs) that were used to communicate with mainframe computers. The use of RBTs allowed users located remote from a mainframe to enter jobs for processing as well as pull system output (Sysout) to the RBT's printer. In addition, many RBTs had tape and disk storage that allowed jobs to be batched for transmission as well as for Sysout to be queued to tape or disk for printing if the printer was being used for local processing.

From the early 1980s, personal computers began to be used as RBT replacements. Several manufacturers introduced synchronous communications adapter cards for installation in the system unit of PCs which allow them to be connected to synchronous modems, such as Bell System 208 and ITU V.27 devices. Other manufacturers offer internal 208 and V.27 type modems which permit PCs to communicate synchronously at 4800 bps over the PSTN or on a leased line by simply connecting a modular plug into the jack built into the rear of the adapter card.

In the late 1980s several third-party vendors introduced 208 type modems with enhanced capabilities. One model, which is manufactured by several vendors, is a 208A/B device that can be used on either the PSTN or on leased lines by simply changing a connector in the modem. A second 208 type modem manufactured by third-part vendors can be obtained with an optional asynchronous to synchronous converter. Through the use of this converter, which also functions as a speed converter, you can use the modem to transmit asynchronous data rates of 1200, 2400 and 4800 bps.

9600 bps

Three common modems that are representative of devices that operate at 9600 bps are the Bell System 209, and the ITU V.29 and V.32 modems.

Bell System 209 and V.29 modems

Modems equivalent to the Bell System 209 and ITU V.29 devices are designed to operate in a full-duplex, synchronous mode at 9600 bps over private lines. The Bell System 209A modem operates by employing a quadrature amplitude modulation technique as previously illustrated in Figure 4.15. Included in this modem is a built-in synchronous multiplexer which will combine up to four data rate combinations for transmission at 9600 bps. The multiplexer combinations are shown in Table 4.10. The use of a multiplexer incorporated into a modem is discussed more thoroughly in Section 4.4. A newer version of the 209A that was offered by AT&T is the 2096A. This modem is noteworthy because it has an EIA RS-449/423 interface with RS-232-C/D compatibility.

Table 4.10 Bell 209A multiplexer combinations

2400–2400–2400–2400 bps
4800–2400–2400 bps
4800–4800 bps
7200–2400 bps
9600 bps

With the exception of Bell System 209-type modems, a large majority of 9600 bps devices manufactured throughout the world adhere to the ITU V.29 standard. The V.29 standard governs data transmission at 9600 bps for full- or half-duplex operation on leased lines, with fallback data rates of 7200 and 4800 bps allowed. At 9600 bps the serial data stream is divided into groups of four consecutive bits. The first bit in the group is used to determine the amplitude to be transmitted while the remaining three bits are encoded as a phase change, with the phase changes identical to those of the V.27 recommendation listed in Table 4.9.

Table 4.11 lists the relative signal element amplitude of V.29 modems, based upon the value of the first bit in the quadbit and the absolute phase which is determined from bits two through four. Thus, a serial data stream composed of the bits 1 1 0 0 would have a phase change of 270° and its signal amplitude would be 5. The resulting signal constellation pattern of V.29 modems is illustrated in Figure 4.28.

Table 4.11 V.29 signal amplitude construction

Absolute phase (degrees)	First bit	Relative signal element amplitude
0, 90, 180, 270	0	3
	1	5
45, 135, 225, 315	0	$\sqrt{2}$
	1	$3\sqrt{2}$

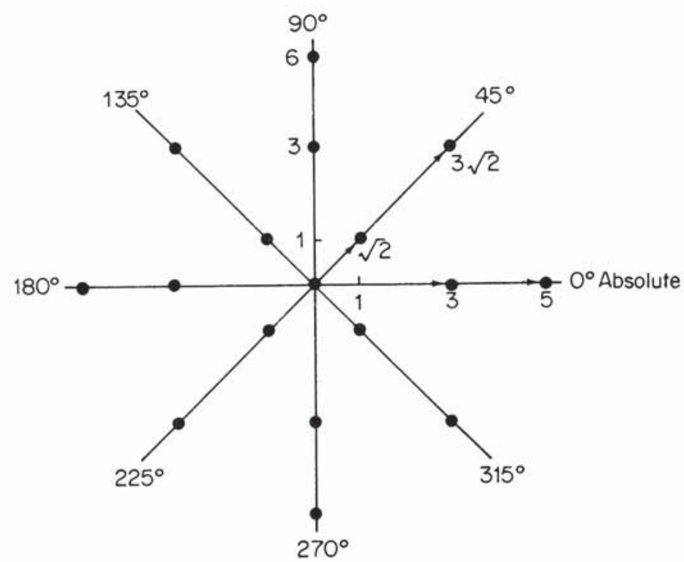

Figure 4.28 V.29 signal constellation pattern

V.29 variants

The basic chip set which provides the functionality of a V.29 modem, which is called a V.29 data pump, is used by many vendors as the basis for developing non-standard modems. Some of these non-standard modems are designed to simulate full-duplex transmission when used on the PSTN by the formation of asymmetrical channels or implementation of a 'ping-pong' scheme is described later in this section. When an asymmetrical channel scheme is used the primary channel employs QAM while the secondary channel uses FSK modulation. One channel will thus operate at 9600 bps while the other channel normally operates at 300 bps.

The primary use of asymmetrical and 'ping-pong' modulation schemes based upon the V.29 data pump is to provide an economical high-speed data transmission capability for personal computer users communicating over the PSTN. Since the vast majority of personal computers have a built-in asynchronous serial port, to support asynchronous data the modified V.29 modems include an asynchronous to synchronous converter. This allows a modem originally designed for synchronous transmission to be used to support asynchronous data transmission requirements.

The driving force behind the development of non-standard modems based upon the use of V.29 data pumps was economics. When these modems reached the market in the late 1980s they had a retail price ranging between $995 and $1295. In comparison, the only true full-duplex 9600 bps modem designed for use on the PSTN, the V.32 modem, had a retail price exceeding $3500. The retail price of V.32 modems significantly declined to the point where their cost became close to V.29 data pump based modems. Due to the significant reduction in the cost differential between V.29 data pump based modems and V.32 modems, the major rationale was removed for the selection of the former category of communications equipment. A secondary reason for the failure of V.29 data pump based modems to achieve a significant installed base resulted from their incompatibility with other vendor products when operating in their 'fast' or high-speed transmission mode. This was because the techniques used to implement asymmetrical transmission or the 'ping-pong' technique differed among vendors, in effect, forcing users to communicate at a lower speed as a V.22 bis device if they wanted to communicate with a modem manufactured by a different vendor.

Although V.29 based asymmetrical modems failed to obtain the market they were developed for, their transmission concept is incorporated into several types of modern digital subscriber line modems which are described in detail later in this chapter.

V.32 modem

V.32 is based upon a modified quadrature amplitude modulation technique and was designed to permit full-duplex 9600 bps transmission over the switched telephone network.

A V.32 modem establishes two high-speed channels in the opposite direction of one another as illustrated in Figure 4.29. Each of these channels shares approximately the same bandwidth, with an echo canceling technique employed to enable transmitted and received signals to occupy the same bandwidth. This is made possible by designing intelligence into the modem's receiver that permits it to cancel out the effects of its own transmitted signal enabling the modem to distinguish its sending signal from the signal being received.

Under the V.32 recommendation synchronous data signaling rates of 2400, 4800 and 9600 bps are supported for asynchronous data entering the modem at those rates. This support is accomplished through the use of an asynchronous to

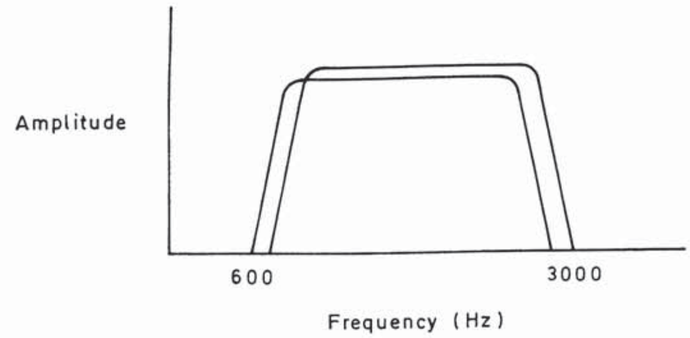

Figure 4.29 V.32 channel derivation

synchronous converter built into the modem. A V.32 modem uses a carrier frequency of 1800 Hz and has a modulation rate of 2400 bauds to support a data transfer rate of 9600 bps. At 9600 bps, the V.32 recommendation specifies two alternative modulation schemes—non-redundant coding and trellis coding.

The use of non-redundant coding results in a 16-point constellation pattern, while trellis coding results in a 32-point constellation pattern. Readers should note that while all V.32 modems must be capable of 'interworking' using the 16-point constellation pattern not all V.32 modems will include trellis coding. Users that require the better immunity to impairments thus afforded by trellis coding should ensure that the V.32 modem they are considering supports that coding alternative.

Non-redundant coding

Under the non-redundant coding technique, the data to be transmitted at 9600 bps is divided into groups of four consecutive data bits. The value of the first two bits is used in conjunction with the value of the two bits last output to generate the value for the next two output bits. These two output bits are then used with the value of bits three and four of the quadbit to select an appropriate signal point. Table 4.12 indicates the dependencies of the output dibit selection upon the value of the input dibit and the previous output dibits. Note that the value of the two input bits ($Q1_n$ and $Q2_n$) and the value of the previous dibit outputs ($Y1_{n-1}$ and $Y2_{n-1}$) are used to determine the phase quadrant change where the signal point will be located; however, this does not actually locate the point in the quadrant. To perform the latter operation the dibit outputs ($Y1_n$ and $Y2_n$) are used in conjunction with the values of the third and fourth bits in the quadbit input to select a position in the quadrant.

To select a quadrant position, the value of the dibit output ($Y1_n$ and $Y2_n$) is used in conjunction with the value of the third and fourth input (Q3 and Q4) bits. Table 4.13 indicates the selection of the non-redundant coding axis position based upon the values of the dibit output and the second dibit input.

To illustrate the use of Tables 4.12 and 4.13, assume a quadbit input has the value 0001 and the previous dibit output was 01. Since $Q1_n$ and $Q2_n$ have the value 00 while $Y1_{n-1}$ and $Y2_{n-1}$ have the value 01, from Table 4.12 the outputs ($Y1_n$ and

Table 4.12 Differential quadrant coding for 4800 bps and non-redundant coding at 9600 bps

Inputs		Previous outputs		Phase quadrant change (degrees)	Outputs		Signal state for 4800 bps
$Q1_n$	$Q2_n$	$Y1_{n-1}$	$Y2_{n-1}$		$Y1_n$	$Y2_n$	
0	0	0	0	+90	0	1	B
0	0	0	1	1	1	1	C
0	0	1	0		0	0	A
0	0	1	1		1	0	D
0	1	0	0	0	0	0	A
0	1	0	1		0	1	B
0	1	1	0		1	0	D
0	1	1	1		1	1	C
1	0	0	0	+180	1	1	C
1	0	0	1		1	0	D
1	0	1	0		0	1	B
1	0	1	1		0	0	A
1	1	0	0	+270	1	0	D
1	1	0	1		0	0	A
1	1	1	0		1	1	C
1	1	1	1		0	1	B

Table 4.13 V.32 non-redundant coding signal-state mappings for 9600 bps

Coded inputs				Non-redundant coding axis position	
Y_1	Y_2	Q3	Q4	X	Y
0	0	0	0	−1	−1
0	0	0	1	−3	−1
0	0	1	0	−1	−3
0	0	1	1	−3	−3
0	1	0	0	1	−1
0	1	0	1	1	−3
0	1	1	0	3	−1
0	1	1	1	3	−3
1	0	0	0	−1	1
1	0	0	1	−1	3
1	0	1	0	−3	1
1	0	1	1	−3	3
1	1	0	0	1	1
1	1	0	1	3	1
1	1	1	0	1	3
1	1	1	1	3	3

$Y2_n$) become 11 and the phase quadrant change will be 90°. From Table 4.13, the axis position will be 3,1 ($Y1 = Y2 = 1$, $Q3 = 0$, $Q4 = 1$), which will place the signal in the lower right position of the quadrant.

Figure 4.30 illustrates the 16-point constellation pattern generated by a V.32 modem using non-redundant coding. The four points that are circled represent valid signal points when the modem operates at 4800 bps. When operating at 4800 bps using non-redundant coding the modem operates on two bits at time, comparing the input bit values to the previous output dibit values to select one of four signal points.

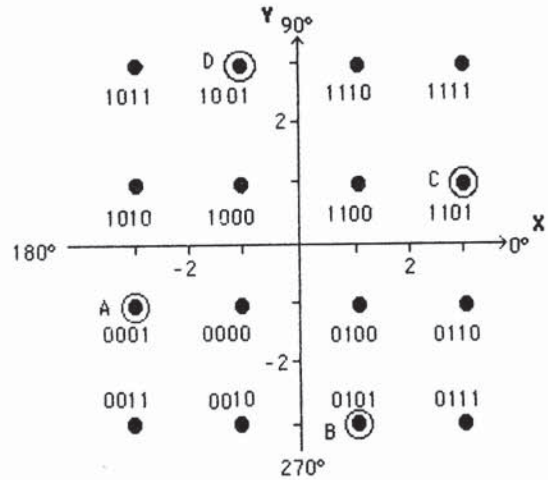

Figure 4.30 V.32 16-point signal constellation. The binary numbers denote $Y1_n$ $Y2_n$ $Q3_n$ $Q4_n$

Trellis coding

When a V.32 modem employs trellis coding, the input data stream is also divided into groups of four bits. Similar to the non-redundant coding method, the first two bits in each group ($Q1_n$ and $Q2_n$) are differentially encoded into $Y1_n$ and $Y2_n$ based upon the previous dibit value output. Table 4.14 indicates the differential encoding used by a V.32 modem operating at 9600 bps employing trellis coding.

Unlike non-redundant coding, under trellis coding the two differentially encoded bits ($Y1_n$ and $Y2_n$) are used as input to a convolutional encoder. The use of the convolutional encoder results in the generation of three bits based upon the two-bit input. Two of the three bits are the differentially encoded bits ($Y1_n$ and $Y2_n$) which are passed by the encoder. The third bit ($Y0_n$) is a redundant bit produced by the convolutional encoding process whose value is based upon the value of $Y1_n$ and $Y2_n$.

At 9600 bps the two passed through output bits ($Y1_n$ and $Y2_n$) and the redundant bit generated by the encoder ($Y0_n$) are used in conjunction with the value of the third and fourth bits in each quadbit to select one of 32 signal state mapping points.

Table 4.14 V.32 differential encoding for use with trellis-coded alternative at 9600 bps

Inputs		Previous outputs		Outputs	
$Q1_n$	$Q2_n$	$Y1_{n-1}$	$Y2_{n-1}$	$Y1_n$	$Y2_n$
0	0	0	0	0	0
0	0	0	1	0	1
0	0	1	0	1	0
0	0	1	1	1	1
0	1	0	0	0	1
0	1	0	1	0	0
0	1	1	0	1	1
0	1	1	1	1	0
1	0	0	0	1	0
1	0	0	1	1	1
1	0	1	0	0	1
1	0	1	1	0	0
1	1	0	0	1	1
1	1	0	1	1	0
1	1	1	0	0	0
1	1	1	1	0	1

Table 4.15 indicates the trellis coding signal points while Figure 4.31 illustrates the constellation pattern formed by a plot of all 32 signal points. When a V.32 modem operates at 4800 bps, as previously explained the device operates upon dibits instead of quadbits. When this occurs, the bits $Q1_n$ and $Q2_n$ are differentially encoded into $Y1_n$ and $Y2_n$ according to Table 4.12 which results in a constellation pattern of four signal points indicated in Figure 4.31 by the letters A, B, C, and D.

Since the V.32 recommendation was promulgated in 1984 numerous modem manufacturers have used the V.32 modulation scheme as a platform for adding enhancements to this modem. These enhancements include error detection and correction capability, data compression, callback security, and other features.

14 400 bps

Currently, the only standardized 14 400 bps modems are the ITU V.32 bis and V.33 recommendations.

V.32 bis modem

The V.32 bis recommendation was promulgated in 1981 and represents the ITU standard for modems operating on the switched telephone network at data rates up to 14 400 bps. Although this modem has been available for over a decade, until 1992 the average retail price of this modem exceeded $1000 and limited its widespread use. Since then, manufacturers have reduced the retail price of a V.32 bis modem to under $100, which has significantly increased its acquisition by personal computer

Table 4.15 V.32 trellis coding at 9600 bps

Coded inputs					Trellis coding	
(YO)	Y1	Y2	Q3	Q4	Re	Im
0	0	0	0	0	−4	1
	0	0	0	1	0	−3
	0	0	1	0	0	1
	0	0	1	1	4	1
	0	1	0	0	4	−1
	0	1	0	1	0	3
	0	1	1	0	0	−1
	0	1	1	1	−4	−1
	1	0	0	0	−2	3
	1	0	0	1	−2	−1
	1	0	1	0	2	3
	1	0	1	1	2	−1
	1	1	0	0	2	−3
	1	1	0	1	2	1
	1	1	1	0	−2	−3
	1	1	1	1	−2	1
1	0	0	0	0	−3	−2
	0	0	0	1	1	−2
	0	0	1	0	−3	2
	0	0	1	1	1	2
	0	1	0	0	3	2
	0	1	0	1	−1	2
	0	1	1	0	3	−2
	0	1	1	1	−1	−2
	1	0	0	0	1	4
	1	0	0	1	−3	0
	1	0	1	0	1	0
	1	0	1	1	1	−4
	1	1	0	0	−1	−4
	1	1	0	1	3	0
	1	1	1	0	−1	0
	1	1	1	1	−1	4

users in the home, business, government, and academia. Although more modern modems provide operating rates up to 33.6 kbps, for the next few years those modems can be expected to have a retail cost up to twice that of a V.32 bis modem. Thus, the popularity of V.32 bis modems can be expected to continue for the foreseeable future.

The V.32 bis recommendation is very similar to the V.32 recommendation in several key areas. Both recommendations specify the use of echo cancellation to obtain a full duplex transmission capability on the two-wire switched telephone network, and both specify the use of trellis coding, which significantly reduces the probability of transmission errors by improving the signal-to-noise ratio of

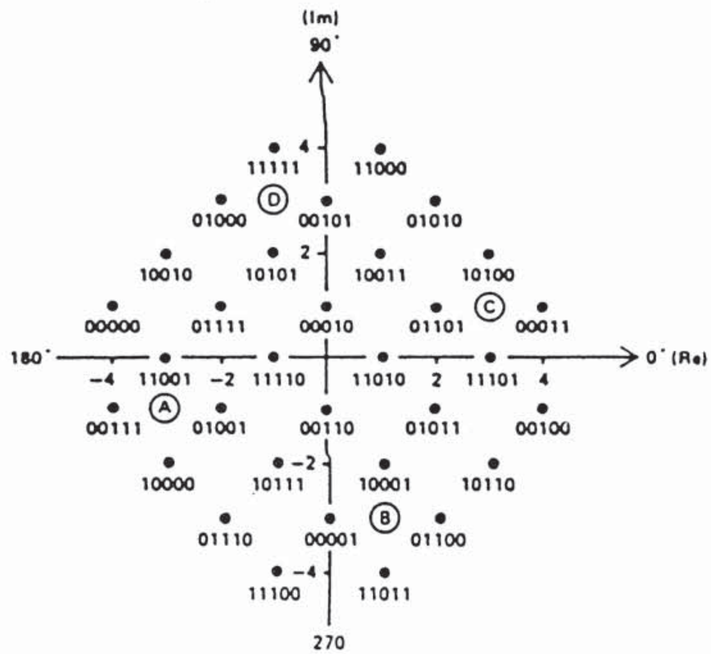

Figure 4.31 V.32 trellis coding signal constellation pattern

Table 4.16 Comparison of features of V.32 and V.32 bis

Feature	V.32	V.32 bis
Operating rates	9600 7200 (optional) 4800	14 400 12 000 9600 7200 4800
Encoded bits/symbol and signal constellation points:		
14 400 bps	—	6 (+TCM) 128
12 000 bps	—	5 (+TCM) 64
9600 bps	4 (+TCM)32	—
Fallback	yes	yes
Fall-forward	no	yes
Retrain	15 seconds	10 seconds

transmission. Major differences between the V.32 bis and V.32 recommendations are in the areas of operating rates supported, the number of encoded bits used per signal change and the resulting constellation pattern, the support of alternative operating rates during a transmission session, and the method and time required for retraining. Table 4.16 summarizes the major differences between the V.32 and V.32 bis recommendations.

As indicated in Table 4.16, the V.32 bis modem supports two operating rates above the maximum operating rate of the V.32 modem. Since those operating rates reflect the data transfer capability of each modem without the effect of data compression, the addition of that feature to each modem significantly increases the difference in throughput achievable through the use of each modem. For example, a V.32 modem that supports V.42 bis data compression, which provides an average compression ratio of 4:1, results in a throughput of 38.4 kbps when the modem operates at 9600 bps. In comparison, a V.32 bis modem operating at 14 400 bps employing V.42 bis compression provides a throughput of 57.6 kbps when the average compression ratio is 4:1. Thus, the operating rate difference of 4800 bps between a V.32 and a V.32 bis can expand to a throughput difference of almost 20 kbps when the effect of compression is considered.

Both V.32 and V.32 bis modems employ trellis coding at operating rates of 9600 bps and above. Each modem operates at 2400 baud and packs either four, five, or six data bits plus the Trellis Coded Modulation (TCM) bit into each signal change. The constellation pattern thus increases from 32 signal points for a V.32 modem operating at 9600 bps to 64 and 128 signal points for a V.32 bis modem operating at 12 000 and 14 400 bps, respectively.

At an operating rate of 14.4 kbps the V.32 bis modem divides data to be transmitted into groups of six consecutive bits. The first two bits in time ($Q1_n$ and $Q2_n$) in each group are differently coded into Y1 and Y2 in a similar manner to that previously described for the coding used by a V.32 modem. However, the outputs based on similar inputs and similar previous output are changed. Table 4.17 provides a summary of the differential encoding used by a V.32 bis modem operating at 14.4 kbps. You can note the differences between the coding of bits $Y1_n$

Table 4.17 V.32 bis differential dncoding for use with trellis coding

Inputs		Previous Outputs		Outputs	
$Q1_n$	$Q2_n$	$Y1_{n-1}$	$Y2_{n-2}$	$Y1_n$	$Y2_n$
0	0	0	0	0	0
0	0	0	1	0	1
0	0	1	0	1	0
0	0	1	1	1	1
0	1	0	0	0	1
0	1	0	1	0	0
0	1	1	0	1	1
0	1	1	1	1	0
1	0	0	0	1	0
1	0	0	1	1	1
1	0	1	0	0	1
1	0	1	1	0	0
1	1	0	0	1	1
1	1	0	1	1	0
1	1	1	0	0	0
1	1	1	1	0	1

and $Y2_n$ from the coding performed by a V.32 modem by comparing the entries in the two rightmost columns of Table 4.17 to the sixth and seventh columns in Table 4.12.

Once the two differentially encoded bits ($Y1_n$ and $Y2_n$) are computed, they are used as input to a convolutional encoder. That encoder generates a redundant bit ($Y0_n$) that is used for the six information-carrying bits ($Y1_n$, $Y2_n$, $Q3_n$, $Q4_n$, $Q5_n$, and $Q6_n$) for mapping into predefined coordinates of the transmitted signal. Figure 4.32 illustrates the signal constellation and mapping for the V.32 bis modem using trellis-coded modulation at 14.4 kbps. Note that the seven-bit binary numbers in the signal constellation reference the sequence of bits $Q6_n$, $Q5_n$, $Q4_n$, $Q3_n$, $Y2_n$, $Y1_n$, and $Y0_n$. Also note that the letters A, B, C, and D reference synchronization signal elements.

Figure 4.32 V.32 bis signal constellation and mapping for trellis-coded modulation at 14.4 kbps. A, B, C and D reference synchronizing signal elements

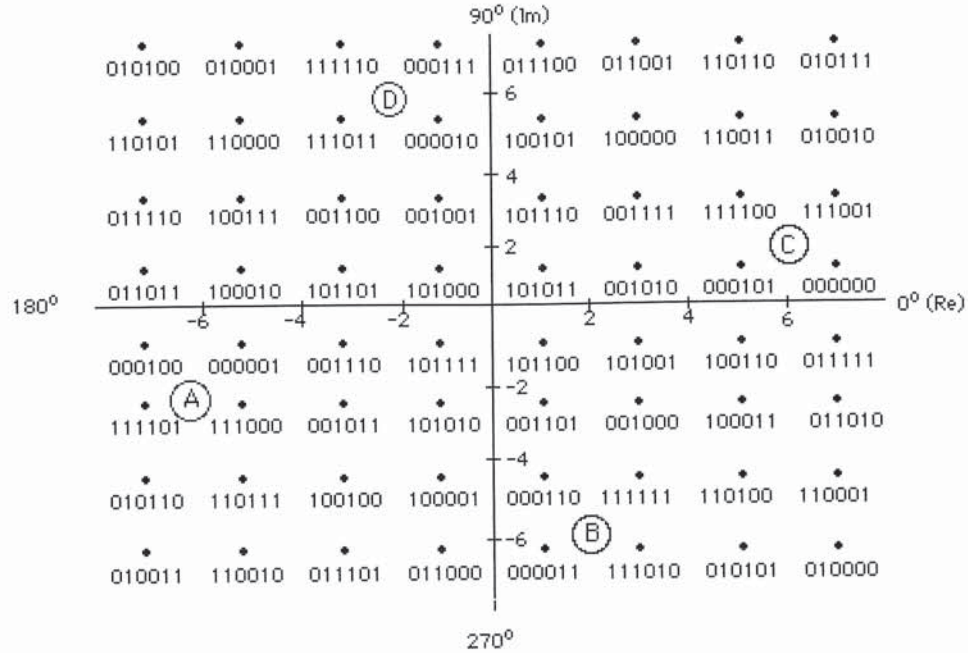

Figure 4.33 V.32 bis signal constellation and mapping for trellis-coded modulation at 12.0 kbps. A, B, C and D reference synchronizing signal elements

In addition to using trellis coding at 14.4 kbps, a V.32 bis modem also supports that coding method at 12.0 kbps, for which Figure 4.33 illustrates the constellation pattern and mapping. At that data rate the convolutional encoder produces a redundant bit that is used with five information-carrying bits ($Y1_n$, $Y2_n$, $Q3_n$, $Q4_n$ and $Q5_n$). Thus, the six binary numbers at each position in the signal constellation reference the bit sequence $Q5_n$, $Q4_n$, $Q3_n$, $Y2_n$, $Y1_n$, and $Y0_n$. The locations labeled A, B, C, and D again reference synchronizing signal elements.

The fallback feature in Table 4.16 refers to the ability of a modem to change its operating rate downward automatically when it encounters a predefined signal-to-noise ratio that would result in an unacceptable error rate if transmission at the current operating rate were maintained. Unfortunately, under the V.32 recommendation there was no provision for restoring the original operating rate if line quality improved after a fallback. Under V.32 bis an automatic fall-forward capability is included in the recommendation. This feature enables a V.32 bis modem to return to a higher operating rate if line quality improves.

Some modem vendors have implemented an enhanced fall-forward capability that can improve the transmission capability of the modem. For example, U.S. Robotics developed a feature known as Adaptive Speed Leveling for its V.32 bis modems. This feature enables the data rate of each direction of transmission to vary. Thus, if the incoming data rate is lowered due to noise encountered in one direction or another impairment, the outgoing data can still be transmitted at the highest operating rate, and vice versa. In comparison, most V.32 bis modem manufacturers implement an auto fall-forward capability symmetrically. The last major difference between V.32 and V.32 bis modems concerns retraining.

The V.32 modem has a 15-second retrain time, which minimizes the effect of a retrain upon modem throughput.

V.33 modem

The V.33 modem can be viewed as a less complex extension of V.32 technology. This is because, although the operating rate of the V.32 bis modem increased to 14.4 kbps from the 9600 bps operating rate of the V.32 modem, the V.33 achieves full-duplex transmission without the use of echo cancellation. This is possible since the V.33 modem is designed to operate on four-wire leased lines which permits the use of two two-wire signal paths.

V.32-compatible modems can operate at data rates of 14.4, 12.0, and 9.6 kbps. When operating at 14.4 kbps, the V.33 modem uses QAM modulation, assigning six data bits to each signal change and a seventh bit for trellis coding. This results in a constellation pattern of 128 signal points, which is the same as the constellation pattern of the V.32 bis modem illustrated in Figure 4.32. Although the V.33 standard is specific in its requirement for operation on four-wire leased lines, a few vendors developed proprietary half-duplex versions of the V.33 modem for use on the two-wire switched network. Those modems never achieved popularity, as most switched network users prefer using V.32 bis modems to obtain a full-duplex transmission capability at an operating rate of 14.4 kbps. In addition, the V.32 bis modem provides those modem users with the ability to communicate with a large potential audience in comparison to the small base of installed modems using a proprietary half-duplex V.33 modulation scheme.

19200 bps

When the V.32 bis modem was ratified in 1991, the ITU formed a study group that was subsequently referred to as vFast. That study group was tasked to look at higher-speed modem standards. The work of that study group was relatively long and resulted in the ratification of the V.34 standard in late 1994 and a modification to that standard in late 1995. While waiting for the standards to be promulgated, several vendors introduced interim technologies in an effort to build market penetration as well as to provide users with access to higher modem operating rates. Three of those interim technologies are referred to as vFast, V.FC, and V.32 terbo.

v.Fast represents the interim standard that preceded V.34. Although similar to V.34, there are some differences between v.Fast and V.34 modems.

The V.FC modem, with FC representing a mnemonic for Fast Class, represents a proprietary modulation scheme developed by Rockwell Semiconductor for communication at data rates up to 28.8 kbps. Early adopters of Rockwell's V.FC products were provided with hardware-based methods to upgrade to full V.34 compliance, and Rockwell now manufactures V.34-compliant chip sets.

AT&T took a different approach from Rockwell, developing an extension of V.32 bis technology in their V.32 terbo modem, which is designed to operate at data rates up to 19.2 kbps. The AT&T approach requires significantly less digital signal processing capability to implement than a V.34 modem, which requires the use of a digital signal processor capable of providing between 35 and 40 million instructions per second (mips). Similar to Rockwell, Paradyne Corporation, which until recently was part of AT&T now manufactures V.34 chip sets, although the V.32 terbo modulation capability is also included in many vendor high-speed modem products to provide transmission compatibility with the base of previously manufactured V.32 terbo modems. Since the V.32 terbo modem represents an extension of V.32 bis technology and provides a maximum operating rate of 19 200 bps, we will focus our attention on that modem in this section.

V.32 terbo

As previously discussed, V.32 terbo represents a relatively simple change to the 14.4 kbps V.32 bis standard. For this reason many people consider V.32 terbo as an enhancement of V.32 bis rather than as a competitor to more complex and higher operating rate V.34 modems.

Although the V.32 terbo modem was championed by AT&T, its specification was not formally adopted by any standards body. In spite of this, its technology was incorporated into products from several modem manufacturers, both as the highest operating rate of a modem and as one of several modulation methods supported by a V.34 modem to provide downward compatibility with V.32 terbo modems. With a retail price only $30 more than a 14.4 kbps modem, the V.32 terbo modem provides a respectable level of price performance. In addition, unlike V.34 modems which have difficulty in reaching their highest-rated speed over the PSTN and fall back to a lower operating rate, a V.32 terbo modem in many cases can be expected to provide a more consistent level of performance.

Operation

The key to the design of the V.32 terbo modem is its modification of the V.32 bis standard to reflect two new data signaling rates: 16.8 and 19.2 kbps. All other characteristics of the V.32 terbo modem are identical to the V.32 bis standard. Thus, a V.32 terbo modem uses two-dimensional trellis coding and echo cancellation.

Signal element coding

At 14 400, 16 800, and 19 200 bps, the scrambled data stream is divided into groups of six, seven, or eight consecutive data bits, respectively. The first two bits in time Q1 and Q2 are differentially encoded and trellis encoded following recommendation V.32 bis to generate three trellis-encoded bits, referred to as Y0, Y1, and Y2. These trellis bits and all remaining information bits, Q3 through Q8, are mapped into the coordinates of the signal elements to form the constellation pattern at a particular operating rate. Figures 4.34 and 4.35 illustrate the signal space diagram and mapping for V.32 terbo operations at 16.8 kbps and 19.2 kbps, respectively. At 14.4 kbps, the signal space diagram is the same as that of the V.32 bis recommendation for operation at 14.4 kbps.

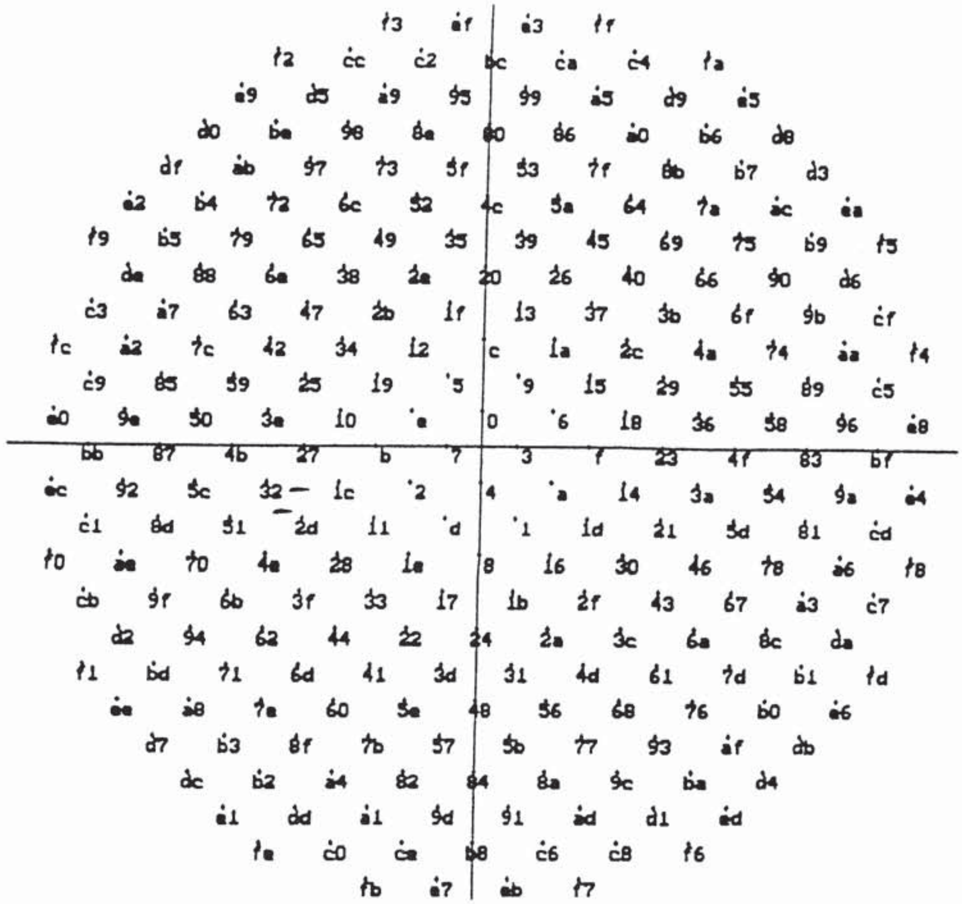

Figure 4.34 V. 32 terbo signal space diagram and mapping for 16 800 bps. Hexadecimal numbers refer to Q7Q6Q5Q4 and Q3Y2Y1Y0. (Diagram courtesy of AT&T Paradyne)

In examining the constellation patterns illustrated in Figures 4.34 and 4.35, note that each signal element is identified in hexadecimal format. Y0 is the least significant bit in the least significant hexadecimal digit. Q4 is the least significant bit in the second hexadecimal digit. At 19 200 bps, Q8 is the least significant bit in the third hexadecimal digit. Also note that the 16 800 bps signal space diagram is a subset of the 19 200 bps signal space diagram.

Non-linear encoding

At the new data rates of 16 800 and 19 200 bps, the V.32 terbo signal space diagrams are modified by non-linear encoding. Non-linear encoding is enabled only during data transmission at these rates and not during other sequence states. Non-linear encoding is disabled during startup and retraining.

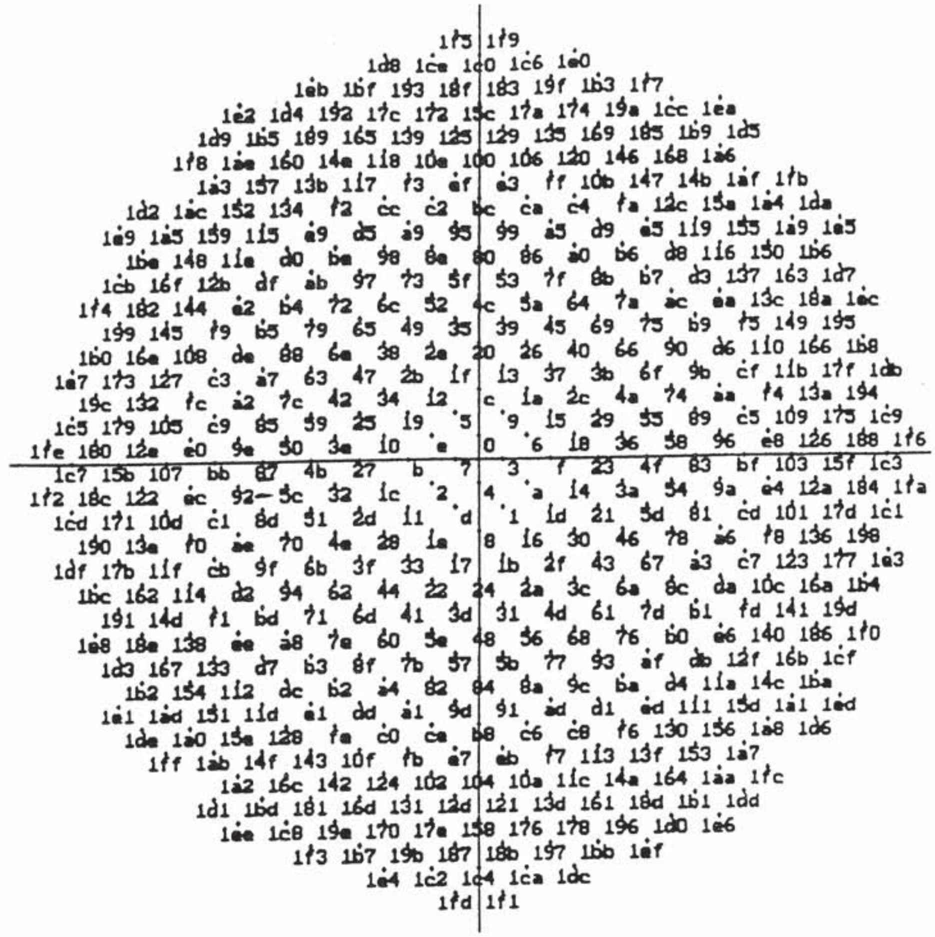

Figure 4.35 V. 32 terbo signal space diagram and mapping for 19 200 bps. Hexadecimal numbers refer to Q8, Q7Q6Q5Q4 and Q3Y2Y1Y0. (Diagram courtesy of AT&T Paradyne)

Each signal space point is non-linear–encoded by

$$X = F(x) \quad \text{and} \quad Y = F(y)$$

where the non-linear encoding projection function is

$$F = K(16\,384 + 2731P + 137P^2 + 3P^3)/16\,384$$

and $P = C(X^2 + Y^2)$. Note that $C = g/R^2$ is a constant, where the non-linear encoding gain is $G = 0.8$, and R^2 is the square of the radius of the signal space at the data signaling rate of 19 200 bps. The transmitter gain correction factor is

$$K = 1 - 1697g/16384$$

All data rates require scaling of the signal space diagrams to achieve equal transmission power. It is recommended that you perform non-linear encoding after this power adjustment. In this case, the radius of the signal space is approximately the same at 16 800 and 19 200 bps. The radius at 19 200 bps can be used to calculate C at both data rates so that identical equations are used.

Non-linear decoding

At the new data rates of 16 800 and 19 200 bps, the signal space diagrams are modified by non-linear decoding at the output of the equalizer in the receiver. Non-linear decoding is enabled only during data reception at these rates and not during other sequence states. Non-linear decoding is disabled during startup and retraining.

The reciprocal non-linear decoding operation is performed at the equalizer output:

$$X = R(x) \quad \text{and} \quad Y = R(y)$$

using the reciprocal function

$$R = 1/K + (2731P + 1229P^2 + 731P^3 + 264P^4)/16\,384$$

where $P = C(X^2 + Y^2)/K^2$ and the constants C and K are as defined in the non-linear encoding section. K is the transmitter gain correction factor. The recovered points x and y are decoded using conventional techniques.

Note that in V.32 terbo the non-linear decoder is only active during data transmission at data rates of 16 800 and 19 200 bps. The non-linear decoder is not active during training. Training is completed before the rate sequence is detected. For this reason, the transmitter gain correction factor K is incorporated in the calculation of the non-linear decoder. The equalizer would remove this constant gain factor if the non-linear decoder were active during training.

The V.32 terbo modem can provide several advantages over the higher operating rate of V.34 and V.34 bis standards that can enable it to serve a niche market for a considerable period of time. Those advantages are based on the problems many V.34 and V.34 bis users will experience when attempting to operate those modems at data rates beyond 19.2 kbps. In addition, from a practical perspective, many computers cannot support the DTE rate required to transmit data into a very-high-speed modem when the modem is in its compression mode of operation. For example, assuming a 4:1 compression ratio, you would want to set the speed of a computer port to 76.8 kbps (19.2×4) to work with a V.32 terbo modem. In comparison, you would want to set the port speed to 134.4 kbps (33.6×4) when using a V.34 bis modem. When working in Windows it is often difficult to achieve an average DTE speed greater than 57.6 kbps, which means that the use of a V.34 bis modem may not result in any noticeable improvement over a V.32 terbo modem unless the transmitted data is non-compressible. Thus, the key advantages of a V.32 terbo modem include its cost and practicality of use.

28 800 bps

There are three modems that have a maximum operating rate of 28 800 bps. Those modems include the two pre-standardization versions of the V.34 recommendation, referred to as F.FC and v.Fast, as well as the V.34 modem. To eliminate a potential degree of confusion, readers should note that the first generation of V.34 modems were limited to a maximum operating rate of 28 800 bps. During 1997 a revision to the V.34 recommendation, instead of an anticipated new recommendation, which many persons referred to as V.34 bis, raised the maximum transmission rate of this modem to 33.6 kbps. Thus, the V.34 modem will be discussed under both 28 800 bps and 33 600 bps headings.

V.FC modem

While waiting for the ITU-T study group to complete its work on finalizing the V.34 standard, Rockwell Semiconductor, a division of Rockwell International, developed a modem to provide customers with a 28.8 kbps transmission capability. This modem, which obtained the mnemonic v.FC for v.Fast Class, represents Rockwell's proprietary technology that was based on the basic features of the proposed ITU v.Fast recommendation during the March–April 1993 time frame. Although the resulting V.FC modem was manufactured as a chipset by Rockwell and incorporated into several modem manufacturer products, the resulting V.34 standard differs from the V.FC specifications enough to preclude interoperability at 28.8 kbps unless the V.34 modem includes V.FC compatibility. Otherwise, the V.34 and V.FC modems will negotiate a V.32 bis connection at 14.4 kbps.

A V.FC modem provides a maximum operating rate of 28.8 kbps as well as V.FC modulation fallback rates of 26.4 kbps and 24.0 kbps. In addition, the V.FC chipset includes full and transparent downward compatibility with V.32 bis (14.4 kbps), V.32 (9.6 kbps), and V.22 bis (2.4 kbps).

Unlike a V.34 modem that can fall back through a range of operating rates, a V.Fc modem will not fall back to V.32 bis. Other differences between a V.FC and a V.34 modem are in the areas of line probing, precoding, and trellis coding, which are discussed when the V.34 standard is described later in this section.

Although the V.FC modem filled an important niche as users waited for the V.34 standard, most 28.8 kbps modems now manufactured are based on the V.34 standard. To provide compatibility with the large base of V.FC modems, many vendors include a V.FC compatibility mode in their products, an important feature you may wish to consider.

v.Fast modem

v.Fast represents the interim standard that preceded V.34. In the haste to get modems to the marketplace as well as for semiconductor firms to develop chipsets for modem manufacturers, several versions of v.Fast were implemented. In fact, the previously described V.FC can be considered to represent v.Fast. Needless to say, different implementations of v.Fast caused some degree of compatibility problems between modems. Fortunately, most modem vendors provided a mechanism for v.Fast products to be upgraded to V.34 compliance. Thus, in the remainder of this section we will focus our attention on the operational characteristics of the V.34 standard that superseded v.Fast products.

V.34 modem

The ITU V.34 standard represents a quantum leap in both complexity and operating rate capability in comparison to the V.32 bis standard. From an operational perspective, the V.34 standard's complexity can be judged by its documentation, which is approximately twice as thick as the V.32 bis standard. The doubling of modem documentation results from the addition of several new features such as an asymmetrical transmission capability, an auxiliary channel, non-linear encoding, precoding, and a significant increase in the trellis-code state to support trellis coding. When such previously developed features as error correction, compression, and different baud rate operations are considered, the V.34 standard offers well over 100 combinations of modulation schemes, baud rates, and other operating parameters. Unfortunately, many of the new features, while standardized for use in the transmitter section of the modem, are specified as options for implementation in the modem's receiver. Thus, interoperability can be a problem as the standard allows chip manufacturers to pick and choose from an extensive array of options.

Modulation

The original V.34 standard specified three mandatory baud rates—2400, 3000, and 3200. Under ideal conditions, a V.34 modem maps nine bits into each symbol, resulting in an operating rate of 3200 baud $\times$ 9 bits/baud, or 28.8 kbps. At this operating rate there are 960 points in the modem's signal constellation. To significantly reduce the probability of errors due to impairments causing a small shift in a point in the signal constellation, the V.34 modem includes non-linear coding, precoding, and a 16-state trellis-coding capability as well as optionally specified 32- and 64-state trellis coding. Each of these features will be discussed later in this section.

In addition to the mandatory baud rates, the V.34 standard specifies optional rates of 2743, 2800, and 3429. The 2743 and 2800 baud rates can be an important consideration when transmission occurs via an infrastructure where voice is digitized using Adaptive Differential Pulse Code Modulation (ADPCM), which uses prediction to reduce the voice digitization rate from PCM's 64 kbps to 32 kbps. You typically encounter ADPCM when communicating via satellite, on terrestrial circuits that are communicating via satellite, or on terrestrial circuits that are routed overseas or to Hawaii, Alaska, and other non-contiguous US locations. Since ADPCM fails when baud rates exceed 3000, the optional 2743 and 2800 baud rates permit relatively high-speed communications to occur over an ADPCM infrastructure.

The V.34 standard specifies two carrier frequencies for both mandatory and optional baud rates. Table 4.18 summarizes the carrier frequencies, bandwidth requirements, and maximum bit rate for the six baud rates included in the V.34 standard.

Also included in Table 4.18 is a column labeled 'Modem protocol', which indicates the difference in the use of carrier frequencies between a V.34 and a V.FC modem.

Table 4.18 V.34 Bandwidth Requirements, Maximum Operating Rates, and Carrier Frequencies

Symbol rate	Modem protocol	Carrier frequency	Bandwidth requirements	Maximum bit rate
2400	V.34	1600 Hz	400–2800 Hz	21 600
	V.34/VFC	1800 Hz	600–3000 Hz	21 600
2473	V.34	1646 Hz	274–3018 Hz	24 000
	VFC V.34	1829 Hz	457–3200 Hz	24 000
2800	V.34	1680 Hz	280–3080 Hz	24 000
	VFC/V.34	1867 Hz	467–3267 Hz	24 000
3000	V.34	1800 Hz	300–3300 Hz	26 400
	V.34/VFC	2000 Hz	500–3500 Hz	26 400
	VFC	1875 Hz	375–3376 Hz	26 400
3200	V.34	1829 Hz	229–3429 Hz	28 800
	VFC	1920 Hz	320–3250 Hz	28 800
3429	V.34	1959 Hz	244–3674 Hz	28 800

In examining the entries in Table 4.18, you will note that the variety of symbol (baud) rates provides the possibility of having a non-integral number of bits per symbol. When this occurs, a shell-mapping algorithm is used to generate the constellation points.

Options

In addition to three baud rates being specified as options, the V.34 standard allows asymmetrical transmission, an auxiliary channel, non-linear encoding, precoding, and two trellis-state codes to be optionally included in a modem's receiver. To appreciate the value of each option, as well as to understand why some V.34 modems implementing all options can cost considerably more than other V.34 modems that defer implementing one or more options, requires a discussion of each option. That discussion follows.

Asymmetrical transmission

The asymmetrical transmission capability of a V.34 modem enables it to send and receive at different operating rates. This feature removes the slowest common denominator acting as a break on the data exchange.

To appreciate the value of asymmetrical transmission, assume two users wish to communicate over the PSTN. Suppose one user's office is very close to a serving central office while the second user is located on a very long local copper loop that limits maximum data transfer capability. Without an asymmetrical transmission capability, communications would occur at the lower rate in both directions.

Auxiliary channel

The V.34 standard specifies a separate transmitted signal for the exchange of information on line quality and modem operating conditions. Unless the auxiliary channel option is included in the receiver, a pair of V.34 modems will not support the use of this option.

The auxiliary channel operates at 200 bps and is multiplexed by the modem over the regular channel, being demultiplexed at the other end of the connection. Thus, this option provides an inband management channel that could be used to transmit or receive management data from a bridge or router during a communications session.

Non-linear encoding

Non-linear encoding represents a technique used to increase the immunity of a modem to interference resulting from the pulse code modulation (PCM) of analog lines by the communications carrier for transport over the carrier's backbone digital infrastructure. Under non-linear coding, constellation points are spaced unequally so that the points in areas most susceptible to noise are farther apart. This technique reduces the probability that adjacent points will be confused with each other and can reduce the error rate when operating at 28.8 kbps by up to 50%.

Precoding

Precoding represents a form of equalization that reduces the amount of high-frequency noise on a line, also reducing the amount of intersymbol interference. This in turn permits the modem to obtain a high baud rate through the use of more bandwidth than might be obtainable if no precoding occurred. One difference between a V.FC and a V.34 modem is in the area of precoding. The V.34 standard, while similar to that of the V.FC, has a change that minimizes the dithering noise of the precoder.

Trellis coding

The use of trellis coding enables a modem to maintain a high baud rate on a noisy line. Under the V.34 specification, 16-, 32-, and 64-state trellis codes are specified for the transmitter; however, only one of the three must be implemented in the receiver.

A 64-state trellis code provides a superior level of performance but is very complex to implement. While the difference between a 16-state and a 64-state trellis code is minimal when transmission occurs on a good quality line, it could prevent a modem from having to drop down to a lower operating rate when line quality deteriorates.

Another difference between a V.FC and a V.34 modem is in their implementation of trellis coding. A V.FC modem uses a two-dimensional trellis code while a V.34 modem uses a four-dimensional code. The V.FC uses the two-dimensional trellis code to produce a 32-state trellis encoder. Although that encoder is more 'powerful' than the 16-state encoder in a V.34 modem, the latter also supports 32- and 64-state encoders.

Operation

A V.34 modem will exchange a test tone with another V.34 modem to profile the channel from 150 Hz to 3750 Hz. This channel profile operation determines the effective bandwidth available and is used to make such configuration decisions as the symbol rate, carrier frequency, constellation shape, encoding scheme, transmission power, and type of trellis encoder to use. Under ideal conditions, a V.34 modem will operate at 3200 baud, mapping nine bits per symbol to obtain a 28.8 kbps operating rate. The modem can drop back in steps of 2.4 kbps to 2.4 kbps and bump up in steps of 2.4 kbps if line quality improves.

Another reason for the incompatibility between V.FC and V.34 modems is a result of their channel-probing methods. A V.FC modem probes at 68 frequencies and uses 50 Hz spacing. In comparison, a V.34 modem uses 21 frequencies with 150 Hz spacing.

Operating limitations

One of the major problems associated with the use of the V.34 modem is the fact that many users will not be able to achieve its maximum operating rate capability of 28.8 kbps. There are two reasons for this, with the first directly related to the second condition, which is associated with a phone line problem.

One reason a V.34 modem fails to achieve a 28.8 kbps transmission rate results from one or both pairs of modems not incorporating the optional features in their receivers. When telephone line quality deteriorates, the lack of previously described optional receiver features makes it difficult, if not impossible, to maintain a 28.8 kbps operating rate.

A second reason why a V.34 modem may not achieve a 28.8 kbps operating rate is due to the bandwidth required to achieve that operating rate. As indicated in Table 4.18, 28.8 kbps requires a bandwidth of 3200 Hz (320 to 3250 Hz). Due to telephone company trunk components and switches, as well as faulty premises wiring and long local loops, the ability to obtain a 3200 Hz bandwidth on an end-to-end basis can become very difficult, resulting in a V.34 modem operating at 26.4 kbps, which requires 3000 Hz bandwidth, or at 24.0 kbps, which requires 2800 Hz bandwidth.

Upgrades

One of the features you will wish to consider when purchasing a V.34 modem is its upgrade capability. This capability is important if you purchase a modem that does not implement all V.34 optional features or if you wish to upgrade to the revised recommendation that provides a maximum data transmission capability of 33.6 kbps. Some vendors also permit a V.34 modem to support one-way 56 kbps transmission which is described later in this section.

Many modem manufacturers offer one of three methods for upgrading a V.34 modem—EPROM, replacement chips, or return to factory.

The use of Erasable Programmable Read Only Memory (EPROM) enables new code to be delivered via a software download. If a modem uses instructions stored in Programmable Read Only Memory (PROM), upgrades are shipped on replacement chips for installation by the user. The third method of upgrade results from the vendor burning code into application-specific integrated circuits (ASICs). The only way to upgrade this type of modem is by shipping it back to the vendor who will swap chip components.

33 600 bps

Not long after the V.34 recommendation was adopted, Rockwell Semiconductor Systems proposed a revision to the standard to provide a maximum operating rate capability of 33 600 bps. The Rockwell proposal was given the name V.34+, with other vendors and writers using the terms V.34 bis and V.34 Plus to describe the expected revision. However, Rockwell withdrew its proposal and the ITU decided to retain the original standard terminology as opposed to the creation of a new standard. Thus, the older V.34 recommendation defines a maximum operating rate of 28.8 kbps, while the revised recommendation defines a maximum operating rate of 33.6 kbps.

The new version of the V.34 recommendation was finalized in the fall of 1996 and added data rates of 33.6 kbps and 31.2 kbps to the original standard. Under the revised V.34 recommendation, the two new operating rates are optional, meaning that you must carefully examine the operating rates supported by a V.34 modem to determine its capability.

When a V.34 modem operates at a connect speed of 33.6 kbps with compression enabled, a theoretical maximum compression ratio of 4:1 means that the serial port from a terminal device must operate at 134.4 kbps. Recognizing the fact that the UART in most personal computers have a maximum transfer rate of 115.2 kbps, some vendors use a parallel port connector on their modems. This is because data transfers via a parallel port can be supported up to approximately 230 kbps. However, since the average compression ratio actually obtained by most users will be below 4:1, you can more than likely use a serial port interfaced V.34 modem without adversely affecting your data transfer capability. In comparison, if you anticipate using a recently introduced 56 kbps modem that is described next, you will more than likely require the use of a parallel interface or an enhanced serial adapter when employing data compression.

56 kbps

The rapid growth in remote access requirements, such as telecommunicating to an organization's LAN-based server or Internet surfing, resulted in the development of a unidirectional so-called '56 kbps' modem technology which was being considered by the ITU for standardization as the V.90 recommendation when this book was prepared. Unlike previously developed standardized modems that have a similar bidirectional operational capability, the 56 kbps modem provides its high-speed operational capability downstream from a source directly connected to a communications carrier's digital network infrastructure. In addition, only one analog-to-digital conversion can occur, otherwise the maximum rate of the modem is that of a V.34. In fact, the 56 kbps modem uses V.34 technology to provide an upstream 33.6 kbps transmission capability as well as a 33.6 kbps downstream capability when it cannot achieve a 56 kbps operating rate.

Operation

A 56 kbps modem achieves its downstream capability by 'intercepting' data in its digital form generated by an access server or similar device connected by a digital line to the central office of a communications carrier. Since digital transmission lines use repeaters instead of amplifiers that build up distortion, a better signal arrives at the central office. This concept is illustrated in Figure 4.36. Note that to achieve a downstream 56 kbps transmission capability, only one analog to digital conversion is permitted. In addition, the access server or similar device at one end must be directly connected to the digital network of the serving carrier, bypassing the normal analog local loop. This means that two 56 kbps modems cannot communicate at 56 kbps when communicating as an analog modem to analog modem since this type of connection requires the use of two analog conversions.

At the time this book was prepared there were two competing 56 kbps technologies that are both proprietary and incompatible with each other. One technology, referred to as X2, was developed by U.S. Robotics which was merged into 3Com Corporation. The second technology, referred to as V.flex2, was developed by Lucent Technologies and Rockwell Semiconductor Systems. Although both technologies support V.34 which enables one modem to interoperate with another at up to 33.6 kbps, they cannot interoperate at 56 kbps. In

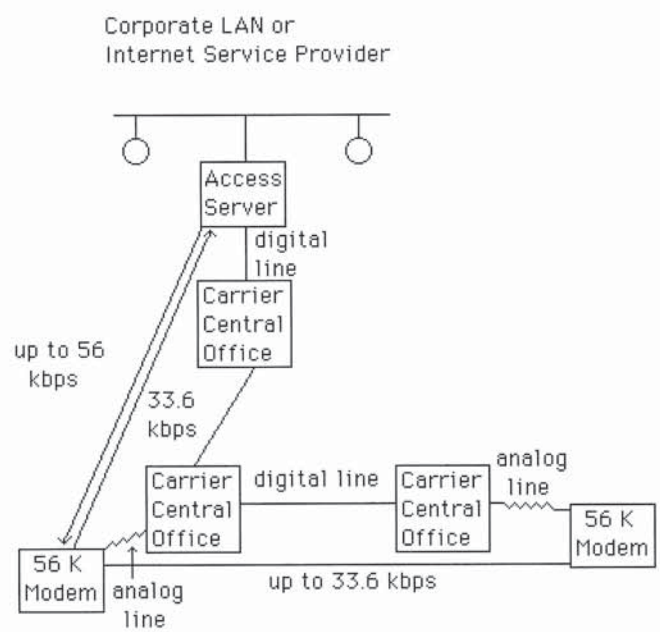

Figure 4.36 56 kbps modem communications require the elimination of one analog to digital conversion from the pair of conversions other modems operate upon. This is accomplished by locating one 56 kbps modem in a carrier's central office

addition, a Federal Communications Commission regulation precludes power levels that result in transmission above 53.3 kbps to protect telephone lines from crosstalk that can occur when the signal rate on an adjacent wire is too high. This means that unless this 20-year-old regulation is lifted, 56 kbps modems can only legally operate at a maximum rate of 53.3 kbps.

Work on the development of a 56 kbps modem standard referred to as the V.90 recommendation and activity to lift or revise the FCC regulation were in progress when this book was written. A summary of the operational characteristics of Bell System and ITU V-series type modems, as well as the previously discussed 56 kbps modem, is provided in Table 4.19.

Table 4.19 Modem operational characteristics

Modem type	Maximum data rate	Transmission technique	Modulation technique	Transmission mode	Line use
Bell System					
103A,E	300	Asynchronous	FSK	Half, full	Switched
103F	300	Asynchronous	FSK	Half, full	Leased
201B	2400	Synchronous	PSK	Half, full	Leased
201C	2400	Synchronous	PSK	Half, full	Switched
202C	1200	Asynchronous	FSK	Half	Switched
202S	1200	Asynchronous	FSK	Half	Switched
202D/R	1800	Asynchronous	FSK	Half, full	Leased
202T	1800	Asynchronous	FSK	Half, full	Leased
208A	4800	Synchronous	PSK	Half, full	Leased
208B	4800	Synchronous	PSK	Half	Switched
209A	9600	Synchronous	QAM	Full	Leased
212	0–300	Asynchronous	FSK	Half, full	Switched
	1200	Asynchronous/ synchronous	FSK	Half, full	Switched
ITU					
V.21	300	Asynchronous	FSK	Half, full	Switched
V.22	600	Asynchronous	PSK	Half, full	Switched/leased
	1200	Asynchronous/ synchronous	PSK	Half, full	Switched/leased
V.22 bis	2400	Asynchronous	QAM	Half, full	Switched
V.23	600	Asynchronous/ synchronous	FSK	Half, full	Switched
	1200	Asynchronous/ synchronous	FSK	Half, full	Switched
V.26	2400	Synchronous	PSK	Half, full	Leased
	1200	Synchronous	PSK	Half	Switched
V.26 bis	2400	Synchronous	PSK	Half	Switched
V.26 ter	2400	Synchronous	PSK	Half, full	Switched
V.27	4800	Synchronous	PSK		Leased
V.29	9600	Synchronous	QAM	Half, full	
V.32	9600	Asynchronous	TCM/QAM	Half, full	Switched/leased
V.32 bis	14 400	Asynchronous	TCM/QAM	Half, full	Switched/leased
V.33	14 400	Synchronous	TCM	Half, full	Leased
v.32 terbo	19 200	Asynchronous	TCM	Half, full	Leased
V.34	28 800	Asynchronous	TCM	Half, full	Switched/leased
V.90	56 000*	Asynchronous	TCM	Half, full	Switched/leased

* 56 kbps is unidirectional and is limited to operating with one analog to digital conversion.

Non-standard modems

There are three types of non-standard modems that deserve mention in this section even though they are obsolete when used as analog modems on conventional voice grade lines. This is because some of the concepts and a technology used by each type of modem are being incorporated into cable and digital subscriber line modems that are described later in this chapter.

Packetized ensemble protocol

One type of non-standard modem that reached the marketplace in the mid-1980s resulted in the development of a technology concept that is being applied to digital subscriber lines during the late 1990s. More formally known as a packetized ensemble protocol modem, this modem incorporated a revolutionary advance in technology due to the incorporation of a high-speed microprocessor and approximately 70 000 lines of instructions built into read-only memory (ROM) chips on the modem board.

Under the packetized ensemble protocol, the originating modem simultaneously transmits 512 tones onto the line. The receiving modem evaluates the tones and the effect of noise on the entire voice bandwidth, reporting back to the originating device the frequencies that are unusable. The originating modem then selects a transmission format most suitable to the useful tones, employing 2-bit, 4-bit or 6-bit quadrature amplitude modulation (QAM), and packetizes the data prior to its transmission. Figure 4.37 illustrates the carrier utilization of a packetized ensemble modem.

As an example of the efficiency of this type of modem, let us assume that 400 tones are available for a 6-bit QAM scheme. This would result in a packet size of 400×6 or 2400 bits. If each of the 400 tones is varied four times per second, a data rate of approximately 10 000 bps is obtained. It should be noted that the modem automatically generates a 16-bit cyclic redundancy check (CRC) for error detec-

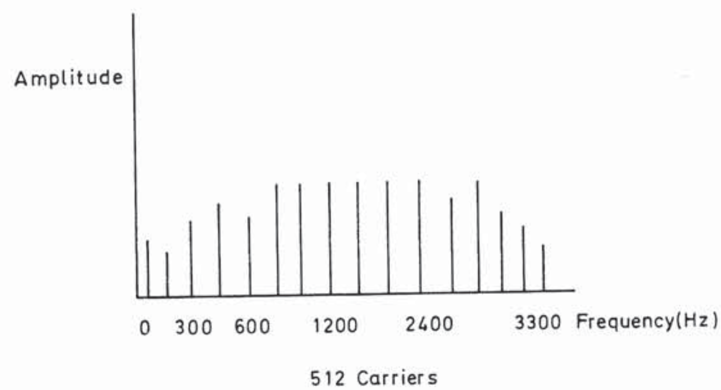

Figure 4.37 Packetized ensemble modem. (Reprinted with permission from *Data Communications Management*, © 1988 Auerbach Publishers, New York)

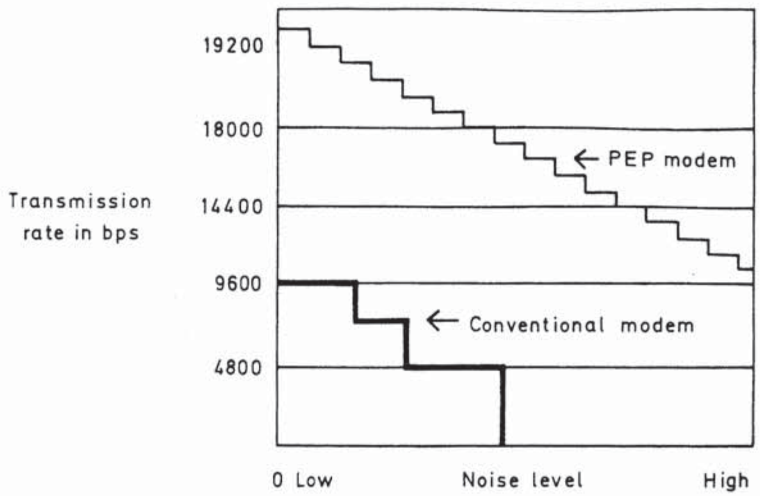

Figure 4.38 Transmission rate versus noise level. (Reprinted with permission from *Data Communications Management*, © 1988 Auerbach Publishers, New York)

tion, which is added to each transmitted packet. At the receiving modem, a similar CRC check is performed. If the transmitted and locally generated CRC characters do not match, the receiving modem will then request the transmitting modem to retransmit the packet, resulting in error correction by retransmission.

Two of the key advantages of a packetized ensemble protocol modem are its ability to automatically adjust to usable frequencies, which greatly increases the use of the line bandwidth, and its ability to lower its fallback rate in small increments. The latter is illustrated in Figure 4.38, which shows how this type of modem loses the ability to transmit on one or a few tones as the noise level on a circuit increases, resulting in a slight decrease in the data rate of the modem. In comparison, a conventional modem, such as a 9600 bps device, is designed to fall back to a predefined fraction of its main data rate, typically 7200 or 4800 bps.

The original Packetized Ensemble Modem was designed by Telebit Corporation and was marketed as the Trailblazer. Several vendors marketed similar modems under license to include a modem card for insertion into the system unit of an IBM PC or compatible personal computer and a stand-alone unit that can be attached to any computer with a standard RS-232 communications port. In addition to being compatible with other packetized ensemble protocol modems, several models of these devices are compatible with V.32, V.22 bis, V.22, 212A, and 103 type modems. This compatibility permits the personal computer user to use the device to access information utilities, other personal computers and mainframes that are connected to industry-standard modems.

Asymmetrical modems

Borrowing an old modem design concept, several vendors introduced asymmetrical modems. These modems in essence contain two channels which, in the early days of modem developments, were known as the primary and secondary channel.

Originally, modems with a secondary channel were used for remote batch transmission, where the primary high-speed channel was used to transmit data to a mainframe computer while a lower speed secondary channel was used by the mainframe to acknowledge each transmitted block. Since the acknowledgements were much shorter than the transmitted data blocks it was possible to obtain efficient full-duplex transmission even though the secondary channel might have one-tenth of the bandwidth of the primary channel.

In the late 1980s, several modem vendors realized that while high-speed transmission might be required to refresh a terminal's screen when the device was connected to a mainframe or for a file transfer, transmission in the opposite direction is typically limited by the user's typing speed or the shortness of acknowledgements in comparison to data blocks of information. Realizing this, modem vendors developed devices which use wide and narrow channels to transmit in two directions simultaneously as illustrated in Figure 4.39. The wide bandwidth channel permits a data rate of 9600 bps while the narrow bandwidth channel is used to support a data rate of 300 bps. Where these asymmetrical modems differ from older modems with secondary channels is in the incorporation of logic to monitor the output of attached devices and to then reverse the channels, permitting an attached terminal device to access the higher speed (wider bandwidth) channel when necessary. Although no standards exist for asymmetrical modems, several manufacturers attempted to formulate the use of common frequency assignments for channels. Because a 9600 bps asymmetrical modem initially had a retail price approximately two-thirds of a V.32 modem, at one time it was thought that asymmetrical technology would find a viable market. However, significant price reductions in V.32 and V.34 modems due to economies of scale have reduced the price of standardized modems below that of asymmetrical technology. Although asymmetrical modems designed for use on regular analog voice circuits failed to gain acceptance, the concept behind this technology was adapted for use with cable and digital subscriber line modems described later in this chapter.

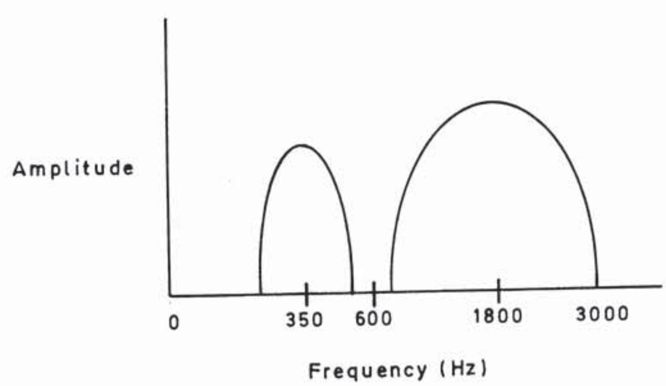

Figure 4.39 Asymmetrical modem channel assignment. (Reprinted with permission from *Data Communications Management*, © 1988 Auerbach Publishers, New York)

Ping-pong modems

Another type of modem operation that simulates full-duplex transmission is the 'ping-pong' or fast-turnaround modem. With this transmission method a modem sends data in one direction and then signals the remote modem when its transmission is completed. This signal informs the remote modem that it can now transmit data, because the originating modem has placed itself in the receive mode.

The faster the two modems are able to turn off their transmitter and turn on their receiver, the closer it appears that full-duplex transmission is occurring. Modems that employ a ping-pong transmission scheme include RAM buffers to hold data as the direction of transmission changes. This permits terminal devices connected to ping-pong operating modems to appear to continuously transmit data to the modem for modulation, although the modem is actually operating as a half-duplex device.

Modem handshaking

Modem handshaking is the exchange of control signals necessary to establish a connection between two data sets. These signals are required to set up and terminate calls, and the type of signaling used is predetermined according to one of three major standards, such as the Electronics Industries Association (EIA) RS-232 or RS-449 standard or the ITU V.24 recommendation. RS-232 and ITU V.24 standards are practically identical and are used by over 95% of all modems currently manufactured. To better understand modem handshaking, let us examine the control signals used by 103-type modems. The handshaking signals of 103-type modems and their functions are listed in Table 4.20, while the handshaking sequence is illustrated in Figure 4.40.

The handshaking routine commences when an operator at a remote terminal dials the telephone number of the computer or uses an intelligent modem to dial a predefined telephone number. At the computer site, a ring indicator (RI) signal at the answering modem is set and passed to the computer. The computer then sends a data terminal ready (DTR) signal to its modem, which then transmits a tone

Table 4.20 Modem handshaking signals and their functions

Control signal	Function
Transmit data	Serial data sent from device to modem
Receive data	Serial data received by device
Request to send	Set by device when user program wishes to transmit
Clear to send	Set by modem when transmission may commence
Data set ready	Set by modem when it is powered on and ready to transfer data; set in response to data terminal ready
Carrier detect	Set by modem when signal present
Data terminal ready	Set by device to enable modem to answer an incoming call on a switched line; reset by adaptor disconnecting call
Ring indicator	Set by modem when telephone rings

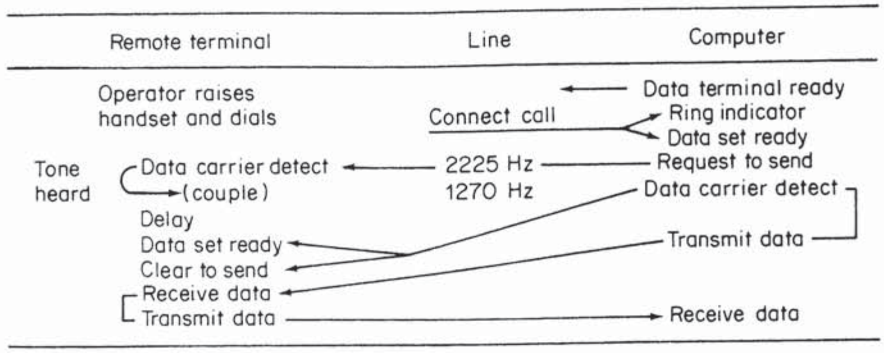

Figure 4.40 Handshake sequence

signal to the modem connected to the terminal. Upon hearing this tone when manually dialing, the terminal operator presses the data pushbutton on the modem. Upon depression of the data button for manually operated modems, the originating modem sends a data set ready (DSR) signal to the terminal, and the answering modem sends the same signal to the computer. At this point in time both modems are placed in the data mode of operation. When a call is dialed by an originating modem, the previously described process occurs automatically without operator intervention.

When accessing a remote computer, the distant device normally transmits a request for identification to the terminal. To do this the computer sets request to send (RTS) which informs its modem that it wishes to transmit data. The modem will respond with the clear to send (CTS) signal and will transmit a carrier signal. The computer's port detects the clear to send and carrier ON signals and begins its data transmission to the terminal. When the computer completes its transmission it drops the DTR signal, and the computer's modem then terminates its carrier signal. Depending upon the type of circuit on which transmission occurs, some of these signals may not be required. For example, on a switched two-wire telephone line, the RTS signal determines whether a terminal is to send or receive data, whereas on a leased four-wire circuit RTS can be permanently raised. For further information the reader should refer to specific vendor literature or appropriate technical reference publications.

Modem testing and problem resolution

Although most modems are extremely reliable devices that can be expected to provide users with years of communications capability, their complexity can result in a variety of problems. Some problems can result from the improper use or setting of one or more of the features built into modems, while other problems can be caused by equipment used with modems. In addition, modems are similar to other electronic devices in that over a period of time an increasing percentage of devices will fail. Regardless of the cause of the problem, modem users have a common goal to identify and correct communications problems.

To assist modem users in resolving problems most devices contain a series of built-in testing capabilities. In addition, all external modems have indicators that can provide users with the ability to resolve many types of communications problems.

Using modem indicators

If you examine the front panel of many external modems you will usually notice a series of mnemonic labels. These mnemonic labels are associated with light emitting diodes (LEDs) that are mounted under the modem's front panel. By understanding the meaning of the illumination or lack of illumination of the LEDs associated with a mnemonic label you can obtain an understanding of the cause of many communications problems as well as ascertain the status of a communications session you previously initiated.

For illustrative purposes we will first examine the front panel indicators of the U.S. Robotics Courier 2400 modem. This popular V.22 bis compatible modem has nine indicators as illustrated in Figure 4.41(a). Although we will primarily focus our attention upon the Courier 2400, you should note that there are no standards that define the inclusion, use, or labeling of modem status indicators. To indicate this fact, Figure 4.41(b) illustrates the front panel indicators for the Telebit T1000 modem. This modem is V.22 bis compatible and like the Courier 2400 is designed for use on the PSTN. As we examine the indicators on the Courier 2400 modem we will compare and contrast them to those on the T1000. This will show that although mnemonics may vary between modems, for most devices the meaning associated with indicators will be very similar. Although the Courier 2400 and Telebit T1000 modems can be considered as relics in an era of 28.8 and 33.6 kbps devices, their indicator panels are very similar in composition to most of the more modern modems. Thus, we can obtain an indication of the use of modem indicators for a wide range of products by reviewing the indicators on those modems.

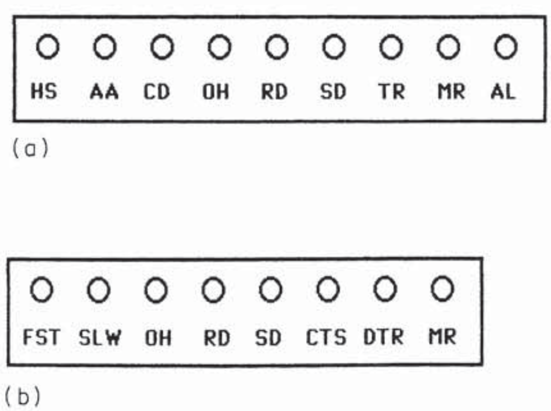

Figure 4.41 Modem indicators: (a) U.S. Robotics Courier 2400; (b) Telebit T1000

HS Indicator

The HS indicator illuminates when the Courier 2400 is communicating with another modem at 2400 bps, which is the highest speed at which that U.S. Robotics modem can operate. Although the illumination of the HS indicator indicates an operating rate of 2400 bps on a Courier 2400, its illumination on a different modem can indicate a different operating rate. For example, some 33.6 kbps V.34 modems have an HS indicator which, when lit, indicates an operating rate of 33.6 kbps. One common problem whose cause can be isolated through the use of the HS indicator is 'slow response'.

Upon occasion, a prior personal computer user may have changed communications software to operate a modem at a lower operating rate, such as at 1200 bps for a particular application. When the next person uses the computer to dial a different modem the prior operating speed selection may force the modem into 1200 bps operation. Other causes of the 'slow response' problem can be a defective answering modem that operates only at 1200 bps, or the user dialing a 1200 bps rotary instead of a 2400 bps rotary. Regardless of the cause, by examining the HS indicator you will note that the reason for slow response is not an overloaded computer system but the local or remote modem. Then, you can examine the data rate setting of your communications program and the telephone number dialed to further isolate the cause of the problem.

In examining the mnemonic status indicators for the T1000 modem, note that this device has two similar indicators to the Courier 2400's HS indicator—FST and SLW. The FST indicator illuminates when the T1000 is operating in its fast transmission mode using its proprietary Packetized Ensemble Protocol. When this indicator is illuminated, it informs you that the modem is communicating with another device that is also using the Packetized Ensemble Protocol. The SLW indicator indicates that the modem is operating in one of several slow modes—Bell System 103, 212A, or ITU V.22 or V.22 bis. Thus, you can use the FST and SLW indicators of the T1000 in a manner similar to that described for the Courier 2400 HS indicator.

AA Indicator

The Courier AA indicator is illuminated when the modem is powered on, on-line with a calling modem, and in its auto-answer mode of operation. This indicator denotes that the modem is set to receive calls instead of originate calls. Since the Courier auto-answer capability is set by placing DIP switch element 5 in an upward position, the failure of the modem to answer calls and the lack of illumination of the AA indicator would denote an improper DIP switch element setting. You can also change the modem's mode to allow it to answer calls by issuing an appropriate command. The Courier 2400, like all Hayes compatible modems, can be placed in an answer mode by setting its S0 register to a non-zero value.

CD indicator

The carrier detect (CD) indicator is illuminated when the local Courier 2400 has received a carrier signal from a distant modem that it recognizes. The iluminator of this indicator tells you that a valid carrier signal exists between the local and remote modem and that data transmission is possible.

If the CD indicator should go off during a communications session its lack of illumination informs you that data transmission is no longer possible. The cause of a lack of continuity between modems can range from noise on a line to the remote modem losing power and you should simply redial whenever you notice that the CD indicator has become extinguished.

CTS Indicator

In comparison to the Courier 2400 note that the T1000 does not have a CD indicator. Instead, you can turn on the modem's speaker to hear the presence or absence of a carrier tone or observe the CTS indicator status.

The clear to send (CTS) indicator is illuminated when the modem is ready to accept data from the attached terminal device for modulation. If no carrier signal is present the modem will not modulate data and its CTS indicator will not be illuminated.

OH Indicator

The off hook (OH) indicator is illuminated when the Courier 2400 has taken control of the telephone line. The illumination of this indicator denotes that the modem has successfully performed a function similar to a person picking up a telephone handset. The lack of illumination of the OH indicator denotes the failure of the modem to take control of the line, a problem usually resulting from someone previously disconnecting the modem from the telephone line. If this indicator does not illuminate when you issue a dialing command you should check the cable from the modem to the telephone wall jack to verify it is correctly connected.

RD Indicator

The receive data (RD) indicator of the Courier modem will flash when a data bit is received from the telephone line or when the modem is sending a result code to an attached terminal device. You can use the RD indicator to isolate several types of communications problems. As an example, assume the RD indicator is flashing but no data is being displayed on an attached personal computer. This would indicate that either the cable from the modem to the computer is defective or the personal computer's serial port has failed. Although the T1000 modem has the same RD indicator, readers should note that on some modems this indicator is labeled RX.

SD Indicator

The send data (SD) indicator functions in a reverse manner to the RD indicator, that is, the SD indicator flashes each time a bit is sent to the Courier 2400 modem from an attached terminal device.

You can use the SD indicator to verify that data is reaching the modem from an attached terminal device once a carrier detect signal is present. This can be accomplished by pressing characters on the terminal keyboard and observing the