DIGITAL TRANSMISSION BOOK

Chapter 1: INTRODUCTION TO DIGITAL TRANSMISSION

This first chapter is a broad introduction to digital transmission that introduces basic concepts and techniques, and their application in major wired and wireless networks. It leaves detailed development to the later chapters. Topics in this chapter cover backbone and access technologies and networks, the rise of data communications and Internet technologies, and the wireless networks that are the current focus of development. The following concise guide may help to tie the sections together.

<u>Section</u>

1.1 Why Digital?
Definition of digital, and explanation of its main advantages of noise immunity and spectral efficiency through digital compression.

1.2 Historical perspective on digital transmission
Morse code, analog to digital conversion, pulse code modulation and the hierarchy of digital transmission rates.

1.3 Microwave, satellite and optical transmission systems
An overview of microwave, optical and satellite transmission systems used in backbone transmission networks and in Direct Satellite Broadcasting.

1.4 Digital access technologies and networks
Voiceband modems, the Integrated Services Digital Network (ISDN), Digital Subscriber Line (DSL), the Hybrid Fiber-Coax (HFC) cable data network, Broadband ISDN (B-ISDN), and the Passive Optical Network (PON).

1.5 The rise of data communications and integrated voice-data networks
Early data networks, packet transmission using Internet Protocol (IP) and Asynchronous Transfer Mode (ATM), and Ethernet.

1.6 Wireless Transmission Systems
Transmission considerations for cellular mobile, from 1G to 5G, and the "WiFi" wireless local area network.

1.1. WHY DIGITAL?

There are still a few exceptions, notably analog AM and FM radio broadcasting which have surprised us all with their longevity, but digital transmission is dominant for all media in wired and wireless communications, terrestrial and satellite television, and in everything that comes through the Internet. Consumers today do not expect anything else and the phrase "analog communication" is probably not understood by most young people. Before embarking on the weightier technical content of this book, which is written under the assumption that we all live in the digital age, it helps our understanding and appreciation of these digital technologies to look back at the motivations, the technical basics, and the social and economic pressures to work with digital representations and transmission of information, including voice and video. We recognize that digital transmission has a very long history and in fact, for long-distance transmission, was used millennia before long-distance analog transmission. The demand for high-quality long-distance transmission was, in fact, one of the main motivators for digitizing analog voice in modern times.

But let's begin thousands of years ago, when people began using drum beats and smoke signals to convey simple messages. It's likely, although we can't know for sure, that information was not sent as continuous variations in smoke or drumbeat volume but rather as the presence or absence of signal or as a coarse quantization of volume such as "lots of smoke" versus "not much smoke" or "no smoke". This reflected, in part, the reality that humans had not yet created signal analogs of information waveforms such as speech although they could have conceived of a signal analog of a continuously-varied parameter such as, for example, the amount of food on hand. But more relevant to contemporary digital communications, this coarse quantization was de facto recognition of the noise immunity advantage of digital signaling. Continuous variations of smoke or sound (or any other detectable-from-a-distance representation) would be difficult to do, degraded by both human error and environmental influences such as wind. Quantized, easily differentiated digital signals could, in contrast, be correctly received and interpreted almost all the time despite (usually) small perturbations caused by these noise sources. This was especially important for information relayed over several hops, where the effects of noise might add and cause even more perturbation of continuous analog signals but would, with an appropriate digital quantization, have only a small probability of causing a detection error for the digital signal. Similarly for information storage, digital representation of information avoids distortions due to the degradation of the medium used for information storage. The information will very likely be as accurate the 1000th time it is read as the first. These properties are as true for electrical communication as they were for smoke signals, and with current achievements in very large scale integration (VLSI) of semiconductor memories, we have an additional advantage of very large memories in very small volumes.

More sophisticated digital transmission systems also evolved quite early. In Greece in 350 BC, Aeneas documented a sight-based relay network in which the information data were represented by different water levels in a visible container [1]. Without telescopes, the water levels would have to be quantized into just a few values readable from a distance. This system had a control mechanism in which the raising of torches signaled the beginning and end of transmission of data, a separate channel demonstrating that today's Software Defined Network (SDN) is not the first example of a system separating control and data planes.

Jumping forward to modern times and coding of analog information waveforms such as speech and video into digital data streams for transmission and/or storage, we recognize that the absence of analog noise perturbation enjoyed by digital transmission does not mean the absence of any noise perturbation. Quantization noise, the difference between a sample value of an analog waveform and its quantized representation, is an aspect of the digital coding process explained below, and the design of a transmission or storage system requires selection of parameters that result in an acceptable quantization noise level. Not surprisingly, lower quantization noise requires higher transmission rate with its attendant cost in bandwidth and other transmission and storage resources.

Another motivation for going digital is the relative ease of implementing coding for secrecy, authentication, and error detection and correction. Secrecy, meaning protection from interception and reading by a third party, always vital in espionage and war, is more important than ever in a global business environment in which hacking into corporate databases has become a common crime. Authentication, proof that a message comes from the claimed source and not a "spoofer", also has great importance to the value and safety of communications. Coding for error detection and correction covers a broad range of techniques that, at the cost of adding a little redundancy to a digital data stream, make it possible to learn that an error has occurred in transmission and, with a little more redundancy, to correct at least occasional errors.

Finally, digital transmission is appropriate and convenient for very low rate channels, initially long wires in the early 19th century and more recently radio channels linking spacecraft with Earth, characterized by low signal to noise/interference ratio (SNR) or severe distortion. An information object, such as a speech segment, photograph or video clip, can be digitally "compressed", meaning removal of redundancy that can consume transmission resources without adding much if any information, and then transmitted at any feasible speed to be reconstructed later and, for voice and video, displayed at a "real time" rate. Of course, there are delay penalties for this processing, and compromises must be made among the available tradeoffs.

1.2. HISTORICAL PERSPECTIVE ON DIGITAL TRANSMISSION

Electrical telegraphy was one of the greatest technical innovations of the 19th century. Although the poorly understood and generally uncompensated transmission characteristics of a long pair of wires forced very low transmission rates, these rates were sufficient for transmission of the most urgent news and personal communications. Samuel Morse devised, in 1838, a variable-length coding from one digital representation (letters and numbers) to another digital representation (brief presses of the telegraph key called "dots" and twice as long presses called "dashes") that achieved a remarkable data compression of the information stream. In Morse's coding scheme, frequently used letters received the shortest representations such as one dot for "e", two dots for "i" and one dash for "t" while the less frequently used (in those days) numbers are represented by various combinations of five dots and dashes.

A •▬	**B** ▬•••	**C** ▬•▬•	**D** ▬••	**0** ▬▬▬▬▬	**1** •▬▬▬▬
E •	**F** ••▬•	**G** ▬▬•	**H** ••••	**2** ••▬▬▬	**3** •••▬▬
I ••	**J** •▬▬▬	**K** ▬•▬	**L** •▬••	**4** ••••▬	**5** •••••
M ▬▬	**N** ▬•	**O** ▬▬▬	**P** •▬▬•	**6** ▬••••	**7** ▬▬•••
Q ▬•▬	**R** •▬•	**S** •••	**T** ▬	**8** ▬▬▬▬▬	**9** ▬▬▬▬•
U ••▬	**V** •••▬	**W** •▬▬	**X** ▬••▬		
Y ▬•▬▬	**Z** ▬•••				

Fig. 1.1. Morse code for basic letters and numbers (there are additional codes for other characters and punctuation).

Morse code does have the drawback of not being prefix free; that is, the representations for some letters could be misinterpreted as the beginnings of other letters, such as "e" (one dot) being a prefix of "a" (two dots) or "s" (3 dots). Telegraphers had to leave a considerable space between letters to avoid ambiguous reception. Huffman coding [2], the best variable-length coding which systematically assigns the number of bits as the log of the inverse of the probability of occurrence, is prefix free, as are all fixed-length codes including the popular ASCII. Fixed length codes do not generate any compression of alphabetic letters.

Alexander Graham Bell, remembered today principally for the telephone, was in fact funded to develop a multitone digital transmission system. The objective was to multiply the capacity of telegraph (i.e., digital) transmission lines by simultaneously communicating using tones of different frequencies. Although Bell retreated to his family's farm in Canada to escape from this work and focus on the telephone, he developed and patented a workable two-tone apparatus, shown in Figure 1.2. This accomplishment, in heated competition with other inventors just as with the telephone, received U.S. patent 174,465 and was an early example of digital FDM (frequency-division multiplexing). Bell's telephone company later was desperate to develop analog FDM for analog voice traffic, which was not really effective until application of vacuum tubes to provide amplification at relay points [3].

Fig. 1.2. Receiver of Alexander Graham Bell's "harmonic telegraph" which used reeds responsive to different frequencies. Photo of replica of original equipment used with permission of Telecommunications Museum, LaSalle, Quebec, Canada.

1.2.1 The Sampling Theorem: Transitioning Media from Analog to Digital

Digital transmission of voice traffic had to wait many more decades, requiring first the discovery of the "sampling theorem". This was a mathematical demonstration, often credited to Harry Nyquist, that an analog (continuous) waveform can be exactly reconstructed from just a set of *samples* taken at a rate equal to (at least) twice the highest frequency component of the analog waveform. Credit is also due to Vladimir Kotelnikov, E.T. Whittaker and others working in this area in the late 1920s and early 1930s. This powerful *sampling theorem* [4] is expressed and proved as follows:

THEOREM: If a continuous time function x(t) is bandlimited to frequencies in the range $-W < f < W$ Hz, then

$$x(t) = (1/2W) \sum_n x(n/2W)\,[\sin 2\pi W(t - n/2W) / 2\pi W(t - n/2W)], \qquad (1)$$

where the samples of x(t) are taken at time intervals of 1/2W.

PROOF: Since the Fourier transform X(f) of x(t) is limited to $|f| < W$, we can represent it (in the frequency domain) by a Fourier series for a periodic function, with period 2W, that equals X(f) in the interval $-W < f < W$:

$$X(f) = \left\{ \begin{array}{l} \sum_n c_n \exp[-j2\pi nf/2W], \ \ |f| \le W \\ 0 \text{ elsewhere} \end{array} \right\} = \sum_n c_n \exp[-j2\pi nf/2W]\ U_W(f), \qquad (2)$$

where $U_W(f)$ equals 1 for $|f| < W$ and is zero otherwise, and the $\{c_n\}$ are Fourier coefficients to be determined below. Recalling that multiplication in the frequency domain corresponds to convolution (indicated by a star *) in the time domain, and denoting the inverse Fourier transform by F^{-1},

$$x(t) = F^{-1}\,[\sum_n c_n \exp[-j2\pi nf/2W]] * F^{-1}\,U_W(f) = \sum_n c_n\, \delta\,[t - n/2W] * 2W\,[\sin(j2\pi Wt) / 2\pi Wt]$$

$$= 2W \sum_n c_n\,[\sin 2\pi W(t - n/2W) / 2\pi W(t - n/2W)], \qquad (3)$$

where we have used the facts that the Fourier transform of the delta time function $\delta\,[t - n/2W]$ is the exponential

$\exp[-j2\pi nf/2W]$, the Fourier transform of the time function $2W\,[\sin(j2\pi Wt) / 2\pi Wt]$ is $U_W(f)$, and the convolution of a delta function with a time function simply picks out a specific sample of that time function. We determine the values of the Fourier coefficients $\{c_n\}$ in equation (3) as follows to complete the proof:

$$x(m/2W) = \int_{-\infty}^{\infty} X(f)\, e^{j2\pi f(m/2W)}\, df = \sum c_n \int_{-W}^{W} e^{j2\pi f(m/2W)}\, df = 2Wc_m\,, \tag{4}$$

where (2) has been substituted for X(f) and it has been noted that all integrals are zero except when n=m. Thus

$$c_m = x(m/2W) / 2W, \tag{5}$$

and substitution into (3) completes the proof.

1.2.2 PCM

Digital transmission was not seriously implemented in the telephone network until the 1960s although the foundation for it was the invention of pulse code modulation (PCM) by Alec Reeves in the U.K. in 1938 [5]. PCM saw limited use during World War II, notably for encrypted communication between President Roosevelt and Prime Minister Churchill. PCM started with the sampling theorem, presuming samples taken at (typically) 8000 samples/sec to represent a telephone voice signal bandlimited to 4 KHz. Reeves' big additional step was to approximate each sample by a digital word, normally eight bits in length, more than adequate for good quality telephone speech. This means that each sample is rounded off to one of 256 quantization levels, each of which is represented by an 8-bit digital word, that with 8000 samples/sec results in a 64 Kbps data stream. Figure 1.3 shows a simplified PCM coding into 8 quantization levels, each represented by a 3-bit digital word. For the waveform shown, the six quantized samples result in the digital data stream 110 011 001 100 001 101.

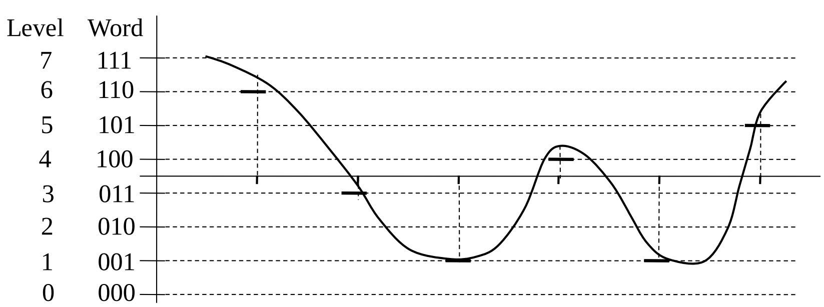


Fig. 1.3. Pulse-code modulation (PCM) rounds off a sample to the nearest quantization level, and represents that quantization level with a fixed-length digital word.

The difference between an actual sample value and the nearest quantization level is the quantization noise. SNR (signal to noise ratio) is a common criterion for transmission quality. When this noise is quantization noise alone, the quantization levels are uniformly spaced as shown in Figure 1.3, and both the signal and the noise can be statistically described, the SNR is the ratio of signal power to noise power expressed as [6]:

$$SNR = \text{signal variance / quantization noise variance} = 3\,(\sigma_x^2/\,x_{max}^2)\;\,2^{2b} \qquad (6)$$

where σ_x^2 is the signal variance, x_{max}^2 is the square of the maximum signal amplitude, and b is the number of bits per digital word, eight in the usual case. This expression makes the important point that although digital transmission is not without noise, it has the great advantage that this noise level is set in advance, as small as possible consistent with the capacity of the transmission channel, and is not increased during transmission or storage. Of course, digital errors can occur during transmission due to channel noise and distortion that the design of digital transmission systems must take into account. Later chapters of this book will describe in detail many of the mechanisms used in modern telecommunications to realize high transmission efficiencies (bits per Hz of bandwidth) while keeping the digital error rate below an acceptable level.

Although digital transmission is commonly associated with efficient use of bandwidth, uncompressed PCM is not necessarily bandwidth efficient. For example, baseband bipolar encoding (that transmits 2 bit/sec/Hz) of a PCM data stream will require 32 KHz bandwidth for a 64 Kbps PCM stream, far more than the 4 KHz bandwidth of the original analog speech waveform. It is only through highly effective compressive coding that we realize rates as low as 8 Kbps for speech signals of reasonably good quality. Together with efficient modulation techniques, this can result in spectrum occupancy significantly lower than 4 KHz.

1.2.3 Digital Transmission Hierarchy and SONET

A major motivation for applying PCM in the public switched telephone network (PSTN) was to realize efficient time-division multiplexed (TDM) transmission for multiple voice channels between telephone offices. For the relatively short distances within a local exchange network, existing copper circuits could carry not just one but 24 simultaneous conversations. TDM operates with a frame containing one word from each of a number of digitized voice streams, as illustrated by the 128 microsecond DS-1 frame containing 24 voice channels (eight bits each) and one synchronization bit shown in Figure 1.4, which transmits data at 1.544 Mb/s. This frame is transmitted synchronously - strictly clocked and regular - 8000 times per second, suitable for regularly sampled signals such as voice and (some) video. This rate was widely implemented on T1 carrier facilities. As time went on and transmission rates increased on microwave and optical facilities, a TDM hierarchy developed with much higher rates

(Table 1) to accommodate far more voice channels on high-speed media, up to the Synchronous Optical Network (SONET) standards extending to the OC-768 rate of almost 40 Gbps.

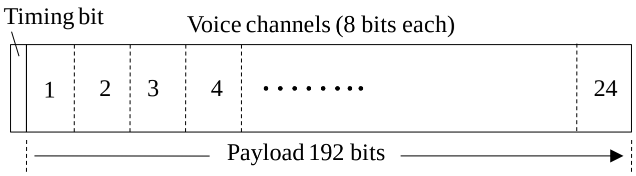


Fig. 1.4. The 128μs DS-1 frame. Each slot contains the bits representing one PCM sample of a particular source.

TABLE 1 - Digital Transmission Hierarchies [7]

DS-0: 64 kbps (one voice channel) DS-1: 1.544 Mbps in North America and 2.048 Mbps in Europe DS-2: 6.312 Mbps DS-3: 44.736 Mbps DS-4: 273 Mbps SONET (Synchronous Optical Network) commonly used rates OC- 1: 51.84 Mbps OC- 3: 155.52 Mbps OC- 12: 622.08Mbps OC- 48: 2.48832Gbps OC-192: 9.95328Gbps OC-768: 39.81312Gbps

Software-controlled electronic switching systems were deployed even before the voice signals themselves were digitized, with "stored program control" of switching functions. And when PCM transmission was introduced, so was fully digital circuit switching. Internally, the telephone network became a fully digitized line-switched system, with analog telephone signals from subscriber access lines converted in the central office channel banks into digitized signals. Only 4 KHz of the bandwidth available on the subscriber access line was used, assured by cutoff filters. Figure 1.5 provides an overview of the organization of the digitized public telephone network, including assembly of traffic from diverse access systems, subscriber and trunk lines, and digital switches in a hierarchy of functions and capacities.

Figure 1.6 illustrates three typical access systems: a two-wire analog telephone line converted into four-wire (separate in each direction) circuits, the upstream one connected to an analog to digital converter and the downstream one to a digital to analog converter ; digital subscriber line (DSL); and passive optical network (PON). Note that a low-pass filter in the analog to digital (A/D) converter limits use of the copper subscriber line to 4 KHz, appropriate for voice but a severe limitation for data transmission. The digitized information streams from these access systems may be combined in the frames used for digital transmission systems, from the modest DS-1 frame at 1.544 Mbps to SONET frames at the high rates described in Table 1.

SONET, an international standard since1988, enabled high speed transmission in the international public backbone network of both synchronous (regularly clocked) traffic, such as PCM speech, and asynchronous traffic such as ATM and IP. SONET originated and is most widely used in North America, while a similar international standard, SDH (synchronous digital hierarchy), is more widely used in Europe.

Figure 1.7 shows the earliest and lowest rate SONET frame. The frame consists of nine rows of 90 bytes, i.e. nine rows of 720 bits each. Frames are transmitted 8000 times per second, so that a single byte per frame can carry a 64 Kbps PCM data stream. The total transmission rate is 8000x9x720 = 51.84Mbps, the OC-1 rate shown in Table 1. With three columns of frame overhead and one of path overhead, the user data payload is 49.536 Mbps. It is carried in a subframe called the synchronous payload envelope (SPE), with nine rows and 87 columns including the path overhead column, leaving 86 columns for user data. The SPE may be contained in a single frame or overlap two frames as illustrated in Figure 1.7. When it overlaps, bytes shown to the right of the frame wrap around into the slots to the left of the SPE. Bytes shown underneath the frame map into columns 4-90 of the next transmitted frame. A multiple of OC-1 frames are byte interleaved to realize a higher SONET rate, but path overhead may be constrained to one column instead of the multiple number of columns. Figure 7 illustrates transport of a single DS-1 stream at 1.544 Mbps, corresponding to 24.125 bytes in each SONET frame, in three columns, a bit wasteful since three columns implies 27 bytes per frame.

Although, for trunk circuits, digital TDM largely replaced the analog frequency division multiplexed (FDM) systems for carrying multiple telephone calls, FDM has returned in optical transmission systems where use of multiple carrier wavelengths can vastly increase the capacity. Each wavelength in such a wavelength division multiplexed (WDM) system is likely to carry one of the SONET TDM rates shown in Table 1.

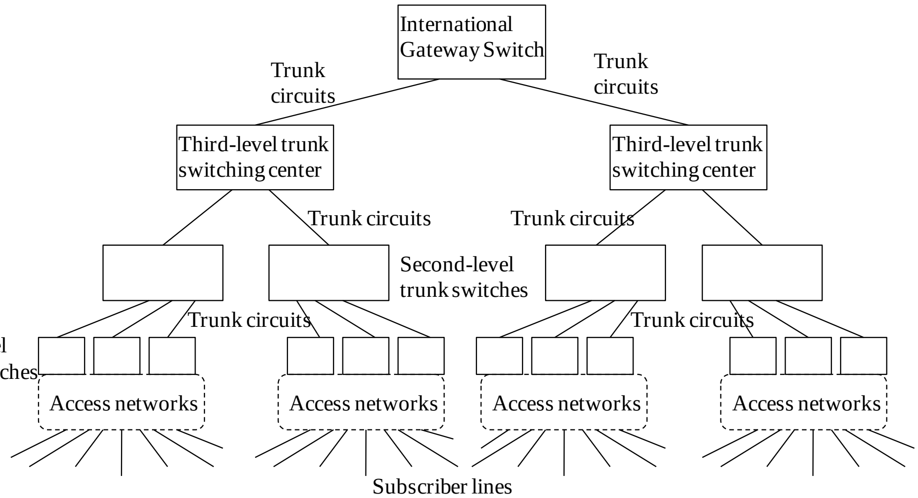


Fig. 1.5. The digitized public network (drawn in part from [8]).

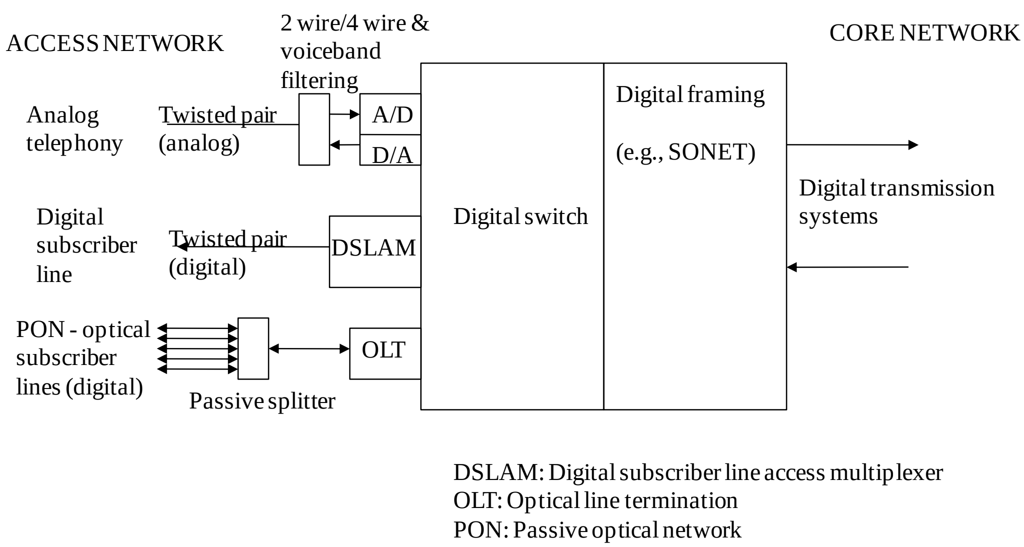


Fig. 1.6. Assembling traffic in a central office for digital transmission systems.

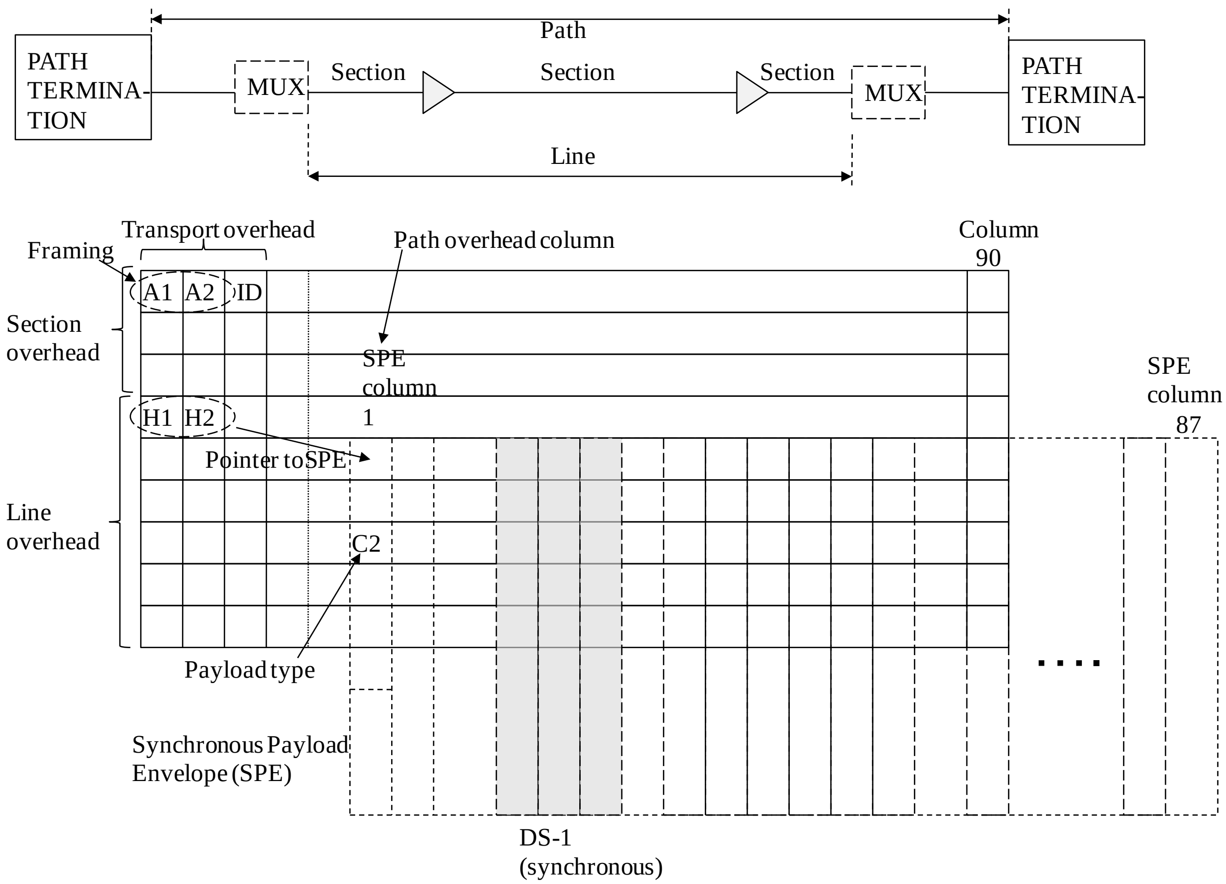


Fig. 1.7. A SONET frame for transmission at the OC-1 rate. Each square represents one byte (8 bits). The rows of 90 bytes are transmitted sequentially. Drawn in part from [9].

Many different kinds of traffic can be multiplexed into a SONET frame, not only PCM voice streams. Although SONET is designed for synchronous traffic only, of which DS-1 is an example, other traffic types such as ATM and IP, both described later in this chapter, can be encapsulated into synchronous streams. Ordinarily a single SPE will carry only one type of traffic since traffic requirements may differ, such as a delay for IP packets that may be unacceptable for synchronous voice traffic, but mixes are possible. The path overhead column allows identification of the traffic type.

1.3 MICROWAVE, OPTICAL AND SATELLITE TRANSMISSION SYSTEMS

Before optical transmission systems became the workhorses of the backbone network, analog voice and the early digitized voice traffic were transmitted over long distances using either coaxial cables or (on land) through fixed microwave transmission systems. A voice channel often used both coaxial and microwave transmission segments.

In 1932 Marconi designed what was probably the first regular microwave relay link for telephony, connecting Vatican City with the Pope's summer residence at Castel Gandolfo (www.seas.columbia.edu/marconi/history.html). However, it was the invention of a good microwave amplifier, the klystron, in the late 1930s, that made microwave carrier systems practical.

The earliest implementation in the public network was probably the 1947 microwave transmission carrier system linking New York and Boston through a sequence of seven microwave relay towers [10]. It used four radio channels in the 4 GHz band, each carrying 480 analog telephone circuits or one analog television signal. Figure 1.8 pictures one of the relay towers. Microwave relay was a relatively low cost option for long-haul transmission and its use expanded rapidly until optical transmission systems, with their vast capacity and high reliability, took over in the 1980s.

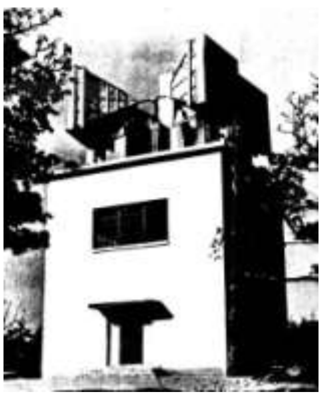


Fig. 1.8. The Bear Hill microwave relay tower near Waltham, Massachusetts in 1947 [www.long-lines.net/documents/latest-word/P06.html].

1.3.1 Optical Transmission in the Backbone Network

The huge and rapidly increasing demands for high rate digital communication, created in large part by streaming of video content, drove introduction of optical transmission systems beginning in the 1970s. Experimentation with fiber optic transmission had begun much earlier. In 1960, glass fibers showed a loss of around 1 dB/meter, not yet usable for communication over a reasonable distance [11]. The laser (Light Amplification by Stimulated Emission of Radiation) needed to generate a strong light carrier signal was invented about the same time, prototyped by Theodore Maiman with reference to the theoretical work of Charles Townes and Arthur Schawlow, who had invented the microwave "maser". The great advance in low-loss optical fiber came from the pioneering experimental work of Charles Kao at the Standard Telecommunications Laboratories in the U.K. He realized that impurities in glass led to high loss but predicted in 1966 that loss could be driven below 20 db/km even though it was still, at that time, of the order of 100db/km. He demonstrated a physically robust structure with a glass core of three to four microns diameter clad in glass of slightly smaller refractive index, with a total waveguide diameter of 300-400 microns. Optical waves were propagated along the interface between the core and the cladding. By 1970 the Corning Glass Works, using fused silica, had realized single-mode fibers with loss below 20dB/km at 633 nm wavelength [11]. Semiconductor diode lasers were developed at about the same time. By the 1980s, fiber loss was well below 1 dB/km and long distance fiber transmission was a reality.

While this history was evolving, it was, for some time, not sure that optical transmission would be adopted rather than millimeter microwave systems of small waveguides, both illustrated in Figure 1.9. The rapid reduction in the transmission loss of single-mode (one coherent optical wave) transmission, together with development of low-loss splicing and other favorable attributes, carried the day in favor of optical transmission.

Fig. 1.9. Transmission media for millimeter microwave and optical transmission systems.

There followed a rapid development of long-haul optical transmission (Figure 1.10), achieving a span of more than 1,000 km with repeaters at 80-100 km intervals and offering an effective bandwidth of several THz when WDM (wavelength division multiplexing) is employed [12]. Tunable lasers also helped. Each wavelength in a set of carriers feeding a WDM long haul optical transmission system is individually modulated by, for example, a Mach-Zehnder modulator, typically at a rate a little over 40 Gbps to accommodate an OC-768 data stream. This rate may already have doubled when this book appears. The modulation initially used on-off keying, but current preference, to achieve very high rates, appears to be for coherent (phase-sensitive) modulation such as quadrature phase-shift keying (QPSK) together with use of quadrature polarizations. Orthogonal frequency division multiplexing (OFDM) is also a serious candidate for optical transmission systems with its potential to "eliminate virtually all inter-symbol interference (ISI) caused by chromatic dispersion and polarization mode dispersion (PMD)", together with its spectrum utilization flexibility that "makes it an ideal candidate for networks with many reconfigurable optical add and drop multiplexers (ROADMs)." [13]

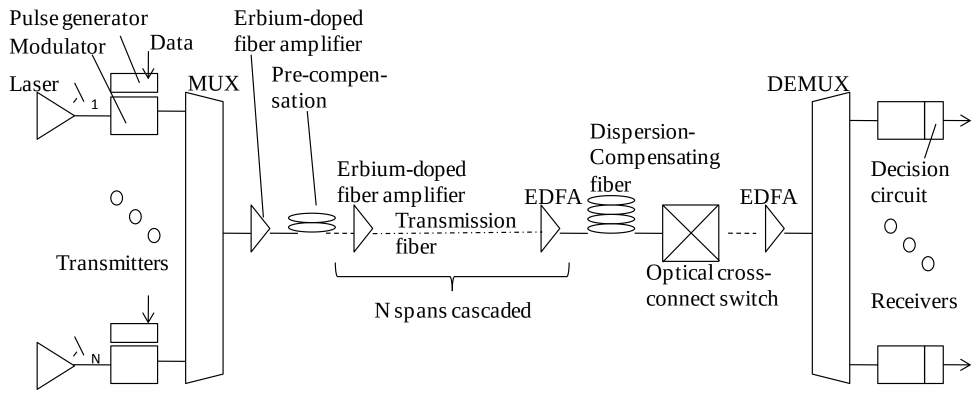


Fig. 1.10. A long-haul WDM optical transmission system (adopted from [12]).

Optical fiber transmission, with its freedom from interference and attenuation beyond that of the fiber, dominates optical transmission but free space optical transmission, used since ancient times, still has a role. It is found in infrared transmission within a room and in campus-scale building to building transmission systems. Outdoor use is constrained by vulnerability to high transmission loss in bad weather.

1.3.2 Communication Satellites for Backbone Transmission and Broadcasting

Satellite applications may be roughly organized in three categories: Backbone transmission similar to that of the microwave and optical systems just described; broadcasting of programming to local distributors (cable operators and telcos) and to end users (Direct Broadcast Satellite or DBS); and point-to-point communications. The British scientist and writer, Arthur C. Clarke, in 1945 envisioned DBS via a system of three space stations in geosynchronous orbit 36,000 km above the surface of the Earth, with a large population of users receiving with 30 cm antennas [14]. Although acceptable for broadcasting purposes, the propagation delay up to a geosynchronous satellite and back down again is a bit more than 1/4 second, which combined with other network delays can be undesirable for real-time voice communication. Echo cancelers mitigate but do not entirely eliminate this disadvantage. For this reason low earth orbiting (LEO) satellite systems were also introduced into the global communications infrastructure although they have been largely overtaken by terrestrial and undersea optical fiber transmission systems.

The first communications satellite, Echo 1, a passive reflective balloon, launched in 1960, was not in geostationary orbit, nor were the active Telstar satellites of the early 1960s [15]. In 1964, when the international consortium Intelsat was formed, the first geostationary satellite, Syncom 3, was put into orbit, followed in 1965 by Intelsat 1

("Early Bird") that supported 240 4KHz voice circuits between the U.S. and Europe.. These satellites and others that followed still relayed analog, not digital signals.

Intelsat satellites began the full operational use of digital time division multiple access (TDMA) in the mid 1980s, with Intelsat-V transponders operating in the C-band, with frequency bands defined in Figure 1.11c). Each transponder was capable of 120.8 Mbps transmission in a 72 MHz bandwidth, and the satellite also performed dynamic beam switching [16].   Spot beams covering relatively small areas on the Earth's surface permit reuse of frequencies for customized content and even individual communications channels to a much larger popular of users than if everyone received the same broadcast content. Digital transmission advanced to a much higher frequency band in 1993 with NASA's ACTS experimental geosynchronous satellite [17].   It operated in a portion of the lightly used Ka band (26-40 GHz) and featured onboard switching and the use of "hopping" (rapidly switched) very thin spot beams (Fig. 1.11a).

Figure 1.11b illustrates a LEO system, with satellites that are not stationary but instead move in several different orbits at low altitudes, typically 800 km above the surface of the Earth. The much shorter transmission path results in short propagation delays that make this system suitable for real-time applications. Signals from the satellite are also much stronger, when received by communications devices on the Earth's surface, than those from the distant geosynchronous satellites. Traffic is relayed among satellites in addition to the links to earth stations. The famous Iridium system, with 66 satellites, was launched in 1998 using C-band uplinks and downlinks for compressed digital voice at 2.4 Kbps. It was a commercial failure in terms of return on the huge investment it required, but continues in operation with large upgrades in data rates underway.

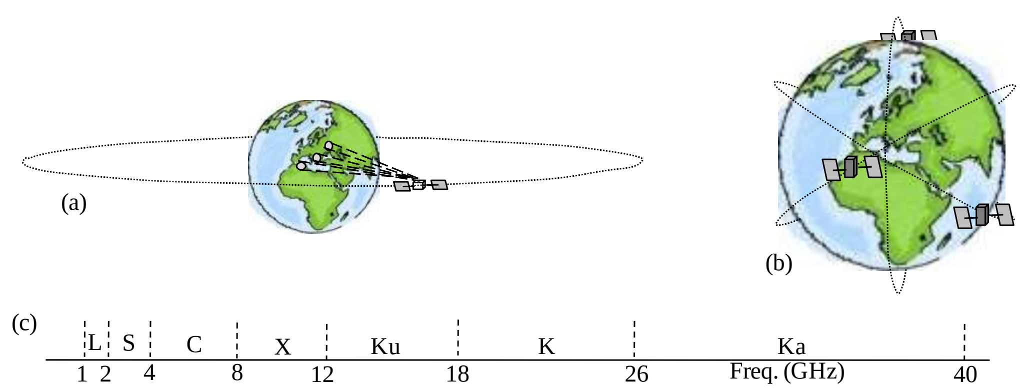


Fig. 1.11. (a) Geosynchronous satellite with spot beams. (b) Low earth orbiting satellite system with moving satellites. (c) Frequency bands in which satellites operate.   K band is sometimes included in Ka band. [www.esa.int/Our_Activities/Telecommunications_Integrated_Applications/Satellite_frequency_bands].

Even before digital transmission, distribution of video programming to cable headends had become a major business for satellite operators. Unlike real-time voice communication, this broadcast application was tolerant of the long propagation time through geosynchronous satellites. By the mid 1980s several C-band geosynchronous satellites were devoted largely to supplying cable programming to cable headends [18]. A single video channel initially required the full 36 MHz bandwidth of an early transponder, but digitally compressed video now has more modest capacity requirements.

Consumers first had direct access to satellites through C-band receiving stations, with relatively large (two meter) and expensive dish antennas, that tapped into programming intended for cable operators. DBS with small (46 cm) antennas began with geosynchronous satellite orbital assignments made by World and Regional Administrative Radio Conferences in the late 1970s and early 1980s, with eight satellites allocated to the United States. Currently, DBS services are available in the C-band (3.7-4.2 GHz), Ku-band (11.7-12.7 GHz) and Ka-band (18.3-18.8 GHz and 19.7-20.2 GHz) (sbca.com/receiver-network/satellite-overview.htm). A high-powered Ku-Band satellite radiates 120 to 240 watts per transponder, ten times as much as a typical C-band transponder, making possible the small dish antennas and reception in moving vehicles. In addition to downstream video and satellite radio programming, DBS satellites also support two-way broadband Internet access that is sometimes the only option for people living in rural areas.

Satellite propagation is usually in line of sight (LOS) channels with consistently reliable performance, but at the higher frequencies rain may significantly attenuate signal strength. Figure 1.12 illustrates attenuation, over time, of a downstream signal at the lower end of the Ka band during an initially heavy rainstorm.

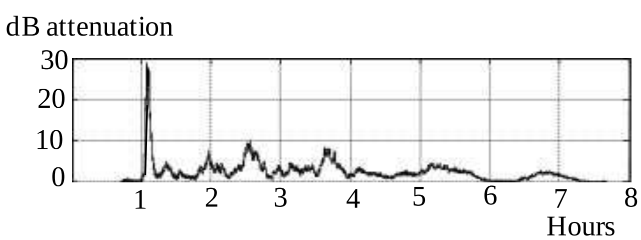


Fig. 1.12. Attenuation of a received 20GHz signal at different times during a rainstorm. [19].

1.4 DIGITAL ACCESS TECHNOLOGIES AND NETWORKS

Digital transmission was deployed in the backbone network well before it appeared in access networks available to consumers. But the need for digital access came early in banking, commercial transaction, and information access applications. The first step was to use the existing analog voice telephone network for this new digital traffic, doing the best one could with facilities designed to optimize analog voice transmission. The device required to do this was the MOdulator-DEModulator (modem).

1.4.1 Voice Channel Modems

End users, both residential and small business, had access only to twisted pair copper access lines running to telephone switches in central offices. A great deal of effort from the 1960s through the 1990s went into developing faster and more adaptive modems for dial-up access despite the severe limitations imposed by telephone network architecture. Cable modems came later, with cable access systems (Subsection 1.4.2).

The largest bottleneck was the constraint on bandwidth imposed on twisted pair telephone subscriber lines. Most twisted pairs, up to about 5 km, can provide about 1 MHz bandwidth, permitting data transmission at speeds of several Mbps. However, the telephone networks included 4 KHz filters in the central offices to support multiplexing of voice channels in both TDM (section 1.3) and analog frequency division multiplexed (FDM) carrier systems, as suggested in Figure 1.13. As a consequence, modem signals were limited to 4 KHz. The earliest modems used frequency-division duplexing (FDD), with completely separate transmission bands in the two directions. Data were initially sent at 300 bps in each direction using frequency-shift keying (FSK), By the 1970s modems incorporated the data-driven echo canceler, shown in Figure 1.14, to cancel self-interference from local transmitted signal and distant reflections of transmitted signal [20]. This permitted use of the same full channel bandwidth in both transmission directions at the same time, opening the door to significantly faster full-duplex performance, reaching 9600 bps in each direction by the late 1970s.

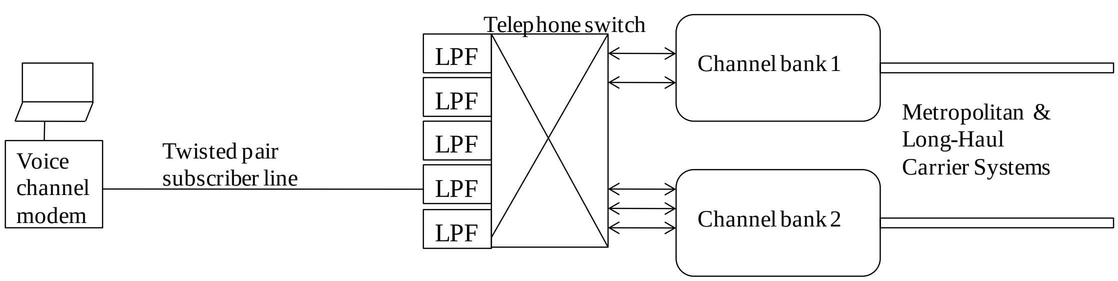


Fig. 1.13. The dial-up access speed bottleneck: voiceband low-pass filters in the central office.

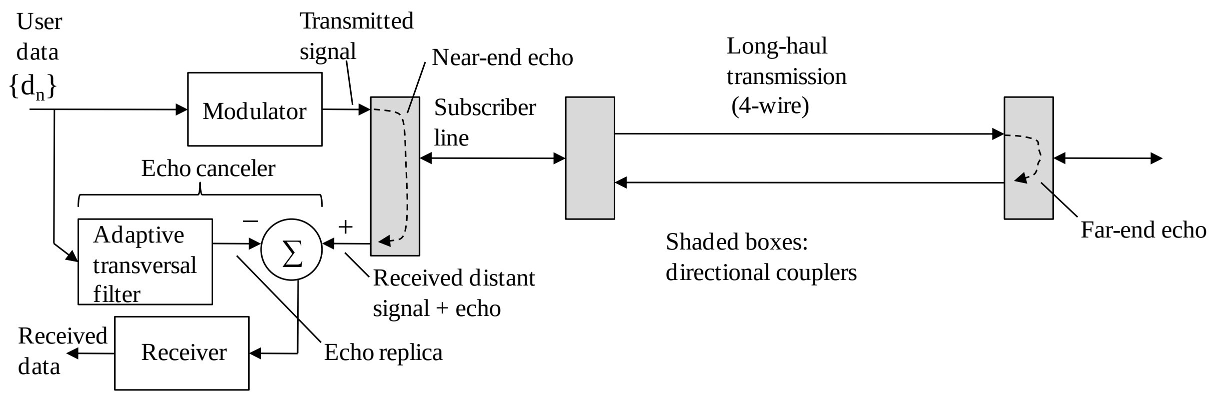


Fig. 1.14. Data-driven echo cancellation supported full-duplex data transmission in the same telephone channel bandwidth. The adaptive transversal filter is similar to that of an adaptive equalizer, Figure 1.19.

Of course, in a low noise environment but with distorted voice channels causing inter-symbol interference (ISI), relatively high rates, that exceeded 50 Kbps full duplex by the 1990s, were nevertheless attained by use of various techniques. These included partial response pulses, multilevel pulse amplitude modulation (PAM), a range of carrier modulation formats including phase-shift keying (PSK) and quadrature amplitude modulation (QAM), and sophisticated channel equalization and detection algorithms. In this brief introductory overview we offer only a few capsule descriptions.

PAM is simply the coding of bits from a string of data into pulse levels. The available bandwidth determines the maximum number of non-interfering pulses transmitted per second, which can be shown to be twice the bandwidth [9]. Given a desired data rate in bits per second (bps), a sufficient number of bits must be coded into each pulse in order to not exceed the limitation on pulse rate. For example, if we are transmitting 8000 pulses per second, also called 8000 baud, requiring 4 KHz bandwidth, and wish to carry data at a rate of 16 Kbps, two bits must be enclosed into each pulse. This corresponds to a choice among four pulse amplitudes as illustrated in figure 1.15. The pulses (Nyquist or Equivalent Nyquist) are designed to not mutually interfere, with their tails taking the value zero at the centers of neighboring pulses. The actual signal waveform is the sum of the pulses shown. The spectral efficiency is 4 bits/Hz. Bandwidth, baud, and data rate, often confused, are three different things.

For phase modulation and QAM, cosine and sine waveforms at a convenient carrier frequency (such as 1800 Hz) within the voiceband are each modulated by a sequence of pulse levels. 16-PSK (phase-shift keying), 16-QAM and 32-CCITT (International Consultative Committee on Telephone and Telegraph) V.32 are represented by the "signal constellations" of Figure 1.16. For each signal point, the horizontal coordinate is the amplitude of the cosine carrier, and the vertical coordinate is the amplitude of the sine carrier. Note that the more signal points to choose from with a fixed power constraint, the closer together the points are and more vulnerable to detection errors due to noise and distortion. It was common for voiceband modems to dynamically select the largest constellation, and thus highest bit rate possible, without exceeding an acceptable bit error rate.

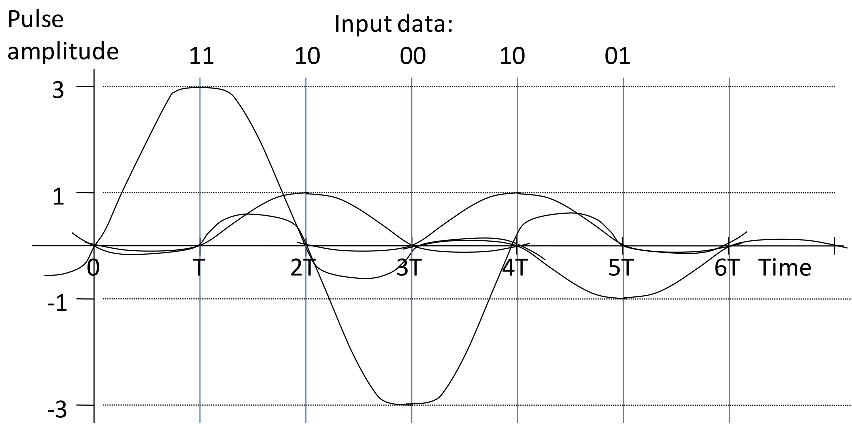


Fig. 1.15. 4-level PAM (pulse-amplitude modulation) coding two bits into a pulse level. T is the symbol interval, e.g., 1/8000.

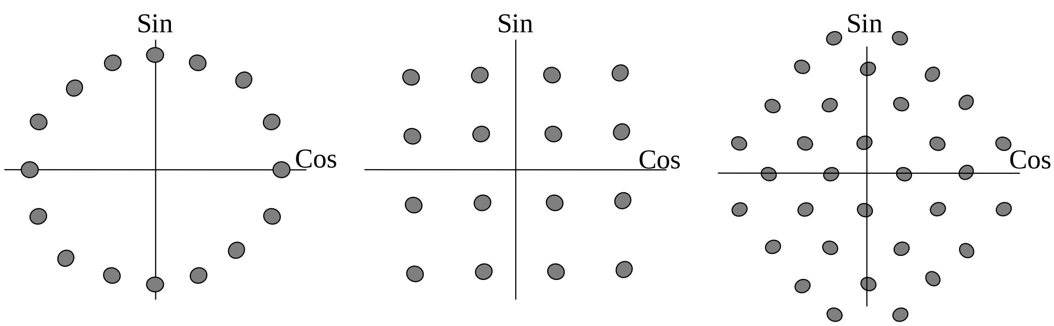


Fig.1.16. 16-PSK, 16-QAM, and 32-point CCITT V.32 signal constellations.

These are "passband" signal constellations because they represent modulation on carrier waveforms. Each again codes four data bits into each symbol. However, a special coding is used for the V.32 constellation with its constellation of doubled size as required for performance-enhancing channel trellis (convolutional) coding [21]. The increased size of the signal constellation implies, as the previous paragraph notes, some cost in immunity to noise and distortion. Nevertheless, the minimum Euclidean distance between coded line signal sequences is significantly increased, realizing a net gain. As a result, data can be reliably sent at the same rate as 16-QAM in channels with lower SNR than 16-QAM would require.

Voice channels in the telephone network are characterized by additive noise and by both linear and nonlinear distortions. Figure 1.17 illustrates the amplitude and group delay (first derivative of phase) characteristics of linear distortion typical in telephone channels of the 1960s-1990s period.

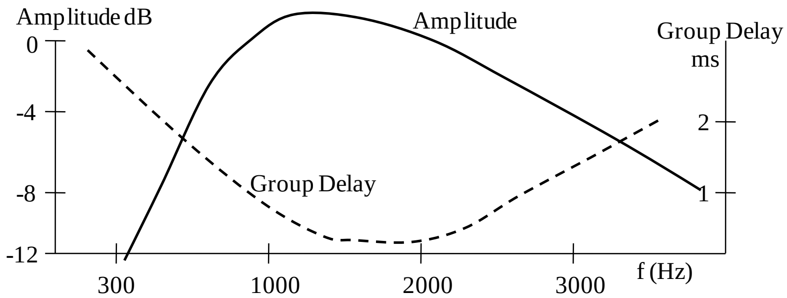


Fig. 1.17. Typical linear distortion of a telephone voice channel (adapted from [9]).

The technical response to linear distortion was both clever pulse design and powerful channel equalizers. The pulse design advances of the 1960s included "partial response" pulses, which caused intentional intersymbol interference between adjacent pulses in order to shape the spectrum so as to minimize the negative effects of channel distortion. This interference was predictable and could be removed during detection. The "duobinary" pulse, illustrated in Figure 1.18, was one of the most successful in this regard.

Adaptive channel equalizers [22], self-tuning linear filters in the receiver of a modem that in large part compensated for linear channel distortion, made possible large gains in voice channel transmission rate, from 1200bps to over 50Kbps, between the late 1960s and the mid 1990s. These instruments usually minimized the residual mean-squared error in the sample value presented to the detector. Figure 1.19 illustrates the basic concept of an adaptive equalizer realized as a tapped delay line, with the delay elements possibly shorter ("fractionally spaced equalization") than in a "synchronous" equalizer which has tap delays equal to the symbol interval T. The simple least mean square (LMS) adaptation algorithm is

$$\mathbf{c}_{n+1} = \mathbf{c}_n - \beta[y(nT+t_0) - \mathbf{r}_n \hat{a}_n]$$

where $\mathbf{c}_n$ is the vector of tap weights at the nth iteration of the algorithm, $y(nT+t_0)$ is the signal sample correspondence to the nth transmitted symbol, β is a small weighting factor, $\hat{a}_n$ is the decision made for the nth transmitted symbol, and $\mathbf{r}_n$ is the vector of voltages at the outputs of the individual delay elements [9].

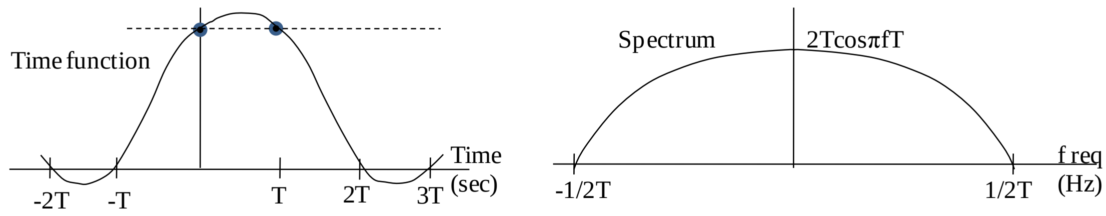


Fig. 1.18. The duobinary partial response pulse, in the time and frequency domains [23].

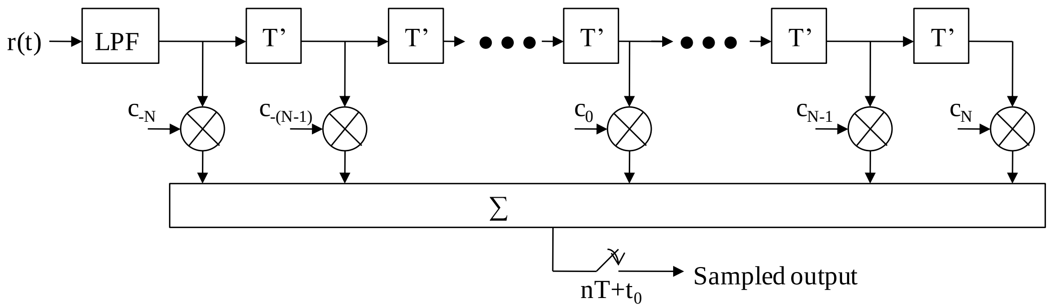


Fig. 1.19. An adaptive equalizer. The delay elements of T' seconds may be a fraction (typically half) of the symbol interval T to permit independent shaping of the filter's spectrum on both sides of the rolloff region [GSW].

Dial-up modems made data services and, in the 1990s, the Internet available to ordinary people. They were the beginning of the transition to relying on communications processing "intelligence" at the edges of the network, rather than only inside the network as was the long tradition of the public telephone network. Wireless devices similarly use modems for transmission over wireless access links, described in Section 1.6 below. Signal modulation, timing and synchronization, detection, and channel equalization, have been important elements in many phases of the history of digital transmission.

1.4.2 ISDN and DSL

The development of high-speed digital switching and digital transmission within the public network left the subscriber access line as the bottleneck, still analog and, from a bandwidth perspective, grossly underutilized. Engineers were acutely aware that, aside from modest-rate modem traffic, digital transmission did not extend all the way to the subscriber. The Integrated Services Digital Network (ISDN) [24] was an attempt to overcome this limitation. The basic ISDN interface shown in Figure 1.20 offered two 64 Kbps channels for PCM voice and one 16Kbps channel for control and data, thought to be sufficient in the 1970s. At the telephone office, the voiceband filters were removed to allow the transmission of the approximately 200 KHz bandwidth required for the upstream and downstream ISDN channels. A "primary" ISDN interface, intended for business users, provided twenty-three 64 Kbps channels for voice and one 16Kbps data channel.

The telephone industry expended considerable effort to deploy ISDN but it never really caught on, being perceived as too little, too late. It was followed, in the 1980s and 1990s by Digital Subscriber Line (DSL), introduced in the next paragraph) and by Broadband ISDN (B-ISDN) described in the following subsection in the context of Asynchronous Transfer Mode (ATM).

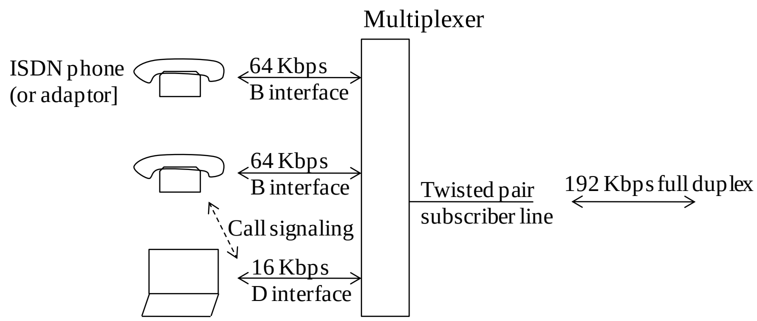


Fig. 1.20. The Basic ISDN interface.

DSL took the same perspective of avoiding voiceband filters in the telephone office and using more of the bandwidth of the twisted pair subscriber line. It went a large step further, offering downstream rates, in Asymmetric Digital Subscriber Line (ADSL), of 1.5-8 Mbps, depending on distance from the central office. ADSL, a concept largely credited to Joseph Lechleider at Bellcore, offers downstream transmission at a rate typically four to 10 times the upstream transmission rate. This reflects the far greater likelihood, at least in the past, that a residential customer will receive a high rate video stream than send one. It makes this service possible on a wide variety of telephone "local loops" (twisted-pair subscriber lines) in spite of severe crosstalk interference, introduced later in this chapter.

The much faster ADSL service doomed ISDN for anything beyond telephony. But it gave the telephone companies a way to become video services providers well before the expected triumph of fiber to the home. Experiments at Bellcore proved the feasibility of linking servers on a high-performance multiservice backbone network, such as an ATM network (Subsection 1.5.2), to provide content for ADSL subscribers. These experiments proved the feasibility of an effective video delivery service even if a truly broadband network could not yet be extended to residential subscribers [25]. Fiber in the home did eventually become practical with the introduction of the Passive Optical Network (PON, Subsection 1.4.4 below), but DSL has itself advanced with much faster speeds and remains a viable technology for residential access to the Internet. Figure 1.21 shows an xDSL access system, where the "x" represents any of several versions including ADSL and very high speed DSL (VDSL). A DSL Access Multiplexer (DSLAM) in the central office terminates a multiplicity of subscriber lines and provides an interface to ATM transport or directly to high-speed routers and the Internet. Table 2 defines many of the xDSL formats.

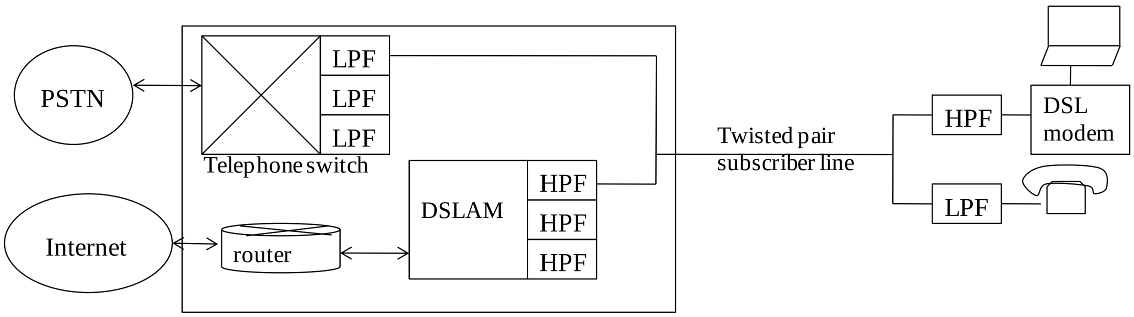


Fig.1.21. xDSL access system.

TABLE 2: xDSL Formats

xDSL	Standard	Downstream	Upstream	Symmetry
ADSL	ITU-T Rec. G.992.1	Up to 8 Mbps	Up to 1 Mbps	Asymmetric
HDSL	ITU-T Rec. G.991.1	784 Kbps, 1.544 Mbps, 2.0 Mbps	784 Kbps, 1.544 Mbps, 2.0 Mbps	Symmetric
SDSL	----------	Up to 2 Mbps	Up to 2 Mbps	Symmetric
VDSL	ITU-T Rec. G.993.1	Up to 100 Mbps	Up to 100 Mbps	Both
VDSL-2	ITU-T Rec. G.993.2	Up to 100 Mbps	Up to 100 Mbps	Both

xDSL historically began with two alternative modulation formats: CAP (carrierless amplitude-phase), very similar to QAM (Subsection 1.4.1), and discrete multitone (DMT), a version of orthogonal frequency division multiplexing (OFDM). DMT now dominates xDSL because of its superior ability to compensate seriously distorted channels, essentially avoiding the placement of energy in segments of the transmission band with very bad characteristics. OFDM/DMT transmits a data stream as a set of slower-rate data streams modulating narrowly-spaced subcarriers in parallel, contiguous frequency channels, a form of frequency division multiplexing in which all the streams are generated in one efficient computational algorithm, the Fast Fourier Transform (FFT) implementation of the Discrete Fourier Transform (DFT), operating on the input data [26]. Figure 1.22 is a simplified illustration of a point-to-point OFDM transmission system. Channel equalization is much simpler in such a system as compared with single-carrier data transmission. OFDM is also used in fourth-generation cellular mobile, WiFi, and wireless over-the-air audio and video broadcasting.

xDSL transmission is hampered by crosstalk interference between different twisted pair subscriber lines that are bundled together. Figure 1.23 illustrates how this crosstalk appears, in both near-end (NEXT) and far-end (FEXT) components.

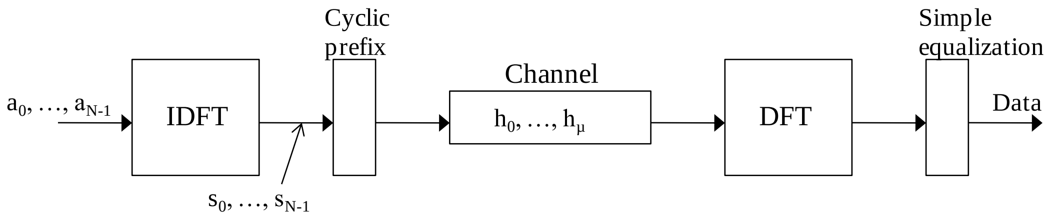


Fig. 1.22. An OFDM transmission system [26].

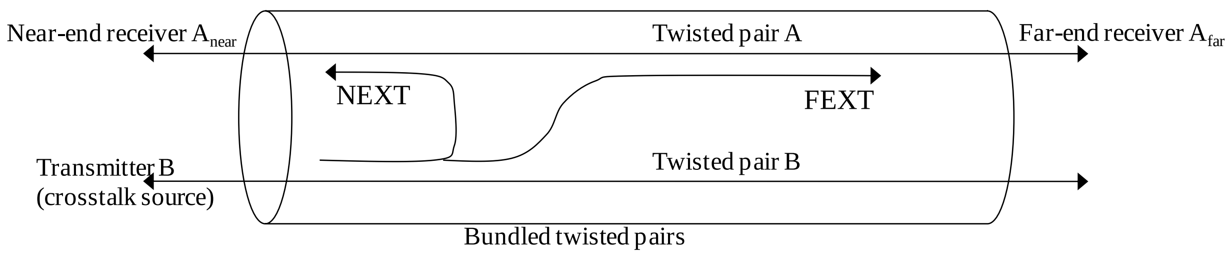


Fig. 1.23. Introduction of crosstalk into xDSL systems.

VDSL can provide very high rates in large part because of vectoring techniques addressing the significant FEXT interference [27]. Vectoring coordinates the detection of the separate VDSL signals on bundled groups of twisted pair subscriber lines, using a process of successive cancellation that is reminiscent of echo cancellation techniques used in modems. This strategy of coordinating multiple communications channels appears in wireless communications as well, in particular current initiatives in cellular mobile systems introduced in Section 1.8.

1.4.3 Cable Data Networks

From a business perspective, the telephone companies' pursuit of xDSL was primarily an effort to compete with the already successful cable television (CATV) industry that not only delivered video programming, but was becoming the main provider of high-speed Internet access. From its origins as a "community antenna" system, cable television quickly acquired satellite-delivered programming including lucrative "pay" channels, and even toyed, quite early, with interactive programming [18].

As a local access system, a typical CATV system operates a tree-like network with a metropolitan area scope. A cable headend may use trunk cables to serve several hubs, which spread out into feeder cables and drop cables.

Originally all of this was in coaxial cable, but reliable transmission was hampered by the need for up to 20 consecutive amplifiers in a large system. It was obvious, as soon as fiber optic cable became available, to replace the trunk cables and (later) even some of the feeder cables with fiber optic lines, leaving the expensive task of replacing all the drop cables to some future time. Thus the common model for a CATV network became the hybrid fiber-cable (HFC) architecture shown in Figure 1.24. Further improvement, such as addressable taps, made it unnecessary to send service personnel to change the filters assigned to individual drop cables when a subscriber's choice of program packages changed.

The higher quality of the transmission system further made it possible to use high spectral efficiency modulation schemes, such as 64-QAM, supporting digital video programming including HDTV with much greater efficiency than had previously been possible for distribution of analog video programs. In particular, each 6MHz channel previously used for one analog television signal could carry up to 30 Mbps digital data, enough for at least six digital television signals or one HDTV signal plus two regular digital television signals. As fiber to the home (FTTH) becomes more practical with the advances in PON (passive optical network, next section), the lower, coaxial cable part of the HFC system may be gradually replaced by the fiber distribution tree shown above the coax portion.

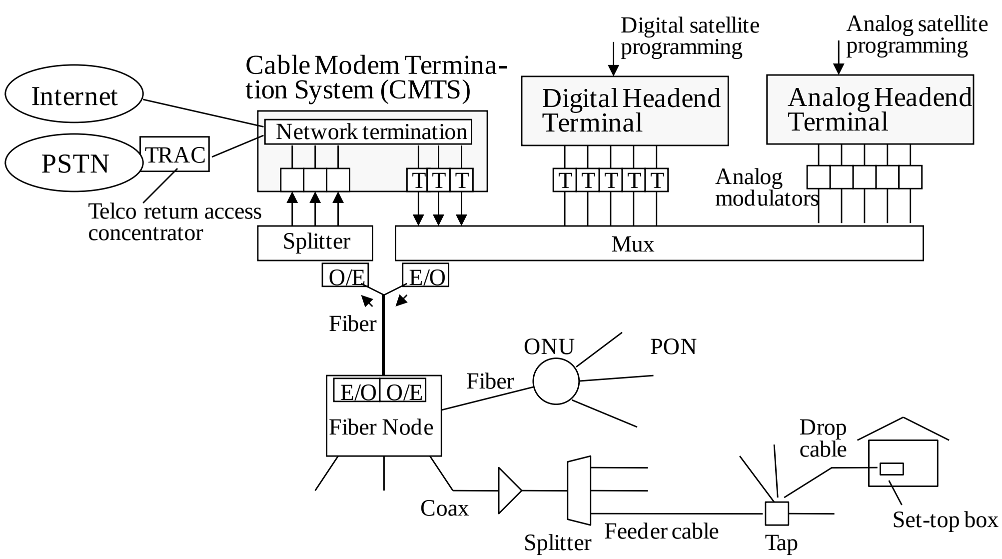


Fig. 1.24. A hybrid fiber-coax (HFC) CATV system.

The first technical standard for OFC systems was the Data Over Cable Service Interface Specification DOCSIS-1, promulgated by Cable Television Laboratories in 2000; DOCSIS 3.1 was the latest version in 2013, available from Cable Labs (www.cablelabs.com). Its objective was bi-directional transparent transfer of IP traffic between a cable headend and customer locations. Within the headend, transmission resources were under the control of the Cable

Modem Termination System (CMTS). In the downstream direction, QAM modulators, using signal constellations of either 64 or 256 points, generated data signals at 27Mbps or more in each 6 MHz channel. For upstream traffic, the rate was much less and a medium access control protocol assigned time slots and resolves contention. The upstream data rate was 0.2-3.2 Mbps. Some parameters have changed as the standard has evolved, particularly faster transmission in both directions, but the architecture is still very much the same.

1.4.4 Passive Optical Network (PON)

Still seeking to bring fiber all the way to residences to beat the performance promises (which may not always be the performance delivered) of cable operators, telephone companies are exploiting the relatively low-cost PON (passive optical network) architecture. PON is a high-performance "last mile" technology, with low-cost passive (not electrically powered) equipment in the field and the possibility of offering rates of up to 1 Gbps to each subscriber, in the upstream as well as the downstream direction [28]. Its reach is actually much more than a mile; it can be 40km in some cases. In addition to serving residential and business subscribers, PON has an important application to backhaul for cellular mobile base stations, metropolitan WiFi networks, and local communication systems such as PBXs (private branch exchanges). It is likely that as small base stations proliferate, the fast, economical backhaul capabilities of PON will become even more in demand. Figure 1.25 shows the generic PON architecture delivering voice, video programming, and Internet access services. and providing backhaul services as well, although it is unlikely that these would all be provided through a single PON.

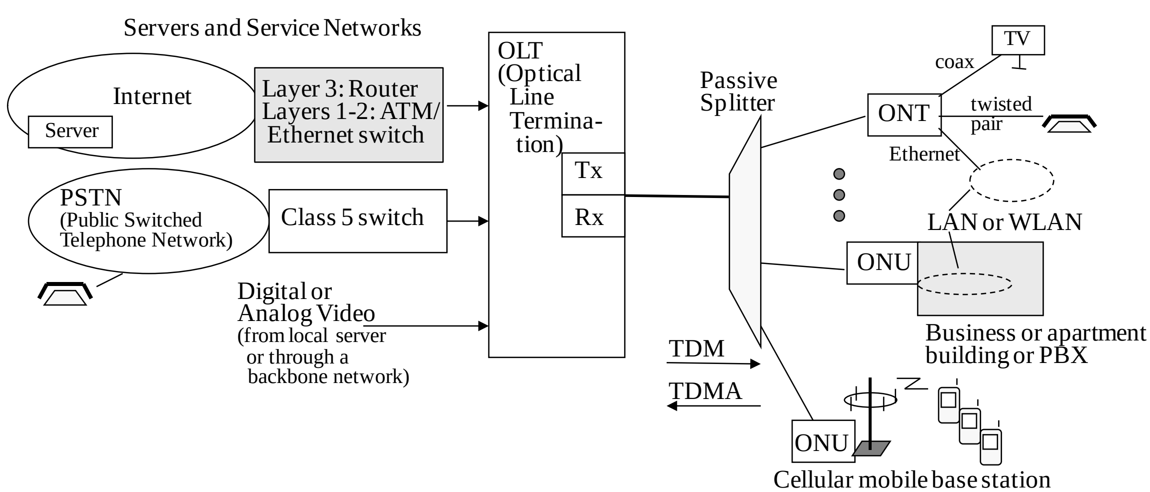


Fig. 1..25 Generic PON (passive optical network) architecture, adapted from [28].

The principal elements of a PON are the OLT (Optical Line Termination) in a central office, the passive splitter which typically shares the power of the downstream signal among 32 outgoing subscriber links (all optical fiber), and the ONT (Optical Network Termination) at a subscriber location.   There are three principal standards: the original broadband PON (BPON), ITU-T G.983.1; Ethernet PON (EPON); IEEE 802.3ah; and gigabit-capable PON (GPON), ITU-T G.984.1 and G.987. GPON and EPON, with 10Gbps shared among the attached users in current implementations, are competitors, with EPON sometimes seen as having a cost edge while GPON, to telephone companies, uses more of the traditional technologies such as SONET. Both use synchronous TDM and TDMA (time division multiple access) technologies, shown in Figure 1.26, for the downstream and upstream directions respectively. EPON additionally can use, for the appropriate kinds of bursty traffic, the original Ethernet CSMA-CD (carrier sense multiple access - collision detection) scheme for contention access, or one of its variants.

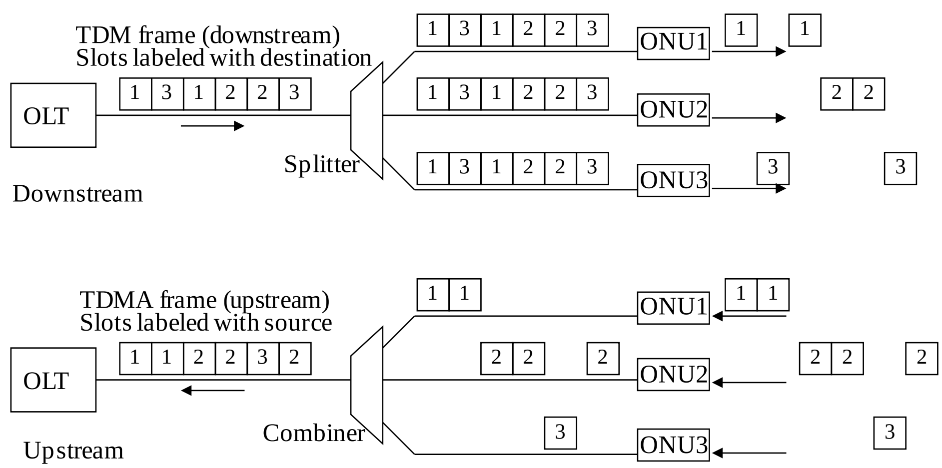


Fig. 1.26. Data transmission by TDM downstream and TDMA upstream.

Wavelength division multiplexed (WDM) PON will become the next generation standard. It can be configured such that each of up to 40 attached user lines gets one wavelength all to itself, at up to 10 Gbps in each direction. In this architecture the field unit is a (passive) wavelength multiplexer/demultiplexer, not a splitter. Of course splitters can also be used so that each wavelength can be shared by multiple users.

In single wavelength (in each direction) PONs, the downstream wavelength is typically 1490nm and the upstream wavelength 1310nm. A second downstream wavelength of 1550nm can be used for special video broadcasting purposes. At these wavelengths in the available single-mode fiber, the physical span of the network depends on the loss budget allowed in view of the input power. The power classes range from the low power class of PONs with transmission loss not exceeding 20dB, a split of as many as 1:16, and a reach of at least 10km, to the extended power budget class of PONs with transmission loss not exceeding 34dB, a split of as many as 1:64, and a reach of at least 20km [www.ieee802.org/3/ad_hoc/ngepon/public/14nov/ngepon_1114_hajduczenia_01.pdf]. Tradeoffs are available, going beyond the bound of one parameter in return for lowering the bound of another.

With the rise of PON solutions, optical transmission now covers the entire span from access network to long haul, including trans-oceanic, network. It is an essential partner of wireless communications which, far from being a separate networking operation, depends heavily upon a very high speed, low-latency backbone network.

1.5 THE RISE OF DATA COMMUNICATION

The dawn of the computer age began, very early, to make special demands on communications. One of the first, when computers were rare and expensive, was the concept of time-shared computing, the "computer utility", for a widely distributed community of users. M.I.T. housed Project MAC beginning with a Defense Advanced Research Projects Agency (DARPA) grant in 1963, and the concept is once again widely used, albeit in an even more distributed way and with a new set of cost-saving justifications, under the banner of "cloud computing". Dedicated private line data communication networks provided access to time-shared computing for large participating institutions, and dial-up modems allowed limited-speed access for individuals, but there was early interest in new communications techniques and networks that were better suited to computer traffic. Packet communications was perhaps the biggest and most revolutionary idea.

1.5.1 Packet Transmission

Computer communications applications are bursty and demand additional flexibility from data communication networks. The time and effort to set up and tear down a switched circuit was impractical for a short keyboard burst, and the uniform-capacity switched circuits of the telephone network were inappropriate for the varying rate, duration and quality requirements of data transmissions.

It was, however, unclear for some time just what mechanisms would be best suited not only for computer-oriented data communications but for the digitized voice and video streams heavily loading today's Internet. The community of researchers and developers interested more in computer-generated traffic than in voice calls favored message or packet transmission, the first defined as store and forward transport of a digital string representing an entire message, and the second defined as segmentation of a digital information object or stream into one or more bundles called packets that are individually routed through the transmission network. A fixed route and transmission resources do not necessarily have to be allocated in advance, so that call setup signaling may not be needed.

Researchers in the 1960s, notably Paul Baran in the U.S. and Donald Davies in the U.K., defined the packet concept and noted its value for "survivable" networks that could route traffic in alternate paths if a portion of a data network were damaged. The best known prototype of a packet network was the ARPAnet, a government-funded research network beginning in the late 1960s and initially linking a few locations, mainly university research centers, over 56 Kbps leased lines.

The story is widely available elsewhere (see in particular the Internet Society's history page, internetsociety.org/internet/what-internet/history-internet/brief-history-internet) and we only note here the conception, by Vinton Cerf and Robert Kahn, of the Transport Control Protocol/Internet Protocol (TCP/IP) mechanism facilitating the movement of packets in the ARPAnet and adopted today in the immense commercial Internet. Packets are transferred from one transmission link to the next, in the general direction of the final destination and possibly in entirely different physical networks, by a router/forwarder. This IP router, unlike a switch in a switched circuit network, does not have to maintain the state of a connection, which is a well-defined path and information flow between endpoints, since there is no connection, just individual packets being forwarded in the right direction. Figure 1.27 shows the generic structure of a typical router/forwarder, in which a routing algorithm, informed about the condition of links to neighboring nodes, decides which output line best conveys the packet closer to its destination. There may be several different possible routes to the destination device, and it is possible for the routing to change between individual packets of a particular flow, although they usually traverse the same route.

There is an important detail in Figure 1.27: Each input and output port includes a line card that exercises the underlying protocols used by the network to which it is connected. Thus one output port might connect to an ATM/optical network, while another might connect to a switched Ethernet, and similarly for input ports. IP packets are encapsulated into lower-level frames or protocol data units of a particular network at an output port of the router, and de-encapsulated at the end of the link to enter the router at the next node. The internal functions of the router are independent of whatever networks (with line cards) are attached, showing why IP is an internetworking protocol that works with all physical networks for which encapsulation protocols exist.

Although the IP service is "best effort", with no guarantee that the packet will not be delayed or dropped because of congestion, it saves a lot of effort in control signaling and offers a great deal of flexibility in routing. Different packets of the same "call" might take different routes. Reliability measures can and frequently are offered at a higher protocol level, notably TCP which establishes a "connection" and retransmits lost packets. Moreover, although communication quality is not guaranteed as in circuit switching, there is the opportunity, through classification of a packet into different quality classes and specification of priorities in the scheduling algorithm, to grant preferential treatment, such as minimized delay, to quality-sensitive traffic types. Packet transmission with this kind of selective treatment allows convenient mixing of traffic at radically different rates while realizing, or at least approaching, the quality demands of all users.

It may be difficult for today's Internet users to believe that IP was hardly visible in early commercial computer communication networks. Table 3 lists a number of these early networks [29]. The mechanisms they employed included X.25[30], a packet forwarding protocol, that, unlike IP, operates over pre-established paths. It implements virtual calls and, as in ATM, virtual circuits as in ATM (see later below). It follows a complex set of rules for realizing quality of service. Leased digital lines interconnected the store and forward nodes. Other networks forwarded whole messages or even used something close to traditional circuit switching. By the late 1970s, these commercial data networks were widely used and important to the global economy long before the commercial Internet appeared in the 1990s.

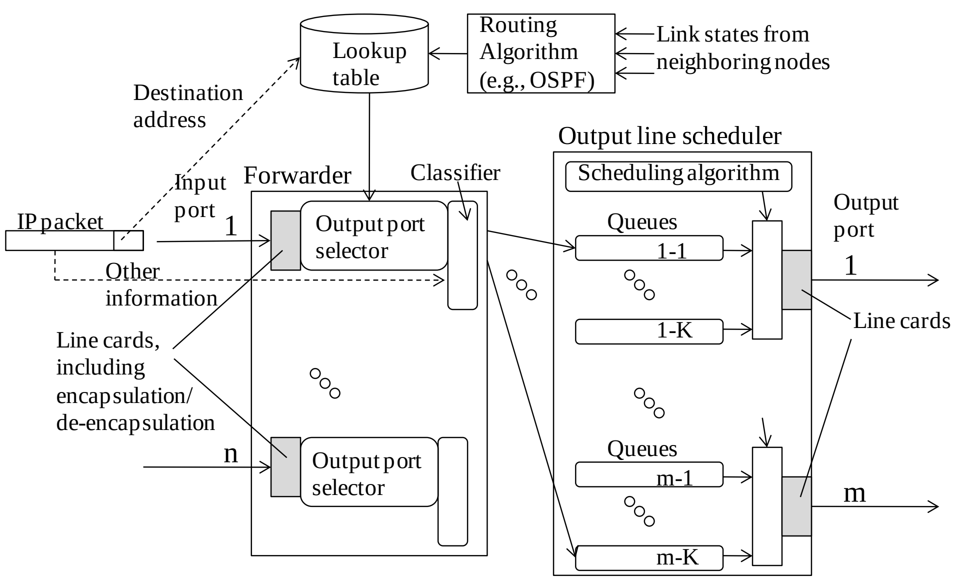


Fig. 1.27. Routing and forwarding at a node of an Internet Protocol network.

TABLE 3: Early Commercial Computer Communication Networks

DATAPAC - Canadian packet switching network using the X.25 host interface packet protocol that offers "reliable", dedicated resources.
TELENET - Packet-switched network that also was a forerunner of the Internet and used X.25.
TYMNET - Computer time sharing. Store-and-forward message switching along a virtual circuit (defined in the next subsection).
DATRAN - Circuit (line) switched.
SITA - Early message-switched network for the air travel industry.
SNA - IBM's System Network Architecture. Packet forwarding along fixed routes and the SDLC (Synchronous Data Link Control) protocol.
SWIFT - The international banking network that employed a proprietary protocol, changing to X.25 in 1990 and an IP network in the early 200s.
CREDIT AUTHORIZATION NETWORKS - Leased-line and dialin facilities, operated by credit card issuers such as banks participating in Visa and Mastercard and other issuers such as American Express, for authorizing credit card purchases. Generally used X.25.

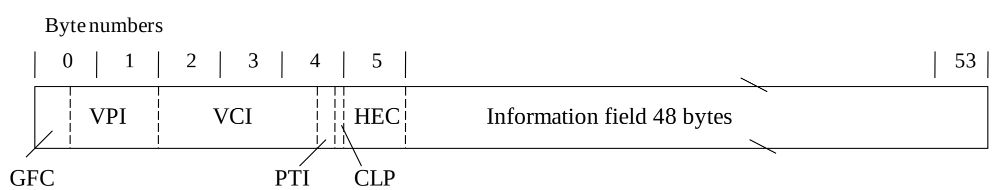
Today packet transmission and switching, dominated by IP, increasingly rules in wireless as well as wired networks. The mixture of different traffic types continues to drive deployment of packet transmission. Interestingly, packet transmission is adapting, under new names, some of the resource reservation and circuit switching aspects that were so successful in maintaining transmission quality in the PSTN. One of the most prominent innovations in this category is Multi-Protocol Label Switching (MPLS) [31] which switches, rather than routes, packets based on identifying labels and pre-allocated transmission resources. It goes part way toward the transmission capacity reservation properties of ATM, introduced in the next section. The gulf between the "telephone" and "computer communication" perspectives is not as large as it once was.

The actual transmission facilities used as links in data networks, including the Internet, are the same optical, microwave, satellite, cellular mobile and copper facilities used by what is left of the legacy telephone network. Packet traffic may, as noted earlier, be included in the SONET and other synchronous frames used over such facilities even though the packet traffic is not synchronous, i.e. packets arrive at input nodes on an irregular schedule. The obvious translation mechanism is buffering in which packets enter buffers at irregular times but are discharged from buffers at regular time intervals. There may, moreover, be an intermediate mechanism called ATM (Asynchronous Transfer Mode), a special packet mode that was, in the 1980s, the telephone companies' answer to the question "how can we efficiently transport all kinds of traffic on the same network".

1.5.2 ATM and B-ISDN

Packet switching was only reluctantly embraced by the telephone industry. Telco managers feared a degradation of the quality inherent in switched circuits with their pre-authorized transmission resources. A kind of compromise was realized in Asynchronous Transfer Mode (ATM) [32] that, in contrast to IP with its variable-length packets, utilized a fixed size packet, called a cell, of 53 bytes including a five-byte header as shown in Figure 1.28. ATM is very effective at providing whatever capacity is needed by different kinds of traffic; a higher-rate source simply issues cells at a higher rate than a lower-rate source and the ATM switch figures out how to mix this traffic in a fair way.

Transmission resources are pre-allocated to calls or sessions via *virtual circuits*. A virtual circuit may be defined as a point-to-point connection in a transmission system with a fixed path and assigned capacity, as in an ordinary circuit, but without dedicating a particular set of physical resources such as particular slots in transmission frames. The necessary capacity is realized as either a fixed rate or a statistical commitment. Furthermore, cell switches in the 1980s were much cheaper, for a given total throughput, than routers handling variable-sized packets. The 48-byte payload was a further compromise. The computer communication community had wanted a larger cell size to accommodate big data transfers, while the telephone industry, especially in Europe, wanted a smaller cell size to avoid cell-loading delays in real-time speech communication that could exacerbate echo problems.



GFC: Generic flow control (to manage flows from different sources)
VPI: Virtual path indicator (for aggregations of virtual circuits between the same end points)
VCI: Virtual circuit indicator (connection-oriented node-to-node channel)
PRI: Payload type indicator (3 bits) identifies data, maintenance, end of message, ….
CLP: Cell loss priority (1 bit)
HEC: Header error correction, computed as a polynomial operator on the header

Fig. 1.28. The ATM cell.

In the 1980s and 1990s, telephone companies had great hopes that the 155Mbps Broadband ISDN (B-ISDN) basic interface for residential subscribers, relying on ATM, would support both ordinary voice telephony and digital video services, especially video on demand to compete with cable operators. Devices performing digital video coding and decoding (codecs) with a high digital compression ratio, taking advantage of in-frame and frame-to-frame redundancies, were not yet available at an acceptable cost. The general presumption at the time was that an uncompressed digital video stream required about 140 Mbps in transmission capacity. Like circuit-switched networks, the virtual circuit ATM networks required an elaborate signaling protocol that appeared as the international standard Q.2931 [33].

B-ISDN was a broadband network consisting largely of optical transmission facilities including optical fiber to subscriber residences, ATM switches, the Q.2931 signaling protocol for setting up and managing virtual circuits and paths, and a set of protocols, above the ATM layer, for encapsulating different types of traffic into ATM cells. The principal rates specified for the user-network interface (UNI) were 155.52 Mb/s and 622.08 Mb/s, including the possibility of asymmetric interfaces with different upstream and downstream rates [34].

The B-ISDN protocol reference model [35], shown in Fig. 1.29, consists of three planes: a user plane, a control plane and a management plane. The physical transmission layer is shared by the user and control planes and consists, for example, of physical transmission media such as optical transmission systems and a framing mechanism using the medium, such as SONET which is able to encapsulate ATM cells. The next layer up is the ATM layer, providing data transfer for all types of traffic. The ATM Adaptation Layer (AAL) above that provides service-dependent functions to higher layers, such as, for example, mapping quality of service (QoS) dependent voice or video traffic into ATM cells, or alternatively encapsulating IP packets into ATM cells. Encapsulation, an important concept in digital transmission, means placing an unchanged protocol data unit (PDU), such as a complete IP packet, into the information fields of a lower-level PDU such as one or more ATM cells.

Above the AAL in the control plane are call control and connection control, while in the user plane there may be mechanisms for creating application-specific PDUs from original sampled audio and video streams. The management plane oversees faults, alarms, and corrective measures; performance issues; network configuration; accounting and security.

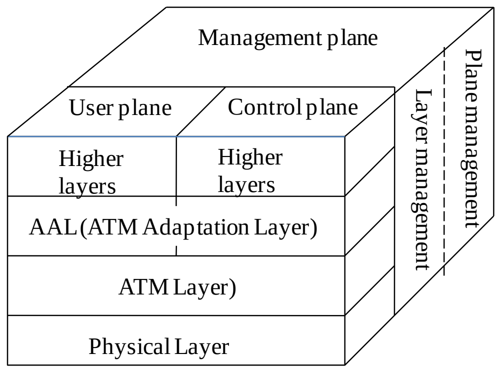


Fig. 1.29. The B-ISDN Reference Model.

When B-ISDN was planned in the 1980s, telephone engineers were optimistic that a large portion of the telephone subscribing public would be using it by the year 2000 and that this would enable telephone companies to successfully compete with well established coaxial cable operators. This failed to happen for a number of reasons, first and foremost that it simply cost too much. Also, by the 1990s cable operators were already introducing optical fiber in HFC (hybrid fiber-copper) systems, described in Subsection 1.4.3, that used fiber transmission to hubs in the field but retained the copper coaxial cables in the part of the system closest to subscribers, thus avoiding the large expense of running fiber to every subscriber location. Another factor in the failure of B-ISDN was that video compression became cheap and available, significantly reducing the required data rate for a video channel to the point (about 1.5Mbps) where more economical services, such as DSL, provided a better entry point into broadband services.

Over a longer period of time, optical access from subscriber locations has become relatively widespread and the (former) telephone companies that operate them have become successful video programming providers, relying heavily on PON (passive optical network) access systems described in Subsection 1.4.3.

1.5.3 Contention Systems: AlohaNet and Ethernet

Most transmission systems provide either guaranteed transport, over circuits or virtual circuits, or "best effort" transport as with the Internet protocol, where packets are queued at routers and a packet is sometimes dropped if there is not enough capacity on the desired outgoing link. There is a third option that has been implemented in several important communication systems, that of contention access on a shared medium, in which a packet or frame of data offered to the shared medium might collide with someone else's, and the collision is resolved by a "backoff" mechanism in which each of the contenders waits a specified time before trying again. This is the basis of the very widely used Ethernet, invented in the early 1970s by Robert Metcalfe at the Xerox Palo Alto Research Center, although it should be noted that many enterprises now use switched Ethernet in which there is no longer any real contention or collision.

AlohaNet, a radio contention system invented by Norman Abramson and others at the University of Hawaii, was a radio contention network linking computers scattered over the Hawaiian Islands. Remarkable for its simplicity as a multiple access system, it began with the "pure" Aloha protocol in which a transmitter would simply begin sending, but if no acknowledgement of successful transmission returned from the designated receiver it would back off for a randomly selected interval and then try again [36]. For light traffic it worked well, but when utilization approached 20 percent there began to be many collisions and performance quickly degraded. The system was enhanced with "slotted Aloha" which offered synchronized transmission time slots that significantly improved the collisions threshold, but still could not be considered a high-utilization network.

Ethernet was invented to aid communication sessions among the advanced Alto personal computers that were one of the pioneering advances at the Xerox Palo Alto Research Center (PARC) [37]. It was quickly generalized to serve any communicating host machine. Its great advance over Aloha was to listen first ("carrier sense") and transmit only if no other station was heard, a protocol labeled Carrier Sense Multiple Access - Collision Detect (CSMA-CD). For two active stations, a collision may nevertheless occur because of the propagation and receiver processing delays. When this collision is detected, each transmitter waits either 0 or 1 (random choice) frame times before transmitting again. If either transmitter should detect another collision, it waits 0, 1, 2 or 3 frame times before transmitting again, and so on, by a power of two, quitting after 16 attempts. Figure 1.30 illustrates a possible collision and retransmission sequence, with the T1 (transmitter 1) frames shown in actual time and the T2 (transmitter 2) frames when they are heard by T1 which will be some time after they are sent. Upon hearing the first collision, T1 chooses randomly between waiting zero frames or one frame and the random choice turns out to be one frame, represented as [1]. Not hearing anyone else, T1 sends its frame again. Transmitter 2, who also may have heard the collision, also waits one frame and, not hearing T1 because of the propagation delay, transmits again, detected by T1 as collision no. 2. This time T1 randomly selects a delay of 3 frames from the choices (0,1,2,3) before its third attempt. A third collision is detected and this time T1 waits 5 frames randomly selected from the set (0,1,2,3,4,5,6,7).

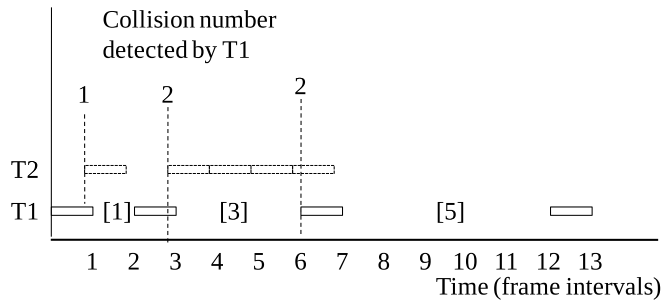


Fig. 1.30. Illustration of operation of the Ethernet retransmission protocol.

Multiple access backoff protocols advanced further, but there are now commercial Ethernet switches that avoid contention altogether, using the Ethernet frame specification for communication through an Ethernet switching hub with individual dedicated channels to the participating computers. Switched resources are requested and granted or denied, with no collisions occurring. Ethernet is almost universally used in residential and enterprise local-area wired networks using cable media..

1.6 WIRELESS TRANSMISSION SYSTEMS

Wireless transmission has a long history in many forms, including the terrestrial microwave relay transport of multiple voice circuits introduced in Section 1.3. The earliest commercial use of digital wireless transmission, beginning around 1900, was for radiotelegraphic ship-to-shore communications, driven by the safety and convenience needs of navies and passenger liners. Marconi was a leading figure in this early application, making many early demonstrations of radio transmission over both land and water in England and Italy, and later in the United States [38]. He was not the first or only implementer of radio transmission; Hertz offered demonstrations in the late 1880s but he did not use it for conveying information. Popov in Russia was another practical innovator like Marconi, demonstrating ship to shore communications to the Russian Navy in the late 1890s; but the Navy's insistence on secrecy may have kept Popov from receiving much recognition in Europe and the U.S. [39]

Fig. 1.31. Hertz, Marconi and Popov.

The spark gap transmitters used in these early systems generated electromagnetic waves over a very wide range of frequencies, so that the dimensions and resonant frequencies of the antennas structures effectively determined the transmitted frequency. Hertz's early demonstrations were at relatively high (60 to 500 Mhz) frequencies. Marconi's long distance transmissions, made much later, used huge antennas that effectively radiated at frequencies ranging from 45KHz to 850 KHz [40]. These low frequencies are appropriate for low loss "ground wave" transmission over long distances and helped make Marconi's experiments great successes.

By 1920 analog voice largely replaced digital telegraphy traffic in wireless systems. It was not until the advent of military high-frequency radio systems in the 1960s that data, including digitized voice, reappeared as the main traffic type. These systems, some of which implemented over-the-horizon transmission of data by low-altitude tropospheric scatter as illustrated in Figure 1.32, suffered from frequency-selective fading that attenuated parts of the sigmal spectrum

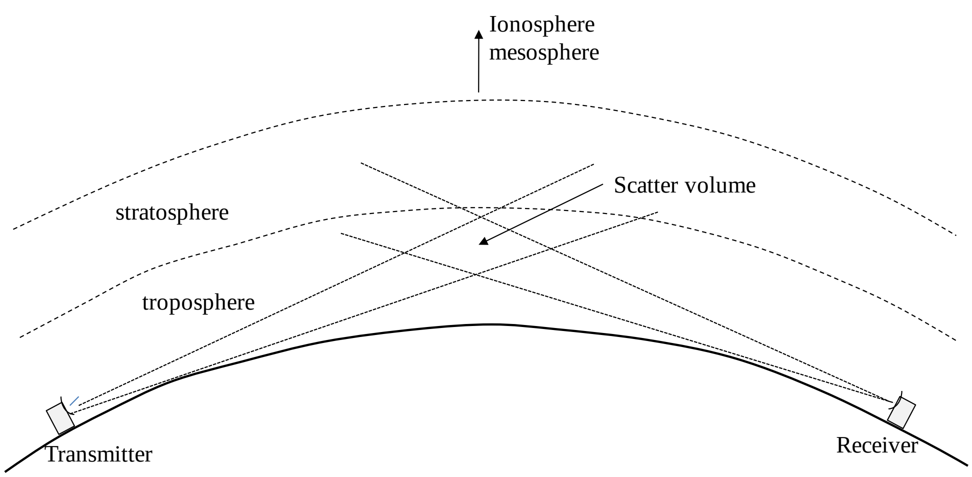


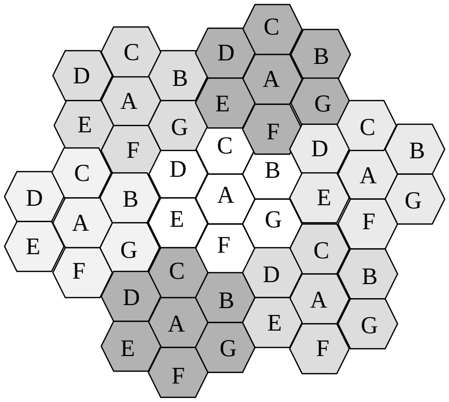
Fig. 1.32. Tropospheric scatter data transmission.

Full-bandwidth channel equalization for channels suffering deep selective fades is very difficult, which is why miltitary VHF data transmission systems were among the earliest users, in the 1960s, of OFDM (orthogonal frequency division multiplexing), introduced in Subsection 1.4.1. The KATHRYN high frequency radio, described in a 1967 paper [41], was a notable example. Its transmitter generated 34 subchannels within a total 3 KHz bandwidth using an analog discrete Fourier transform (DFT). Deep fades within the passband could be addressed by simply not putting energy into the lossy subbands, the same basic principle that later was so successful in DSL.

1.6.1 Cellular mobile: 1G to 4G

The first generation (1G) of cellular mobile communication systems, called Advanced Mobile Phone System (AMPS) in North America, used analog FM (frequency modulation) in 30 KHz channels and accommodated multiple users through FDMA (frequency division multiple access) in the 800-900 MHz band allocated by the 1976 WARC [World Allocation Radio Conference) [42]. The critical cellular structure (Figure 1.33a) came into commercial practice with the 1979 NTT deployment in Japan, followed by the 1981 Nordic mobile Telephone service and the 1983 AT&T AMPS (Advanced Mobile Phone Service) in the U.S. The cellular structure made possible multiple re-use of frequencies in different (separated) cells in contrast with earlier mobile systems that often had one antenna for an entire city and a very limited number of simultaneous channels . The separation between cells with the same carrier set is sufficient to attenuate signals to a non-interfering level. Additional frequency reuse within a single cell is possible with directive antennas at the vertices, a system that is called *sectorization*, with directional antennas represented by arrows in Figure 1.33b.

Wireless *digital* transmission came with 2G, the second generation cellular mobile systems deployed in the early 1990s exemplified by the North American IS-54 and IS-136 standards using TDMA (time-division multiple access), the European Group Speciale Mobile (GSM) standard that also used TDMA, and the North American IS-95 code-division multiple access (CDMA) standard. By use of digital voice compression, the TDMA standards supported three times as many voice channels in the same bandwidth as the old analog systems. In GSM, the first deployed of the TDMA standards, carriers separated by 200 KHz each carried "frames" of 8 time slots each, with a time slot illustrated in Figure 1.34. This provided just under 198 Kbps for traffic data, mainly multiple compressed digital voice streams but also including data from the 9.6 Kbps voiceband data modems of that time, and from ISDN. The upstream and downstream signals were conveyed full duplex on entirely separate carrier frequencies. IS-54, in contrast, kept the 30KHz spacing of carrier frequencies and implemented a 40ms frame containing three slots. Both systems implemented several compressive voice coders with data rates ranging from about 7 Kb/s to 13 Kb/s.



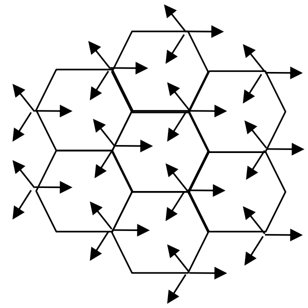


Fig. 1.33. (a) Cellular structure based on a reused seven-cell pattern. Each letter represents a different set of carrier frequencies. Adapted from [Stuber]. (b) Sectorization, increasing frequency reuse with directional antennas.

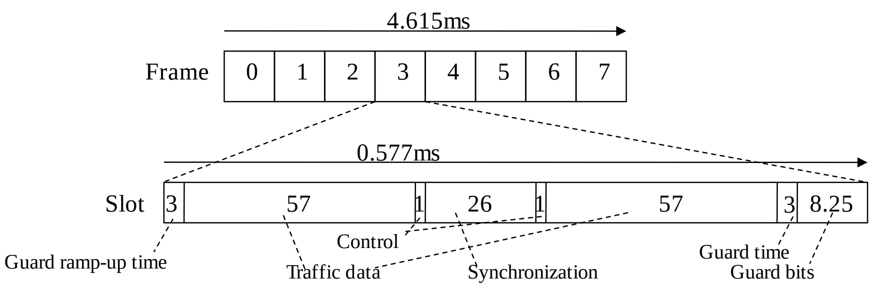


Fig. 1.34. GSM time slot is carried in an 8-slot frame on one of the carrier frequencies spaced at 200 KHz. Adapted from [42].

In the alternative IS-95 CDMA system [43] the user information stream (primarily compressed voice), at 9.6 Kb/s, was spread over a much wider transmission bandwidth by multiplying it with a PN (pseudo-noise) bit sequence with a clock, or chip, rate 128 times faster, or 2.2288 Mchips/sec. The PN sequences of different users were mutually orthogonal making detection possible by multiplication with the correct, synchronized PN sequence. With a very large number of available orthogonal PN sequences, increasing numbers of simultaneous users were possible, with conversations already in progress experiencing a little more background noise each time a new user was added. This accommodation of more users led to forecasts of huge gains in spectral efficiency. In practice, for comparable quality, IS-95 showed gains of six to ten times over 1G AMPS compared to IS-54's factor of three. There are additional levels of coding beyond the scope of this brief overview and additional measures were required to regulate signal levels to overcome the near-far effect, in which mobile units close to a base station overpower the signals from mobile units farther away, and to meet other special needs of CDMA. As an historical note, CDMA appears to have been described in the 1930s by a Russian researcher, Dmitri Ageev [44].

The channel between a transmitter and a receiver is typically not the ideal unobstructed line of sight, but rather a complex set of paths resulting from scattering, resulting in frequency-dependent fading. Figure 1.35 illustrates nonisotropic scattering in a city street between tall buildings. The phase differences between multipath components arriving at the receiver may, due to small variations in the path delays, result in constructive or destructive combination. Addressing the fading wireless channel is a major topic for the rest of this book.

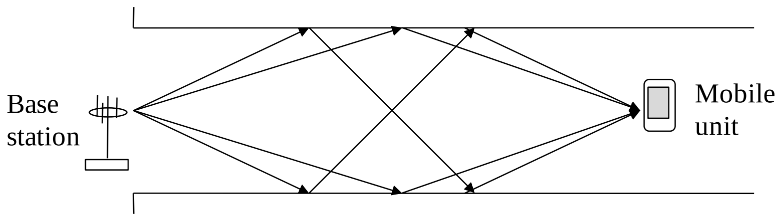


Fig. 1.35. Multipath propagation between rows of buildings [42].

The motivation for 3G cellular mobile was to move beyond the heavily voice-oriented architecture of 2G to provide at least equal emphasis on delivering data and media (audio and video) content. The IMT-2000 concept, with spectral allocations assigned by the 1992 WARC, had specific goals for transmission rates, including 144 Kb/s in vehicles, 384 Kb/s for pedestrians, and (asymmetric) 2 Mb/s downstream indoors. A second phase defined data rates as high as 20 Mb/s although in practice this was left to 4G systems.

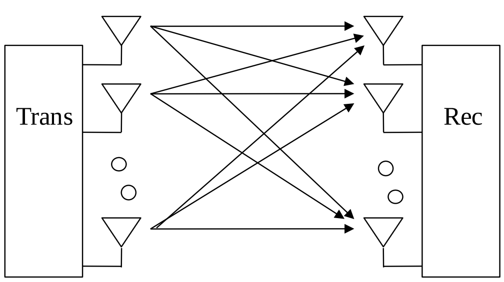
There are two major 3G standards: Universal Mobile Telecommunications System (UMTS), a successor to the 2G GSM promulgated by the 3GPP industry consortium, and CDMA2000 system, promoted by the 3GPP2 industry consortium, which is largely a successor to the IS-95 CDMA standard. Both of these standards operate in frequency bands from 1850-2200 MHz and utilize CDMA, but in different configurations. UMTS employs wideband CDMA (W-CDMA), with carrier spacing starting at 5 MHz and asynchronous operation of base stations, while CDMA2000 has carrier spacing starting at 1.25 MHz, the same as IS-95, and operates base stations synchronously. W-CDMA has implemented several upgrades with increasing speeds, notably High Speed Packet Access (HSPA) with 14.4 Mb/s downstream rate. CDMA2000 has evolved to similar rates.

Fourth generation (4G) cellular mobile employs a range of techniques to realize a large increase in capacity, with the goal of full accommodation of Internet applications migrating from computers to personal handsets. Peak data rate can reach 3 Gbps in the downlink and 1.5 Gbps in the uplink. Carriers have acquired or reassigned additional bandwidth for 4G services in segments between 700MHz to 2700 MHz. This is still within the familiar frequency range, so that propagation conditions are similar to those for the previous generations of cellular mobile systems.

The characteristics of 4G cellular mobile, defined in 3GPP's Long Term Evolution - Advanced (LTE-A) specifications, include use of OFDM with its resilience against multipath delay and pulse dispersion; carrier aggregation through use of multiple frequency channels for one communication session; multiple antennas in-multiple antennas out (MIMO) spatial multiplexing, a way to profit from relatively uncorrelated multipath propagation [45]); and cooperative multipoint (CoMP), the collaboration of multiple base stations in communicating with a mobile unit [46]. Relay nodes, essentially microcells embedded in ordinary cells near their boundaries to improve performance, are also supported. Either frequency-division duplex (FDD) or time-division duplex (TDD) operation can be used.

Carrier aggregation groups up to five component carriers, each with a bandwidth up to 20 MHz for a maximum of 100 MHz. The component carriers may come from different overlapping service cells offering different frequency channel sets, not the same as the spatially disjoint cells in the basic cellular mobile concept.

MIMO, a major enhancer of high speed digital wireless communication, provides the physical environment for spatial coding. For each transmitter-receiver pair, a transmitter uses multiple transmitting antennas to send appropriately designed signals to multiple receiving antennas, as illustrated in Figure 1.36. This figure also shows a possible channel variance matrix when the matrix channel contains two very small scatterers, two larger ones, and one large scattering cluster. LTE-A supports up to 8x8 MIMO in the downlink and 4x4 MIMO in the uplink. However, a range of available transmission modes supports a variety of different user equipments.



$$\begin{bmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{bmatrix}$$

Fig. 1.36. (a) MIMO with multiple antennas at transmitter and at the receiver. (b) A possible channel variance matrix for a 5x5 MIMO system [47].

Although 4G is still at the beginning of its deployed lifecycle, the relentless demand for greater bandwidth and flexibility drives efforts to devise technologies and techniques for reliable wireless access in much higher frequency bands, as described in the next subsection.

1.6.2 5G Cellular Mobile

There are many views of what 5G is, since, at the time of writing, it is still not defined in a standard, which 3GPP anticipates will be issued in 2020. There is agreement that the scope of 5G includes the smooth integration of different network types of wireless access and local-area networks and the use of new frequency allocations in much higher bands than those used for currently deployed cellular systems. As defined in METIS, a major European project, "the goal is a system that supports 1000 times higher mobile data volume per area (10 to 100 times higher number of connected devices and 10 to 100 times higher use data rate), 10 times longer battery life for low power massive machine communication (MMC), and 5 times reduced end-to-end latency, all of them at a similar cost and energy dissipation as today. " [48] In addition to going to frequency bands up to 86 GHz, the 5G system will exploit "new network topologies and technologies such as moving networks, multi-hop communications, self-configuration networks, and direct device-to-device communications."

Transmission studies for the development of 5G focus on understanding of the transmission characteristics of the higher frequency bands and the extremely high speed digital logic required to use them. As noted in [49], "spectrum at 38 GHz, 38 GHz and 70-80 GHz looks especially promising for next-generation cellular systems." Although oxygen in the air attenuates electromagnetic waves at 60 GHz, adjacent frequency bands experience relatively low attenuation, as illustrated in Figure 1.37. Rain also degrades propagation at higher frequencies, as much as 20 dB/km above 50 GHz for a heavy rain, compared with negligible attenuation in the 2G-4G frequency bands. However, high-speed wireless communication is becoming shorter range through deployment of microcells and picocells to reduce load on the macrocellular base stations, making the attenuation associated with higher frequencies a less significant limitation. The 60 MHz region is a serious candidate for LAN and small-cell communication because the high attenuation is actually a benefit in avoiding inter-cellular interference.

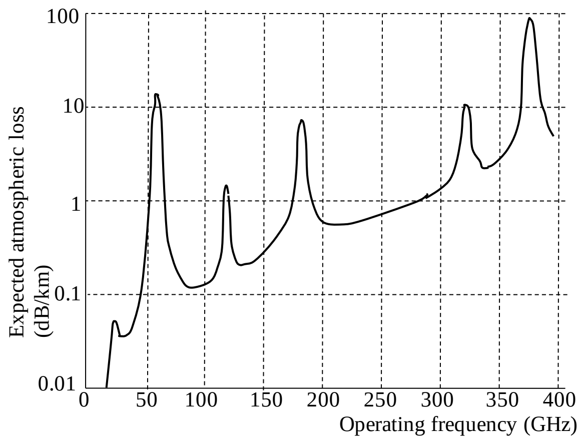


Fig. 1.37. Atmospheric path loss as a function of frequency [49].

In urban environments where 5G is expected to be heavily used, the millimeter-wave frequencies are characterized by long delay spreads from the multiple reflection paths. These long spreads can be partly avoided using directional beamforming, but designers must keep in mind the complexity tradeoff between beamforming (which reduces the need for equalization of long delay spreads) and channel equalization (which reduces the need for directional transmission).

Channel modeling, for simulation studies and system design, is an important part of current work on millimeter-wave communication. Several models are described in [48] and [49]. Ray tracing "uses computer simulation to model and discretize the energy radiated in space as it interacts with a computer model of the physical environment [49]." A model can include reflections, scattering, and atmospheric attenuation. The METIS map-based model, based on simplified ray tracing, begins with a simple geometric description of the propagation environment using geographic maps or 3D models of indoor environments. Figure 1.38 illustrates such a simplified geometric description for a rectangular city grid. Modeling also can be done through empirical measurements, such as channel soundings in selected environments such as city streets and university campuses. Analytical models form a third category, beginning with the simple log-distance path loss model

$$P_r(d) = P_t K_{fs} (d_0/d)^{\alpha} \tag{7}$$

for a propagation distance d, a close-in (but in the far field) free space path loss reference distance d_0, and a constant K_{fs} and path loss exponent α that are selected in accordance with empirical measurements [49].

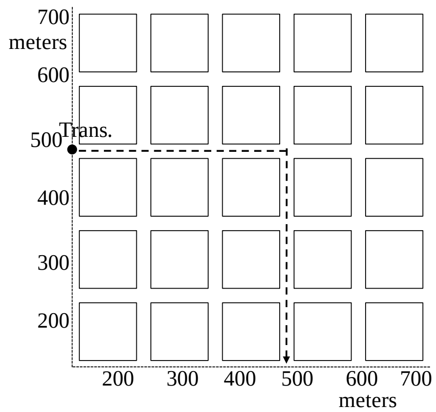


Fig. 1.38 Example of METIS simplified ray tracing model [48]

1.6.3 Wireless Transmission in the Cloud

Aside from transmission through literal clouds, the precipitation mentioned above, wireless communications is increasingly taking advantage of the economies and performance gains of shared distributed resources. The RAN (Radio Access Network) is an unseen but vital part of the wireless digital transmission system. Traditionally, baseband radio signal processing, mapping information streams from the wired backhaul network into radio signals destined for user handsets and the reverse for radio signals coming from the wireless handsets, occurs at the base station. These signal processing actions are computing tasks that might be handled more efficiently by shared computing resources located somewhere within the computing resources cloud. As noted in [50], "This transition from distributed to centralized infrastructure for baseband processing can have significant benefits: saving the operating expenses due to centralized maintenance; improving network performance due to advanced coordinated signal processing techniques; reducing energy expenditure by exploiting the load variations." This is the essence of the Cloud-RAN, a concept particularly applicable to the many very small base stations that are promising to become ubiquitous in 5G access networks.

Figure 1.39 illustrates the Cloud-RAN concept and further introduces the possibility of SDN (software defined network) control of network and service configuration and management. There are two clouds, one handling the immediate signal processing needs of a several-base station RAN, and the other the SDN functions that may take in multiple RANs and include, for example, association of base stations with RANs, managing handoff between RANs, performance monitoring, usage measurement for billing, distributing signal processing load among multiple servers within the wireless network cloud, and facilitating redundant functionality for robustness.

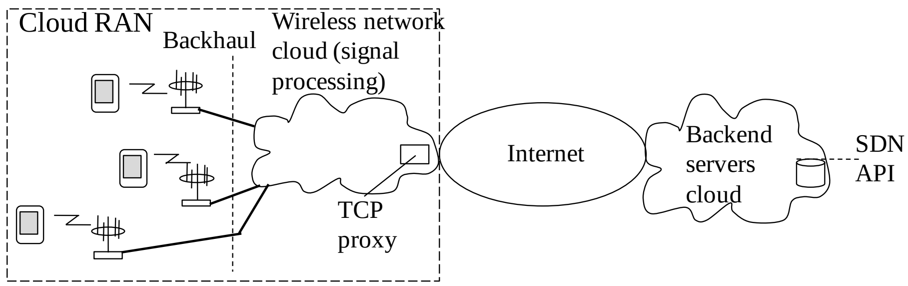


Fig. 1.39. The Cloud-RAN with signal processing in the wireless network cloud rather than at each base station, and software-defined network functions in a backend servers cloud (adapted from [50]).

The backhaul from base stations, which may only contain power amplifiers and antennas, is likely to be RoF (radio over fiber), carrying either analog or digitized RF (radio frequency) or intermediate frequency (IF) data streams. RoF backhaul transmission requires considerably larger capacity than in conventional backhaul networks where only user data are carried and all RF processing occurs in the base stations, but the cost savings from sharing signal processing functions in the wireless network cloud are likely to offset the increased cost of transmission.

1.6.4 IEEE 802.11Wireless LANs and Some Alternatives

IEEE 802.11 ("WiFi") networks are primarily one-hop local area systems in which all mobile units communicate directly with a single access point connected to an access network. But extensions through relay nodes are also possible, as Figure 1.40 illustrates, and such self-organizing networks can significantly extend the range and utilization rate of the access point. The relay nodes may be dedicated network extension nodes, or user devices with added relay capabilities. A repeater is a simpler range extension device, lying between an access point and client terminals, which simply retransmits everything it hears. This increases the useable distance between access points and client terminals but reduces the overall traffic capacity of the access node because the repeater retransmits data frames, duplicating traffic and possibly halving the total capacity [51]. A repeater also requires powering that a battery-powered relay node does not.

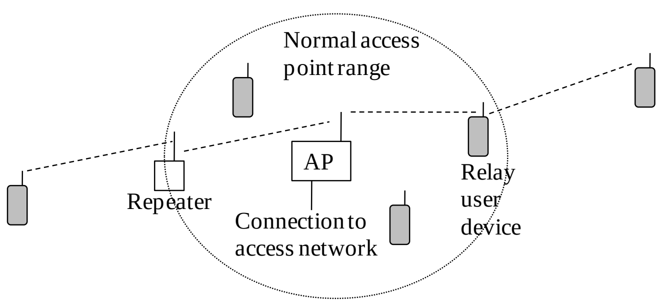


Fig. 1.40. An IEEE 802.11 network with relay and repeater extensions, adapted from [52].

IEEE 802.11 is, like Ethernet, a contention system with users competing for transmission time on each of the available carrier frequencies, primarily within the unlicensed (ISM and RF devices) portions of the 2.4 GHz and 5 GHz frequency bands [53]. The link-level protocol used to resolve the contention is carrier sense multiple access/collision avoidance (CSMA-CA) with an explicit acknowledgement and exponential backoff, differing from Ethernet's CSMA-CD by avoiding collisions rather than detecting them and trying again. A sender listens, backs off (delays) transmission according to the exponential protocol similar to that of Ethernet, transmits a frame, waits for an acknowledgement (ACK), and, if there is a timeout waiting for the ACK, goes again to the backoff procedure. Several generations of IEEE 802.11 standards, listed in Table 4, have steadily increased data rate.

Table 4: IEEE 802.11 major standards

Standard	Modulation	Band	Max. Data Rate
802.11b	Direct-sequence spread spectrum	2.4 GHz	11 Mbps
802.11a	OFDM/64QAM	5 GHz	54 Mbps
802.11g	OFDM/64QAM	2.4 GHz	54 Mbps
802.11n	MIMO/OFDM/64QAM	2.4 & 5 GHz	200 Mbps
802.11ac	multiuserMIMO/OFDM/256QAM	5 GHz	1300 Mbps

Propagation conditions determine range as in wireless access networks. Table 5 illustrates typical in-building losses over distance and through walls for IEEE 802.11g and 802.11n signals. Note that increasing distance and walls increases signal loss, but at a decreasing rate.

TABLE 5: Distance and absorption indoor losses for IEEE 802.11 signals, adapted from [54].

Distance (meters)	10	15	20	25	30
Number of walls	1	2	4	5	6
Absorption (dBm)					
802.11g	-10	-16	-20.8	-22.5	-24
802.11n	-9	-14	-25	-26.8	-28

IEEE 802.11 is generally not used for very short distances, e.g., for personal computer and smartphone peripherals such as earpieces, speakers, mice and keyboards. These and a range of rising body-area applications are more likely to use the very low powered Bluetooth [55] transmission system. Bluetooth employs a frequency-hopping system in short time slots are sent on different subchannels in a random hopping pattern.

Free-space infrared optical communication is another alternative for room-area communications, for applications as simple as television remote controls but going beyond to higher rates. Recent research suggests that OFDM, with signals modified to accommodate the non-negative nature of optical signals, might be a successful technique to combat the serious multipath interference seen in room-area environments [56].

1.7 ALL THE REST

The brief introduction in this chapter to digital transmission history and the major digital transmission areas sets the stage for the in-depth discussions of the rest of the book. To keep this broad overview to a reasonable size several topics, such as undersea transmission systems, were regrettably not included. We hope the reader will understand and find this book a helpful guide to what digital transmission is, how it works, and how it is likely to advance in the years to come.

REFERENCES

[1] G. J. Holzmann & B. Pehrson, The Early History of Data Networks, Wiley-IEEE Computer Society Press, 1994, ISBN: 978-0-8186-6782-4

[2] D.A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes", Proc. of the I.R.E., September 1952.

[3] G. Squier, "Multiplex Telephony and Telegraphy by Means of Electric Waves Guided by Wires", Trans. AIEE, vol. 30 pt. 2, 1911, pp. 1675–80.

[4] H.D. Luke, "The origins of the sampling theorem", IEEE Communications magazine, vol. 37 #4, April, 1999.

[5] A. H. Reeves, "The Past, Present, and Future of PCM", IEEE Spectrum, May 1965.

[6] M. Konate & P. Kabal, "Quantization Noise Estimation for Long-PCM", Proc. IEEE CCECE 2011, 24th. Annual Canadian Conference on Electrical & Computer Engineering. May 8-11, 2011, Niagara Falls, Ontario, Canada.

[7] D.R. Smith, Transmission Systems, 3rd edition, Springer, 2004, ISBN 978-1-4613-4726-2.

[8] R.J. Chapuis, 100 years of telephone switching (1878-1978), North Holland Publishing Company, 1982.

[9] R. Gitlin, J. Hayes & S. Weinstein, Data Communications Principles, Plenum, 1992, ISBN 0-306-43777-5

[10] AT&T Bell Labs staff, A History of Engineering and Science in the Bell System: Transmission Technology, AT&T Bell Labs, 1985, ISBN 0932764088.

[11] J. Hecht, City of Light: The Story of Fiber Optics, Oxford University Press, New York, 1999, ISBN 0-19-510818-3.

[12] S. Pachnicke, Fiber-Optic Transmission Networks, Springer-Verlag, Berlin, 2012.

[13] S. Jansen, I. Morita, T. Schenk & D. van den Borne, "Optical OFDM - A candidate for future long-haul optical transmission systems", Proc. IEEE OFC/NFOEC 2008.

[14] A.C. Clarke, 'Extra-Terrestrial Relays", Wireless World (U.K.), Oct. 1945.

[15] A Brief History of Communications, IEEE Communications Society, 2012, ISBN 0-7803-9825-4

[16] D. Martin, Communication Satellites, 4th edition, Aerospace Corporation, 2000, ISBN-13: 9781884989094.

[17] R. Gedney et.al., "Advanced Communications Technology Satellite (ACTS)", Acta Astronautica, Volume 26, Issue 11, November 1992

[18] S. B. Weinstein, Getting the Picture: A Guide to CATV and the New Electronic Media, IEEE Press, 1984, ISBN 0-87942-197-5.

[19] J. Riera et.al., "Ka-band propagation characterization: Madrid Experiment", Proc. IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting, Bilbao, Spain, 2009.

[20] S. Weinstein & V. Koll, "Simultaneous two-way data transmission over a two-wire circuit", IEEE Trans. on Comm., vol. COM-21 #1, Feb., 1973.

[21] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets", Part 1, Introduction, and Part 2, State of the Art, IEEE Commun. Mag., vol. 25 #2, Feb.1987.

[22] R. Lucky, "Automatic equalization for digital communications", Bell System Technical Journal, vol. 44, April, 1965.

[23] A. Lender, "Correlative level coding for binary data transmission", IEEE Spectrum, vol. 3, Feb., 1966.

[24] [O. Inoue, "Implementation in Japan (ISDN)", IEEE Commun. Mag., Aug. 1992, vol. 30 #8.

[25] A. Gelman et.al., "A store and forward architecture for video on demand service", Proc. IEEE ICC 91, Denver, June, 1991.

[26] S. Weinstein, "The History of OFDM", IEEE Communications Magazine, Nov., 2009.

[27] G. Ginis and J. Cioffi, “Vectored Transmission for Digital Subscriber Line Systems”, IEEE Journal on Selected Areas in Communications, vol. 20, no. 5, June, 2002.

[28] S. Weinstein, T. Wang and Y. Luo, The ComSoc Guide to Passive Optical Networks: Enhancing the Last Mile Access, Wiley-IEEE Press, 2011, ISBN: 978-0-470-16884-4.

[29] M. Schwartz, Computer-Communication Network Design and Analysis, Prentice-Hall, 1977, ISBN 0-13-165134-X.

[30] S.K. Schlar, Inside X.25: A Manager's Guide, McGraw-Hill Information, 1990, ISBN 0070553270.

[31] IETF RFC 3031, "Multiprotocol Label Switching Architecture", Jan. 2001, available at https://tools.ietf.org/html/rfc3031.

[32] M. de Prycker, Asynchronous Transfer Mode: Solution for Broadband ISDN, Prentice Hall, 1995, ISBN 0133421716.

[33] B.G. Lee & W. Kim, Integrated Broadband Networks: TCP/IP, ATM, SDH/SONET and WDM/Optics, Artech House, 2002, ISDN 1-58053-163-6.

[34] [I.432] CCITT recommendation I.432, "B-ISDN User-Network Interface - Physical Layer Specification", Geneva, 1991.

[35] CCITT recommendation I.321, "B-ISDN Protocol Reference Model and its Application", Geneva, 1991.

[36] N. Abramson, "The ALOHA System - Another alternative for computer communications", Proc. ACM 1970 Fall Joint Computer Conference, November, 1970.

[37] C.E. Spurgeon & J. Zimmerman, Ethernet: The Definitive Guide, 2nd Edition, O'Reilly books, ISBN-10 1449361846, 2014.

[38] J.A. Fleming, The Principles of Electric Wave Telegraphy and Telephony, Longmans, Green & Co., London, 1910, available at books.google.com.

[39] O. Vendik, "Significant contribution to the development of wireless communication by professor Alexander Popov", IEEE Commun. Mag., vol. 48 #10, October, 2010.

[40] J. S. Belrose, "Fessenden and Marconi: Their Differing Technologies and Transatlantic Experiments During the First Decade of this Century", IEEE International Conf. on 100 Years of Radio, London, 5-7 September 1995, available at www.ieee.ca/millennium/radio/radio_differences.html.

[41] M. Zimmerman & A.L. Kirsch, "The AN/GS-10 (KATHRYN) Variable Rate Data Modem for HF Radio", IEEE Trans. on Commun Technology, vol. 15 #2, Feb. 1967.

[42] G. Stuber, Principles of Mobile Communication, 2nd ed., Kluwer Academic Publishers, 2001, ISBN 0-7923-7998-5.

[43] A.J. Viterbi, CDMA: Principles of Spread Spectrum Communication (1st ed.). Prentice Hall, 1995, ISBN 0-201-63374-4.

[44] N. Tokareva, Bent Functions: Results and Applications to Cryptography, Academic Press, 2015, ebook ASIN: B014IXH3Y0.

[45] A. Paulraj, R. Naar & D. Gore, Introduction to Space-Time Wireless Communications, Cambridge Univ. Press, 2003, ISBN 0-521-82615-2.

[46] J. Wannstrom, "LTE-Advanced", 3GPP, June, 2013, available at www.3gpp.org/technologies/keywords-acronyms/97-lte-advanced.

[47] V. Veeravalli, Y. Liang & A. Sayeed, "Correlated MIMO wireless channels: Capacity, optical signaling, and asymptotics", IEEE Trans. on Infor. Theory, vol. 51 #6, June 2005.

[48] L. Raschkowski et.al., "METIS Channel Models", deliverable D.1.4, European Seventh Framework Programme, METIS (Mobile and wireless communications Enablers for the Twenty-twenty Information Society) project, April, 2015, available at https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.4_v2.pdf.

[49] T. Rappaport, R. Heath, R. Daniels & J. Murdock, Millimeter Wave Wireless Communication, Prentice-Hall, 2015, ISBN 0-13-217228-3.

[50] Y. Cai, F. Yu & S. Bu, "Cloud Radio Access Networks (C-RAN) in Mobile Cloud Computing Systems", Proc. 2014 IEEE INFOCOM Workshop on Mobile Cloud Computing.

[51] J. Geier, Designing and Deploying 802.11n Wireless Networks, Cisco Press, 2010, ISBN 1-58705-889-8.

[52] D. Zhao & T. Todd, "Real-time traffic support in relayed wireless access networks using IEEE 802.11", IEEE Wireless Communications, April, 2004.

[53] "FCC Online Table of Frequency Allocations", rev. Aug. 13, 2015, available at https://transition.fcc.gov/oet/spectrum/table/fcctable.pdf.

[54] S. Japertas et.al., "Research of IEEE 802.11 standard signal propagation features in multi partition indoors", IEEE Second Int. Conf. on Digital Information Processing & Communications, 2012.

[55] J-S. Lee, Y-W. Su & C-C. Shen, "A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee and Wi-Fi", IEEE Industrial Electronics Society 33rd Annual Conf., Taipei, Nov., 2007.

[56] D.J.F. Barros, S.K. Wilson & J.M. Kahn, "Comparison of Orthogonal Frequency-Division Multiplexing and Pulse-Amplitude Modulation in Indoor Optical wireless Links", IEEE Trans. on Commun., vol. 60 #1, Jan., 2012.