

# Chapter 4

## Information I: Communication, Transmission, and Information Theory

*The importance of information is directly proportional to its improbability.*

Jeremia Eugene Pournelle

*Intelligence is not information alone but also judgment, the manner in which information is collected and used.*

Carl Sagan

**Abstract** Information is a basic element of life and society involved in all areas of human, scientific, technological, economic, and developmental activity. Information storage, flow, and processing are inherent processes in nature and living organisms. Information transmission and communication/networking techniques contribute to the development of modern society, including social, economic, business, scientific, and technological operations and activities. This chapter covers at a conceptual level the following issues of information: definition, historical landmarks of its manifestations, communication models, modulation/demodulation, computer networks, multimedia, informatics/telematics, Shannon information entropy, source and channel coding/decoding, and theorems of information theory. The above sets of information/communication models, techniques, and technologies are affecting, and will continue to increasingly affect, the social, economic/business, and developmental activities of people in the short- and long-term future.

**Keywords** Information · Communication systems and models · Modulation Demodulation · Analog modulation/demodulation · Frequency modulation/demodulation · Phase modulation/demodulation · Pulse modulation/demodulation Information theory/theorems · Shannon’s entropy · Coding/decoding Source coding · Channel coding · Hamming distance · Convolutional codes Error detecting and correcting codes

### 4.1 Introduction

The term “*information*” is very diverse and is used in almost all areas of human scientific, technical, economic, and societal activity. In particular, information covers communication theory, information theory, information technology,

informatics, documentation science, and information systems that are the result of the interactions between technology and human processes.

Very broadly, the development of human communication and information dissemination has spanned three main periods:

- Speech and language development era.
- Development of writing era.
- Information revolution era.

Speech and language enabled humans to exchange knowledge with others who were physically at the same place. Writing has enabled communication between people at different places and times and initiated the development of civilization. The information revolution is still on-going and expanding with advancements in computer science and engineering, telecommunication systems, knowledge science, and computer networks, the biggest and most powerful of which is the Internet.

The above shows that “*information*” is indeed one of the “*basic pillars*” of human life and society, as will be addressed in this book. In the present chapter, we will guide the reader to the concept of information, including a listing of the key historical landmarks of its particular manifestations and branches. The core of the chapter is the tour to communication and information transmission theory, namely, communication models, modulation/demodulation, Shannon’s information entropy, source coding/channel coding, and theorems of information theory. The material in the chapter is not intended to constitute a complete treatment of the subject, but it is sufficient for the purposes of this book which principally aims at highlighting the impact of the “*information pillar*” on human life and society (Chap. 11). The fields of information documentation science, information science, information technology, and information systems will be discussed in the next chapter.

## 4.2 What Is Information?

According to the Oxford English Dictionary (1989), the two main contexts in which the term *information* is used are as follows:

- The act of molding the mind.
- The act of communicating knowledge.

The two processes, “*molding*” and “*communicating*”, are intrinsically related and, in most cases, occur inseparably (although not in a clearly known way).

Other frequently used interpretations of the term information include the following (Dictionary.com):

- Knowledge communicated or received concerning a particular fact or circumstance.
- Knowledge gained through study, communication, research, instruction, experience, etc.
- The act or fact of informing.
- An indication of the number of possible choices or messages.

- Important or useful facts obtained as output from a computer via the processing input data with a program.
- Data at any stage of processing (input, storage, output, transmission, etc.).
- A numerical measure of the uncertainty of an experimental outcome.
- The result of applying data processing to data, giving it context and meaning. Information can then be further processed to yield knowledge.
- Directory assistance.

The concept of information as “*knowledge communicated*” has a dominant position in modern human life and society and has been fully developed after World War II with the advancement and use of information science, informatics, computer science, computer networks, information theory, information technology, and information systems. From an epistemological viewpoint, the information concept has also found extensive and important use in biology, physics, psychology, etc.

The term *information* has a Latin origin (*informatio, informo*) [1, 2] and is composed by the prefix “*in*” and the word “*formatio*” or “*formo*” which has the meaning of giving a form to something. The prefix “*in*” here is used to make stronger the act of this form giving. There are two primary contexts in which the term *information* or *informo* has been used, the first is a *tangible (corporaliter)* context and the second an *intangible (incorporaliter)* concept. The tangible context refers to low-level processes, and the intangible context to high-level processes (moral, pedagogical, spiritual, mental). Capurro [1] has studied the Greek origins of the term *informatio* and its subsequent evolution. The Greek origin is evident in the works of Cicero (106–43 B.C.) and Augustine (354–430 A.D.) [1, 2].

The Greek term for information is *pliroforia* (πληροφορία) composed by the two words “*pliro*” (πλήρης = complete) and “*foria*” (φέρω = carry/bring). The word *pliroforia* indicates that the common meaning (sense) of a simple or complex symbol consisting of two or more subjects is completely carried. Conceptually, the term *pliroforia* (information) signals the content that is complete and clear (in whatever form). In computer science, the information is reflected in the qualitative value of the “*bit = binary digit*” 0 or 1 or the **quantum bit** (*qubit*) in quantum computers. The computer processes the data (sequences of 0s, 1s) and provides processed data. The human gives *meaning* to these processed data and converts them into “information”.

It must be remarked that today the word *information* is used in almost all scientific fields with various different meanings depending on the subject of each field and the processes studied therein (e.g., decrease in entropy, physical organization, a communication pattern, a form of feedback control, the meaning of a linguistic term, the probability of a signal transmitted over a communication channel, etc.). According to Bogdan [3]: “There seems to be no unique idea of information upon which these various concepts converge and hence no proprietary theory of information.”

The issue of whether the concept of “information” should include necessarily a human knower or an interpretative component or exclude mental processes and user-oriented intentions and address only an objective magnitude or property of human beings is of continuous concern by scientists and philosophers [4]. Several approaches that belong between these two extremes, including the need for a unified theory of information, have been proposed [5].

In [6], *Machlup* and *Mansfield* present their view that: “Information is addressed to human minds and is received by human minds.” All other senses are *metaphoric* and *anthropomorphic*. That is, the basic senses in information deal with “telling something or with the something that is being told.” Therefore, according to [6], the term information is not appropriate for use in the context of signal transmission. In overall, information is a human phenomenon that involves individuals transmitting and receiving messages in the framework of their possible decisions and actions.

In [7], *Capurro* defines *information* as an *anthropological class* referring to the phenomenon of human messages with vertical and horizontal structures related to the Greek concept of *message* (αγγελία: *aggelia* or μήνυμα: *menima*) and the philosophic discourse (λόγος: *logos*). The debate about the unification and naturalization of the term *information* goes back to Boltzmann, Neumann, Nyquist, Wiener, and Hartley.

In [8], *R.V.L. Hartley* states: “it is desirable to eliminate the psychological factors involved [in electrical transmission systems] and to establish a measure of information in terms of purely physical quantities” (This is because electrical transmission systems have to do not with humans but with machines).

According to *C.S. Pierce* (1839–1914) and *C.W. Morris* (1901–1979), the information transmitted through a communication channel between an emitter and a receiver involves three levels:

- Syntactic level.
- Semantic level.
- Pragmatic level.

At the *syntactic level*, the information deals with the formal bonds that exist between the various elements that make up the information, the rules of the communication code, the capacity of the channels, and the system design and coding methods for the transmission, processing, and storage of the information.

The *semantic level* is concerned with the ways of expressing the meaning of the information (e.g., written or nonwritten, cultural, societal, or moral rules or traditions valid in the human group or society concerned).

At the *pragmatic level*, the information is related to its utility. This level is determined to a large extent by the background of the person(s) who receive the information (political, social, economic, psychological, and moral factors).

The above three levels have a hierarchical structure in which the information can be managed (transmitted, processed, stored) at the syntactic level without the need to necessarily know its conceptual content at the semantic level or its practical utility at the pragmatic level.

Indeed, the word *information* in *Shannon’s* “Mathematical Theory of Communication” is used in a sense that must not be confused with *meaning* [9]. *Shannon* and *Weaver* state: “the semantic aspects of communication are irrelevant to the engineering aspects” [10]. Of course, as *Capuro* and *Hjorland* note [2]: “this does not mean that the engineering aspects are necessarily irrelevant to the semantic aspects.

The philosophical discussion about the concept of information was originated by *Norbert Wiener* who stated that “*Information is information, not matter or energy. No materialism which does not admit this can survive at the present day*” [2]. In the twentieth century, the notions of information and communication were applied at higher levels of abstraction and not to the communication of human knowledge as expressed in the above Wiener quotation. According to Titze [11], information is not a metaphysical principle but refers to a natural tendency for order and evolution. According to Stonier [12], structural and kinetic information is an intrinsic element of the universe, which is independent of the existence or not of an intelligent agent that can perceive it or not. Information may be manifested as “*infons*” which are comparable to “*photons*” [13]. Stonier distinguishes and separates the syntactic from the semantic features of information, and adopts the emergence of a global brain called “*noosphere*” by *Teilhard de Chardin* (from the Greek “*νοῦς*” = noos = mind, and *σφαίρα* = sphere = domain/globe) [14].

### 4.3 Historical Landmarks

The concept of information is generic and today spans all branches of science, technology, and human society that deal with the generation, acquisition, organization, storage, retrieval, interpretation, transformation, processing transmission, and utilization of information. Therefore, the *history of information* includes the evolution of computers, (theoretical) computer science, computation, information theory, information technology information science, information systems, multimedia, and informatics. Clearly, a detailed exposition of the combined history of all these parallel and overlapping branches, which in one or the other way involve the concept of information, needs too much space, and so here only the main historical landmarks will be given.

On the basis of the principal technology employed for the input, processing, output, and communication processes, the history of *information technology and systems* can be divided into the following four periods [15–17]:

- Pre-mechanical period (3000 B.C.–1450 A.D.).
- Mechanical period (1450–1840 A.D.).
- Electromechanical period (1840–1940).
- Electronic period (1940–Present).

#### 4.3.1 Pre-mechanical Period

Humans have been attempting to facilitate calculation using mechanical devices and to find ways to communicate for thousands of years. The communication of people using pictures and symbols (alphabet) was started by the Phoenicians

(around 3500 B.C.). Around 2900 B.C., the Egyptians develop *hieroglyphic writing*. Around 2000 B.C., the Greeks enhance the Phoenician alphabet through the addition of vowels. The Egyptians were writing on papyrus and around 2600 B.C. Chinese make paper from rags on which the modern paper making is based. Around 600 B.C., religious and other books were used, to permanently store information, made by folding papyrus vertically into leaves. Egyptians developed the first number system 1, 2, ..., 9 as vertical lines (with the number 10 as a U circle, etc.), a number system similar to the present was developed in India (around 10–200 A.D.), and the zero number was added in around 870 A.D. The ancient Greeks constructed some sophisticated analog computers such as the *Antikythera mechanism* (involving rusted metal gears and pointers) which has discovered in 1901 on the island of Antikythera [18]. Around 200 B.C., human messengers on foot or horseback started to be used in Egypt and China with messenger relay stations available. Fire signals were used on many occasions instead of human messengers.

### 4.3.2 Mechanical Period

Principal dates in this period are as follows:

- **10–20 A.D.:** Romans establish postal services and use mirrors for sending messages (heliographs).
- **100–200 A.D.:** First wooden printing presses used in China and the first bound books.
- **1450:** Movable metal-type printing process (Johannes Gutenberg, Germany).
- **1600:** Slide rule; an early form of analog computer (William Oughtred, England).
- **1641:** Blaise Pascal's adding machine
- **Late 1600s:** Gottfried Wilhelm Leibniz's adding machine.
- **1822:** Charles Babbage engines (difference engine, analytical engine).
- **1830–1840:** Parts and processes similar to modern computers (storage, punch card, binary logic, fixed program, real-time concept). It appears that the first programmer is *Ada Augusta Byron* (Countess of Lovelace), a friend of Babbage, who wrote a report on Babbage's machine. The name of the *Ada* programming language was chosen to honor her.

### 4.3.3 Electromechanical Period

This period was characterized by the conversion of information and knowledge into electric pulses and the rise of mathematics.

### 1. The start-up of telecommunications:

- **Late eighteenth century:** Voltaic battery.
- **Early 1800s:** Telegraph.
- **1835:** Samuel Morse develops the telegraphic Morse code.
- **1876:** Telephone (Alexander Graham Bell).
- **1894:** Invention and development of radio (Guglielmo Marconi).

### 2. Electromechanical computing

- **1880–1940:** Census tabulation machine (Herman Hollerith), early punch cards, punch card workers, IBM Mark 1.

### 3. Some landmarks of mathematics

- **1928:** The German mathematician David Hilbert poses the following questions: (a) Is mathematics complete (i.e., can every mathematical statement be either proved or disproved?), (b) Is mathematics consistent (i.e., is it actually true that statements like  $0 = 1$  cannot be proven by a valid method?), and (c) Is mathematics decidable (i.e., does there exist a mechanical way that can confirm the validity or not of a mathematical statement?).
- **1931:** The answers to two of Hilbert's questions were given by Kurt Gödel who proved that every sufficiently powerful formal system is either inconsistent or incomplete, and also that the consistence of a consistent axiom system cannot be proven within this system. The third question remains unanswered.
- **1936:** Alan Turing (1912–1954) gave an answer to the third question of Hilbert, via his formal model of a computer, known as the *Turing machine*. He showed that his machine would not be able to solve any problem (e.g., the question: given a Pascal program, does it halt on all inputs?—the so-called halting problem).

#### 4.3.4 Electronic Period

The general-purpose electronic digital computer was developed during World War II. One of the major needs in this period was the automatic calculation of ballistic equations.

- **1940:** At Iowa State University, an electronic computer was built for solving systems of linear equations (John Vincent Atanasoff and Clifford Berry).
- **1945:** Development of EDVAC (Electronic Discrete Variable Computer).
- **1946:** Based on the ideas of Atanasoff and Berry, the ENIAC (Electronic Numerical Integrator and Computer) system was built, originally intended for artillery calculations. This was the first fully working high-speed general-purpose computer using vacuum tubes.

- **1948:** Construction of Manchester Mark I (the first stored-program computer).
- **1949:** Construction of EDSAC (Electronic Delay Storage Automatic Calculator).
- **Late 1940:** UNIVAC (Universal Automatic Computer). This is the first general-purpose computer for commercial use.

We now give a brief description of the development of the four generations of digital computing and information processing.

**1950–1958: First generation.** Logic circuits which use vacuum tubes, punch cards for storage of external data, internal storage of data and programs using magnetic drums, machine language, assembly language, and the need for a compiler.

**1959–1963: Second generation.** Logic circuits using transistors designed on semiconductors, magnetic disks and tapes, magnetic cores, and high-level languages such as FORTRAN and COBOL.

**1964–1979: Third generation.** Integrated circuits (ICs) instead of individual transistors, magnetic tapes and disks, metal–oxide–semiconductor (MOS) memories, operating systems, and advanced languages, e.g., BASIC.

**1980–Present: Fourth generation.** Large-scale integrated (LSI) and very large-scale integrated (VLSI) circuits, central processing units (CPUs) on a single chip leading to the development of personal computers (PCs), e.g., Apple II (1977), Apple Mac (1984), IBM PC (1981), Microsoft Disk Operating System (MS-DOS), graphical user interfaces (GUIs) for PCs (1980), and MS Windows (version 3, 1990)).

### 4.3.5 *Information Theory Landmarks*

The field of *information theory* as we know it today was initiated by Shannon’s celebrated paper: “A Mathematical Theory of Communication” [9]. Shannon realized and adopted the need to have a communication theory, in which the communication signals must be utilized separately from the meaning of the messages they transmit. He also realized that a signal can be transmitted arbitrarily close to the theoretical channel capacity, even if the signal is contaminated by noise. These initial ideas have inspired and guided information, communication and computer engineers ever since. Information theory overlaps considerably with communication theory, but it is principally concerned with the basic and theoretical constraints on the processing and communication of information and not with the design and operation of individual components and communication devices. The principal historical landmarks of information theory are the following [9, 19–23]:



- **1929:** The publication of Leó Szilard on the decrease in entropy caused by the intervention of intelligent agents [21].
- **1948:** The introduction by Shannon of the entropy concept in information theory as an expected value that expresses the “information content” in a message (in units, such as bits) [9].
- **1949:** Publication of the book of Robert Fano concerning the transmission of information [23].
- **1956:** The publication of the book of Leon Brillouin about information theory and his ambitious attempt to incorporate all of scientific endeavor within the framework of Shannon’s information theory [22].
- **1957:** The paper of Edwin Jaynes on maximum entropy principle (MEP) [24].
- **1961:** Publication of the book of Myron Tribus in which he tried to formulate the laws of thermodynamics via information theory [25].
- **1988:** Publication of the work of George Saridis on the application of MEP methods [24, 26] to control [27, 28].
- **2006:** Publication of the book of George Klir [29] in which he develops a “generalized information theory” (GIT) by viewing uncertainty as a manifestation of some information deficiency, and information as the capacity to reduce uncertainty.
- **2007:** Publication of the book of Ben-Naim [30] on the benefits of the use of the concept (term) information or uncertainty, instead of the term entropy, in statistical mechanics and thermodynamics.

More detailed historical landmarks of information theory, especially regarding the development of error-correcting codes and lossless data compression, can be found in Wikipedia [31].

### 4.3.6 *Computer Networks, Multimedia, and Telematics Landmarks*

The development of computer networks and multimedia has, over the years, led to advanced and very accurate uses of transmitted information elements of any type for the benefit of present human society [32]:

**Computer networks** Before the widespread adoption of internetworking, which led to the Internet, the majority of communication networks were allowing communications only between the stations of the network. One of the dominating methods of computer networking was based on the central mainframe concept, in which its terminals were enabled to be connected via long-leased lines. This method was in use, for example, during the 1950s for research communication purposes between Pittsburgh (Pennsylvania) and Santa Monica (California). During the 1960s, a period in which the telephone network was the primary communication network in the world, many groups were working toward enhancing and implementing “packet switching”, especially in the defense field. The origin of the

Internet was the development of the *Advanced Research Project Agency Network (ARPANET)*, the first ARPANET link being realized between the University of California and Stanford (21 November 1969). ARPANET was enhanced with ideas drawn from the ALOHA net and grew rapidly. The ARPANET development was based on the Request for Comments (RFC) process, which continues to be used until this day.

Some of the primary landmarks in the history of computer networking are the following [33]:

**1971:** The *ALOHA net*, a type of TDMA transmission system and protocol for terrestrial and satellite random access communication (developed at the University of Hawaii) became operational in June 1971. ALOHA is a Hawaiian word (symbol) meaning hello, or goodbye or love and coming from “*Alo*” = presence/front/face and “*ha*” = breath.

**1974:** The *X.25 network* developed by the *International Telecommunication Union (ITU)* was used on the British *SERCnet* network of academic and research sites (later it became *JANET*). The first ITU standard on X.25 was approved in March 1976.

**1982:** The TCP/IP (Transmission Control Protocol and Internet Protocol) is established as the standard for ARPANET.

**1986:** TCP/IP is offered on workstations and PCs.

**1989:** The number of hosts exceeds 100,000.

**1990:** Several search tools (ARCHIE, Gopher, WAIS, etc.) are starting to enter the market after the official shut down of ARPANET.

**1991:** Development of the *World Wide Web (WWW)* by Jim Berners-Lee at CERN (European Center for Nuclear Research).

**1992–Present:** The WWW and Internet explode into the world [34].

**Multimedia** is the branch of computer science and information technology which deals with the computer-controlled integration of text, graphics, drawing, static and moving images (video), audio animation, and any other media suitable for representing, storing, transmitting, and digitally processing every kind of information. A multimedia application is any application that employs a collection of multiple media sources (e.g., texts, graphics, images, audio, animation, and/or video). Hypermedia is one of the multimedia applications [35].

The *multimedia* term is used in contrast to the term *media*, which indicates that only conventional types of printed or hand-produced material are used. The term *multimedia* was introduced by Bob Goldstein at Southampton, Long Island, in July 1966. Some landmark events in the area of multimedia are [36]:

- **1945:** Vannevar Bush publishes the first landmark paper that describes what amounts to a hypermedia system called “*Memex*”.
- **1960:** Ted Nelson coined the concept of “*hypertext*”.
- **1969:** Development of the hypertext editor at Brown (Nelson and Van Dam).
- **1985:** Negroponte and Wiesner formed the MIT Media Laboratory.
- **1989:** Tim Berners-Lee proposal for the World Wide Web to CERN.

- **1990:** Kristine Hooper Woolsey headed the Apple Multimedia Laboratory (with more than 100 scientists).
- **1992:** *JPEG* was adopted as the international standard for digital image compression. Also, the first *M-Bone audio* multicast on the Net was made.
- **1993:** Production of the first full-fledged browser, *Mosaic* of NCSA (The Illinois National Center for Supercomputing Applications).
- **1994:** Development of *Netscape* by Jim Clark and Marc Andreessen.
- **1995:** Introduction of JAVA for platform-independent application development (The first applet created is called Duke).
- **1996:** Introduction of Microsoft's *Internet Explorer*.

Further historical information pertaining to multimedia can be found in [37].

**Telematics** Telematics is the synergy of telecommunications and informatics (*telematics* = **tele** communications + inform**atics**). Therefore, as an overall term, “*telematics*” refers to the long-distance transmission and computer processing of information. The term “telematics” was coined in French by Simon Nora and Alain Minc as “*telematique*” (**tele** communication et **informatique**) in their 1978 report entitled “L’ Informatisation de la Société” [38]. This report was requested by the president of France *Valery Giscard d’ Estaing* in 1976 to explore how the computer and informatics applications could be extended to the organizational, economic, and developmental issues of modern society. In their report, Nora and Minc made the following remark: “Today, any consumer of electricity can instantly obtain the electric power needs without worrying about where it comes from or how much it costs. There is every reason to believe that the same will be true in the future of telematics.” Today, computers become smaller and smaller with ever lower energy requirements, and computing devices gradually become mobile and can accompany us wherever we go. These developments have led to the so-called mobile computing, ubiquitous computing, or pervasive computing. The historic landmarks of *telematics* involve the parallel and combined landmarks of computers and telecommunications. On the computer side, the two major landmarks are as follows [39, 40]:

- **1971:** Development of the first microprocessor, the Intel 4004.
- **1981:** Release of Osborne 1, the first portable computer.

On the telecommunications side, we mention the following:

- **1860:** Invention of the telephone by Antonio Meucci.
- **1895:** Invention of the wireless telegraph by Guglielmo Marconi.
- **1946:** AT&T introduces the first mobile telephone.
- **1958:** Launching of the first communication satellite, SCORE.
- **1969:** The ARPANET goes online and links for the first time two computers (University of California and Stanford).
- **1992:** The WWW is released by CERN worldwide.
- **2000:** The first commercial UMTS network is introduced in Japan.

We now present a summarized description of the basic concepts and methods of various manifestations of information that have been discussed in this section from a historical point of view, namely,

- Communication systems.
- Information theory.
- Computers and computer science.
- Informatics.
- Telematics.
- Multimedia.
- Information systems.
- Information science.

## 4.4 Communication Systems

### 4.4.1 General Issues

In communication systems theory, signals represent information. Information is used neatly packaged in analog or digital form. Communication systems are used for both manipulating and transmitting information. The following cases are the main types:

- *Point-to-point communication* (from a single place to another place).
- *Broadcast communication* (from one place to many other places).
- *Multipoint to multipoint communication* (teleconferencing, chat rooms).

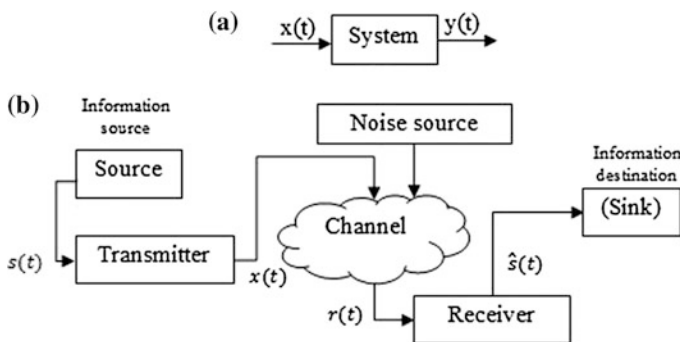
Communication systems can be *analog* (e.g., via radio waves) or *digital* (e.g., computer networks). The question arising here is: Which is better, analog or digital communication? This question has been answered by Shannon, in his work on information theory [9], who suggests the use of digital representation of signals and the digital communication strategy. This means that all information-bearing signals are converted into digital ones (discrete-time, amplitude-quantized signals) applying the sampling theorem which shows the condition that must be met for this conversion to be an accurate one. As mentioned in Sect. 4.3.5. Shannon has shown that in digital form, a properly designed system can communicate digital information without error despite the presence of communication noise in all transmission channels. This result has a fundamental importance and value not only in communications, but also in all areas where digital information is handled, e.g., compact discs (CDs) and digital video disks (DVDs).

### 4.4.2 Shannon–Weaver Communication Model

One of the fundamental concepts in communication and information theory is the “communication model” of Shannon and Weaver [10], which is illustrated in Fig. 4.1b. Figure 4.1a shows the graphical representation (called *block diagram*) of a system operating on a receiving signal  $x(t)$  and producing an overall output signal  $y(t)$ . In Fig. 4.1b, we have the following components (blocks):

- The *information source* that selects a desired signal (message)  $s(t)$ , out of a set of possible messages, which is to be sent to the destination (sink). The message can have several forms, e.g., speech, music, strings of letters as in telegraph or teletype, or characters typed on the keyboard of a PC, etc.
- The *transmitter* that receives the message  $s(t)$  and produces a signal  $x(t)$  suitable for transmission over the channel. The signal  $x(t)$  is either modulated or encoded, depending on the message’s physical nature. In telephony, the transmitter’s operation consists of changing sound pressure into electrical current. In telegraphy, an encoding operation takes place that generates a sequence of dots, dashes, and spaces corresponding to the message. In a multiplexed, pulse-coded, modulation system, the various speech functions are sampled, compressed, quantized, and encoded, and finally properly interleaved to construct the signal  $x(t)$ .
- The *channel* is the medium over which the signal is transmitted to the receiver. In the channel, the transmitted signal is typically corrupted by noise or distorted or attenuated by various phenomena [giving the corrupted message  $r(t)$ ].
- The *receiver* is a sort of inverse transmitter, changing the transmitted signal back into a received message  $\hat{s}(t)$  that must resemble  $s(t)$  as much as possible.
- The *destination* (or information sink) that is the person (or thing) for whom the message is intended (and who uses it for the desired purpose).

The “noise” contaminating the transmitted message  $x(t)$  may be *internal* (i.e., coming from the receiver’s attitudes, beliefs, or knowledge) or *external* (i.e., caused



**Fig. 4.1** a Block diagram of system receiving an input signal  $x(t)$  and producing an output signal  $y(t)$ , b Shannon–Weaver communication model

by other sources). This internal or external noise may either *strengthen* the intended outcome of the messages (if the information confirms the message) or *weaken* the intended outcome (the information in the noise disconfirms the original message).

The implementation of the *Shannon–Weaver communication model* needs the following:

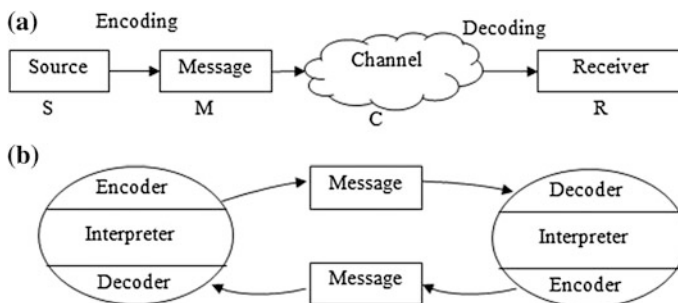
- To understand signals and signal generation.
- To understand the nature of the information these signals represent.
- To know how information is transformed between analog and digital forms.
- To know how information can be processed by systems working on information-bearing signals.

These requirements are met through *electrical engineering* (which studies the ways signals are represented and manipulated electrically/electronically) and *signal engineering* (which studies the structure of signals, their information content, and the capabilities that this structure imposes upon communication systems), independently of what the signal sources are.

#### 4.4.3 Other Communication Models

The Shannon–Weaver linear communication model is applicable to pure machine communication, i.e., it is not intended to match human communication. This model is focused on the technical problems associated with the selection and arrangement of discrete units of information and not with the semantic “content” of the messages. Weaver himself pointed out that: “*It is surprising but true that, from the present view point, two messages, one heavily loaded with meaning and the other pure nonsense, can be equivalent as regards information.*”

Two other communication models devised for human communication are the *Berlo model* of communication [43] and the *Schramm model* of communication [44], shown in Fig. 4.2a, b [45]:



**Fig. 4.2** **a** Berlo’s SMCR model: a source encodes a message for a channel to a receiver who decodes the message, **b** Schramm’s communication model

Berlo's model is an enhanced adaptation of the Shannon–Weaver model so that, to cover the human communication, it provides a mechanism in the source for the following:

- Communication skills.
- Attitudes.
- Knowledge.
- Social and cultural issues.

In other words, the source is sufficiently flexible to include oral, written, electronic, or any other form of “symbolic” generator of messages. The “*message*” is considered to be the core element of the model, stressing the transmission of ideas, specifically content, elements, treatment, structure, and code. The receiver is recognized to be very important to communication, since it is actually the target. Like the source, the receiver takes into account and interprets communication skills, attitudes, knowledge, and social and cultural issues. The channel, i.e., the communication medium, involves all the human senses (hearing, seeing, touching, smelling, and tasting). The concepts of “encoding” and “decoding” emphasize the psychophysical problem every person has in translating his/her own thoughts in words or symbols and providing them in an understandable way to other people. Here, it is tacitly assumed that human communication is similar to machine communication (e.g., telephone signal, television signal, and computer information transmission, etc.). Clearly, the accuracy of human communication using this model depends on choosing the “proper” symbols, preventing interference, and sending efficient messages. However, even if the proper symbols are used, people may misunderstand each other. But all these issues are human-centered depending on agreements, beliefs, shared values, and attitudes.

The Schramm's model provides the additional features of “*feedback field expertise*” and “*psychological reference frame*” for the interacting humans. It is noted that Schramm's model is less linear than the Shannon–Weaver or the Berlo's model. But again, it is suitable only for bilateral communication between two partners (i.e., it does not cover the case of multilevel communication). Three nonlinear models of communication are the following:

- **Dance's helical spiral (1967)**: This spiral depicts communication as a dynamic process, where the helix represents how communication evolves in a human from his/her birth up to the present [41].
- **Westley and MacLean's conceptual model (1957)**: This model is based on the fact that the communication of a person begins when he/she starts to respond selectively to the immediate surroundings and not when he/she starts to talk [41].
- **Becker's mosaic model (1968)**: Here, it is assumed that most communicative acts link message elements from several social situations (not just from one situation). These complex communicative events are linked to the activity of a receiver who moves through a permanently varying cube or mosaic of information [41]. This model adds a third dimension, but human communication

involves many more dimensions. Therefore, many researchers attempted to develop multidimensional communication models. Two such models are the following [41]:

- **Ruesch and Bateson functional model (1951)**: This involves four levels of analysis, i.e., *level 1* (intrapersonal process), *level 2* (interpersonal), *level 3* (group interaction), and *level 4* (cultural level).
- **Barnlund's transactional model (1970)**: A very systematic functional model where the key assumptions on which it is based are shown explicitly. Its most important property is that it includes no simple or linear directionality in the interplay between itself and the surrounding world.

#### 4.4.4 Transmitter–Receiver Operations

We now briefly discuss the operations of the *transmitter* and *receiver* in the Shannon–Weaver communication model (Fig. 4.1b).

**Modulation** The transmitter performs the *modulation* operation, namely, the superposition of the information (message) onto an electronic or optical carrier signal. Modulation can be applied to direct current (mainly by turning it on or off), alternating current, and optical signals. The *Morse code*, used in telegraphy and presently in amateur radio, uses a *binary* (i.e., two state) digital code similar to the code used in modern computers. In today's typical radio and telecommunication systems, the carrier is an alternating current (AC) sent over a given frequency band. Standard modulation methods are as follows:

- **AM (Amplitude modulation)**, where the voltage superimposed on the carrier modulation signal is time varying.
- **FM (Frequency modulation)**, where the frequency of the carrier waveform is varied (modulated) in suitably small amounts.
- **PM (Phase modulation)**, where the natural flow of the alternating signal is delayed temporarily.

The above are called *continuous (analog) signal modulation* methods to discriminate them from **PCM (Pulse-Code Modulation)** which is employed to encode analog and digital signals in a binary (digital) way. In general, the modulation techniques can be classified as follows:

- Analog versus digital modulation.
- Baseband (low pass) versus bandpass (passband) modulation.
- Binary versus M-ary modulation.
- Linear versus nonlinear modulation.
- Memoryless modulation versus modulation with memory.
- Power-efficient versus bandwidth-efficient modulation.
- Constant envelope versus nonconstant envelope modulation.

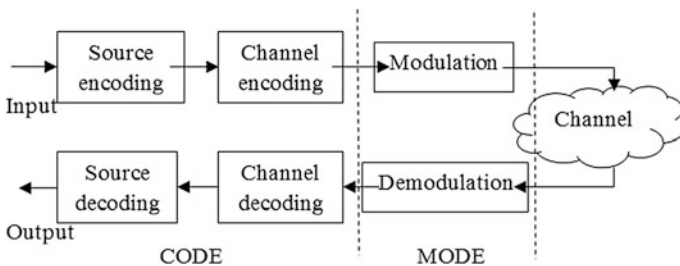


- Equivalent digital modulation methods (amplitude-shift keying: ASK, frequency-shift keying: FSK, phase-shift keying: PSK). Two-way radios employ FM although some use *single sideband* (SSB). A combination of ASK and PSK is the *quadrature-amplitude modulation* (QAM).

**Demodulation** This is the inverse operation of modulation, which extracts the original information-bearing signal from the modulated carrier signal. An electronic device used to perform the recovery of the original message from the modulated carrier signal is called a *demodulator*. To each modulation technique, there corresponds an appropriate demodulation technique and an associated demodulator. For example, in an AM signal (which encodes the message into the carrier wave by varying its amplitude in proportion to the analog message sent), we have two kinds of demodulators. These are the *envelope detector* (which uses a *rectifier* allowing the current to flow only in one direction) and the *product detector*. The latter multiplies the incoming AM-modulated signal by the signal of a local oscillator with the same frequency and phase as the carrier used. After filtering, the original message is obtained. For FM, many typical forms of demodulators exist. For example, a *quadrature detector* phase shifts the signal 90° and multiplies it with the unshifted version. Among the terms produced by this operation is the original message which is selected and amplified. Another FM demodulator employs two AM demodulators, one tuned to the high end of the frequency band and the other to the lower end. The two outputs are then passed to a difference amplifier.

A pair of *transmitter* (coder, modulator) and *receiver* (decoder, demodulator) is called a *transceiver*. The general structure of modern communication systems involving a *CODEC* (Coder/Decoder) and *MODEM* (Modulator/Demodulator) has the form shown in Fig. 4.3.

**Multiplexing** To convey more information in a given amount of time, the bandwidth of a signal carrier is divided such that more than one modulated messages can be sent simultaneously on the same carrier. This is called *multiplexing*, where the carrier is sometimes called the *channel* and its separate message signal carried is referred to as *subchannel*. The multiplexer is the device that puts the individual signals onto the carrier and takes off received transmissions—separately. Typical forms of multiplexing are as follows:



**Fig. 4.3** General structure of a digital communication system

**FDM:** Frequency-division multiplexing.

**TDM:** Time-division multiplexing.

FDM which divides the principal frequency of the carrier into separate sub-channels is commonly employed in analog communication systems, whereas TDM, which divides the principal signal into time slots, each one carrying a separate signal, is used in digital communication systems. A digital cellular phone technology based on *TDMA (time-division multiple access)*, which was developed in the 1980s and is predominant in Europe, is the *Global System for Mobile (GSM)* communications. GSM is also used worldwide. GSM operates in the 900 MHz and 1.8 GHz bands in Europe, and the 1.9 GHz PCS band in the United States. It is based on a circuit-switched system, which divides each 200 kHz channel into eight 25 kHz time slots.

#### 4.4.5 Analysis of Analog Modulation–Demodulation

Here, we outline the mathematical analysis of analog modulation. The carrier wave is a sinusoid of the following form:

$$c(t) = A \cos(\omega_c t + \phi)$$

where the *amplitude*  $A$ , the *carrier cyclic frequency*  $\omega_c = 2\pi f_c$  ( $f_c$  is the frequency), and the *phase*  $\phi$  are the parameters that can be varied according to the message-bearing signal:

$$m(t) = M \cos(\omega_m t + \theta_0)$$

where  $\theta_0$  is a constant and, without loss of generality, it can be assumed  $\theta_0 = 0$ .

##### 4.4.5.1 Amplitude Modulation

In amplitude modulation (AM), the amplitude  $A$  of the carrier signal is modulated in proportion to the message-bearing (low frequency) signal  $m(t)$ , to give

$$\begin{aligned} x(t) &= (A + M \cos(\omega_m t)) \cos(\omega_c t + \phi) \\ &= A(1 + \mu \cos \omega_m t) \cos(\omega_c t + \phi) \end{aligned}$$

where

$$\mu = M/A$$

is called the “*modulation index*” of AM, and for demodulation purposes (i.e., for the recovery of  $m(t)$  from the above modulated carrier signal) is typically selected less than one (i.e.,  $M < A$ ).

Using the well-known trigonometric identity:

$$\cos(\theta) \cos(\psi) = (1/2)[\cos(\theta + \psi) + \cos(\theta - \psi)],$$

the amplitude modulated signal  $x(t)$  can be written as

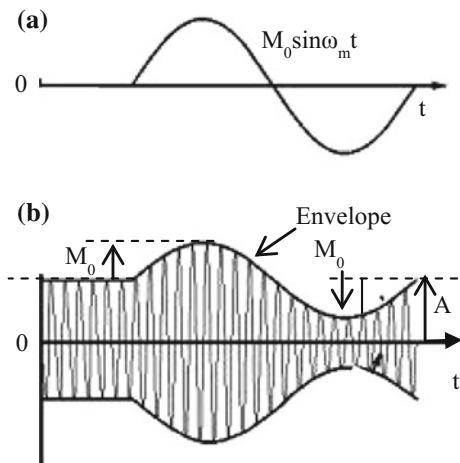
$$x(t) = A \cos(\omega_c t + \phi) + \frac{A\mu}{2} [\cos(\omega_c + \omega_m)t + \phi] + \frac{A\mu}{2} \cos[(\omega_c - \omega_m)t + \phi]$$

This shows that  $x(t)$  has three additive components with frequencies  $\omega_c$ ,  $\omega_c + \omega_m$  and  $\omega_c - \omega_m$ . The frequency  $\omega_c + \omega_m$  is called the *upper side frequency*, and  $\omega_c - \omega_m$  is called the *lower side frequency*. A typical AM signal with  $\mu_0 = 0.8$  is shown in Fig. 4.4.

In the type of AM described above, both side frequencies  $\omega_c - \omega_m$  and  $\omega_c + \omega_m$  are transmitted. It is therefore called *Double Sideband AM* (DSB-AM). But for more efficiency, only one of the sidebands must be transmitted, in which case we have the so-called *Single Sideband AM* (SSB-AM). In particular, SSB-AM is called *Lower Sideband AM* (LSB-AM) or *Upper Sideband AM* (USB-AM), if only the lower frequency or the upper side frequency is transmitted, respectively. The case where the modulation index is  $\mu > 1$  is called overmodulation. Now, let us calculate the power of a modulated wave. The power is proportional to the square of the voltage (amplitude). Therefore, the power of the carrier wave is

$$\text{Carrier power} = KA^2$$

**Fig. 4.4** Typical amplitude signal with  $\mu_0 = 0.8$



where the factor  $\frac{1}{2}$ , required because  $A$  is the amplitude (peak value) of  $c(t)$ , is included in the proportionality constant  $K$ .

The total sideband power is equal to

$$\begin{aligned} \text{Total side - band power} &= KA^2 \left(\frac{\mu}{2}\right)^2 + KA^2 \left(\frac{\mu}{2}\right)^2 \\ &= \frac{\mu^2}{2} \times \text{Carrier power} \end{aligned}$$

Therefore, the total power of the AM signal is equal to

$$P = KA^2(1 + \mu^2/2)$$

i.e., equal to the sum of the constant carrier power  $KA^2$  and the power of the two sidebands, which depends on the value of  $\mu$ . For example, if  $\mu = 1$ ,  $P = KA^2 \times (3/2) = KA^2 \times 150\%$ .

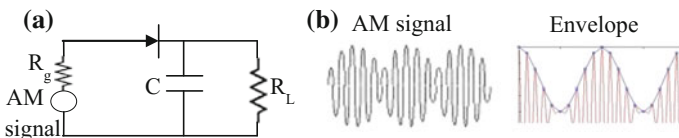
#### 4.4.5.2 Amplitude Demodulation

As mentioned in Sect. 4.4.4, for AM we have two kinds of demodulators (detectors): the *envelope demodulator* and the *product demodulator*. The signal

$$v(t) = A(1 + m(t))$$

is called the *envelope* of the AM signal. Clearly, if the envelope  $v(t)$  is extracted, the transmitted message  $m(t)$  can be recovered. A simple way to obtain the envelope is to use the “*envelope demodulator circuit*” shown in Fig. 4.5.

The envelope detector is of the so-called *noncoherent* (asynchronous) type. Better performance can be obtained by *coherent* (synchronous) detector in which both the frequency and the phase of the carrier are known at the demodulator. The amplitude of the carrier affects only the level of the demodulated signal, which can be changed by a simple amplifier. Therefore, the amplitude of the carrier is not important in the demodulation. The carrier signal is restored at the receiver by a circuit called the *carrier-recovery circuit*. In the *product detector*, the amplitude



**Fig. 4.5** **a** Envelope AM demodulator. **b** The AM signal passing through the demodulator provides the envelope signal. This is done by the capacitor which removes the carrier frequency and leaves only the envelope

modulated signal is multiplied by the signal provided by a local oscillator. The simplest case is obtained if the local oscillator has the same frequency as the carrier, i.e.,  $\omega_c$ . Then, the result of the multiplication is the sum of the original message and another AM signal at frequency  $2\omega_c$ . A low-pass filter is used to suppress this second component. The block diagram of the product detector is shown in Fig. 4.6.

In the above block diagram, we have

$$\begin{aligned} x(t) c(t) &= A(1 + m(t)) \cos(\omega_c t) \cos(\omega_c t) \\ &= \frac{1}{2} A(1 + m(t))(1 + \cos(2\omega_c t)) \end{aligned}$$

where, without loss of generality, the phase  $\phi$  of the carrier was omitted and use was made of the same trigonometric identity as in the AM modulation. After low-pass filtering, the original message is recovered.

The carrier signal can be available to the product detector in two ways:

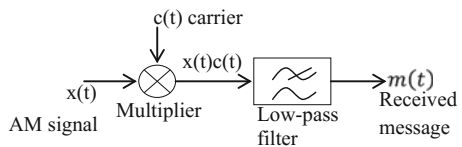
- Through transmission of the carrier.
- Through recovering of the carrier.

The carrier recovery can be achieved by using a transmitted pilot signal outside the passband of the modulated signal as shown in Fig. 4.7.

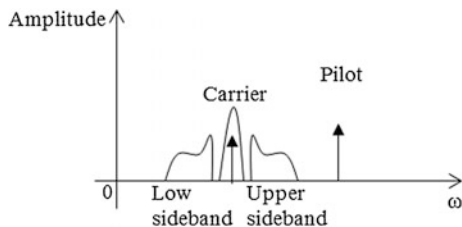
To recover the carrier signal, two techniques can be applied: (i) recovery by a bandpass filter (Fig. 4.8a), and (ii) recovery by a phase-locked loop—PLL (Fig. 4.8b).

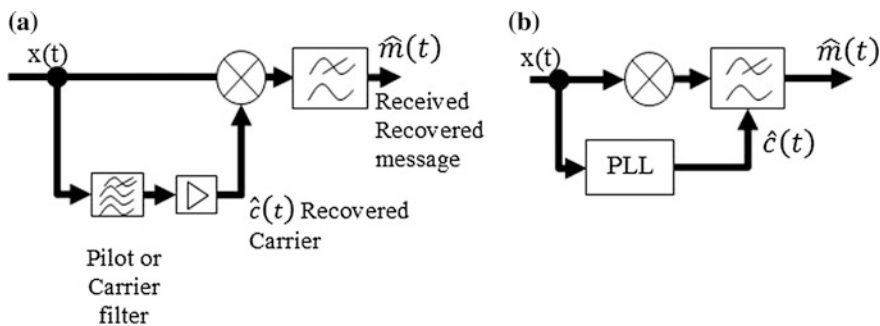
It is noted that, with the product demodulator, we can also decode overmodulated AM, SSB, and AM with suppressed carrier, in addition to the standard DSB-AM.

**Fig. 4.6** Block diagram of a product detector



**Fig. 4.7** Pilot signal outside the AM signal





**Fig. 4.8** **a** Recovery of carrier by a bandpass filter. **b** Recovery of carrier by a phase-locked loop (PLL). Here, we use the symbols  $\hat{c}(t)$  and  $\hat{m}(t)$  to indicate approximate copies of  $c(t)$  and  $m(t)$

### 4.4.5.3 Frequency Modulation

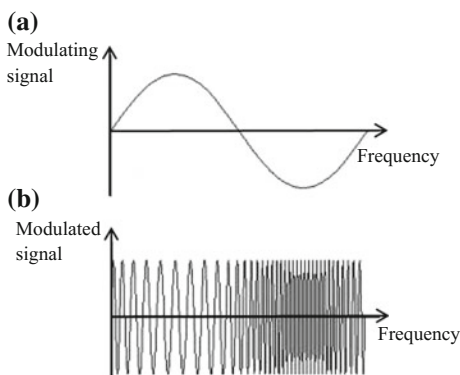
The principal advantages of *frequency modulation* (FM) over AM are as follows: (i) better signal-to-noise ratio, (ii) less interference effects between neighboring stations, and (iii) less radiated power. The drawbacks are as follows: (i) It requires much more bandwidth than AM (up to 20-times more), and (ii) The transmitter and receiver devices are more complex. In FM, the frequency of the carrier wave is changed in proportion to the message signal  $m(t)$ , i.e., we have

$$\omega(t) = \omega_c(1 + \mu \cos(\omega_m t))$$

where  $f_c = \omega_c / 2\pi$  is called the *center frequency*, and  $\mu$  is the degree of modulation. A frequency-modulated signal has the form shown in Fig. 4.9b.

Frequency modulation is popular in a variety of radio transmission applications from broadcasting to general point-to-point transmissions. It is also widely employed for broadcasting via very high frequency (VHF) because it provides a very good medium for high-quality transmissions. FM is also largely used in

**Fig. 4.9** Frequency modulation. **a** The modulating signal. **b** The modulated signal



several mobile applications. Here, because  $\omega$  is time varying, the FM signal is given by

$$\begin{aligned} A \cos \left[ \int_0^t \omega(t) dt \right] &= A \cos \left[ \int \omega_c (1 + \mu \cos(\omega_m t)) dt \right] \\ &= A \cos \left[ \frac{\omega_c \mu}{\omega_m} \sin(\omega_m t) + \omega_c t + \theta_0 \right] \end{aligned}$$

where the constant of integration  $\theta_0$  can be neglected as a constant angle. The quantity  $\Delta\omega = \omega_c \mu$  is called the *cyclic frequency deviation*.

The parameter

$$\frac{\omega_c \mu}{\omega_m} = \frac{\Delta\omega}{\omega_m} = \mu_f$$

is called the *FM modulation index* (or *deviation ratio*) and has a value in radians that is different for each value of  $\omega_m$ . On the basis of this, the expression for an FM signal becomes

$$x(t) = A \cos(\omega_c t + \mu_f \sin(\omega_m t))$$

Expanding this signal using and the identity for  $\cos(\theta + \psi)$ , we obtain

$$\begin{aligned} x(t) &= A \cos(\omega_c t + \mu_f \sin(\omega_m t)) \\ &= \cos(\omega_c t) \cos(\mu_f \sin(\omega_m t)) - \sin(\omega_c t) \sin(\mu_f \sin(\omega_m t)) \end{aligned}$$

The two complex functions  $\cos(\mu_f \sin(\omega_m t))$  and  $\sin(\mu_f \sin(\omega_m t))$  can be expressed as an infinite series of Bessel functions, namely [46, 47],

$$\begin{aligned} \cos(\mu_f \sin(\omega_m t)) &= J_0(\mu_f) + 2J_2(\mu_f) \cos(2\omega_m t) \\ &\quad + 2J_4(\mu_f) \cos(4\omega_m t) + 2J_6(\mu_f) \cos(6\omega_m t) + \dots \\ \sin(\mu_f \sin(\omega_m t)) &= 2J_1(\mu_f) \sin(\omega_m t) + 2J_3(\mu_f) \sin(3\omega_m t) \\ &\quad + 2J_5(\mu_f) \sin(5\omega_m t) + 2J_7(\mu_f) \sin(7\omega_m t) + \dots \end{aligned}$$

where the Bessel functions  $J_k$  ( $i = 1, 2, \dots$ ) of the first kind are defined as

$$J_k(\mu_f) = \sum_{q=0}^{\infty} \frac{(-1)^q}{q!(q+k)!} \left(\frac{\mu_f}{2}\right)^{k+2q}$$

and are provided in corresponding mathematical tables and computer-mathematics libraries. Using the trigonometric identities for  $\cos(\theta) \cos(\psi)$  and  $\sin(\theta) \sin(\psi)$ , we get

$$\begin{aligned}
 x(t) = A \{ & J_0(\mu_f) \cos(\omega_c t) - J_1(\mu_f) [\cos(\omega_c - \omega_m t) - \cos(\omega_c + \omega_m) t] \\
 & + J_2(\mu_f) [\cos((\omega_c - 2\omega_m) t) + \cos((\omega_c + 2\omega_m) t)] \\
 & + J_3(\mu_f) [\cos((\omega_c - 3\omega_m) t) - \cos((\omega_c + 3\omega_m) t)] \\
 & + J_4(\mu_f) [\cos((\omega_c - 4\omega_m) t) + \cos((\omega_c + 4\omega_m) t)] - \dots
 \end{aligned}$$

We see that the FM signal involves the center frequency  $\omega_c$  and an infinite number of side frequency pairs, each pair spaced by an amount equal to the modulating frequency  $\omega_m$ .

The amplitude of the successive side frequency pairs is determined by the Bessel function coefficients. Although theoretically the FM signal covers the entire frequency spectrum with sidebands, the  $J$  coefficients decrease relatively quickly and the series converges rapidly. So, the bandwidth actually required is finite. Because  $\omega_m$  and  $\mu_f$  are inversely proportional ( $\mu_f = \Delta\omega / \omega_m$ ), a small  $\omega_m$  will result in more side frequencies of significant value, than those obtained in the case of a high  $\omega_m$  as shown in Fig. 4.10.

It is noted that  $J_0(\mu_f) = 0$  for  $\mu_f = 2.405, 5.520, 8.654, 11.79, 14.93$ , etc. Therefore, for these values of  $\mu_f$ , the center frequency component is zero (i.e., it does not exist).

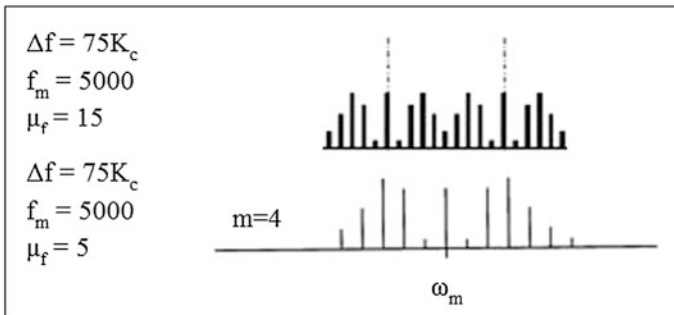
A practical rule of thumb is *Carson's rule* which says that almost all ( $\sim 98\%$ ) of the power of an FM signal is contained in a bandwidth  $B_c$  equal to

$$B_c = 2(f_m + \Delta f)$$

where  $\Delta f = \mu_f f_c$  is the frequency deviation  $\Delta\omega / 2\pi$  from the central frequency.

#### 4.4.5.4 Frequency Demodulation

In FM, it happens that signals with a large frequency deviation are supporting higher quality transmissions at the expense of occupying a wider bandwidth. Thus,



**Fig. 4.10** Frequency spectra of two FM signals with  $\Delta f = 75K_c/s$



in practice, several levels of frequency deviations (or modulation index values) are employed, according to the application that is used. Those with low deviation are known as *narrowband frequency modulation* (NBFM), with typical values  $\pm 3$  kHz. NBFM is generally used for point-to-point communications. Higher frequency deviations are required for broadcasting and are called *wideband frequency modulation* (WBFM). Typical values of WBFM frequency deviation are  $\pm 75$  kHz.

To receive FM, a receiver must be sensitive to the frequency variations of the incoming signal, which may be either NBFM or WBFM, and insensitive to amplitude variations. To assure that any amplitude variations are removed, the signals are amplified to such a degree that the amplifier goes into limiting. The demodulator must be frequency-dependent in order to be able to linearly convert frequency variations into voltage variations. This suggests that the FM demodulator must have an S-type voltage–frequency characteristic as shown in Fig. 4.11.

The principal types of FM demodulator circuits are as follows [48]:

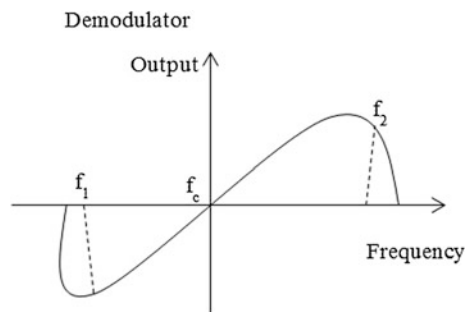
- Slope detector.
- Ratio detector.
- Foster–Seeley detector.
- Phase-locked loop (PLL) detector.
- Quadrature detector.

**Slope detector** This is the simplest form of FM detector. It is essentially a tank circuit that is tuned to a frequency slightly higher or lower than the carrier frequency.

**Ratio detector** This detector is very popular because it provides a higher level of amplitude variation rejection. This has the result of a greater level of noise immunity (since most noise is amplitude noise), and a satisfactory operation with lower levels of limiting circuitry is required. The ratio detector consists of a frequency-sensitive, phase-shift circuit with a transformer and two diodes in series. The transformer performs the detection of the frequency variations of the incoming signal.

**Foster–Seeley detector** This detector, also called a *phase-shift detector*, was invented in 1936 by Dudley Foster and Stuart Seeley [49]. It employs a double-tuned, radio frequency transformer to convert frequency variations of its

**Fig. 4.11** S-curve characteristic of an FM demodulator



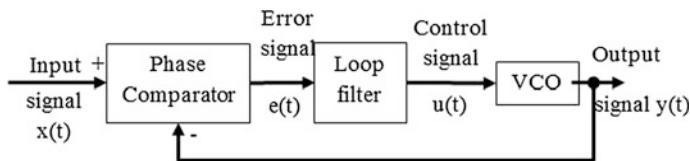
input signal to amplitude variations. These amplitude variations are then rectified and filtered to give a dc output voltage that varies in both amplitude and polarity as the frequency of the input signal varies. The output voltage is zero when the input frequency is equal to the carrier frequency  $f_c$ . A typical response curve of the Foster–Seeley detector has the S-form shown in Fig. 4.11, where the frequency band of operation is  $[f_1, f_2]$ .

**Phase-locked-loop detector (PLL)** This detector produces a fixed (locked) relation to the phase of a “reference” input signal and responds to both the frequency and the phase of the input signals automatically, so that it matches the reference in both frequency and phase. Actually, a PLL detector is an example of a control system working under negative feedback. PLL FM detectors are very popular and used in several types of radio equipment ranging from broadcasting receivers to high-performance communications equipment. Their wide use started when integrated circuit technology had developed to allow the manufacture of radio frequency analog circuits. The basic structure of a PLL detector is shown in Fig. 4.12 [42].

The phase comparator compares the phase of the output signal with that of the input signal and sends the phase-error signal to the loop filter, where it is filtered and becomes the control signal  $u(t)$  that drives the *voltage-controlled oscillator* (VCO). The key issue in designing a PLL detector is the loop filter, which must have sufficiently wide bandwidth to allow it to follow the anticipated variations of the frequency-modulated signal.

**Quadrature detector** The quadrature detector shifts the incoming signal,  $90^\circ$  and multiplies it with the unshifted signal. It does not need a center-tapped transformer and so it can easily be integrated into a single LSI chip, unlike the other detectors. The  $90^\circ$ -phase shift is achieved using a high-reactance capacitor and passes the phase-shifted signal to an LC circuit tuned at the carrier frequency. Then, the frequency changes produce an additional leading or lagging phase shift within the multiplier. The quadrature FM detector circuit is shown in Fig. 4.13. The operation of the quadrature FM detector is based on the property that multiplying two periodic signals with the same frequency generates a DC voltage that is directly proportional to the signal-phase difference.

At the resonance frequency, the phase of the LC circuit is zero, below resonance the phase is positive and above resonance the phase is negative, i.e., the phase changes with the variations of the frequency, and the same is also true for the output voltage.



**Fig. 4.12** Basic structure of a PLL detector

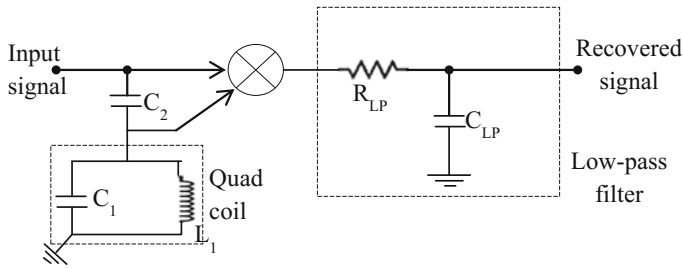


Fig. 4.13 Quadrature detector circuit (The capacitor  $C_2$  shifts the phase of the input signal  $90^\circ$ )

### 4.4.5.5 Phase Modulation

For a carrier  $c(t) = A \cos(\omega_c t + \phi_c)$  and a modulating signal  $m(t) = \mu \cos(\omega_m t)$ , the phase modulated (PM) signal  $x(t)$  has the following form:

$$x(t) = A \cos(\omega_c t + \psi(t)), \quad \psi(t) = m(t) + \phi_c$$

Here, we directly see that, as  $m(t)$  increases or decreases over time, so does the phase shift of the modulated signal. The phase shift can also be viewed as a change of the carrier frequency. Therefore, phase modulation is equivalent to frequency modulation, and so it can be analyzed by the methods presented in Sect. 4.4.5.3. We note again that, for small amplitude-modulating signals, we have the undesired result of sidebands and poor efficiency. For very large, single, sinusoid modulating signals, the bandwidth of PM is approximately given by Carson’s rule as in FM, i.e.,

$$B_C = 2(1 + \mu_{PM})f_m$$

where  $\mu_{PM}$  is the *PM modulation index* defined as  $\mu_{PM} = \Delta\phi$  with  $\Delta\phi$  being the maximum phase shift. An example of PM signal is shown in Fig. 4.14.

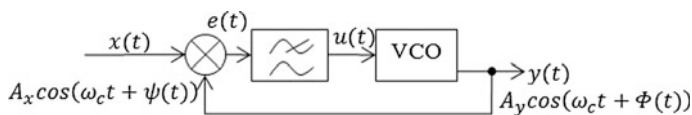
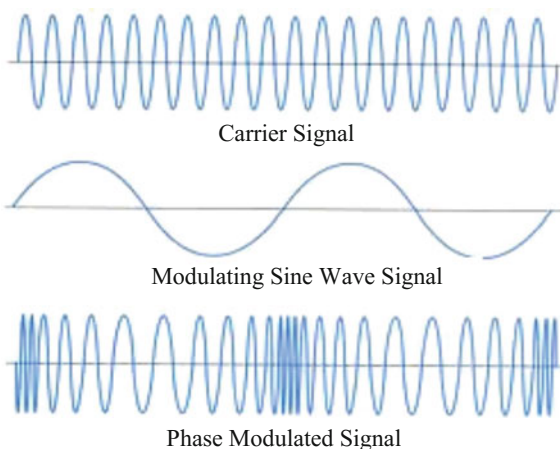
### 4.4.5.6 Phase Demodulation

Because of the equivalence of FM and PM (the carrier frequency is modulated by the time derivative of the phase shift), phase demodulation can be performed by any FM demodulator. Here, we will outline a phase demodulator of the PLL type, called a “*sinusoidal phase detector*”, which is shown in Fig. 4.15 [42].

In the phase-locked loop detector of Fig. 4.13, the VCO tends to phase lock to the input in “*quadrature*”, i.e., with  $90^\circ$ -phase difference ( $\phi(t) \rightarrow \psi(t) + \pi/2$ ). This means that we can define  $\phi(t)$  as  $\phi(t) = \theta(t) + \pi/2$  with  $\theta(t) \rightarrow \psi(t)$  as  $t \rightarrow \infty$ .

In Fig. 4.18, the output  $e(t)$  of the multiplier is equal to

**Fig. 4.14** Typical PM signal



**Fig. 4.15** Sinusoidal phase detector

$$\begin{aligned}
 e(t) &= x(t)y(t) = (A_x A_y / 2) [\cos(\psi(t) - \phi(t)) \\
 &\quad + \cos(2\omega_c + \psi(t) + \phi(t))] \\
 &= (A_x A_y / 2) [\cos(\psi(t) - \theta(t) - \pi/2) \\
 &\quad + \cos(2\omega_c + \psi(t) + \theta(t) + \pi/2)] \\
 &= (A_x A_y / 2) [\sin(\psi(t) - \theta(t)) - \sin(2\omega_c t + \psi(t) + \theta(t))]
 \end{aligned}$$

Now, assuming that the low-pass filter removes the second (high frequency) term, the output  $u(t)$  of the filter is found to be

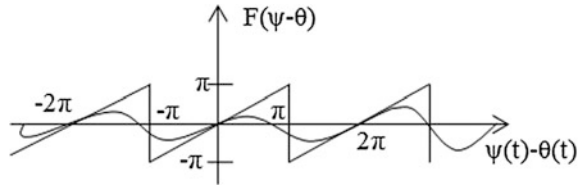
$$u(t) = (A_x A_y / 2) \sin(\psi(t) - \theta(t))$$

where  $\theta(t)$  is considered to be the output phase. Here, instead of a sawtooth function  $F(\cdot)$ , we have the sinusoid function:

$$F(p) = (A_x A_y / 2) \sin(p(t)), \quad p(t) = \psi(t) - \theta(t)$$

which has the plot shown in Fig. 4.16.

**Fig. 4.16** The sinusoidal gain function  $F(\cdot)$ . For comparison, the sawtooth function is also shown



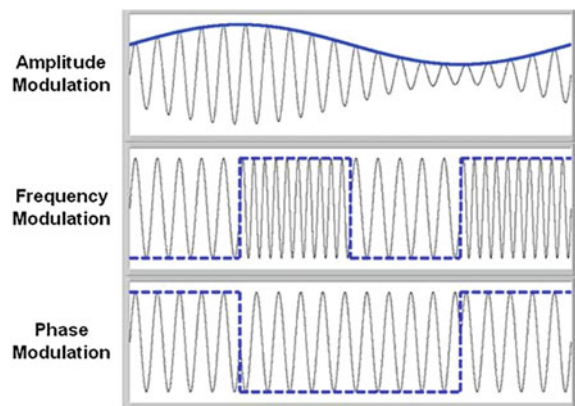
Here,  $\theta(t) \rightarrow \psi(t)$ , and so the phase  $\phi(t)$  of the detector tends to lock in quadrature with the input. Figure 4.17 shows in the same plot the AM, FM, and PM signals obtained using the same modulating signal and the same carrier in all cases. This plot can be generated by using the proper signal generation (SigGen) modules available in [50].

### 4.4.6 Pulse Modulation and Demodulation

**General Issues** *Analog modulation* (AM, FM, PM) is used for transferring an analog low-pass (baseband) signal, such as an audio signal or TV signal, over an analog passband channel, e.g., a restricted radio frequency band or a cable TV network channel. *Digital modulation* is used for transferring a digital bit stream over an analog passband channel, such as a limited radio frequency band or a public switched telephone network (where a bandpass filter restricts the frequency band to between 300 and 3400 Hz).

*Pulse modulation* is used to transfer a narrowband analog signal, such as a phone call, over a wideband baseband channel, or in some cases, as a bit stream over another digital transmission system. It is also used in neuron modeling and circuit design. Analog modulation/demodulation techniques were discussed in previous sections. Here, we will briefly discuss the basic pulse modulation schemes. Digital modulation schemes will be presented in the next section. Pulse modulation

**Fig. 4.17** AM, FM, and PM signals compared. The carrier and modulating signals are shown superimposed on the top



schemes use as a carrier signal a pulse train. The form of the pulses can be selected from among several types that differ in terms of energy and spectral content consumption. Examples of pulse types are square pulses and raised-cosine pulses. The five basic pulse modulation methods are the following, according to the pulse train parameter, that is, modulated (amplitude, width/duration, frequency, and position of leading edge):

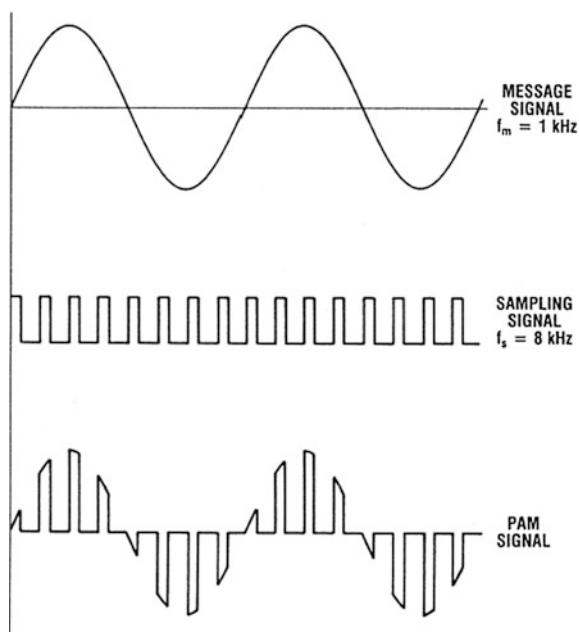
- Pulse-amplitude modulation (PAM).
- Pulse-width modulation (PWM).
- Pulse-frequency modulation (PFM).
- Pulse-position modulation (PPM).
- Pulse-code modulation (PCM).

**Pulse-Amplitude Modulation and Demodulation** In pulse-amplitude modulation (PAM), the amplitude (height) of individual pulses in the pulse train is varied from its normal value in accordance with the instantaneous amplitude of the modulating signal at sampling intervals. The width, frequency, and position of the pulses are kept constant. In other words, the information carried by the modulating signal is carried on a train of pulses being encoded in the amplitude of the pulses.

Figure 4.18 shows a typical PAM signal.

The PAM transmitter design is simple since it is a standard sampler with constant sampling period. Similarly, the receiver (demodulator) is simple.

**Fig. 4.18** Example of PAM in which the amplitude of a square pulse carrier train is modulated by a sinusoid signal

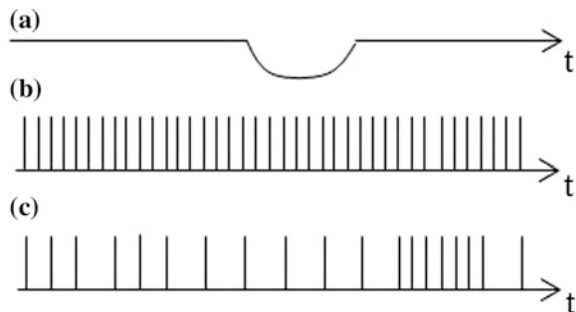


**Pulse-Width Modulation and Demodulation** In *pulse-width modulation* (PWM) or *pulse-duration modulation* (PDM) or *pulse-length modulation* (PLM), the modulating signal changes the width of individual pulses from their normal values, in proportion to the change at each sampling interval. The amplitude, frequency, and position of the pulses are kept constant. PWM is extensively used in power electronics and power control applications, because it is a very good method of providing intermediate quantities of electric power between *fully on* and *fully off*. It is noted that a standard power switch provides full power when switched on, and zero power when switched off. In communication applications, the width of the pulses corresponds to specific data values encoded at one end and decoded at the other end. Pulses of various widths (lengths) that represent the transmitted information are sent at regular intervals determined by the carrier frequency. There is no need to use a clock signal, because the leading edge of the pulse train plays the role of a clock. But, to avoid a data value with a zero-width pulse, the addition of a small leading edge offset is required.

**Pulse-Frequency Modulation and Demodulation** In *pulse-frequency modulation* (PFM), the instantaneous amplitude of the modulating signal changes the frequency of the carrier-pulse train, leaving unaltered the amplitude and width of the pulses. PFM is analogous to PWM since the magnitude of the information-bearing (modulating) signal is encoded in the duty cycle of the square pulse train. Compared to PAM, PFM has the advantage of better immunity to noise interference. The drawback is that the design of transmitter and receiver is more complex. For this reason, in practice, PAM or PWM is more commonly used. An example of PFM is shown in Fig. 4.19.

**Pulse-Position Modulation** In *pulse-position modulation* (PPM), the variation of the instantaneously sampled values of the modulating signal changes the position of each pulse with respect to the position of a periodic reference pulse. The amplitude and width of the pulses are kept constant, and so the required transmitter power is constant. PPM has the drawback that it depends on the synchronization of the transmitter–receiver pair. PPM is widely used in optical communications. PPM and FSK (frequency-shift keying) modulation are two canonical forms of orthogonal modulation. PPM and FSK are actually dual modulation techniques in the sense that

**Fig. 4.19** Examples of PFM: **a** modulating signal, **b** carrier periodic pulse train, and **c** PFM pulse train



in FSK each signal gets a different slice of the frequency band, whereas in PPM each signal gets a different slice of the signaling interval. A PPM signal can be produced by feeding a PWM signal into a differentiating circuit, which provides positive-and-negative-polarity pulses at the rising and falling edges of the PWM pulses. Passing these alternating polarity pulses to a rectification circuit, which cuts the positive fixed (non-modulated) pulses, we obtain the negative pulse train which is the desired PPM pulse signal.

**Pulse-Code Modulation and Demodulation** In *pulse-code modulation (PCM)*, the amplitude of the analog modulating signal is sampled with a fixed sampling rate, and then it is quantized to a set of symbols, typically a binary code. The PCM principle is illustrated in Fig. 4.20.

The four-bit coded values of the sinusoid modulating signals at the sampling instants one up to 12 are as shown in Table 4.1.

The general structure of a PCM modulator has the form shown in Fig. 4.21.

This structure is implemented in several ways, using suitable, single-integrated circuits (known as analog-to-digital (A/D) converters).

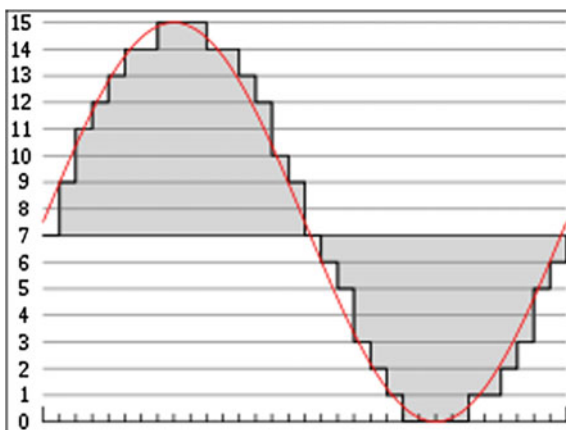
*Pulse-coded demodulation* reproduces the analog input (message) from the digital output using a reverse sequence of operations. A PCM demodulator circuit is known as digital-to-analog (D/A) converter (see Fig. 4.22).

Pulse-code modulation has two important features:

- Noise contamination is almost completely eliminated when the pulse signals exceed the noise levels by at least 20 dB.
- The signal can be received and retransmitted as many times as desired with no distortion of the signal.

**Other Pulse Modulation Techniques** Besides the five types of pulse modulation just discussed, over the years, several other techniques with associated benefits and drawbacks have been developed. These are as follows [51]:

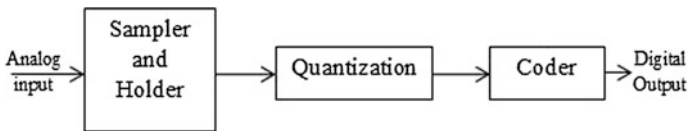
**Fig. 4.20** The principle of PCM. Sampling and binary coding using 16 quantized levels (i.e., four-bit quantization)



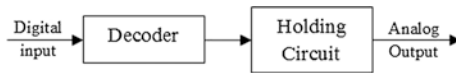


**Table 4.1** PCM of a sinusoid using the four-bit binary code 1-2-4-8

Sampling instant	Sampled value	Binary coded value
1	9	1001
2	11	1011
3	12	1100
4	13	1101
5	14	1110
6	14	1110
7	15	1111
8	15	1111
9	15	1111
10	14	1110
11	14	1110
12	13	1101



**Fig. 4.21** The pulse-code modulator performs the operations: sampling–holding, quantization, and binary (digital) coding



**Fig. 4.22** The PCM demodulator (or D/A converter) is performed by a binary decoder and an analog holder

- **ASK:** *Amplitude-Shift Keying*, where a finite number of amplitudes are used.
- **FSK:** *Frequency-Shift Keying*, where a finite number of frequencies are used.
- **PSK:** *Phase-Shift Keying*, where a finite number of phases are used.
- **BPSK:** *Binary-Phase Shift Keying*.
- **QPSK:** *Quadrature-Phase Shift Keying*.
- **QAM:** *Quadrature-Amplitude Modulation*.
- **ADPCM:** *Adaptive-Differential PCM*.
- **PDM:** *Pulse-Density Modulation*.

Some more specific schemes of quantized modulation (QM) are the following:

- **QAM:** *Quantized-Amplitude Modulation*
- **QFM:** *Quantized-Frequency Modulation*.
- **QPAM:** *Quantized-Pulse-Amplitude Modulation*.
- **QPM:** *Quantized-Phase Modulation*.

- **QPPM:** *Quantized-Pulse-Position Modulation*.

For descriptions and details about them, the reader is referred to modern textbooks of communication systems [42, 52].

## 4.5 Information Theory

### 4.5.1 General Issues

As we saw in Sect. 4.3.5, information theory was formally initiated by Claude Shannon (1948) who coined the concept of “*entropy*” as the average “*information content*” in a message, measured in bits (binary digits) [9]. The term “*information*” was originally coined by Ralph Hartley in his 1928 paper “*Transmission of Information*”, where he treated “*information*” in a technical sense as a measurable quantity, reflecting the receiver’s ability to recognize the sequence of symbols sent by the sender (without any concern about the meaning or semantic issues of these symbols). Hartley’s formula for information is [8]

$$I = \log S^n = n \log S$$

where  $S$  is the number of possible symbols and  $n$  is the number of symbols in a transmission. The information  $I$  is measured in *decimal digits*, also called *Hartley units*.

Information theory is a mathematical theory principally concerned with coding–decoding. Its primary mathematical tools are probability theory and statistics. Shannon’s formula for entropy is

$$H(X) = - \sum_{k=1}^n p_k \log_2 p_k$$

where  $p_k = p(x_k)$  are discrete probabilities of the random process (message)  $X$ , with possible values  $x_1, x_2, \dots, x_n$ , which express the probabilities that a particular message is transmitted.  $H(X)$  is a measure of how much information is contained in the transmitted message.

Shannon gave the name “*entropy*” to the quantity  $H(X)$  upon the suggestion of John von Neumann. Originally, he thought to call it “*information*” but since this word was overly used, he decided to call it “*uncertainty*”. But, in a private conversation with von Neumann, he was advised to call it “*entropy*”. “You should call it entropy, John von Neumann suggested, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.”

In the following, we will present two derivations of the concept of entropy,  $H(X)$ , and the four fundamental theorems of information theory, namely,

- Nyquist–Shannon sampling theorem.
- Shannon’s source coding theorem.
- Shannon’s channel capacity theorem.
- Landau–Pollack bandwidth signal dimensionality theorem.

### 4.5.2 Information Theory’s Entropy

The following two ways of introducing information theory’s entropy will be discussed:

- Entropy as the information content of a measurement.
- Direct definition of entropy.

#### 4.5.2.1 Entropy Derivation via the Information Content of a Measurement

The present derivation is as follows [53]. A real number  $N$  is typically expressed by the infinite series:

$$N = d_{-n}10^n + \dots + d_{-2}10^2 + d_{-1}10 + d_0 + d_110^{-1} + \dots + d_n10^{-n} + \dots$$

where the  $d_i$ ’s (digits) are integers between 0 and 9, and  $d_{-n} \neq 0$  for  $n > 0$ . In shorthand,  $N$  is written in the well-known decimal form:

$$N = d_{-n}d_{-n+1} \dots d_{-1}d_0.d_1d_2 \dots d_n$$

For each number, there exists a unique decimal expansion of this form, provided that we agree to exclude expansions like 2.36999... where nines repeat indefinitely, and express numbers like 2.5000... as 2.5 (omitting the infinitely repeating zeros).

Rational numbers are defined to be ratios of integers, i.e.,

$$a/b = d_n \dots d_1d_0.d_1d_2 \dots d_n \quad (b > 0)$$

A decimal expression represents a fraction if and only if some sequence of digits in the expression repeats indefinitely. For example,

$$1/4 = 0.2500, \dots, \quad 1/6 = 0.166 \dots 6 \dots$$

$$13/70 = 0.1857142857142857142857 \dots$$

With other numbers, the situation is more complex. For example, the decimal expressions for the numbers  $\sqrt{2}$  and  $\pi$  (the ratio of the circumference of a circle to its diameter) are as follows:

$$\sqrt{2} = 1.41421356\dots, \pi = 3.14159265\dots$$

where all digits must be calculated one by one. These classes of numbers differ in the amount of information they convey and in the effort required to obtain that information.

Although in some cases (like the ones mentioned above) we know explicitly all the digits of the expansion of a number, this is not so when we observe or measure a certain physical quantity. What one can do here is to obtain successive refined intervals that bound the actual value  $m$  of the quantity  $M$  at hand. Suppose that we get an estimate of the true value  $m$  in the interval  $\hat{m}_1 \in (a_1, b_1)$ . If this is not satisfying, a second more precise measurement is performed which gives

$$\hat{m}_2 \in (a_2, b_2) \text{ with } a_1 < a_2 < b_2 < b_1$$

Clearly, the new bounding interval  $(a_2, b_2)$  provides some additional number of digits in the decimal expansion of  $m$ . The number of decimal digits in the original decimal expansion is approximately given by  $\log_{10}(b_1 - a_1)$ , and the number of digits of the more accurate expansion is  $\log_{10}(b_2 - a_2)$ .

The gain  $I_{10}$  in information when we go from the first interval  $(a_1, b_1)$  to the second interval  $(a_2, b_2)$  is equal to

$$\begin{aligned} I_{10} &= \log_{10}(b_1 - a_1) - \log_{10}(b_2 - a_2) \\ &= \log_{10}[(b_1 - a_1)/(b_2 - a_2)] \quad \text{decimal digits} \end{aligned}$$

If the number  $N$  is represented in binary form:  $N = c_{-n} \cdots c_{-1} c_0 \cdot c_1 c_2 \cdots c_n$  where  $c_i = 0$  or  $1$  (for all  $i$ ), the information gain is equal to

$$I_2 = \log_2[(b_1 - a_1)/(b_2 - a_2)] \text{ bits}$$

Since  $b_1 - a_1$  is the length of the interval with limits  $a_1$  and  $b_1$ , and  $b_2 - a_2$  is the corresponding length of the second interval with limits  $a_2$  and  $b_2$ , the ratio  $(b_2 - a_2)/(b_1 - a_1)$  represents the probability  $p$  that a random point in the interval  $(a_1, b_1)$  lies in the interval  $(a_2, b_2)$ , i.e.,

$$p = (b_2 - a_2)/(b_1 - a_1)$$

Thus, the gain in information can be expressed as

$$I_2 = \log_2[(b_1 - a_1)/(b_2 - a_2)] = \log_2\left(\frac{1}{p}\right), \quad 0 \leq p \leq 1$$

Now, consider a system that can be in one of the finite sets of states  $x_1, x_2, \dots, x_n$  (arbitrarily ordered), and let  $p_k$  be the probability of the system to be in state  $x_k$ . Clearly, the probabilities  $p_k$  must satisfy the total probability condition:

$$\sum_{k=1}^n p_k = 1$$

Suppose that an observation of the system reveals that the system is in state  $x_k$ . To calculate how much information was gained after this observation, we assume that our observation system associates with each state a probability interval. For example, state  $x_1$  with probability  $p_1$  corresponds to the interval  $0 < x < p_1$ , state  $x_2$  with the interval  $p_1 < x < p_1 + p_2$ , and so on. In general, the  $k$ th state  $x_k$  is associated with the interval:

$$p_1 + p_2 + \cdots + p_{k-1} < x < p_1 + p_2 + \cdots + p_k$$

with limiting condition for the  $n$ th state of

$$p_1 + p_2 + \cdots + p_{n-1} < x < p_1 + p_2 + \cdots + p_n = 1$$

After observing that the system is in state  $x_k$ , the length of the measurement interval becomes  $p_k$ , whereas, at the beginning (without any measurement), we have  $0 < x < 1$ , i.e., a length interval equal to one. Therefore, the gain in information compared with the case before the observation is

$$\log_2(1/p_k)$$

The average information  $\bar{I}$  gained when the system changes from one state to another, up to the state  $x_n$ , is equal to

$$\begin{aligned} \bar{I} &= p_1 \log_2(1/p_1) + p_2 \log_2(1/p_2) + \cdots + p_n \log_2(1/p_n) \\ &= \sum_{k=1}^n p_k \log_2(1/p_k) \\ &= - \sum_{k=1}^n p_k \log_2 p_k = H \end{aligned}$$

where  $H$  is exactly the *entropy* introduced by Shannon in his 1948 paper [9]. Here, if  $p_k = 0$  for some  $k$ , the value of the term  $0 \log_2 0$  is taken to be 0, consistent with the limit  $\lim_{p \rightarrow 0^+} p \log_2 p = 0$ .

#### 4.5.2.2 Direct Definition of Entropy

This direct way of defining entropy is due to Shannon and can be found in most textbooks on information theory or communication systems. Consider a random variable  $X$  that carries an infinite amount of information if it is continuous in amplitude range. Each realization (presentation) of  $X$  can be considered to represent

a message. Actually, in a physical or biological system, it is not realistic to assume that the amplitude measurements can have infinite accuracy. Therefore, it seems to be realistic to assume that the value of  $X$  can be uniformly *quantized* to a finite set of values, in which case  $X$  is *discrete* random variable:

$$X = \{x_k; k = 0, \pm 1, \pm 2, \dots, \pm n\}$$

where  $2n + 1$  is the total number of discrete levels. Now, suppose that the event  $X = x_k$  occurs with probability:

$$p_k = P(x_k) = P(X = x_k)$$

under the conditions:

$$0 \leq p_k \leq 1 \quad \text{and} \quad \sum_{k=1}^n p_k = 1$$

If for some  $k$ , the event  $X = x_k$  occurs with probability  $p_k = 1$ , in which case all other  $p_i$ 's for  $i \neq k$  are zero, the occurrence of the event  $X = x_k$  does not add any “*information*” and does not incur any “*surprise*”, since we know surely what the message is. However, if the various discrete levels occur with different probabilities, the occurrence of some event  $X = x_k$  conveys some *information* (or surprise), which is higher when  $p_k = p(x_k)$  is lower (i.e., the *uncertainty* about the occurrence of  $X = x_k$  is higher). Thus, the terms “*uncertainty*”, “*surprise*”, and “*information*” can be used interchangeably in the information theory framework. In particular, the amount of information (or surprise) is related to the inverse of the probability of occurrence.

On the basis of this discussion, the amount of information obtained after the observation of the event  $X = x_k$  with probability  $p_k$  is defined as

$$I_B(x_k) = \log_B(1/p_k) = -\log_B p_k$$

where the base  $B$  of the logarithm is arbitrary and when the base  $B = 2$ ,  $I(x_k)$  is measured in bits, when  $B = 10$   $I(x_k)$  is measured in decimal digits (Hartley's), and when  $B = e$  (natural logarithm)  $I(x_k)$  is measured in natural units (nats).

The quantity  $I(x_k)$  defined here has the following properties:

- $I_B(x_k) = 0$  for  $p_k = 1$  (i.e., no information is gained if the occurrence of  $X = x_k$  is absolutely certain).
- $I_B(x_k) \geq 0$  for  $0 \leq p_k \leq 1$  (i.e., a loss of information never occurs).
- $I_B(x_k) > I(x_i)$  for  $p_k < p_i$  ( $k \neq i$ ) (i.e., the less probable an event is, the more information is gained after its occurrence).

Clearly, the quantity  $I_B(x_k)$  is a random variable with probability  $p_k$ . Therefore, the mean value  $\bar{I}_B$  of  $I_B(x_k)$  over the entire range of  $2n + 1$  discrete values is equal to

$$\begin{aligned}\bar{I}_B(X) &= E[I(x_k)] = \sum_{k=-n}^n p_k I(x_k) \\ &= - \sum_{k=-n}^n p_k \log p_k = H(X)\end{aligned}$$

and is called the *entropy* of the discrete random variable  $X$ , which takes a finite set of  $(2n + 1)$  values. The entropy  $H(X)$  provides a measure of the average amount of information provided per message. The basic properties of the entropy  $H(X)$  are as follows:

- $H(X)$  is continuous (a small change of the probabilities yield a small change in  $H(X)$ );
- $0 \leq H(X) \leq \log_B(2n + 1)$ ;
- $H(X) = 0$  if and only if  $p_k = 1$  for some  $k$  (i.e., if there is no uncertainty);
- $H(X) = \log_B(2n + 1)$  if and only if all discrete levels are equally probable ( $p_k = 1 / (2n + 1)$  for all  $k$ );
- $H(X)$  is additive (i.e., the entropy is independent to how the process is divided into parts). This allows us to calculate the entropy of a system via the entropies of the subsystems, if we know the interactions between the subsystems. That is if the system is divided into  $M$  blocks (subsystems) with  $b_1, b_2, \dots, b_M$  elements each, we have

$$\begin{aligned}H\left(\frac{1}{2n+1}, \dots, \frac{1}{2n+1}\right) &= H_M\left(\frac{b_1}{2n+1}, \dots, \frac{b_M}{2n+1}\right) \\ &\quad + \sum_{k=-n}^n \frac{b_k}{2n+1} H_{b_k}\left(\frac{1}{b_k}, \dots, \frac{1}{b_k}\right);\end{aligned}$$

- $H(X)$  remains unchanged if the events  $X = x_k$  are re-ordered, i.e.,

$$H(x_n, \dots, x_0, x_1, \dots, x_n) = H(x_{-n+1}, x_n, \dots, x_0, x_1, \dots, x_n);$$

- For equal probability events,  $H(X)$  increases with the number of observations, i.e.,

$$H\left(\frac{1}{2n+1}, \dots, \frac{1}{2n+1}\right) < H\left(\frac{1}{2n+2}, \dots, \frac{1}{2n+2}\right);$$

- The addition or removal of an event with probability zero does not change the value of the entropy:

$$H\left(\frac{1}{2n+1}, \dots, \frac{1}{2n+1}, 0\right) = H\left(\frac{1}{2n+1}, \dots, \frac{1}{2n+1}\right).$$

### 4.5.2.3 Differential Entropy

The concept of entropy thus far discussed refers to discrete random variables. Here, we will introduce the corresponding concept for the case of a continuous random variable  $X$  that has a probability density function  $f(x)$ . This concept is the following:

$$\begin{aligned} h(X) &= - \int_{-\infty}^{\infty} f(x) \log_B f(x) dx \\ &= - E[\log_B f(x)] \end{aligned}$$

and is called the *differential entropy* of  $X$  (in contrast to the absolute entropy  $H(X)$ ). Of course, it is remarked right from now that  $h(X)$  does not provide in any way a measure of the randomness of  $X$ . This expression for the differential entropy is justified by using the mean value theorem:

$$f(x_k) \delta x = \int_{k\delta x}^{(k+1)\delta x} f(x) dx$$

which leads to the Riemannian approximation:

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{\delta x \rightarrow 0} \sum_{k=-\infty}^{\infty} f(x_k) \delta x$$

Define

$$\begin{aligned} H_\delta &= - \sum_{k=-\infty}^{\infty} f(x_k) \delta x \log_B [f(x_k) \delta x] \\ &= - \sum_{k=-\infty}^{\infty} f(x_k) \delta x \log_B f(x_k) - \sum_{k=-\infty}^{\infty} f(x_k) \delta x \log_B \delta x \end{aligned}$$

Then, as  $\delta x \rightarrow 0$ :

$$\lim_{\delta x \rightarrow 0} H_\delta = - \int_{-\infty}^{\infty} f(x) \log_B f(x) dx - \lim_{\delta x \rightarrow 0} (\log_B \delta x) \int_{-\infty}^{\infty} f(x) dx$$

Or



$$h(X) = - \int_{-\infty}^{\infty} f(x) \log_B f(x) dx = \lim_{\delta x \rightarrow 0} (H_\delta + \log_B \delta x)$$

where the relation  $\int_{-\infty}^{\infty} f(x) dx = 1$  was used.

Now,  $\lim_{\delta x \rightarrow 0} \log_B \delta x = -\infty$ , which implies that the entropy of a continuous variable is infinitely large. This is indeed justified by the fact that a continuous random variable may take any value in the interval  $(-\infty, \infty)$ , and so the uncertainty associated with the variable is infinite. Regarding the infinite offset,  $\log_B 0 = -\infty$ , as a *reference*, the quantity  $h(X)$  is actually a *differential entropy*. Clearly, the differential entropy is not a limit of the Shannon entropy for  $n \rightarrow \infty$ , but differs from this limit by the infinite offset  $\log \delta x \rightarrow -\infty$  as  $\delta x \rightarrow 0$ .

#### 4.5.2.4 Joint Entropy, Conditional Entropy, and Mutual Information

**Joint entropy** Consider two discrete random variables  $X$  and  $Y$ . The entropy of their pairing  $(X, Y)$  is called *joint entropy*  $H(X, Y)$ , i.e.,

$$\begin{aligned} H(X, Y) &= E_{(X,Y)}[-\log_B p(x, y)] \\ &= - \sum_{x,y} p(x, y) \log_B p(x, y) \end{aligned}$$

**Conditional entropy** The *conditional entropy* of the random variable  $X$  given the random variable  $Y$  is defined as

$$\begin{aligned} H(X|Y) &= E_Y[H(X|Y)] \\ &= - \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_B p(x|y) \\ &= - \sum_{x,y} p(x, y) \log_B [p(x, y) / p(y)] \end{aligned}$$

i.e.,  $H(X|Y)$  is the average conditional entropy  $E_Y[H(X|Y)]$  over  $Y$ . It follows that

$$H(X|Y) = H(X, Y) - H(Y)$$

**Mutual information** The *mutual information* (or *transinformation*) is defined to be the amount of information that can be obtained about one random variable  $X$  via the observation of another variable  $Y$ , i.e.,

$$I(X; Y) = E_{X,Y}[I_s(x, y)] = \sum_{x,y} p(x, y) \log_B \frac{p(x, y)}{p(x)p(y)}$$

where  $I_s(x, y)$  is the point-wise (specific) mutual information. It follows that  $I(X; Y)$  has the following property:

$$I(X; Y) = H(X) - H(X|Y)$$

This means that, if we know  $Y$ , we can have a saving of  $I(X; Y)$  bits in encoding  $X$ , compared to the case in which we do not know  $Y$ . The mutual information has the following property:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) - [H(X, Y) - H(Y)] \\ &= H(X) + H(Y) - H(X, Y) \\ &= H(Y) - [H(X, Y) - H(X)] \\ &= H(Y) - H(Y|X) = I(Y; X) \end{aligned}$$

i.e., the mutual information is *symmetric*.

### 4.5.3 Coding Theory

#### 4.5.3.1 General Issues

Coding theory is the branch of information theory that deals with the transmission of information across communication channels (noiseless, noisy) and the recovery of the messages. Coding theory aims to make messages easy to read, and should not be confused with *cryptology*, which is concerned with making messages hard to read. One of the principal goals of coding theory is to remove information redundancy (i.e., compress data) and correct the errors that may occur. Specifically, coding theory studies the properties of codes and their suitability for particular applications. Here, a brief outline of coding theory will be given. Full presentations can be found in the respective literature (e.g., [54–66]).

Coding is categorized into three categories:

- Source coding.
- Channel coding.
- Combined source and channel coding.

*Source encoding* (or *data compression*) performs data compression on the information sent by a source, so that the transmission is more efficient. *Channel encoding* adds extra bits to assure a more robust transmission regarding disturbances and noise present on the transmission channel.

### 4.5.3.2 Source Coding

*Source coding (compression)* is the process of encoding information using fewer information-bearing units (e.g., bits) compared to a non-coded representation, through the employment of particular encoding schemes. Compression is very useful since it contributes to the reduction of the consumption of expensive resources. The development of source coding techniques and the design of equipment must be based on compromises among several factors, such as the degree of compression, the level of noise and distortion involved, and the computational resources needed for the coding and recovery of the data.

The two ways of data compression are as follows:

- Lossless data compression, where the data must be recovered precisely.
- Lossy data compression, where the bits required for recovering the data with the desired reliability, as measured by a distortion index, are added.

### 4.5.3.3 Channel Coding

*Channel coding* is concerned with maximizing the information rate that the channel can convey reliably, i.e., with acceptable error probability. To this end, codes must be designed that allow fast transmission of data, involve many valid codewords, and can detect and correct various errors. This again requires a number of trade-offs between several conflicting issues. Although source coding tries to remove as much redundancy as possible, channel coding designs and investigates several error-correcting codes, which add sufficient redundancy (i.e., error correction) that assures efficient and faithful transmission of information across a noisy channel.

### 4.5.3.4 Error Detecting and Correcting Codes

**Error detection** It deals with the detection of errors introduced by noise and other disturbances across the transmission channel from the transmitter to the receiver. The general way is to add some extra data bits (i.e., redundancy) to a message, which makes possible the detection of any errors in the conveyed message. The additional bits are called *check bits* and are determined by a suitable algorithm applied to the message bits. To detect any existing errors, the receiver applies the same algorithm to the received data bits and compares the output to the received check bits. If no matching of the values occurs, an error has taken place at some point of the transmission.

**Error correction** It deals with both the detection of errors and the recovery of the initial, error-free data. The three primary ways for design the channel code for error correction are as follows:

- **ARQ** (*Automatic Repeat reQuest*): The receiver requests the retransmission of the data that are assumed to contain errors.
- **FEC** (*Forward Error Correction*): The transmitter sends the encoded message with an error-correcting code. The receiver decodes the received message into the most likely data, without sending any request for retransmission.
- **Hybrid ARQ and FEC**: Here minor errors are restored without retransmission, and major errors are corrected by sending a request for retransmission.

Three ARQ schemes that are suitable for varying or unknown capacity communication channels (such as the Internet) are as follows:

- Stop-and-Wait ARQ.
- Go-back-N ARQ.
- Selective Repeat ARQ.

**FEC** is typically employed in lower layer communication and in storage devices (CDs, DVDs, Dynamic RAM) and are classified into two classes:

- *Block codes* which are processed in blocks. In this class, we have, among others, repetition codes, Hamming codes, and multidimensional parity-check codes.
- *Convolutional codes* which are processed on a bit-by-bit basis. Here, the so-called *Viterby decoder* performs optimal decoding.

#### 4.5.3.5 Block Codes

An  $(n, k)$  block code, where  $k$  is the number of input bits and  $n$  is the number of output bits, is characterized by the following parameter:

$$\text{Code rate : } r = k/n$$

$$\text{Channel data rate : } R_0 = rR_n,$$

where  $R_n$  denotes the bit rate of the information source.

A binary block code  $C$  of block length  $n$  is a subset of the set of all binary  $n$ -tuples

$\bar{\mathbf{x}} = [x_0, x_1, \dots, x_{n-1}]$ , where  $x_i = 0$  or  $1$  for  $i = 0, 1, \dots, n-1$ . *Code vector* or *code word* is called an  $n$ -tuple belonging to the code.

- *Hamming weight*  $w([x_0, x_1, \dots, x_{n-1}])$  is the number of nonzero components of the  $n$ -type.
- *Hamming distance*  $d(\mathbf{x}, \mathbf{x}')$  between two  $n$ -tuples  $\mathbf{x}$  and  $\mathbf{x}'$  is defined as the number of positions in which their components differ. From this definition, it follows that

$$d(\mathbf{x}, \mathbf{x}') = w(\mathbf{x} - \mathbf{x}'),$$

where

$$\mathbf{x} - \mathbf{x}' = [x_0, x_1, \dots, x_{n-1}] - [x'_0, x'_1, \dots, x'_{n-1}] = [x_0 - x'_0, x_1 - x'_1, \dots, x_{n-1} - x'_{n-1}]$$

**Minimum Hamming distance**,  $d_{\min}$ , of the block code  $C$  is the smallest Hamming distance between pairs of distinct code words.

Some theorems on the detection and correction ability of block codes are the following [61].

**Theorem 1** A code  $C$  can detect all patterns of  $s$  or fewer errors if and only if  $d_{\min} > s$ .

**Theorem 2** A code  $C$  can correct all patterns of  $s$  or fewer errors if and only if  $d_{\min} > 2s$ .

This theorem suggests the following implicit decoding rule: “Decode  $\mathbf{c}$  (the corrupted received codeword of an actual codeword  $\mathbf{a}$ ) to the nearest codeword in terms of the Hamming distance, i.e.,  $d_{\min} = \min w(e) = \min w(\mathbf{c} - \mathbf{a}) = \min d(\mathbf{c}, \mathbf{a})$ ”

**Definition** A binary code  $C$  is linear if, for  $\mathbf{a}$  and  $\mathbf{a}'$  in  $C$ , the sum  $\mathbf{a} + \mathbf{a}'$  is also in  $C$ . For example, the code  $C = \{[0, 0, 0] + [1, 1, 0] + [1, 0, 1] + [0, 1, 1]\}$  is a linear code because  $[1, 1, 0] + [101] = [011]$ , and so on.

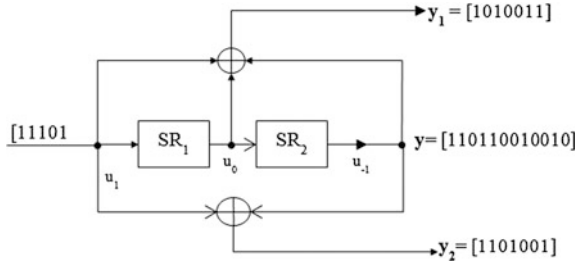
- The *minimum Hamming distance* of a linear block code is equal to the smallest weight of the nonzero code words in the code.

#### 4.5.3.6 Convolutional Codes

A *binary convolutional code* is characterized by a 3-tuple  $(n, k, m)$ , where  $k$  is the number of input (message) bits,  $n$  is the number of generated (output) bits, and  $m$  is the number of memory registers (memory order). As in block codes, the parameter  $r = k/n$  is called the *code rate*. Typically,  $k$  and  $n$  vary from 1 to 8,  $m$  from 2 to 10, and the code rate  $r$  from  $1/8$  to  $7/8$ . In deep space applications, code rates  $r$  as low as  $1/100$  or smaller have been used. In many commercial cases, the convolutional codes are described as  $(r, L)$ , where  $r$  is the code rate and  $L$  is the *constraint length*  $L = k(m - 1)$ , which represents the number of bits in the encoder memory that affect the generation of the  $n$  output bits.

A binary convolutional encoder is represented pictorially by a set of shift registers and modulo-2 adders, where the output bits are modulo-2 sums of selective shift register contents and present input bits. The diagram of a  $(2, 1, 2)$  convolutional code  $r$  is shown in Fig. 4.23.

**Fig. 4.23** Binary (2, 1, 2) convolutional encoder (SR<sub>*i*</sub> = *i*th shift register)



This is a rate  $r = 1/2$  code. Each input bit is coded into two output bits. The constraint length of the code is  $L = k(m - 1) = 1 \times (2 - 1) = 1$ .

The choice of the bits to be added for generating the output bit is determined by the so-called *generator polynomial* ( $g$ ) for that output. In Fig. 4.23, the first output bit has a generator polynomial of 111. The second output has a generator polynomial of 101. The output bits of  $y_1$  and  $y_2$  are given by

$$y_{1,i} = \text{mod } 2[u_1 + u_0 + u_{-1}]_i = u_{1,i} \oplus u_{0,i} \oplus u_{-1,i}$$

$$y_{2,i} = \text{mod } 2[u_1 + u_{-1}]_i = u_{1,i} \oplus u_{-1,i}$$

The generator polynomial gives a unique error protection quality to the code. For example, one code (3, 2, 2) may have entirely different properties from another, depending on the selected polynomial. However, it is noted that not all polynomials lead to good error correction performance. A list of good polynomials for rate 1/2 codes is given in Table 4.2 [54].

The encoders of convolutional codes are represented by multivariable (MV) linear time-invariant (LTI) systems as shown in Fig. 4.24.

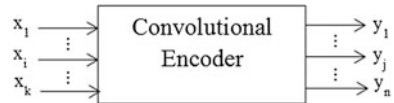
The  $y_j$  output in Fig. 4.21 is given by

$$y_j = \sum_{i=1}^k x_i * g_j^{(i)}$$

**Table 4.2** Efficient generator polynomials for 1/2 codes

Constraint length	G <sub>1</sub>	G <sub>2</sub>
3	110	111
4	1101	1110
5	11010	11101
6	110101	111011
7	110101	110101
8	110111	1110011
9	110111	111001101
10	110111001	1110011001

**Fig. 4.24** LTI representation of a MV convolutional encoder



where “\*” denotes the *convolution operator*, and  $g_j^{(i)}$  is the *impulse response* of the *i*th input sequence with respect to the *j*th output. The *impulse responses* are the *generator polynomials (sequences)* of the encoder, previously discussed.

The impulse response for the binary (2, 1, 2) convolutional code of Fig. 4.20 is

$$g_1 = [1, 1, 1, 0, \dots] = [1, 1, 1]$$

$$g_2 = [1, 0, 1, 0, \dots] = [1, 0, 1]$$

The outputs  $y_1$  and  $y_2$  corresponding to the input vector  $\mathbf{x} = [1\ 1\ 1\ 0\ 1]$  are equal to

$$y_1 = [1\ 1\ 1\ 0\ 1] * [1\ 1\ 1] = [1\ 0\ 1\ 0\ 0\ 1\ 1]$$

$$y_2 = [1\ 1\ 1\ 0\ 1] * [1\ 0\ 1] = [1\ 1\ 0\ 1\ 0\ 0\ 1]$$

As in block codes, the convolutional codes can be generated by a generator matrix multiplied by the input (information, message) vectors.

Let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$  and  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  be the input and output sequences. Arranging the input sequences as

$$\mathbf{x} = [x_{1,0}, x_{2,0}, \dots, x_{k,0}; \dots; x_{1,k}, x_{2,k}, \dots, x_{k,k}]$$

$$= [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_p, \dots]$$

and the output sequences as

$$\mathbf{y} = [y_{1,0}, y_{2,0}, \dots, y_{n,0}; \dots; y_{n,0}, y_{n,1}, \dots]$$

$$= [\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_p, \dots]$$

the *convolutional encoder* is described by the matrix-vector equation:

$$\mathbf{y} = \mathbf{x}\mathbf{A}$$

where  $\mathbf{A}$  is the generator matrix of the code:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_m & & \\ & \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{m-1} & \mathbf{A}_m & \\ & & \mathbf{A}_0 & \cdots & \mathbf{A}_{m-2} & \mathbf{A}_{m-1} & \mathbf{A}_m \\ & & & \ddots & & & \end{bmatrix}$$

where the  $k \times n$  blocks  $\mathbf{A}_p$  are given by

$$\mathbf{A}_p = \begin{bmatrix} \mathbf{g}_{1,p}^{(1)} & \cdots & \mathbf{g}_{n,p}^{(1)} \\ \mathbf{g}_{1,p}^{(2)} & \cdots & \mathbf{g}_{n,p}^{(2)} \\ \cdots & \cdots & \cdots \\ \mathbf{g}_{1,p}^{(k)} & \cdots & \mathbf{g}_{n,p}^{(k)} \end{bmatrix}$$

and  $g_{jp}^{(i)}$  the impulse response of the  $i$ th input with respect to the  $j$ th output:

$$\mathbf{g}_j^{(i)} = [g_{j,0}^{(i)}, g_{j,1}^{(i)} \cdots g_{j,p}^{(i)} \cdots g_{j,m}^{(i)}]$$

The output vector  $\mathbf{y}$  of the binary encoder (2, 1, 2) of Fig. 4.30 is given by

$$\mathbf{y} = \mathbf{x}\mathbf{A}$$

where  $\mathbf{x} = [1 \ 1 \ 1 \ 0 \ 1]$ , and

$$\mathbf{A} = \begin{bmatrix} 11 & 10 & 11 & & & & \\ & 11 & 10 & 11 & & & \\ & & 11 & 10 & 11 & & \\ & & & 11 & 10 & 11 & \\ & & & & 11 & 10 & 11 \end{bmatrix}$$

i.e.,

$$\mathbf{y} = [11, 01, 10, 01, 00, 10, 11]$$

as shown in Fig. 4.30.

A better understanding of the operation of an encoder can be obtained using one or more of the following graphical representations [59–64]:

- State diagram.
- Tree diagram.
- Trellis diagram.

The definition and the way these encoder diagrams can be drawn and used are given in the literature [54–57].

For the decoding of convolutional codes, there are available several different approaches that can be grouped into two basic classes [59–64]:

- Sequential decoding (*Fano* algorithm).
- Maximum-likelihood decoding (*Viterbi* algorithm).

Sequential decoding was one of the earliest techniques developed for decoding convolutionally coded bit streams. It was first coined by Wosencraft and refined by Fano. Maximum-likelihood decoding is a good alternative methodology that is best



implemented by the Viterbi algorithm. For full presentations of the above decoding techniques, the reader is referred to the bibliography [54–64].

### 4.5.4 Fundamental Theorems of Information Theory

Here, four fundamental theorems of information theory will be reviewed and their roles and usefulness in practice will be explained. These theorems are the following:

- Nyquist–Shannon sampling theorem.
- Shannon’s source coding theorem.
- Shannon’s noisy channel coding and capacity theorem.
- Landau–Pollack bandwidth signal dimensionality theorem.

#### 4.5.4.1 Nyquist–Shannon Sampling Theorem

This theorem, which was first formulated by Harry Nyquist in 1928 [67] and formally proved by Shannon in 1949 [68], states: “*To be able to reconstruct perfectly a sampled analog signal  $x(t)$  from its sampled version  $x(kT)$ ,  $k = 0, 1, 2, \dots$  the sampling frequency  $1/T$  Hz (where  $T$  is the sampling period) must be greater than twice the highest frequency  $W$  of  $x(t)$ .*”

If the sampling frequency is less than this limit, then frequencies in the original analog signal that are greater than half the sampling frequency will be “*aliased*” and will appear in the resulting signal as lower frequencies. If the sampling frequency is exactly twice the highest frequency of the analog input, then “*phase mismatching*”, between the sampler and the signal, will distort the signal. Therefore, in practice, an analog low-pass filter must be used before the sampler to guarantee that no components with frequencies above the sampling frequency remain. This is called an “*anti-aliasing filter*” and must be very carefully designed, because a poor filter causes phase distortion and other effects. The minimum sampling frequency  $2W$  that permits exact reconstruction of the original signal is known as the *Nyquist frequency* (or *Nyquist rate*), and the time spacing between samples is known as “*Nyquist time interval*”. To express the above concepts mathematically, let a signal  $x(t)$  have a Fourier transform:

$$F[x(t)] = X(f) = 0 \quad \text{for } |f| > W.$$

Then,  $x(t)$  is completely determined by giving the value of the signal at a sequence of points, spaced  $T = 1/2W$  apart. The values  $x_k = x(k/2W) = x(kT)$ ,  $T =$  sampling period, are called the samples of  $x(t)$ .

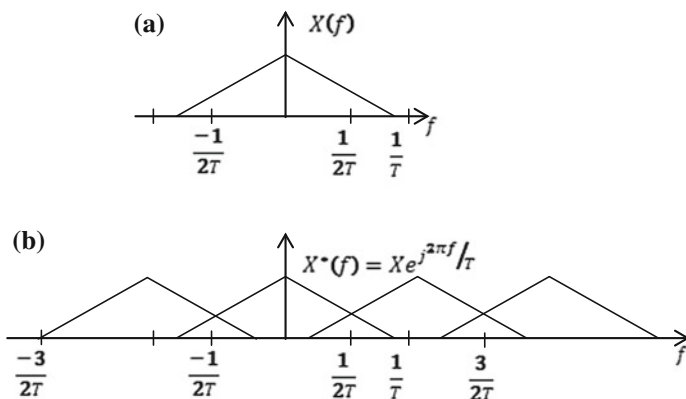
The sampled signal  $x^*(t)$  is expressed as the amplitude modulation via  $x(t)$  of a  $\delta$ -pulse (Dirac) train  $\sum_{k=-\infty}^{\infty} \delta(t - kT)$ , i.e.,

$$x^*(t) = x(t) \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

Since multiplication in the time domain is expressed by convolution “\*” in the frequency domain, we have

$$\begin{aligned} X^*(f) &= X(f) * \left[ \frac{1}{T} \sum_{j=-\infty}^{\infty} \delta(f - j/T) \right] \\ &= \frac{1}{T} \int_{-\infty}^{\infty} X(s) \sum_{j=-\infty}^{\infty} \delta(f - s - j/T) ds \\ &= \frac{1}{T} \sum_{j=-\infty}^{\infty} X(f - j/T) = X(e^{j2\pi f/T}) \end{aligned}$$

Figure 4.25 shows pictorially the Fourier transform  $X^*(f) = X(e^{j2\pi f/T})$  of the sampled version  $x^*(t)$  of a signal  $x(t)$  that has the Fourier transform  $X(f)$  [42]. In this case,  $X(f) \neq 0$ , outside the frequency region determined by the Nyquist frequency  $W = 1/2T$ . Therefore, “aliasing distortion” appears, which is due to the overlap of the various periodically repeated sections of  $X^*(f)$  in the frequency domain.



**Fig. 4.25** **a** Fourier transform of a continuous-time signal, **b** the corresponding Fourier transform of the sampled version with sampling frequency  $1/T$ . The overlaps indicate that in this case we have “aliasing effects”

### 4.5.4.2 Shannon's Source Coding Theorem

Consider a source that assigns to each outcome (sample)  $\mathbf{x}$  a binary word, which as we have seen is called the *code*. Here, the vector  $\mathbf{x}$  is an outcome of a discrete random vector  $\mathbf{X}$ . The *asymptotic equipartition theorem* (which is the basis for the entropy's interpretation) states that for  $\mathbf{x} \in \mathbf{X} = \{X_1, X_2, \dots, X_n\}$ , where  $X_i$  are independent trials of  $\mathbf{X}$ , as  $n \rightarrow \infty$  (i.e., asymptotically), there is a set of “*typical*” outcomes  $S$  for which

$$P_X(\mathbf{x}) = 2^{-nH(\mathbf{X})}, \quad \mathbf{x} \in S$$

and the total probability that the outcome is in  $S$  is almost one. Since the “*typical*” outcomes are all equiprobable, this means that there must be approximately  $2^{nH(\mathbf{X})}$  outcomes in  $S$ . As  $n$  becomes larger, this approximation becomes more accurate [42].

In the information source, when  $n$  is large, we can assign only the “*typical*” outcomes and omit the “*non-typical*” ones. By using  $nH(\mathbf{X})$ -bit code words, we are able to encode each of the  $2^{nH(\mathbf{X})}$  typical outcomes with a unique binary word, for an average of  $H(\mathbf{X})$  bits per component of the vector  $\mathbf{x}$ . Since  $H(\mathbf{X})$  represents the average information obtained from the observation, each outcome of  $\mathbf{X}$  needs an average of  $H(\mathbf{X})$  bits. This is true only for an average of  $n$  components, not for an individual component.

The source coding theorem states that [42]:

If a source can be modeled as repeated independent trials of a random variable  $\mathbf{X}$  at  $r$  trials per second, then the source can be encoded by a source coder into a bit stream with bit rate less than  $R + \epsilon$ , for any  $\epsilon > 0$ , where  $R = rH(\mathbf{X})$  is the so-called “rate of the source”.

This source coding theorem establishes the limits of data compression. Stated in another way, Shannon's source coding theorem says that [69]:

The minimum average number of bits,  $C$ , needed to encode  $n$  symbols (which are treated as  $n$  independent samples of a discrete random vector  $\mathbf{X}$  with probability  $p_{\mathbf{X}}(\mathbf{x})$  and entropy  $H(\mathbf{X})$ ) satisfies the relation:

$$H(\mathbf{X}) \leq C < H(\mathbf{X}) + 1/n.$$

In practice,  $p_{\mathbf{X}}(\mathbf{x})$  is not exactly available, but we only have an estimate  $q_{\mathbf{X}}(\mathbf{x})$  for use in the source coding process. In this case, the corresponding minimum  $C_q$  satisfies

$$H(\mathbf{X}) + KL(p||q) \leq C_q < H(\mathbf{X}) + KL(p||q) + 1/n$$

where  $KL(p||q)$  is the relative entropy (or the *Kullback–Leibler*: KL) divergence, defined as

$$KL(p||q) = \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \log[p_{\mathbf{X}}(\mathbf{x})/q_{\mathbf{X}}(\mathbf{x})] \geq 0$$

We note that  $KL(p||q)$  is the difference between the two probability distributions,  $p_{\mathbf{X}}(\mathbf{x})$  and  $q_{\mathbf{X}}(\mathbf{x})$ . Clearly,  $KL(p||q) = 0$  if and only if  $p_{\mathbf{X}}(\mathbf{x}) = q_{\mathbf{X}}(\mathbf{x})$ .

Constructing practical codes that are close to  $R$  is difficult. However, constructing good suboptimal codes is usually easy (see Sect. 4.5.3). In words, the source coding theorem (as expressed in the above two alternative ways) says that no source coding scheme can be better than the “*entropy of the source*”, or bit stream rate less than the “*source rate*”.

*Example* Let a binary random variable  $\mathbf{X}$  with set of values (alphabet)  $A = \{0, 1\}$ . Its entropy is equal to

$$H(\mathbf{X}) = -a \log_2 a - (1 - a) \log_2(1 - a)$$

where  $a$  denotes the probability of taking the value 1,  $a = p_{\mathbf{X}}(1)$ . The entropy  $H(\mathbf{X})$  is a function of  $a$ , i.e.,  $H = H(a)$  which has the graphical representation shown in Fig. 4.26 [42].

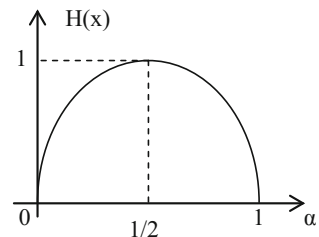
We observe that when  $a = 1 - a = 1/2$  (equiprobable values 0 and 1),  $H(\mathbf{X})$  takes a maximum, in agreement with the theory. The maximum of  $H(\mathbf{X})$  is one bit. Also,  $H(\mathbf{X})$  is zero when either  $a = 0$  ( $1 - a = 1$ ) or  $a = 1$ , again in agreement with the theory. The fact that, for  $a = 1 - a = 1/2$ , the value of the entropy is  $H(\mathbf{X}) = 1$  means that, to encode repeated samples of  $\mathbf{X}$ , we need on average one bit per sample (outcome). Here, this is also sufficient for each sample, not just on average, since the source is binary. A source coder that can achieve rate  $R$  transmits samples of  $\mathbf{X}$  unaltered (exactly). Now, suppose that  $a = 0.1$ , in which case

$$H(\mathbf{X}) = -0.1 \log_2(0.1) - 0.9 \log_2(0.9) = 0.47$$

This means that, to encode repeated samples of  $\mathbf{X}$ , we need 0.47 bits per outcome. Clearly, there are coding methods with average number of bits greater than 0.47 but less than unity.

In information theory, when dealing with a language, we speak about the entropy or rate of the language. For example, in the case of the English language, the *alphabet* consists of 26 letters (symbols) plus some additional symbols such as

**Fig. 4.26** Graphical representation of the entropy  $H(\mathbf{X})$  of a binary function in terms of the probability  $a = p_{\mathbf{X}}(1)$



space, comma, period, etc. These symbols can be treated as  $n$  independent samples of a random variable  $\mathbf{X}$  transmitted via the communication channel, each having a probability  $p_{\mathbf{X}}(\mathbf{x})$ , and entropy  $H(\mathbf{X}) = -\sum_{\bar{\mathbf{x}} \in \mathbf{X}} p_{\mathbf{X}}(\bar{\mathbf{x}}) \log p_{\mathbf{X}}(\bar{\mathbf{x}})$ . For example,  $p_{\mathbf{X}}(\mathbf{x} = \text{“s”})$  is much higher than  $p_{\mathbf{X}}(\mathbf{x} = \text{“z”})$ . To minimize the code length for the language, we assign shorter (more compressed) codes for symbols of higher probabilities (here, a shorter code for the letter “s” than that of the letter “z”). As length code for a symbol  $\mathbf{x}$  with  $p_{\mathbf{X}}(\mathbf{x})$ , we can use its “surprise”  $-\log p_{\mathbf{X}}(\mathbf{x})$ .

### 4.5.4.3 Shannon’s Noisy Channel Coding and Capacity Theorem

Shannon’s noisy channel coding and capacity theorem deal with the maximum possible efficiency of error-correcting codes (see Sect. 4.5.3) versus levels of data corruption and noise interference [10, 68]. This theorem establishes that a randomly designed error-correcting code is basically as good as the best possible code, and states [70]:

“The capacity of a discrete-memoryless channel is given by

$$C = \max_{p_{\mathbf{X}}(\mathbf{x})} \{I(X; Y) | p_{\mathbf{X}}(\mathbf{x})\} \quad (\text{bits/symbol})$$

where  $I(X; Y)$  is the *mutual information* between the channel input  $X$  and the output  $Y$ :

$$I(X; Y) = H(X) - H(X|Y)$$

If the transmission rate  $R$  is less than  $C$ , then, for any  $\varepsilon > 0$ , there exists a code with block length  $n$  large enough whose error probability is less than  $\varepsilon$ . When  $R > C$ , the error probability of any code with any block length is bounded away from zero. If the channel is used  $m$  times per second, then the channel capacity in bits per second is  $C' = mC$  ”.

*Example* Let a binary symmetric channel with cross-over probability 0.1. Then,  $C \approx 0.5$  bits per transmission. Thus, we can send reliably through the channel 0.4 bits/per channel. This can be achieved by taking (for example) 400 input information bits and map them into a code of length 1000 bits. The whole code is then transmitted through the channel, in which case 400 information bits can be decoded correctly, but 100 bits may be detected incorrectly. Now, let us consider a continuous-time additive white-Gaussian channel with fully uncorrelated signal and noise. In this case, the channel capacity is found to be [70]:

$$C = W \log_2 \left( 1 + \frac{S}{N_0 W} \right) \frac{\text{bits}}{\text{s}}$$

where  $S$  is the upper bound of the power of the input signal  $\mathbf{x}$  (measured in Watt),  $N_0/2$  is the Gaussian noise power spectral density, and  $W$  is the channel bandwidth

in Hz. Clearly,  $N_0W = N$  where  $N$  is the total noise or interference power over the bandwidth  $W$ ?

*Example* Using the above formula for the capacity of a Gaussian channel, we can compute the capacity of the voice band of a telephone channel. Typical values in this case are as follows:

$$W = 3000\text{Hz}, S/N = 1000 \text{ or } 10 \log_{10}(1000) = 30 \text{ db.}$$

Thus,  $C = 3000 \log_2(1 + 1000) \approx 30 \text{ kbits/s}$

This means that using this model one cannot design modems faster than 30 kbits/s. Because the signal-to-noise ratio is large, one can expect to be able to transmit  $C/W = 10 \text{ bits/s/Hz}$  across the telephone channel.

*Remarks*

- (a) If the noise and signal are not fully uncorrelated (as in the case of non-white additive noise), the signal-to-noise ratio  $S/N$  is not constant with frequency over the bandwidth. In this case, we can assume that the channel is modeled as a multitude of narrow independent Gaussian channels in parallel, and so  $C$  is given by

$$C = \int_0^W \log_2 \left( 1 + \frac{S(f)}{N(f)} \right) df$$

where  $S(f)$  and  $N(f)$  are the signal and noise power spectrums, respectively, which are functions of the frequency  $f$ .

- (b) For large  $S/N$  ratio (i.e.,  $S/N \gg 1$ ), we get  $C \approx 0.332 W (S/N, \text{ in db})$ , where  $S/N \text{ in db} = 10 \log_{10}(S/N)$ .
- (c) For very small  $S/N$  (i.e.,  $S/N \ll 1$ ), we obtain  $C \approx 1.44W(S/N) = 1.44W(S/N_0W) = 1.44(S/N_0)$ , i.e., the channel capacity in this case is (approximately) independent of the noise bandwidth.

#### 4.5.4.4 Landau–Pollack Bandwidth Signal Dimensionality Theorem

This theorem is a consequence of the Nyquist–Shannon sampling theorem and states [42]: “A signal cannot be both band-limited and time-limited.”

Actually, a band-limited signal is not time-limited, because its energy cannot be entirely restricted to any finite interval of time. Similarly, a time-limited function is not band-limited because its energy cannot be totally confined to a finite band of frequencies. However, one can assume that a band-limited signal is approximately time-limited, and a time-limited signal is approximately band-limited.

Quantitatively, the Landau–Pollack theorem says: “The signal space of all finite-energy signals is infinite dimensional, but the subset of such signals that are band limited to  $W$  Hz and approximately time limited to  $[0, t_0]$ ,  $t_0$  sufficiently large, is approximately finite dimensional with dimension  $2Wt_0 + 1$ .”

This means that there exists a set of  $2Wt_0 + 1$  orthonormal basis functions  $\Psi_i(t)$ , such that for any finite-energy signal  $x(t)$  with energy  $E_x$  that is band limited to  $|f| < W$ , for any constant  $\varepsilon$  with  $0 < \varepsilon < 1$ , and for any  $t_0$  sufficiently large, the following relations hold:

$$\int_0^{t_0} |x(t)|^2 dt > E(1 - \varepsilon)$$

$$\int_{-\infty}^{\infty} \left\{ x(t) - \sum_{i=0}^{2Wt_0} x_i \Psi_i(t) \right\}^2 dt < 12\varepsilon E_x$$

where  $x_i$ ,  $i = 0, 1, \dots, 2Wt_0$  are the  $2Wt_0 + 1$  expansion coefficients. In other words, if outside the interval  $[0, t_0]$ , the maximum fraction of the band-limited-signal’s energy is  $\varepsilon$ , then this signal can be approximately expressed by a linear combination of a set of  $2Wt_0 + 1$  orthonormal basis functions, with an energy error less than a fraction  $12\varepsilon$  of the signal’s energy. As  $t_0$  gets larger, the fraction of energy outside the dimension  $2Wt_0 + 1$  of this signal subspace becomes smaller.

To verify that the Landau–Pollack theorem is a consequence of the Nyquist–Shannon sampling theorem, we reason as follows [71]. Suppose that a signal is both band-limited and time-limited exists and that this signal is sampled at a frequency greater than the Nyquist frequency. These finitely many time-domain coefficients should represent the entire signal. In the same way, the entire spectrum of the band-limited signal should be represented through the finitely many time-domain coefficients resulting from the signal’s sampling. Mathematically, this would require that a (trigonometric) polynomial can have infinitely many zeros, since the band-limited signal must be zero on an interval beyond a critical frequency that has infinitely many points. But we know from the fundamental theorem of algebra that a polynomial cannot have more zeros than its order. This contradiction is due to our incorrect assumption that a signal that is both band-limited and time-limited exists.

### 4.5.5 Jayne’s Maximum Entropy Principle

The *maximum entropy principle* (**MEP**) was first formulated by Jaynes [26] and is actually a generic optimization problem, namely,

“Find a probability distribution  $p_X(x)$ ,  $x \in X$ , that maximizes the Shannon’s entropy  $H(X)$  subject to a set of given constraints  $c_1, c_2, \dots, c_n$  which express

partial information about the probability distribution  $p_X(x)$  sought, and also the typical axioms (constraints) of probability theory.”

The most common constraints in practice are expected (mean) values of one or more random variables and/or random functions, or several marginal probability distributions of an unknown joint distribution. More specifically, this optimization problem is formulated as follows:

Suppose we are given a random variable  $X$  taking the values  $\{x_1, x_2, \dots, x_n\}$  where  $n$  can be finite or infinite, and the mean (average) values of various functions  $f_1(X), f_2(X), \dots, f_m(X)$  where  $m < n$ . The problem is to determine the probability assignment  $p_i = p(x_i)$  which satisfies the given data (constraints):

$$\sum_{i=1}^n p_i = 1, p_i \geq 0$$

$$\sum_{i=1}^n p_i f_k(x_i) = E[f_k(X)] = \Xi_k, \quad k = 1, 2, \dots, m$$

and maximizes Shannon’s entropy:

$$H(X) = - \sum_{i=1}^n p_i \log p_i$$

The solution to this mathematical problem can be determined using the well-known method of Lagrange multipliers, which however has the drawback that it does not make clear whether a true (global) maximum of  $H(X)$  has been obtained. Here, without loss of generality, we will develop the solution for the case  $f_k(X) = x_k$ , in which the constraints about  $f_k(X)$  are reduced to

$$\sum_{k=1}^n p_k x_k = E(X) = \zeta$$

Also, for simplicity, the logarithm  $\log$  in  $H(X)$  is assumed to be the natural logarithm ( $\log_e = \ln$ ) [29]. We start the solution, by converting the above-constrained optimization problem into the equivalent unconstrained optimization problem, with respect to  $p_k (k = 1, 2, \dots, n)$  and the Lagrange multipliers  $\lambda$  and  $\mu$ , of the Lagrange function:

$$L = - \sum_{k=1}^n p_k \ln p_k - \lambda \left( \sum_{k=1}^n p_k - 1 \right) - \mu \left( \sum_{k=1}^n p_k x_k - \zeta \right)$$

Then, we write down the associated canonical (partial derivative) equations:



$$\begin{aligned} \partial L / \partial p_k &= -\ln p_k - 1 - \lambda - \mu x_k = 0, \quad k = 1, 2, \dots, n \\ \partial L / \partial \lambda &= 1 - \sum_{k=1}^n p_k = 0 \\ \partial L / \partial \mu &= \zeta - \sum_{k=1}^n p_k x_k = 0 \end{aligned}$$

The first  $n$  equations can be rewritten as

$$\begin{aligned} p_1 &= e^{-1-\lambda-\mu x_1} = e^{-(1+\lambda)} e^{-\mu x_1} \\ p_2 &= e^{-1-\lambda-\mu x_2} = e^{-(1+\lambda)} e^{-\mu x_2} \\ &\quad \text{-----} \\ &\quad \text{-----} \\ p_n &= e^{-1-\lambda-\mu x_n} = e^{-(1+\lambda)} e^{-\mu x_n} \end{aligned}$$

which, if divided by the sum  $p_1 + p_2 + \dots + p_n = 1$ , gives

$$p_k = e^{-\mu x_k} / \sum_{i=1}^n e^{-\mu x_i}, \quad k = 1, 2, \dots, n$$

Now, multiplying both sides of the  $p_k$  equation by  $x_k$  and adding, we get

$$\zeta = \sum_{k=1}^n x_k e^{-\mu x_k} / \sum_{k=1}^n e^{-\mu x_k}$$

from which we obtain

$$\sum_{k=1}^n x_k e^{-\mu x_k} - \zeta \sum_{k=1}^n e^{-\mu x_k}$$

Finally, multiplying this equation by  $e^{\mu \zeta}$  gives

$$\sum_{k=1}^n (x_k - \zeta) e^{-\mu(x_k - \zeta)} = 0$$

This is a nonlinear equation (with respect to  $\mu$ ) and can be solved numerically for  $\mu$ . Introducing this value of  $\mu$  into the  $p_k$  equations ( $k = 1, 2, \dots, n$ ), we find the desired probabilities  $p_k$ ,  $k = 1, 2, \dots, n$ .

As a simple illustration of the *maximum entropy principle*, let us consider an unbiased (“honest”) die. Here,  $x_k = k$  ( $k = 1, 2, 3, \dots, 6$ ), and so  $\zeta = (1 + 2 + 3 + 4 + 5 + 6) / 6 = 21 / 6 = 3.5$ . The last equation here becomes

$$\begin{aligned}
 & - 2.5e^{2.5\mu} - 1.5e^{1.5\mu} - 0.5e^{0.5\mu} \\
 & + 0.5e^{-0.5\mu} + 1.5e^{-1.5\mu} + 2.5e^{-2.5\mu} = 0
 \end{aligned}$$

The solution of this equation is  $\mu = 0$ , and so we find

$$p_k = 1/6 \text{ for } k = 1, 2, \dots, n$$

*Remark* In this example, the mean value  $\zeta = E(X)$  was known and used as a constraint on the corresponding probabilities. If  $E(X)$  were not known, then no constraint would be on the probabilities. In this case, the maximum entropy principle would give equal probabilities (the uniform or equiprobable distribution  $p_x(x)$ ), which are the only probabilities for which the entropy takes its absolute maximum value (see the properties of the entropy in Sect. 4.5.2.2). If the probabilities are subject to constraints, the MEP principle gives a maximum entropy of limited value, which is usually smaller than the entropy of the uniform distribution.

## 4.6 Concluding Remarks

This chapter is the first of three chapters that deal with the “*information pillar*” of human life and society. The topics that have been covered include the general definition of the information concept, the historical landmarks of its manifestations, namely, communication (speech, writing, printing, telegraph, telephone, computing, and computers), information theory, computer networks, multimedia, telematics, and informatics, and a technical review of communication and information theory.

Information storage, flow, and processing are inherent processes in nature and living organisms. Organisms do not live separately from other organisms but interact with each other within the ecosystem in which they exist. Many of the existing technological communication and modulation/demodulation models were inspired by life on Earth, or they can be used to model and better understand biological communication models. The exploration of the physical and biological world provides unexpected surprises and discoveries that increase steadily our information entropy.

In a similar way, modern information transmission and communication techniques are affecting and will continue to increasingly affect the social and economic/business activity of people over coming decades. An example of this is electronic commerce (e-commerce), which reduces substantially the sales and operational costs in comparison with the traditional stores. The Internet expands and moves e-commerce into an open and global market with big changes in the market structure, size, and players.

More generally, inter-networked information technology enables people to act in their own self-interests and affect the experiences of other people. Overall, it is anticipated that the prospect of *return-on-investment (ROI)* research in communications, information theory, and information technology is very promising, with directly measurable results in economic strength and social benefits.

## References

1. R. Capurro, *Information: A Contribution to the Foundation of the Concept of Information Based on Its Etymology and in the History of Ideas* (Saur, Munich, 1978). <http://www.capurro.de/info.html>
2. R. Capurro, B. Hjørland, The concept of information. *Ann. Rev. Inf. Sci. Technol.* **37**, 343–411 (2003). (B. Cronin, ed., Chap. 8)
3. R.J. Bogdan, *Ground for Cognition: How Goal-Guided Behavior Shapes the Mind* (Lawrence Earlbaum, Hillsdale, N.J., 1994)
4. G. Ropohl, The concept of information in the framework of the culturalist struggle. *Ethik und Sozialwissenschaften* **1**, 1–12 (2001)
5. W. Hofkirchner (ed.), The quest for a unified theory of information, in *Proceedings, 2nd International Conference on the Foundations of Information Science* (Gordon and Breach, Amsterdam, 1999)
6. F. Machlup, U. Mansfield (eds.), *The Study of Information: Interdisciplinary Messages* (Wiley, New York, 1983)
7. R. Capurro, On the genealogy of information, in *Information: New Questions to a Multidisciplinary Concept*, ed. by K. Kornwachs, K. Jacoby (Akademie Verlag, Berlin, 1996), <http://www.capurro.de/cottinf.hbm>
8. R.V.L. Hartley, Transmission of Information. *Bell Syst. Tech. J.* **7**, 535–563 (1928)
9. C. Shannon, A mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423, 623–656 (1948)
10. C. Shannon, W. Weaver, *The Mathematical Theory of Communication* (The University of Illinois Press, Urbana, IL, 1949) (also 1972)
11. H. Titze, *Is Information a Principle? (Ist Information ein Prinzip?)* (Hain, Meisenheim, 1971)
12. T. Stonier, Towards a new theory of information. *J. Inf. Sci.* **17**, 257–263 (1991)
13. T. Stonier, Information as a basic property of the universe. *Biosystems* **38**, 135–140 (1997)
14. P.T. de Chardin, *The Future of Man* (Harper and Row, New York, 1964) (originally published in 1959)
15. S. Augarten, *BIT by BIT: An Illustrated History of Computers* (Ticknor and Fields, New York, 1984)
16. R. Moreau, *The Computer Comes of Age: The People, the Hardware, and the Software* (Trans. by J. Howlett) (MIT Press, 1984)
17. J.G. Butler, Information Technology History-Outline, <http://www.tcf.ua.edu/AZ/ITHistory.Outline.htm>
18. The Antikythera Mechanism Research Project, [www.antikythera-mechanism.gr/project/general/the-project.html](http://www.antikythera-mechanism.gr/project/general/the-project.html)
19. Information Theory, <http://www.eoht.info/page/information+theory>
20. History of Information Theory, [http://en.wikipedia.org/wiki/History\\_of\\_information\\_theory](http://en.wikipedia.org/wiki/History_of_information_theory)
21. L. Szilard, On the decrease in entropy in a thermodynamic system by the intervention of intelligent beings. *Z. Angew. Phys.* **53**, 840–856 (1929)
22. L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956)
23. R.M. Fano, *Transmission of Information* (MIT Press, Cambridge, MA, 1949)

24. E.T. Jaynes, Information theory and statistical mechanics: (I). Phys. Rev. **106**, 620 (1957); (II), *ibid.*, **108**, 171 (1957)
25. M. Tribus, *Thermostatistics and Thermodynamics: An Introduction to Energy, Information, and States of Matter* (Van Nostrand Reinhold, New York, 1961)
26. E.T. Jaynes, On the rationale of maximum entropy methods. Proc. IEEE **70**(9), 939–952 (1982)
27. G.N. Saridis, Entropy Formulation for Optimal and Adaptive Control. IEEE Trans. Autom. Control **33**(8), 713–721 (1988)
28. G.N. Saridis, *Entropy in Control Engineering* (World Scientific, London, 2001)
29. G.J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory* (Wiley-Interscience, Hoboken, NJ, 2006)
30. A. Ben-Naim, *Farewell to Entropy: Statistical Thermodynamics Based on Information* (World Scientific, London, 2007)
31. Wikipedia: Timeline of Information Theory, [http://en.wikipedia.org/wiki/Timeline\\_of\\_information\\_theory](http://en.wikipedia.org/wiki/Timeline_of_information_theory)
32. History of Computer Networks (Hammed Haddadi). <http://www.ee.ucl.ac.uk/~hamed/transfer/thesis/node10.html>
33. Computer Networking History-Networking Computer Tips, <http://www.onlinecomputertips.com>
34. Internet World Stats, <http://www.internetworldstats.com/stats.htm>
35. R. Albarino, Goldstein's Light Works at Southampton, Variety **213**(12) (1966) (Aug 10)
36. D. Marshall, History of multimedia systems, <http://www.cs.cf.ac.uk/Dave/Multimedia/node8.html>
37. History of Multimedia (Calgary University) <http://www.acs.ucalgary.ca/~edtech/688/hist.html>
38. S. Nora, A. Minc, *The Computerization of Society. A Report to the President of France* (MIT Press, Cambridge, MA, 1980) (translation of *L'Informatisation de la Société*. Rapport à M. le Président de la République, La Documentation Française, Paris, 1978)
39. The History of Telematics. Telematique, <http://www.telematique.eu/telematics/>
40. Telematics Valley, <http://telematicsvalley.org/>
41. C.D. Mortensen, *Communication: The Study of Human Communication* (McGraw-Hill, New York, 1972)
42. J.R. Barry, E.A. Lee, D.G. Messerschmidt, *Digital Communication* (Kluwer, Boston, 2004)
43. D.K. Berlo, *The Process of Communication* (Holt Rinehart and Winston, New York, 1960)
44. W. Schramm, How Communication Works, in *The Process and Effects of Communication*, ed. by W. Schramm, (University of Illinois Press, Urbana, Illinois, 1954)
45. Communication models, <http://www.shkaminski.com>
46. B.G. Korenev, *Bessel Functions and their Applications* (CRC Press, Boca Raton, FL, 2002)
47. E. Weisstein, Bessel Function of the First Kind, Wolfram Research, <http://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html>
48. D.E. Foster, S.W. Sceleay, Automatic tuning, simplified circuits, and design practice. Proc. IRE **25**(3), 289–313 (1937) (part 1)
49. FM Detectors (Discriminators), <http://www.ycars.org/EFRA/Module%20B/fmdet.htm>
50. Amplitude, Frequency and Phase Modulation, <http://www.aubraux.com/dsp/dsp-modulation-all.php>
51. A. Schwope, Modulation, [http://www.andreas-schwope.de/ASIC\\_s/Schnittstellen/Data\\_Lines/body\\_modulation.html](http://www.andreas-schwope.de/ASIC_s/Schnittstellen/Data_Lines/body_modulation.html)
52. J. Proakis, M. Salehi, *Digital Communications* (Mc Graw-Hill, New York, 2007)
53. H.L. Resnikoff, *The Illusion of Reality* (Springer, Berlin, 1989)
54. W.W. Reterson, E.J. Weldon Jr., *Error Correcting Codes*, 2nd edn. (The MIT Press, Cambridge, MA, 1972)
55. V. Press, *Introduction to the Theory of Error-Correcting Codes*, 3rd edn. (Wiley, New York, 1998)

56. W. Huffman, V. Pless, *Fundamentals of Error-Correcting Codes* (Cambridge University Press, Cambridge, 2003)
57. J.H. Van Lint, *Introduction to Coding Theory* (Springer, Berlin, 1981)
58. N.J.A. Sloane, *A Short Course on Error Correcting Codes*, vol. 188, CISM Courses and Lectures (Springer, Berlin, 1975)
59. A.M. Michelson, A.H. Levesque, *Error Control Techniques for Digital Communication* (Wiley, New York, 1985)
60. C. Schlegel, L. Perez, *Trellis Coding* (IEEE Press, New York, 1997)
61. Error Correcting Codes, <http://www.eie.polyu.edu.hk/~enmzwang/adc/1-notes/node3.html>
62. C. Langton (ed.), Convolution Codes, <http://www.complextoreal.com/convo.htm>
63. Y.S. Han, *Introduction to Binary Convolution Codes*, Graduate Institute of Communication Engineering, National Taipei University, Taiwan. [http://web.ntpu.edu.tw/~shan/intro\\_in\\_cod.pdf](http://web.ntpu.edu.tw/~shan/intro_in_cod.pdf)
64. C. Fleming, A Tutorial on Convolutional Coding with Viterbi Decoding. <http://home.netcom.com/~chip.f/viterbi/tutorial.html>
65. F.J. Mc Williams, N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1977)
66. R. Yates, A Coding Theory Tutorial, <http://www.digitalsignallabs.com/tutorial.pdf>
67. H. Nyquist, Certain topics in telegraph transmission theory. Trans. AIEE **47**, 617 (1928)
68. C. Shannon, Communication in the presence of noise. Proc. IRE **37**(1), 10–21 (1949). <http://www.stanford.edu/class/ee104/shannonpaper.pdf>
69. R. Wang, Shannon's source coding theorem (Dec 10, 2009). <http://fourier.eng.hmc.edu/e161/lectures/compression/node7.html>
70. B. Aazhang, Shannon's Noisy Channel Coding Theorem (May 19, 2004). <http://cnx.org/content/m10180/latest/>
71. Nyquist-Shannon Sampling Theorem Signal Frequency Mhz, <http://www.economicexpert.com/a/Nyquist:Shannon:sampling:theorem.htm>