ELECTRONIC CONCEPTS

AN INTRODUCTION

JERROLD H. KRENZ
University of Colorado, Boulder

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 2000

Printed in the United States of America

Typeface Sabon 10/13 pt. and Futura *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Krenz, Jerrold H., 1934-
Electronic concepts : an introduction / Jerrold H. Krenz.
p. cm.
Includes bibliographical references (p.
ISBN 0-521-66282-6 (hb.)
1. Electronics. 2. Electronic circuits – Computer simulation.
I. Title.
TK7816.K73 1999
621.381 – dc21 99-30407
CIP

ISBN 0 521 66282 6 hardback

CONTENTS

ELECTRONIC SYSTEMS: A CENTURY OF PROGRESS

Our daily lives are shaped by electronic systems. In the home we have a myriad of electronic accessories: radios, TVs, VCRs, hi-fis, camcorders, cassette and CD players, telephone answering machines, microwave ovens, and personal computers. Not so obvious but just as much a part of our lives are sophisticated electronic controls such as the microprocessor engine control of our car. We utilize a telephone system that functions with electronic devices to amplify and transfer telephone signals. Our conversations are carried around the world using a combination of microwave or fiber-optic links and satellites. Electronic radar systems are relied on for a safe flight from one airport to the next, and electronic sensors and computers "fly" a modern jet airplane. Modern medical practice depends on extremely complex diagnostic and monitoring electronic systems. Moreover, the commercial and industrial sectors could no longer function without electronic communications and information processing systems. The video monitor is a pervasive reminder of the new electronic world.

For better and at times for worse, electronics has changed our lives. Although we are in constant touch with what is happening around the world, we are also at the peril of weapons of unimaginable destructive power that rely on electronic developments. An understanding of electronics is imperative not only for designing and using electronic systems but for directing the evolution of electronic systems so that they serve to improve the human condition.

It has been stated that to move forward we must know where we have been. The 20th century is the era of electronics – it was only after 1900 that the devices we now describe as electronic appeared. The use of the term *electronics* in the current sense did not occur until 1930 (Süsskind 1966). This introductory chapter starts with a very brief overview of electronic devices and is followed by a discussion of wireless systems: radio. The first application of electronic devices, the vacuum tube diode invented in 1904 and the triode invented in 1906, was for radio receivers. Radio communications was not only nearly a decade old at the time the tube was invented, but most of the systems of the first decade of the 1900s did not use tubes. The vacuum tube, without exaggeration, can be described as having revolutionized radio communications, resulting in the generation of coherent transmitting signals and highly sensitive and selective receivers. The vacuum tube, following its first telephone use in 1913, also became an important component of telephone systems. With vacuum tube amplifiers and multiplexing circuits, long-distance telephone service greatly expanded. With the development of digital systems made possible by the transistor and integrated circuits in the latter half of the 20th century, telephone switching and transmission systems were again significantly improved.

The development of electronic devices, on the one hand, depended on a knowledge of basic physical principles: the behavior of electrons in a vacuum and the interaction of electrons with matter. On the other hand, electron devices were frequently developed to fulfill perceived needs. The characteristics of electronic devices dictated those applications that could be realized. Television, discussed in Section 1.4, illustrates the interrelatedness of the development of electronic devices and circuits with a particular application. An analog television system was developed in the 1930s and was commercially introduced in the late 1940s. Over the rest of the 20th century, television was based on this analog system, and the only enhancement was the introduction of a subcarrier for color information. At the close of the 20th century, a digital system, totally different, and therefore incompatible with the analog system, was developed. Although this digital system, from a transmission perspective, is considerably more efficient, the signal processing required is very complex. Without the development of very-large-scale integrated (VLSI) circuits during the 1980s that could do the encoding and decoding, digital TV would not have been possible.

The electromagnetic spectrum (Section 1.5) is used for a variety of radio, TV, and other communications services. Although early radar systems can be traced back to the 1930s, it was the impetus of World War II that resulted in a rapid development of this technology. New electronic devices capable of transmitting and detecting extremely high-frequency signals ($f > 1000$ MHz) were invented. Communications satellites, first launched in the 1960s, also relied on these extremely high-frequency (microwave) devices.

Digital electronic circuits have revolutionized computing. Early computers, until about the mid-1960s, relied on vacuum tube circuits. These computers, from today's perspective, not only had minuscule processing capabilities, but, owing to the limited reliability of vacuum tubes, were frequently down. Solid-state devices resulted not only in a tremendous improvement in reliability but made possible machines with much greater computing capabilities. With ultra-large-scale integrated circuits, desktop computers emerged with a computing capability that a decade earlier was available only in large mainframe machines.

Needless to say, electronic devices and circuits have become common for many applications in addition to those discussed. Power electronics is dependent on electronic switching devices and circuits. Frequency and voltage transformations, as well as alternating-to-direct-current and direct-to-alternating-current conversions can often be efficiently achieved using electronic systems. In medical electronics, a variety of electronic sensing circuits have been developed along with computer systems to process and display the data. Furthermore, electronic systems, such as heart pacemakers, have been perfected to augment body functions. Electronic sensing and control systems dependent on simple microprocessors are now used in applications ranging from programmable thermostats to automobile ignition and fuel systems. More complex sensing and control systems involving large computing capabilities are used for automated manufacturing systems. Although it is beyond this introductory chapter to discuss these and other applications, it should be recognized that similar electronic devices and circuits are often used by these different systems. A knowledge of basic concepts, the subject of this text, is a prerequisite for understanding both the simplest and the most esoteric of electronic systems.

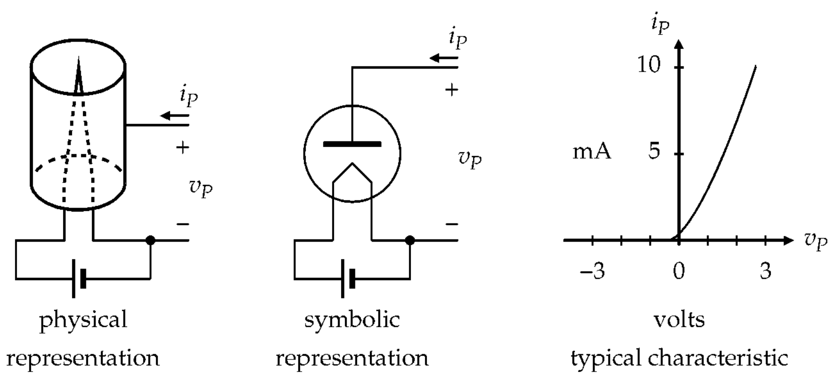
1.1 ELECTRONIC DEVICES: AN OVERVIEW

The thermionic valve or vacuum tube was developed in Great Britain by Sir John Ambrose Fleming (Pierce 1950; Shiers 1969). This tube relied on what is known as the *Edison effect*, a current being produced by the hot filament of a light bulb. Fleming, through a series of experiments with bulbs having an electrode near the hot filament, deduced that this current was due to negative electric charges. We now understand the current to be due to electrons emitted by the hot filament that are collected by the electrode. To the extent that only electrons are responsible for this current, the current to the electrode is only in one direction; in a high-vacuum tube, a current corresponding to the movement of positive charges does not occur.

THE DIODE

Fleming's valve consisted of a hot filament (corresponding to the incandescent filament of a light bulb) heated by a current produced by an external battery. The emitted electrons were then collected by a plate surrounding the filament (Figure 1.1). Even though the physical current is that due to electrons traveling from the filament to the plate, the plate current i_P, is, by convention, a positive quantity because a current is defined in terms of the movement of hypothetical positive charges. A positive plate voltage v_P attracts electrons, thus increasing the current, whereas a negative plate voltage repels electrons, yielding either

Figure 1.1: Vacuum tube diode and typical characteristic.



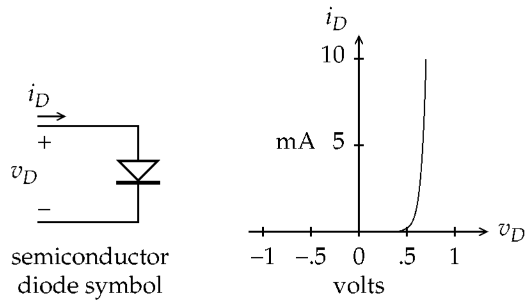


Figure 1.2: Semiconductor diode and a typical characteristic.

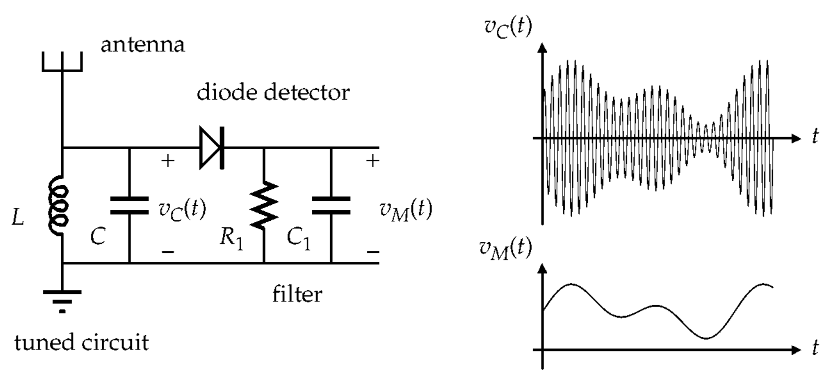
a very small or zero current. This nonlinear effect results in a current in only one direction ($i_P \geq 0$). For significant negative voltages, the current of a well-evacuated tube is essentially zero.

At about the same time that Fleming introduced his vacuum tube, Greenleaf W. Pickard was experimenting with a point-contact semiconductor detector (Douglas 1981). This device may be considered the precursor of modern solid-state devices. In addition to the detector using silicon developed by Pickard, a similar detector using Carborundum was developed by Henry H. C. Dunwoody in 1906. Point-contact diodes were extensively used until the junction semiconductor diode was introduced in the 1950s.

A semiconductor diode has a nonlinear characteristic, as does the vacuum tube (Figure 1.2). The current of the diode i_D increases very rapidly with diode voltage v_D (for an ideal semiconductor diode it may be shown that the current has an exponential dependence on the diode voltage). The rectification property of a diode, which allows a current in only one direction, was first used for the detection of radio signals. The detection problem provided the impetus for the development of vacuum tube and semiconductor diodes. Represented in Figure 1.3 is a basic radio receiver with a typical amplitude-modulated carrier signal. Although carrier frequencies of 50 to 100 kHz were common for early communications systems, the present radio broadcast band consists of signals with carrier frequencies of 540 to 1600 kHz. For an *on–off* system (continuous wave or CW), the carrier is simply keyed on and off to form a pattern of dots and dashes. However, for amplitude modulation (AM), the amplitude of the carrier signal is varied in accordance with the modulating signal; for example, that of a voice signal produced with a microphone.

It should be noted that the period corresponding to the carrier frequency is generally much smaller than that associated with the time scale over which appreciable variations in the modulating signal occur. In a radio receiver, the

Figure 1.3: An elementary diode radio detector.



energy received by the antenna is coupled to the tuned circuit which, ideally, excludes all other signals with different carrier frequencies. A diode rectifier is then used to convert the carrier signal $v_C(t)$ to a signal with a single polarity. For the circuit shown, the capacitor C_1 tends to smooth the detected signal. Without the capacitor, a signal similar to the top half of $v_C(t)$ would result.

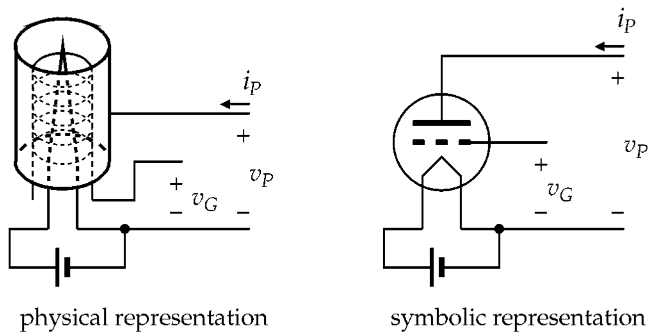


Figure 1.4: A triode vacuum tube.

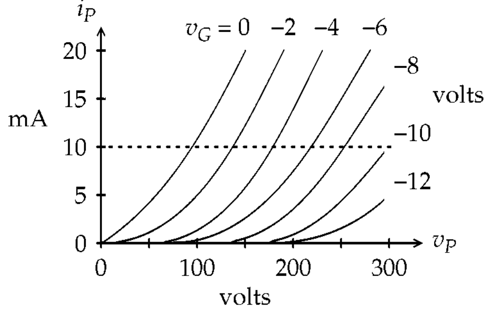
THE VACUUM TRIODE

The next significant development that, in effect, ushered in the electronics age, was Lee De Forest's addition of a control electrode or grid to Fleming's vacuum diode. This resulted in the triode vacuum tube. A sketch of the physical device, which is referred to as a triode because it has three elements, is presented in Figure 1.4. The third element, the grid, is a cagelike wire structure surrounding the filament of the tube. An externally applied grid potential regulates the plate current of the tube.

For normal operation, the grid is at a negative potential (relative to that of the filament), which tends to repel electrons emitted by the hot filament. The more negative the grid potential v_G, the smaller the plate current i_P for a given plate voltage v_P (Figure 1.5). Because electrons are repelled by a negative grid potential, the grid current is essentially zero. (The exceedingly small grid current that does occur is due to positive ions produced by ionizing electron collisions with the air molecules of the imperfect vacuum. Although the grid current of De Forest's early tube may have been significant, those of later tubes with good vacuums were truly negligible.) As a result of this essentially zero grid current, the power utilized by the grid circuit is extremely close to zero. Herein lies the worth of the triode vacuum tube. Its plate current and voltage are not only controlled by the grid voltage, but essentially zero power is required to do the controlling. It is not a perpetual-motion device (a power source is required for the plate circuit) but, for many applications, it is the next best thing!

To illustrate the utility of a vacuum tube triode, consider the typical characteristic of Figure 1.5 and suppose that a constant current source of 10 mA is connected between the filament and plate of the tube ($i_P = 10$ mA). For a particular value of grid voltage, the resultant plate voltage corresponds to the intersection of the curve corresponding to that grid voltage with the 10-mA coordinate (shown as a

Figure 1.5: The plate characteristic of a typical triode vacuum tube.



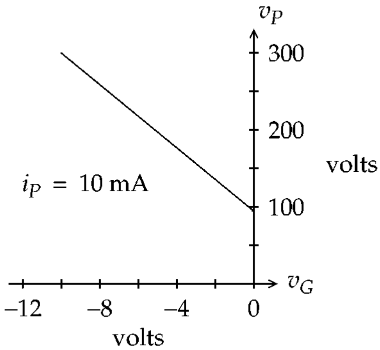


Figure 1.6: The transfer characteristic of the triode of Figure 1.5.

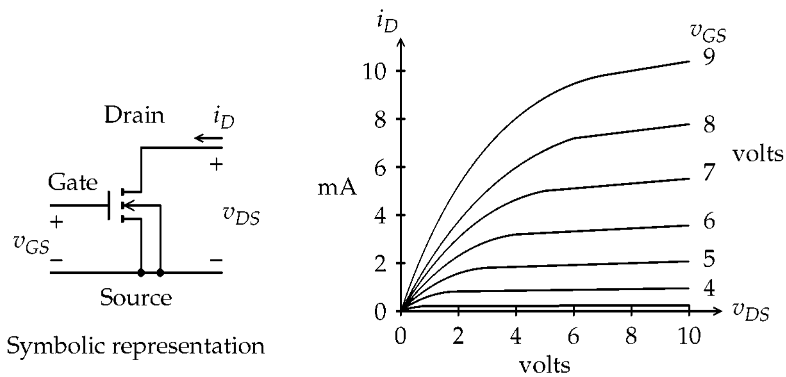
dashed line in Figure 1.5). A grid voltage of −4 V, for example, results in a plate voltage of 180 V; a grid voltage of −6 V in a plate voltage of 220 V, and so forth. The transfer characteristic of Figure 1.6 is thus obtained. Of particular importance is that a relatively small change in grid voltage results in a fairly large change in plate voltage. The slope of the characteristic of Figure 1.6 is approximately −20. This implies that a 1-V change in v_G results in a change of −20 V in v_P. The minus sign signifies that an increase in v_G results in a decrease in v_P. This circuit therefore has a voltage gain with a magnitude of approximately 20.

The first triode vacuum tube of De Forest was used to detect radio signals (in place of the diode of Figure 1.3); it was initially described as an oscillation valve. However, because vacuum tube triodes have the ability to amplify as well as to detect radio signals, tubes were soon used for a multitude of applications, including the generation of high-frequency radio signals.

THE TRANSISTOR AND INTEGRATED CIRCUITS

Solid-state devices, transistors, have replaced vacuum tubes for most, but not all, electronic applications. The symbolic representation and typical characteristic of a modern metal-oxide semiconductor field-effect transistor (MOSFET) are given in Figure 1.7. For the device shown, free electrons from the source of the MOSFET semiconductor device flow to its drain. In a manner analogous to that of the grid of the vacuum tube, the free-electron current is controlled by the gate potential of the MOSFET device. The gate current, like the grid current of a triode, is essentially zero. The free electrons, however, are produced by a doped semiconductor rather than by a hot filament, thus resulting in a much more efficient device. Furthermore, the voltage levels required for a typical MOSFET application are considerably smaller than those of a typical triode vacuum tube circuit.

Figure 1.7: The metal-oxide semiconductor field-effect transistor (MOSFET).



In addition to MOSFET devices, the bipolar junction transistor (BJT) is also extensively used in modern electronic circuits. Germanium bipolar junction transistors were developed shortly after the invention of the point-contact transistor in 1948. With the development of silicon processing techniques during the 1950s, germanium and silicon transistors tended to replace vacuum tubes for most applications by the 1960s. It was, however, the introduction of the integrated circuit, a single semiconductor wafer initially limited to a few tens of transistors, that has had the most profound effect on electronic systems. This effect has been characterized by some as revolutionary (Noyce 1977).

Vacuum tubes generally consisted of only one, two, or possibly three electronic devices enclosed by a single glass envelope. These tube circuits were generally mounted on a metal chassis that had sockets relying on spring contacts to hold the vacuum tubes. This permitted vacuum tubes to be readily replaced – an all-too-frequent need. Connections between the sockets and other components were achieved through hand-soldered wires. Small components, such as resistors and capacitors, were often supported directly by their leads while forming connections between components.

Even the earliest commercially produced transistors, introduced during the 1950s, were considerably more reliable than the vacuum tubes they replaced. Hence, transistors could be wired directly into a circuit, thereby eliminating the need for sockets. This led to the printed circuit board utilizing copper foil conductors bonded to a phenolic base. Transistors, as well as other components, were mounted directly on the printed circuit board, and a dip-type soldering process was used for electrical connections to the copper foil. Because transistors are much smaller than vacuum tubes and tend to dissipate considerably less power, a much higher density of components was possible.

A batch process was soon developed in which several transistors were simultaneously fabricated on a single semiconductor wafer. The wafer was then cut to obtain individual transistors, leads were attached, and the transistors were encapsulated in a package suitable for their application. During the assembly process, individual transistors were tested, and faulty ones were discarded. With the improvement of processing techniques, the yield of well-functioning devices greatly increased.

In retrospect, it now seems obvious to question why the individual transistors of a semiconductor wafer were separated. Why not develop a process for electrically isolating the devices from each other to replace the isolation that had been achieved by cutting them apart? The devices could then be interconnected on the semiconductor wafer to form what we now refer to as an integrated circuit. At the end of the 1950s, this idea was realized (Meindl 1977). As is often the case, several individuals working independently were involved in developing the earliest integrated circuits. However, Jack Kilby is frequently credited with having "invented" the integrated circuit (Kilby 1976). In 1958, he demonstrated a hand-fabricated phase-shift oscillator and a flip-flop using germanium transistors. Resistors consisted of appropriately doped semiconductors, whereas capacitors utilized reverse-biased semiconductor junctions. These demonstration circuits established the feasibility of a concept that was rapidly exploited.

1.2 WIRELESS COMMUNICATION: A NEW ERA

The first use of the triode vacuum tube was for wireless communication. Lee De Forest, its inventor, described the tube as an oscillation valve – that is a device for detecting wireless or radio signals. (As an aside, it should be noted that Lee De Forest's autobiography has the subtitle of *Father of Radio*. This parentage is not widely accepted.) A close relationship of electronic devices to radio characterized the first half of the 20th century. The related professional organization in the United States was the Institute of Radio Engineers founded in 1912. It was not until 1963 that the designation "radio" was dropped when this organization merged with the Institute of Electrical Engineers to form the Institute of Electrical and Electronic Engineers (IEEE).

Maxwell's equations, the kernel of electromagnetic theory, provide the basis on which wireless communication, that is radio, is based. James Clerk Maxwell built on the work of Coulomb, Oersted, Ampère, Henry, Faraday, and Gauss in formulating these now well-known equations. Through a series of experimental observations and theoretical deductions, Heinrich Rudolf Hertz demonstrated the validity of Maxwell's equations. Hertz published the first text on electrodynamics in 1892 *Untersuchungen über die Ausbreitung der elektrischen Kraft* (*Electric Waves*, the title of an English translation by D. E. Jones). Following the death of Hertz in 1894, the lectures on the studies of Hertz by Oliver Joseph Lodge laid the groundwork for a much wider understanding of electromagnetic principles. Lodge and Ferdinand Braun were responsible for developing the concept of resonant tuning and demonstrating the importance of having the transmitter and receiver of a system tuned to the same frequency (Aitken 1976, Jolly 1975, Kurylo and Süsskind 1981, McNicol 1946). Concurrently, Oliver Heaviside is credited with putting Maxwell's equations into their presently utilized form (Nahin 1988, 1990).

A difficulty encountered in performing early electromagnetic experiments was that of obtaining a suitable detector of high-frequency signals. An early detector was the coherer, basically a small glass tube filled with loosely packed metal filings. The operation of this device relied on the nonlinear nature of the resistance of the filings. For small currents the filings had a high resistance, whereas for larger currents the filings tended to cohere, resulting in a small resistance. A mechanical tapping of the coherer was necessary to restore the high resistance after the termination of a large current. For a receiver, the alternating current produced by an electromagnetic signal caused the filings to cohere. This effect was detected by a low-voltage direct-current circuit connected to the coherer. Edouard Branly developed several different coherers and appears to have been the first to use the term radio (in this context) by proposing the name *radioconductor* for the coherer.

It was Guglielmo Marconi who in 1895 refined and assembled the appropriate apparatus and demonstrated that it could be used for signaling (Jolly 1972, Masini 1995). Not being successful in interesting his Italian government in this new means of communication, he traveled to England, where the British post office was receptive. Recognizing the commercial importance of wireless

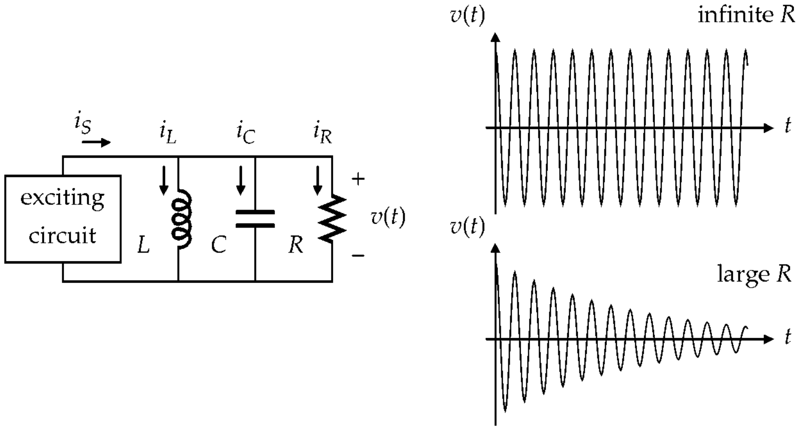


Figure 1.8: A resonant circuit.

telegraphy, he took out patents and formed the Marconi Wireless Signal Company. Progress was rapid: in 1901 he succeeded in sending a wireless signal across the Atlantic.

ELECTRICAL TUNING

An important aspect of radio communication is that of tuning; that is, to utilize a circuit that has an optimal response at a particular signal frequency. This is generally achieved with an inductor–capacitor circuit such as that of the parallel circuit of Figure 1.8. The resistance R is included to account for circuit losses (the resistance of the inductor) and energy that might be radiated as a result of an antenna connected to the circuit. Early wireless transmitters used a current impulse i_S produced by a spark gap to initiate the voltage oscillations of the circuit. Consider the case for which the circuit has previously been excited and the current i_S is zero. This implies that the sum of the currents of the individual elements must be zero, as given by the following:

$$\begin{aligned} &i_S(t) = i_L + i_C + i_R = 0 \\ &i_L = \frac{1}{L}\int v\,dt, \qquad i_C = C\frac{dv}{dt}, \qquad i_R = \frac{v}{R} \end{aligned} \tag{1.1}$$

These two equations may be combined and then differentiated to produce a single second-order differential equation:

$$\begin{aligned} &\frac{1}{L}\int v\,dt + C\frac{dv}{dt} + \frac{v}{R} = 0 \\ &\frac{d^2v}{dt^2} + \frac{1}{RC}\frac{dv}{dt} + \frac{v}{LC} = 0 \end{aligned} \tag{1.2}$$

For an ideal circuit with no loss ($R \to \infty$), a constant-amplitude oscillating voltage is a valid solution of the differential equation as follows:

$$v(t) = V_m \cos\omega_0 t, \qquad \omega_0 = \frac{1}{\sqrt{LC}} \tag{1.3}$$

Hence, once this lossless circuit is excited by an external current, its voltage will continue to oscillate indefinitely.

For a circuit with loss (finite R), damped sinusoidal oscillations occur given by

$$v(t) = V_m\, e^{-\alpha t} \cos \omega_0 t$$
$$\alpha = \frac{1}{2RC}, \qquad \omega_0 = \sqrt{\frac{1}{LC} + \left(\frac{1}{2RC}\right)^2} \tag{1.4}$$

The current impulse of a spark was used for the earliest wireless transmitters. Modern transmitters (radio and TV stations, citizens band transceivers, cellular telephones, etc.) rely on essentially the same principle except that an electronic exciting circuit is utilized that generally provides a current impulse for each oscillating cycle.

How does this circuit manage to continue to oscillate when the exciting current no longer exists? To answer this question, we must recall that inductors and capacitors store electrical energy. Let e_C and e_L be the instantaneous stored energies of the capacitor and inductor, respectively.

$$e_C = \frac{1}{2}Cv^2, \qquad e_L = \frac{1}{2}Li^2 \tag{1.5}$$

Consider the idealized case ($R \to \infty$) for which the amplitude of the voltage is constant (Eq. (1.3)).

$$i_L = \frac{1}{L}\int v\, dt = \frac{V_m}{\omega_0 L}\sin \omega_0 t$$
$$e_L = \frac{1}{2}\frac{V_m^2}{\omega_0^2 L}\sin^2 \omega_0 t = \frac{1}{2}CV_m^2 \sin^2 \omega_0 t \quad \text{because} \quad \frac{1}{\omega_0^2} = LC \tag{1.6}$$

The total energy $e_C + e_L$ is constant for this circuit:

$$e_C + e_L = \frac{1}{2}CV_m^2 = \frac{1}{2}\frac{V_m^2}{\omega_0^2 L} \tag{1.7}$$

It will be noted that when the stored energy of the capacitor is a maximum, that of the inductor is zero and vice versa (Figure 1.9).

In effect, there is an interchange of energy between the capacitor and the inductor of the circuit. For a circuit with a finite resistance, the electrical energy is gradually dissipated by the resistor; that is, the electrical energy is converted to thermal energy (or radiated if the resistor represents the effect of an antenna).

VACUUM TUBE CIRCUITS

Following its invention, the vacuum tube triode was extensively improved, and numerous electronic circuits were developed that greatly increased the tube's utility. Armstrong's invention of regeneration in 1912, the use of positive feedback to increase the gain of a circuit, increased the sensitivity of receivers. For example, using Armstrong's regeneration principle, it is possible to build a shortwave receiver with but a single vacuum tube (or transistor) that is capable of receiving signals from all over the world. A modification of this circuit was also used

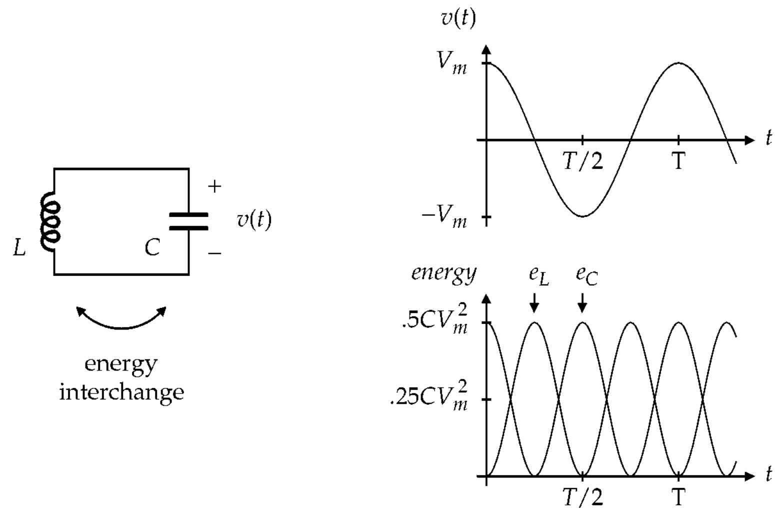


Figure 1.9: The energy interchange of a resonant circuit.

to produce radio-frequency oscillations that allowed the replacement of radio transmitters relying on either spark-gap or mechanical alternator generators.

Although Armstrong was the first to submit a patent application for regeneration, the application was immediately challenged by De Forest. It was claimed by De Forest that this effect was discovered a year earlier in his laboratory, albeit the technician's curt notebook entry for the circuit was "no good." A lifelong animosity surrounding legal challenges over patents ensued between these two radio pioneers. Armstrong was initially granted a patent for regeneration and successfully resisted the early challenges of De Forest. Eventually, however, De Forest won through challenges that were carried all the way to the U.S. Supreme Court. Nevertheless, Armstrong is generally accepted as the circuit's inventor, and the historical record indicates that Armstrong had a better understanding of the circuit than De Forest. The Institute of Radio Engineers (IRE) honored Armstrong for the regeneration invention with its medal of honor in 1918. When Armstrong attempted to return the medal in 1934 after losing De Forest's patent challenge, the IRE board of directors not only refused to accept the return of the medal (a unanimous decision) but reaffirmed its initial citation (Lewis 1991).

THE SUPERHETERODYNE RECEIVER

Among Armstrong's numerous inventions is the superheterodyne radio receiver. His earlier regenerative receiver, although sensitive, was prone to behave erratically (it frequently burst into oscillation). Tuned circuits, such as the parallel resonant circuit of Figure 1.8, are required to select a desired radio signal and to reject other signals. In addition, tuned circuits are used to enhance the gain of radio-frequency amplifiers. To tune a given circuit, its capacitance, inductance, or both must be changed. Several amplifiers, each with a tuned circuit, are often needed. The tuning of a radio thus required the simultaneous adjustment of several circuits – a tuning knob was needed for each circuit.

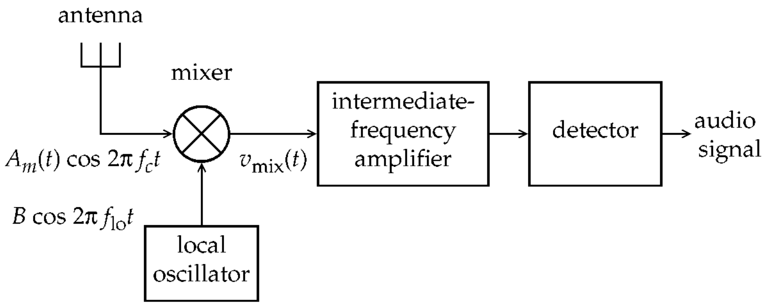


Figure 1.10: A superheterodyne receiver.

Armstrong recognized that the carrier frequency of a signal could be changed through a nonlinear mixing process (Figure 1.10). Consider the case for an amplitude-modulated signal $A_m(t)\cos 2\pi f_c t$ derived from an antenna system. A second high-frequency signal generated by the local oscillator of the receiver, $B\cos 2\pi f_{lo} t$, is also required. Suppose, initially, that the mixer results in an output voltage $v_{mix}(t)$ that is the product of its two inputs (a standard multiplier symbol is shown in Figure 1.10) as expressed by

$$\begin{aligned} v_{mix}(t) &= A_m(t)\cos 2\pi f_c t \cdot B\cos 2\pi f_{lo} t \\ &= \frac{1}{2} A_m(t) B[\cos 2\pi(f_{lo} + f_c)t + \cos 2\pi(f_{lo} - f_c)t] \end{aligned} \quad (1.8)$$

The preceding result was obtained using the trigonometric identities for the cosine of the sum and difference of two angles as follows:

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\ \cos(\alpha-\beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta \\ \cos\alpha\cos\beta &= \frac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)] \end{aligned} \quad (1.9)$$

The output voltage of the mixer consists of two signals, one multiplied by $\cos 2\pi(f_{lo} + f_c)t$ and the other multiplied by $\cos 2\pi(f_{lo} - f_c)t$. These signals are two distinct amplitude-modulated signals, one having a carrier frequency of $f_{lo} + f_c$ and the other of $f_{lo} - f_c$. The amplitude of each is proportional to the amplitude of the original signal, that is, $A_m(t)$.

Consider the case for a typical AM broadcast receiver that might be tuned to receive an amplitude-modulated signal with a carrier frequency of 1350 kHz. Suppose that its local oscillator is generating an 1800-kHz signal. The output of the mixer would consist of two amplitude-modulated signals, one with a carrier frequency of 450 kHz and the other with a carrier frequency of 2250 kHz. If the intermediate-frequency amplifier is tuned to a frequency of 450 kHz, the component with a carrier frequency of 450 kHz would be amplified, whereas the 2250-kHz carrier signal would be lost. The 450-kHz signal would be detected after being amplified, thus yielding an audio output signal corresponding to the amplitude modulation $A_m(t)$ of the received signal. (A level shifting, generally achieved with a coupling capacitor, is also necessary to recover the audio signal.)

What is the advantage of a superheterodyne receiver? Again, consider the broadcast receiver with an intermediate-frequency amplifier tuned to a fixed frequency of 450 kHz. The carrier frequency of the signal to which the receiver responds depends on the receiver's local oscillator frequency. To receive a signal of 550 kHz (the lower end of the broadcast band), a local oscillator frequency of 1000 kHz is required. This results in signals with carrier frequencies of 450 kHz and 1550 kHz being produced by the mixer. The 450-kHz signal is amplified, and the 1550-kHz signal is rejected. To receive a 1600-kHz signal (the upper end of the broadcast band), a local oscillator frequency of 2050 kHz is required, which, in turn, produces mixer output signals with frequencies of 450 kHz and 2500 kHz. The advantage of this receiver is that tuning is achieved by changing the local oscillator frequency (a range of 1000 to 2050 kHz is required). Although this necessitates that the inductance, capacitance, or both of the circuit be changed, the resonant frequency of only a single circuit needs to be changed. Even for an improved receiver, in which a tuned circuit is employed for the input of the mixer, a mechanical tracking system is used to tune the two circuits simultaneously with a single tuning knob.

1.3 THE TELEGRAPH AND TELEPHONE: WIDE-SCALE INTERCONNECTIONS

The telephone, invented by Alexander Graham Bell in 1876, predated electronic devices by over a quarter of a century (Bruce 1973, Sharlin 1963, Pupin 1926). By the time of the invention of the vacuum tube triode in 1906, the telephone was widely used throughout urban areas. The operation of the telephone was predicated on an earlier electrical communication system, the telegraph. Bell was attempting to develop a multiplexing system to transmit several telegraph signals simultaneously on a single telegraph line when he went "astray" and invented the telephone. It is not, however, inappropriate that Bell should be associated with the telephone because he, his father, and his grandfather were highly respected speech specialists (elocution experts).

THE TELEGRAPH

The telegraph system of Morse, invented in 1837, depended on the earlier work of Volta, Oersted, and Ampère, among others. Volta (after whom the voltage unit is named) devised the first battery, an "electrochemical pile" consisting of zinc and silver discs separated by brine-soaked cloth or paper. Oersted observed that an electric current produces a magnetic field, and Ampère established the mathematical theory relating magnetic fields to electric currents. This led to the development of the electromagnetic responder, the basis of the telegraph receiver. An elementary telegraph system is shown in Figure 1.11. Morse's success depended on the then available electrical devices, which he assembled into a telegraph system. His prime contribution, that for which he is generally remembered, was a binary coding system for signaling. An *on* state corresponded to a current (the key depressed), and an *off* state corresponded to no current. Furthermore,

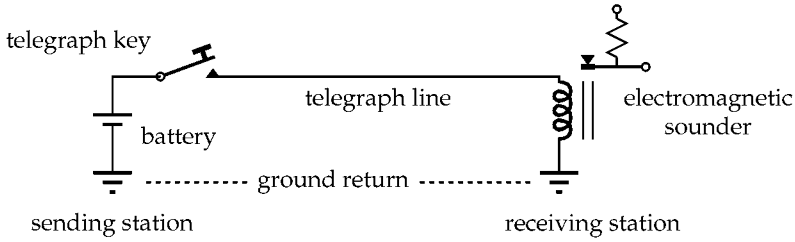


Figure 1.11: An elementary telegraph system.

on periods were broken into short and long intervals, that is, the dots and dashes that we now refer to as Morse code.

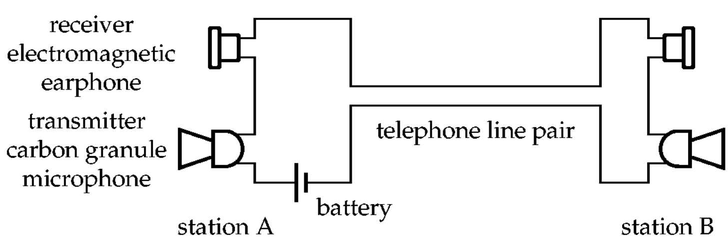
Morse's original telegraph system utilized a receiver consisting of a pencil activated with an electromagnet, which made a trace on a moving strip of paper. A telegraph operator would then decode the marks, thus recovering the original message. Operators, however, soon discovered that they could directly decode the message by listening to the clicks of the printer – the marked tape was unnecessary. The electromagnetic sounder was developed to optimize the decoding, one type click being associated with the electromagnet's being activated and the other with its being deactivated, thus distinguishing the *off*-to-*on* from the *on*-to-*off* transition of the current.

A revised code, the American Morse code, was used for wired telegraph systems in the United States, whereas a second version of this code, the International Morse code, was adopted for wireless communication and is still extensively used for shortwave radio communications by radio amateurs and others. The same *on–off* signaling principle utilized in the early telegraph forms the basis of today's modern fiber-optic systems. The light from a light-emitting laser, which is turned on and off, is transmitted through an optical fiber to a receiver, a light-detecting diode. Not only is the speed of the fiber-optic system much greater, but the electrical signals, such as those produced by a telephone, are directly encoded into *on–off* signals.

BASIC TELEPHONE SYSTEM

The telephone of Figure 1.12 works on a similar principle to that of the telegraph. In place of the key, a microphone is used to modulate the current of the circuit. Bell initially utilized a microphone that consisted of a diaphragm attached to a needle immersed in an acidic solution. The motion of the needle, the result of sound waves striking the diaphragm, caused the resistance of the circuit to

Figure 1.12: A basic telephone system.



fluctuate. The fluctuating resistance, in turn, resulted in circuit current fluctuations. As for the telegraph system, an electromagnet was used for the receiver. In place of metallic contacts to produce the clicks of a sounder, an iron diaphragm acted upon by a magnetic field was used. A permanent magnet and an electromagnet energized by the fluctuating current of the telephone circuit jointly produced the magnetic field.

An improved telephone microphone in which carbon granules were used in place of the acidic solution was soon introduced. The motion of the microphone's diaphragm produced pressure fluctuations on the granules, which, in turn resulted in a fluctuation in resistance. Although modified versions of the carbon granule microphone have been introduced, this type of microphone is still used for phones over 100 years after being first introduced.

The circuit of Figure 1.12 is bidirectional, with two transmitters and receivers, so that either party can transmit while the other listens. In place of the earth ground return of the telegraph system, a second wire is used to complete the telephone circuit. A ground return, although reducing the amount of wire required, resulted in erratic and unpredictable effects. Today, a pair of wires is universally used for all local telephone connections (for example, to connect one's home phone to the telephone office). With a two-wire circuit, interference (cross talk) caused by electric and magnetic coupling between different telephone circuits is greatly reduced.

In addition to radio applications, early triode vacuum tubes were also used for long-distance repeater telephone amplifiers. As a result of insulation and wire losses, telephone signals are attenuated, that is, they become weaker as the length of the telephone line is increased. Before the advent of the vacuum tube amplifier, a mechanical-type amplifier was developed that consisted of a tightly coupled telephone receiver and a carbon granule microphone. Unfortunately, it badly distorted the telephone signal. The triode vacuum tube amplifier, first used as a telephone repeater in 1913, was the ideal solution for extending long-distance telephone service. Triode vacuum tube amplifiers were used in 1915 for the first U.S. transcontinental telephone line, although a set of mechanical-type amplifiers were held in reserve (Fagen 1975).

ANALOG TELEPHONE SIGNALS

The basic telephone system of Figure 1.12 is an analog system because the voltage differences of the circuit may take on any value within a set of prescribed limiting values. Circuit voltages fluctuate in accord with the audio signal produced by the speaker's microphone. This differs from the telegraph system in which signaling depends only on an *on–off* voltage condition.

An important characteristic of analog signals is their frequency spectrum (Figure 1.13). A sinusoidal signal has only a single frequency component, its periodic frequency, whereas a nonsinusoidal periodic signal has a fundamental frequency component as well as a set of harmonic components. The amplitude and phase of each frequency component depend on the periodic waveform of the signal: the more rapid the variations of the signal, the larger the harmonic amplitudes. From an information perspective, a periodic signal is not very interesting because, if

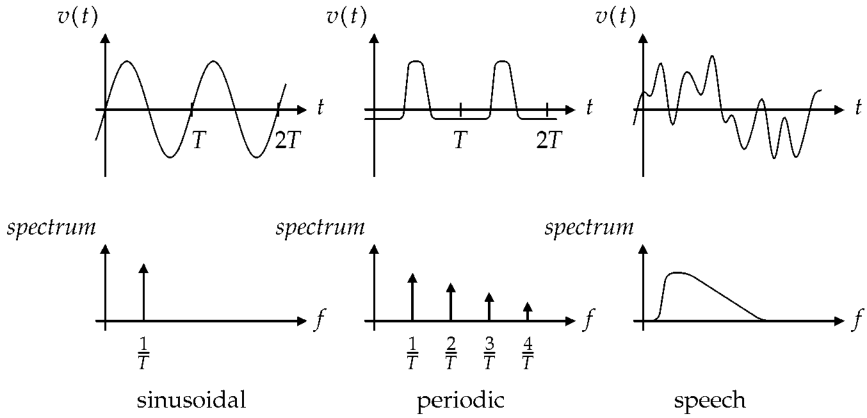


Figure 1.13: Time-dependent signals with the relative amplitudes of their frequency spectrum.

one has "seen" one period of the signal, one has "seen" them all. The nonperiodic speech signal has a frequency spectrum that tends to be continuous. As a result of electrical limitations of a telephone system, those that are unavoidable as well as those intentionally introduced, the frequency spectrum of telephone signals is generally limited. For the U.S. system, a spectrum of approximately 300 to 3400 Hz is utilized, and frequency components outside this range are filtered out.

An electronic system, such as an amplifier, must be capable of responding to all desired frequency components of a signal. Because a signal may be considered to be composed of a multitude of sinusoidal signals (based on a Fourier series representation of a periodic signal or a Fourier transform of a nonperiodic signal), an electronic system may be designed to reproduce sinusoidal signals faithfully. Furthermore, testing is usually done with sinusoidal signals, and operating specifications are given in terms of sinusoidal signals. For the speech signal of Figure 1.13, an electronic system must have a uniform frequency response over the frequency spectrum of the signal and be capable of responding to the amplitude range of the signal without significant distortion.

Generally, a dedicated pair of wires is used to connect a subscriber's telephone to a local telephone switching office. If the subscriber is calling a second subscriber connected to the same office, a direct connection is established through the switching equipment of the telephone office. Until the 1970s, switching was primarily through mechanically positioned contacts. For a subscriber calling a different office, for example an individual in another city, an interoffice connection was necessary. Depending on the circumstances, this may have required intermediate switching connections to reach the final destination.

In addition to amplifying signals, electronic systems are used for multiplexing telephone signals. A single pair of wires, or any other type of transmission system, is generally used to carry several telephone signals simultaneously. This, before the advent of digital telephone circuits, was achieved by translating the frequency spectrum of individual telephone signals (Figure 1.14). This process is similar to that employed for radio systems in which a high-frequency carrier is used to "carry" a lower-frequency modulating signal. The modulating signals

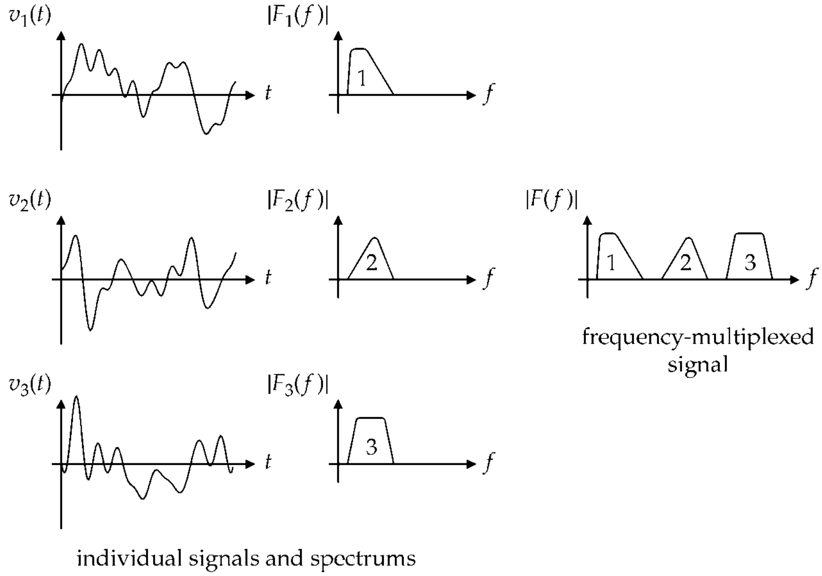


Figure 1.14: Frequency multiplexing of telephone signals.

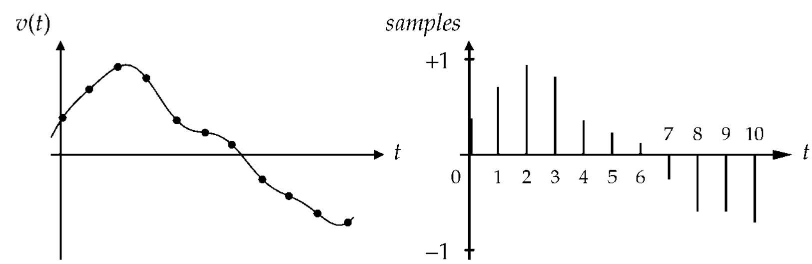
are the individual telephone signals. In the United States, Bell Telephone Laboratories was responsible for much of the early improvements in vacuum tubes and vacuum tube circuits (Fagen 1975). Hence, a parallel, simultaneous development of electronic systems occurred for both early telephone and wireless systems.

DIGITAL TELEPHONE SYSTEMS

Although an analog connection is still generally used for connecting a subscriber and a telephone office, digital signals are used within switching offices and for transmitting telephone signals between offices. Analog-to-digital and digital-to-analog converters are used at each subscriber's office connection. With digital signals, switching by entirely electronic means is readily achieved, thereby eliminating erratic connections that may occur with mechanically activated contacts. With the development of specialized integrated circuits, the wide-scale usage of all-electronic digital telephone systems is now common.

As indicated in Figure 1.15, a sampling process is used to convert an analog signal to a digital signal. The U.S. telephone system uses 8000 samples per second; that is, the telephone signal is sampled every 125 μs. The resultant samples are

Figure 1.15: Sampling of a telephone signal.



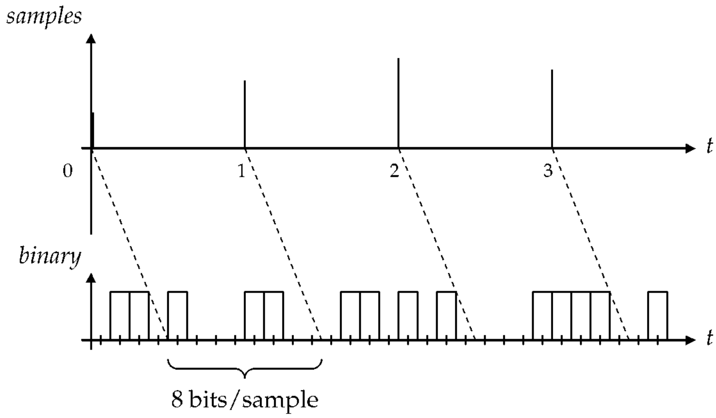


Figure 1.16: The conversion of sample amplitudes to a binary coded signal. An expanded time scale has been utilized for clarity.

a set of amplitudes that represent the analog signal. It may be shown that if the sampling rate is at least twice the highest frequency component of the analog signal, the samples contain all the information of the original analog signal; that is, the samples may be used to regenerate the original analog signal without error.

For the U.S. telephone system, the sample amplitudes are converted to a set of eight binary quantities. The normalized range of ± 1 for the samples of $v(t)$ of Figure 1.15 is divided into 256 (2^8) discrete levels, and each sample is assigned the level to which it is the closest. A nonuniform set of levels is utilized so that high-level and low-level signals tend to be reasonably well preserved. The digital representation of the signal of Figure 1.15 is indicated in Figure 1.16. Because 8 bits are used for each sample, the bits have to be squeezed into the 125-μs (1/8000 s) sampling interval. This results in a time interval of 15.625 μs being available for each bit, which corresponds to 64,000 bits per second. Although the digital signal requires a much higher transmission rate than the analog signal it represents, only an *on–off* condition needs to be transmitted.

Two digital signal paths, one for each direction of the telephone connection, are required. Electronic logic circuits are utilized for "switching" these signals. A modern digital telephone office is, in essence, a specialized computer with a very large number of input and output connections. Efficient modern telephone transmission systems utilize time-multiplexed digital signals. The time of each group of 8 bits representing a sample of a signal is reduced. For a 50-percent reduction in time, the bits of a second signal, similarly modified, could be inserted between the bits of the first signal. At the end of the transmission system, the signals that occur at different time intervals can be separated.

A typical first-level time-multiplexing system combines 24 digital telephone signals. One frame of the resultant digital signal consists of the 8 bits corresponding to a single sample of each signal (a total of 192 bits) plus one framing bit used to identify the frame. If an *on–off* sequence (*on* one frame, *off* the next, etc.) is used for the framing bit, the beginning of a frame may be readily identified at the receiving end of a time-multiplexed transmission system. The overall bit rate of the time-multiplexed 24 telephone signals is therefore 1,544,000 bits per