CHAPTER ONE

Information Transfer Technology

1.1 INTRODUCTION

The design of radio frequency (RF) circuits borrows from methods used in low frequency audio circuits as well as from methods used in design of microwave circuits. However, there are also important departures from audio and microwave frequency methods, so that design of radio frequency circuits requires some specialized techniques not found in these other frequency ranges. The radio frequency range for present purposes will be taken to be approximately somewhere between 300 MHz and 3 GHz. It is this frequency range where much of the present day activity in wireless communication occurs. In this range of frequencies, the engineer must be concerned with radiation, stray coupling, and frequency response of circuit elements that, from the point of view of lumped, low frequency analysis, might be expected to be independent of frequency. At the same time, the use of common microwave circuit elements such as quarter wave transformers is impractical because of the long line lengths required. The use of monolithic circuits have enabled many high frequency designs to be implemented with lumped elements, yet the frequency response of these "lumped" elements still must be carefully considered. The small size of lumped elements in integrated circuits has provided practical designs of filters, transformers, couplers, etc. in lumped element form. Therefore discussion of designs for low noise amplifiers, power amplifiers, oscillators, mixers, and phase lock loops will be addressed with both lumped and distributed elements. Several of the numerical examples given in the text use computer programs. Source code for these programs are available

Radio Frequency Circuit Design, Second Edition, by W. Alan Davis

on the web*. However, before getting into the details in the design of radio frequency circuits, it is important to understand that the purpose for these circuits is to transmit information.

1.2 INFORMATION AND CAPACITY

What exactly is information? *Random House Dictionary* 1966 states that "information" is "knowledge communicated or received concerning a particular fact or circumstance. …" A narrower technical definition more closely aligns with the focus given here is that "information" is an "indication of the number of possible choices of messages, expressible as the value of some monotonic function of the number of choices, usually log to the base 2." *Information* then is a term for data that can be coded for digital processing.

Some examples of data that illustrate the meaning of information is helpful. If a signal were sent through a communication channel that never changed, then it would be conveying no information. There must be change to convey a message. If the signal consisted of 1 0 1 0 1 0 1 0 … , there would be changes in the signal but still no information is conveyed because the next bit would be perfectly predictable. So while change is important, it is not the sole criterion for information. There is one last example. If a signal in an amplitude modulation system consists of purely random voltage fluctuations, then again no information is being transmitted. It is simply noise, and the receiver is no more knowledgeable after having heard it.

A communication system consists of a transmitter, a receiver, and a channel. The channel is capable of carrying only a certain limited amount of information. A water pipe can be seen as a rough analogy to a communication channel. The limitation in a communication channel is given the technical term *capacity*. It refers to the amount of information that is transmitted over a time interval of T seconds. The time interval can be broken up into short time intervals, each of duration τ. Clearly, the more distinct time intervals τ there are in the total time span T, the more information that can be transmitted. The minimum size of τ is determined by how well one pulse in one time frame can be distinguished from a pulse in a neighboring time frame. The limitation on how short a time frame can be is related to the channel bandwidth. In the water pipe analogy, the channel bandwidth corresponds to the pipe diameter.

In addition, the signal voltage will have a maximum amplitude that is limited by the available power in the system. This voltage range can be divided into many levels, each level representing a bit of information that is distinguished from another bit. The voltage range cannot be split indefinitely because of the noise that is always present in the system. Clearly, the more voltage intervals in a given time frame τ, the more information capacity there is in the system. Just as the flow of water through a pipe is limited by the amount of

*http://www-ee.uta.edu/online/adavis/rfsoftware pressure on the water, by the friction on the walls of the pipe, and by the diameter of the pipe, so the capacity of a transmission system is limited by the maximum voltage level, by the noise in the system that tends to muddle the distinction between one voltage level and another, and by the bandwidth of the channel, which is related to the rise time of a pulse in the system.

In one of the time intervals, τ, there are n voltage levels. The smaller that τ is and the larger n is, the more information that can be transmitted through the channel. In each time interval, there are n possible voltage levels. In the next time interval there are also n possible voltage levels. It is assumed that the voltage level in each time frame is independent of what is going on in other time frames. The amount of information transmitted in a total of T seconds corresponds to the products of the possibilities in each interval:

$$n \cdot n \cdot n \cdot n \cdots n = n^{T/\tau} \tag{1.1}$$

The total information, H, transmitted intuitively is directly proportional to the total time span T, and is defined as the log of the above product. By convention, the base 2 logarithm is used.

$$H = T/\tau \log_2 n \tag{1.2}$$

The system capacity is simply the maximum *rate* of transmission (in bits/s) through a system:

$$C = H/T = 1/\tau \log_2 n \tag{1.3}$$

System capacity is inversely proportional to the minimum time interval over which a unit of information can be transmitted, τ. Furthermore, as the number of voltage levels increases, so does the capacity for more information.

Information can be transmitted through a channel in a variety of different forms, all giving the same amount of information. For example, suppose that a signal can take on any one of eight different voltage levels, 0,1, … , 7, in a given time interval τ. But the eight-level signal could also equally be sent with just two levels, 0,1. However, for every interval that has eight possible levels, three intervals will be needed for the two-level signal. A convenient conversion between the two systems is shown in Table 1.1.

Clearly, a 16-level signal could be transmitted by a sequence of 4 binary signals, and a 32-level signal with a sequence of 5 binary signals, and so on. For n levels, $\log_2 n$ bits are needed. The information content of a signal is defined then to be the number of binary choices, or bits, that are needed for transmission. A system that is designed to transmit speech must be designed to have the capacity to transmit the information contained in the speech. While speech is not the total of what humans communicate, in a communication system, it is that with which engineers have to work. A decision must be made as to what level of fidelity the speech is to be transmitted. This translates to the bandwidth

TABLE 1.1 Eight-Level and Two-Level Systems

$n = 8$	$n = 2$
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

requirement of an analog system, or the number of voltage levels available in a given total voltage range. Ultimately the restriction is always present even if sophisticated coding techniques are used. The capacity of the system must be greater than or equal to the rate of information that is to be transmitted. Beyond this, system cost, power levels, and available transmission media must be considered.

1.3 DEPENDENT STATES

The definitions of the preceding section imply that the voltage level in each time interval, τ, is independent of the voltage level in other time intervals. However, one very simple example where this is not the case is the transmission of the English language. It is known in the English language that the letter *e* is much more likely to appear than the letter *z*. It is almost certain that the letter *q* will be followed by the letter *u*. So in transmitting a typical message in English, less information is being actually sent than there would be if each letter in the alphabet were equally likely to occur. A way to express this situation is in terms of probability. The total number of signal combinations that could occur in a message T seconds long if the value in each interval is independent of the others is $n^{T/\tau}$. On average, every possible message T seconds long would have a probability of occurrence of $1/n^{T/\tau}$.

The probability takes the form

$$P = \frac{\text{number of occurrences of a particular event}}{\text{total number of events}} \tag{1.4}$$

Information can be measured in terms of probability. The probability is $P = 1/n$ if there are n possible events specified as one of n voltage levels, and each of these events is equally likely. For any one event, the information transmitted is written $H_1 = -P\log_2 P$. For m intervals, each τ seconds long, there will be m

times more information. For m intervals, the information written in terms of probability is

$$H = \frac{T}{\tau}\log_2 n = -m\log_2 P \quad \text{bits} \tag{1.5}$$

Consider a binary system, where a number 0 occurs with a probability of p and the number 1 occurs with a probability of q. Knowing that $p + q = 1$, the information content of a message consisting of 0's and 1's is found. The total information is the sum of the information carried by the 0's and that of the 1's:

$$H = -\frac{T}{\tau}(p\log_2 p + q\log_2 q) \quad \text{bits} \tag{1.6}$$

If the probabilities of p and q were each 0.5, then the total information in T seconds is T/τ. If, for example, $p = 0.25$ and $q = 0.75$, then

$$H = -\frac{T}{\tau}(0.25\log_2 0.25 + 0.75\log_2 0.75) \quad \text{bits}$$

$$H = \frac{T}{\tau}(0.5 + 0.3113) = 0.8113\frac{T}{\tau} \quad \text{bits} \tag{1.7}$$

Hence, when there is a greater probability that an expected event will occur, there is less information. As p approaches 1 and q approaches 0, the near certainty of an event with probability p will give 0 information. Maximum information occurs when $p = q = 0.5$.

This scenario can be generalized for n signal levels in a given signal interval τ. Assume that each of these n signal levels, s_i, have a probability of occurrence of P_i where

$$P_1 + P_2 + \cdots P_n = \sum P_i = 1 \tag{1.8}$$

Assume further that the probability of finding a given signal level is independent of the value of the adjacent signal levels. The total information in T/τ intervals or in T seconds is

$$H = -\frac{T}{\tau}\sum_i^n P_i\log_2 P_i \quad \text{bits} \tag{1.9}$$

The capacity required to transmit this amount of information is then

$$C = -\frac{1}{\tau}\sum_i^n P_i\log_2 P_i \quad \text{bits/s} \tag{1.10}$$

In the case where each level is equally likely, $P_1 = P_2 = P_3 = \cdots P_n = 1/n$, then for the n level signal,

$$H = -\frac{T}{\tau}\sum_{i}^{n} P_i \log_2 P_i = \frac{T}{\tau}\log_2 n \quad \text{bits} \tag{1.11}$$

More details on information may be found in specialized texts; a short introduction is given by Schwartz [1]. In this study of radio frequency (RF) design the primary focus will be on the fundamental hardware design used in transmitters and receivers. Other topics that are of great interest to communication engineers such as programming digital signal processing chips, various modulation schemes, or electromagnetic propagation problems are more fully explored in specialized texts in those areas. In this book these areas will be referred to only as needed in illustrations of how systems may be implemented.

1.4 BASIC TRANSMITTER–RECEIVER CONFIGURATION

Analog RF and digital designs are both found in typical communication systems. There are many systems where digital signal processing is playing a large role along with advanced RF circuit design. A typical superheterodyne radio transmitter and receiver are shown in Fig. 1.1. An actual system would

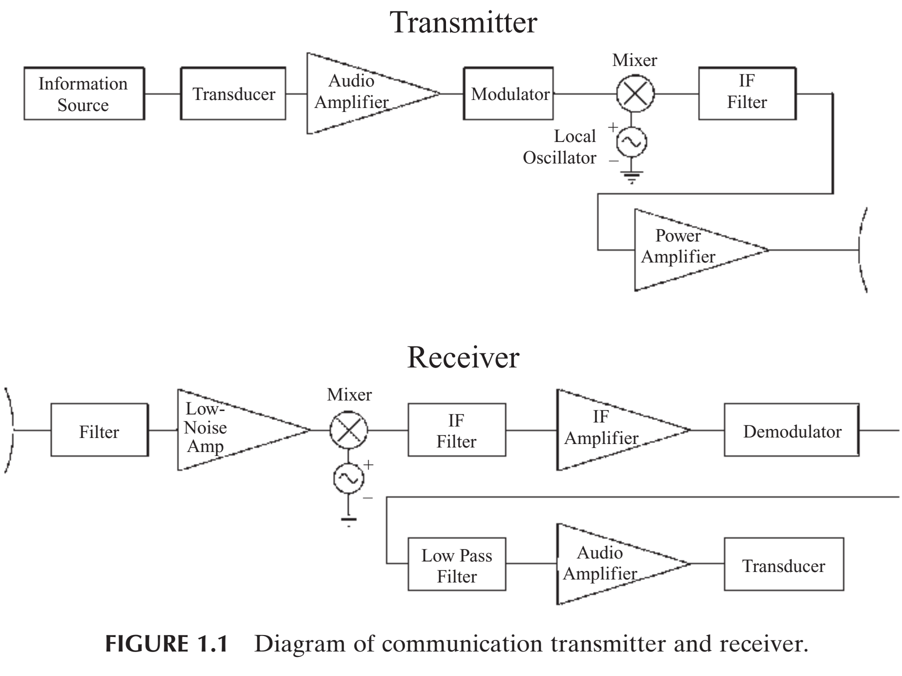

FIGURE 1.1 Diagram of communication transmitter and receiver.

be optimized for cost, noise immunity, fading, available bandwidth, bandwidth efficiency (the ratio of the throughput data rate per hertz in a given bandwidth), power efficiency (which measures the ability of a system to preserve the message under low-power conditions), intermodulation products, adjacent channel interference, and so on. The modulator and demodulator shown in the figure symbolize a large range of design options, often making use of digital techniques. Clearly, the circuits in Fig. 1.1 are only an outline of actual transmitters and receivers.

The transmitter in Fig 1.1 starts with some information source, which could be sound or a visual image. This is then converted to an electrical signal in the transducer, which may require amplification. The modulator codes the information and must be compatible with the demodulator. The modulator can be either analog or digital, and it comes in a wide variety of forms. It encodes the message in a certain way so as to meet the communication channel and receiver requirements. For example, if a video signal is being transmitted, the signal must carry information about the sweep time, intensity, and often color as well as the actual intelligence. The commonly used analog modulation techniques of amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM) encode the carrier wave by changing its amplitude, frequency, or phase, respectively. Multiple signals can share the same channel if the signals are at different frequencies as in *frequency division multiple access* (FDMA), or at different time slots as in *time division multiple access* (TDMA), or with different digital codes as in *code division multiple access* (CDMA).

The mixer circuit is the first component in this discussion that breaks into the RF range, and it provides two necessary functions. First, it raises the carrier frequency that in AM and FM systems is distinct from neighboring transmitters. The second function of the modulator is that it translates the message information to a much higher frequency. This allows antennas to be made a manageable size since their mechanical size normally corresponds to the wavelength of the signal. A great deal of effort has gone into making electrically small antennas, but there are always design compromises. Chapter 11 is devoted to mixers.

The mixer is accompanied by a local oscillator that in some cases is carefully tuned to different frequencies or is fixed as in broadcast stations. The quality of an oscillator is judged on how low its phase noise is or how much its frequency will drift over time with temperature or age. Oscillators can be designed to be manually or electrically tuned to different frequencies. Techniques that are used to stabilize an oscillator include using high Q elements such as quartz crystals, dielectric resonators, or using a constant-temperature oven. Phase-lock loops can be used to stabilize a high frequency with a stable low-frequency oscillator. Design of oscillator circuits is considered in Chapter 10 and phase-lock loops in Chapter 12.

The filter that follows the mixer is required because the nonlinear multiplication process of the mixer produces unwanted frequencies. In addition, providing appropriate impedance levels to the mixer and the following amplifier often requires impedance matching. Radio-frequency filters and transformers are the primary subject in Chapters 3, 5, and 6 and are used in the design of amplifiers in Chapters 8 and 9.

The final stage of the transmitter before reaching the antenna is the power amplifier. Since this component uses the greatest amount of power, high efficiency becomes important. In FM systems, class C amplifiers are often used since in practice they can produce efficiencies as high as 70%. For AM systems, class A or B amplifiers are often used because of the required linearity of AM signal transmission. However, class A amplifiers typically have efficiencies of only 30 to 40%. In the transmission of digital modulated signals, linearity of the power amplifier becomes very important because of the need to minimize co-channel interference. In all these cases, it is clear that designing the amplifier for maximum power transfer so that the load impedance is conjugately matched to the amplifier output impedance would mean half the power would be dissipated in the transistor itself. The power amplifier must be designed for maximum efficiency where the internal output impedance is small relative to the external load.

The receiver is usually more complicated than the transmitter, and its purpose is to unravel the signal from the transmitter after the signal has acquired some noise and other distractions while going through the channel. If the received signal is strong enough, it can be put directly into the mixer. However, as will be seen in a later chapter, the overall noise response of the amplifier is greatly enhanced by using a low-noise amplifier for the front end. The design of the low-noise amplifier is described in detail in Chapter 8.

1.5 ACTIVE DEVICE TECHNOLOGY

The first RF vacuum devices made their appearance in the 1930s and today are still found to be the most reliable and efficient high-power amplifiers with power levels reaching up to 30 MW. Their demise is not likely to occur soon as is made evident by such things as the ubiquitous microwave oven. New device designs and new materials continue to improve the quality of vacuum tubes used in amplifiers and oscillators.

The solid-state entrance to the RF arena began with two-terminal diodes. These included the Gunn diode, the impact avalanche transit time (IMPATT) diode, the trapped plasma avalanche triggered transit (TRAPATT), the tunnel diode, and even the *pn* junction (varactor diode) used in parametric amplifiers. The three terminal GaAs *metal semiconductor field-effect transistor* (MESFET) soon displaced the diodes in most applications. Even though the MESFET did not have as low a noise figure as the parametric amplifier or the power (at the time) of an IMPATT, its stability and efficiency was superior. Furthermore, its noise level was low enough for many practical applications. Subsequent arrivals were the AlGaAs/GaAs *heterojunction bipolar transistor*

(HBT) and the *high electron mobility transistor* (HEMT) all based on GaAs or other III–V materials. These classes of devices in some cases still provide the best performance for a variety of high-power, high-frequency applications. Engineers are starting to make use of GaN and SiC for high-power RF applications. The wide band gap of GaN (3.4 eV), high break down voltage, high drift velocity, and high thermal conductivity of these materials make them attractive for high-power *heterojunction field-effect transistors* (HFET) devices. While self-heating and high flicker noise has been a problem with the GaN devices, some resolution with the flicker noise problem has been accomplished.

However, the world is made of silicon. Silicon has the advantage of being cheaper to manufacture than its GaAs cousins, has good thermal characteristics, and most important has an entrenched manufacturing infrastructure. Silicon soon surpassed its predecessor, germanium. Within a few years the *complementary metal–oxide semiconductor* (CMOS) technology found favor in digital circuits because of its ability to integrate a large number of transistors in a small space. The desire to integrate digital and analog applications on the same chip as well as to provide cost reduction relative to the GaAs devices has spawned much interest in RF CMOS designs. The progress in making small gate-length high-speed CMOS devices has provided the ability to make RF devices using CMOS technology. However, the mixed signal designs have required the sacrifice in the Early voltage, which is important in many analog circuits. *Laterally diffused metal–oxide semiconductor* (LDMOS) has also been used in power amplifiers. Their high gain, linearity, and reliability have made them the best choice in many cellular base station applications. More recently, the SiGe *heterojunction bipolar transistor* (HBT) has been found to have many advantages over straight CMOS. These include superior flicker noise, broadband noise, Early voltage, transconductance, and better tracking of V_{be} relative to the V_t of the MOSFET. The SiGe HBT does well with linearity, though not quite as well as the CMOS device.

In summary, there are a wide variety of devices available to the analog RF designer and with them a variety of specialized processing and circuit design techniques. It is the goal of the following chapters to provide basic circuit design techniques that can be applied to a wide variety of active devices.

PROBLEMS

1.1. A pulse train is being transmitted through a channel at the maximum channel capacity of 25×10^3 bits/s. The pulse train has 16 levels.

a. What is the pulse width?

b. The pulse width is doubled and sent back on the same channel. What number of levels is required?

1.2. A system can send out a signal at six different levels: 0, 1, 2, 3, 4, 5, each 1 ms long. The probability of each of these levels occurring is 1/8, 1/8, 1/16, 1/4, 3/8, 1/16, respectively. Each pulse value is independent of any previous pulse values. What is the total amount of information conveyed in 1 second?

REFERENCE

1. M. Schwartz, *Information Transmission, Modulation, and Noise*, 3rd ed., New York: McGraw-Hill, Chapter 1, 1980.