Of related interest...

NOISE REDUCTION TECHNIQUES IN ELECTRONIC SYSTEMS

Henry W. Ott

This comprehensive reference book solves the perplexing problem of noise suppression. Emphasizes shielding, grounding, balancing, and filtering. Features include a checklist of the more commonly used noise reduction techniques, practical examples of solutions to the problem, and bibliographical references.

1976   (0 471-65726-3)   294 pp.

LOW-NOISE ELECTRONIC DESIGN

C. D. Motchenbacher & F. C. Fitchen

Offers practicing electrical engineers and technicians complete coverage of the problems of low-noise design in a simplified, practical manner. Special emphasis is on a logical approach to the subject and specific techniques useful in the design of low-noise systems. Among the materials presented are a computer program for the calculation and integration of noise, new information on noise in passive components, and many practical design examples.

1973   (0 471-61950-7)   358 pp.

ELECTRONICS
Circuits and Devices, 2nd Edition

Ralph J. Smith

A teachable treatment of integrated circuits, digital devices, microprocessors, and operational amplifiers. Includes the first presentation of microprocessors in an introductory text, an exceptionally clear explanation of physical concepts involved in the operation of diodes, FETs and BJTs that uses mathematics to formalize the results.

1980   (0 471-05344-9)   494 pp.

NOISE IN RECEIVING SYSTEMS

Raoul Pettai

CONTENTS

10 MULTI-RESPONSE TRANSDUCERS   157

11 MEASUREMENT OF NOISE PARAMETERS   207

12 SIGNAL AND NOISE IN CRYSTAL DETECTORS   243

2 | OVERVIEW OF COMMON NOISE SOURCES

Now that we have seen what some of the ultimate limitations of communication channels are, we should back off somewhat and take a closer look at the entire range of electrical noise, intrinsic or otherwise. This is not only instructive, but will place the principal object of study in this book—thermal noise—in proper perspective.

Figure 2.1 shows a classification of various types of noise commonly encountered. Obviously, this classification is neither all-inclusive nor the only one possible [1]. Depending on the emphasis and the interests of the reader, the same material could be arranged in many different ways. The arrangement was chosen because of its applicability to the daily work of a communications engineer who must be concerned with the noise generated within the receiving system, the noise generated in the usually distant and inaccessible source, and the ever-present noise arising in the environment. All three types of noise are important in the design and operation of most RF communication systems, although the techniques to combat them may vary widely. Under each major group in Fig. 2.1 several subdivisions are shown. As noted above, the list does not pretend to be complete. There is even some overlap in it as, for example, in the case of sky noise. The latter originates in the environment, but could also be a significant contributor to the source noise whenever the beamwidth of the receiving antenna happens to include a strong radio source in the sky.

Let us now review Fig. 2.1 in some detail. Starting with the environmental noise, we find four broad categories. Man-made noise includes the myriad of
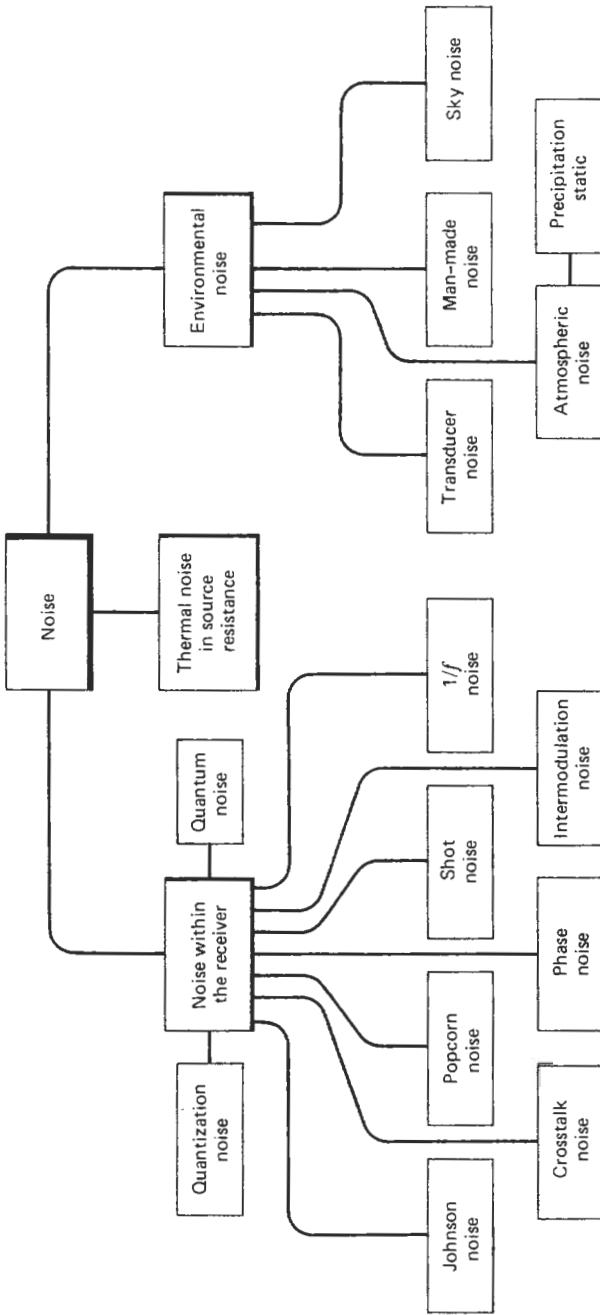

FIGURE 2.1. Principal sources of noise.

electrical disturbances found in a modern civilization, from the noise radiated by high tension wires and the ignition noise from car engines, to the interference caused by your neighbor's defective toaster. As irritating and disruptive as this noise can be, it is at least theoretically under our control. Appropriate shielding either at the source or at the receiver could be used to combat it.

Transducer noise includes those cases where a mechanism exists to convert acoustic or mechanical noise into electrical noise. Pertinent examples would be a loose antenna cable swaying in the wind or a microphonic electronic component. Again, the problem can be alleviated or eliminated entirely by an improved design and/or a better selection of components. We shall therefore not dwell any longer on these two topics. The reader is advised to consult specific literature in these areas [2].

Atmospheric noise and precipitation static both arise in earth's atmosphere. They will be discussed briefly in Section 2.8, but they are not a factor at UHF and microwave frequencies.

The next category, sky noise, is of much greater concern since it contains such strong intrinsic, wideband noise sources as the Sun, the radio stars, and the various diffuse radiations of galactic origin. A receiving antenna pointed at a satellite, for example, will also pick up whatever background sky noise exists at the satellite's location. If this radiation falls within the receiver passband and is of sufficient strength, considerable interference could result. We shall discuss sky noise and its effect in greater detail in Section 2.8. It suffices here to note that sky noise can be an important contributor to the second major category in Fig. 2.1—noise radiated by the source. It is well to note that it is the source noise which sets the minimum noise floor in a receiving system *even* if the transmission channel and the receiver were ideal and contributed no noise of their own.

The third block in Fig. 2.1 includes the large variety of noises encountered within the receiver itself. For our purposes the most important one is thermal noise, which will also be the main topic of this book. The reason for this choice is primarily that the book is intended for the system engineer involved in the design and analyses of high-frequency and microwave communication links where the lower limit of signal detection is usually set by thermal noise. The other noise mechanisms shown under the third block will therefore remain largely outside the scope of this book. We shall, however, give at least an overview of each noise type in the following sections. For a detailed treatment of a particular noise, the reader is referred to the many excellent sources in the literature.

2.1. INTERMODULATION AND CROSSTALK NOISE

Intermodulation (IM) noise can be a severe problem wherever unintentional or intentional nonlinearities exist such as saturated active devices, mixers, detector diodes, or even poorly conducting or corroded joints in passive installations

(antennas, waveguide flanges, etc.). Such nonlinearities generate either higher harmonics when a single-frequency signal is present, or a large number of $mf_1 \pm nf_2$ type in-band and out-of-band products for two or more signals of different frequencies [3]. Since these products, discrete as they may be, would nevertheless interfere with the desired signals, they are classified as noise. The obvious, but not necessarily practical or even possible, cure for the IM problem is to avoid all unnecessary nonlinear elements. More often appropriate filters are used to block the IM products, or the frequency plan is adjusted to shift the troublesome interference products outside the operating band.

Crosstalk noise is interference caused by unwanted electromagnetic coupling of signal energy from one circuit to another. Sometimes the term crosstalk is also applied to interference caused by intermodulation, particularly when the IM products fall into an adjacent signal channel. However, strictly speaking, crosstalk means coupled interference only [4].

Note that in both cases a distinct mechanism is required to generate the interfering signals. The implication is, of course, that it should be possible through appropriate design techniques to reduce such interference to as low a level as desired. In this sense intermodulation noise and crosstalk noise do not constitute intrinsic noise types.

2.2. QUANTIZATION NOISE

Quantization noise arises in pulse code modulation systems, where the amplitudes of the signal samples are not sent directly but are quantized according to a fixed number of quantum levels which, in turn, are encoded for transmission by using some binary code. The quantized signal is therefore not a true replica of the initial message,* and will differ from it to a greater or lesser degree depending on the coarseness of the quantizing process. This difference between the initial and the quantized versions of the message can be regarded as noise. It can be shown that if the peak-to-peak range of the original signal had been quantized into equal intervals of magnitude S each, the root-mean-square quantization error would be proportional to S [5]. We thus note that the quantization error is also not an intrinsic noise, but is a byproduct of the modulation process. Various techniques are available (signal companding, etc.) which lessen the impact of quantization on the signal-to-noise ratio of the system.

2.3. PHASE NOISE

Phase noise, as the name implies, is a spurious fluctuation, either discrete or random, in the phase rather than the amplitude of a signal. It is of great

*According to the sampling theorem, if a band-limited signal of $f \leqslant f_{max}$ is sampled at a rate of $f_s \geqslant 2f_{max}$, then the signal may be reconstructed from these samples without any distortion.

importance in phase-modulated systems, or wherever the frequency or phase stability must meet stringent requirements, such as in oscillators, frequency synthesizers, and so on. The output spectrum of such devices should ideally be an infinitesimally narrow line in the frequency domain. In practice we find a more or less extensive spread of energy around the desired frequency. Loosely speaking, phase noise plays the same role in the phase domain as thermal noise does in the amplitude domain. Note, however, that the ultimate mechanism for both is the same, namely, random motion of thermally agitated electrons. In the case of phase noise, these fluctuations affect the frequency-determining elements of a circuit, causing a random fluctuation in output phase of the signal generated by or transmitted through these elements. It is only as a result of the convention that the term "thermal noise" has been reserved for the amplitude fluctuations.

Phase noise is thus an intrinsic type of noise which cannot be eliminated by clever circuit design or shielding. Its effect can, however, be decreased by judicious design techniques, just as is true for the thermal amplitude noise. Phase noise constitutes an important field of study in view of its impact on angle-modulated and digital systems [6, 7].

2.4.   POPCORN NOISE

Popcorn noise derives its name from the audible effect heard when this noise is fed to a loudspeaker. It is also called burst noise and is encountered in semiconductors and integrated circuits [8]–[10]. Popcorn noise occurs in aperiodic bursts, with a repetition rate varying widely from hundreds of pulses per second to less than one pulse per minute [11]. The width of the pulse could also vary widely from microseconds to seconds, although for a given semiconductor device the pulse amplitudes remain the same. The amplitudes are typically many times higher than those of thermal noise. Burst noise is thought to be due to certain defects in the semiconductor junction which can, to an extent, be controlled by manufacturing processes. The power spectral density of popcorn noise is proportional to $1/f^a$ where $1 < a < 2$. Additional studies have shown [8] that the spectrum has the form of

$$\frac{\text{constant}}{1 + f^2\tau^2} \tag{2-1}$$

where f is frequency and τ is a parameter inversely proportional to the number of bursts per second. This expression shows that popcorn noise tends to level off at very low frequencies, contrary to the flicker or $1/f$ noise to be described in the next section.

2.5.   FLICKER OR $1/f$ NOISE

Flicker or $1/f$ noise is a low-frequency phenomenon which is typically encountered in nonequilibrium systems, that is, those in which a dc current is

flowing [12]. The name flicker noise stems from early vacuum-tube work where a flicker-type fluctuation was noticed in the anode current. Flicker noise has the interesting characteristic that its spectral density is inversely proportional to frequency. Theoretically, the noise power density would approach infinity as frequency approaches zero. Indeed, the $1/f$ characteristic has been found to hold at frequencies as low as a few cycles per day [13, 14].

A more precise description of this noise would be $1/f^n$ where the value of n is typically $0.8 < n < 1.3$, although $n \approx 1$ seems to be the more common value [15]. The $1/f$ noise is a poorly understood phenomenon. There is no one theory that would fit all observed cases [16]. This noise seems to be related to surface properties of materials, and is also associated with imperfect contact between conductors, such as carbon granules in a microphone or in a carbon resistor [17]. In the latter case the term "excess noise" is often used to emphasize that the noise is over and above the thermal noise normally obtained from the resistor. Wire-wound resistors, for example, do not exhibit $1/f$ noise. Interestingly enough, there is evidence that $1/f^n$ type noise is quite prevalent throughout nature. It has been observed in certain biological systems and other nonelectrical processes where n may range as high as 2.7 (cosmic radio noise at VHF [15]).

In electronic circuits the sensitivity of audio and dc amplifiers is ultimately limited by flicker noise. In this connection it is well to note that the measurement accuracy of a system limited by flicker noise cannot be improved by extending the measuring time if $n \geq 1$. This follows from Rayleigh's energy theorem which relates the energy of a signal to the integral of the square of its amplitude spectrum. Thus

$$E = \int_\varepsilon^\infty \left(\frac{1}{f^n}\right)^2 df = \frac{T^{(2n-1)}}{2n-1} \tag{2-2}$$

where $\varepsilon = 1/T$ designates an arbitrary lower limit of frequency and T is the corresponding period. Clearly, for $n \geq 1$ the total accumulated energy of the flicker noise increases at least as fast as the measuring time.

Van der Ziel [18] gives the following expression for the mean-square noise current per unit bandwidth:

$$\overline{i^2} = K\left(\frac{I^a}{f^n}\right) df \tag{2-3}$$

where K is a constant that depends on the type of material, I is the average dc current, the exponent a is close to 2, and n is usually close to unity. This expression seems to hold for a wide variety of cases, including semiconductors, carbon microphones, photoconductors, crystal diodes, and so on. The frequency above which the $1/f$ noise could be neglected depends to a large degree on the

physical parameters involved. In general, the flicker or $1/f$ noise is of no consequence at radio frequencies.

2.6. SHOT NOISE

Shot noise was the earliest electrical noise mechanism recognized to be intrinsic in the sense that its origin was not man-made. The discovery came in the course of working with early vacuum tubes where shot noise was observed in the form of spontaneous current fluctuations superimposed on the dc plate current. This phenomenon was soon determined to be due to the particle character of the electric current itself [12]. As the electrons are emitted from the cathode at a *random* rate, each carrying a charge q, the resultant flow of current is really granular in a microscopic sense, giving rise to the observed random fluctuation around the average dc value. One may wonder why shot noise is not observed in a conductor. A rather lucid explanation is given by Bennett [19] who points out that since electrical neutrality must be preserved within the conductor, for any electron leaving a section of the conductor, another electron must enter that section. On the other hand, for electron streams in a vacuum, the electrons can leave the conductor (consisting of the cathode, the anode, and the connecting external circuitry) independently at random. This results in random fluctuations of current as measured in the external circuit.

We thus observe that shot noise arises in systems not in equilibrium, that is, a flow of discrete electrical charges is required. This is in contrast to thermal noise as we shall soon see. In the latter case the system is in a thermodynamic equilibrium.

W. Schottky in his pioneering work (ref. [2] in Chapter 1) showed that the mean-square value of the shot noise in a temperature-limited thermionic diode can be expressed as

$$\overline{i^2} = 2qI_s \, df \quad (\text{ampere}^2) \tag{2-4}$$

where q is the electron charge, 1.59×10^{-19} coulombs; I_s the saturation current in amperes; and df the bandwidth in Hz.

Equation (2-4) shows that the power spectral density of shot noise is given by

$$P_n' = 2qI_s \quad (\text{ampere}^2/\text{Hz}) \tag{2-5}$$

that is, the spectrum is flat with frequency [12]. The spectrum of shot noise would theoretically extend to infinity. Clearly, in physical world the spectral density must fall off eventually. Indeed, the reason for this nonphysical

conclusion is the idealized assumption used in deriving (2-4), that the transit times of the electrons flying from the cathode to the anode are infinitesimally short. The current pulse corresponding to each electron flight is therefore taken as a delta or impulse function. Such a function does have an infinite or "white" spectrum. In practice it is the finite transit time which causes a roll-off in the spectrum of shot noise.

Equation (2-4) is subject to certain limitations. First, it is true only for the temperature-limited case. That is, the anode voltage must be high enough so that *all* electrons emitted by the cathode at the particular operating temperature are collected by the anode. Second, there is no space charge between the cathode and the anode, that is, the cathode-to-anode voltage is the only force acting on the electrons. Third, the electrons are emitted randomly, and the correlation between the times of emission of individual electrons is therefore zero. In fact, the shot noise actually observed may be greater or less than that predicted by (2-4) depending on the degree of correlation.

We have been referring to thermionic diodes as examples of shot noise sources. Actually, according to Schottky, (2-5) holds quite generally whenever the current consists of independent random events, each carrying a charge of q. It is therefore applicable also to p-n junction diodes, bipolar transistors, metal–semiconductor (Schottky-barrier) diodes, and so on, where charges are carried across potential barriers [20]. It is of interest to note that (2-4), when applied to a p-n junction, can be shown [21] to lead to the familiar expression $P = kTB$ of thermal noise when the applied voltage and hence the total current through the diode are zero. This is not surprising because in this case the diode is at thermal equilibrium and would behave like any other resistor. The next chapter will show that the expression given for P above is the thermal noise power available from a resistor at temperature T.

Shot noise is important to us for two reasons. First, as will be shown in Chapter 11, a source having a white noise spectrum similar to that of thermal noise is very useful in measuring the noise temperature or noise figure of an amplifier or any linear receiver component for that matter. Second, according to (2-4) the rms value of the shot noise can be easily calculated from the measured dc current I_s. Furthermore, the level of shot noise can be varied by simply varying the current.

The topic of shot noise is an important one and a vast body of literature exists on the subject (see, in particular, ref. [1]).

2.7. QUANTUM NOISE

As will be shown in Section 3.2, thermal noise has a flat noise density spectrum which at room temperature extends well into the infrared. As the frequency increases, however, a point is reached where the flat spectrum begins to fall off. As this happens, we would theoretically need less and less signal power to transmit a message at a given signal-to-noise ratio. This follows from the

Shannon–Hartley theorem [22] according to which the channel capacity of a white band-limited gaussian channel is

$$C = B \log_2\left(1 + \frac{S}{N}\right) \quad \text{(bits/second)} \tag{2-6}$$

where B is the channel bandwidth, S the signal power, and N the total noise power within the bandwidth B. It is noteworthy, however, that in this region another type of noise—the quantum noise—makes itself felt. This noise has its roots in the quantum physics, and in Heisenberg's uncertainty principle. Its power spectral density is given by

$$W(f) = hf \quad \text{(watts/Hz)} \tag{2-6a}$$

where h is Planck's constant, 6.63×10^{-34} joule-second, and f the frequency in Hz. The quantum noise is thus entirely negligible at frequencies normally encountered in engineering applications, but becomes increasingly significant at optical frequencies. It is of interest to note that the concept of quantum noise leads in turn to the minimum noise temperature achievable in a linear amplifier. Excellent discussions of these topics can be found in Oliver [21] and Heffner [23].

2.8. SKY NOISE AND NOISE IN THE EARTH'S ATMOSPHERE

For simplicity, sky noise, atmospheric noise, and precipitation static are grouped together in this section. Thus in a broad sense we are considering here sky noise to be due to a variety of terrestrial and extraterrestrial sources.

The most obvious type is the atmospheric noise, caused mainly by lightning discharges. As such it depends on time of day, location, and seasons, being higher in tropical areas and decreasing in intensity as one moves to higher latitudes. Atmospheric noise is of importance mainly below 20 MHz [24]. It does not affect operation at UHF and microwave frequencies.

Another form of atmospheric noise is precipitation static encountered in rain, hail, snow, and dust storms in the vicinity of the antenna. This noise is important to perhaps 10 MHz or so. It can be reduced by eliminating sharp points on and near the antenna.

When a directional antenna is pointed toward the sky, we find that random noise of widely varying intensity is received. It was 1932 when K. G. Jansky of Bell Telephone Laboratories first discovered [25] the existence of extraterrestrial noise coming from a galaxy. He was also the first to investigate it quantitatively. In view of the origin of this noise, it was also called cosmic noise. Cosmic noise is caused by a number of discrete and distributed sources, the principal ones being the Sun, the ionized interstellar gas clouds in our Galaxy, and a vast number of galactic and extragalactic sources such as supernova remnants, radio galaxies, quasars, and so on [26]. The quasars are

particularly strong radiators, emitting enormous amounts of energy, the origin of which is still not completely understood. All these sources radiate in a continuous mode, although some quasars are known to be variable.

Two broad classes of emission are found—thermal and nonthermal. The first type is associated with encounters between electrons and ions in ionized gas clouds, mostly hydrogen. The second type, also called the synchrotron radiation, originates as a result of electrons moving in magnetic fields [26]. High-level nonthermal radiation is therefore associated with discrete cosmic radio sources where intense magnetic fields are found. Weak nonthermal radiation is distributed throughout the Galaxy, indicating that high-energy electrons and magnetic fields are present even in interstellar space.

The wavelength dependence of cosmic noise is a function of whether the noise is thermal or nonthermal in origin [26]. The physics of cosmic noise and radio astronomy in general are very broad topics and will not be treated here. To us the importance of cosmic noise lies in the fact that it becomes effectively the generator or source noise whenever a directional antenna is pointed toward the sky to receive a satellite signal. The beamwidth of the antenna is obviously much larger than the angular diameter of the satellite seen from the ground. The antenna thus receives not just the desired signal from the satellite, but background sky noise as well. Radio-noise maps have been compiled for the entire sky. Some examples are shown in ref. [24]. Note in passing that because of the diurnal motion of the sky, a given background is not constant with time, even for a stationary ground antenna pointed toward a geostationary satellite.

The most intense discrete cosmic noise source is, of course, the Sun. Its noise temperature varies quite substantially with frequency [27]. The noise temperature of the quiet (no sunspot activity) Sun reaches some $700,000°K$ at 200 MHz, but only $6000°K$ at 30 GHz (Fig. 2.2). It should be realized that these values are measured only if the beamwidth of the antenna is smaller than the angle subtended by the Sun's disc (approximately 0.5 degrees). Otherwise the antenna sees a part of the much colder sky background, and the measured average temperature over the beamwidth goes down accordingly. An 85 ft diameter antenna dish has been measured to have a noise temperature of $40,000°K$ at 400 MHz [24] when pointed directly at the Sun. The same antenna had a noise temperature of $600°K$ when looking at the cosmic radio source Cassiopeia A.

Broadly speaking, the cosmic radio noise extends from approximately 15 MHz to 100 GHz. It is dominant in the 40–250 MHz region (below about 20 MHz other contributions, such as atmospheric noise, take over). Hogg and Mumford [27] give the following approximate limits for the sky-noise temperature in the VHF range:

Maximum:	$T_s = 1450\lambda^2$	(2-7)
Average:	$T_s = 100\lambda^{2.4}$	(2-8)
Minimum:	$T_s = 58\lambda^2$	(2-9)
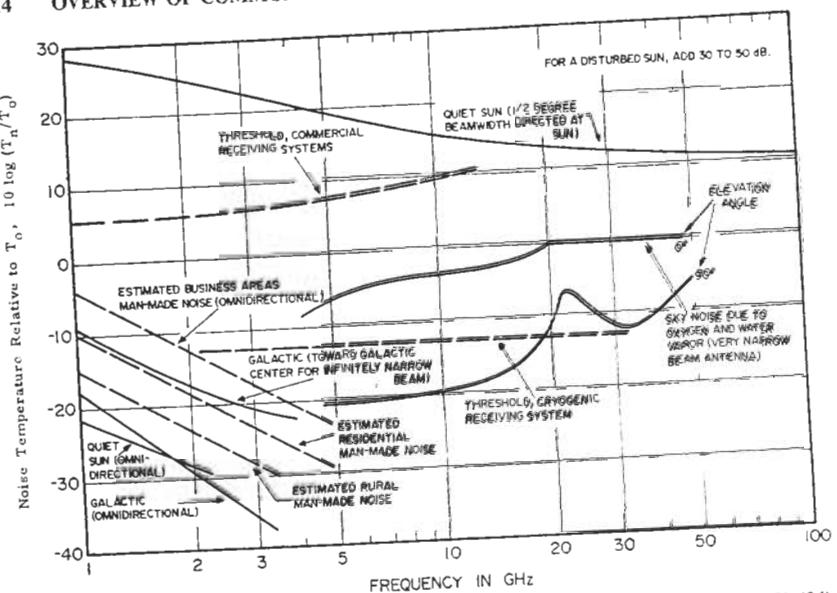
FIGURE 2.2.   Typical noise temperatures for various radio-noise sources (from MIL-HDBK-416).

FIGURE 2.3.   Atmospheric attenuation.

where the wavelength is expressed in meters. The cosmic noise also depends on the direction of the antenna, being highest when looking toward the galactic center of the Milky Way.

As the frequency increases, the cosmic noise decreases roughly in accordance with (2-8), reaching the standard or "room" temperature of 290°K somewhere around 300 MHz. At still higher frequencies, in the commonly used microwave region of 1–10 GHz, the cosmic noise from extraterrestrial sources (except certain discrete radio sources and the Sun) could be anywhere between approximately 3° and perhaps 100°K. The lower limit quoted above, discovered by A. A. Penzias and R. W. Wilson in 1965 and theoretically (and independently) predicted from cosmological considerations by R. H. Dicke, is thought to be the lingering echo from the primordial fireball of the "big bang" creation of the universe some 15 billion years ago.

The actual sky-noise temperature observed in the higher microwave region with a narrow-beam terrestrial antenna is noticeably higher than these numbers would indicate. The characteristic of the sky-noise temperature has a broad minimum as seen in Fig. 2.2, and begins to rise again as the frequency increases beyond 10 GHz. This behavior is due to increasing absorption of radio waves in the earth's atmosphere as a consequence of water vapor and oxygen. The end result is thus also dependent on temperature and atmospheric pressure. The increase in noise temperature as the intervening atmospheric losses increase will be understood more clearly when Chapter 8 is covered.

Figure 2.3 illustrates the atmospheric absorption effect at SHF and EHF frequencies.

Returning to Fig. 2.2, it is noted that there is a marked dependence of sky noise on the elevation angle of the antenna. This is not surprising, because at low elevation angles the antenna looks through much more atmosphere and the attenuation varies accordingly.

Finally, the measured results are also a function of the antenna sidelobe level. If the antenna has prominent sidelobes which are pointed toward the warm earth or the warm atmosphere layer near the horizon, then the measured noise temperature of the sky could be substantially higher than expected, even though the main beam of the antenna is directed toward a cold portion of the sky. We shall discuss this in greater detail in Section 5.2, but it should be apparent that very low sidelobes in addition to a high antenna directivity are required for accurate measurements of sky noise.

REFERENCES

1. *Electrical Noise: Fundamentals and Sources*, M. S. Gupta (Ed.), IEEE Press, New York, 1977, p. 125.

2. H. W. Ott, *Noise Reduction Techniques in Electronic Systems*, Wiley, New York, 1976.

3. B. Oliver, "Distortion and Intermodulation," Hewlett-Packard Application Note No. 15.

4. *Reference Data for Radio Engineers*, 6th ed., Sams/ITT, Indianapolis, 1975, Chapter 35.

5. H. Taub and D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, New York, 1971, pp. 201–203.

6. W. P. Robins, *Phase Noise in Signal Sources*, Peter Peregrinus Ltd., London, 1982.

7. "Understanding and Measuring Phase Noise in the Frequency Domain," Hewlett-Packard Application Note No. 207, October 1976.

8. A. van der Ziel, *Noise: Sources, Characterization, Measurement*, Prentice-Hall, Englewood Cliffs, N.J., 1970, pp. 115–117.

9. W. H. Card and P. K. Chaudhari, "Characteristics of Burst Noise," *Proc. IEEE*, vol. 53, pp. 652–653, June 1965.

10. C. D. Motchenbacher and F. C. Fitchen, *Low-Noise Electronic Design*, Wiley, New York, 1973, pp. 95–97.

11. H. W. Ott, ref. [2], pp. 209–210.

12. J. R. Pierce, "Physical Sources of Noise," *Proc. IEEE*, vol. 44, pp. 601–608, May 1956.

13. J. E. Firle and H. Winston, *Bull. Am. Phys. Soc.*, vol. 30, No. 2, 1955.

14. W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960, pp. 101–109.

15. D. Halford, "A General Model for f^a Spectral Density Random Noise with Special Reference to Flicker Noise $1/f$," *Proc. IEEE*, vol. 56, pp. 251–257, March 1968.

16. A. Ambrózy, *Electronic Noise*, McGraw-Hill, New York, 1982, p. 113.

17. H. W. Ott, ref. [2], p. 209.

18. A. van der Ziel, *Noise*, Prentice-Hall, Englewood Cliffs, N.J., 1954, p. 209.

19. W. R. Bennett, ref. [14], p. 87.

20. A. van der Ziel, ref. [8], pp. 14–15 and 90–100.

21. B. M. Oliver, "Thermal and Quantum Noise," *Proc. IEEE*, vol. 53, pp. 436–454, May 1965.

22. H. Taub and D. L. Schilling, ref. [5], pp. 421–423.

23. H. Heffner, "The Fundamental Noise Limit of Linear Amplifiers," *Proc. IRE*, vol. 50, pp. 1604–1608, July 1962.

24. Reference [4], Chapter 29.

25. K. G. Jansky, *Proc. IRE*, vol. 25, p. 1517, December 1937.

26. E. L. Schatzman, *The Structure of the Universe*, McGraw-Hill, New York, 1968.

27. D. C. Hogg and W. W. Mumford, "The Effective Noise Temperature of the Sky," *Microwave J.*, pp. 80–84, March 1960.

3 | THERMAL NOISE

We shall now turn our attention to thermal noise, which is the main topic in this book. Thermal noise is the earliest intrinsic noise to be studied in detail. It was in 1828 when Robert Brown published his famous paper on what came to be known as the Brownian movement—the irregular agitation of microscopic solid particles suspended in liquid, observable under a powerful microscope. At first the reason for this strange activity was thought to lie within the particles themselves. It took a long time before it was finally recognized that the real reason for the movement was the random impact of the thermally agitated molecules of the fluid on the suspended particles. Quantitative analyses of the phenomenon did not appear until the end of the 19th century.

Once this breakthrough had been achieved, the investigation extended rapidly to other areas. In particular, it was realized that in accordance with the kinetic theory of heat, free electrons within a conductor had to be in continual random motion also as a consequence of collisions with thermally agitated molecules. Visualize one electron set in motion by such a collision. As the electron accelerates, the changing magnetic field will set other electrons in motion, and so on, until the agitation is spread throughout the entire electron gas within the conductor. Since moving charges constitute electric current, such motion of electrons results in minute, random current pulses. Indeed, this was found to be the case except that the advent of high-gain vacuum–tube

amplifiers was required before the existence of these weak currents could be demonstrated.

The number of randomly moving electrons within a conductor is exceedingly large. Since each flight of each electron between successive collisions with the molecules generates a current pulse, and since the mean-square velocity of the moving electrons is proportional to the absolute temperature, a noise-like, temperature-dependent voltage appears at the terminals of the conductor. The average value of this noise voltage must be zero because otherwise there would be an accumulation of charges at one end of the conductor. The entire system is therefore at thermal equilibrium. The thermal noise has no dc component, but its ac portion does have a well-defined root-mean-square (see Section 3.1) value.

Chapter 2 outlined a number of other noise mechanisms which exist in nature. If we eliminated all man-made noises, all environmental noises (through perfect shielding), and all those noises due to manufacturing deficiencies in our devices, then at commonly encountered temperatures and frequencies the ultimate minimum background noise level would be set by the thermal noise. The latter is the limiting factor in any transmission of low-level electrical signals and therefore plays a very important role in the design of any electronic communication system.

3.1. BASIC CONCEPTS

Before beginning a detailed discussion of thermal noise, it is necessary to refresh our understanding of such fundamental concepts as the root-mean-square value of a waveform, and correlation. Both terms were introduced in previous sections, but an explanation was not given. We shall now describe these in greater detail.

Root-Mean-Square Value

In elementary ac circuits theory, the concept of power associated with an ac waveform, $V\cos \omega t$, is handled by finding an equivalent dc value which, when impressed across a resistance R, yields the same heating effect as the ac wave. This equivalent dc value is calculated by averaging a squared sine wave over its period (see Fig. 3.1):

$$\frac{V_{eq}^2}{R} = \frac{V^2 \int_0^T \cos^2 \omega t\, dt}{TR} \tag{3-1}$$

or

$$V_{eq} = \frac{V}{\sqrt{2}} \equiv V_{rms} \tag{3-2}$$
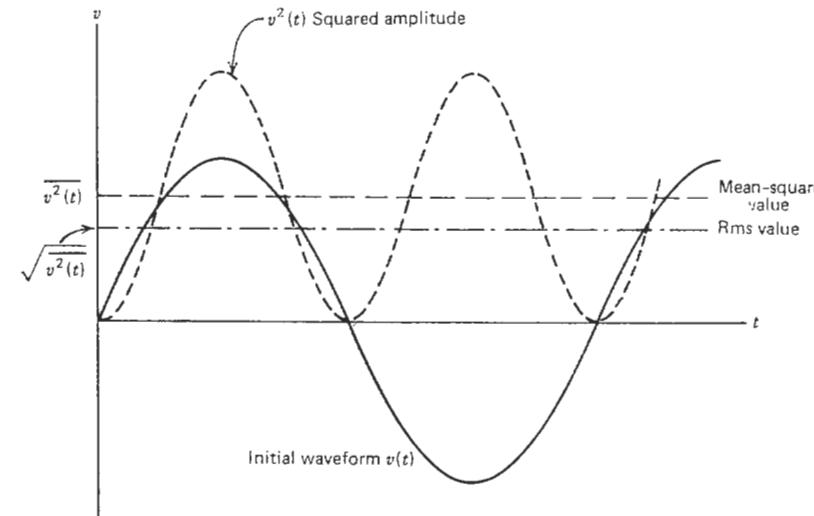
FIGURE 3.1.   Mean-square and rms values of a sinusoid.

Since V_{eq} was obtained by a multiple operation as illustrated in Fig. 3.1, the result is called the root-mean-square value of the given voltage, current, or whatever function we happen to be considering. Symbolically, it is written as

$$V_{rms} = \sqrt{\overline{v^2}} \qquad I_{rms} = \sqrt{\overline{i^2}} \tag{3-3}$$

The term "mean square" is simply the square of (3-3), or

$$\overline{v^2} = V_{rms}^2 \qquad \overline{i^2} = I_{rms}^2 \tag{3-4}$$

The power dissipated in a resistor by a sinusoidal signal $V\cos \omega t$ can now be calculated very simply from $P = V_{rms}^2/R$ just as in a dc case. Thus, regardless of the waveform, once the rms value of the waveform is known, the usual rules of linear circuit theory can be used to calculate the rest. The mean-square value of either voltage or current, $\overline{v^2}$ or $\overline{i^2}$, can also be considered as power dissipated in a 1-ohm resistor. For this reason, we shall from time to time refer to $\overline{v^2}$ and $\overline{i^2}$ as the normalized power of $v(t)$ and $i(t)$.

In case of a deterministic waveform, such as a sine wave, there are two choices. We could describe the wave either by means of an explicit function of some variable, typically time, or we could use its rms value. When dealing with nondeterministic, random phenomena, such as thermal noise voltages, we no longer have the luxury of a closed analytical expression. The instantaneous value of the waveform is totally unpredictable and an explicit functional

relationship does not exist. We can, however, still talk about the "heating" effect or power of such a wave by carrying out, conceptionally at least, a squaring and averaging operation on its recorded time characteristic. There is no identifiable period T any more, but we get around this problem by taking a sufficiently long period of time t (or whatever independent variable applies) to ensure the statistical validity of the calculated average. The rms technique thus provides a convenient and physically meaningful measure of random functions, and allows us to characterize such waveforms without knowing their precise instantaneous amplitudes.

It may appear from this discussion that the amplitude information of a random waveform remains an elusive quantity. We shall see in Chapter 4 that this is not so. As it turns out, random processes, such as the thermal noise, obey well-defined statistical rules which allow us to calculate their amplitude distribution within a definite probability of being correct.

Correlation

This section will give a broad overview of the concept of correlation. The latter term plays an important role in signal and noise analysis, and it is therefore necessary to have at least a cursory understanding of what correlation means. The following discussion will not always be mathematically rigorous, but it is hoped that this deficiency would be outweighted by a greater clarity in presentation which makes it easier to grasp the fundamentals. The way would thus be paved for those who wish to seek additional details in more advanced texts.
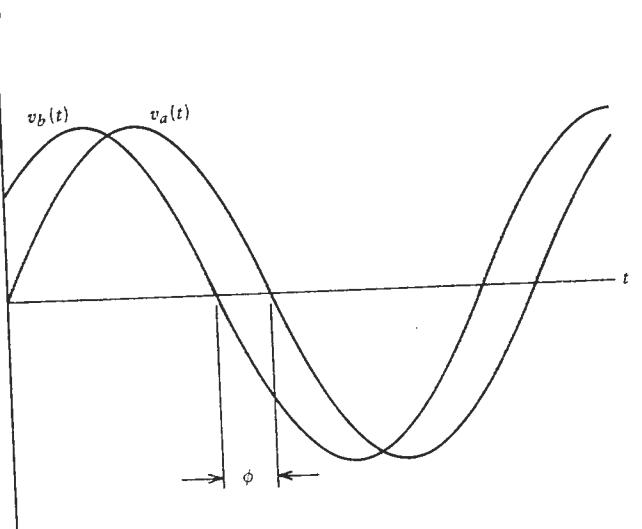

FIGURE 3.2.   Two sinusoidal voltages.

In a broad sense, correlation is a measure of similarity between two comparable entities. Before giving a more formal definition of correlation, let us consider a simple example. Take two sinusoidal voltages $v_a(t)$ and $v_b(t)$ as in Fig. 3.2, which have the same frequency, but may have different amplitudes. If an arbitrary phase shift $\phi = \omega\tau$ is allowed between them, then

$$v_a(t) = A \cos \omega t \tag{3-5a}$$

and

$$v_b(t) = B \cos(\omega t + \phi) \tag{3-5b}$$

Suppose these two voltages are combined across a unit resistance and we wish to calculate the normalized average power dissipated in the resistance. In accordance with the previous section, we must first calculate the mean-square value of the total voltage. This is given by

$$\overline{v_T^2} = \overline{(v_a + v_b)^2} = \overline{v_a^2} + 2\overline{v_a v_b} + \overline{v_b^2} \tag{3-6a}$$

$$= \frac{1}{T}\int_0^T \left[A^2\cos^2\omega t + 2AB\cos\omega t\cos(\omega t + \phi) + B^2\cos^2(\omega t + \phi) \right] dt$$

$$\tag{3-6b}$$

where the bar denotes the time average of the particular term.*

At this point it is necessary to digress briefly to Chapter 4 where statistical parameters such as the mean and the variance are discussed. As will be shown there, the mean or average of a sum is always equal to the sum of the means. It is this rule that permits us to break up (3-6a) in the manner shown. However, the mean of a *product*, $\overline{xy}$, is equal to the product of the means, $\bar{x}\,\bar{y}$, *only* if the functions x and y are statistically independent. An alert reader may have been forewarned of this. Although Fig. 3.2 clearly shows both means, $\bar{v}_a$ and $\bar{v}_b$, to be zero, the mean of the product in (3-6b) will not be zero as we shall see 'shortly. This is because the two sinusoids are obviously not independent. Given a particular value of v_a, that is, a value of t from (3-5a), the corresponding value of v_b can be calculated from (3-5b).

The first and last terms in (3-6b) integrate as expected to $\frac{1}{2}A^2$ and $\frac{1}{2}B^2$, respectively, the mean-square values of the two individual voltages v_a and v_b. But, what about the middle term? Completing the integration yields

$$\frac{2AB}{T}\int_0^T \cos\omega t\cos(\omega t + \phi)\,dt = AB\cos\phi \tag{3-7a}$$

---

*This is done for simplicity. The usual symbol for time average is $\langle v_T \rangle$, the bar sign being used for ensemble averages (Section 4.4). In our case the meaning is clear from the text and no confusion should result.

and therefore

$$\overline{v_T^2} = \tfrac{1}{2}A^2 + \tfrac{1}{2}B^2 + AB\cos\phi \qquad (3\text{-}7b)$$

If $\phi = 0$, the two sinusoids will coincide as is apparent from Fig. 3.2, and

$$\overline{v_T^2} = \frac{A^2}{2} + \frac{B^2}{2} + AB = \left(\frac{A}{\sqrt{2}} + \frac{B}{\sqrt{2}}\right)^2 \qquad (3\text{-}8)$$

For this special case only, the rms value of the total voltage equals the sum of the individual rms values. The waveforms are said to be *completely correlated*. Another synonymous term often encountered is that the waveforms are coherent.

If $\phi = \tfrac{1}{2}\pi$, the term $\cos\phi$ in (3-7b) becomes zero and

$$\overline{v_T^2} = \tfrac{1}{2}A^2 + \tfrac{1}{2}B^2 \qquad (3\text{-}9)$$

This time the mean-square value of the total voltage equals the sum of the mean-square values of the component voltages. Since the mean-square value of mean-square is equal to its normalized power, the total power dissipation in the a sinusoid is simply the sum of normalized powers of v_a and v_b. The waveforms are now said to be *completely uncorrelated* or incoherent.

In the general case $0 < \phi < \tfrac{1}{2}\pi$, which implies that one wave may be shifted by an *arbitrary* amount. From the preceding discussion it seems logical that the correlation between v_a and v_b, however expressed, should now lie somewhere between these two extremes. In other words, what is needed, is a more quantitative measure of correlation. Let us therefore define [1] a correlation coefficient ρ such that

$$\rho = \frac{\overline{xy}}{\sqrt{\overline{x^2}\,\overline{y^2}}} \qquad (3\text{-}10)$$

where x and y are arbitrary functions, either periodic or random. In our specific example,

$$\rho = \frac{\overline{v_a v_b}}{(v_a)_{\mathrm{rms}}(v_b)_{\mathrm{rms}}} \qquad (3\text{-}11)$$

Using (3-10), (3-6a), and (3-7b),

$$\overline{v_T^2} = \overline{v_a^2} + \overline{v_b^2} + 2\rho(v_a)_{\mathrm{rms}}(v_b)_{\mathrm{rms}} \qquad (3\text{-}12a)$$

$$= \frac{A^2}{2} + \frac{B^2}{2} + 2\left(\frac{A}{\sqrt{2}}\right)\left(\frac{B}{\sqrt{2}}\right)\cos\phi \qquad (3\text{-}12b)$$

from which, for this particular case,

$$\rho = \cos\phi \qquad (3\text{-}13)$$

We can now see from (3-8) and (3-9) that the correlation coefficient ρ becomes $+1$ for fully correlated waveforms where the individual waveforms can be added arithmetically, and becomes zero for completely uncorrelated cases where the mean-square values must be added. The case of $\rho = -1$ also implies full correlation, but the two waveforms now subtract from each other [2] as seen from Fig. 3.2 when $\phi = \pi$. It is to be emphasized that the range of ρ, namely, from $+1$ to -1, is a general property of ρ and not just due to this example where ρ happened to equal $\cos\phi$.

The parameter ρ provides the desired measure of correlation and we have seen what happens when ρ is either 1 or 0. Clearly, then, $0 < |\rho| < 1$ means that a *partial* correlation between the waveforms $x(t)$ and $y(t)$ exists. It can be shown [3] that in this case one function or waveform can be split in two parts—one part fully correlated with the other function and the other part completely uncorrelated. Although a detailed treatment of this topic is outside the scope of this book, we shall have an occasion to use the technique in Section 9.9.

The correlation coefficient ρ has been briefly covered here, because it is often encountered in papers on thermal noise. We should, however, go one step farther. When Eqs. (3-9) and (3-10) are examined, it becomes evident that it is the term $\overline{xy}$, that is, the time average of the product of $x(t)$ and $y(t)$, that holds the key as far as correlation is concerned. The rms values $\sqrt{\overline{x^2}}$ and $\sqrt{\overline{y^2}}$ in (3-10) pertain to the respective functions only and are independent of how $x(t)$ might be related to $y(t)$. For this reason, the specific mathematical tool for analyzing statistical correlation between two arbitrary functions is the cross-correlation function given by

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\,y(t+\tau)\,dt \qquad (3\text{-}14)$$

Note that this is again in the form of a time averaging of the product of $x(t)$ and $y(t)$, but two important features have been added which render (3-14) much more powerful than the simple time average of the product used in (3-6b). In the first place, the limiting process $T \to \infty$ permits the analysis of nonperiodic signals, and, more importantly, the parameter τ has been explicitly introduced. The purpose of this parameter is to provide a scanning "window" so that the correlation could be examined for *all* possible time-shifted positions of x and y. The cross-correlation function, through its independent variable τ, enables us to uncover any time-shifted similarities between x and y which the one-time calculation of ρ may miss. In (3-13) the role of τ was played by the angle ϕ. We saw in our example that as ϕ varied from 0 to 2π, the correlation between v_a and v_b varied cyclically from full

(positive or negative) to none at all. It is therefore premature to conclude on the basis of one value of ρ or τ alone that a correlation either does or does not exist.

Consider therefore two waveforms, not necessarily periodic, and write the expression for the normalized power S_{ab} as in (3-6b), but use (3-14) instead:

$$S_{ab}(\tau) = \overline{v_T^2} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} [v_a(t) + v_b(t + \tau)]^2 \, dt$$

$$= \lim_{T \to \infty} \frac{1}{T} \left[\int_{-T/2}^{T/2} v_a^2(t) \, dt + \int_{-T/2}^{T/2} [v_b(t + \tau)]^2 \, dt \right.$$

$$\left. + 2 \int_{-T/2}^{T/2} v_a(t) v_b(t + \tau) \, dt \right] \tag{3-15}$$

In view of (3-14) this becomes

$$S_{ab}(\tau) = S_a + S_b + 2R_{ab}(\tau) \tag{3-16}$$

where because of $T \to \infty$ (meaning that the entire time axis is eventually included) we have

$$\lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} [v_b(t + \tau)]^2 \, dt = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} v_b^2(t) \, dt = S_b \tag{3-17}$$

Equation (3-16) yields the important result that two waveforms are unconditionally uncorrelated only if $R_{ab}(\tau) = 0$ for all τ.

We have been talking rather loosely about statistical independence and correlation without regard to their mutual relationship. A brief clarification is therefore in order, particularly when it is sometimes carelessly assumed that one implies the other. For a complete discourse of statistical dependence and independence the reader should consult specific texts on probability and statistics. For our purposes it suffices to state here (and we shall say more about it in Chapter 4) that if two functions $x(t)$ and $y(t)$ are *statistically independent*, that is, the occurrence or value of x does not influence the occurrence or value of y and vice versa, then they are also uncorrelated (see Section 4.4). The converse is *not* necessarily true. We can easily devise situations where the correlation is zero, but the two processes or functions $x(t)$ and $y(t)$ are very much dependent. A simple example would be our two sinusoidal waves in Fig. 3.2 at a relative phase angle $\phi = \frac{1}{2}\pi$. Uncorrelated? Certainly. Independent? No.

Before leaving this topic, let us also consider the autocorrelation function which measures the correlation of a waveform with itself. In other words, if a replica of a given waveform is time shifted with respect to the original by a time interval τ, the autocorrelation function will tell us what degree of

similarity may exist between the two at some arbitrary value of τ. Mathematically, the autocorrelation function is stated as

$$R(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} v(t) v(t + \tau) \, dt \tag{3-18}$$

from which it follows directly that for $\tau = 0$,

$$R(0) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} v^2(t) \, dt = S \tag{3-19}$$

where S is the average normalized power. This is not surprising because, after all, maximum similarity or correlation would be expected whenever there is a perfect coincidence.

It is not difficult to show [4] that for a periodic waveform the power spectral density (which is in the frequency domain) and the autocorrelation function (in τ domain) form a Fourier pair.* A rather simple illustration would again be a single-frequency sinusoid $v = V\cos \omega_0 t$. Its power spectral density in frequency domain could be readily calculated from the Fourier transform (see, e.g., ref. [4], p. 19). This is unnecessary because from elementary considerations we already know that any sinusoid of constant amplitude and of unlimited duration in time domain is represented in frequency domain by an impulse[†] function located at $\omega = \omega_0$. According to the relationship just stated, the autocorrelation $R(\tau)$ is calculated by taking the inverse Fourier transform of this impulse function, with the parameter t replaced by τ. But in doing so we are retracing our earlier steps, because by the same argument, namely, that an impulse in one domain implies constancy in the other, we expect to get—and we do—a constant-amplitude sinusoid for the autocorrelation function when the starting point is an impulse. The result therefore is that $R(\tau)$ of a periodic function is also periodic. A simple analysis would show that the period of $R(\tau)$ is the same as that of $v(t)$. To see this graphically, turn to Fig. 3.2, equalize the amplitudes (although this is not essential) and consider what happens to the value of (3-7a) as ϕ varies through 2π. It is then easily seen that the autocorrelation function $R(\tau)$ will be just another sinusoid with a period of $T = 1/\omega_0$.

It is possible to show [5] that the result described above holds for random waveforms, such as thermal noise, as well. Again using a qualitative argument, we can deduce that since the power spectral density of thermal noise is ideally

---

*This is also known as the Wiener–Kinchine theorem.

[†] It should be emphasized that we are taking considerable liberties with mathematics here in order to keep the discussion simple. The spectrum of a sinusoid $V\cos \omega_0 t$ consists actually of *two* impulses at $+\omega_0$ and $-\omega_0$, respectively, since the Fourier transform results in the customary complex representation, having a two-sided spectrum. The reader is urged to review pp. 53–59 in ref. [8], which contain an excellent discussion of the nature and the use of impulse functions.

flat [see (3-28)], the autocorrelation function of thermal noise must be an impulse function. Physically this means that for an *unlimited* band of thermal noise there is no correlation whatever between the noise and its shifted replica except for the infinitesimally brief time interval when the two coincide perfectly, that is, at $\tau = 0$. Thus if we took two different samples of thermal-noise waveform at, say, τ_a and τ_b, these samples would always be completely uncorrelated, no matter what the τ are. Moreover, anticipating Chapter 4, we could add that the samples are also statistically independent.

In the real world we do not have unlimited bands of noise, the given sample of thermal noise being always more or less limited to a finite bandwidth $\Delta\omega$. The autocorrelation function is then no longer an infinitesimal spike, but assumes [6] a shape proportional to

$$\frac{\sin \tau \left(\frac{1}{2}\Delta\omega\right)}{\frac{1}{2}\Delta\omega} \tag{3-20}$$

This is not hard to see either. Consider decreasing the bandwidth of a noise waveform gradually until in the limit only one frequency remains. The autocorrelation function, originally an impulse function, degenerates into a less and less rapidly varying $(\sin x)/x$ function. In the limit, as the noise bandwidth becomes very narrow, the autocorrelation function approaches a steady, equal-amplitude sine wave as we saw in the example.

It is hoped that this brief tutorial introduction has given the reader some insight into the terminology and techniques used in the analysis of noise. As stated above, we shall return to some of these concepts in Chapter 4, but in the meantime the material just presented should answer at least a few of the basic questions.

3.2. THERMAL NOISE IN LUMPED CIRCUITS

A significant milestone in the study of thermal noise was reached in 1927 when J. B. Johnson of Bell Telephone Laboratories first observed thermal noise in the output of an amplifier. Shortly thereafter, in 1928, H. Nyquist completed his definitive theoretical treatment of thermal noise. Starting with the principle of the equipartition of energy, he showed that the rms voltage of the thermal noise appearing across a resistor R may be expressed for a narrow bandwidth df as

$$\sqrt{\overline{e_n^2}} = \sqrt{4kTR\,df} \quad \text{(volts)} \tag{3-21}$$

where R is the resistance in ohms; T the temperature of the resistor in °K; k the Boltzmann's constant, 1.38×10^{-23} joules/°K; and df the bandwidth in Hz. R is not necessarily the value of the physical resistor imbedded within a

general, complex, two-terminal network, but the value of the real part of the complex impedance appearing at the terminals of the network. Reactive components do not generate thermal noise, but they have an obvious effect on the $\text{Re}(Z_{\text{in}})$. In general, therefore, the R in (3-21) should be written as $R(f)$.

From (3-21) the mean-square noise voltage is simply

$$\overline{e_n^2} = 4kTR\,df \quad \text{(volt}^2\text{)} \tag{3-22}$$

and over a given band of frequencies, f_1 to f_2, the total mean-square voltage is calculated from

$$\overline{e_n^2} = \int_{f_1}^{f_2} 4kTR(f)\,df \tag{3-23}$$

From (3-22) and the usual maximum available power definition, the latter is obtained as

$$P_n = \frac{\overline{e_n^2}}{4R(f)} = kT\,df \quad \text{(watts)} \tag{3-24}$$

which is seen to be independent of R. The same noise power is therefore available regardless of the value of R. At this point it is necessary to understand clearly what is meant by the term "available." From elementary network theory, maximum power is transferred from a generator to the load when the load impedance is the complex conjugate of the generator impedance:

$$\text{Re}(Z_g) = \text{Re}(Z_L)$$

and   (3-25)

$$\text{Im}(Z_g) = \text{Im}(Z_L^*)$$

Thus the available power is equal to the power delivered to R_L *only* if a conjugate match exists between the generator and its load. Under these conditions (Fig. 3.3),

$$(P_{\text{av}})_{\text{max}} = \frac{V_{\text{rms}}^2}{4R_g} \tag{3-26}$$

Using (3-21) a noise model for the physical resistor R can be drawn as shown in Fig. 3.4—a noise-free resistor in series with a zero-impedance voltage generator of rms voltage $\sqrt{4kTR\,df}$. The equivalent current source representation shown in Fig. 3.4c is easily derived from Norton's theorem. It consists of a
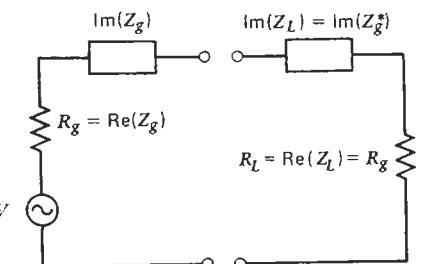
FIGURE 3.3. Circuit conditions for maximum available power.

noise-free shunt conductance G in parallel with an infinite-impedance current generator having an rms current of

$$\sqrt{\overline{i_n^2}} = \sqrt{4kTG\,df} \qquad (3\text{-}27)$$

Since thermal-noise calculations in RF applications usually involve noise *power* rather than rms voltages or currents, it is common practice in the literature to use mean-square voltage $\overline{e_n^2}$ and mean-square current in $\overline{i_n^2}$ generators in circuit models of Fig. 3.4 rather than the more appropriate rms generators. The convenience of not having to drag along the square-root sign is partly responsible for this convention.

From (3-24) the power spectral density of thermal noise is simply

$$P_n' = \frac{P_n}{df} = kT \quad (\text{watts/Hz}) \qquad (3\text{-}28)$$

In other words, the power spectrum of thermal noise is flat or "white," similar to that of shot noise discussed in Section 2.6. This could have been surmised from the nature of the underlying mechanism. The extremely short-lived flight of a given electron results in a very narrow current pulse, similar to an impulse function, which implies a very broad frequency spectrum.

(a)   (b)   (c)
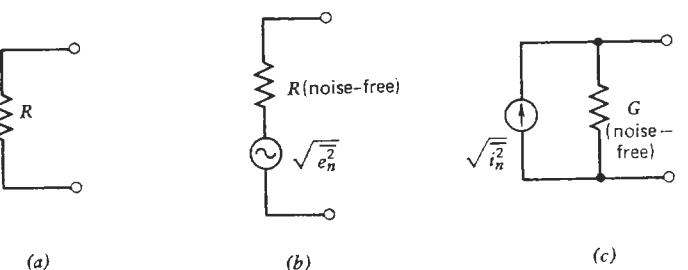
FIGURE 3.4.   Noise model for a physical resistor.

Equations (3-21) and (3-24) are noteworthy for several reasons. First, as a rather obvious fact, thermal noise can be decreased by decreasing the temperature. All molecular motion will cease and the thermal noise will vanish at $0°K$. This is why cooled amplifiers are used in extreme low-level signal reception.

Second, we notice from (3-21) that thermal noise is proportional to $\sqrt{R}$. Does an open circuit, $R \to \infty$, imply an infinite voltage? No, because in practice every physical resistor has some stray capacitance across its terminals. This capacitance, together with the resistance R of the resistor, forms a low pass filter whose cutoff frequency and noise bandwidth (see Section 3.3) decrease as R increases [7, 8]. Thus the shrinking bandwidth limits the magnitude of the noise voltage in accordance with the df in (3-21).

Finally, an inspection of (3-24) reveals that this expression must surely be nonphysical. If one attempted to calculate the total noise power by integrating (3-24) from $f = 0$ to $f = \infty$, the answer would be infinity. Results such as this one were extremely disturbing to theoretical physicists in the pre-quantum-physics days. The paradoxes were not resolved until Max Planck showed in 1900 using quantum mechanics that the term kT in the equipartition law should be replaced by the more complete

$$\frac{hf}{e^{hf/kT} - 1} \quad (\text{watts/Hz}) \qquad (3\text{-}29)$$

where h is Planck's constant, 6.63×10^{-34} joule-second. In other words, the equipartition law is a good approximation, valid only at low frequencies.

Equation (3-24) should therefore be written

$$P_n(f) = kTp(f)\,df \qquad (3\text{-}30)$$

where $p(f)$ is called the Planck factor, given by

$$p(f) = \frac{hf/kT}{e^{hf/kt} - 1} \qquad (3\text{-}31)$$

Clearly, as long as $hf/kT \ll 1$, we may approximate the exponential using only the first two terms of its power series. Thus

$$P_n(f) \approx kT\left(\frac{hf/kT}{1 + hf/kT - 1}\right)df = kT\,df \qquad (3\text{-}32)$$

Using numerical values for h and k, it is easily shown that for $f < 4.3 \times 10^9 T$, the term hf/kT is less than 0.207. Consequently,

$$P_n(f) > 0.9kT\,df$$

and the error is small. At $T_0 = 290°K$, which has been accepted as the reference or standard temperature,* (3-24) will remain valid up to almost 1300 GHz, well above current operating frequencies. However, beware of cryogenic temperatures! At 3°K the assumption of a flat noise density spectrum breaks down already at 12 GHz.

If we repeated our earlier integration, but using (3-30) instead of (3-24), the total thermal-noise power available from a resistor would become

$$(P_n)_T = kT \int_0^\infty \frac{hf/kT}{e^{hf/kT} - 1} \, df$$

This expression can be reduced to a tabulated integral through the substitution $hf/kT = -\ln p$. Thus

$$(P_n)_T = -\frac{(kT)^2}{h} \int_0^1 \frac{\ln p}{1-p} \, dp = \frac{(kT)^2}{h} \frac{\pi^2}{6} \qquad (3-33)$$

or approximately 4×10^{-8} watts for $T = 290°K$.

3.3. TRANSMISSION-LINE APPROACH

The presence of thermal noise was in the previous section accounted for by a lumped-element equivalent circuit. We had a resistor R in series with an equivalent noise generator whose rms voltage was given by (3-21). This model is therefore tied to a specific impedance level R and becomes difficult to use wherever such an impedance cannot be defined as in the case of waveguides.

The difficulty can be circumvented by shifting the emphasis from a lumped-element viewpoint to a transmission-line viewpoint [9]. Instead of voltages and currents we shall consider waves traveling along a transmission line. The key to this is Eq. (3-24) which gives the maximum available noise power from a resistance R. From the definition of maximum available power this would be the noise power delivered by R to a load resistance $R_L = R$. Note again that (3-24) is independent of the value of R. Consider therefore a resistance R_1 connected to a lossless transmission line of characteristic impedance $Z_0 = R_1$ as in Fig. 3.5. We can now talk about a wave of noise power $P_n = kT\,df$ flowing to the right regardless of the value of R_1, provided $Z_0 = R_1$. Thus if we had a waveguide which is terminated in a matched load at its input end, we could immediately say that the noise power radiated by the load and propagating down the waveguide is $P_n = kT\,df$ even though the numerical "resistance" of the waveguide load, the characteristic impedance of the waveguide, and the associated voltage and current amplitudes all remain undefined [10].

We have not yet considered what happens to the noise power as it flows away from R_1. Clearly, if the transmission line is terminated at the far right by a similar matched load $R_2 = Z_0$, all of the noise power is absorbed by it. At

*The reason is simply because at 290°K, which is close to typical room temperature, the term kT is very close to a round number, 4×10^{-21} watt-second.
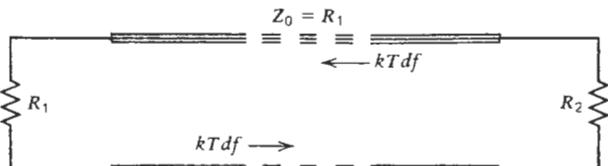
FIGURE 3.5.   Noise on transmission lines.

the same time, the right-hand load resistance R_2 is radiating an equal noise power $P_n = kT\,df$ toward the left where it will be completely absorbed by R_1. We assume, of course, that R_1 and R_2 are at the same temperature. In this manner the thermal equilibrium is preserved.

Now, the load at the right could be a complex one such as a reactive waveguide discontinuity jX in front of a resistive termination as shown in Fig. 3.6. A portion of the noise power arriving from the left would then be reflected back toward R_1 while the remainder would be dissipated in R_2. Obviously, the same happens to the noise originating in R_2. It, too, is split in two parts—one reflected back toward R_2 by the discontinuity and the rest flowing toward R_1 where it is absorbed.

The point in all of this is to illustrate that the concept of noise power waves equal to $P_n = kT\,df$ flowing to the right and to the left permits a straightforward analysis of fairly complicated situations. In fact, since much of this book will deal with RF and microwave applications, the concept of noise power flowing along a transmission line will be found very helpful. We shall make extensive use of it along with the more conventional lumped-element concept of noise voltage or noise current generators. Both methods are important, and it depends on the type of circuit and on the problem at hand which of the two viewpoints is easier to apply.

Before closing this section, it is well to say a few words about the blackbody radiation and its relationship to the thermal noise. The preceding discussion dealt with thermal noise in electrical conductors and transmission lines. Since there is a well-defined interaction between currents in conductors (such as antennas) and electromagnetic waves in space, there should also be a corresponding noise phenomenon associated with electromagnetic radiation. This is
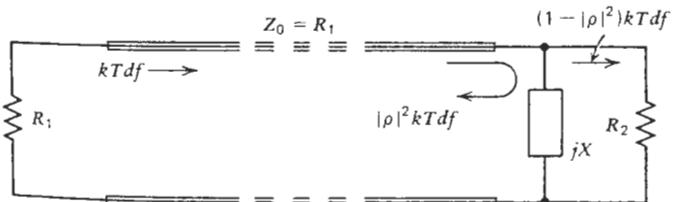

FIGURE 3.6.   Noise on a transmission line with reflective discontinuity.

indeed so, the noise being the blackbody radiation. Although any surface at a temperature $T > 0°K$ is a source of blackbody radiation, a commonly used illustration is a homogeneous space inside a large enclosure whose walls are at a uniform temperature T and are thermally insulated from the outside. It can then be shown [11, 12] that thermal noise and blackbody radiation are merely different manifestations of the same basic mechanism. The resultant noise energy density of unpolarized radiation per unit volume of the enclosed space is

$$\rho(f) = \frac{8\pi h f^3 \, df}{c^3 (e^{hf/kT} - 1)} \approx \frac{8\pi f^2 kT}{c^3} \qquad (3\text{-}34)$$

where c is the velocity of propagation and other terms are as defined earlier. Equation (3-34) is also known as the Rayleigh–Jeans law.

3.4. NOISE BANDWIDTH

The reader has noticed that throughout the preceding discussion the bandwidth was designated by the symbol df, implying that it was infinitesimally narrow. This was done in recognition of the fact that all practical transducers and networks are band limited, and that their transfer characteristics are functions of frequency. Thermal noise, generated within or passing through such circuits, has therefore a power spectral density that is no longer flat with frequency. An ideal resistance with no reactive elements associated with it would yield over a bandwidth f_1 to f_2 a mean-square noise voltage which is given by

$$\overline{e_n^2} = 4kTR(f_2 - f_1) \qquad (3\text{-}35)$$

However, because of intentional or unintentional reactive elements, R in practice is often the real part of some complex impedance Z, and consequently acquires a frequency dependence. Thus, depending on the bandwidth under consideration, (3-35) may or may not hold. In general, an integration, as in (3-23), is required.

Consider next an amplifier having an available power gain (Section 7.5) $G(f)$ as shown in Fig. 3.7. If the input signal to this amplifier were a thermal noise source, the available output noise power over an incremental bandwidth would be

$$P_n = kTG(f) \, df \qquad (3\text{-}36)$$

and the available output power over a full bandwidth would be given by the integral

$$P_n = kT \int_{f_1}^{f_2} G(f) \, df \qquad (3\text{-}37)$$
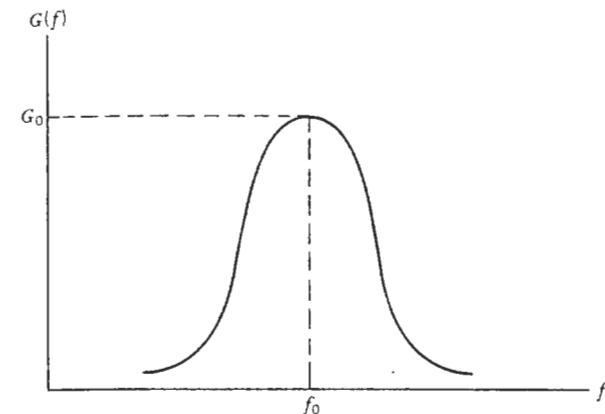
FIGURE 3.7.   Typical gain response of an amplifier.

Clearly, if we had transducers with ideal, rectangular passbands as shown in Fig. 3.8 and implied in (3-35), the problem would be greatly simplified. Unfortunately, such rectangular transfer functions do not exist in the real world, but we can define an *equivalent* rectangular bandwidth B_{eq} such that the total integrated noise power transmitted by the equivalent bandwidth is equal to that transmitted by the actual response of the transducer. This equivalence is illustrated in Fig. 3.9, and can be expressed in mathematical terms as

$$P_n = G_0 kT B_{eq} = kT \int_0^\infty G(f) \, df \qquad (3\text{-}38)$$
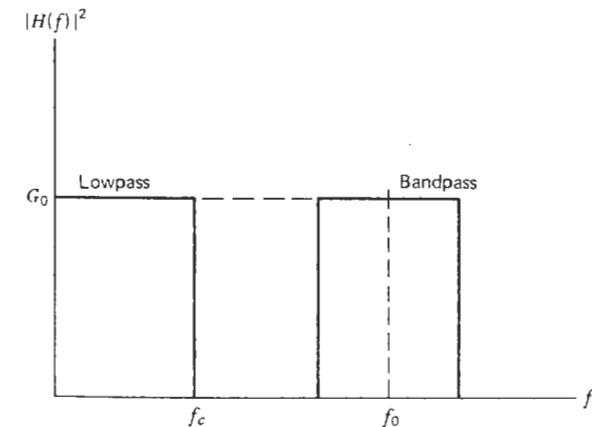
FIGURE 3.8.   Idealized transfer functions.
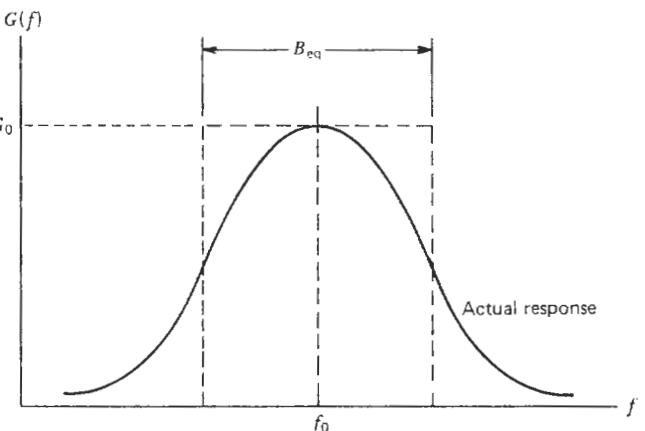
FIGURE 3.9. Definition of a noise bandwidth.

from which

$$B_{eq} = \frac{1}{G_0} \int_0^\infty G(f) \, df \tag{3-39}$$

The parameter B_{eq} is called the equivalent noise bandwidth of the transducer and can be used in lieu of performing the integration (3-37) each time. One must recognize, however, that this convenience is possible only because the input noise is assumed to have a flat spectral noise density. We acknowledged this tacitly in (3-38) by leaving the kT outside the integral sign. Note that the concern here is not the distant roll-off caused by the Planck factor (3-31), but the much more mundane distortion of the input noise characteristic by other band-limited circuits which may precede the transducer under study. As long as the input noise source has a flat bandwidth much greater than that of the transducer it is driving, the commonly used statement that "the system noise bandwidth is set by the component with the narrowest bandwidth" is certainly true. However, if several circuits having roughly equal bandwidths B are connected in cascade, the overall noise bandwidth could become appreciably narrower than B.

The noise bandwidth B_{eq} can be determined by direct integration if an analytical expression for the transfer function $H(f)$ [defined as the generally complex ratio of the output voltage (or current) to the input voltage (or current)] is available. If not, and we only have a measured characteristic, then graphical methods must be applied. The procedure is first to plot the response on a linear graph paper, where it is important to remember that the $|H(f)|^2$ must be expressed either as a numerical ratio, or as the output power in watts. It cannot be in dB or dBm.

The area under the resultant curve is then determined by dividing it into sufficiently small squares, and then counting the squares. The total area, when divided by the midband value of the $|H(f)|^2$, yields the desired noise bandwidth B_{eq}.

The calculation of noise bandwidth is straightforward for symmetrical response shapes such as that shown in Fig. 3.7 where the location of the center frequency coincides with the maximum gain G_0 and both are thus well defined. When the power gain characteristic is asymmetrical, or has several peaks as in Fig. 3.10, the situation becomes less clear. It is not obvious, for example, which frequency should be selected as the center frequency. As a consequence, the value of the "peak" gain remains equally vague. Note, however, that since the gain-bandwidth product of the transfer function remains invariant, any of the selections shown in Fig. 3.10 could be used. The noise bandwidth calculated in each case would be different, but the final value of P_n as calculated from

$$P_n = kTG_0 B_{eq} \tag{3-40}$$

would not be affected because G_0 would vary accordingly.

The equivalent noise bandwidth is a useful artifice in practice. Once the B_{eq} has been determined for a given band-limited transducer, it remains unchanged even if the gain of the transducer varies (Fig. 3.11). This follows from (3-39) since multiplying $G(f)$ by a constant factor multiplies G_0 also by the
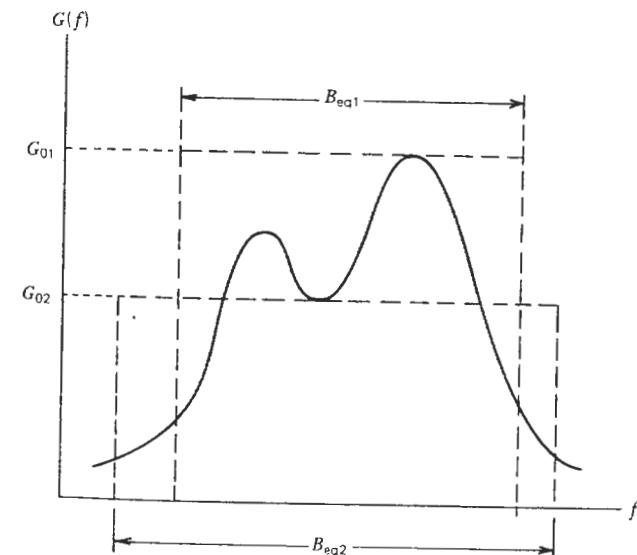

**FIGURE 3.10.** Noise bandwidth for irregular gain response.
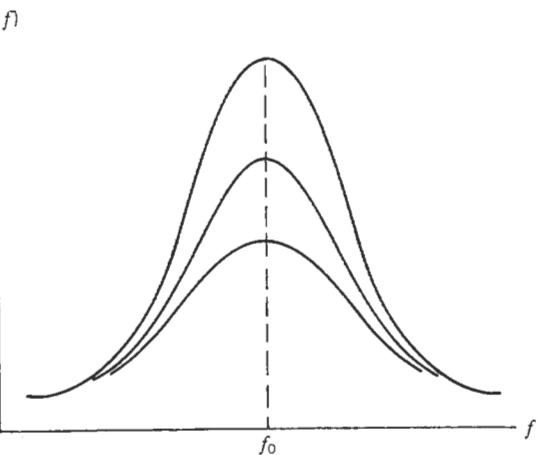
FIGURE 3.11.   Amplifier response for varying gain.

same factor. Note that through the use of B_{eq}, the two parameters, gain and bandwidth, have effectively been separated which was not the case in (3-37).

Equation (3-40) suggests an experimental method of determining the equivalent noise bandwidth of an amplifier or other transducer. All that is required is to apply a flat noise spectrum to the input of the unit, and to measure the total noise P_n at the output terminals. The noise bandwidth B_{eq} can then be calculated directly from (3-40). The one important requirement is, however, that the instrument used to measure P_n must have a flat response well beyond the expected B_{eq}. This becomes clear from Fig. 3.9, because the calculated value of B_{eq} will be correct only if all the noise under the actual response is included in P_n.

As is obvious from Fig. 3.8, an ideal rectangular transfer function has only one definition of bandwidth. Practical transducers on the other hand have responses shaped as in Fig. 3.9, and these are commonly characterized by the half-power, or 3-dB, bandwidth. The question therefore arises: How is the equivalent bandwidth B_{eq} related to the 3-dB bandwidth?

As every practicing engineer knows, the two bandwidths are usually taken to be equal, because the 3-dB bandwidth of a response is easily measured and hence well known, while the noise bandwidth is more difficult to determine. How justified is this assumption?

Consider a single-pole RC network as in Fig. 3.12. The voltage transfer function V_{out}/V_{in} is given by

$$H(f) = \frac{f_c}{f_c + jf} \qquad (3-41)$$
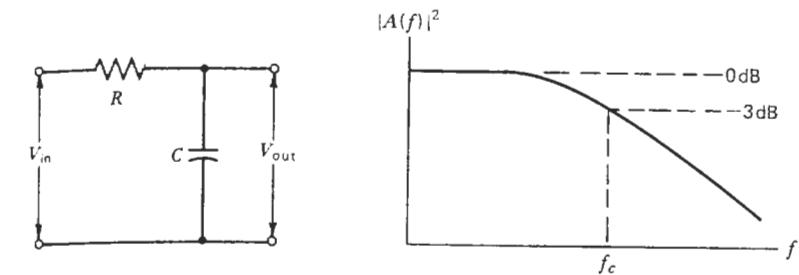
FIGURE 3.12.   Single-pole RC network.

where f is the frequency and

$$f_c = \frac{1}{2\pi RC} \qquad (3-42)$$

is the 3-dB cutoff frequency, and for this low-pass prototype also the 3-dB bandwidth. Let us now apply the definition of B_{eq} by substituting (3-42) into (3-39). The result is

$$B_{eq} = \int_0^\infty \left(\frac{f_c}{\sqrt{f_c^2 + f^2}} \right)^2 df = f_c \int_0^{\pi/2} d\theta = \frac{\pi}{2} f_c \qquad (3-43)$$

where the trigonometric substitution $f = f_c \tan\theta$ was used. The noise bandwidth for a single-pole RC network is therefore 1.57 times the 3-dB bandwidth. Table 3.1 gives the ratios of $B_{eq}/B_{3\,dB}$ for circuits with an increasing number of *identical* poles. For an mth-order Butterworth low-pass filter the same can be calculated from [8]

$$\frac{B_{eq}}{B_{3\,dB}} = \frac{\pi}{2m \sin(\pi/2m)} \qquad (3-44)$$

TABLE 3.1.   Ratio of Noise Bandwidth to 3-dB Bandwidth for Networks with Identical Poles

Number of Poles	$B_{eq}/B_{3\,db}$
1	1.57
2	1.22
3	1.15
4	1.13
5	1.11

As the order (or the number of poles) of the transducer or network increases, the equivalent noise bandwidth is seen to approach the 3-dB bandwidth. This, then, is the justification for using the 3-dB bandwidth in place of the more correct B_{eq}. It should not be forgotten, however, that this is an approximation only, and that the error is not negligible for low-order networks. It may be necessary to apply a correction in accordance with Table 3.1.

3.5.   DESIGNATION OF BANDWIDTH AND SPECTRAL DENSITY

The equivalent noise bandwidth when known permits a calculation of the total available noise power density in watts/Hz, or the total mean-square noise voltage in volts squared through simple multiplication. The expressions kTB_{eq} and $4kTRB_{eq}$ are complete as is and no further integration is required. For simplicity, the subscript on B_{eq} will henceforth be dropped and the symbol B will be used throughout the book to denote the noise bandwidth of the circuit or transducer under study. The symbol df will, however, be retained where it is advantageous to emphasize the frequency dependence of a given noise parameter. The incremental bandwidth df is taken narrow enough so that the parameter is constant over the df. The total value over a *finite* bandwidth is then calculated by the usual integration in the manner of (3-23) or (3-37). We could, of course, circumvent the whole issue and work directly with the noise density concept as in (3-28). The disadvantage of this method from the tutorial viewpoint is that the role of bandwidth in various expressions is obscured and may not be properly appreciated.

As we shall see in Chapters 8, 9, and 10, the important parameters of noise temperature and noise factor can be expressed either as spot or average values. The spot values, as the name indicates, are functions of frequency and are based on either noise density or an incremental "spot" bandwidth df. The average values, on the other hand, are based on a finite bandwidth, typically the noise bandwidth B. The average values are what we typically measure and deal with in practice, but they lack the flexibility of spot parameters (see Sections 8.3 and 9.8). For these reasons we shall use df as a token indication of bandwidth while retaining thereby the advantages of the spot representation. All three designations, noise density, df, and B will be used in this book depending on the point to be emphasized. Since it will be clear from the discussion why one or the other alternative was chosen, there should be no confusion.

3.6.   SUMMARY

(a)   Available noise power from resistor R over a bandwidth df:

$$P_n = kTp(f)\,df = \frac{hf\,df}{e^{hf/kT} - 1} \approx kT\,df.$$

(b)   Mean-square voltage: $\overline{e_n^2} = 4kTR\,df$.

(c)   Mean-square current: $\overline{i_n^2} = 4kTG\,df$.

(d)   Noise bandwidth is the equivalent rectangular bandwidth such that

$$B = \frac{1}{G_0} \int_0^\infty G(f)\,df.$$

(e)   Noise bandwidth is not equal to the 3-dB bandwidth, but the approximation improves as the number of poles (or the order) of the network increases.

REFERENCES

1. H. C. Montgomery, "Transistor Noise in Circuit Applications," *Proc. IRE*, vol. 40, pp. 1461–1471, November 1952.

2. F. R. Connor, *Noise*, Edward Arnold, London, 1982, p. 21.

3. H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE*, vol. 44, pp. 811–818, June 1956.

4. H. Taub and D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, New York, 1971, pp. 34–35.

5. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

6. A. Ambrózy, *Electronic Noise*, McGraw-Hill, New York, 1982, p. 79.

7. C. D. Motchenbacher and F. C. Fitchen, *Low-Noise Electronic Design*, Wiley, New York, 1973, pp. 95–97.

8. A. B. Carlson, *Communication Systems*, McGraw-Hill, New York, 1968, p. 126.

9. A. E. Siegman, "Thermal Noise in Microwave Systems," *Microwave J.*, pp. 81–90, March 1961; pp. 66–73, April 1961.

10. C. G. Montgomery, R. H. Dicke, and E. M. Purcell, *Principles of Microwave Circuits*, vol. 8, Rad. Lab. Series, McGraw-Hill, New York, 1948, p. 37.

11. W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960, p. 114.

12. B. M. Oliver, "Thermal and Quantum Noise," *Proc. IEEE*, vol. 53, pp. 436–454, May 1965.

4 | RANDOM VARIABLES AND PROCESSES

wavelengths. A uniform energy density on wavelength scale implies a nonuniform distribution on frequency scale. Nevertheless, the term "white noise," referring to a uniform energy distribution in terms of *frequency*, has become widely accepted, and we shall use it in this sense also.

The other term used in conjunction with thermal noise is "gaussian." This designation attests to the statistical nature of thermal noise and refers to a particular amplitude distribution within the noise wave. More precisely, it relates to the probability that the noise voltage (or current) exceeds a given amplitude. It is regrettable that in careless usage the terms white and gaussian are often being uttered in the same breath, without regard to them being entirely different concepts. If a waveform with white energy density is passed through a linear filter, the output is clearly no longer white, but the amplitude distribution remains gaussian as before. Conversely, if gaussian noise is applied to a nonlinear circuit such as a limiter, the output becomes nongaussian, but it can be made to retain its white spectral characteristic. Hence one does in no way imply the other.

In order to place our understanding of thermal noise on a firmer footing, it is necessary to review some concepts and techniques of the theory of probability. In particular, there should be at least a cursory understanding of such concepts as the random variable, the variance, probability distributions, and so on. We shall therefore review some of the analytical tools required to handle thermal noise and random waveforms. The latter can be defined as those signals whose properties (amplitude, phase, frequency) vary in time in an entirely unpredictable manner. We note, of course, that such a definition is too broad. Any waveform designed to carry intelligence must necessarily be unpredictable to a degree. After all, if a received waveform is predictable in advance, no new information is conveyed and the transmission of such a message would be pointless. The important difference between this and what we call noise is that the former signal is deterministic; that is, the *sender* could write an explicit function in time which would completely describe the message waveform at any given time t. The unpredictability of the signal at the *receiving* end is thus a prerequisite for useful information transmission. On the other hand, any *undesirable* randomness in the received waveform is detrimental since it distorts or masks the intended message in an unpredictable manner. It is this kind of randomness that we call noise.

Some basic relationships which apply specifically to thermal noise were presented in Chapter 3. We had the familiar noise model of a resistor consisting of a voltage generator in series with a noise-free resistor such that the mean-square voltage was given by

$$\overline{e_n^2} = 4kTR\,df \tag{4-2}$$

From this the available noise power was obtained as

$$P_n = kT\,df \tag{4-3}$$

It was shown in Chapter 3 that the thermal noise arising in a resistor, or for that matter in any conductor, can be described in terms of available noise power which is proportional to the absolute temperature of the conductor and to the noise bandwidth (Section 3.4) of the system. It was also established in (3-32) that for temperatures normally encountered in practice the power spectral density of thermal noise remains constant over a very wide band of frequencies. Only at extreme cryogenic temperatures does the deviation from this relationship become noticeable at microwave frequencies. The power spectral density is for all practical purposes flat, or saying it differently, the mean noise power per unit bandwidth, $P_n' = kT$, is constant. Thermal noise therefore contains a very large number of frequency components just as the white light consists of many different colors (frequencies). This semblance is the basis for calling thermal noise "white."

Strictly speaking, the analogy is not correct. As pointed out by Bennett [1], in optics the term "white" implies equal energy per unit *wavelength*, not frequency. Since $f = c/\lambda$, and by differentiating

$$df = -\frac{c}{\lambda^2}\,d\lambda \tag{4-1}$$

it is seen that for equal wavelength intervals the corresponding frequency intervals shrink in size as one moves toward lower frequencies and longer

Armed with these expressions, one could learn a great deal about thermal noise. We could calculate signal-to-noise ratios at various points within our system, determine how the gains and losses affect the observed noise level, and so on. In fact, the concepts of noise temperature and noise factor, which are the cornerstones of noise analysis, could be developed starting solely with (4-2) and (4-3).

Having such convenient mathematical models already available, why should we bother with the fine-grain statistical properties of thermal noise? The answer is that while the average power and the mean-square voltage or current description of thermal noise is very valuable in its conceptual simplicity, it represents only a time average of the noise wave, and leaves us short of a complete understanding of such random signals. It is therefore advisable to devote some time to a review of the theory of probability and random processes. We obviously cannot undertake an in-depth treatment of the subject, but we can and should consider those aspects of the probability theory which apply directly to our study of thermal noise. Even then it is possible to cover only the essential framework, but this will suffice to answer most practical questions.

4.1. RANDOM SIGNALS AND PROBABILITY FUNCTIONS

In contrast to deterministic signals, for which the behavior is known for all time, the thermal noise voltage is characterized by randomness. It is clearly not possible to express such a voltage as an explicit function of time and therefore we cannot predict with certainty its exact amplitude, phase, or frequency at any instant. This property of a random waveform is what noise is all about. If we had advance knowledge of the interfering signal, we could devise appropriate countermeasures. Unfortunately, the randomness of the thermal noise precludes such a direct attack.

All is not lost, however. While the instantaneous amplitude of a random waveform cannot be predicted with certainty, the randomness itself may obey well-defined statistical rules. In other words, the outcome of the random process may well be predictable with a certain probability of being correct. Consider, for example, the experiment of tossing a coin. If the coin is not weighted, there is no way to predict what the next outcome would be. However, let the experiment be repeated a number of times, N, and let N_a designate the total number of tails and N_b the total number of heads as observed. Define further the ratios N_a/N and N_b/N as the relative frequencies of occurrence of the outcomes "tail" and "head," respectively. We then expect both ratios to approach $\frac{1}{2}$ as N becomes very large. The extension of this argument to cases having more than two outcomes, such as the rolling of a six-sided die, is straightforward. Thus the probability of observing the outcome "5" is $\frac{1}{6}$.

The point is that if a unique limiting value exists for a given frequency of occurrence N_x/N, then a probability of that event, $P(x)$, can be defined as

follows:

$$P(x) = \lim_{N \to \infty} \frac{N_x}{N} \qquad (4\text{-}4)$$

Before proceeding with the rest of this discussion, we should understand the concept of random variable. Suppose that we performed a series of experiments resulting in outcomes $a_1, a_2, \ldots, a_n$ which could be numbers, names, or whatever. We shall now assign a symbol X to designate a function or a rule to transform the outcomes of the experiments into a set of real numbers $x_1, x_2, \ldots, x_n$ which are points on the real axis $-\infty < x < +\infty$. This rule is called the random variable or variate. The general designation for the random variable is thus X, with the x_j representing the specific numerical values.

In our case we shall be investigating voltage or current amplitudes of a random noise waveform. These are already numerical data and there is really no reason why we should not use them as is. The rule X therefore simplifies to a multiplication by unity, or to express it even more simply, the measured voltage amplitude v_j itself could be considered the random variable. In that case it might even be more logical to use the symbol V as the general symbol while the specific numerical outcomes could be designated as v_j. The complete shorthand notation for the probability that V lies between two values v_m and v_n (inclusive) would then be

$$P(v_m \leqslant V \leqslant v_n) \qquad (4\text{-}5)$$

The random variable could be either discrete or continuous. As examples of the discrete type we could cite the outcomes of tossing a coin, rolling a die, or playing a numbers game. On the other hand, when we consider the measured diameters of supposedly identical ball bearings, we would find a continuous spread, a continuum of outcomes centered around the nominal diameter D. Another classical example of a random variable is, of course, the instantaneous amplitude of the thermal noise voltage—the object of study in this book.

Having thus gained a basic understanding of the nature of the random variable, we shall now consider two important probability functions—the cumulative distribution function and the probability density function. A firm grasp of the meanings and of the principal properties of these functions is essential in order to understand what the gaussian probability density function is all about.

4.2. CUMULATIVE DISTRIBUTION FUNCTION (CDF)

We shall begin by first introducing the cumulative distribution function, or CDF in short. The CDF is defined as

$$F(x) \equiv P(X \leqslant x) \qquad (4\text{-}6)$$

In other words, $F(x)$ is the probability that the observed value of the random variable X will be less than or equal to x. Clearly, the CDF has the following property:

$$0 \leqslant F(x) \leqslant 1 \tag{4-7}$$

since as a probability it must be bounded by 0 and 1. Furthermore

$$F(-\infty) = 0 \quad \text{and} \quad F(+\infty) = 1 \tag{4-8}$$

since $F(-\infty)$ implies total exclusion of all possible outcomes, while $F(+\infty)$ implies that all outcomes have been included.

Let us for a moment return to (4-4) and consider the case of *several* possible outcomes of an experiment where the individual outcomes are mutually exclusive. By this we mean that the occurrence of one outcome precludes the occurrence of other outcomes. Let $x_1 \cdots x_M$ be M mutually exclusive outcomes (such as $1 \cdots 6$ in rolling a die), and let the corresponding numbers of occurrences in a large number of experiments be $n_1 \cdots n_M$. As we already saw, (4-4) would readily give us the probability of any single outcome, say x_1, as

$$P(x_1) = \lim_{N \to \infty} \left(\frac{n_1}{N} \right) \tag{4-9}$$

What, then, would be the probability of occurrence of *either* x_1 *or* x_2? The answer is given by

$$P(x_1 + x_2) = \lim_{N \to \infty} \left(\frac{n_1 + n_2}{N} \right) = \lim_{N \to \infty} \left(\frac{n_1}{N} \right) + \lim_{N \to \infty} \left(\frac{n_2}{N} \right)$$

$$= P(x_1) + P(x_2) \tag{4-10}$$

Now let us consider two arbitrarily selected outcomes x_1 and x_2 such that $x_1 < x_2$. The probability that the outcome, or stating it differently, the random variable X, has the value equal to or less than x_2 can be written as a sum of two terms:

$$P(X \leqslant x_2) = P(X \leqslant x_1) + P(x_1 < X \leqslant x_2) \tag{4-11}$$

These terms follow from (4-10) since the intervals $X \leqslant x_1$ and $x_1 < X \leqslant x_2$ are mutually exclusive. Rearranging (4-11) yields

$$P(x_1 < X \leqslant x_2) = P(X \leqslant x_2) - P(X \leqslant x_1) \tag{4-12}$$

and therefore from (4-6),

$$P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1) \tag{4-13}$$

In other words, the probability that the value of the random variable falls between the limits x_2 and x_1 is equal to the difference in the value of the CDF at x_2 and x_1. Thus, knowing the cumulative distribution function of a random process allows us to calculate any of the probabilities associated with X. Furthermore we know from (4-7) that probability is always a positive number. This, combined with (4-13), shows that the CDF is a nondecreasing function of x.

Discrete Random Variable

The concept of cumulative distribution function can be applied to discrete or continuous random variables. A very simple example is shown in Fig. 4.1 which describes the results of rolling a die. The random variable X is discrete —the number of points, 1 to 6, shown on the die after each roll. The details of Fig. 4.1 are self-explanatory. For obvious reasons, $F(0) = P(X \leqslant 0) = 0$. At $x = 1$ the function jumps abruptly to $\frac{1}{6}$ since the probability $P(X \leqslant 1)$ of getting a 1 with the first roll is $\frac{1}{6}$. The $F(x)$ then remains constant until $x = 2$ where it increases to

$$P(1 \text{ or } 2) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}, \text{ etc.}$$

Another useful way of looking at Fig. 4.1, and at the CDF in general, is to consider an experiment of rolling a die a very large number of times, say $N = 6000$. If we counted the number of outcomes for which the result is $x = 1$, $\leqslant 2$, $\leqslant 3$, and so on, the results would look something like those in Table 4.1. After normalizing these results with respect to $N = 6000$, and plotting them against the x_j, Fig. 4.1 would again result.
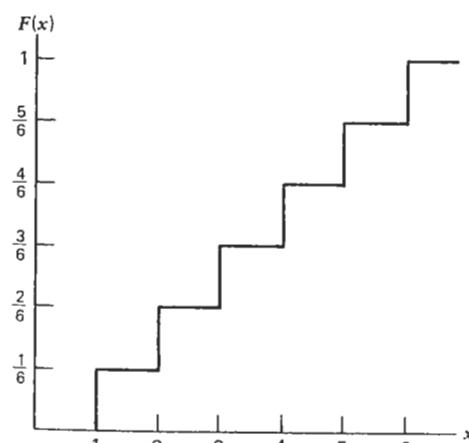

FIGURE 4.1.   Cumulative distribution function for a discrete random variable.

TABLE 4.1.   Result of Rolling a Die 6000 Times

Outcome x_j:	1	$\leqslant 2$	$\leqslant 3$	$\leqslant 4$	$\leqslant 5$	$\leqslant 6$
Number of rolls for which $x \leqslant x_j$:	≈ 1000	≈ 2000	≈ 3000	≈ 4000	≈ 5000	6000

The CDF of a discrete random variable is thus discontinuous at each x_j, with the height of each step equal to the probability of occurrence of x_j.

Continuous Random Variable

We mentioned earlier the measured diameters of N ball bearings as an example of a continuous random variable. We also noted that if N is very large, a continuum of diameters, clustering around the nominal diameter D, would be observed. Using the technique described in connection with Table 4.1, a CDF shown in Fig. 4.2 would be generated. As before, the ordinate of the curve at a given diameter represents the probability of an outcome being less than or equal to d_j. However, the curve is now smooth because any value of d is possible. Using (4-13) we could calculate the probability of finding a diameter within a given interval. Note, however, that while we could calculate the probability of *one discrete outcome*, for example, the number 4 on a die, by
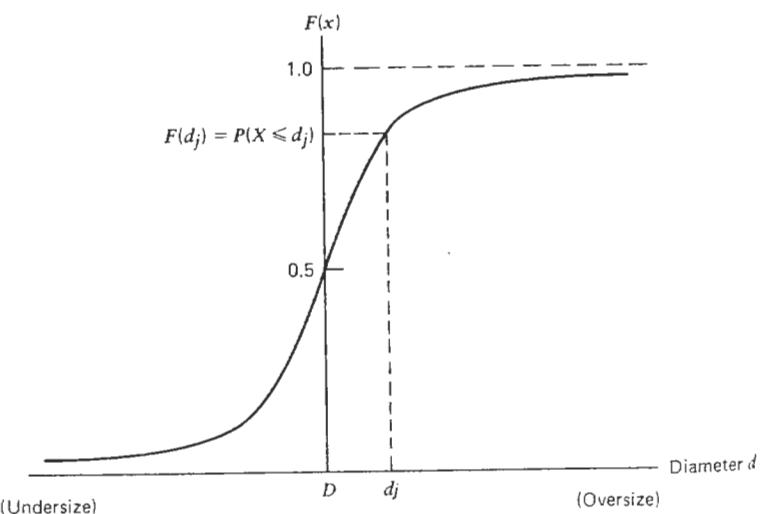

FIGURE 4.2.   Cumulative distribution function for a continuous random variable.

evaluating the difference

$$P(4) = P(5) - P(4) = P(X \leqslant 5) - P(X \leqslant 4)$$

$$= \tfrac{5}{6} - \tfrac{4}{6} = \tfrac{1}{6}$$

the same process, when applied to the continuous case in Fig. 4.2, would yield a probability of zero for any specific outcome:

$$P(X = x_j) = \lim_{\varepsilon \to 0} P\big[(x_j - \varepsilon) < X \leqslant x_j\big]$$

$$= P(X \leqslant x_j) - \lim_{\varepsilon \to 0} P\big[X \leqslant (x_j - \varepsilon)\big] = 0$$

This result is not as strange as it seems. After all, the probability of finding a ball bearing of an *exact*, specific diameter d_j is indeed vanishingly small.

4.3.   PROBABILITY DENSITY FUNCTION (PDF)

The probability density function or PDF in short, is defined as the derivative of the cumulative distribution function CDF; that is,

$$p(x) = \frac{dF(x)}{dx} \tag{4-14}$$

This seems simple enough except a question arises as to how to apply it to a discontinuous CDF. It suffices to state that the problem is solved through the use of impulse functions, or by defining a function which is zero everywhere except at x_j where it assumes the value $P(x_j)$. We shall not pursue this any farther since our objective in this book is the study of random noise—a decidedly continuous random process. We shall therefore turn our attention from now on to continuous random variables for which the probability density function $p(x)$ is easily defined in accordance with Eq. (4-14).

Starting with the CDF in Fig. 4.2 and using (4-14), it is easily seen that the resultant PDF would appear as in Fig. 4.3. Let us note some of the properties of this function. Clearly, the PDF is nonnegative everywhere because the slope of the CDF is positive for all x. From (4-14) it follows that

$$F(x) = \int_{-\infty}^{x} p(s)\, ds = P(X \leqslant x) \tag{4-15}$$

where a dummy variable is used since the integral is a function of x. The result (4-15) is an important one. It says that the *area* under the PDF from $-\infty$ to x is the probability that the observed value or outcome is less than or equal to x.

From this it follows that the probability of observing X in a given range x_1 to x_2 is

$$P(x_1 < X \leqslant x_2) = \int_{x_1}^{x_2} p(x)\, dx \qquad (4\text{-}16)$$

Finally, from (4-8),

$$\int_{-\infty}^{\infty} p(x)\, dx = F(+\infty) - F(-\infty) = 1 - 0 = 1 \qquad (4\text{-}17)$$

which is intuitively seen to be true since the value of the random variable X must fall somewhere between $-\infty$ and $+\infty$.

4.4. MEAN, VARIANCE, AND STANDARD DEVIATION

Figure 4.3 gave a qualitative description of the probability density function. We shall now seek parameters that could be used to describe the given PDF (and the random process it represents) in more quantitative terms without having to resort to a detailed, explicit equation of the curve itself. We first observe that since the basis for the PDF in Fig. 4.3 is a very large (theoretically infinite) population of experimental outcomes, the latter must have a mean or average value. Indeed, the mean of a continuous random variable, also called the expected value or expectation (designated by the letter E), is given by

$$E[x] = \bar{x} = \int_{-\infty}^{\infty} xp(x)\, dx \qquad (4\text{-}18)$$

$p(x)$
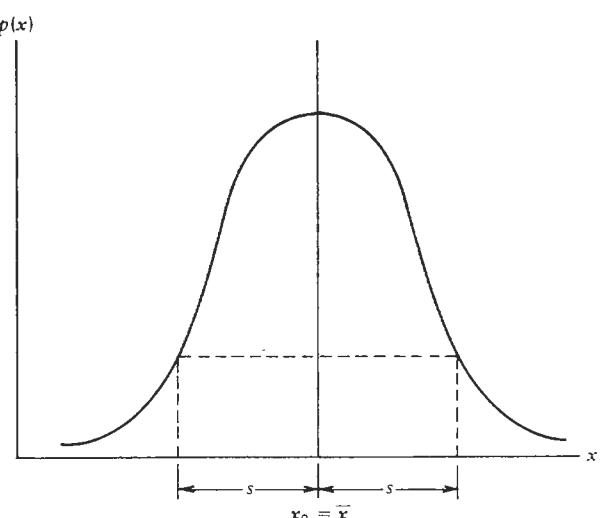
FIGURE 4.3.   Probability density function for a continuous random variable.

where $p(x)$ is the PDF of the given statistical process. If the PDF is symmetrical about some $X = x_0$, then $x_0 = \bar{x}$. This follows from the observation that the integral (4-18) is for a symmetrical curve just an infinite sum of pairs of terms of the form

$$(x_0 - s)p(x)\, dx \quad \text{and} \quad (x_0 + s)p(x)\, dx$$

where x_0 is the value of x at the center of the PDF and s is an arbitrary distance from x_0 (see Fig. 4.3). In the summation process all terms

$$sp(x)\, dx \quad \text{and} \quad -sp(x)\, dx$$

mutually cancel, leaving

$$\bar{x} = x_0 \int_{-\infty}^{\infty} p(x)\, dx = x_0 \qquad (4\text{-}19)$$

In Section 3.1 the mean of the sum of random variables was mentioned. Consider two random variables X and Y, and form their sum $Z = X + Y$. Let the corresponding means be $\bar{x}$, $\bar{y}$, and $\bar{z}$, each defined as in (4-18). Then

$$\bar{z} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x + y)p(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} p(x, y)\, dy\, dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} p(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} xp_x(x)\, dx + \int_{-\infty}^{\infty} yp_y(y)\, dy = \bar{x} + \bar{y} \qquad (4\text{-}20)$$

where $p(x, y)$ is the joint PDF. Thus the mean of a sum is equal to the sum of the means, and this relationship holds regardless of whether X and Y are statistically independent or not.

The *product* of two random variables also has a mean associated with it. This time the relationship is not that simple. As mentioned in Section 3.1,

$$\overline{xy} = (\bar{x})(\bar{y}) \qquad (4\text{-}21)$$

is true *only* if X and Y are statistically independent. Otherwise (4-21) does *not* hold.

The mean $\bar{x}$ is a useful parameter, but it alone is not sufficient to describe a particular statistical process, or to state it differently, a particular population of outcomes. This is illustrated in Fig. 4.4, where two random processes have the same mean value $\bar{x}$ but they obviously have different spreads of observed values around this mean. What we need is a parameter to measure the "width" of the PDF, that is, the spread of the outcomes, in a unique and convenient

$p(x)$
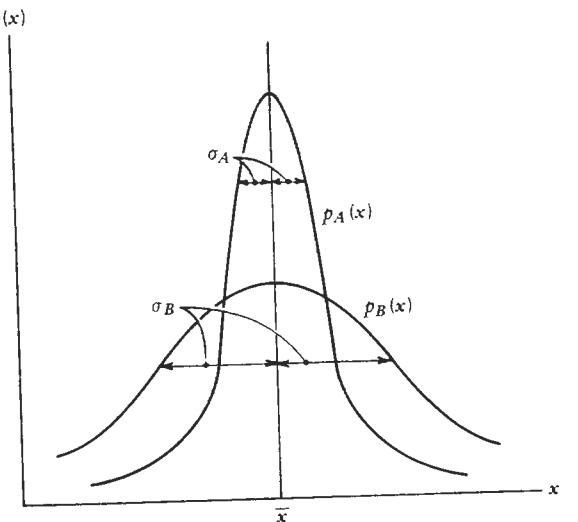
FIGURE 4.4.   Probability density functions with different spreads of outcome.

manner. We could devise a function involving the algebraic deviation from the mean, $(x - \bar{x})$, but this would not do very well because the deviation is obviously positive and negative, and like terms on opposite sides of $\bar{x}$ would cancel each other out. A better choice is therefore the variance of the random variable, defined by

$$\sigma^2 \equiv E\left[(x - x)^2\right] = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)\, dx \qquad (4\text{-}22)$$

where the squaring process has now taken care of the sign problem.

Since

$$(x - \bar{x})^2 = x^2 - 2x\bar{x} + (\bar{x})^2$$

the integral in (4-22) becomes simply

$$\sigma^2 = \overline{x^2} - 2(\bar{x})^2 + (\bar{x})^2 = \overline{x^2} - (\bar{x})^2 \qquad (4\text{-}23)$$

as a consequence of (4-18) and the fact that

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

The term $\overline{x^2}$ is also called the mean square of x. Note carefully that the mean square $\overline{x^2}$ is not the same as the mean value squared, $(\bar{x})^2$. One quantity

stems from the averaging process whereas the other represents a squaring process.

It is of interest to calculate the variance of the sum of two random variables X and Y. To us the most pertinent example would be thermal-noise waves from two physically separate resistors. Since $Z = X + Y$, Eq. (4-23) yields

$$\sigma_z^2 = E\left[(z - \bar{z})^2\right] = \overline{z^2} - (\bar{z})^2$$

$$= \overline{(x + y)^2} - (\bar{x} + \bar{y})^2 = \overline{x^2} + 2\overline{xy} + \overline{y^2} - (\bar{x})^2 - 2\bar{x}\bar{y} - (\bar{y})^2$$

$$= \overline{x^2} - (\bar{x})^2 + \overline{y^2} - (\bar{y})^2 + 2\overline{xy} - 2\bar{x}\bar{y}$$

$$= \sigma_x^2 + \sigma_y^2 + 2\left(\overline{xy} - \bar{x}\bar{y}\right) \qquad (4\text{-}24)$$

Equation (4-24) shows that if X and Y are statistically independent, that is, if

$$\overline{xy} = (\bar{x})(\bar{y}) \qquad (4\text{-}25)$$

only then is the variance of the sum equal to the sum of the variances.

The positive square root of variance is called the standard deviation σ:

$$\sigma = \sqrt{(x - \bar{x})^2} \qquad (4\text{-}26)$$

which is also the rms value of $x - \bar{x}$.

The standard deviation σ is the parameter we have been seeking. It is a convenient measure of the spread of observed values around the mean. If the mean value x is zero, then $\sigma^2 = \overline{x^2}$ and

$$\sigma = \sqrt{\overline{x^2}} \qquad (4\text{-}27)$$

that is, the standard deviation is equal to the rms value of x.

Before attempting to apply these concepts to random waveforms such as thermal noise, we should clearly understand where we are at this point. We started with the ball-bearing experiment and obtained a continuous PDF for the measured diameters. For this PDF, a mean, variance, and standard deviation were defined which described the underlying random process. The important thing to note is that these parameters pertain to an *ensemble* of outcomes, that is, to a large number of measured diameters. The nature of the parameters suggests that they should be equally useful in the study of random waveforms where over a period of time the amplitude, frequency, and phase all vary in a perfectly random manner. If we consider the instantaneous amplitude as the random variable, we can certainly visualize collecting enough data so

that a meaningful average, variance, and so on, could be calculated for the data points, that is, for the instantaneous amplitudes. Despite this apparent simplicity, the road from here to there is not as straight as one might expect. The problem is that we are about to cross the boundary between the ensemble averages and the time averages.

Consider a very large number of generators as in Fig. 4.5, each producing a random waveform $v_j(t)$. If the instantaneous amplitudes of all the waves were recorded at a particular time $t = t_a$, we could plot these outcomes and obtain the associated CDF and PDF. From the latter, the mean, the variance, and other statistical parameters could be determined. It must be remembered, however, that what we obtain are an *ensemble* average and *ensemble* variance, which describe the population of outcomes but at a specific instant of time only, namely, t_a. There is no a priori reason to assume that ensemble parameters determined for another time, $t = t_b$, would equal those at t_a. Indeed, they would in general *not* be equal.

This difficulty leads us to an important concept in statistical analysis. We shall assume from now on that the processes generating our random waveforms are *stationary*, meaning that the statistical characteristics of $v_j(t)$ do not change with time. This is not just an idealistic assumption adopted to render the mathematics pretty, but, as it turns out, in many cases, certainly in the case of thermal noise, the underlying process *is* stationary. This means an important simplification for us. It implies that the ensemble parameters, $E[x]$, σ^2, and so on, remain constant for all values of t, the time of observation. Once a PDF has been determined, it remains invariant and need not be
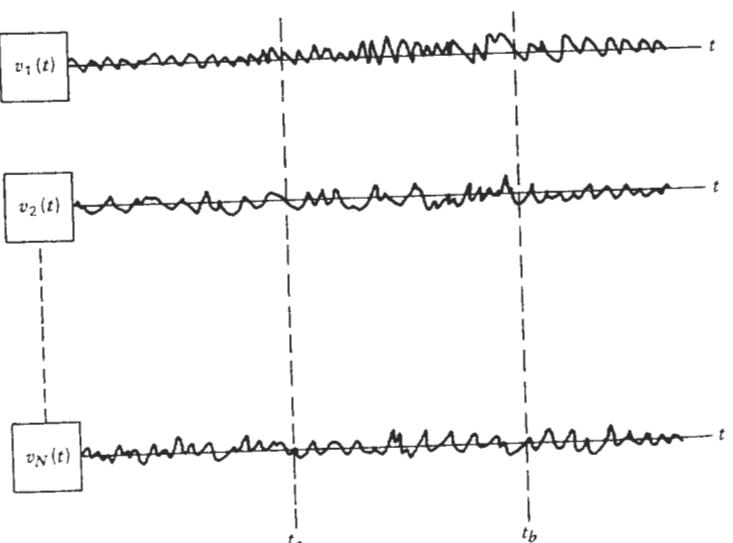
checked or remeasured again. We could move anywhere along the time axis in Fig. 4.5, measure the ensemble PDF and calculate the ensemble averages, and the answer would always be the same.

But we are not there yet. Our ultimate goal, it will be remembered, was to obtain the *time averages* of a given random waveform, because, as we already saw in Section 3.1, the rms and mean-square parameters are directly related to important physical characteristics of the wave, such as the power. It was easy with deterministic waveforms to obtain the rms and the mean-square values through direct integration as in (3-1) and (3-6b). The problem is that random waveforms do not have explicit, analytical expressions, and all information regarding the power or expected amplitudes of such waveforms must be derived using statistical tools, which are the only mathematical descriptions of the wave that we have. Hence the urgency to tie together the ensemble and the time-related properties of a random waveform $v_j(t)$.

The crucial link between the ensemble parameters and the time-average parameters is provided by ergodicity. The ergodic process is one in which *the time and the ensemble averages are identical*. It should be added also that an ergodic process is always stationary while the converse is not necessarily true. We shall not dwell on this subject much longer except to state that from now on only ergodic processes will be considered in this book. The reason is obvious—thermal noise has a random waveform which is ergodic. This is extremely fortuitous. We now have the important relationship that once it is established that a given random waveform meets the conditions of ergodicity, we can immediately equate its ensemble and time averages. In practical terms it means that once the PDF of a given statistical process is known, and consequently the ensemble parameters have been determined, it is possible to identify these parameters as follows [2]:

(a) The squared mean, $(\bar{v})^2$, is the power in the dc component of the $v_j(t)$.

(b) The mean-square value, $\overline{v^2}$, is the total average power of $v_j(t)$.

(c) The variance $\sigma^2 = \overline{v^2} - (\bar{v})^2$ is the power in the time-varying component of $v_j(t)$. It therefore represents the ac power of the waveform.

We should, however, keep in mind that this convenient state of affairs represents a special case only, with ergodicity as its cornerstone. If the latter is missing, then the identities given above do not hold either.

4.5. GAUSSIAN PROBABILITY DENSITY

With this introduction we are ready to consider the gaussian probability density function, also referred to as the normal distribution function. The gaussian PDF is of great importance since many natural phenomena can be modeled by this particular function. In particular, a gaussian distribution is

FIGURE 4.5. An ensemble of random waveforms.
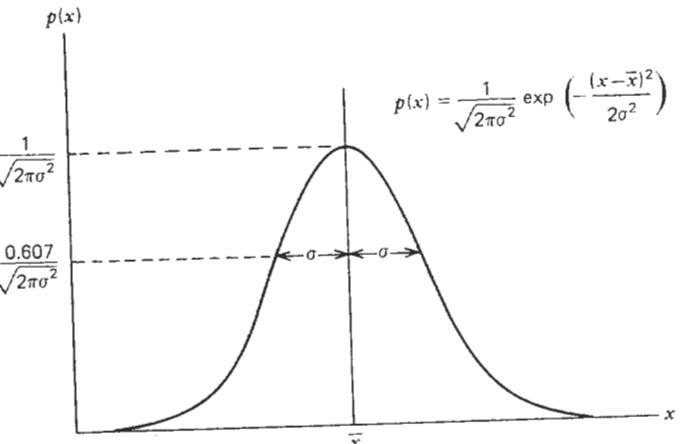
FIGURE 4.6.   Gaussian probability density function.

obtained whenever a large number of independent random causes produces a cumulative effect.* This is the case with thermal noise which is the superposition of the effects of a very large number of electrical charges in spontaneous and random motion within a conductor. Similarly, the shot noise (Section 2.6) follows the gaussian distribution since it, too, is generated by a large number of electrons emitted from the cathode at random times.

The gaussian[†] probability density is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) \tag{4-28}$$

and is plotted in Fig. 4.6. The terms $\bar{x}$ and σ are the mean and the standard deviation defined in Section 4.4. The curve has an even symmetry about $\bar{x}$ which implies

$$P(X \leqslant \bar{x}) = P(X > \bar{x}) = 0.5 \tag{4-29}$$

Since thermal noise has a zero average value,[‡] we shall from now on set $\bar{x} = 0$, and use

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{4-30}$$

*More formally, this follows from the central limit theorem of statistics, which states that if there are N independent random variables, and if we designate M as their sum, then the distribution of M approaches the gaussian distribution as M becomes very large [3].
†The equation was first derived by DeMoivre in 1733, but named in honor of K. F. Gauss (1777–1855) who used it while studying the statistics of errors.
‡The shot noise does have a dc component—the anode current of the tube. This is subtracted out in the analysis.
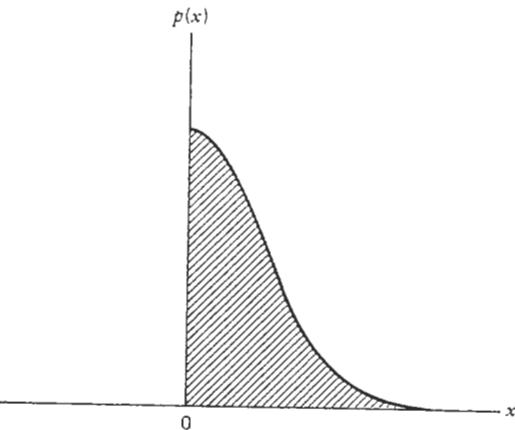
FIGURE 4.7.   Idealized representation of rectified gaussian noise.

As an exercise, let us apply (4-22) to (4-30). The answer, after some algebra, is quite as expected. The variance of the gaussian distribution is σ^2, which is why the symbol is already included in (4-28). If (4-18) is tried on (4-30), we get $\bar{x} = 0$. Again, this is expected and not very informative. Let us, however, pass a gaussian-noise voltage through a half-wave linear detector. By definition this means that the output contains only those noise peaks which are, say, in the positive direction (see Fig. 4.7). We now have for gaussian noise:

$$\text{Average} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty x \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

$$= \frac{\Gamma(1)}{\sigma\sqrt{2\pi}\, 2(1/2\sigma^2)} = \frac{\sigma}{\sqrt{2\pi}} \tag{4-31}$$

Since the rms value of a gaussian waveform is σ, the ratio of the rms to the average value of thermal noise is

$$(\text{Ratio})_{\text{noise}} = \frac{\text{rms}}{\text{average}} = \sqrt{2\pi} = 2.507 \tag{4-32}$$

For a pure sine wave, $V\sin\theta$, the rms value is $V/\sqrt{2}$, and the average value (for half-wave rectification) is

$$\frac{1}{2\pi} \int_0^\pi V\sin\theta\, d\theta = \frac{V}{\pi} \tag{4-33}$$

Hence

$$(\text{Ratio})_{\text{sine wave}} = \frac{\pi}{\sqrt{2}} = 2.221 \tag{4-34}$$

which is different by $20 \log(2.507/2.221) = 1.05$ dB. The significance of this will be seen in Chapter 11.

Returning to (4-30), we note that (4-15) and (4-16) cannot be applied directly because the gaussian function (4-30) is not integrable in closed form. The expression for the gaussian PDF is, however, related to the error function defined as

$$\text{erf}(u) \equiv \frac{2}{\sqrt{\pi}} \int_0^u e^{-\lambda^2} \, d\lambda \qquad (4\text{-}35)$$

and to the complementary error function

$$\text{erfc}(u) \equiv \frac{2}{\sqrt{\pi}} \int_u^\infty e^{-\lambda^2} \, d\lambda = 1 - \text{erf}(u) \qquad (4\text{-}36)$$

Using these functions it can be shown [4] that the probability $P(X \leqslant x)$ is given by

$$P(X \leqslant x) = 1 - \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2\sigma^2}}\right)$$

$$= \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2\sigma^2}}\right)\right] \qquad (4\text{-}37)$$

for $x \geqslant 0$. For $x \leqslant 0$, the proper expression is

$$P(X \leqslant x) = \frac{1}{2} \text{erfc}\left(\frac{|x|}{\sqrt{2\sigma^2}}\right) \qquad (4\text{-}38)$$

A short table of the erfc(x) is included in Appendix A.

A frequent practical problem is to calculate the probability of an outcome falling within a given range centered around the mean (Fig. 4.8). In particular,


$\bar{x} = 0$
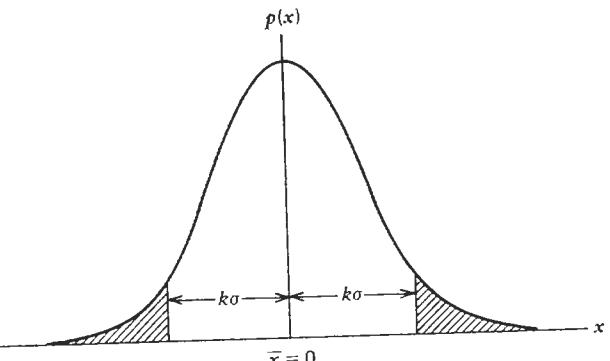
FIGURE 4.8. Definition of $\pm k\sigma$ limits.

TABLE 4.2. Probability of an Outcome Falling Within $\pm k\sigma$

k	$P_{\pm k\sigma}$
0.5	0.383
1.0	0.683
1.5	0.866
2.0	0.955
2.5	0.988
3.0	0.997
3.5	0.9995
4.0	0.99994

a commonly encountered range is $\pm k\sigma$, where k is some positive integer. We thus seek the probability $P(-k\sigma \leqslant X \leqslant +k\sigma)$. This is calculated most easily by utilizing the symmetry of the gaussian PDF (see Fig. 4.6) and excluding the two areas outside the $\pm k\sigma$ limits:

$$P_{\pm k\sigma} = P(-k\sigma \leqslant X \leqslant +k\sigma)$$

$$= 1 - 2P(X > k\sigma) = 1 - 2\int_{k\sigma}^\infty \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}} \, dx \qquad (4\text{-}39)$$

Now make a change in variable, $u = x/\sqrt{2}\,\sigma$ and $du = dx/\sqrt{2}\,\sigma$. Then

$$P_{\pm k\sigma} = 1 - \int_{k/\sqrt{2}}^\infty \frac{2e^{-u^2}}{\sqrt{\pi}} \, du = 1 - \text{erfc}\left(\frac{k}{\sqrt{2}}\right) = \text{erf}\left(\frac{k}{\sqrt{2}}\right) \qquad (4\text{-}40)$$

Table 4.2 shows $P_{\pm k\sigma}$ for several values of k. We can see from the table that the familiar $\pm 3\sigma$ limit does indeed include nearly all of the possible outcomes.

4.6. PEAK FACTOR

A useful application of the preceding material is the evaluation of the peak factor for gaussian noise. For a sinusoidal wave, $V_p \cos \omega t$, the peak factor, defined as the ratio

$$\frac{\text{peak value}}{\text{rms value}} = \frac{V_p}{V_p/\sqrt{2}} = \sqrt{2} \qquad (4\text{-}41)$$

is (in decibels) $20 \log\sqrt{2} = 3$ dB.

For thermal noise the instantaneous amplitude follows the gaussian distribution. A unique peak value cannot be defined since, theoretically at least, any peak value is possible. We must therefore agree on an arbitrary limit for amplitude, beyond which the probability of occurrence is reasonably close to zero. In other words, we select a "peak" exceeded by the noise amplitude a very small percentage of time only. This limit is commonly taken to be 0.01%, that is, we shall include all amplitudes within the 0.9999 probability. According to Table 4.2, this corresponds nearly to $\pm 4\sigma$. Since σ represents the rms value of the noise, then the peak factor for gaussian noise becomes, from the definition given in (4-41),

$$\text{Peak Factor} = \frac{4\sigma}{\sigma} = 4 \text{ or } 12 \text{ dB}. \tag{4-42}$$

Thus if we are to select an amplifier to pass *undistorted* a combination of sinusoidal signals *plus* gaussian noise, the dynamic range of the amplifier must be far wider than would appear from considering the sinusoidal signal alone.

4.7. SUMMARY

(a) Probability of a random variable X being less than or equal to x: $P(X \leqslant x)$.

(b) Cumulative distribution function (CDF) $\equiv F(x) = P(X \leqslant x)$.

(c) Probability density function (PDF) $\equiv p(x) = dF(x)/dx$.

(d) Ensemble average or expectation: $E[x] = \bar{x} = \int_{-\infty}^{\infty} x p(x)\, dx$.

(e) Ensemble variance: $\sigma^2 = E[(x - \bar{x})^2] = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)\, dx$.

(f) The standard deviation is the positive square root of the variance.

(g) For an ergodic, stationary process the ensemble average is equal to the time average.

(h) Thermal noise is ergodic and stationary.

(i) Gaussian PDF or distribution:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right).$$

(j) For a gaussian distribution:

$$P(X \leqslant x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2\sigma^2}}\right)\right] \qquad (x \geqslant 0),$$

$$= \frac{1}{2}\text{erfc}\left(\frac{|x|}{\sqrt{2\sigma^2}}\right) \qquad (x \leqslant 0).$$

(k) For a half-wave rectification:

$$\frac{\text{rms}}{\text{average}} = 2.22 \text{ (6.93 dB)} \quad \text{(sine wave)},$$

$$= 2.51 \text{ (7.98 dB)} \quad \text{(gaussian noise)}.$$

(l) Peak factor:

$$\frac{\text{peak value}}{\text{rms value}} = \sqrt{2} \text{ (3 dB)} \quad \text{(sine wave)},$$

$$= 4 \text{ (12 dB)} \quad \text{(gaussian noise)}.$$

REFERENCES

1. W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960. pp. 8–35.

2. A. B. Carlson, *Communication Systems*, McGraw-Hill, New York, 1968, p. 109.

3. A. B. Carlson, ref. [2], p. 103.

4. H. Taub and D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, New York, 1971, pp. 60–61.

FIGURE 5.1.   Noisy resistors in series.

If all resistors are at the same physical temperature T, then

$$\left(\overline{e_n^2}\right)_T = 4kBT \sum_{j=1}^{N} R_j = 4kBTR_T \tag{5-4}$$

where R_T is the total series resistance. Again, it can be easily shown that the dual version

$$\left(\overline{i_n^2}\right)_T = 4kTBG_T \tag{5-5}$$

also holds. The case of series resistances or parallel conductances is therefore fairly trivial. Even if instead of pure resistances we had a number of complex impedances in series, the same conclusion would hold. We would sum all resistances to get R_T, and all reactances to get jX_T, and the equivalent circuits shown in Fig. 5.2 would result as a consequence of (5-4) and (5-5). If now a conjugate load of $R_T - jX_T$ were connected to the output terminals $a-a'$ of the circuits in Fig. 5.2, an average power of $kT\,df$ would be delivered to the load R_T as before. As mentioned in Section 3.2, the total resistance seen at the input of a complex network is the real part of the complex input impedance, and hence will be a function of frequency. In general, therefore, we may have

$$R_T = R_T(f) \tag{5-6}$$

(a)   (b)
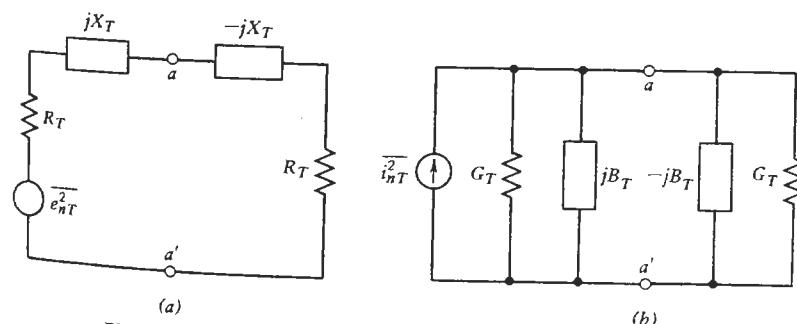
FIGURE 5.2.   Load immittances for maximum available noise power.

5 | SINGLE-PORT NETWORKS

We saw in Chapter 3 that the random motion of electrons within a conductor under thermal agitation results in a voltage at the open ends of that conductor. In particular, a conductor of resistance R at an absolute temperature T could be represented by an equivalent noise voltage generator e_n of zero internal resistance, in series with a noise-free resistor R (Fig. 3.4b). The generator accounts for the noise observed in the bandwidth B. A dual representation using a current generator and a shunt conductance is also available. Thus

$$\overline{e_n^2} = 4kTBR \tag{5-1}$$

and

$$\overline{i_n^2} = 4kTBG \tag{5-2}$$

5.1.   EQUAL TEMPERATURES

Suppose now that we had several resistors, R_1 to R_N, connected in series as in Fig. 5.1. The noise originating in one resistor is completely uncorrelated with those from other resistors. The total mean-square noise voltage can therefore be written as the sum of the squares of individual noise voltages:

$$\left(\overline{e_n^2}\right)_T = \sum_{j=1}^{N} \left(4kT_jBR_j\right) = 4kB \sum_{j=1}^{N} \left(T_jR_j\right) \tag{5-3}$$

FIGURE 5.3. Noise model of a linear, bilateral, passive network.

We shall consider next a general two-terminal or single-port network consisting of an arbitrary number of shunt and series resistors connected together in a perfectly arbitrary manner (Fig. 5.3). Assume that all resistors are at the same temperature T. What noise voltage and noise power would we expect to measure at terminals $a-a'$?

The first step is to replace each resistor with its noise equivalence,* $\overline{e_n^2}$ and R. One would then seek a Thevenin equivalent circuit for the overall network by calculating the open-circuit mean-square voltage and the equivalent resistance at $a-a'$. This is accomplished by determining the mean-square contribution of each individual generator at $a-a'$ and, since all generator outputs are uncorrelated, by adding the individual voltages to get $(\overline{e_n^2})_T$. The equivalent Thevenin resistance follows in the usual manner by replacing each voltage generator with its internal resistance (zero in our case), and then evaluating the total resistance R_{eq} seen at $a-a'$.

A simple example should clarify this. Consider Fig. 5.4a where the three resistors R_1, R_2, and R_3 have been replaced with their thermal noise equivalents. The mean-square open-circuit noise voltage at $a-a'$ contributed by R_1 is

*We shall for simplicity refer to the *voltage* generator model of noise only. Clearly, the current model or a combination of the two could well be more advantageous to use in some instances.

given by

$$\overline{e_1^2} = \frac{\overline{e_{n1}^2}\,R_2^2}{(R_1 + R_2)^2} = \frac{(4kTBR_1)\,R_2^2}{(R_1 + R_2)^2} \qquad (5\text{-}7)$$

Similarly, the contribution from R_2 is

$$\overline{e_2^2} = \frac{\overline{e_{n2}^2}\,R_1^2}{(R_1 + R_2)^2} = \frac{(4kTBR_2)\,R_1^2}{(R_1 + R_2)^2} \qquad (5\text{-}8)$$

The contribution from R_3 is particularly simple:

$$\overline{e_3^2} = \overline{e_{n3}^2} = 4kTBR_3 \qquad (5\text{-}9)$$

The total voltage at $a-a'$ is the sum of the contributions (5-7) to (5-9):

$$\left(\overline{e_n^2}\right)_T = 4kTB\left(\frac{R_1 R_2^2}{(R_1 + R_2)^2} + \frac{R_1^2 R_2}{(R_1 + R_2)^2} + R_3\right)$$

$$= 4kTB\left(\frac{R_1 R_2}{R_1 + R_2} + R_3\right) \qquad (5\text{-}10)$$

The equivalent resistance R_{eq}, now considered noise-free, is simply

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2} + R_3 \qquad (5\text{-}11)$$

The complete equivalent circuit is then given by Fig. 5.4b and Eqs. (5-10) and (5-11).
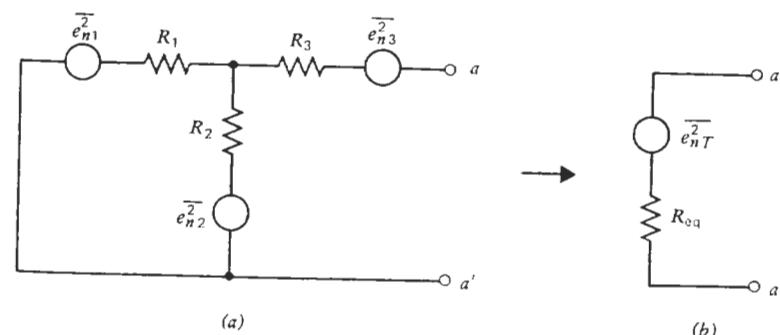
FIGURE 5.4. Sample calculation of a T-network.

A more general way of looking at this is to define a transmittance function which is a (usually complex) ratio of some arbitrary current or voltage to another arbitrary current or voltage associated with the given network. Thus in Fig. 5.5 the ratios $Y(f) = I_L/V_b$, $A(f) = V_L/V_b$, $Z(f) = V_b/I_1$, and so on, are examples of transmittance functions. Now recall what was done in deriving (5-7) to (5-9). There were three mean-square voltage generators $\overline{e_{n1}^2}$, $\overline{e_{n2}^2}$, and $\overline{e_{n3}^2}$ for which the resultant mean-square output voltages at a–a' were sought. Using the transmittance-function concept, we can express this process in general as

$$\overline{e_1^2} = \overline{e_{n1}^2} \left| A_1(f) \right|^2$$

$$\overline{e_2^2} = \overline{e_{n2}^2} \left| A_2(f) \right|^2 \qquad (5\text{-}12)$$

$$\overline{e_3^2} = \overline{e_{n3}^2} \left| A_3(f) \right|^2, \text{ etc.}$$

The summed open-circuit mean-square voltage at a–a' is therefore given by

$$\left(\overline{e_n^2} \right)_T = \left| A_1(f) \right|^2 \overline{e_{n1}^2} + \left| A_2(f) \right|^2 \overline{e_{n2}^2} + \left| A_3(f) \right|^2 \overline{e_{n3}^2} + \cdots \qquad (5\text{-}13)$$

as in Fig. 5.3b. If the temperature is the same throughout the network, then

$$\overline{e_{n1}^2} = 4kTR_1\,df$$

$$\overline{e_{n2}^2} = 4kTR_2\,df, \text{ etc.} \qquad (5\text{-}14)$$

and therefore (5-13) becomes, for a narrow bandwidth df,

$$\left(\overline{e_n^2} \right)_T = 4kT\,df \left[\left| A_1(f) \right|^2 R_1 + \left| A_2(f) \right|^2 R_2 + \cdots \right] \qquad (5\text{-}15)$$
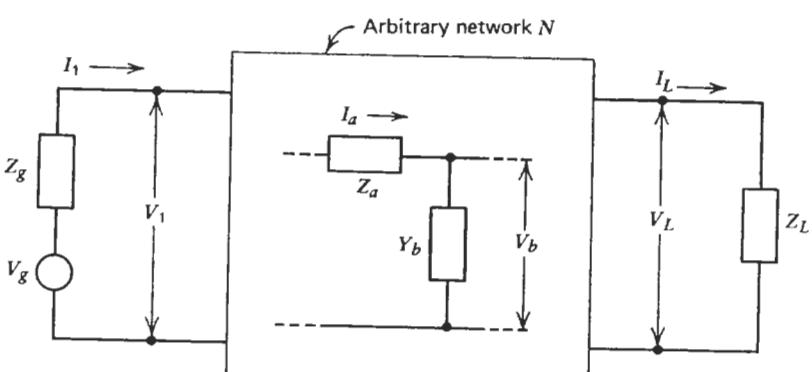

FIGURE 5.5.   Definition of a transmittance function.

from which in turn the mean-square noise voltage over any arbitrary bandwidth f_1 to f_2 can be calculated as

$$\left(\overline{e_n^2} \right)_T = 4kT \int_{f_1}^{f_2} \left[\left| A_1(f) \right|^2 R_1 + \left| A_2(f) \right|^2 R_2 + \cdots \right] df \qquad (5\text{-}16)$$

Let us apply this to Fig. 5.4a. Since in this example the network is purely resistive, all transmittance functions should be purely real and independent of frequency. Indeed, as seen from (5-7) to (5-9), where the squared transmittance functions were calculated, this is so. Thus

$$A_1^2 = \frac{R_2^2}{(R_1 + R_2)^2} \qquad (5\text{-}17a)$$

and

$$A_2^2 = \frac{R_1^2}{(R_1 + R_2)^2} \qquad (5\text{-}17b)$$

$$A_3^2 = 1 \qquad (5\text{-}17c)$$

Substituting these into (5-15), and remembering that for this resistive case the actual bandwidth equals the noise bandwidth, we obtain (5-10) as before.

The utility of the transmittance method may not be apparent from this example. We shall, however, show in the next section how the transmittance functions are related to the very useful Pierce's rule and to the analysis of noise from components at unequal temperatures.

Now continuing with (5-10) and (5-11), and using the basic relationship (3-24), we obtain the available output power* at a–a':

$$N_o = \frac{\left(\overline{e_n^2} \right)_T}{4R_{eq}} = kTB \qquad (5\text{-}18)$$

as expected. We also suspect from (5-10) and (5-11) that the result

$$\left(\overline{e_n^2} \right)_T = 4kTBR_{eq} \qquad (5\text{-}19)$$

holds in general. This is indeed correct. That is, given the equivalent input resistance R_{eq} (or the conductance G_{eq}) of a linear, bilateral, passive two-

---

*In anticipation of the forthcoming discussion of signal and noise, the symbol for noise power will henceforth be N, with the appropriate subscript to designate the particular noise, input, output, excess, and so on.

terminal network, the mean-square noise output is immediately given by (5-19) or by its dual

$$\left(\overline{i_n^2}\right)_T = 4kTBG_{\text{eq}} \tag{5-20}$$

provided, of course, that all resistances within the network are at the same temperature T. A proof of general validity of this result, based on thermal equilibrium considerations, is given by Bennett [1]. Finally, it must not be forgotten that for complex networks $R_{\text{eq}} = R_{\text{eq}}(f)$ as noted in the beginning of this section.

5.2.   UNEQUAL TEMPERATURES

Section 5.1 assumed that all resistances were at the same physical temperature T. Suppose this were not true. Assume, for example, that R_1 is at T_1, R_2 at T_2, and so on. What then?

A direct application of Thevenin's theorem would not suffice, for the equivalent resistance R_{eq} alone would be useless unless we could somehow calculate for R_{eq} an effective temperature T_{eff} to represent the weighted contributions of the individual resistors at unequal temperatures. Of course, the parameter R_{eq} would be calculated as before, but the equivalent Thevenin generator now becomes

$$\overline{e_n^2} = 4kT_{\text{eff}}R_{\text{eq}}\,df \tag{5-21}$$

and we could not proceed until T_{eff} is known.

The starting point for determining T_{eff} is (5-13). Figure 5.6 shows an arbitrary passive network containing a number of resistances R_1 to R_N which are at temperatures T_1 to T_N, respectively. The network will in general also contain reactive elements, and the equivalent resistance $R_{\text{eq}}(f)$ at terminals a–a' is therefore a function of frequency. Since each resistor R_j is a source of thermal noise, it has a mean-square voltage generator $\overline{e_{nj}^2} = 4kT_jR_j\,df$ in series with it. The combined effect of these generators is represented by the overall voltage generator $(\overline{e_n^2})_T = 4kT_{\text{eff}}R_{\text{eq}}(f)\,df$ at a–a'.

If a complex conjugate load $R_{\text{eq}}(f) - jX_{\text{eq}}(f)$ were connected to terminals a–a', the available power would be

$$N_o = \frac{\left(\overline{e_n^2}\right)_T}{4R_{\text{eq}}(f)} = kT_{\text{eff}}\,df \tag{5-22}$$

Using (5-22) and (5-19), and noting that all temperatures are now different, we can write

$$N_o = kT_{\text{eff}}\,df = \frac{\left(\overline{e_n^2}\right)_T}{4R_{\text{eq}}(f)}$$

$$= \frac{k\,df\left[|A_1(f)|^2R_1T_1 + |A_2(f)|^2R_2T_2 + \cdots + |A_N(f)|^2R_NT_N\right]}{R_{\text{eq}}(f)} \tag{5-23}$$

$$= k\,df\left(\frac{|A_1(f)|^2R_1}{R_{\text{eq}}(f)}T_1 + \frac{|A_2(f)|^2R_2}{R_{\text{eq}}(f)}T_2 + \cdots + \frac{|A_N(f)|^2R_N}{R_{\text{eq}}(f)}T_N\right) \tag{5-24}$$

$$= k\,df\left(\alpha_1T_1 + \alpha_2T_2 + \cdots + \alpha_NT_N\right) \tag{5-25}$$

Comparing (5-25) to (5-23) yields

$$T_{\text{eff}} = \alpha_1T_1 + \alpha_2T_2 + \cdots + \alpha_NT_N \tag{5-26}$$

The effective temperature T_{eff} can therefore be calculated once the coefficients α_1 to α_N are known. For these we must calculate the corresponding
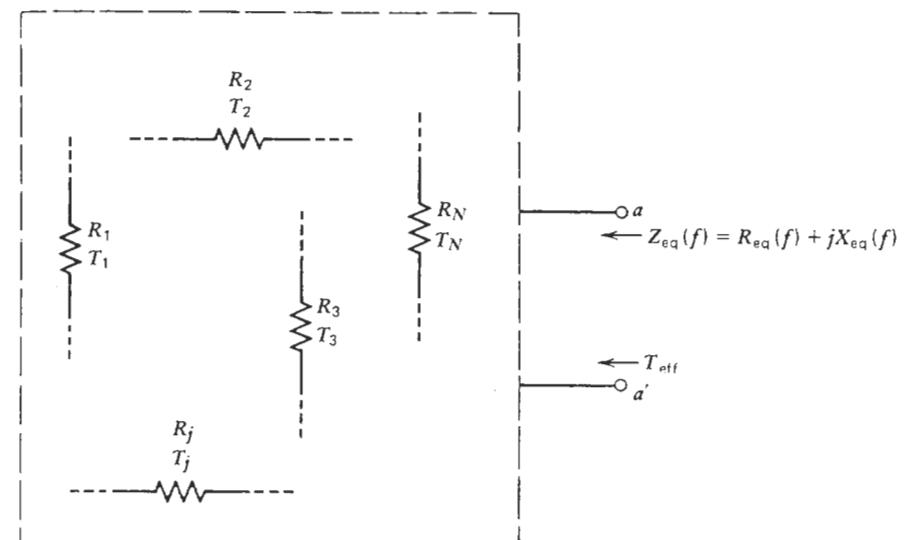

FIGURE 5.6.   Equivalent noise temperature of a network at unequal

transmittance functions which could be a laborious task. Fortunately, at times a much easier method, based on Pierce's rule, can be applied. This rule was first proposed by J. R. Pierce [2] for determining the noise temperature of an antenna. The proof, although fairly simple, will not be repeated here (see, e.g., ref. [3]). In essence, Pierce's rule postulates a reverse situation where power is fed *into* the network at $a-a'$ (Fig. 5.6). Let us therefore connect a generator of an internal impedance $Z_g = Z_{eq}^*$ to terminals $a-a'$. It is clear that the power from the generator will then be completely absorbed by the resistances within the network. Pierce's rule is now rephrased for our purposes as follows: If a unit of power is delivered to a composite, passive, and linear two-terminal network, and if a fraction α_1 is absorbed by the resistance R_1 at temperature T_1, a fraction α_2 by a resistance R_2 at temperature T_2, and so on, then the effective input temperature T_{eff} at $a-a'$ is given by (5-26).

An example of the application of Pierce's rule follows. Fig. 5.7a shows a matched attenuator of loss L at a temperature T_x. A resistor R_p at a temperature T_p is connected to terminals b-b'. What is the T_{eff} at terminals $a-a'$?

Let a unit of power be applied at $a-a'$. The fraction of power dissipated in R_p is then

$$\alpha_p = \frac{1}{L} \tag{5-27}$$

on account of the loss being L. The remainder of the input power must be dissipated in the attenuator. Since the total of all such dissipated powers must add up to unity, we can write

$$1 = \alpha_p + \alpha_a = \frac{1}{L} + \alpha_a$$
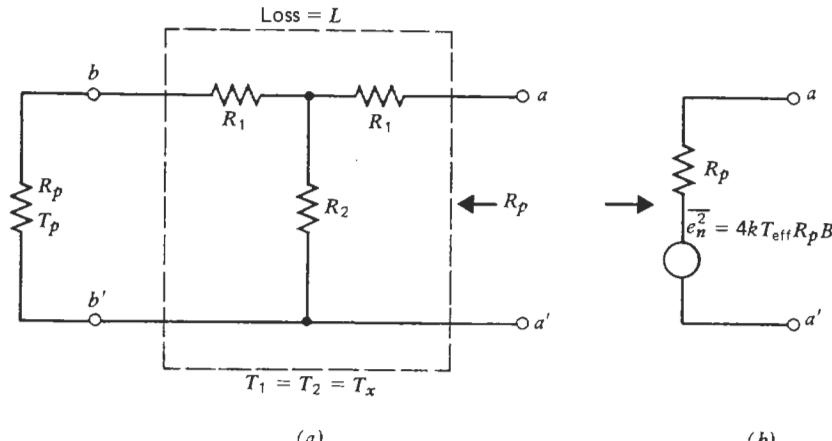
FIGURE 5.7.   Noise temperature of matched attenuator.

where α_a designates the fraction of power dissipated in the attenuator. From this,

$$\alpha_a = 1 - \frac{1}{L} = \frac{L-1}{L} \tag{5-28}$$

Combining (5-26)–(5-28) yields directly

$$T_{eff} = \frac{T_p}{L} + \left(\frac{L-1}{L}\right) T_x \tag{5-29}$$

and the final equivalent circuit is shown in Fig. 5.7b. Equation (5-29) is an important relationship which we shall encounter again in later chapters.

One should always remember that a network contains generally resistive as well as reactive elements. Thus both T_{eff} and R_{eq} are likely to be functions of frequency. This is easily seen from (5-24).

Finally, (5-13) is a rather generally applicable expression *provided* that the individual noise generators $\overline{e_n^2}$ are uncorrelated (see Section 3.1).

5.3. SUMMARY

(a) Thermal noise from a lossy two-terminal (single-port) network at uniform temperature T:
  (1) Voltage representation: $e_{nT} = \sqrt{4kTBR_{eq}}$.
  (2) Current representation: $i_{nT} = \sqrt{4kTBG_{eq}}$.
  (3) Available noise power: $P_n = kTB$.
(b) Same network but with individual resistances R_j at different temperatures T_j:
  (1) Voltage and current representations same as above, but with T replaced by the equivalent noise temperature T_{eff} where $T_{eff} = \sum_{j=1}^{N}(\alpha_j T_j)$ and the α_j are fractions of power dissipated in R_j when power is flowing *into* the network.
  (2) Available noise power: $P_n = kT_{eff}B$.
(c) For a matched attenuator L at T_x, terminated in a load at T_p:

$$T_{eff} = \frac{T_p}{L} + \left(\frac{L-1}{L}\right) T_x$$

REFERENCES

1. W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960, pp. 8–35.
2. J. R. Pierce, "The General Sources of Noise in Vacuum Tubes," *IRE Trans. ED*, vol. ED-1, pp. 134–167, December 1954.
3. W. W. Mumford and E. H. Scheibe, *Noise Performance Factors in Communication Systems*, Horizon House, Dedham, 1968.

6 | TWO-PORT NETWORKS

The preceding chapter dealt with two-terminal, or single-port networks. We considered several useful techniques which permit the derivation of equivalent thermal-noise models for any linear, passive, one-port network at a given pair of terminals. In an analogous manner we can develop noise models for two-port networks. Before proceeding, it is advisable to review briefly some properties and common representations of linear two-ports.

6.1. REVIEW OF LINEAR TWO-PORTS

A linear two-port, containing no internal sources, can be represented by the familiar open-circuit impedance or short-circuit admittance matrices

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \times \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = [Z][i] \tag{6-1}$$

and

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = [Y][v] \tag{6-2}$$

where the z_{mn} and y_{mn} are in general complex functions of frequency. The terminal voltages and currents are defined as in Fig. 6.1. Note that the currents are defined to be positive if they flow into the network. For reciprocal networks, $z_{mn} = z_{nm}$ and $y_{mn} = y_{nm}$. If the network is lossless (reactive components only), then all z_{mn} and y_{mn} become purely imaginary.

If the given two-port is an active one containing internal sources, it is possible to bring their effect out to the input and output terminals by an extension of Thevenin's theorem [1], first derived by L. C. Peterson [2]. In particular, if the network is described in terms of open-circuit z parameters, the internal sources are represented by two fictitious voltage sources of zero internal resistance, placed in series with the input and the output, respectively, as shown in Fig. 6.2a. Alternatively, we could use the dual representation when the short-circuit y parameters are given. In this case current generators of zero internal conductance are connected in shunt across the input and the output as shown in Fig. 6.2b. The impedance and admittance matrices now become

$$v_1 = z_{11}i_1 + z_{12}i_2 + E_1$$
$$v_2 = z_{21}i_1 + z_{22}i_2 + E_2 \tag{6-3}$$

and

$$i_1 = y_{11}v_1 + y_{12}v_2 + I_1$$
$$i_2 = y_{21}v_1 + y_{22}v_2 + I_2 \tag{6-4}$$

The two equivalent generators E_1 and E_2, or I_1 and I_2, are generally not independent. This is easily seen by considering the example below. Take for simplicity a T-attenuator containing voltage sources in series with each resistor (Fig. 6.3). Write the loop equations for an impedance representation to obtain

$$v_1 = i_1(R_1 + R_3) + i_2 R_3 + (e_1 + e_3)$$

and

$$v_2 = i_1 R_3 + i_2(R_2 + R_3) + (e_2 + e_3)$$
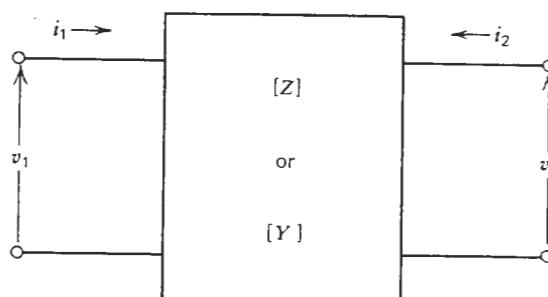
FIGURE 6.1. Definition of impedance and admittance matrices.

(a)

(b)
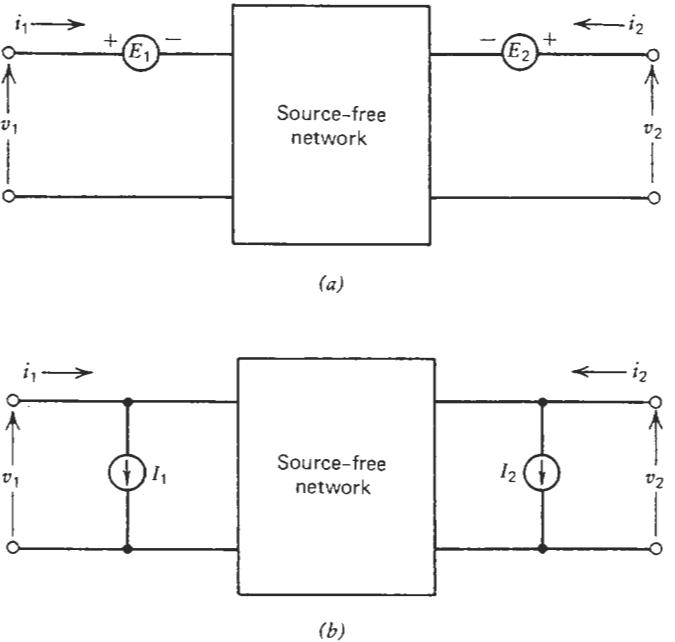
FIGURE 6.2. Representation of networks with internal sources.

From these the equivalent input/output generators are seen to be

$$E_1 = e_1 + e_3 \quad \text{and} \quad E_2 = e_2 + e_3$$

These include one common term, e_3, which provides the dependence between E_1 and E_2. The values of E_1 and E_2 for a given two-port can be determined by open-circuiting both ports of the network and measuring the resultant
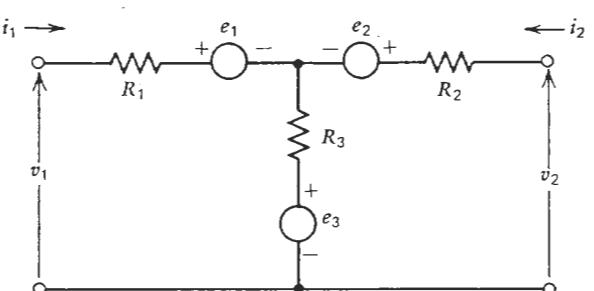

FIGURE 6.3. T-attenuator with internal sources.

open-circuit voltages. This follows directly from (6-3) if $i_1 = i_2 = 0$. Thus

$$E_1 = v_1|_{i_1 = i_2 = 0} \quad \text{and} \quad E_2 = v_2|_{i_1 = i_2 = 0} \tag{6-5}$$

Alternatively, both ports could be short-circuited and the resultant short-circuit currents measured to determine I_1 and I_2:

$$I_1 = i_1|_{v_1 = v_2 = 0} \quad \text{and} \quad I_2 = i_2|_{v_1 = v_2 = 0} \tag{6-6}$$

The equivalent representations shown in Fig. 6.2 are not the only possible ones. We could adopt various combinations of voltage and current generators. A particularly useful representation is based on the ABCD matrix (Appendix B):

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \tag{6-7}$$

which leads to the following description of the internal sources (Fig. 6.4):

$$v_1 = Av_2 + Bi_2 + V \quad \text{and} \quad i_1 = Cv_2 + Di_2 + I \tag{6-8}$$

The internal sources are now represented by a voltage and a current generator, both located at the input of the network. As before, the generators V and I are in general not independent. The advantage of this representation is that the physical network is split in two distinct parts—one containing the effects of all internal sources and the other one being completely passive. This is particularly useful when the internal sources are noise generators. By bringing the internal noise generators to the input of the network, we have separated out the noise-producing elements and rendered the otherwise unchanged network noise-free. This in turn means that in signal-to-noise ratio (SNR) calculations we can simply ignore the entire noise-free portion of the original two-port, because the SNR is not affected when the signal and the noise both pass through a noise-free network. The gain of the network, and even the input match, no longer enter the calculations. It is entirely sufficient to deal only with the two equivalent input noise generators e_n and i_n. We shall describe in Chapter 11 the technique of measuring these quantities for a given
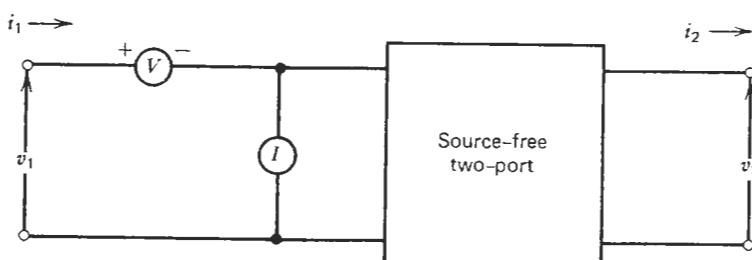

FIGURE 6.4. Description of internal sources based on the $ABCD$ ma

noisy two-port. At this time we merely note that for a general active two-port the V and the I in (6-8) cannot be defined or measured as easily as the E_1, E_2, I_1, and I_2. The best method is to apply the impedance matrix representation to Fig. 6.4:

$$v_1 = (i_1 - I)z_{11} + i_2 z_{12} + V$$

$$v_2 = (i_1 - I)z_{21} + i_2 z_{22} \tag{6-9}$$

and then to compare these expressions to (6-3) from which we can identify

$$E_1 = V - Iz_{11} \quad \text{and} \quad E_2 = -Iz_{21}$$

Hence

$$V = E_1 - E_2\left(\frac{z_{11}}{z_{21}}\right) \quad \text{and} \quad I = -\frac{E_2}{z_{21}} \tag{6-10}$$

If (6-6) had been used, the corresponding equations would be

$$V = \frac{-I_2}{y_{21}} \quad \text{and} \quad I = I_1 - I_2\left(\frac{y_{11}}{y_{21}}\right) \tag{6-11}$$

6.2. APPLICATION TO NOISY TWO-PORTS

The basic procedure of representing a noisy two-port by an equivalent circuit of the type shown in Fig. 6.5 will now be outlined. The development follows essentially that of Rothe and Dahlke [3], Montgomery [1], and Haus et al. [4, 5]. The two equivalent generators V and I in Fig. 6.4 become two noise generators e_n and i_n which are usually correlated to a degree. It is this aspect of our model which we must properly take into account. As mentioned in Section 3.1, either the e_n or the i_n can be split in two parts—one fully
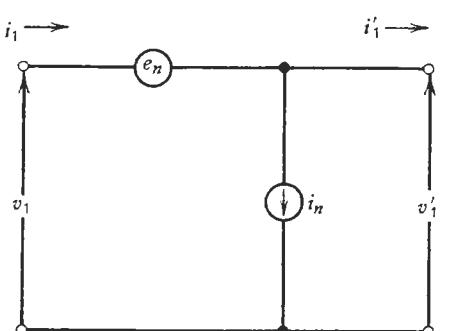

FIGURE 6.5. Representation of a noisy two-port using equivalent input generators.

correlated and the other uncorrelated. Since the two configurations are duals of one another, we shall for simplicity consider only one of them. Thus let the series voltage generator e_n be given by

$$e_n = e_u + e(i) \tag{6-12}$$

where e_u designates the uncorrelated portion of e_n and $e(i)$ designates the fully correlated (with i) part. In view of full correlation between $e(i)$ and i_n, the former must be proportional to i_n, with the constant of proportionality having the dimensions of a complex impedance. Thus

$$e_n = e_u + i_n Z_c \tag{6-13}$$

where $Z_c = R_c + jX_c$ is called the correlation impedance. The latter is related to the correlation coefficient introduced in (3-10):

$$\rho = \frac{\overline{e_n i_n^*}}{\sqrt{\overline{|e_n|^2}}\sqrt{\overline{|i_n|^2}}} = \frac{\overline{(e_u + i_n Z_c)i_n^*}}{\sqrt{\overline{|e_n|^2}}\sqrt{\overline{|i_n|^2}}} = \frac{\sqrt{\overline{|i_n|^2}}\, Z_c}{\sqrt{\overline{|e_n|^2}}} \tag{6-14}$$

where the complex conjugate form is required on account of the voltage and current being generally complex quantities. Equations (6-14) also used the fact that for any complex quantity X,

$$\overline{XX^*} = \overline{|X|^2}$$

and e_u and i_n^* are uncorrelated. The time average of their product is therefore zero.

Returning to Fig. 6.5, we note that

$$i_1 = i_n + i_1' \tag{6-15}$$

and

$$v_1 = e_n + v_1' = e_u + i_n Z_c + v_1' \tag{6-16}$$

Hence from (6-15),

$$v_1 = e_u + (i_1 - i_1')Z_c + v_1' \tag{6-17}$$

Equation (6-17) leads to the equivalent circuit shown in Fig. 6.6 where both correlation impedances Z_c and $-Z_c$ are noise-free, that is, their noise temperature is considered to be $0°K$. The derivation of an analogous dual equivalent circuit using correlation admittances and an uncorrelated current source i_u is left as an exercise.

Now, $\overline{e_u^2}$ and $\overline{i_n^2}$ could be modeled as an equivalent resistance and conductance at the reference temperature T_0 in accordance with (3-21):

$$\overline{e_u^2} = 4kT_0 R_u\, df \tag{6-18}$$

and

$$\overline{i_n^2} = 4kT_0 G_n\, df \tag{6-19}$$

We thus obtain the final noise equivalent circuit shown in Fig. 6.7. This model describes completely the noise properties of the original two-port as far as the input and output terminals are concerned. Note that for a complete noise description *four* quantities are required: R_u, G_n, R_c, and X_c.

As a further convenience we shall derive the expression for an equivalent total input resistance R', which would account for the total noise observed at terminals $a-a'$. We are basically asking for an equivalent resistance which would yield the same open-circuit noise voltage at $a-a'$ as the network shown to the left of $a-a'$ in Fig. 6.7a. This voltage is easily obtained by summing at $a-a'$ the mean-square contributions (they are all uncorrelated!) from R_u, G_n, and R_g. It will be remembered that Z_c was defined to be at $0°K$, and hence produces no noise. Thus

$$\overline{e_{oc}^2} = \overline{e_g^2} + 4kT_0 R_u\, df + \overline{i_n^2}|Z_c + R_g|^2 \tag{6-20}$$

Recall that e_u is a generator of zero internal impedance. Hence the last term in (6-20) shows voltage drops across Z_c and R_g only. From (6-20) we now get (Fig. 6.7b)

$$4kT_0 R'\, df = 4kT_0\, df\left(R_g + R_u + G_n|Z_c + R_g|^2 \right)$$
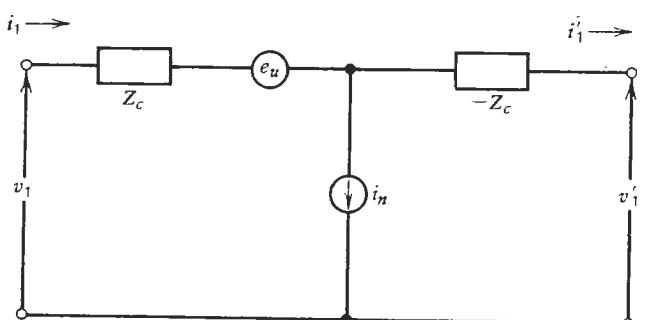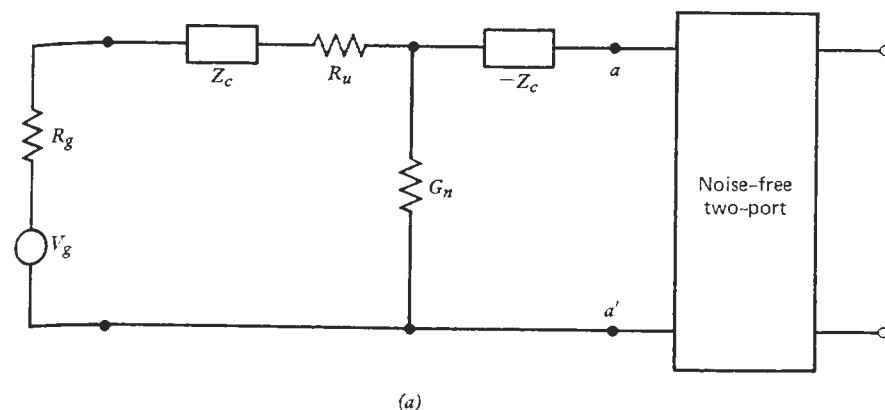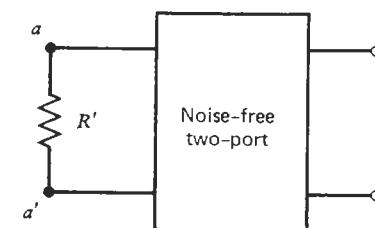

FIGURE 6.6.   Complete noise model of a two-port showing correlation effects.

(a)

(b)

FIGURE 6.7.   The use of equivalent noise resistances in the noise model of a two-port.

or

$$R' = R_g + R_u + G_n|Z_c + R_g|^2 \tag{6-21}$$

The value of the noise model shown in Fig. 6.7 will become more apparent in Section 9.9 where we shall treat the dependence of the noise performance of a given two-port on the generator impedance Z_g.

6.3.   SUMMARY

(a)   Internal sources of a two-port network can be represented by either of the following:

(1)   Two equivalent series voltage generators at the input and output terminals, respectively.

(2)   Two equivalent shunt current generators at the input and output terminals, respectively.

(3)   A voltage and a current generator, both either at the input or the output terminals.

(b) The equivalent generators are generally not independent. Their dependence can be expressed in terms of a correlation impedance $Z_c = R_c + jX_c$ or admittance $Y_c = G_c + jB_c$.

(c) A noisy two-port can be separated into two parts—one consisting of the equivalent circuit given in (a)(3) above and including all noise contributions and the other consisting of the initial network, now considered noise-free.

REFERENCES

1. H. C. Montgomery, "Transistor Noise in Circuit Applications," *Proc. IRE*, vol. 40, pp. 1461–1471, November 1952.
2. L. C. Peterson, "Signal and Noise in Microwave Tetrode," *Proc. IRE*, vol. 35, pp. 1264–1272, November 1947.
3. H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE*, vol. 44, pp. 811–818, June 1956.
4. H. A. Haus, chairman, et al., "Representation of Noise in Linear Twoports," *Proc. IRE*, vol. 48, pp. 69–74, January 1960.
5. H. A. Haus, chairman, et al., "IRE Standards on Methods of Measuring Noise in Linear Twoports," *Proc. IRE*, vol. 48, pp. 60–68, January 1960.

7 | DEFINITION OF GAIN

The concept of available power gain is an important cornerstone of noise analysis, and must be thoroughly understood to avoid pitfalls in analyzing various practical circuits. This statement may sound self-evident, but it really is not for two reasons. First, the power gain of a two-port network could be defined in a number of ways of which the available power gain G_a is only one of several possibilities. Furthermore, it is not uncommon in the literature to see the term "available" being dropped altogether. Thus a casual student may well conclude that any gain, however defined or measured, is usable and valid when dealing with thermal noise and noise figure problems. Such is definitely not the case! This lack of rigor is particularly unfortunate because it could lead to gross errors in the analysis.

Second, even if one did accept the use of G_a, it may not be obvious why some other gain definition would not do. The question deserves an answer, particularly since for some reason the subject is rarely discussed in textbooks. We shall try in this chapter to clarify the issue by starting with the basic concepts. Much of the following material is undoubtedly familiar to the reader. Still, for the sake of completeness and of having brought it together for easy reference, a detailed presentation seems warranted. At the very least, it would constitute a useful review and would give the reader a better background [1].

We shall start by describing and analyzing seven gain definitions:

Voltage gain or transfer function, A_v.
Current gain or transfer function, A_i.
Direct power gain G_p.
Insertion power gain G_i.
Transducer power gain G_t.
Available power gain G_a.
Signal gain G_s.

It should then become apparent why the G_a, and to a lesser extent the G_t and G_s, are the only definitions suitable for our purposes. In subsequent sections the available power gain G_a will be calculated for a few simple networks. It will then be seen that some of the results are not so obvious—a fair warning to those who may be tempted to apply preconceived ideas of gain to G_a.

7.1. VOLTAGE AND CURRENT GAINS A_v AND A_i

The simplest definitions of gain for a two-port network follow from the concept of the transmittance function described in Section 5.1. In particular, we are interested in the ratios of output voltage to input voltage, or output current to input current. Another name for these is the voltage and current transfer functions, or voltage and current gains, in short. Both parameters are expressed in terms of the familiar admittance and impedance matrices given in (6-1) and (6-2). By writing these in their algebraic forms, and referring to Fig. 7.1, we get

$$i_1 = y_{11}v_1 + y_{12}v_2$$

$$i_2 = y_{21}v_1 + y_{22}v_2$$

$$i_1 \longrightarrow \qquad\qquad \longleftarrow i_2$$

FIGURE 7.1.

---
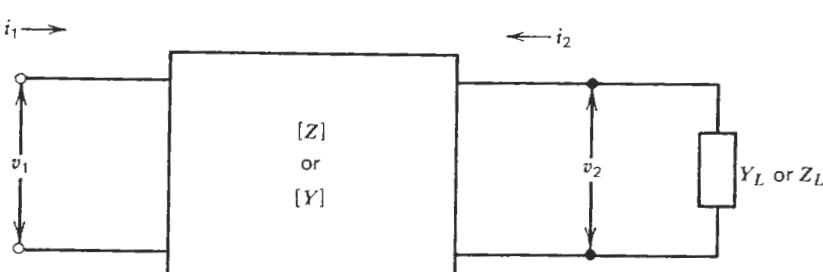
and

$$v_1 = z_{11}i_1 + z_{12}i_2$$

$$v_2 = z_{21}i_1 + z_{22}i_2$$

From these the voltage gain is

$$A_v = \frac{v_2}{v_1} = -\frac{y_{21}}{y_{22} + Y_L} \tag{7-1}$$

and the current gain is

$$A_i = \frac{i_2}{i_1} = -\frac{z_{21}}{z_{22} + Z_L} \tag{7-2}$$

Note that the minus sign is the consequence of the assumed direction of the output current, that is, $v_2/i_2 = -Z_L = -1/Y_L$.

These definitions of gain are limited in applicability. True, Eqs. (7-1) and (7-2) are valid measures of the voltage and current gains of the network, but, as will be shown shortly, the often carelessly used ratios

$$\left|\frac{v_2}{v_1}\right|^2 \quad \text{and} \quad \left|\frac{i_2}{i_1}\right|^2 \tag{7-3}$$

do *not* automatically represent "power gain" unless certain specific conditions are met. The quantities in (7-3) have their places in circuit analysis, and we shall use them, but they do not lead directly to a power gain usable for our purposes.

At times one encounters the open-circuit voltage gain A_{vo} and the short-circuit current gain A_{io}. These are merely special cases of (7-1) and (7-2), with $Z_L = 0$ and $Y_L = 0$.

7.2. DIRECT POWER GAIN G_p

If we were to ask for the simplest definition of power gain for a two-port network, such as that in Fig. 7.2, the answer would probably be

$$\text{Power Gain} = \frac{\text{power out at port 2 into } R_L}{\text{power in at port 1}} \tag{7-4}$$

We shall call this ratio the direct power gain G_p. There is noth
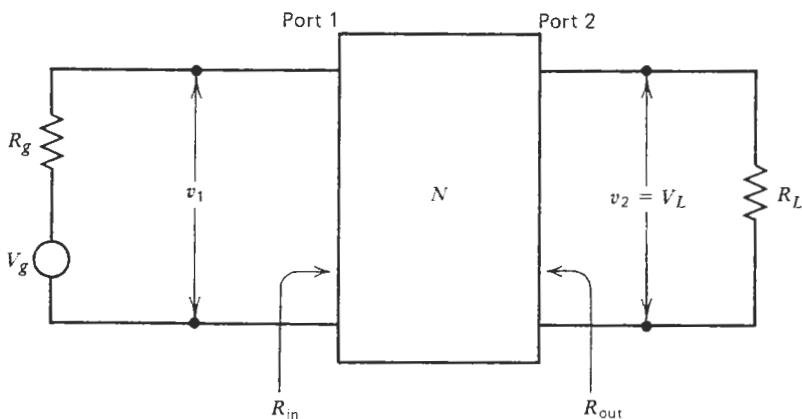
FIGURE 7.2. Definition of circuit parameters.

given in Fig. 7.2, the G_p can be easily expressed as

$$G_p = \frac{|v_2|^2 R_{in}}{R_L |v_1|^2} = |A_v|^2 \frac{R_{in}}{R_L} = \left|\frac{v_2^2}{V_g}\right| \frac{(R_g + R_{in})^2}{R_L R_{in}} \qquad (7\text{-}5)$$

where A_v is defined in (7-1). At this point we shall introduce a simplification. Since the purpose of this chapter is to review the concepts rather than to present a general theory of two-ports, all impedances will be assumed real. That is, throughout the rest of this chapter we shall take $Z_L = R_L$, $Z_{in} = R_{in}$, all z_{mn} = real, and so on. This also allows us to drop the absolute-value signs around A_v. The final conclusions are not affected by this, but unnecessary complexity is avoided.

Since the definition of (7-4) is based on the input power, it is not surprising that G_p depends on the input resistance R_{in} of the network. This creates a problem. Clearly, if the R_{in} happens to be high, the direct power gain could approach infinity because for $R_{in} \gg R_g$,

$$G_p \approx \left(\frac{v_2}{V_g}\right)^2 \frac{R_{in}}{R_L} \qquad (7\text{-}6)$$

This is hardly a useful result. By the way, (7-5) also illustrates why the square of either the voltage or the current gain is not necessarily equal to the power gain—the equality holds only if $R_L = R_{in}$.

The second major difficulty stems from the fact that a thermal noise source is characterized by its *available* output power, $P_n = kTB$. Suppose that we had the circuit shown in Fig. 7.2, and had been asked to calculate the *output* noise power reaching the load R_L (assuming, for simplicity, that the network N itself is noise-free). It should be fairly obvious that the answer is not neces-

sarily equal to the product of kTB and G_p because in general, that is, for $R_g \neq R_{in}$, the power delivered to R_L is not equal to $(P_L/P_{in})P_{av}$.

7.3. INSERTION POWER GAIN G_i

The insertion power gain of a two-port is defined as

$$G_i = \frac{\text{power delivered to the load with network in}}{\text{power delivered to the load with network out}} \qquad (7\text{-}7)$$

Again, in terms of the parameters in Fig. 7.3, the insertion power gain G_i becomes

$$G_i = \left(\frac{v_2}{V_g}\right)^2 \left(\frac{R_g + R_L}{R_L}\right)^2 = (A_v)^2 \left(\frac{R_g + R_L}{R_L}\right)^2 \left(\frac{R_{in}}{R_g + R_{in}}\right)^2 \qquad (7\text{-}8)$$

The insertion power gain suffers from two shortcomings. First, it cannot in general be cascaded. The overall insertion gain $(G_i)_T$ of N two-ports in


Network out

Network in
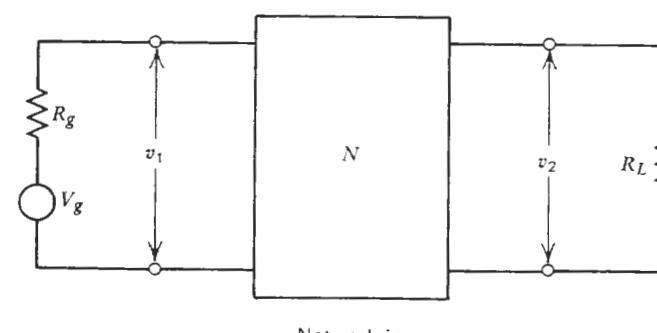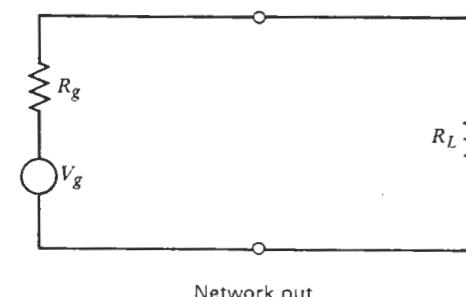
FIGURE 7.3. Definition of insertion power gain.

cascade, having arbitrary input and output impedances, is not easily express-ible in terms of the individual insertion gains. It certainly is *not* equal to the product of individual gains. Second, the definition is not compatible with the available power concept used in defining kTB. The argument here is analogous to that presented in Section 7.2.

Finally, for a complete power transfer the insertion power gain could assume almost any value. This is contrary to our usual notion of power flow through passive networks. To see this, consider Fig. 7.4. Here we have a generator of internal resistance R_g and a load resistance R_L. It is obvious that a complete power transfer is obtained only if the network N consisted of an impedance transformer having a turns ratio of $(n_1/n_2)^2 = R_g/R_L$. The input resistance R_{in} then becomes

$$R_{in} = \left(\frac{n_1}{n_2}\right)^2 R_L = R_g \tag{7-9}$$

and the voltage gain A_v is

$$A_v = \frac{v_2}{v_1} = \frac{n_2}{n_1} = \sqrt{\frac{R_L}{R_g}} \tag{7-10}$$

Combining (7-8)–(7-10) yields

$$G_i = \frac{(R_g + R_L)^2}{4R_L R_g} \tag{7-11}$$

which for $R_g \gg R_L$ becomes

$$G_i \approx \frac{R_g}{4R_L} \tag{7-12}$$
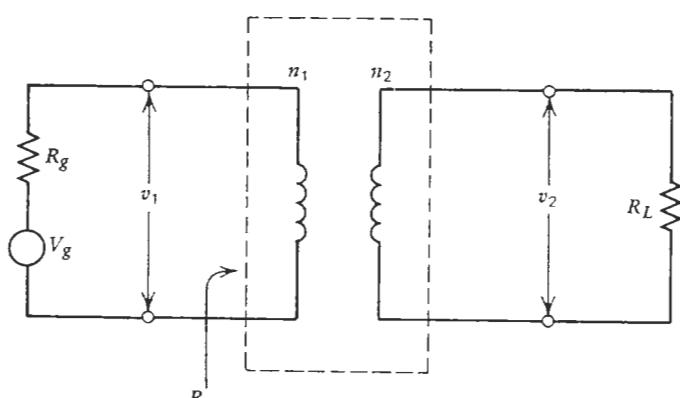
FIGURE 7.4. Example for insertion power gain.

Thus at complete power transfer the G_i could become arbitrarily large depending on the ratio of R_g to R_L. Since the transformer has corrected the initial gross mismatch, there is indeed a "gain" of sorts. In fact, a somewhat similar argument is used in defining the power gain of an antenna. There the "gain" is due to the directivity of the antenna which concentrates the trans-mitter power in the desired direction rather than allowing it to scatter over the entire 4π steradians. While the antenna gain is a useful concept and is universally accepted, the insertion power gain greater than unity for a passive network seems odd and is not much used.

An alert reader has probably noticed that the difficulties encountered with the G_p and G_i were due to an ill-chosen reference condition. In the first case, that reference was the input power which for high-impedance devices could be near zero. In the second case, the initial mismatched state of the circuit constituted the reference. Could a better reference condition be devised to avoid such eccentric results while still retaining a simple, concise definition for power gain? The answer is yes if the *available* power of the input generator is used as the reference. The following two definitions of power gain will therefore explore this possibility.

7.4. TRANSDUCER POWER GAIN G_t

As we saw earlier, the maximum power that an ideal signal generator of open-circuit voltage V_g and internal impedance $Z_g = R_g + jX_g$ can deliver into a conjugate load impedance is given by

$$P = \frac{V_g^2}{4R_g} \tag{7-13}$$

This is called the maximum available power, or available power, in short. Now define the transducer power gain of a two-port as

$$G_t = \frac{\text{power actually delivered to the load}}{\text{power available from the generator}} \tag{7-14}$$

In terms of the circuit parameters of Fig. 7.2,

$$G_t = \left(\frac{v_2}{V_g}\right)^2 \frac{4R_g}{R_L} = (A_v)^2 \frac{R_{in}^2}{(R_g + R_{in})^2}\left(\frac{4R_g}{R_L}\right) \tag{7-15}$$

The transducer power gain G_t does have the property that it becomes unity for complete power transfer in a passive circuit. A substitution of (7-9) and (7-10) into (7-15) leads immediately to $G_t = 1$, which is a much more palatable result, considering our ingrained ideas of power relationships in passive

networks. More importantly, however, the G_t allows us to calculate directly the power delivered to the output load when the input signal is characterized in terms of available power:

$$P_L = G_t P_{av} = \left(\frac{P_L}{P_{av}}\right) P_{av} = P_L \qquad (7\text{-}16)$$

It will be remembered that this could not be done with either G_p or G_i.

The one problem still remaining is that G_t cannot in general be cascaded. Note that if G_{t1} and G_{t2} represent the transducer gains of two arbitrary networks (Fig. 7.5), their product [using (7-14)]

$$G_{t1}G_{t2} = \left(\frac{P_{L1}}{P_{av1}}\right)\left(\frac{P_{L2}}{P_{av2}}\right) \qquad (7\text{-}17)$$

is not necessarily equal to the overall transducer gain

$$(G_t)_T = \frac{P_{L2}}{P_{av1}} \qquad (7\text{-}18)$$

unless

$$P_{L1} = P_{av2} \qquad (7\text{-}19)$$

Here P_{L1} and P_{L2} are the power delivered to the load by networks 1 and 2, respectively; P_{av1} and P_{av2} are the available power from the generator driving networks 1 and 2, respectively.

In turn, (7-19) is true only if

(a) $R_{out1} = R_{in2} = R_{L1}$.
(b) G_{t2} was measured with R_{out1} as the generator resistance.
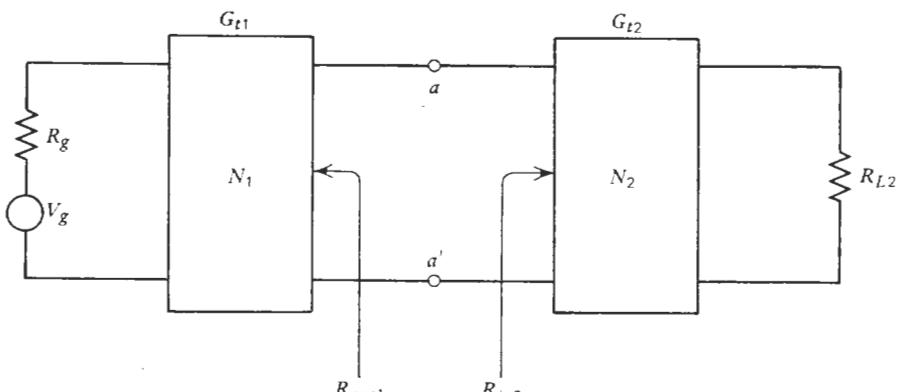
FIGURE 7.5. Two networks in cascade.

Except for this problem, the G_t is a useful definition and, in view of (7-16), quite applicable to calculations of thermal noise. If, however, a cascading of networks is required, it is the available power gain that solves the problem, as we shall see in the next section.

Another useful feature of G_t is that its value reflects the effects of both input and output mismatches. Clearly, if $R_{in} \neq R_g$, the network will not receive the full available power from the generator, and P_L will suffer accordingly. If $R_{out} \neq R_L$, there is further degradation in P_L.

7.5. AVAILABLE POWER GAIN G_a

The available power gain G_a is defined as

$$G_a = \frac{\text{available power from the network}}{\text{available power from the generator}} \qquad (7\text{-}20)$$

$$= \left(\frac{V_{out}}{V_g}\right)^2 \left(\frac{R_g}{R_{out}}\right) \qquad (7\text{-}21)$$

where V_{out} is the open-circuit output voltage of the network and R_{out} the output resistance of the network.

The first important observation is that G_a does not depend on the load. Hence it gives no information on the output circuit or its match. However, G_a *does* depend on the generator resistance R_g and on the manner the generator is coupled to the network. Note that this ultimately determines the V_{out}. Thus the available power gain G_a is not a unique characteristic of the network but depends on the generator used in measuring G_a, although a match between R_g and R_{in} is neither required nor implied. This is a very important aspect of G_a (and also of G_t), and should be fully appreciated. Another way of expressing this is to say that for N networks in cascade, the G_a of the kth network is determined by and must be measured with the output resistance of the $(k-1)$th network as the generator. Unless this rule is observed, the cascading equation (7-22) does not hold. We shall soon illustrate (Section 7.7) the correct procedure of calculating the G_a.

Assuming that the G_{ak} were correctly calculated or measured, it follows from (7-20) that for N two-ports in cascade the total gain $(G_a)_T$ is

$$(G_a)_T = G_{a1}G_{a2} \cdots G_{aN} \qquad (7\text{-}22)$$

$$= \frac{P_1}{P_{av}}\frac{P_2}{P_1} \cdots \frac{P_N}{P_{N-1}} = \frac{P_N}{P_{av}} \qquad (7\text{-}23)$$

where P_k is the available output power of the kth two-port and P_{av} is the available power from the generator.

Before leaving this subject, it is well to note a subtle difference between G_a and G_t. Let us consider Fig. 7.2 again and assume that both gains are known for the network N. Furthermore, take the general case for which $R_{out} \neq R_L$. We can now calculate the output power from either (7-16) using G_t or from (7-23) using G_a. What is the difference? Clearly, if we seek that output power which is actually dissipated in R_L, then G_t applies. If, on the other hand, the *maximum available* output power is of interest, that is, the optimum case, then G_a should be used. The distinction vanishes when $R_L \rightarrow R_{out}$.

7.6. POWER GAIN IN MATCHED CIRCUITS

In view of the many possible definitions of power gain and the substantial differences between them, one may well wonder how do we generally manage to get correct results despite the obvious pitfalls on the way. After all, one rarely gives much thought to the intricacies of the power gain when performing routine laboratory tests. The answer is that in most cases, particularly in RF and microwave work, there is an impedance match throughout the system. The signal generators, the connecting cables, the loads, and so on, are usually 50-ohm devices, or there are matched waveguide interfaces. Table 7.1 tabulates the results for the case where

$$R_g = R_{in} = R_L = R_{out} \tag{7-24}$$

In the case of G_a we note that since $R_{out} = R_L$, we get $v_2 = \frac{1}{2} V_{out}$. Hence

$$G_a = \left(\frac{V_{out}}{V_g}\right)^2 = 4\left(\frac{v_2}{V_g}\right)^2 \tag{7-25}$$

TABLE 7.1. Summary of Power Gains for a Linear Two-Port

Type of Gain	General Case	Complete Match $(R_g = R_{in} = R_{out} = R_L)$
Power gain G_p	$\left(\dfrac{v_2}{V_g}\right)^2 \dfrac{(R_g + R_{in})^2}{R_L R_{in}}$	$4\left(\dfrac{v_2}{V_g}\right)^2 = (A_v)^2$
Insertion gain G_i	$\left(\dfrac{v_2}{V_g}\right)^2 \left(\dfrac{R_g + R_L}{R_L}\right)^2$	$4\left(\dfrac{v_2}{V_g}\right)^2 = (A_v)^2$
Transducer gain G_t	$\left(\dfrac{v_2}{V_g}\right)^2 \dfrac{4R_g}{R_L}$	$4\left(\dfrac{v_2}{V_g}\right)^2 = (A_v)^2$
Available gain G_a	$\left(\dfrac{V_{out}}{V_g}\right)^2 \dfrac{R_g}{R_{out}}$	$\left(\dfrac{V_{out}}{V_g}\right)^2 = 4\left(\dfrac{v_2}{V_g}\right)^2 = (A_v)^2$

Furthermore, using the input match condition $R_g = R_{in}$, it follows that

$$V_g = 2v_1$$

and all gains become

$$G = 4\left(\frac{v_2}{2v_1}\right)^2 = \left(\frac{v_2}{v_1}\right)^2 = (A_v)^2 \tag{7-26}$$

Thus for a completely matched system, all power-gain definitions converge and yield the same result. The power gain is then simply the voltage gain squared. Note, however, that this simplification is valid *only* for a special case, and is not at all true in general. Where (7-24) does not hold, either in entirety or in part, caution is in order. The available power gain must then be used and carefully calculated to ensure correctness in subsequent analysis.

7.7. CALCULATION OF AVAILABLE POWER GAIN

The preceding sections have hopefully demonstrated that it is the available power gain G_a which is best suited for the analysis of thermal noise. We shall now calculate G_a for a few simple cases to illustrate the procedure [2].

Shunt Resistor

Consider the case of a shunt resistor R connected to a generator V_g of internal resistance R_g as shown in Fig. 7.6. The available input power from the generator at terminals $a-a'$ is

$$P_1 = \frac{V_g^2}{4R_g} \tag{7-27}$$
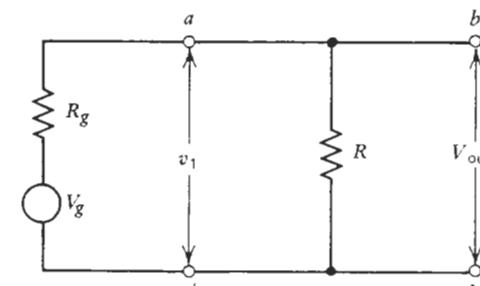
FIGURE 7.6.   Calculation of G_a for a shunt resistor.

The available power at terminals b–b' is

$$P_2 = \frac{\left[V_g R/(R_g + R)^2\right]}{4RR_g/(R_g + R)} \qquad (7\text{-}28)$$

The ratio of (7-28) to (7-27) yields the G_a which is

$$G_a = \frac{R}{R_g + R} \qquad (7\text{-}29)$$

This result may seem unexpected at first, for there is no square term in (7-29) despite our intuitive feel that power gain must have or be "something squared." To be sure, a square term $(V_{\text{out}}/v_1)^2$ is involved, but since the ratio is unity the square term is not apparent.

Shunt Reactance

Consider next a lossless shunt reactance jx (Fig. 7.7). The output impedance of the "network" is now complex and in evaluating the available power at the output we must use

$$P_2 = \frac{|V_{\text{out}}|^2}{4\,\text{Re}(Z_{\text{out}})} \qquad (7\text{-}30)$$

since by definition the available power is that power which a source of a given internal impedance can deliver to a conjugate load.

Applying (7-30) to Fig. 7.7 yields

$$P_2 = \frac{(V_g x)^2/(R_g^2 + x^2)}{(R_g^2 + x^2)/4R_g x} = \frac{V_g^2}{4R_g} \qquad (7\text{-}31)$$
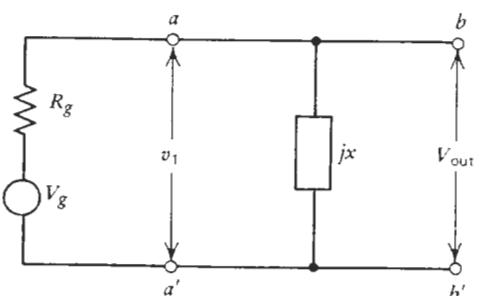
FIGURE 7.7. Calculation of G_a for a shunt reactance.

But this is the same as the available power from the input generator. It therefore follows that for a shunt reactance, $G_a = 1$. The same is true for a series reactance. In fact, as shown in Appendix B, the available power gain of any lossless two-port is always equal to unity.

As a comparison, it is instructive to calculate the transducer gain G_t for the same case. Consider Fig. 7.8 and apply (7-15):

$$|V_L|^2 = \frac{V_g^2}{\left(1 + R_g/R_L\right)^2 + \left(R_g/x\right)^2} \qquad (7\text{-}32)$$

and

$$P_L = \frac{V_g^2}{R_L\left[\left(1 + R_g/R_L\right)^2 + \left(R_g/x\right)^2\right]} \qquad (7\text{-}33)$$

Since $P_1 = V_g^2/4R_g$, the ratio of (7-33) to P_1 [in accordance with (7-15)] is the transducer power gain G_t, or

$$G_t = \frac{4R_g}{R_L\left[\left(1 + R_g/R_L\right)^2 + \left(R_g/x\right)^2\right]} \qquad (7\text{-}34)$$

and for $R_g = R_L$,

$$G_t = \frac{1}{1 + \left(R_g/2x\right)^2} \qquad (7\text{-}35)$$

Thus for any value of x other than infinity, G_t will always be less than unity. The G_a, on the other hand, is always equal to unity for lossless networks as we saw earlier.

Amplifier

As the last example, we shall consider an active network, such as an amplifier shown in Fig. 7.9. Let the input impedance of the amplifier be infinite and let
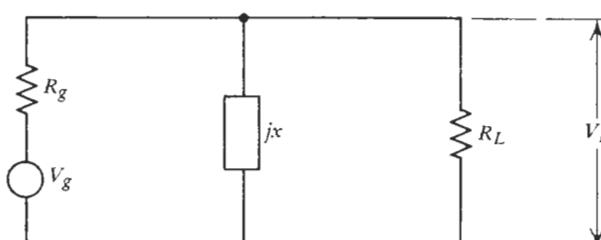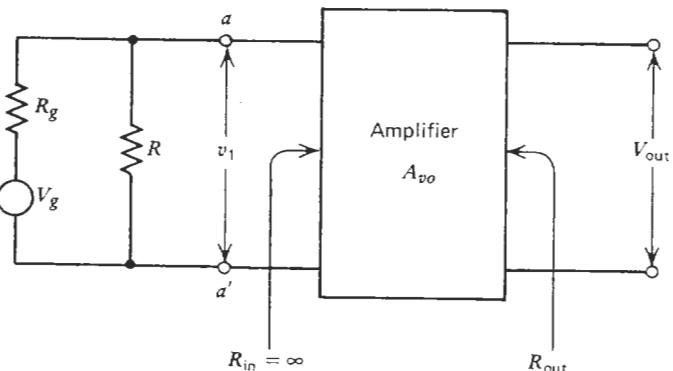

FIGURE 7.8. Calculation of G_t for a shunt reactance.

FIGURE 7.9. Calculation of G_a for a combined network.

A_{vo} be the open-circuit voltage gain of the device. A resistor R is connected in shunt across the input terminals. We shall evaluate the total G_a by two methods—direct calculation and cascading.

Direct Calculation. This method uses the definition of G_a, (7-20), directly. The available power from the generator is given by

$$P_1 = \frac{V_g^2}{4R_g}$$

as before.

The available output power P_2 is

$$P_2 = \frac{V_{out}^2}{4R_{out}} = \frac{(A_{vo}v_1)^2}{4R_{out}} = \frac{(A_{vo}V_g R)^2}{4R_{out}(R_g + R)^2} \tag{7-36}$$

Hence

$$G_a = \frac{P_2}{P_1} = \frac{A_{vo}^2 R_g}{R_{out}} \left(\frac{R}{R + R_g} \right)^2 \tag{7-37}$$

where A_{vo} is as discussed in Section 7.1.

Cascading. In view of (7-22), we should get the same answer by cascading the available power gain of the shunt resistor and that of the amplifier in accordance with

$$G_a = (G_a)_R (G_a)_{amp} \tag{7-38}$$

We already calculated $(G_a)_R$ in (7-29). It is only necessary to find $(G_a)_{amp}$, keeping in mind that the available power gain of the amplifier must be evaluated using the output impedance of the preceding stage as the generator. In our case, everything to the left of $a-a'$ constitutes the generator for the amplifier. The generator resistance is therefore not just R_g, but R_g in parallel with R. Similarly, the effective generator voltage is not V_g but V_g', the Thevenin equivalent appearing at $a-a'$. Thus

$$V_g' = \frac{V_g R}{R_g + R} \qquad R_g' = \frac{R R_g}{R_g + R}$$

and therefore the available power of the "generator" is

$$P_1 = \frac{(V_g')^2}{4R_g'} = \frac{V_g^2 R}{4R_g(R_g + R)} \tag{7-39}$$

The available power at the *output* of the amplifier was calculated in (7-36). The ratio of (7-36) to (7-39) now yields the desired $(G_a)_{amp}$:

$$(G_a)_{amp} = \frac{(A_{vo}V_g R)^2}{4R_{out}(R_g + R)^2} \left(\frac{4R_g(R_g + R)}{V_g^2 R} \right)$$

$$= \frac{A_{vo}^2 R_g R}{R_{out}(R_g + R)} \tag{7-40}$$

We are now ready to cascade the two power gains by forming the product of $(G_a)_R$ from (7-29) and $(G_a)_{amp}$ from (7-40):

$$G_a = \left(\frac{R}{R_g + R} \right) \left(\frac{A_{vo}^2 R_g R}{R_{out}(R_g + R)} \right) = \left(\frac{A_{vo}^2 R_g}{R_{out}} \right) \left(\frac{R}{R_g + R} \right)^2 \tag{7-41}$$
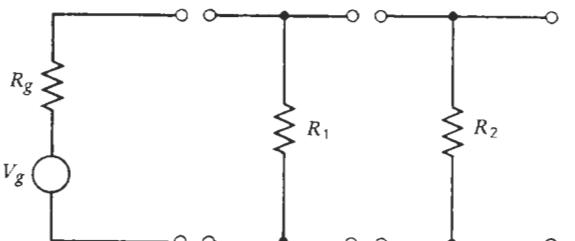
which is indeed the same as (7-37).

As a final exercise, the reader should work out the overall G_a for two shunt resistors in cascade as in Fig. 7.10. It will be quickly evident that the answer is not equal to the product of

$$\frac{R_1}{R_g + R_1} \quad \text{and} \quad \frac{R_2}{R_g + R_2}$$

as a hasty application of (7-22) would suggest.

Since the majority of cases in this book will use the available power gain, we shall for simplicity drop the subscript on G_a and the symbol G will henceforth

FIGURE 7.10.   Calculation of G_a for two shunt resistors in tandem.

be used to designate the available power gain. Wherever the symbol G is used for other purposes, or if other definitions of power gain are applicable, proper identification will be given to avoid confusion. For example, the transducer power gain will continue to be designated by G_t.

7.8.   SIGNAL GAIN G_s

The preceding definitions of gain have all referred to two-port networks having a single response, that is, a single input frequency is associated with a given output frequency. In cases involving multi-response networks (see Chapter 10) a parameter called the signal gain is used and is defined as follows [3]:

> The signal gain G_s is the ratio of (a) the signal power *delivered* at the specified output frequency into the output circuit under operating conditions, to (b) the signal power *available* at the corresponding input frequency or frequencies to the system under operating conditions at its accessible input terminations.   □

We shall discuss this topic later in Chapter 10. It suffices here to state that a multi-response system is one in which several different input frequencies contribute to one output frequency. A typical example would be a superheterodyne receiver.

The concept of signal gain is best understood when a broadband radiometry measurement is considered. Here the input signal is in the form of broadband noise distributed over the signal as well as the various image responses of a superheterodyne receiver. The latter converts all these contributions to one IF frequency. If the input signals in various responses are uncorrelated, then the signal gain G_s can be mathematically expressed as

$$G_s = \frac{S_{i1}G_{t1} + S_{i2}G_{t2} + \cdots + S_{iN}G_{iN}}{S_{i1} + S_{i2} + \cdots + S_{iN}} \tag{7-42}$$

where the S_{ij} are the uncorrelated input signal powers and G_{tj} are the

corresponding transducer power gains of the individual responses. For a single-response system, all gains other than G_{t1} are zero, and the signal gain is simply equal to the transducer gain. Even in multi-response systems, if the gains are all equal, that is, $G_{t1} = G_{t2} = \cdots = G_{tN}$, Eq. (7-42) reduces to the simple form $G_s = G_t$. When this is not true, then the G_s must be calculated from (7-42) for the particular set of conditions. It should be noted that G_s is maximum when the received signal is entirely in that response which has the largest G_t. If the same signal were distributed over several responses, the G_s would always be less than this maximum.

If the various portions of input signal received through the different responses are either partially or totally correlated, the situation is much more complicated and (7-42) becomes a complex function. In principle, the G_s can still be determined, but the details will have to remain outside the scope of this book.

7.9.   SUMMARY

(a)   Voltage gain:

$$A_v = \frac{\text{voltage out}}{\text{voltage in}}$$

(b)   Current gain:

$$A_i = \frac{\text{current out}}{\text{current in}}$$

(c)   Direct power gain:

$$G_p = \frac{\text{power delivered to load}}{\text{power in}}$$

(d)   Insertion power gain:

$$G_i = \frac{\text{power into load, network in}}{\text{power into load, network out}}$$

(e)   Transducer power gain:

$$G_t = \frac{\text{power delivered to load}}{\text{power available from source}}$$

(f)   Available power gain:

$$G_a = \frac{\text{power available from the network}}{\text{power available from the source}}$$

(g) G_a and G_t are best suited for the noise analysis because thermal noise power is defined in terms of *available* power.

(h) G_a is independent of the load, but does depend on the source impedance.

(i) G_a can be cascaded, but G_t can, in general, not be cascaded.

(j) For fully matched systems, all power-gain definitions yield the same result, that is, A_v^2.

(k) G_a of a lossless network is always unity.

(l) Signal gain G_s relates the total signal power at output, received through several responses (images, etc.), to the total signal power at input.

REFERENCES

1. "IRE Standards on Electron Tubes: Definitions of Terms, 1957," *Proc. IRE*, vol. 45, pp. 983–1010, July 1957.

2. H. Goldberg, "Some Notes on Noise Figure," *Proc. IRE*, vol. 36, pp. 1205–1214, October 1948.

3. "IRE Standards on Electron Tubes: Definitions of Terms, 1962," *Proc. IRE*, vol. 51, pp. 434–435, March 1963.

8 | NOISE TEMPERATURE

This chapter will give a detailed discussion of the concept of noise temperature and the various extensions and applications of this idea [1], including the effective input noise temperature T_e, and the operating noise temperature T_{op} of a linear transducer. Note the change in terminology here. Previous chapters dealt primarily with passive, mostly lossy, circuits. Hence the terms "circuit" or "network" were used. Since the emphasis will henceforth slowly shift to active circuits, amplifiers, receivers, and so on, the term "transducer" will be used to designate the black box under study. A transducer can be defined as a device that is actuated by power from one system and supplies power in any other form to a second system [2]. In our case the first system is typically the dc bias or B+ voltage applied to the device, and the second system is the RF. This would exclude passive circuits and such nonlinear circuits as mixers. However, we shall use the term "transducer" in a rather loose sense to designate any device used for signal transmission, passive or otherwise, which has clearly defined input and output ports.

It must be emphasized at the outset that this chapter covers only single-response transducers. Examples include amplifiers and most two-port networks. Frequency converters, such as mixers, even though they may physically be two-port devices (RF in, RF out), are electrically multi-response transducers because several input frequencies (images, etc.) may contribute to a single output frequency. Although the specifics of multi-response operation will be

covered in Chapter 10, the basic concepts outlined here will apply to both versions and should be thoroughly understood. These concepts will form the foundation for much that is to follow.

8.1. INPUT NOISE TEMPERATURE

We saw in (3-24) that the noise power available from a resistor is directly related to its physical temperature; that is, given a resistance at temperature T, the available power of thermal noise is simply $N_o = kTB$. Later in Section 5.2, it was shown that the temperature term can assume a more abstract meaning. An *effective* noise temperature T_{eff} was defined in (5-21) and (5-22), and in Fig. 5.6, which allowed the output noise power from an arbitrary source of white noise to be expressed in terms of a fictitious temperature, not necessarily equal to any physical temperature within the given device. Let us develop this idea further.

Consider a linear two-port transducer M having an available power gain G (Fig. 8.1). The available input noise power per unit bandwidth generated by the source is

$$N_i' = kT_g \tag{8-1}$$

where the prime signifies that a unit bandwidth is used, that is, the noise density is being used. We are doing this here solely for convenience. No other meaning is implied.

In an ideal case, the resultant available *output* noise density would be $N_o' = kT_g G$. We know, of course, that in practice the measured output noise is higher because of added noise from the transducer. This added noise appearing at the output is called the excess noise of the transducer, denoted by N_{no} if the bandwidth is finite or by N_{no}' if noise density is used. The total available output noise density can therefore be written

$$N_o' = kT_g G + N_{no}' = kT_g G + kT_n \tag{8-2}$$
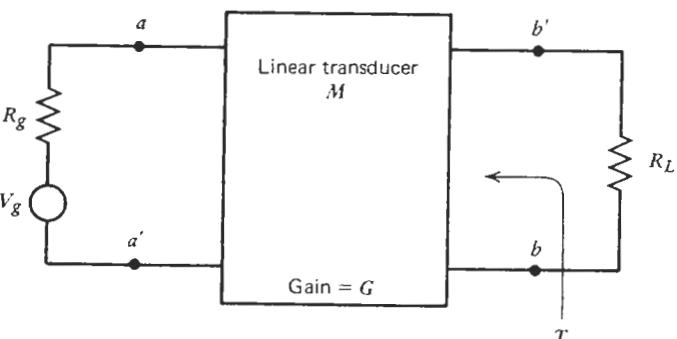
FIGURE 8.1.   Effective output noise temperature of a linear transducer.

where T_n is the effective noise temperature corresponding to the observed N_{no}'. Note that the separation of N_o' into two terms is permissible only because the generator noise and the transducer noise are uncorrelated.

We can now go one step farther and refer the N_{no}' and hence the kT_n to the input of the transducer. In essence, we postulate that the transducer is noise-free and the observed excess noise N_{no}' is due to a fictitious noise generator connected to the input terminals. Thus

$$N_o' = kT_g G + kT_n = kG(T_g + T_e) \tag{8-3}$$

The T_e is called the effective input noise temperature, or simply the input noise temperature of the transducer M, referred to the input terminals a–a':

$$T_e = \frac{N_{no}'}{kG} \tag{8-4}$$

Equation (8-3) suggests the noise model shown in Fig. 8.2 where the transducer M is considered noise-free and both noise contributions, the generator or source noise and the internal noise, appear as inputs to the transducer. The bandwidth df in Fig. 8.2 merely indicates that the total available output noise power, measured over a narrow bandwidth, is the sum of two input noise powers, each measured in the same bandwidth df and referred to the output.

Note that through this artifice we have separated the thermal noise aspects and the network aspects of the transducer. This is similar to separating a physical resistor R into two components as in Fig. 3.4—a noise-free R in series with a mean-square noise voltage generator $\overline{e_n^2}$. The input noise temperature is expressed in degrees Kelvin even though the origin for the observed noise could well be something other than a hot resistor.

The input noise temperature T_e of a two-port, single-response, linear transducer is therefore defined as follows:

T_e is that input source temperature which, when connected to a noise-free equivalent of the transducer, would result in an available output noise power per unit bandwidth equal to that of the actual, noisy transducer having a noise-free input source at $0°K$.   ☐
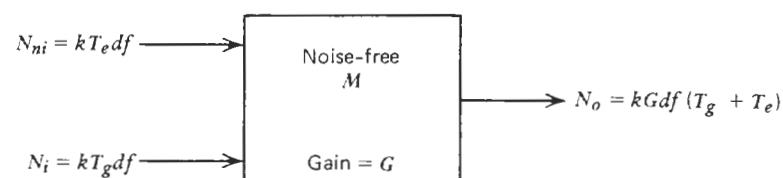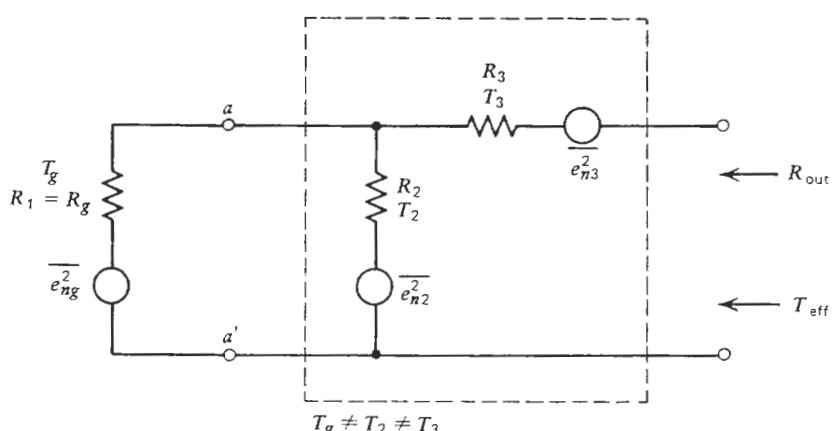

FIGURE 8.2.   Noise model of linear transducer using input noise tempe

Another way of stating this is that kT_eG yields the same *excess* output noise as that actually observed. The preceding definition underlines the general applicability of the noise temperature model to a wide variety of thermal noise problems. The effective output noise temperature T_{eff} of a two-terminal network was one example. Another example would be a gas discharge or solid-state noise source which, as we shall see in Chapter 11, could easily have an output noise temperature T_n exceeding 20,000°K. Obviously, nothing in the noise source has a physical temperature remotely approaching such a high value. Yet the thermal noise output of the device does equal that of a hypothetical resistor at T_n.

The concept of T_e is very useful in characterizing the noise properties of not just two-port transducers, but, as will be shown in Chapter 10, of general multi-response transducers as well. T_e is defined more broadly than the other widely used measure of thermal noise—the noise figure (Chapter 9), which must be based on a specific reference temperature. In this sense, the T_e is a more basic yardstick of noise than the noise figure, particularly in cases where the source or the background noise temperature is something other than the standard reference temperature of 290°K.

The T_e is not dependent on the noise temperature of the source. It does, however, depend on the source impedance. This is a very important characteristic and must be well understood. The interrelationship can be best illustrated by considering the example in Fig. 5.4, but redrawing it as in Fig. 8.3. The resistance R_1 is taken as the generator resistance R_g. Each resistor is represented by its noise model; that is, the resistor is considered noise-free and the noise contribution is lumped in the respective mean-square voltage generator in series with it. The available output noise can be written in the most general case as

$$N_o = kT_{\text{eff}}\,df \qquad (8\text{-}5)$$

$T_g \neq T_2 \neq T_3$

But using (5-25), this can also be expressed as

$$N_o = k\,df\left(\alpha_g T_g + \alpha_2 T_2 + \alpha_3 T_3\right) = k\,df\,G\left(T_g + T_e\right) \qquad (8\text{-}6)$$

From this the input noise temperature T_e of the network at terminals a–a' in Fig. 8.3 can be written as

$$T_e = \frac{\alpha_2 T_2 + \alpha_3 T_3}{G} = \left(\frac{R_g}{R_2}\right)T_2 + \frac{\left(R_g + R_2\right)^2 R_3}{R_2^2 R_g}T_3 \qquad (8\text{-}7)$$

where the α_2 and α_3 were obtained from (5-25), (5-24), and (5-17a)–(5-17c), and the gain G of the network inside the dotted lines in Fig. 8.3 came from (7-21) as

$$G = \frac{R_2^2}{\left(R_g + R_2\right)^2}\left(\frac{R_g}{R_{\text{out}}}\right) \qquad (8\text{-}8)$$

Equation (8-7) clearly shows that T_e is indeed a function of the source resistance R_g. Note further that since, in general, the α_j can be complex quantities, they become functions of frequency, which renders T_e a function of frequency also.

In defining T_e, the output load is taken as noise-free over the output frequency band when evaluating the output noise power in accordance with (8-4). This means that any noise N_L generated in the load and reflected back by the possibly mismatched output impedance of the transducer, is excluded from the definition regardless of its magnitude [3]. The reason for this is that it is useful to relate the input noise temperature T_e of a system to the system's noise factor F (Chapter 9). Since the definition of noise factor does not include the effect of load noise, the latter is excluded from T_e also.

There is one case, however, where T_e may depend on the impedance and the noise temperature of the termination at the output port. This happens when the transducer generates at its output other frequencies beside the specific output frequency. An example would be an upconverting mixer which generates an image frequency on the other side of the local oscillator frequency. The load noise at the *image* frequency, radiated back into the output port of the transducer, may be converted to noise at the *desired* output frequency. It would then appear as part of the excess output noise N_{no}, and would thus affect T_e. This is, however, a rather special case.

Before leaving this section, the noise-temperature ratio, noise ratio, or output noise ratio should be mentioned. These terms are occasionally encountered in the literature. They all mean the same thing and are defined as

$$t = \frac{T_n}{} \qquad (8\text{-}9)$$

where T_n is the noise temperature measured looking into a given pair of terminals. In other words, t is the ratio of the actual available output noise power of a device or network at these terminals, kT_nB, to the available noise power that would be measured there if T_n were equal to $T_0 = 290°$K. Sometimes (8-9) is modified to

$$t - 1 = \frac{T_n}{290} - 1 \qquad (8\text{-}10)$$

which is called the excess noise, or excess-noise ratio [not to be confused with the excess noise N'_{no} in (8-2)!]. This expression is widely used in characterizing noise sources such as noise diodes and gas-discharge tubes as we shall see in Chapter 11.

8.2. OPERATING NOISE TEMPERATURE

Returning to (8-2) and (8-3), and noting that they consist of two parts, leads to two ways of analyzing thermal noise. First, if we limit ourselves to the transducer only, then the input noise temperature T_e provides an obvious means of comparing two arbitrary transducers. The one with the higher T_e (other parameters being equal) clearly must add more noise to the signal passing through it. This excess noise, contributed by the transducer itself, is given by

$$N_{no} = kT_e G\,df \qquad (8\text{-}11)$$

Second, we could consider the *total* noise power $kG\,df(T_g + T_e)$ available at the output terminals of the transducer M as in Fig. 8.2. This quantity is important because it determines the output signal-to-noise ratio* SNR$_o$ of the system. Since the total noise depends on *both* T_e and T_g, it follows that the ultimate sensitivity of the system is determined by the source noise $kT_g\,df$ even if the receiver itself were perfect, that is, $T_e = 0°$K. We have some control over T_e, but in most cases there is no control over the incoming source noise. For this reason it is advantageous to seek a parameter which would characterize the overall noise performance of the system under actual operating conditions, that is, including both noise contributions, the source and the receiver. Such a parameter is the operating noise temperature T_{op} given by

$$T_{op} = \frac{N'_o}{kG} = \frac{N_o}{kG\,df} \qquad (8\text{-}12)$$

*The signal-to-noise ratio SNR can be defined on a full bandwidth or a unit bandwidth basis. In the first case the noise is expressed in terms of the total noise N_o. In the second case the noise per Hz, N'_o (the noise density), is used. We then have a normalized SNR. The symbols used in this book will be SNR and SNR', respectively, and a subscript SNR$_i$ or SNR$_o$ may be additionally used to identify either the input or the output SNR. Obviously,

$$\text{SNR} = (\text{SNR}')(\text{noise bandwidth})$$

The expression given above is the one commonly used in practice. The formal definition of T_{op} is somewhat broader. The IEEE definition [4, 5] gives the operating noise temperature as

$$T_{op} = \frac{N'_{oL}}{kG_s} \qquad (8\text{-}13)$$

where N'_{oL} is the output noise power per unit bandwidth at a specified output frequency *delivered* into the output load and includes the load noise N'_L. G_s is the signal gain defined in Section 7.8. Its use becomes clearer when Section 10.2 is covered. The load noise N'_L may or may not be negligible, depending on the gain of the transducer, the noise temperature of the load, and the mismatch between Z_{out} of the transducer and the load Z_L. The total delivered output noise power, particularly in view of its inclusion of the load noise, is therefore a more meaningful [6] quantity from the operational viewpoint than the available output power N'_o which represents an optimum condition. Fortunately, in microwave work impedance match usually exists between the transducer output and the load. The distinction between the delivered and the available power then disappears, rendering $N'_{oL} = N'_o$. In view of the output match, there is no contribution from the noise generated in the load because nothing is reflected back to the load by the Z_{out}.

Finally, as explained in Section 7.8, the signal gain G_s of a single-response transducer is equal to its transducer gain G_t, and this in turn is equal to the available gain G under matched output conditions. Thus (8-13) reduces to (8-12). One should, however, be aware that there is a difference between the popular, but restricted, definition of T_{op} and the more general definition given by the IEEE. The latter must be used if the assumptions given above do not hold.

Returning to (8-12), we can write the following for the specific case of a matched single-response transducer:

$$N_o = kG(T_g + T_e)\,df \qquad (8\text{-}14)$$

and therefore from (8-12) we have

$$T_{op} = \frac{kG(T_g + T_e)\,df}{kG\,df} = T_g + T_e \qquad (8\text{-}15)$$

For simplicity, the matched condition will be assumed throughout the remainder of this book, which allows the use of (8-12) and (8-15).

Equations (8-14) and (8-15) show that the actual operating conditions of the system are set by T_{op} as far as thermal noise is concerned. In systems where the source consists of a matched signal generator connected to the receiver through a physical transmission line (cable, wire, etc.), the source temperature T_g is

usually close to the standard temperature 290°K. In radio systems, on the other hand, the receiving antenna may look at a wide variety of source or background temperatures. These could range from less than 5°K for cold portions of the sky, to perhaps thousands of degrees for the Sun if the antenna is pointed directly at it, and the beamwidth of the antenna is small enough to be covered by the subtended angle of the Sun's disc. The T_{op} could thus assume a wide range of values and the noise performance of the system could vary accordingly.

Equations (8-14) and (8-15) enable the designer to trade off a number of design parameters. Suppose the task were to upgrade a communication system consisting of a transmitter and a low-noise receiver. The output SNR is to be improved. The system presently uses a receiver with a low-noise amplifier (LNA) of $T_e = 150°K$. Let the background noise be a relatively high 400°K, and consider either of two options:

(a)   Buy for the receiver a more expensive LNA of $T_e = 50°K$.

(b)   Increase the transmitter output power.

The SNR improvement achievable with the lower-noise amplifier is

$$\frac{S_{o2}/N_{o2}}{S_{o1}/N_{o1}} = \frac{N_{o1}}{N_{o2}} = \frac{T_{op1}}{T_{op2}} = \frac{400 + 150}{400 + 50} = 1.22 \text{ or } 0.9 \text{ dB}$$

which is not particularly impressive. In view of this it may be more economical to raise the transmitter output power instead. In contrast, had the system background noise been only 40°K, the improvement would have been

$$\frac{40 + 150}{40 + 50} = 2.11 \text{ or } 3.3 \text{ dB}$$

and the investment in a 50°K LNA versus doubling the transmitter power would surely appear in a different light.

In conclusion, the primary purpose of the operating noise temperature concept is that through the use of T_{op} the output SNR can be quickly calculated. This can be seen from the following. Consider a single-response system and use the same assumptions as stated earlier in conjunction with (8-13). Then

$$\text{SNR}_o = \frac{S_o}{N_o} = \frac{S_o/G}{N_o/G} = \frac{S_i}{kT_{op}\,df} \tag{8-16}$$

Thus, knowing the operating noise temperature T_{op} of a system, the important output SNR, SNR_o, can be calculated with ease. Furthermore, as shown in (8-16), kT_{op} is numerically equal to the power per unit bandwidth required of the input signal to render the normalized output SNR equal to unity.

8.3.   AVERAGE NOISE TEMPERATURE

The preceding discussion used either the available output noise power per unit bandwidth, that is, the noise density, or the noise in a very narrow bandwidth df in developing the concepts of T_e and T_{op}. The resultant terms are therefore called the "spot input noise temperature" and the "spot operating noise temperature," respectively, to emphasize that they measure the noise characteristics of a given system at a single frequency, or at most over a bandwidth narrow enough so that there is no variation over the df. Unfortunately, as we move through the operating band of frequencies, the spot parameters will vary from frequency to frequency. The situation is analogous to measuring the gain response of a transducer. We could, conceptually at least, use a very narrow-band instrument, and measure and plot the T_e and T_{op} as functions of frequency. However, in practice we are invariably dealing with finite bandwidths when measuring and characterizing the noise performance of physical transducers. The quantities of interest are therefore the *average* T_e and the *average* T_{op} over a band of frequencies [7]. These are designated $\bar{T}_e$ and $\bar{T}_{op}$, respectively. The original definition of T_e, Eq. (8-4), now becomes

$$\bar{T}_e = \frac{N_{no}}{k\bar{G}B_0} \tag{8-17}$$

where N_{no} is the total available output noise power and $\bar{G}$ is the average available power gain, both over the given bandwidth B_0, such that

$$\bar{G}B_0 = \int_{f_1}^{f_2} G(f)\,df \tag{8-18}$$

As noted in Fig. 8.4, this definition is a general one as far as B_0 is concerned. Theoretically, (8-17) gives the $\bar{T}_e$ for any bandwidth we may
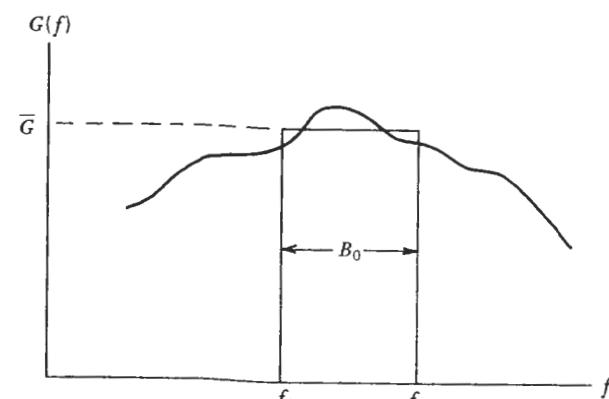

FIGURE 8.4.   Definition of B_0 for average noise temperature.

choose. In practice, the B_0 is usually taken to be the noise bandwidth B, defined in (3-39) and Fig. 3.9.

Similarly, the formal definition of the average operating noise temperature $\overline{T}_{op}$ follows from (8-13) and [8] as

$$\overline{T}_{op} = \frac{N_{oL}}{k\overline{G}_s B_0} \qquad (8\text{-}19)$$

where N_{oL} is the total output noise power (including the load noise) over the bandwidth B_0, delivered to the load, and $\overline{G}_s$ is the average signal gain over B_0. We shall again use the same assumptions as in Section 8.2. Hence

$$\overline{G}_s = \overline{G} \qquad N_{oL} = N_o \qquad (8\text{-}20)$$

and therefore

$$\overline{T}_{op} = \frac{N_o}{k\overline{G}B_0} \qquad (8\text{-}21)$$

In a rigorous sense, therefore, it is nearly always the average $\overline{T}_e$ and $\overline{T}_{op}$ which are measured and quoted in practice. At times the bandwidth B_0 may be narrow enough or the noise temperatures may be such weak functions of frequency that $T_e(f) \approx T_e$ and $T_{op}(f) \approx T_{op}$ can be assumed with negligible error. This, however, is not generally true as can be easily seen from Fig. 8.5, which shows a hypothetical variation of N'_{no} and G with frequency. In
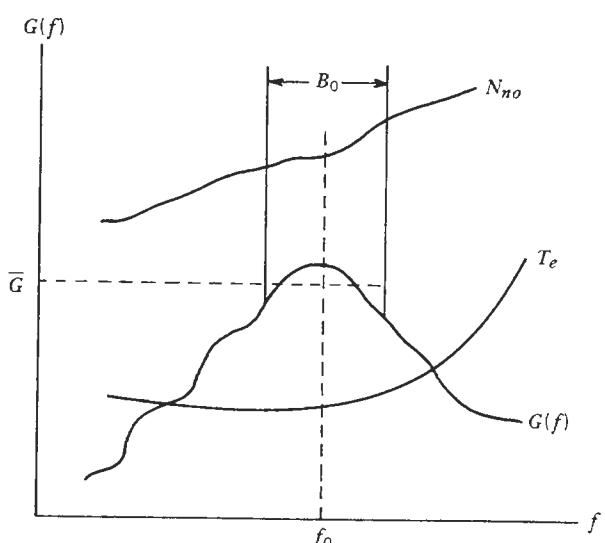
FIGURE 8.5. Dependence of the average noise temperature on bandwidth.

accordance with (8-4), the spot noise temperature T_e at any frequency f_0 between f_1 and f_2 is proportional to the ratio N'_{no}/G. This corresponds to looking through an infinitesimally narrow window df. Clearly, if the window is broadened to B_0, the ratio of the integrated total N_{no} and the average gain $\overline{G}$, which equals $\overline{T}_e$, will deviate from the spot value T_e depending on the bandwidth B_0 used. As f_0 is varied, the resultant plot of $\overline{T}_e$ as a function of frequency will also change as a function of B_0 unless the frequency variations of N_{no} and G are equal.

8.4.   MEASUREMENT OF T_e

Although Chapter 11 will present a detailed discussion of the various techniques and test equipment required to measure noise parameters of arbitrary linear, noisy transducers, we shall give here a brief description of how the T_e is measured. The intent is to show that the input noise temperature, although a fictitious quantity, can nevertheless be easily measured and is therefore a very useful and versatile parameter in the characterization of receivers and electronic components in general.

Consider a transducer, such as an amplifier, which is shown in Fig. 8.6. Let us first connect a noise source of high noise temperature T_h to its input terminals. The total output noise power measured over an incremental bandwidth df is then given by

$$N_h = kT_h G\,df + kT_e G\,df = k\,G\,df(T_e + T_h) \qquad (8\text{-}22)$$

Now replace the hot noise source with one having a very low noise temperature T_c and measure the output noise again:

$$N_c = kG\,df(T_e + T_c) \qquad (8\text{-}23)$$

Define further a parameter Y such that

$$Y = \frac{N_h}{N_c} \gg 1 \qquad (8\text{-}24)$$
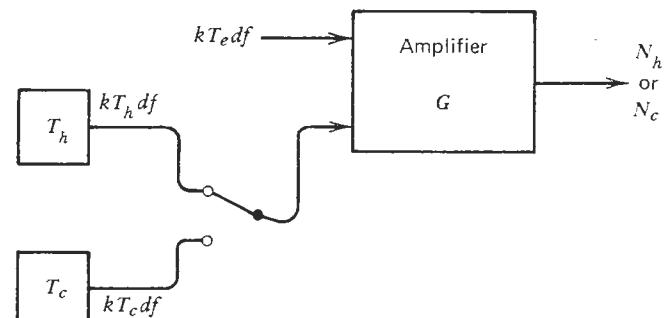
FIGURE 8.6. Measurement of input noise temperature.

and equate (8-22) and (8-23) using (8-24):

$$kG\,df(T_e + T_h) = YkG\,df(T_e + T_c)$$

From this it follows that

$$T_e = \frac{T_h - YT_c}{Y - 1} \tag{8-25}$$

At this point it is essential to recall from Section 8.1 that T_e is not an intrinsic property of the transducer alone, but depends on the internal impedance of the input generator used with the transducer. The measurement of T_e must therefore be carried out with noise sources having the *same* generator impedance as that found in the actual system where the transducer is to be used. Therefore, as broadly applicable as the T_e may appear, its use is nevertheless governed by definite rules which must be observed.

8.5. NETWORKS IN CASCADE

We saw in Section 8.1 that the noise properties of a linear, single-response, two-port transducer could be modeled in terms of an equivalent input noise temperature. If we had a number N of such transducers, amplifiers for example, each characterized by its T_e and G, and if we connected them all in cascade as in Fig. 8.7a, what would be the overall input noise temperature T_{eT}?

(a)

(b)
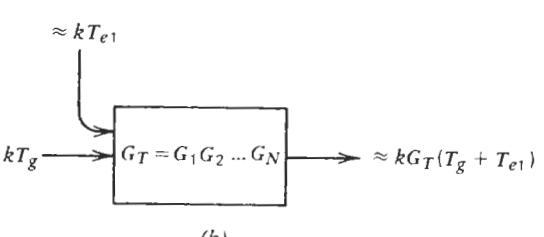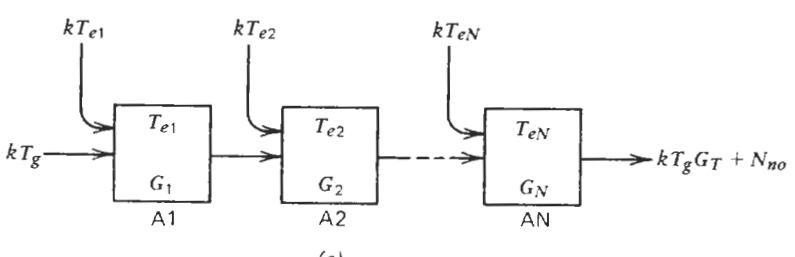
FIGURE 8.7.   Derivation of the cascade formula for noise temperature.

The answer is easily obtained from the definition of T_e, Eq. (8-4). First, assuming that the available power gains of the individual amplifiers, G_1 to G_N, have been properly defined (see Section 7.5), the overall gain G_T can be written as the product $G_T = G_1 G_2 \cdots G_N$ in accordance with (7-22). The excess output noise per unit bandwidth, N'_{no}, must be determined next. Starting from the left of Fig. 8.7a, the individual contributions are summed as follows:

$$N'_{no} = (kT_{e1}G_1G_2 \cdots G_N) + (kT_{e2}G_2 \cdots G_N) + \cdots + (kT_{eN}G_N)$$

(8-26)

Substituting the G_T and (8-26) into (8-4) yields

$$T_{eT} = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1G_2} + \cdots + \frac{T_{eN}}{G_1G_2 \cdots G_{N-1}} \tag{8-27}$$

This result illustrates the very important and well-known property of cascaded amplifiers: If the gain of the first stage is high enough ($G_1 \gg 1$), the noise contributions of succeeding stages, $T_{e2}, T_{e3}, \ldots, T_{eN}$ no longer matter (Fig. 8.7b), unless, of course, some of the gains happen to be losses ($G < 1$) instead. Another way of saying this is that the system noise performance depends primarily on the quality and the gain of the first amplifier in the chain. For this reason the input amplifier of a sensitive receiver system is usually selected to have the lowest possible input noise temperature T_{e1}, and a gain G_1 high enough to minimize all noise contributions from the rest of the system. In practice, where economics and design difficulties must also be considered, the choice may not be as clear cut. One would have to weigh the gain G_1 and the noise T_{e1} against the same parameters of the succeeding stages. For example, the expense of extra gain in $A1$ (Fig. 8.7a) may not be worth the marginal improvement in the suppression of T_{e2} and T_{e3}. Perhaps a better choice would be to use a quieter second stage $A2$ instead, and so on.

It should also not be overlooked that if the source noise T_g is high, costly improvements in individual amplifier stages may not buy much in terms of the final SNR (see Section 8.2).

In conclusion, it should again be emphasized that (8-27) is valid *only* if all T_{ej} were measured or calculated with the ($j - 1$)th stage as the generator. This is a direct consequence of T_e being a function of the source impedance as pointed out in Section 8.1 and Eq. (8-7).

8.6.   INPUT NOISE TEMPERATURE OF A MATCHED ATTENUATOR

Suppose we had a matched pad of attenuation L in series with a transmission line (Fig. 8.8). A signal power S_i is applied to the input at left. Clearly, the output signal S_o is given by S_i/L. As we increase L, we could attentuate S, to
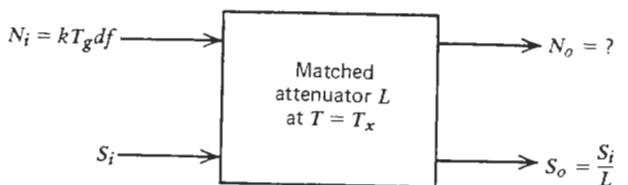
FIGURE 8.8.   Effect of the attenuator on the signal-to-noise ratio.

as low a value as desired. Thus in dB,

$$S_o \, (\text{dB}) = S_i \, (\text{dB}) - L \, (\text{dB})$$

Now assume that noise power N_i from a source resistance R_g at T_g is also entering the same attenuator. Since noise could be considered as just another signal, does it follow that after passing through the attenuator, the output noise power is also reduced to N_i/L? We know intuitively that this could not be true, for it would then be possible to suppress thermal noise to an arbitrarily low level by simply using sufficiently high attenuation.

The correct answer is provided by Fig. 5.7a and Eq. (5-29) where we considered a matched attenuator of loss L at a physical temperature T_x, terminated by an input load R_p at T_p. In our case T_p corresponds to the noise T_g of the input generator. The T_{eff} in (5-29), or more correctly kT_{eff}, represents the noise power N_o' per unit bandwidth at the output of the attenuator. Rewriting (5-29) in terms of power density yields

$$kT_{\text{eff}} = N_o' = \frac{k}{L}\Big[T_g + (L-1)T_x\Big] \tag{8-28}$$

But according to (8-3) the noise properties of the attenuator (it is just another network!) could be expressed in terms of an equivalent input noise temperature T_e as $N_o' = kG(T_g + T_e)$. Equating (8-3) and (8-28) yields

$$\frac{k}{L}\Big[T_g + (L-1)T_x\Big] = kG\big(T_g + T_e\big) \tag{8-29}$$

from which the input noise temperature T_e of a matched attenuator L at a physical temperature T_x is obtained as

$$T_e = (L-1)T_x \tag{8-30}$$

since the available power gain of the attenuator is $G = 1/L$. Fig. 8.9 illustrates this important case and shows the thermal noise model of a matched attenuator. Thus, to reiterate, the total output noise of a matched attenuator is the
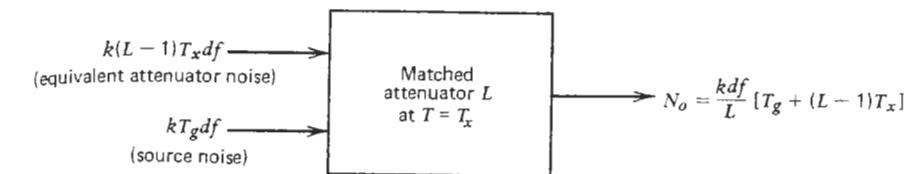
FIGURE 8.9.   Noise model of a matched attenuator.

ing within the attenuator. For a bandwidth df,

$$N_o = \frac{N_i}{L} + N_{no} = \frac{kT_g \, df}{L} + \frac{kT_e \, df}{L}$$

$$= \frac{k \, df}{L}\Big[T_g + (L-1)T_x\Big] \tag{8-31}$$

If (8-31) is arranged in the following manner,

$$N_o = k \, df\left(\frac{T_g - T_x}{L} + T_x\right) \tag{8-32}$$

it is easily seen that for the low loss case, $L \approx 1$, or when the source noise is high $(T_g/L \gg T_x)$,

$$N_o = k \, df\left[\frac{T_g}{L} - \frac{T_x}{L} + T_x\right] \approx k \, df\frac{T_g}{L} = \frac{N_i}{L}$$

and the output noise is indeed close to N_i/L. If on the other hand, L is high, then (8-32) approaches

$$N_o \approx k \, df \, T_x \tag{8-33}$$

and the output noise will be essentially noise from the attenuator.

A simple numerical example would illustrate this (Fig. 8.10). Assume that we are receiving a signal $S_i = -70$ dBm. The receiving antenna temperature is $T_a = 100°\text{K}$. The noise bandwidth of the system is 1 MHz and all noise temperatures are for simplicity assumed to be constant with frequency. This permits us to write $\bar{T} = T$. The input noise power from the antenna is therefore

$$N_i = kT_a B = 1.38 \times 10^{-23} \times 100 \times 10^6$$

$$N_i \longrightarrow$$
$$N_{ni} = N_{no}L \longrightarrow$$

Matched attenuator
$L = 4$ dB
$T_x = 290°$K

$\longrightarrow N_o = -115.3$ dBm

$S_i = -70$ dBm $\longrightarrow$

$\longrightarrow S_o = -74$ dBm

$N_i = -118.6$ dBm (antenna noise)
$N_{ni} = -112.2$ dBm (equivalent attenuator noise)
$S_i = -70$ dBm
$B = 1$ MHz
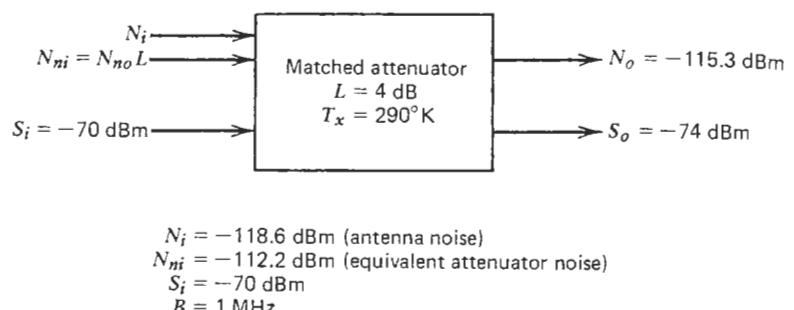
FIGURE 8.10. Sample calculation of SNR after an attenuator.

The input SNR is

$$\text{SNR}_i \text{ (dB)} = -70 - (-118.6) = 48.6 \text{ dB}$$

Both inputs from the antenna pass through an attenuator of $L = 4$ dB which is at $T_x = 290°$K. The input *signal* is attenuated directly and becomes

$$S_o \text{ (dB)} = -70 - 4 = -74 \text{ dBm}$$

However, in accordance with (8-31), the output *noise* power will be the sum of two terms

$$N_o = \frac{N_i}{L} + \frac{k(L-1)T_xB}{L} = 5.5 \times 10^{-16} + 24.1 \times 10^{-16}$$

$$= 29.6 \times 10^{-16} \text{ watts} \quad \text{or} \quad -115.3 \text{ dBm.}$$

This clearly shows that the attenuator contribution predominates and that the output noise N_o is *not* equal to $N_i/L = -122.6$ dBm. Note in passing that since the attenuator is a linear network, it does not matter whether the contributions from the source and the attenuator are added at the input or at the output, that is,

$$N_o = \frac{N_{no}L + N_i}{L} = N_{no} + \frac{N_i}{L}$$

Note also that in Fig. 8.10 the N_i and the equivalent input noise $N_{no}L$, although shown in dBm, must be converted to absolute power (watts or milliwatts) before they can be added.

The total available noise power at the output of an attenuator can thus be represented as in Fig. 8.11. For high values of N_i where the attenuator contribution is negligible, the curve follows N_i/L as it should. Eventually, however, the attenuator noise begins to predominate and the N_o levels off at

$$N_o = kT_x df\left(\frac{L-1}{L}\right) \tag{8-34}$$
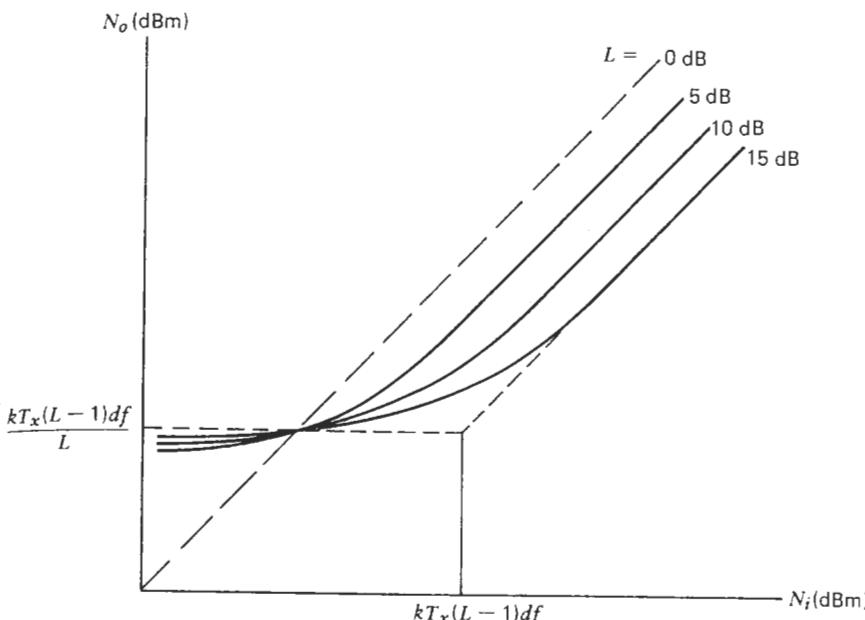
FIGURE 8.11. Noise relationships in an attenuator.

The 3-dB deviation occurs at

$$N_i = kT_x df(L-1) \tag{8-35}$$

An interesting case is $T_g = T_x$, that is, both the source and the attenuator are at the same temperature T. From (8-31),

$$N_o = \frac{k\,df\,T}{L}(1 + L - 1) = k\,df\,T = N_i$$

regardless of L. Under this condition the increased noise from the increasing L [see (8-30)] precisely compensates for the decrease in generator noise as the latter is increasingly attenuated by L.

8.7. EFFECT OF INPUT LOSS ON T_{op}

Equation (8-28) provides a means of calculating the operating noise temperature T_{op} for a practical situation where the generator at a noise temperature T_g is connected to the transducer not directly, but through a loss L. This situation arises whenever the receiver is located some distance from the antenna and the connection must be made through a length of lossy cable as in Fig. 8.12. Since the cable can be considered an attenuator, we no longer see T at the input of
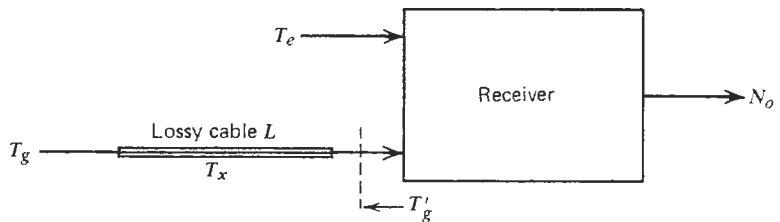
FIGURE 8.12. A receiver with a lossy input cable.

the receiver, but an effective source temperature T_g' instead. The latter is obtained from (8-28) as

$$T_g' = \frac{T_g}{L} + \left(\frac{L-1}{L}\right)T_x \qquad (8\text{-}36)$$

The operating noise temperature, originally $T_{op} = T_g + T_e$, is correspondingly affected and becomes

$$T_{op}' = T_g' + T_e = \frac{T_g}{L} + \left(\frac{L-1}{L}\right)T_x + T_e \qquad (8\text{-}37)$$

which is higher than T_{op} (unless T_x is very low). In a typical situation the antenna may look at a cold sky while the physical temperature of the cable could be high. It is instructive to consider what happens when the loss L is small. Assume therefore $T_x \gg T_g$, let $L = 1 + \varepsilon$, and use these in (8-36). The result is

$$T_g' \approx T_g(1 - \varepsilon) + \varepsilon T_x = T_g + \varepsilon(T_x - T_g) \approx T_g + \varepsilon T_x \qquad (8\text{-}38)$$

Let us consider a numerical example. Take $T_g = 20°K$ and $T_x = 300°K$, and $L = 1.072$ (0.3 dB). From (8-38),

$$T_g' = 20 + (0.072)(300) = 41.5°K$$

Thus the source noise level has been degraded by $10 \log T_g'/T_g = 3.2$ dB by an input line loss of a mere 0.3 dB! If this result is coupled with (8-16), then the adverse effect of L on the SNR_o is easily seen.

8.8. SYSTEM APPLICATIONS

The noise temperature concept, both input and operating, finds its most important use in receiver systems. Here the designer is usually faced with a specified SNR_o which must be realized at the lowest possible cost. Assuming that the designer has taken care of all other threats to signal fidelity such as man-made electrical noises, amplitude and phase nonlinearities, intermodula-

tion distortion, and so on, there remains the basic hurdle of thermal noise which cannot be eliminated, but which can be neutralized to a degree. The designer is therefore vitally interested in such factors as the overall gain and the input noise temperature T_e of the system which determine the ultimate signal and noise levels at the user's end of the receiver. The designer is constantly battling the unavoidable losses in transmission lines connecting the antenna to the receiver, and the equally unavoidable sky (source) noise. Furthermore, the antenna gain G_R and the system bandwidth B_{sys} must be selected with care. Any bandwidth beyond that required to pass the signal will simply admit more noise in accordance with kTB and is therefore undesirable. All these parameters must be carefully weighed and traded off against the cost and complexity of the proposed system.

A number of basic tools for this task were developed in Chapters 3 and 5 and in Sections 8.1 through 8.7. We shall now apply them to the analysis of a typical front end of a receiving system. Figure 8.13 illustrates the layout. The system usually begins with a receiving antenna having a gain G_R and an equivalent sky-noise temperature T_s. At this point it is important to understand the difference between the sky temperature T_s and the antenna temperature T_a. If the beamwidth of the antenna is very narrow and if the sidelobes are far down, then the T_a is given by the average sky temperature T_s seen within the beamwidth of the main lobe. Sometimes, however, the sidelobes are appreciable and they may see the warm earth or warm atmosphere near the horizon even when the main beam is pointed toward a cold portion of the sky. The actual antenna temperature T_a may then be significantly higher than T_s. For example, if the gain in a particular sidelobe [which corresponds to α_j in (5-26)] and the associated noise temperature T_{nj} are high enough, then in accordance with (5-26),

$$T_{eff} = T_a = \alpha_1 T_s + \alpha_j T_{nj}$$

which could mean an unpleasant degradation.
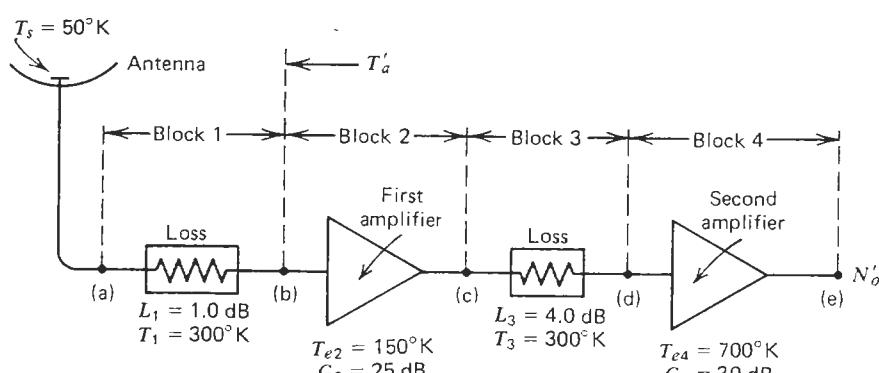
FIGURE 8.13. Parameters of a receiving system.

The losses within the antenna feed, pedestal, and the associated transmission lines are for simplicity lumped together in the attenuator shown as block 1. The first LNA (block 2) has an input noise temperature T_{e2} and an available power gain G_2. A cable link of loss L_3 (block 3) connects the two amplifiers. The second-stage amplifier (block 4) has a noise temperature and gain of T_{e4} and G_4, respectively. The task is to calculate the final noise density N_o' in dBm/Hz at the output of the second amplifier, point (e) in Fig. 8.13. The necessary numerical data are given in Fig. 8.13 except for the noise temperatures of the two cable losses. These are easily calculated from (8-30):

$$T_{e1} = (L_1 - 1)T_1 = (1.259 - 1)300 = 77.7°K \qquad (8-39)$$

and

$$T_{e3} = (L_3 - 1)T_3 = (2.512 - 1)300 = 453.6°K \qquad (8-40)$$

The problem will be solved in four ways to illustrate the different methods. These are:

Cascade formula (8-27)
Walk-through method
Summation method
Pierce's rule

Before getting started, it is not superfluous to recall that the rule given at the end of Section 8.5 must be strictly observed. Otherwise a particular value used for T_{ej} is not applicable and no further conclusions regarding the overall T_{eT} or the final output noise of the cascaded system can be drawn.

Cascade Formula

In this method the combined input noise temperature T_{eT} of the blocks 2, 3, and 4 at point (b) is calculated first, using the familiar cascade formula (8-27). The result is then combined with the equivalent antenna temperature also referred to point (b) in Fig. 8.13. This yields the operating noise temperature T_{op} which now lumps all noise information together at one point, namely, (b). From here the total noise can be referred to the output, point (e), by adding and subtracting (on a dB basis) the gains and losses between (b) and (e). The result is the desired output noise density N_o'.

Thus, applying (8-27) and (8-40) to blocks 2, 3, and 4 in Fig. 8.13:

$$T_{eT} = 150 + \frac{453.6}{316.2} + \frac{700}{316.2}(2.51) = 150 + 1.4 + 5.6 = 157.0°K \qquad (8-41)$$

Now, looking toward the antenna at point (b), the equivalent antenna temperature is obtained from (8-36) as

$$T_a' = \frac{50}{1.26} + \frac{1.26 - 1}{1.26}(300) = 101.4°K \qquad (8-42)$$

Combining the answers from (8-41) and (8-42) yields the operating noise temperature and the equivalent noise density at (b):

$$kT_{op} = 1.38 \times 10^{-23}(157.0 + 101.4) = 1.38 \times 10^{-23} \times 258.4$$

$$= 3.57 \times 10^{-21} \text{ watts/Hz} \quad \text{or} \quad -174.5 \text{ dBm/Hz} \qquad (8-43)$$

Hence from Fig. 8.13, allowing for the various losses and gains,

$$N_o' = -174.5 + 25 - 4 + 30 = -123.5 \text{ dBm/Hz} \qquad (8-44)$$

It should be noted how the use of T_{op} has simplified the calculations. Also, the noise density is best expressed in dBm/Hz rather than watts/Hz because the gains and losses are usually expressed in dB.

The selection of the point (b) as the reference should perhaps be explained. The reason is that (8-36) applies only as long as the corresponding network is reciprocal. We did not use point (c), for example, because there the equivalent antenna noise temperature T_a' is more awkward to calculate (one must take into account the T_{e2} and the G_2 as well). As a result, the reference location for T_{op} is typically chosen to be the input of the first active element in the chain. Here it is still valid to use (8-36) which simplifies the calculations. In fairness, Fig. 8.13 shows a very simple antenna circuit, and is not really representative of a general situation. We shall treat a more complicated case in Fig. 8.15.

Walk-Through Method

In this method one starts with the input noise, in our case the sky noise, and adds to it the effective input noise of the first element (block 1). The result is modified by the gain or loss of the block and thereby referred to its output. To this number the effective input noise of the second block is added, and the result is transferred to the output of the second block, and so on, until we finally arrive at the output where the last number is the desired N_o'. We thus effectively "walk through" the system, accumulating and adding noise contributions as they occur, and modifying each result by appropriate loss or gain encountered along the way (Table 8.1).

The noise contribution of each individual block or component is first calculated in accordance with Fig. 8.14 using the data in Fig. 8.13 and in Eqs.

TABLE 8.1.   Summary of Walk-Through Method (All Numbers in dBm / Hz)

Block	Input A	Input B	Sum	Output
1	−181.6	−179.7	−177.5	−178.5
2	−178.5	−176.8	−174.6	−149.6
3	−149.6	−172.0	−149.6	−153.6
4	−153.6	−170.2	−153.5	−123.5 $= N_o'$

(8-39) and (8-40). Hence

$$kT_s = (k)(50) \quad = 6.90 \times 10^{-19} \text{ mW/Hz} \quad \text{or} \quad -181.6 \text{ dBm/Hz}$$

$$kT_{e1} = (k)(77.7) \quad = 10.72 \times 10^{-19} \text{ mW/Hz} \quad \text{or} \quad -179.7 \text{ dBm/Hz}$$

$$kT_{e2} = (k)(150) \quad = 2.07 \times 10^{-18} \text{ mW/Hz} \quad \text{or} \quad -176.8 \text{ dBm/Hz}$$

$$kT_{e3} = (k)(453.6) = 6.26 \times 10^{-18} \text{ mW/Hz} \quad \text{or} \quad -172.0 \text{ dBm/Hz}$$

$$kT_{e4} = (k)(700) \quad = 9.66 \times 10^{-18} \text{ mW/Hz} \quad \text{or} \quad -170.2 \text{ dBm/Hz}$$

At the input to block 1 we thus have the contributions −181.6 dBm/Hz from the sky and −179.7 dBm/Hz from the antenna losses. When these are added, the result (still at the *input* of the block!) becomes 17.62×10^{-19} mW/Hz or −177.5 dBm/Hz. At the *output* of block 1 this figure becomes −178.5 dBm/Hz which is the input to block 2. Here we combine −178.5 and −176.8 to obtain −174.6 dBm/Hz at the input, or

$$-174.6 + 25 = -149.6 \text{ dBm/Hz}$$

at the output of the block 2 (Table 8.1). The remaining blocks are treated similarly and the final answer at point (e) is again $N_o' = -123.5$ dBm/Hz as we had before in (8-44). It is interesting to observe as one proceeds toward the output of the system how the individual noise contributions become less and less significant. The noise reaching blocks 3 and 4 from previous stages is several orders of magnitude above that contributed by these blocks themselves. This is simply what is observed in using the cascade formula: As one proceeds down the cascaded chain, the noise contributions become less and less important. In view of the high gains in blocks 2 and 4, we could just take the noise level at the output of the block 2, add algebraically the remaining gains and losses, and get the final answer within a tenth of a dB.

The walk-through method is useful if it is desired to study the build-up of noise throughout the system. The effect of each contribution, gain, loss or noise, can be clearly seen. The apparently awkward back-and-forth conversion between dBm and mW can be accomplished quickly on any calculator.
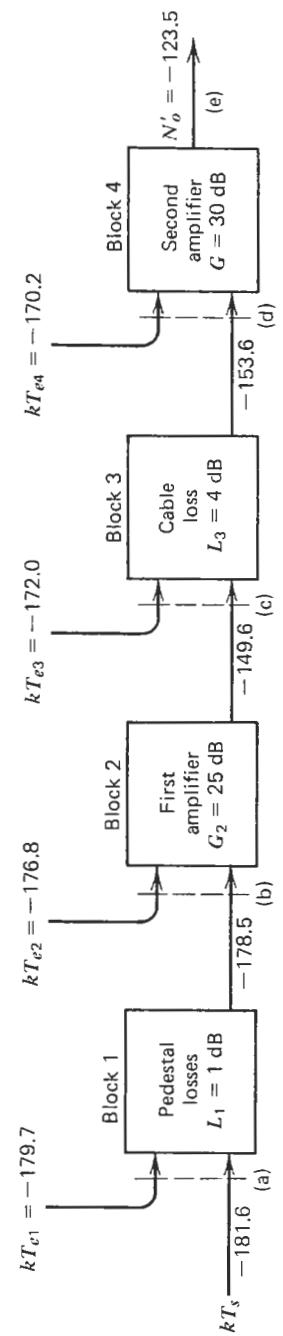
FIGURE 8.14. Calculation of output noise using the walk-through method. (All noise densities expressed in dBm/Hz.)

Summation Method

In this method each noise contribution, kT_s, kT_{e1}, and so on, shown in Fig. 8.14, is *individually* taken through the various gains and losses to the output, that is, to point (e), where all are summed.

Thus the sky contribution kT_s and the contribution from the first block become

$$-181.6 - 1 + 25 - 4 + 30 = -131.6 \text{ dBm/Hz}$$

and

$$-179.7 - 1 + 25 - 4 + 30 = -129.7 \text{ dBm/Hz}$$

respectively. Block 2 yields

$$-176.8 + 25 - 4 + 30 = -125.8 \text{ dBm/Hz, etc.}$$

Table 8.2 summarizes the remaining steps and gives the final result. As expected, the answer is once again $N_o' = -123.5$ dBm/Hz.

The summation method is useful in that one can easily see the relative importance of each contribution. Thus the major contributor is clearly the first amplifier in block 2. In view of its high gain G_2, the effects of the cable loss L_3 and the second-stage amplifier noise T_{e4} are negligible.

Section 8.7 discussed the significant effect that any input loss would have on system performance. Let us explore this using Table 8.2. If we could somehow reduce the line loss L_1 to 0 dB (i.e., delete the block 1 entry altogether in Table 8.2), then the numbers would change as listed in the second column of Table 8.2. The improvement in N_o', $(-123.5) - (-124.4) = 0.9$ dB, is approximately equal to the reduction in loss.

Now, Table 8.2 used a first-stage amplifier of $T_{e2} = 150°$K. Had a lower-noise, say 60°K, amplifier been used, the numbers would have been as in Table

TABLE 8.2. Results of the Summation Method (First Amplifier $T_{e2} = 150°$K)

Noise Contributor	Noise Density at Output (dBm/Hz)	
	$L_1 = 1$ dB	$L_1 = 0$ dB
Sky	−131.6	−130.6
Block 1	−129.7	—
Block 2	−125.8	−125.8
Block 3	−146.0	−146.0
Block 4	−140.2	−140.2
Total N_o' at output	−123.5	−124.4

TABLE 8.3. Results of the Summation Method (First Amplifier $T_{e2} = 60°$K)

Noise Contributor	Noise Density at Output (dBm/Hz)	
	$L_1 = 1$ dB	$L_1 = 0$ dB
Sky	−131.6	−130.6
Block 1	−129.7	—
Block 2	−129.8	−129.8
Block 3	−146.0	−146.0
Block 4	−140.2	−140.2
Total N_o' at output	−125.3	−126.9

8.3. The improvement in the final N_o', $(-125.3) - (-126.9) = 1.6$ dB, is now *more* than the decrease in L_1. It should be apparent why in low-noise receiver installations every effort is made to decrease the front end losses. The first amplifier is usually mounted right at the antenna feed to minimize L_1. If instead of eliminating L_1 we had kept it at 1 dB, and had sought the same improvement in N_o' through a decrease in T_{e2} from 60°K to some lower number, the latter would have to go down to 9.1°K—a very costly proposition!

Pierce's Rule

Typically the antenna circuit is much more complicated than the simple configuration shown in Fig. 8.13. Figure 8.15 illustrates a more representative situation. The sky temperature is still 50°K (block 1), but its contribution comes through a warmer atmosphere at a physical temperature $T_2 = 200°$K and a loss of 0.6 dB (block 2). The antenna has a radome at $T_3 = 320°$K and a loss of $L_3 = 1.0$ dB (blocks 3 and 3'). The coupling to the antenna is not perfect; that is, only 60% of the input signal is received through the main lobe. The sidelobes, which are pointed toward the ground (block 1', at 300°K), account for 40% of the received signal (block 4). Finally, there are the losses in the antenna pedestal (block 5) which are $L_5 = 1.0$ dB at a physical temperature $T_5 = 315°$K.

Equation (8-36) is obviously not adequate to calculate T_a' for this case. The problem can be solved using the general form of Pierce's rule (Section 5.2). For this purpose the signal flow in the chain in Fig. 8.15 is reversed and a unit of power (1 watt) is applied to the input, point (a). Next, the fractions of input power dissipated in each block are calculated, resulting directly in the values of α_j required by (5-26). This calculation is straightforward as shown in Fig. 8.15 and tabulated in Table 8.4. As a check, we note that the total of all α_j is unity as it should be. The final result is the effective noise temperature looking into the antenna terminals, $T_a' = 222.1°$K. This number correctly accounts for all individual noise contributions, and constitutes the weighted T_a' analogous to that calculated in (8-42) for the simpler case.
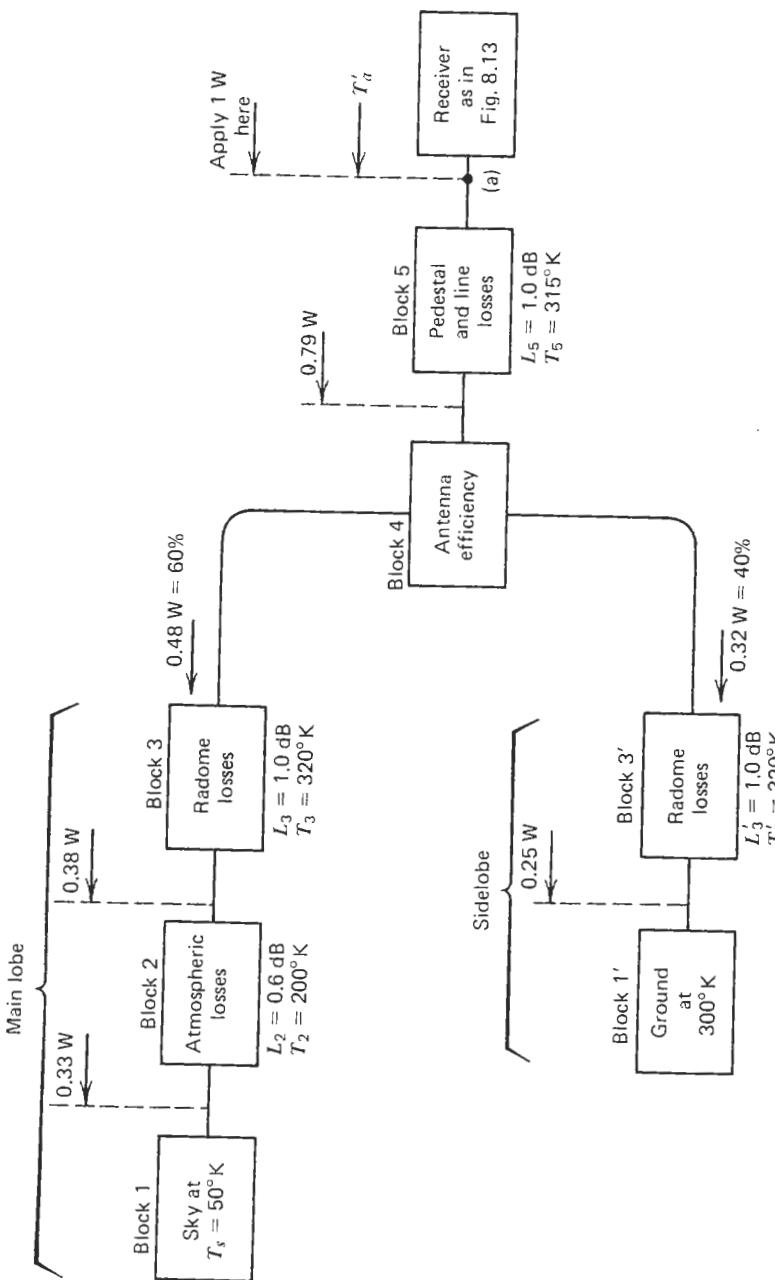
FIGURE 8.15. Calculation of a more complex receiving system.

TABLE 8.4. Calculation of Effective Antenna Temperature Using Pierce's Rule

Block	$P_{\text{dissipated}} = \alpha_j$	T_j (°K)	$\alpha_j T_j$
5	0.21	315	66.2
3	0.10	320	32.0
3'	0.07	320	22.4
2	0.05	200	10.0
1	0.33	50	16.5
1'	0.25	300	75.0
$T_{\text{eff}} = T_a' = \Sigma(\alpha_j T_j) =$			222.1°K

The rest of the calculation proceeds as before, that is, the T_a' could either be combined with the T_{eT} as in (8-43), or it could be taken through the system as in the walk-through method, or the summation method could be applied to it.

8.9. THE G/T RATIO

The widely used ratio G/T arises in link budget calculations of microwave and satellite communication systems where it is an important and concise measure of the overall SNR. Consider the situation shown in Fig. 8.16 which defines a typical satellite/ground station interface. A satellite transmitter having an output power P_T is beamed toward a ground station. The carrier power received at the input to the receiver is given by

$$C = \frac{P_T G_T A_R}{4\pi R^2 L_a} = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2 L_a} \tag{8-45}$$

FIGURE 8.16

where P_T is the output power of the satellite transmitter; G_T the gain of the transmitting antenna in the satellite (over isotropic); A_R the effective area of the receiving antenna $= G_R(\lambda^2/4\pi)$; G_R the gain of the receiving antenna; λ the wavelength of the transmitted signal; R the distance from the satellite to the ground station; and L_a the propagation losses in excess of free space path loss (atmospheric losses, rain attenuation, etc.).

The *noise* density at the input of the first amplifier was shown in Section 8.2 to be $N' = kT_{op}$, and the noise is assumed to have a flat spectral density. The carrier-to-noise density ratio then becomes

$$\frac{C}{N'} = \frac{P_T G_T \lambda^2}{(4\pi R)^2 L_a k}\left(\frac{G_R}{T_{op}}\right) = \frac{P_T G_T}{(4\pi R/\lambda)^2 k L_a}\left(\frac{G_R}{T_{op}}\right) \qquad (8\text{-}46)$$

Equation (8-46) is usually expressed in dB. Thus

$$\left(\frac{C}{N'}\right)_{dB} = 10\log(P_T G_T) - 20\log\left(\frac{4\pi R}{\lambda}\right) + 10\log\left(\frac{G_R}{T_{op}}\right) - 10\log(kL_a)$$

$$(8\text{-}47)$$

where $10\log(P_T G_T)$ is the effective isotropic radiated power (EIRP) of the transmitter; $20\log(4\pi R/\lambda)$ the free space path loss; and $10\log(G_R/T_{op})$ the G/T ratio of the receiving station.

The term G/T, when expressed in dB, thus serves to simplify the calculations on receiving systems and their SNRs. It is a figure of merit which combines the quality of the antenna and the effect of the total system noise. As indicated above, the G/T is usually expressed in dB/°K. Typical values for satellite ground terminals are shown in Table 8.5.

TABLE 8.5. G/T Ratios for Common Satellite Ground Terminals

System	Diameter of Antenna (ft)	Gain (dB)	Frequency (GHz)	G/T (dB/°K)
Intelsat A	95–105	$\geqslant 57$	3.7–4.2	+40.7
B	33–39	$\geqslant 49$		+31.7
C	56–59	≈ 64	11.0–11.7	+39
DSCS	60	≈ 57	7.25–7.75	+37
	40	≈ 53		+34
	20	≈ 51		+27
Canadian Anik, HR	97	62^a	3.7–4.2	+37
NTV	33	52^a		+28

^{a}Measured at 6 GHz.

The G/T is calculated with both quantities G and T_{op} referred to the same reference plane. Typically, since T_{op} is measured at the input of the first low-noise amplifier [point (b) in Fig. 8.13], the G_R must also be referred there. This means that G_R must be decreased by the amount of losses in the antenna and the input transmission line unless such losses were already included in the initial determination of G_R.

The reader is probably wondering if there is anything unique about the stated selection of the reference point, or, to phrase it differently, does the G/T depend on the location of the reference point? The answer is no. The G/T remains invariant for linear circuits. The G and T_{op} taken singly would obviously depend on the location where they are measured. However, the ratio does not. This could be easily shown, but the details are left as an exercise.

8.10. SUMMARY

For single-response transducers:

(a)   Input noise temperature $T_e = N'_{no}/kG$ defined on the basis of *available* output power.

(b)   Operating noise temperature $T_{op} = N'_{oL}/kG_s$ defined on the basis of *delivered* output power.

(c)   The average $\bar{T}_e$ and $\bar{T}_{op}$ are defined using *average* gain over the given finite bandwidth.

(d)   Neither T_e nor $\bar{T}_e$ include load noise generated at the output frequency.

(e) T_{op} and $\bar{T}_{op}$ *do* include possible load noise reflected by the output impedance of the transducer.

(f)   Input signal per unit bandwidth for unity SNR_o: $S_i = kT_{op}$.

(g)   For several transducers in cascade:

$$T_{eT} = T_{e1} + \frac{T_{e2}}{G_1} + \cdots + \frac{T_{eN}}{G_1 G_2 \cdots G_{N-1}}$$

This relationship holds *only* if the T_{ej} is measured with a generator impedance equal to the output impedance of the $(j-1)$th stage.

(h) T_e and $\bar{T}_e$ are functions of the generator (source) resistance.

(i)   For a matched attenuator:

(1)   Input noise temperature: $T_e = (L-1)T_x$.

(2)   Output noise temperature: $T_n = (1/L)[T_g + (L-1)T_x]$.

(j)   The G/T ratio is the ratio of the antenna gain to the operating noise temperature, both referred to the same reference plane.

(k)   For linear systems the G/T is invariant to the location of the reference plane.

(l)   Noise-temperature ratio: $kT_nB/kT_0B = T_n/290$.

(m)   Excess-noise ratio: $(T_n/290) - 1$.

REFERENCES

1. "IRE Standards on Electron Tubes: Definition of Terms, 1962," *Proc. IEEE*, vol. 51, pp. 434–435, March 1963.

2. *Webster's Seventh New Collegiate Dictionary*, G. C. Merriam Co., Springfield, 1967.

3. "IRE Standards," ref. [1].

4. Ibid.

5. H. A. Haus, chairman, et al., "Description of the Noise Performance of Amplifiers and Receiving Systems," *Proc. IEEE*, vol. 51, pp. 436–442, March 1963.

6. Ibid.

7. "IRE Standards," ref. [1].

8. Ibid.

9 | NOISE FACTOR AND NOISE FIGURE

As shown in Chapter 8, the equivalent noise temperature is a very useful measure of thermal noise and can be applied to a variety of problems. In particular, the input noise temperature T_e allows us to characterize the noisiness of a given transducer, provided, of course, that the dependence of T_e on the source impedance is properly understood and is taken into account. If, in addition, the source noise N_i, the available power gain G of the transducer, and the noise bandwidth B of the system are also known, we could calculate the available output noise power and hence the output SNR. But this is precisely the drawback of T_e. Taken as a single number, it is not a direct measure of the effect that the transducer would have on signals passing through it. Further calculations are necessary before this information becomes available. We know that the noise originating within the transducer, an amplifier, for example, sets the lower limit for signal detection, and T_e is clearly a measure of that noise. Yet there is no simple yardstick for comparing the output conditions of an amplifier, that is, both signal and noise outputs, to conditions at the input of the device.

The problem could be approached several ways. We could, for example, specify the so-called minimum detectable signal which is defined as that input signal in dBm (or milliwatts) which yields an output signal just discernible above the noise when viewed on a spectrum analyzer. Such a measurement is, unfortunately, quite dependent on the bandwidth of the measuring system, and, even worse, is quite subjective. Results obtained by different observers

could show a spread of several dB. Another measure of amplifier quality would be the drop in output noise level as the input terminals are first connected to the source and then short-circuited (thus eliminating the source noise). This method, too, fails to answer our original question.

9.1. DEFINITIONS AND PROPERTIES

The early pioneers in this area, E. W. Herold and D. O. North of RCA, and H. T. Friis of Bell Telephone Laboratories made an extensive study of the problem in the 1940s and 1950s [1–4]. By this time the operating frequencies of receivers had reached UHF and microwave regions where the usual atmospheric and man-made noise was negligible compared to thermal noise originating in the receiver itself. A simple, yet unambiguous method of characterizing the noise properties of such receivers was therefore needed. After extensive deliberation and review, the concept of noise factor was developed and formalized by the IRE in 1952 [5]. A reworded definition was published in 1957 [6]. According to this definition, the noise factor of a transducer at a specified input frequency is the ratio of:

(a) the *total* noise power per unit bandwidth at a corresponding output frequency available at the output port when the noise temperature of its input termination is standard (290°K) at all frequencies, to

(b) that portion of (a) engendered at the input frequency by the input termination at the standard noise temperature of 290°K.

Before proceeding with the mathematical formulation of this definition, let it be emphasized that the name, noise figure or noise factor, was deliberately left to the choice of the user because both terms had become entrenched in practical usage. Tradition has it that in most cases the term noise factor is reserved for the numerical ratio, and noise figure denotes the decibel version of it, that is,

$$\text{Noise Figure } NF = 10\log(\text{noise factor } F) \tag{9-1}$$

The rule is, however, not always followed, and the terms tend to be used interchangeably. There is no unanimity in the symbols for F and NF either. At times noise figure may even be denoted by $F_{\log}$ or F_{dB}, or just plain F. Usually, it is clear from the discussion what is meant and no confusion should arise. Nevertheless, the user should be aware of the ambiguity found in the literature. In this book we shall use the convention as illustrated in (9-1), and all references and symbols will be in accordance with it.

Returning to the IRE definition, the noise factor can be written in mathematical terms as (Fig. 9.1a)

$$F = \frac{N_o'}{N_1'G} = \frac{N_o'}{kT_0G} = \frac{N_o'}{290kG} \tag{9-2}$$
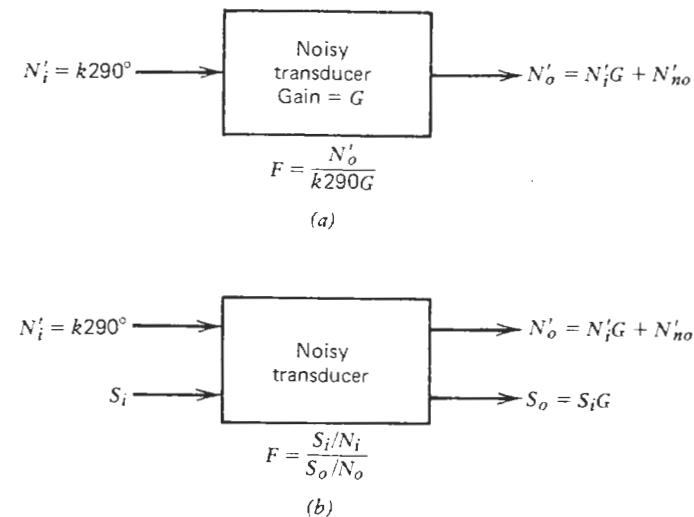
FIGURE 9.1. Definitions of noise factor.

where T_0 designates the standard reference temperature 290°K, and N_o' and G are the total available output noise density (noise per unit bandwidth) and the available power gain, respectively. Note that (9-2) corresponds to the spot noise factor in analogy to the spot noise temperature in (8-4). The average noise factor $\overline{F}$, which relates to a finite bandwidth, will be discussed in Section 9.8.

The first very important observation concerning the noise factor is that it must be based on a reference temperature. This reference temperature is by definition taken as 290°K. The noise factor is thus not an absolute parameter in the sense of the noise temperature, which was derived directly from the observed noise power in accordance with the kT relationship. Instead, the noise factor F as defined is meaningful only as long as everybody uses the same reference. Unless this rule is followed, a given amplifier or transducer could be made to have any value of F through a mere change in T_0—clearly a meaningless situation. To appreciate this even more, let us consider the other definition of noise factor, originally proposed by H. T. Friis [1], who used the term noise figure. The wording is modified slightly from the original to normalize all noise powers to the unit bandwidth as in (9-2). Thus the noise factor of a transducer is defined as the ratio of:

(a) the available SNR at the signal generator terminals per unit bandwidth when the temperature of the input termination (generator or source) is 290°K and the bandwidth is limited by the transducer, to

(b) the available SNR per unit bandwidth at the output terminals of the transducer.

In accordance with Fig. 9.1b, this can be expressed as

$$F = \frac{S_i/N_i'}{S_o/N_o'} = \frac{N_o'}{(S_o/S_i)N_i'} = \frac{N_o'}{kT_0G} \qquad (9\text{-}3)$$

Friis' definition is thus equivalent to the IRE definition provided the receiver is linear, that is, the power gain is the same for signal and noise. Equation (9-3) illustrates rather clearly the dependence of F on a specific available input noise N_i, that is, on a specific noise temperature T_0. For this reason the above ratio cannot officially be called the noise factor unless 290°K is used for T_0.

The noise factor is not a measure of excellence of the output SNR, but merely a measure of degradation in the SNR as the signal and noise pass through the given transducer. An ideal transducer would add no noise of its own and the noise factor would therefore be unity (0 dB).

Several other properties of noise factor should be noted. First, F is a function of the generator (source) impedance Z_g. This dependence, as in the case of T_e in Section 8.1, is first a consequence of G being included in (9-2) and (9-3), which is a function of Z_g (see Section 7.5). Second, the impedance match at the input of the transducer may affect the *internal* noise originating within the transducer quite differently from the way the *source* noise is affected as the latter passes through the transducer. One can see this from (9-5): The excess noise N_{no}' at the output is the sum total of many individual contributions from within the transducer. Clearly, since some of these contributions also emerge at the input terminals and are reflected back by the generator impedance, they become dependent on Z_g, hence rendering the N_o' dependent on Z_g also. It follows that if a given amplifier, for example, is to be used with a generator of internal impedance Z_g, its noise factor F must be based on and be measured with the same Z_g. Otherwise the value of F, obtained with a different Z_g, is not applicable and cannot be used in system calculations. As it happens, it is not possible in general to infer the new noise factor F' corresponding to a new Z_g' from the knowledge of F and Z_g alone. We shall discuss the dependence of F on source impedance in greater detail in Section 9.9. The same section will also treat the subject of minimizing F through proper input matching.

Second, F in its basic definition is independent of the bandwidth of the transducer and of the magnitude of the gain. This can be seen from (9-3) since bandwidth and gain are implicitly included in the numerator. The effect therefore cancels out [see Eq. (9-5)]. We must note, however, that this statement is true only as long as both N_o' and G are constant over the incremental bandwidth over which they are measured. In (9-2) we were dealing with noise density, that is, a single- (or spot-) frequency case, and the problem did not arise. The case of a finite bandwidth and the concept of average noise factor $\bar{F}$ will be discussed in Section 9.8. It will then be appreciated that unless the

excess available noise power density N_{no}' of a physical transducer has the same frequency dependence over the given finite bandwidth as that of G, the measured average noise factor could be a function of bandwidth.

The above statement concerning the independence of F from G may bother some readers because they probably already know (Section 9.4) that the noise factor of a matched attenuator L is very much dependent on its "gain", that is, on $1/L$. The explanation is that in the case of an active transducer (amplifier, etc.) the degradation in the SNR is due to noise mechanisms not directly related to the gain of the device. The noise contribution of an attenuator is a direct consequence of the resistive loss, and the dependence of F on L is therefore not surprising.

Third, in anticipation of Chapter 10 where multi-response transducers are discussed, it is important to note now that for heterodyne systems the denominator of (9-2) includes *only* that input termination noise which appears at the output via the principal-frequency transformation [7]. For example, a down-converting mixer converts source noise into the IF output from at least *two* channels—the desired signal $f_s = f_{LO} + f_{IF}$ *and* the image $f_i = f_{LO} - f_{IF}$. By the definition stated above, *the latter noise is specifically excluded* unless the legitimate signal occupies both channels, as would be true for broadband noise in radiometry applications. The numerator of (9-2), on the other hand, does include the total noise observed at the output—signal channel(s), image channel(s), and noise originating in the transducer itself. A thorough understanding of this is mandatory if future confusion regarding single- and double-sideband noise factors is to be avoided. Many hours of engineering effort have been lost arguing and agonizing over this subject. It could all have been avoided if only the IRE definition of noise factor had been correctly understood and followed from the start!

9.2. RELATIONSHIP TO T_e AND THE CASCADE FORMULA

Let us now examine (9-2) and (9-3) more closely. The total available output noise density N_o' is made up of two components as we saw in connection with (8-3)—noise generated internally and noise from the input generator termination. Thus

$$N_o' = N_i'G + N_{no}' \qquad (9\text{-}4)$$

where N_i' is the available noise density at the input and N_{no}' is the excess-noise density at the output. Substituting (9-4) into (9-2) yields

$$F = \frac{N_i'G + N_{ni}'G}{N_i'G} = 1 + \frac{N_{ni}'}{N_i'} = 1 + \frac{N_{ni}'}{kT_0} \qquad (9\text{-}5)$$

where N_{ni}' is the excess-noise density N_{no}', referred to the *input* terminals of the transducer. We saw previously in (8-3) how the equivalent input noise could be expressed in terms of a fictitious input noise temperature T_e:

$$\frac{N_{no}'}{G} = N_{ni}' = \frac{kGT_e}{G} = kT_e \tag{9-6}$$

Combining (9-6) and (9-5) yields

$$F = 1 + \frac{T_e}{T_0} \tag{9-7}$$

and, consequently,

$$T_e = (F - 1)T_0 \tag{9-8}$$

Equations (9-7) and (9-8) give the desired relationships between the noise factor and the input noise temperature.

Another interesting and useful relationship can be derived from (9-7). If we rewrite it as

$$F = \frac{T_0 + T_e}{T_0} \tag{9-9}$$

and then compare the numerator to (8-15), the noise factor is seen to be closely related to the operating noise temperature when $T_g = T_0$, that is,

$$F = \frac{T_{op}}{T_0} \tag{9-10}$$

for this special case.

If (9-8) is combined with (8-27), we obtain the well-known expression for the overall noise factor F_T of N cascaded stages:

$$F_T = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}} \tag{9-11}$$

Again, it must be emphasized that as in the case of (8-27), all F_j must be evaluated using the output impedance of the $(j-1)$th stage as the generator impedance. This was explained in Section 9.1, but needs to be mentioned again. If the rule is violated, then (9-11) does *not* hold because the F_j would be inconsistent with interstage impedances existing in the cascaded system. One additional warning: Eq. (9-11) holds only for single-response transducers. We shall see in Chapter 10 how the cascade formula is affected by the presence of additional responses.

9.3. NOISE MODEL IN TERMS OF F

Figure 8.2 illustrates the noise model of a linear, two-port transducer in terms of the input noise temperature T_e. A similar noise model can be easily derived in terms of the noise factor F. It only requires the use of (9-8), and we can immediately draw the new model as in Fig. 9.2. The total output noise now becomes

$$N_o = kT_g G\, df + (F - 1)kT_0 G\, df = kG\, df \left[T_g + (F - 1)T_0 \right] \tag{9-12}$$

or using (8-9),

$$N_o = kGT_0\, df\, (t + F - 1) \tag{9-13}$$

Equation (9-13) also shows that as long as the source noise temperature in a system is equal to the standard temperature T_0, the change in output noise power would be given by the ratio of the noise factors:

$$\frac{N_{o2}}{N_{o1}} = \frac{F_2 kT_0 G\, df}{F_1 kT_0 G\, df} = \frac{F_2}{F_1} \tag{9-14}$$

If the generator temperature is other than T_0, such as an antenna having a noise temperature $T_a \neq T_0$, then

$$\frac{N_{o2}}{N_{o1}} = \frac{t_2 - 1 + F_2}{t_1 - 1 + F_1} \tag{9-15}$$

where t is defined as in (8-9) and (9-13).

Equation (9-15) again demonstrates the importance of background noise when attempting to upgrade a receiving system. Clearly, if the source noise is high, that is, t is large, a decrease in F may not change the SNR_o substantially, and the marginal improvement in performance may not be worth the added cost of a lower-noise receiver. Conversely, if the source noise is low, t will be less than unity and the system improvement will be greater than the ratio F_2/F_1 would suggest.

$(F-1)kT_0df \longrightarrow$ [Linear, noise-free transducer Gain $= G$] $\longrightarrow N_o = kG\,df[(F-1)T_0 + T_g]$

$kT_g df \longrightarrow$ $= kGT_0 df(t + F - 1)$
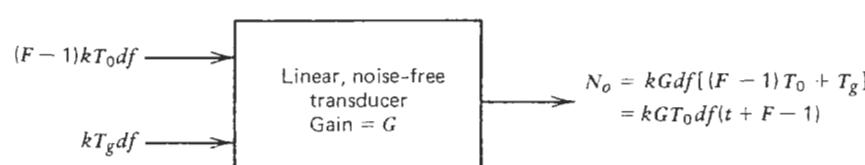
FIGURE 9.2

When analyzing the noise properties of either individual transducers or entire systems, one may wonder when to use F and when to use T_e. Obviously, both are correct, but T_e is easier to use when the source temperature is other than 290°K. Also, if the overall operating noise temperature or the G/T ratio are of interest, T_e would be the natural choice. On the other hand, when $T_g = T_0$, the use of the noise factor will simplify calculations, particularly when dealing with passive attenuators or line losses as we shall see shortly. Another criterion as to which parameter to use is the noisiness of the transducer itself. At high noise levels T_e will quickly reach awkwardly high numbers. For example, a 10-dB noise figure corresponds to a $T_e = 2610$°K. Hence in this case the noise factor is easier to handle. Conversely, in the cases of very low noise the NF becomes a fraction of a dB. Here the noise temperature is a more convenient measure, 50°K being equal to a $NF = 0.69$ dB, for example.

9.4.   APPLICATION OF THE NOISE FACTOR

In this section we shall illustrate the application of the noise factor concept to several single-response components and circuits used in communication systems. Equation (9-2) will be our starting point. For simplicity, noise density will be used, that is, all calculations will be on a unit bandwidth basis. There is no loss of generality because the final result, F, does not depend on bandwidth, and intermediate results, such as the output noise power density, can be easily converted to total noise through the use of the applicable noise bandwidth B.

The first step in noise-factor calculations is to evaluate the available output noise N_o' in terms of pertinent circuit parameters. This yields the numerator of (9-2). The available power gain G must be calculated next. The necessary information is now complete and the noise factor F follows directly from (9-2).

Shunt Resistor

Let us begin by calculating the noise factor of a shunt resistor R at temperature T_0, shown in Fig. 9.3a [8]. The available output noise N_o' at terminals $a–a'$ is contributed by two sources, R_g and R. The two contributions are uncorrelated and can be added on a mean-square basis. We could, of course, begin by assigning a noise generator, either $\overline{e_n^2}$ or $\overline{i_n^2}$, to each resistor, then summing their effects at $a–a'$. In this very simple example it is easier to combine R_g and R in parallel since they are at the same temperature. The total resistance seen at $a–a'$ is

$$R_T = \frac{RR_g}{R + R_g} \qquad (9\text{-}16)$$
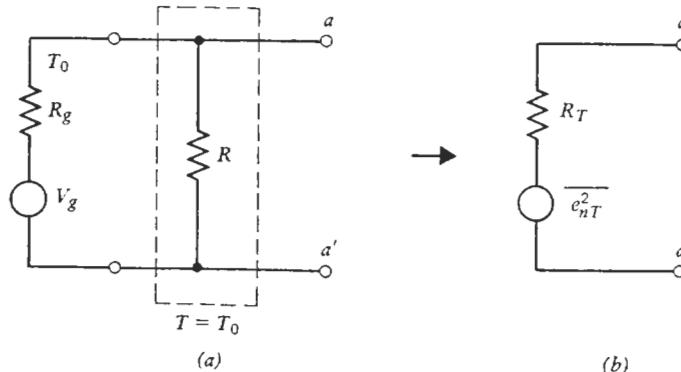
FIGURE 9.3.   Noise factor of a shunt resistor.

The noise model then becomes as shown in Fig. 9.3b. The available output power at $a–a'$ is

$$N_o' = \frac{\overline{e_{nT}^2}}{4R_T} = \frac{4kT_0R_T}{4R_T} = kT_0 \qquad (9\text{-}17)$$

as expected. The available power G of a shunt resistor was calculated earlier in (7-29) as $G = R/(R_g + R)$. Substituting (7-29) and (9-17) into (9-2) yields

$$F = \frac{kT_0(R_g + R)}{kT_0 R} = 1 + \frac{R_g}{R} \qquad (9\text{-}18)$$
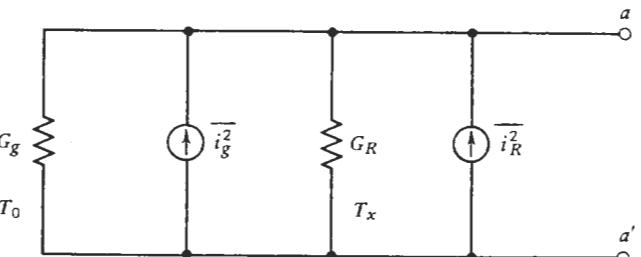
This expression reduces to unity, as it should when R approaches an open circuit. The "transducer" then becomes simply a noise-free straight-through connection.

Now consider the physical temperature of the resistor to be $T_x \neq T_0$. The denominator of (9-2) remains unchanged, but we must account for the change in noise output at $a–a'$, which is due to unequal temperatures of R_g and R. The noise model changes to that shown in Fig. 9.4. We can no longer combine the two resistors on account of unequal temperatures. Thus summing the mean-square contributions at $a–a'$ individually yields

$$N_o' = \frac{\overline{i_{nT}^2}}{4G_T} = \frac{4k(T_0G_g + T_xG_R)}{4(G_g + G_R)} \qquad (9\text{-}19)$$

where $G_R = 1/R$. Therefore

$$F = \frac{N_o'}{kT_0G} = \frac{(T_0G_g + T_xG_R)(R_g + R)}{T_0(G_g + G_R)R} = 1 + \left(\frac{T_x}{T_0}\right)\left(\frac{R_g}{R}\right) \qquad (9\text{-}20)$$

FIGURE 9.4.   Noise factor of a resistor at $T_x \neq T_0$.

Matched Attenuator

Consider a matched attenuator of loss L at a physical temperature T_x as in Figs. 8.7 and 8.8. The total available noise power density at output is given by (8-31) as

$$N_o' = \frac{k}{L}\big[T_0 + (L-1)T_x\big] \qquad (9\text{-}21)$$

where T_g was set equal to T_0 in accordance with the definition of the noise factor. The gain of the attenuator is simply $1/L$. Combining everything in (9-2), we obtain

$$F = 1 + (L-1)\left(\frac{T_x}{T_0}\right) \qquad (9\text{-}22)$$

This is a very handy relationship because in many cases the physical temperature T_x is equal or close to T_0. If this is true, then

$$F = L \qquad (9\text{-}23)$$

Hence the familiar statement that the noise figure of a matched attenuator is equal to its loss in dB. Note, however, that *this holds only* for $T_x = T_0$. If the physical temperature of a 10-dB pad varied from 100°K to 400°K, for example, the noise figure would vary from 6.1 dB to 11.0 dB—certainly not a trivial change.

Now consider a matched attenuator L ahead of an amplifier (Fig. 9.5). Assume that there is an impedance match also between the attenuator and the amplifier. Assume further that $T_x = T_0$. What is the overall noise factor?

The easiest method to calculate this is obviously the cascade equation (9-11) combined with (9-23):

$$F_T = L + \frac{F_a - 1}{1/L} = L + L(F_a - 1) = LF_a \qquad (9\text{-}24)$$
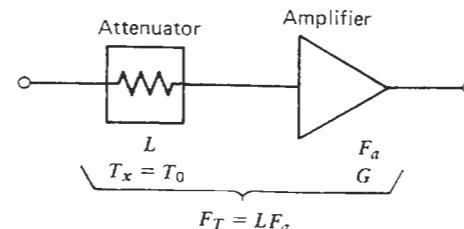
FIGURE 9.5.   Noise factor of an amplifier preceded by an attenuator.
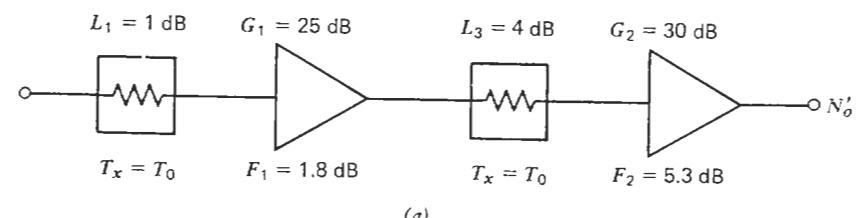
The total noise figure therefore is

$$NF = L \text{ (dB)} + NF_a \qquad (9\text{-}25)$$
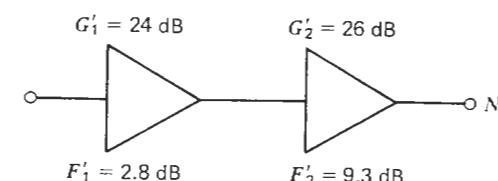
where F_a and NF_a pertain to the amplifier.

This shows why the noise factor is the preferred method of calculating the noise performance of composite circuits containing attenuators and amplifiers. If an amplifier of the noise figure NF is preceded by a lossy input transmission line or some other lossy component, the total noise figure can be written simply as the sum of NF and the loss L in dB in the matched system, *provided* that the temperature conditions $T_x = T_0$ is met. In comparison, the input noise temperature of an attenuator, (8-30), and the cascade formula (8-27) do not lead to such a convenient expression.

In practical situations, such as shown in Fig. 9.6a, the circuit can be quickly simplified as illustrated in Fig. 9.6b. The two amplifiers have noise figures of

(a)

(b)

FIGURE 9.6.   The combining of amplifier and attenuator parameters.

1.8 and 5.3 dB, respectively as can be verified by applying (9-7) and (9-1). When these numbers are combined with the two losses in accordance with (9-25), we are left with two amplifiers, with their noise figures and gains appropriately modified. The gains are decreased by the amount of the particular front-end loss, and the combined noise figure from (9-25) takes the place of the earlier amplifier noise figure. This procedure of combining the losses with the amplifier data is usually much quicker than a straightforward application of the cascade equation. However, as a prerequisite, (9-23) *must hold*.

Directional Coupler

Figure 9.7 shows a directional coupler of coupling value C. The coupler is at a physical temperature T_x. The definition of C is that

$$C = \frac{P_{in}}{P_c} > 1 \qquad (9\text{-}26)$$

A part of P_{in}, P_c, is coupled out of the main line. As a result

$$P_{out} = P_{in}\left(\frac{C-1}{C}\right) \qquad (9\text{-}27)$$

We shall assume an ideal coupler of infinite directivity, that is, the coupled portion of P_{in} will all go to port 3 and no power will reach the internal load. If this coupler is used in a system, what is its noise factor when the signal flow is (a) from port 1 to port 2 and (b) from port 1 to port 3?

The easiest way to solve this is to use (9-2) again. Let us therefore calculate the N_o' reaching port 2:

$$N_{o2}' = \underbrace{\frac{kT_0(C-1)}{C}}_{\substack{\text{noise from}\\\text{source}}} + \underbrace{\frac{kT_x}{C}}_{\substack{\text{noise from}\\\text{internal}\\\text{termination}}} \qquad (9\text{-}28)$$
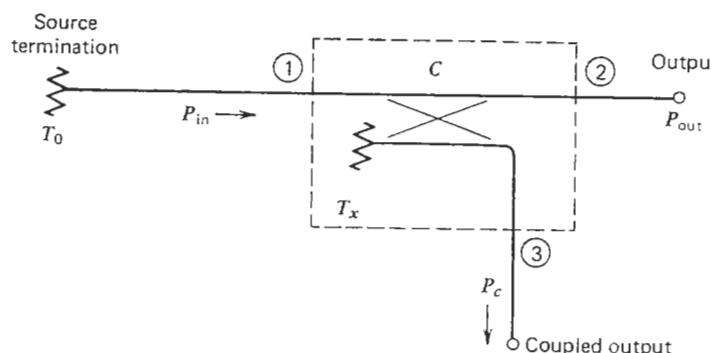
FIGURE 9.7. A directional coupler.

The gain from port 1 to port 2 is given by (9-27). Hence

$$F_{1\text{-}2} = \frac{(k/C)[T_0(C-1) + T_x]}{kT_0[(C-1)/C]} = 1 + \left(\frac{T_x}{T_0}\right)\left(\frac{1}{C-1}\right) \qquad (9\text{-}29)$$

If $T_x = T_0$, F becomes equal to $C/(C-1)$, which is simply the insertion loss between ports 1 and 2. Thus, since a matched coupler can be considered a loss in the main line on account of the fraction of power coupled out to port 3, the result is not surprising if one recalls (9-23).

The other path, port 1 to port 3, yields an analogous result:

$$F_{1\text{-}3} = 1 + \left(\frac{T_x}{T_0}\right)(C - 1) \qquad (9\text{-}30)$$

Ideal Isolator

An ideal isolator has the property that power incident at port 1 (Fig. 9.8) will flow unimpeded to port 2, while power entering port 2 will be dissipated in the internal load on port 3. In practical ferrite isolators there will be a small forward insertion loss $L_{1\text{-}2}$, usually less than 1 dB for typical microwave units. Since the power flowing in the opposite direction is diverted to the internal termination, the reverse insertion loss $L_{2\text{-}1}$, called the isolation, is ideally infinite. Again, in practice this value is typically 20–30 dB. We could sum the total noise power N_o' at port 2 in the usual manner, such as in (9-21), to obtain

$$N_o' = \left(\frac{k}{L_{1\text{-}2}}\right)[T_0 + (L_{1\text{-}2} - 1)T_x]$$
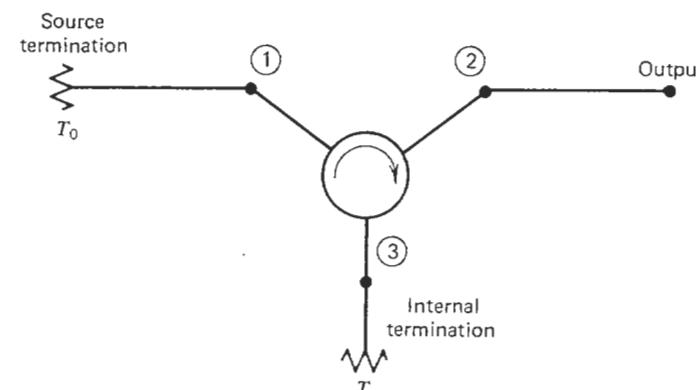
FIGURE 9.8. Ideal isolator.

and therefore

$$F_{1-2} = 1 + (L_{1-2} - 1)\left(\frac{T_x}{T_0}\right) \qquad (9\text{-}31)$$

which is no different from (9-22). After all, an isolator, considered as a black box, behaves in the forward direction as an attenuator of loss L_{1-2}.

9.5.   SYSTEM APPLICATION

The noise factor can be used in the noise analysis of large systems in much the same manner as the noise temperature. Using (9-8), the noise model shown in Fig. 8.2 could very simply be rewritten in terms of F (as in Fig. 9.2), and the calculations in Section 8.8 could be carried out as before.

We would gain additional insight, however, if (9-5) were used as the starting point instead. To repeat

$$F = 1 + \frac{N'_{ni}}{N'_i} = 1 + \frac{N_{ni}}{N_i} \qquad (9\text{-}32)$$

where N'_{ni} is the equivalent noise density which accounts for noise originating within the transducer, but referred to the *input* of the latter; N'_i is the available noise density from the source as measured at the same point; and N_{ni} and N_i are the corresponding terms for a finite noise bandwidth df. In order to understand these terms clearly, it is suggested that the derivation of (9-5) be reviewed.

When (9-32) is solved for N'_{ni}, we get

$$N'_{ni} = (F - 1)N'_i = (F - 1)kT_0 \qquad (9\text{-}33)$$

or

$$N_{ni} = (F - 1)kT_0 \, df \qquad (9\text{-}34)$$

for a finite noise bandwidth.

Equations (9-33) and (9-34) lead directly to the noise model shown earlier in Fig. 9.2. The noise contributed by the transducer and that from the source have now been separated, just as was done in Section 8.1 and Fig. 8.2. Note carefully, that the expression for the transducer excess noise, $(F - 1)kT_0 \, df$, uses the standard temperature T_0. This is a direct consequence of F having been defined in terms of T_0. To use any other value in this expression renders the latter meaningless. The noise contributed by the *source*, $kT_g \, df$, could be due to any arbitrary noise temperature and hence T_g can assume any value. The source could, for example, be a very "hot" gas discharge or solid-state
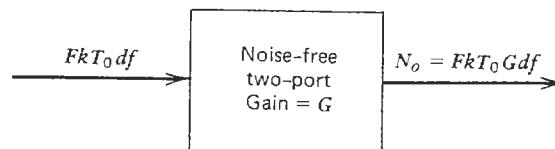
FIGURE 9.9.   Noise relationship when source is at T_0.

noise source, or it could be a very cold portion of the sky. In contrast, once we use F to characterize the noisiness of a transducer, T_0 becomes mandatory in (9-33) and (9-34), and in Fig. 9.2. Otherwise, the numerical value of F would be inconsistent.

If $T_g = T_0$, that is, the source is at the reference temperature, the effective total input noise in Fig. 9.2 becomes

$$N_{ni} + N_i = (F - 1)kT_0 \, df + kT_0 \, df = FkT_0 \, df \qquad (9\text{-}35)$$

and the total available *output* noise N_0 is obtained easily (see Fig. 9.9) as

$$N_o = FkT_0G \, df \qquad (9\text{-}36)$$

This is not surprising—we could have obtained (9-36) directly from (9-2). In other words, given a two-port transducer of noise factor F, connected to a source at T_0, the total available output noise is then, *and only then*, given by (9-36). This expression is often encountered in practice, usually in the decibel form:

$$N_o \, (\text{dBm}) = -174 + NF + 10\log df \, (\text{Hz}) + G \, (\text{dB}) \qquad (9\text{-}37)$$

The application of F to system calculations is as straightforward as that of noise temperature. The first three methods described in Section 8.8 remain basically unchanged. If the amplifiers in blocks 2 and 4 and the attenuators L_1 and L_3 in Fig. 8.13 had been characterized in terms of their noise figures, Fig. 8.13 would become as shown in Fig. 9.10 [recall (9-22) and the effect of $T_x \neq T_0$ on NF_1 and NF_3]. Following the procedure in Section 8.8, we could calculate the total noise factor for the blocks 1 to 4 at point (a) from the cascade formula (9-11). Instead, note that for the attenuators $T_x \approx T_0$ and the assumption of (9-25) is therefore permissible. The circuit in Fig. 9.10 can now be converted to that in Fig. 9.11 with a negligible error, and the noise factor F_T at point (a) becomes

$$F_T = 1.91 + \frac{8.51 - 1}{251.2} = 1.94 \quad \text{or} \quad 2.9 \text{ dB}$$

The resultant model is illustrated in Fig. 9.12. The excess-noise density N'_{ni} contributed by the two attenuators and the two amplifiers follows from (9-33):

$$N'_{ni} = (F_T - 1)kT_0 = 3.80 \times 10^{-18} \text{ mW/Hz} \quad \text{or} \quad -174.2 \text{ dBr}$$
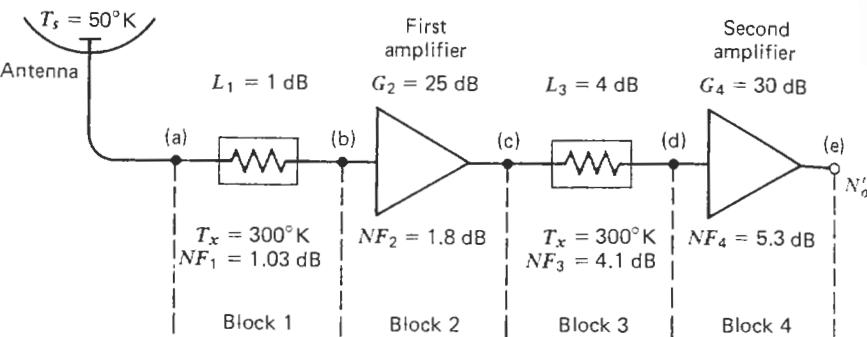
FIGURE 9.10.   Parameters of a receiving system in terms of noise factors.

Note that in contrast to Fig. 8.13 we are here using point (a) as the reference for calculations. There is no special significance in this. The reason is merely that the technique in Fig. 9.11 allowed an easy calculation of F_T at (a) instead of (b).

The total input noise density at (a) is obtained by adding the source (sky) contribution to N'_{ni}:

$$N'_i + N'_{ni} = kT_s + (F_T - 1)kT_0 = (6.9 \times 10^{-19} + 3.8 \times 10^{-18}) \text{ mW/Hz}$$

$$= 4.49 \times 10^{-18} \text{ mW/Hz} \quad \text{or} \quad -173.5 \text{ dBm/Hz}.$$

Since our end goal is the noise density N'_o at the output of the chain, point (e), we have to apply the total gain to the input noise density figure:

$$N'_o = -173.5 + 50 = -123.5 \text{ dBm/Hz}$$

which is the same as that in (8-44).
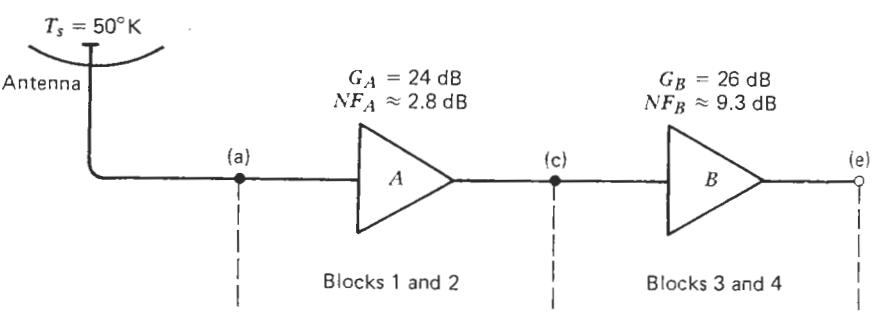
FIGURE 9.11.   Combining of system components.
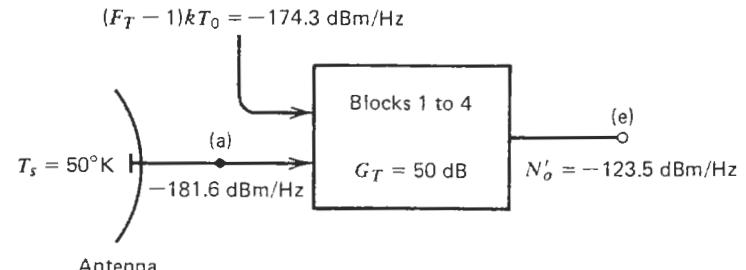
FIGURE 9.12.   Final output noise of a receiving system.

The walk-through and the summation methods discussed in Section 8.8 are also directly applicable. The noise factor of each building block in Fig. 9.10 is converted into its equivalent input noise density in accordance with (9-33) and tabulated as in Fig. 8.14. The remaining calculations and results are identical to those presented in Section 8.8 and in Tables 8.1, 8.2, and 8.3, and need not be repeated here. As an exercise, the reader should recalculate these tables, but starting with the noise figures given in Fig. 9.10.

9.6.   RELATIONSHIP OF F TO THE DYNAMIC RANGE

All receiver systems are bounded by two power limits. At high signal levels, saturation in amplifiers is encountered. At low levels the system noise will eventually mask the signal, and the so-called noise floor will set the minimum level for signals that can be detected. The difference between these limits is called the dynamic range of the system. Figure 9.13 illustrates the transfer characteristic of an ideal system, and shows the effects that the two limits will have on it. Figure 9.13 is plotted in dBm, that is, using logarithmic scales. This yields a convenient straight-line relationship between P_{out} and P_{in}:

$$P_{out} \text{ (dBm)} = P_{in} \text{ (dBm)} + G \text{ (dB)} \qquad (9\text{-}38)$$

The upper limit is easily measured by noting that input power P_{in} at which the deviation from the ideal transfer characteristic is, say, 1 dB. The lower limit where the noise begins to predominate is more difficult to define quantitatively. True, we could watch the output signal on a spectrum analyzer or oscilloscope and note the level at which the signal just disappears into the background noise. This "minimum discernible signal" point could be defined as the lower limit of the dynamic range. The problem, however, is the same described in Section 9.1—different observers may reach different conclusions. The method is simply too loose and the outcome overly subjective.

Fortunately, it is possible to relate the noise factor to the lower limit in a formal manner [9]. We are thus able to *calculate* a unique value for the lower limit of the dynamic range. Consider Fig. 9.14 which shows the low power end of Fig. 9.13 for an amplifier of gain G. The amplifier has an input noise
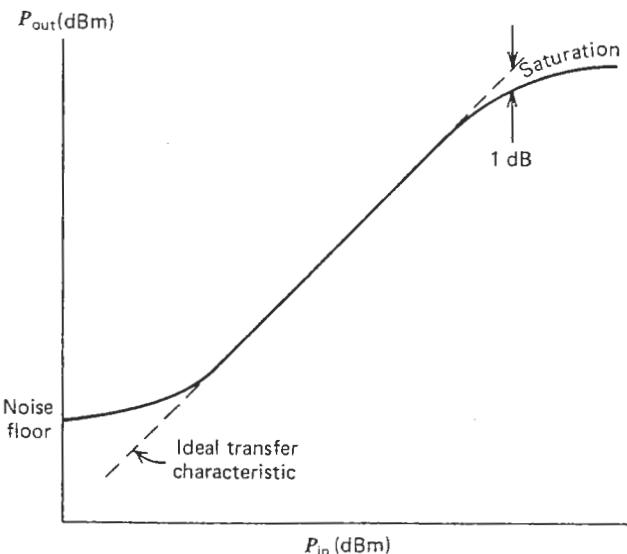
FIGURE 9.13.   Typical transfer characteristic of an amplifier.
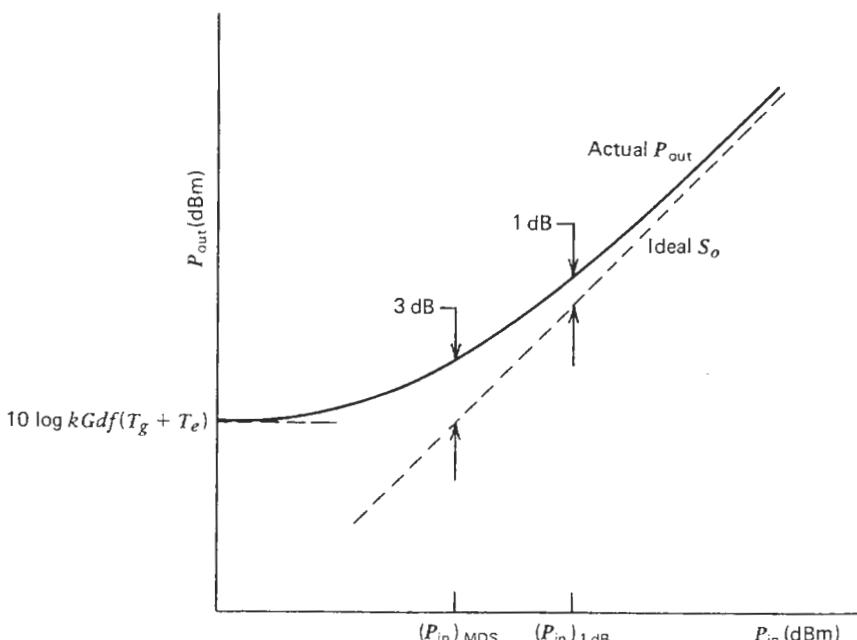
FIGURE 9.14.   Definition of deviation from the linear transfer characteristic.

temperature T_e. It is clear from (9-38) and Fig. 9.14 that if the input source and the amplifier were noiseless, the total output power for this ideal case would be just the signal power given by

$$S_o = S_i G \tag{9-39}$$

In practice there is noise at the output contributed by the source as well as by the amplifier itself. Thus

$$P_{\text{out}} = kG\,df\left(T_g + T_e\right) + S_i G \tag{9-40}$$

The ratio of (9-40) to (9-39) therefore measures the deviation from the ideal characteristic:

$$M = \frac{P_{\text{out}}}{S_o} = \frac{kG\,df\left(T_g + T_e\right) + S_i G}{S_i G} = 1 + \frac{k\,df\left(T_g + T_e\right)}{S_i} \tag{9-41}$$

What does (9-41) mean? Let S_i be very small such that $M \approx k\,df(T_g + T_e)/S_i$. In view of (9-39) we can write

$$P_{\text{out}} \approx S_o \frac{k\,df\left(T_g + T_e\right)}{S_i} = kG\,df\left(T_g + T_e\right) = \text{const} \tag{9-42}$$

Equation (9-42) clearly shows that below a certain S_i the P_{out} no longer decreases. The relationship is shown in Fig. 9.14. If (9-41) is now solved for S_i, we obtain

$$S_i = \frac{k\,df\left(T_g + T_e\right)}{M - 1} \tag{9-43}$$

or in terms of the noise factor,

$$S_i = \frac{k\,df\,T_0\left[\left(T_g/T_0\right) + F - 1\right]}{M - 1} \tag{9-44}$$

from which we could calculate that S_i at which a given deviation of P_{out} from the ideal S_o is observed. For example, for the popular minimum discernible signal (MDS), the noise and signal powers are equal, $M = 2$. Thus

$$(S_i)_{\text{MDS}} = k\,df\left(T_g + T_e\right) \tag{9-45}$$

Since the upper, saturation limit of the dynamic range is usually specified in terms of the 1-dB deviation point, we may choose to do the same at the other end. Hence

$$(S_i)_{1\text{ dB}} = \frac{k\,df\left(T_g + T_e\right)}{0.259} \tag{9-46}$$
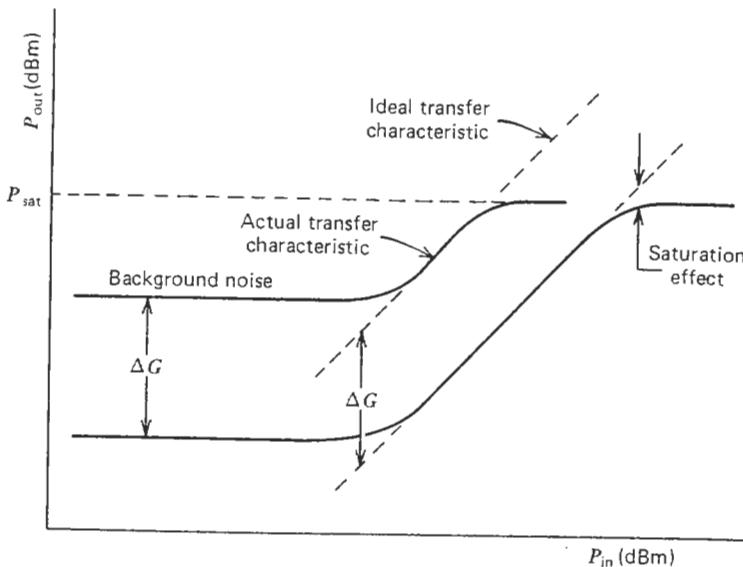
FIGURE 9.15.   Saturation of an amplifier by its own noise.

If $T_g = T_0$, the expressions are simplified and we obtain

$$(S_i)_{MDS} = k\,df\,T_0 F \tag{9-47}$$

and

$$(S_i)_{1\,dB} = k\,df\,T_0\left(\frac{F}{0.259}\right) \tag{9-48}$$

respectively. In logarithmic terms, (9-48) becomes

$$(S_i)_{1\,dB} = -174 + 10\log(df) + NF + 5.9 \tag{9-49}$$

since $10\log(kT_0) = -174$ dBm/Hz.

Note in passing that a given amplifier or receiver system can saturate under its own noise if the gain is sufficiently high. This can be seen from Fig. 9.15. As the gain is increased (ΔG), the background noise level will eventually reach the saturation level P_{sat}.

9.7.   SYSTEM SENSITIVITY FUNCTION (SSF)

Consider a noisy transducer and represent its noise contribution as in the noise model of Fig. 9.2. The *total* output noise density, referred to the input, is given by

$$N_{oi}' = (F-1)kT_0 + kT_g \tag{9-50}$$

If $T_0 = T_g$, this becomes $N_{oi}' = FkT_0$, and the ratio of "sensitivities" of two arbitrary receivers is given by the ratio of their noise factors, as we saw in (9-14). If, on the other hand, $T_g \neq T_0$, then (9-15) applies and it is advantageous to define [10] a system sensitivity function (SSF) by rearranging (9-50):

$$N_{oi}' = kT_0\left(\frac{T_g}{T_0} + F - 1\right) = (\text{const}) \times (\text{SSF}) \tag{9-51}$$

Clearly, a sensitive receiving system must have a low SSF. If we express the SSF in decibels, then

$$(\text{SSF})_{dB} = 10\log\left(\frac{T_g}{T_0} + F - 1\right) \tag{9-52}$$

Figure 9.16 is a plot of (9-52). It shows that if the background noise temperature T_g is low, a 1-dB improvement in the receiver noise figure NF
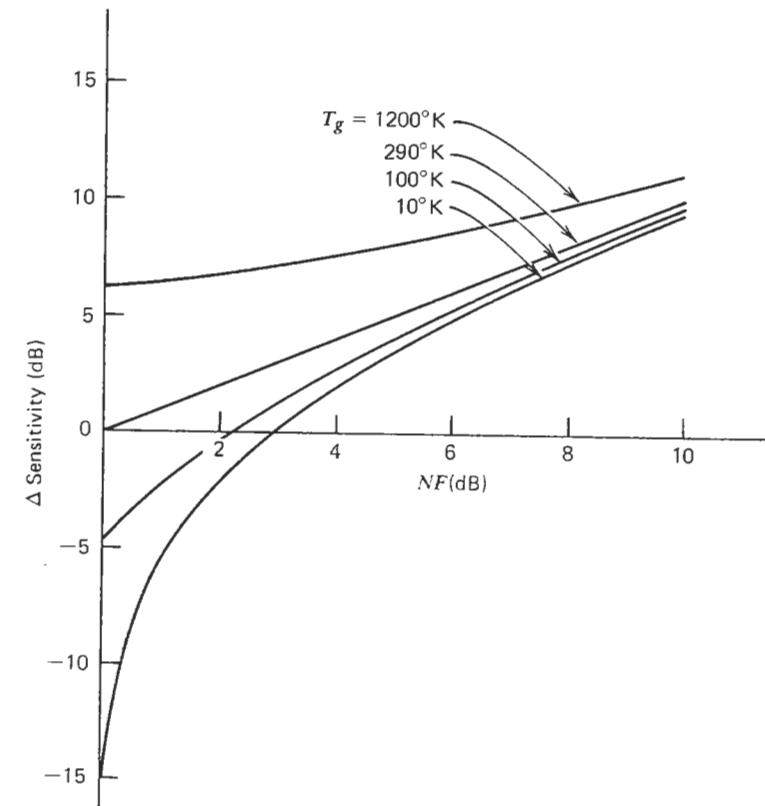
FIGURE 9.16.   Improvement in system noise performance as a function of noise factor and source noise temperature.

could mean several dB of improvement in system sensitivity. Conversely, if the T_g is high, an improvement in receiver noise figure would have little effect.

9.8.   AVERAGE NOISE FACTOR

The two definitions of noise factor, (9-2) and (9-3), are effectively single-frequency or spot values based on a unit bandwidth. The important point is that whatever the bandwidth, it is assumed sufficiently narrow so that all parameters may be considered constant over this bandwidth. Remember that the spot noise factor is a function of frequency since N_o' and G both vary as one moves from one narrow frequency slot df_1 to another df_2. We should therefore properly write that $F = F(f)$.

One could, conceptually at least, determine F at successive df using a very narrowband receiver, and then plot the results as a function of frequency. The drawback would be that the noise performance of a given transducer would no longer be given by one convenient number. It will be remembered that the very purpose of noise factor was to provide a single-number characterization of noise. In practice, when measuring (Chapter 11) and using noise factors, we are, more often than not, working with finite bandwidths, usually broad enough so that the F *does* vary over the bandwidth. We saw in Section 8.3 how an average noise temperature could be defined over a finite bandwidth. In an analogous manner the same could be done for the noise factor, resulting in an average noise factor $\bar{F}$. The starting point is (9-2), and we shall now consider how this definition would be extended to an arbitrary bandwidth B_0. Since it is recognized that both N_o' and G vary with frequency, the procedure clearly would be to evaluate the following expression:

$$\bar{F} = \frac{\int N_o'(f)\, df}{\int kT_0 G(f)\, df} \qquad (9\text{-}53)$$

But using (9-2), this becomes (integrating over B_0)

$$\bar{F} = \frac{\int F(f) G(f)\, df}{\int G(f)\, df} \qquad (9\text{-}54)$$

which yields the relationship between the spot and the average noise factors.

If we compare F to $\bar{F}$, it is apparent that $\bar{F}$ does not provide as much information as F. We could rewrite (9-53) as follows

$$\bar{F} = \frac{\overline{N_o} B_0}{kT_0 \overline{G} B_0} = \frac{\overline{N_o}}{kT_0 \overline{G}} = (\text{const})\left(\frac{\overline{N_o}}{\overline{G}}\right) \qquad (9\text{-}55)$$

which says that $\bar{F}$ is proportional to the ratio of average output noise to average power gain, the averages being taken with respect to the bandwidth B_0. The fine-grain variation of F has thus been lost.

Additionally, in order to evaluate $\bar{F}$, we need more information than was required for F. The frequency distribution of both $N_o'(f)$ and $G(f)$ must be known before $\bar{F}$ can be calculated from (9-53).

In practice the bandwidth B_0 is frequently taken to be the equivalent noise bandwidth B (see Section 3.4) of the transducer. Also, since in real life power *delivered* to the output load is easier to measure than the available power, $\bar{F}$ is usually defined in terms of transducer gain rather than the available gain (see Section 7.4). The value of $\bar{F}$ is unaffected by this choice since $\bar{F}$ is a power ratio. However, for the sake of consistency, G_t must then be used in (9-54).

The average noise factor $\bar{F}$ is thus the desired single-number measure of the noise performance of a practical transducer, and as such it shares the usual advantages and disadvantages of representing an inherently variable parameter by its constant average value. For example, we found that the basic definitions of F did not depend on bandwidth. The value of $\bar{F}$, on the other hand, could depend on bandwidth if the frequency distribution of noise, N_o', and that of gain, G, are different (see Fig. 8.5). Still, with this proviso, $\bar{F}$ is a very useful concept and is widely used in practical work. As a matter of fact, the term noise factor, as encountered in popular usage, refers nearly always to some kind of average noise factor $\bar{F}$, even for narrowband measurements, because, after all, it is the average value of F which is measured in practice.

9.9.   DEPENDENCE OF F ON SOURCE IMPEDANCE

It was pointed out in Section 9.1 that the noise factor of a linear, noisy two-port depends on the source impedance, and consequently, a given F is meaningful only if the source impedance is specified. Furthermore, the relationship between F and Z_g is not a simple linear one. Mumford and Scheibe [11] give a good qualitative argument to illustrate this point: One should visualize two unilateral amplifiers with gain in the forward direction, but infinite isolation in the reverse direction. Let both amplifiers have initially the same noise factor $F = 1 + N_{no}/N_i G$, but assume the internal noise sources in the first amplifier to be located near the output terminals, while in the second amplifier they are near the input terminals. Consider next what happens when the generator resistance R_g is varied, creating a variable mismatch at the input. The gain of both amplifiers would vary equally. However, while the term N_{no} of the first amplifier would not be affected at all by the input mismatch, the situation with the second amplifier would clearly be different. To compound this difficulty, this dependence of N_{no} on R_g is a function of whether the internal noise sources are represented as voltage generators, current generators, or a combination of the two, and what, if any, correlation there is between them (Chapter 6).

So, this section will investigate quantitatively the dependence of F on Z_g, starting with the noise model developed in Chapter 6. We established there that the full description of noise properties of a linear two-port transducer requires four quantities [12]. They are the two equivalent noise generators, e_n and i_n, and the real and imaginary components of the correlation impedance. This representation is very useful since the noise effect is now lumped into the two generators at the input side of the transducer, and the generators describe completely the internal noise sources as far as their contributions to terminal voltages and currents are concerned. We thus have a separation of the given two-port into a noisy network, followed by a noise-free network. Since the SNRs are not affected by linear noise-free networks, we can simply disregard the noise-free portion of the transducer model in noise factor calculations because F involves only ratios of power.

For convenience, the equivalent circuit developed in Section 6.2 is repeated in Figure 9.17a. As before, $Y_c = G_c + jB_c$ is the complex correlation admittance and Y_g is the complex generator admittance. The Y_c is considered to be at 0°K and therefore generates no noise. Before proceeding, it is advantageous

(a)

(b)
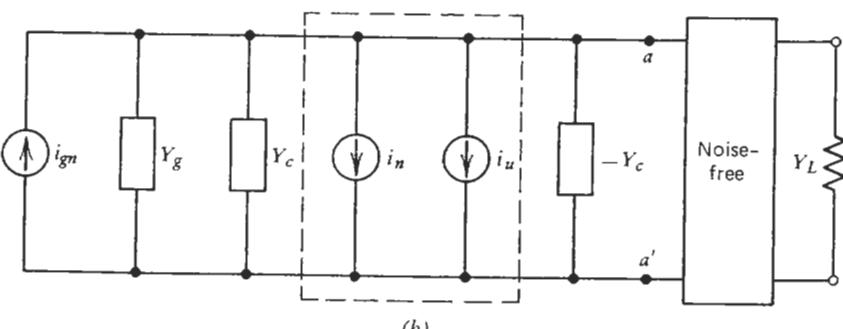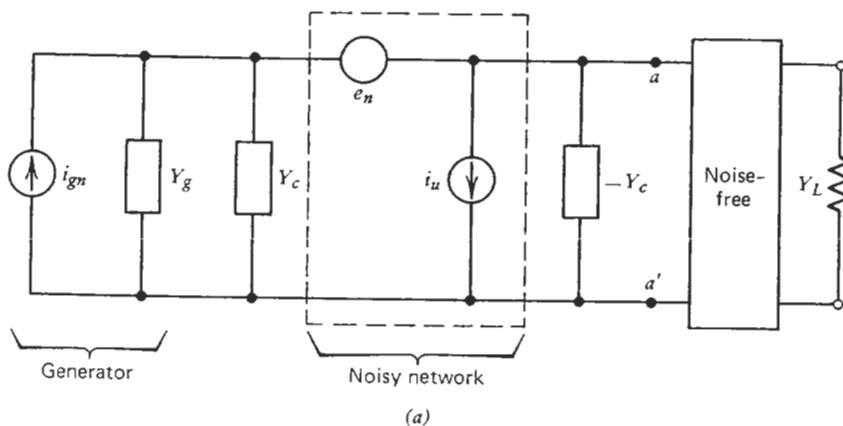
FIGURE 9.17. Noise model of a noisy transducer, including the source noise.

to convert the series voltage generator e_n into an equivalent shunt current generator by applying Norton's theorem to the network comprised of e_n, Y_g, and Y_c. Thus

$$i_n = e_n(Y_g + Y_c) \tag{9-56}$$

Note, by the way, that Fig. 9.17a is the dual of that considered* in Fig. 6.7a. There we split the e_n into two parts—one fully correlated with i_n and one uncorrelated. Here it is the i_n which is treated in this manner, resulting in i_u and $e_n Y_c$.

The circuit now becomes as shown in Fig. 9.17b. The noise factor is derived by applying (9-5) to Fig. 9.17b. The noise available at a–a', originating in the generator conductance G_g, is given by

$$N_i = \frac{\overline{i_{gn}^2}}{4(G_g + G_c - G_c)} \tag{9-57}$$

The noise contributed by the noisy portion of the network is

$$N_i = \frac{\overline{i_n^2} + \overline{i_u^2}}{4(G_g + G_c - G_c)} = \frac{\overline{i_u^2} + \overline{e_n^2}|Y_g + Y_c|^2}{4(G_g + G_c - G_c)} \tag{9-58}$$

Hence from (9-5),

$$F = \frac{\overline{i_{gn}^2} + \overline{i_u^2} + \overline{e_n^2}|Y_g + Y_c|^2}{\overline{i_{gn}^2}} \tag{9-59}$$

Note carefully that we are permitted to write the numerator of (9-59) as a sum of mean-square terms *only* because all three noise generators are uncorrelated. Equation (9-59) can now be simplified using (3-22) and (3-27) and Fig. 9.18:

$$F = 1 + \frac{4kT_0BG_u + 4kT_0BR_n|Y_g + Y_c|^2}{4kT_0BG_g}$$

$$= 1 + \frac{G_u}{G_g} + \frac{R_n}{G_g}|Y_g + Y_c|^2$$

$$= 1 + \frac{G_u}{G_g} + \frac{R_n}{G_g}\left[(G_g + G_c)^2 + (B_g + B_c)^2\right] \tag{9-60}$$

*Th
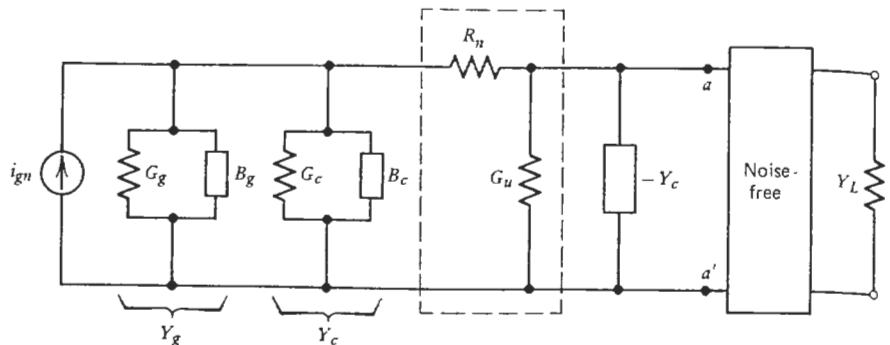
FIGURE 9.18.   Noise model of a noisy transducer using equivalent noise resistances.

We first observe that the noise factor could be decreased by "noise tuning", that is, by adjusting the generator susceptance B_g such that

$$B_g \equiv B_{g0} = -B_c \qquad (9\text{-}61)$$

Equation (9-60) now becomes

$$F = 1 + \frac{G_u}{G_g} + \frac{R_n}{G_g}(G_g + G_c)^2 \qquad (9\text{-}62)$$

which has a minimum when

$$G_g \equiv G_{g0} = \sqrt{\frac{G_u}{R_n} + G_c^2} \qquad (9\text{-}63)$$

where G_{g0} and B_{g0} represent the optimum values of generator conductance and susceptance, respectively. Combining (9-63) with (9-62) yields the optimum noise factor F_0:

$$F_0 = 1 + 2R_n(G_{g0} + G_c) \qquad (9\text{-}64)$$

and finally

$$F = F_0 + \frac{R_n}{G_g}\left[(G_g - G_{g0})^2 + (B_g - B_{g0})^2\right] \qquad (9\text{-}65)$$

Thus we see again that contrary to popular belief, F alone is *not sufficient* for a complete characterization of a linear, noisy two-port. *Four* parameters are required; for the noise factor, they are F_0, R_n, G_{g0}, and B_{g0}. Once these are given, F is completely determined for any generator admittance.

If one had complete control over the generator admittance Y_g, it would be a simple matter to adjust B_g to B_{g0} and G_g to G_{g0}. The noise factor would be
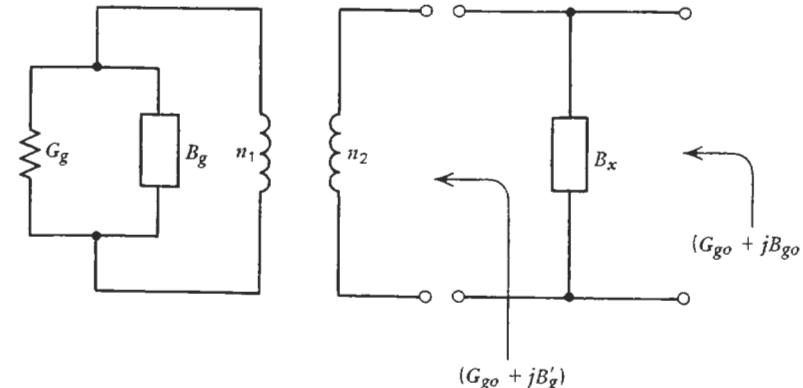
FIGURE 9.19.   The realization of noise match using a transformed source admittance.

minimized and the job would be over. In practice Y_g cannot usually be altered at will. We must therefore resort to a matching transformer to realize the required G_{g0}. From Fig. 9.19,

$$\left(\frac{n_2}{n_1}\right)^2 = \frac{G_g}{G_{g0}} \qquad (9\text{-}66)$$

This yields the correct generator conductance G_{g0}, but also transforms B_g to an equivalent generator susceptance of

$$\frac{B_g}{B_g'} = \left(\frac{n_2}{n_1}\right)^2 \quad \text{or} \quad B_g' = B_g\left(\frac{n_1}{n_2}\right)^2 \qquad (9\text{-}67)$$

Therefore, combining the two operations requires that for correct "noise tuning" the equivalent generator susceptance should satisfy (9-61), that is, $B_{g0} = -B_c$. This is achieved by providing an appropriate susceptance B_x such that (see Fig. 9.19)

$$B_x + B_g' = B_{g0} = -B_c \qquad (9\text{-}68)$$

We have now simultaneously met both conditions—noise tuning and proper generator conductance. As a result, $F = F_0$ follows.

The argument given above forms the basis of the noise-matching technique whereby a transformer of appropriate turns ratio is inserted between the generator and the transducer to improve the noise factor. Note, however, that *the noise match does not necessarily imply a circuit match.* This can be seen qualitatively by noting that the input admittance Y_{in} of an arbitrary two-port (which determines the input circuit match) is given by (Appendix B)

$$Y_{\text{in}} = \frac{C + DY_L}{A + BY_L} \qquad (9\text{-}69)$$

that is, Y_{in} is a function of Y_L, the load admittance. The derivation of (9-63), on the other hand, is independent of load admittance connected to $a-a'$ in Fig. 9.17a.

The result in (9-65) was based on the equivalent circuit shown in Fig. 9.17. A dual version where the correlation is expressed in terms of a complex correlation impedance is also possible. The corresponding equivalent circuit was discussed in Section 6.2 and the final results are:

$$R_{g0} = \sqrt{\frac{R_u}{G_n} + R_c^2} \tag{9-70}$$

$$F_0 = 1 + 2G_n(R_{g0} + R_c) \tag{9-71}$$

and

$$F = F_0 + \frac{G_n}{R_g}\left[(R_g - R_{g0})^2 + (X_g - X_{g0})^2\right] \tag{9-72}$$

The quantities G_{g0}, B_{g0}, R_n, and F_0, which appear in (9-65), and R_{g0}, X_{g0}, and G_n, in (9-72), can all be determined experimentally from noise factor measurements. The optimum noise factor F_0 will be the same in both cases. It does not depend on whether the generator has an impedance or admittance configuration.

The methods of measuring these parameters will be covered in Section 11.4.

9.10.   SUMMARY

(a)   Noise factor:

(1)   According to the IRE definition:

$$F = \frac{N_o'}{kT_0G} = \frac{N_o'}{290kG}$$

(2)   According to H. T. Friis' definition:

$$F = \frac{S_i/N_i'}{S_o/N_o'}$$

(b)   The reference temperature T_0 for the noise factor is always 290°K.

(c) F depends on the source impedance, but does not depend on the bandwidth or the gain of the transducer.

(d) F is a measure of *degradation* in the signal-to-noise ratio (SNR), not a measure of excellence of the output SNR.

(e)   The input noise temperature T_e and the noise factor are related through

$$T_e = (F - 1)T_0$$

(f)   Overall noise factor of N cascaded stages is

$$F_T = F_1 + \frac{F_2 - 1}{G_1} + \cdots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}}$$

(g)   If the source noise temperature is T_0, the output noises of two transducers relate as

$$\frac{N_{o2}}{N_{o1}} = \frac{F_2}{F_1}.$$

(h)   Noise factor of a matched attenuator L at temperature T_x is

$$F = 1 + (L - 1)\left(\frac{T_x}{T_0}\right)$$

and

$$F = L \quad only \text{ if } T_x = T_0.$$

(i)   Lower limit of dynamic range for $10 \log M$ dB deviation from linearity:

$$(S_i)_M = \frac{k\,df\,(T_g + T_e)}{M - 1}$$

$$= \frac{k\,df\,T_0\left[(T_g/T_0) + F - 1\right]}{M - 1}.$$

(j)   Average noise factor over a bandwidth B_0:

$$\bar{F} = \frac{\int_{B_0} F(f)G(f)\,df}{\int_{B_0} G(f)\,df}.$$

(k)   Noise factor as a function of source admittance:

$$F = F_0 + \frac{R_n}{G_g}\left[(G_g - G_{g0})^2 + (B_g - B_{g0})^2\right]$$

where G_{g0} and B_{g0} are the optimum generator conductance and susceptance, respectively, and R_n the equivalent input noise resistance.

REFERENCES

1. H. T. Friis, "Noise Figures of Radio Receivers," *Proc. IRE*, vol. 32, pp. 419–422, July 1944.

2. H. T. Friis and D. O. North, "Discussion on Noise Figures of Radio Receivers," *Proc. IRE*, vol. 33, pp. 125–127, February 1945.

3. E. W. Herold, "An Analysis of the Signal-to-Noise Ratio in Ultrahigh Frequency Receivers," *RCA Rev.*, vol. 6, pp. 302–331, January 1942.

4. D. O. North, "The Absolute Sensitivity of Radio Receivers," *RCA Rev.*, vol. 6, pp. 332–344, January 1942.

5. "Standards on Receivers: Definition of Terms, 1952," *Proc. IRE*, vol. 40, pp. 1681–1685, December 1952.

6. "IRE Standards on Electron Tubes: Definitions of Terms, 1957," *Proc. IRE*, vol. 45, pp. 983–1010, July 1957.

7. "IRE Standards on Electron Tubes: Definition of Terms, 1962," *Proc. IEEE*, vol. 51, pp. 434–435, March 1963.

8. H. Goldberg, "Some Notes on Noise Figure," *Proc. IRE*, vol. 36, pp. 1205–1214, October 1948.

9. J. J. Sie, private communication, 1960.

10. L. A. Blackwell and K. L. Kotzebue, *Semiconductor-Diode Parametric Amplifiers*, Prentice-Hall, Englewood Cliffs, N.J., 1961.

11. W. W. Mumford and E. H. Scheibe, *Noise Performance Factors in Communication Systems*, Horizon House, Dedham, 1968.

12. H. A. Haus, chairman, et al., "Representation of Noise in Linear Twoports," *Proc. IRE*, vol. 48, pp. 69–74, January 1960.

10 | MULTI-RESPONSE TRANSDUCERS

Chapters 8 and 9 dealt exclusively with single-response transducers. The name "single response" means that to each output frequency there corresponds only one input frequency. The simplest example of a single-response transducer is a linear amplifier which merely amplifies an input signal without any frequency translation, conversion, or generation of new frequencies. In contrast, a multi-response transducer has several input frequencies which all contribute either equally or unequally to a single output frequency. An example here would be a superheterodyne receiver or any downconverter where as a minimum the signal *and* its image on the other side of the local oscillator frequency both contribute to the same IF output (Fig. 10.1). This constitutes a dual-response system where the two bands of frequencies are customarily called the signal response and the image response. Strictly speaking, the upconverting mixer, which converts an IF signal into two RF signals located symmetrically with respect to the local oscillator frequency, could also be called a multi-response transducer. It has several *output* frequencies for one input frequency. However, in this book the discussion will be limited to the downconverting type of the multi-response transducer only because of the importance of the latter from the viewpoints of the receiver design and noise analysis. It should be noted that since the image and the signal responses behave electrically as if they were two separate inputs, the term "port" is frequently applied to them, particularly in microwave work. We thus refer to the signal port and the image port of a multi-port transducer although both frequencies may enter the device physi-
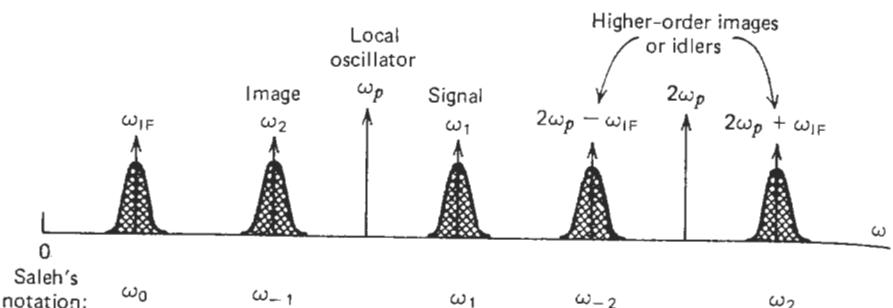
FIGURE 10.1.   Definition of frequencies in multi-response systems.

cally through the same connector or waveguide opening. The terms multi-response and multi-port will therefore be used interchangeably. It is also important to note when talking about responses or ports of a transducer that the local oscillator or pump ω_p input is not considered as one of them. The assumption is made that the local oscillator (LO) voltage is large enough to drive the nonlinear element of the transducer (e.g., a mixer diode) completely from one extreme state (conducting) to the other state (nonconducting). Furthermore, the LO voltage is assumed to be much greater than any of the applied signal voltages. Under these conditions the laws of linear network theory continue to apply even though a highly nonlinear circuit element, characterized by a multiplicity of frequency translations, is involved. The LO is thus not a response or input port in the information transmission sense. It merely provides the coupling mechanism between the various signal and image frequencies.

Returning to the downconverter illustrated in Fig. 10.1, we note that there are two principal responses—one centered around $\omega_1 = \omega_p + \omega_{IF}$ and the other around $\omega_2 = \omega_p - \omega_{IF}$. The design of the system determines which of these is used as the signal and which is the image. The designations shown in Fig. 10.1 are therefore arbitrary and for our purpose the choice is irrelevant. The important point is that signals and noise are converted to the output IF band from *both* responses. If the local oscillator has strong harmonics, additional spurious responses would exist at $2\omega_p \pm \omega_{IF}$, $3\omega_p \pm \omega_{IF}$, and so on, and signals present at these freqencies could also be converted to the IF output. Each response has its own gain G_r and noise bandwidth B_r. It is important to note that even for a sinusoidal applied LO waveform the situation is quite complex. The nonlinear element of the transducer will itself generate harmonics of the LO and the applied signal frequencies, plus a multitude of mixing products, all of which are modified by the respective source and load terminations, and reflected back into the transducer. Thus, not only are unwanted external signals (and noise) being coupled into the system through the spurious image bands, but the self-generated mixing products at image frequencies, even if confined wholly within the transducer circuitry, will have an effect on the overall operation. Putting it more succinctly, the image

terminations seen by the transducer can sensitively affect both the noise performance and the conversion gain (or loss) from one response to the other.

One is therefore faced with a multi-dimensional problem, requiring a complex matrix representation [1–3] to account for the many interactions and interrelationships. Since the purpose of this book is to present the fundamentals of noise analysis as applied to typical single- and multi-response transducers encountered in engineering practice, it is necessary from time to time to adopt certain assumptions and simplifications to bring out clearly the main points, and to keep them from being obscured by overly complicated mathematics. Once the basics of multi-response transducers are thoroughly understood, the reader can pursue any particular topic to a greater depth, using the large number of sources available in the literature, some of which are listed in the references and the general bibliography sections.

Let us begin by considering a typical superheterodyne receiver where only the signal and the first image, either $\omega_p + \omega_{IF}$ or $\omega_p - \omega_{IF}$, are of interest. For now, designate the signal with the subscript 1 and the image with 2. We shall adopt a more formal designation later. The available power gains for higher-order image responses, G_3, G_4, and so on will be assumed negligible compared to G_1 and G_2, although at times these terms will be retained for the sake of generality. Further simplifications are possible if the IF frequency is very low or the RF bandwidth of the receiver is sufficiently broad to include both the signal ω_1 and the image ω_2. The gains G_1 and G_2 could then be equal. Frequently, however, this is not so, because the image gain may by design be lower than that of the desired signal.

The overall system bandwidth of superheterodyne receivers is usually set by the IF amplifiers which are much narrower in bandwidth than the front end RF circuits. Consequently, we may set

$$B_1 = B_2 = B_{IF} \equiv B \qquad (10\text{-}1)$$

Again, caution is in order, for this need not always be true. Although these assumptions simplify the mathematics of multi-response transducers without an undue loss of generality, the added complication of noise conversion from other spurious (image) responses into the desired IF output channel does, as expected, impact the definitions and mathematical formulation of noise factor and noise temperature. Even such familiar expressions as the noise factor of cascaded networks, (9-11), will have to be modified. Confusing as this may seem, it is nevertheless a straightforward procedure *provided* one keeps the IRE definitions of noise factor and noise temperature firmly in mind. As emphasized in Chapter 9, much confusion can be avoided if this advice is heeded, especially when dealing with single-sideband and double-sideband operations in a multi-response environment.

Finally, it should be reiterated that the transmission from the signal port to the output (IF) port is a function of the image termination(s). The image could, for example, be terminated reactively with a filter preceding the input port and

passing only the signal frequency, or resistively if the image sees the same source resistance as the signal. In each case the effect is different. Conventional microwave mixers are usually quite broadband (unless they contain a built-in image filter), and a resistively terminated image response is therefore of considerable practical interest. Physically, a downconverting mixer has one RF input connector, but electrically this constitutes two input ports—signal and image. In a general case, these ports could well see different source resistances and/or source noise temperatures. We shall consider the effects of image termination in greater detail in Section 10.7, but in the following sections a broadband unit is assumed, with resistive signal and image terminations, because this is the configuration most often encountered in practice.

One may wonder why the abstract and perhaps awkward designation "multi-response transducer" is retained while in practice such a device is commonly nothing more than the familiar downconverting mixer. The explanation is that we shall strive first to gain a thorough understanding of the fundamentals of multi-response networks and systems before encumbering ourselves with the specifics of physical mixer circuits.

10.1.   OUTPUT NOISE AND NOISE TEMPERATURE

Our starting point is Eq. (8-3), which for the single-response case yields the output noise in a bandwidth df in accordance with

$$N_o = kG(T_g + T_e)\, df \qquad (10\text{-}2)$$

where G is the available power gain of the transducer. Equation (10-2) consists of two terms. The first one, $kGT_g\, df$, is merely the noise from the source resistance at T_g, modified by the available gain of the transducer. The second term, $kGT_e\, df$, represents noise contributed by the transducer itself, the so-called excess noise. It is expressed in terms of a fictitious noise generator connected to the input terminals of the noise-free transducer and having a noise temperature T_e. The value of the latter is a measure of the noise power actually observed. This equivalent, lumped noise generator thus accounts for the internally generated noise even though the physical noise mechanism within the transducer could well be distributed in nature. Through this artifice we have separated out the "noisiness" of the given transducer, and the transducer itself can now be considered noise-free (Fig. 8.2).

The noise model for a multi-response transducer can be developed in an analogous manner. Again, we shall be concerned with two principal contributions of noise at the designated output port—the source noise and the excess noise. The difference is, however, that we must now take into account contributions from *several* responses instead of just one.

The total *source* noise, as measured at the output of a multi-response transducer, is given simply by

$$(N_i)_o = k \sum_{r=1}^{R} (T_{gr}G_r B_r) \qquad (10\text{-}3)$$

where G_r is the available power gain of the rth response; B_r the noise bandwidth of the rth response; and T_{gr} the noise temperature of the source resistance of the rth response. Notice the designation $(N_i)_o$ which means the *input* noise from the source as measured at the *output* of the transducer.

The *excess* noise in multi-response transducers is handled in much the same manner as the excess noise in single-response cases. We start again with the total excess noise observed at the output, but then it is argued that this noise is contributed by all responses, each having the *same* equivalent input noise temperature T_e, although the gains and noise bandwidths of the responses could be different:

$$N_{no} = kT_e \sum_{r=1}^{R} (G_r B_r) \qquad (10\text{-}4)$$

This statement is in accordance with the IRE definition of T_e for multi-response transducers [4], which states effectively the following:

The effective input noise temperature T_e of a multiport transducer, with one port designated as the output port, is that temperature which, when assigned simultaneously to the specified impedance terminations at the set of frequencies contributing to the output, at all accessible ports except the designated output port of noise-free equivalent of the transducer, would yield the same available noise power in a specified output band as that of the actual transducer connected to noise-free equivalents of the same input terminations.   □

For reasons which will become clearer shortly, we shall call T_e the multi-response noise temperature.
From (10-4),

$$T_e = \frac{N_{no}}{k \displaystyle\sum_{r=1}^{R} (G_r B_r)}$$

$$= \frac{N_{no}}{k\overline{G}_1\overline{B}_1\mathscr{R}} \qquad (10\text{-}5)$$
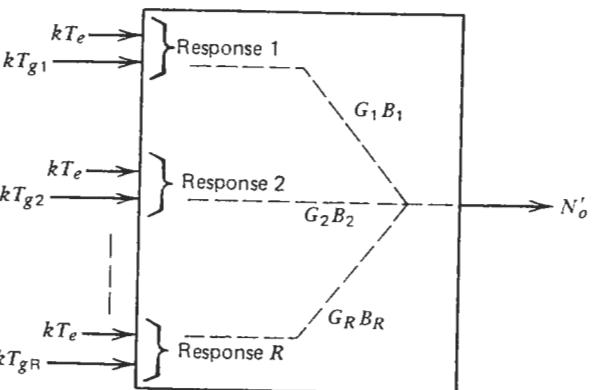
FIGURE 10.2. Noise model of a multi-response transducer using equivalent input noise temperature.

where

$$\mathscr{R} = 1 + \frac{G_2 B_2}{G_1 B_1} + \frac{G_3 B_3}{G_1 B_1} + \cdots + \frac{G_R B_R}{G_1 B_1} \tag{10-6}$$

is called the response factor [5].

Combining the two expressions, (10-3) and (10-4), yields the total output noise

$$N_o = (N_i)_o + N_{no} = k \sum_{r=1}^{R} \left[G_r B_r (T_{gr} + T_e) \right] \tag{10-7}$$

If we assume all source temperatures T_{gr} to be equal, then (10-7) can be rewritten

$$N_o = k G_1 B_1 (T_g + T_e) \mathscr{R} \tag{10-8}$$

where $\mathscr{R}$ is the response factor as before. The noise model for the general case is therefore as shown in Fig. 10.2.

It must be remembered that T_e is still a function of the source impedances connected to the various responses. This, however, raises a question. How can all responses have the same T_e when their source impedances may well be unequal? After all, it is entirely possible that the source impedance at the image frequency might differ from that at the signal frequency. This difficulty is resolved by recalling how the input noise temperature T_e was obtained in a single-response case. Clearly, it is not possible to measure T_e directly by connecting some hypothetical test instrument to the input port of a transducer. Instead, T_e is found indirectly by first determining the available excess noise

N_{no} at the *output* of the transducer. The N_{no} is then divided by the available power gain G and thereby referred to the input.

Finally, the value of T_e is calculated from (see Fig. 8.2)

$$T_e = \frac{N_{no}/G}{kB} = \frac{N_o}{kGB} - T_g \tag{10-9}$$

Note that the key to this operation is the power gain G. The latter is a function of the source impedance (Section 7.5) and thus ties the T_e directly to Z_g.

For multi-response transducers the situation is analogous. The total excess noise at output $(N_{no})_T$ is again referred to the input except that this time we recognize it to be a sum of contributions from *several* responses, each contributing to the output in accordance with its power gain G_r [assume all noise bandwidths to be equal in (10-1)]:

$$(N_{no})_T = N_o - \underbrace{kB \sum_{r=1}^{R} (G_r T_{gr})}_{\substack{\text{total source} \\ \text{noise}}} = k T_e B \sum_{r=1}^{R} G_r \tag{10-10}$$

Note that we are at liberty to assign the *same* T_e to all responses despite the possible differences in source impedances, provided the numerical value of T_e satisfies (10-10). The ΣG_r is what accounts for the differences in input terminations (Fig. 10.2).

A relationship analogous to (10-9) for the general case of unequal source temperatures could be easily derived from (10-10), but the result is awkward and adds little to our understanding. It behooves us therefore to consider certain special cases of (10-7) instead. Equation (10-8) already took care of the equal-source-temperature case. Frequently, there are only two significant responses—signal and image—that is, G_r is very small for all $R > 2$ and we have a dual-response system. From (10-8),

$$N_o = k G_1 B_1 \left(1 + \frac{G_2 B_2}{G_1 B_1} \right) (T_g + T_e) \tag{10-11}$$

If, furthermore, the conditions $G_1 = G_2$ and $B_1 = B_2$ hold, then

$$N_o = 2kGB(T_g + T_e) \tag{10-12}$$

and the total available output noise is seen to be twice that of a single-response system of the same gain and noise bandwidth. Keep in mind, however, that the number 2 is in no way sacred, being merely a consequence of special circuit conditions. The general expression is still (10-11) or even (10-8), corresponding to Fig. 10.2.

10.2.   OPERATING NOISE TEMPERATURE

How is the concept of operating noise temperature applied to multi-response transducers? Let us first recall that the rationale for using T_{op} in the first place was that the operating noise temperature provided a convenient means of calculating the output SNR of a system. In single-response systems there was only one channel to worry about and (8-15) followed rather naturally from basic definitions. In multi-response systems the situation is more complicated. In addition to *noise* contributions from several responses, the *signal* could also be present simultaneously in more than one response. In communication and radar receivers the signal occupies typically only the "signal" response, but in radiometry receivers the signal is broadband noise which covers many, if not all, responses.

Let us first consider the radar case. There is only one input signal S_i and hence only one output signal $S_o = S_i G_1$. Using (10-8) and (10-1), the normalized SNR_o is

$$\mathrm{SNR}'_o = \frac{S_o}{N'_o} = \frac{S_i G_1}{N'_o} = \frac{S_i}{k(T_g + T_e)\mathscr{R}'} \tag{10-13}$$

where

$$\mathscr{R}' = 1 + \frac{G_2}{G_1} + \frac{G_3}{G_1} + \cdots + \frac{G_R}{G_1} \tag{10-14}$$

But from (8-16),

$$\frac{S_o}{N'_o} = \frac{S_i}{kT_{\mathrm{op}}} \tag{10-15}$$

Comparing (10-13) to (10-15),

$$T_{\mathrm{op}} = (T_g + T_e)\mathscr{R}' \tag{10-16}$$

The result in (10-16) is not unexpected. The operating noise temperature, which measures the overall noisiness of the entire system, should be higher for a multi-response system because of additional noise entering through the spurious image responses. Indeed, if we had a dual-response system and if $G_1 = G_2$, the operating noise temperature would be twice that of (8-15), and the SNR'_o would degrade by 3 dB.

In order to develop an expression for T_{op} for the radiometry case, we start again with the normalized SNR_o, recognizing that the output signal is now contributed by all responses:

$$\mathrm{SNR}'_o = \frac{S_{oT}}{N'_o} = \frac{S_{o1} + S_{o2} + \cdots + S_{oR}}{N'_o}$$

$$= \frac{S_{i1}G_1 + S_{i2}G_2 + \cdots + S_{iR}G_R}{N'_o} \tag{10-17}$$

Equation (10-17) is not very useful because it is not in terms of total *input* signal power, $S_{iT} = S_{i1} + S_{i2} + \cdots + S_{iR}$. To remedy this, we introduce a new term, called the signal gain G_s, which was defined in Section 7.8 and (7-42) as

$$G_s = \frac{S_{i1}G_1 + S_{i2}G_2 + \cdots + S_{iR}G_R}{S_{i1} + S_{i2} + \cdots + S_{iR}} \tag{10-18}$$

Using (10-18) and (10-8) in (10-17) yields

$$\mathrm{SNR}'_o = \frac{(S_{i1} + S_{i2} + \cdots + S_{iR})G_s}{N'_o}$$

$$= \frac{S_{iT}G_s}{k(T_g + T_e)(G_1 + G_2 + \cdots + G_R)} \tag{10-19}$$

As before in (10-15), we let

$$\mathrm{SNR}'_o = \frac{S_{iT}}{kT_{\mathrm{op}}} \tag{10-20}$$

and by comparing it to (10-19) we get

$$T_{\mathrm{op}} = \frac{(T_g + T_e)(G_1 + G_2 + \cdots + G_R)}{G_s} \tag{10-21}$$

which is a rather noninformative relationship. However, when the simplifying assumption is made that the input signal power is the same in all responses (usually true in practice), that is, $S_{i1} = S_{i2} = \cdots = S_{iR} \equiv S_i$, then for R responses we have

$$G_s = \frac{S_i(G_1 + G_2 + \cdots + G_R)}{RS_i} = \frac{G_1 + G_2 + \cdots + G_R}{R}$$

and

$$S_{iT} = RS_i$$

and therefore from (10-20) we have

$$\mathrm{SNR}'_o = \frac{S_i}{k(T_g + T_e)} \tag{10-22}$$

which is the same as (8-15).

Equation (10-22) says that if we have a multi-response system with spurious images which contribute noise and degrade the SNR_o, it is possible to restore

the single-response SNR_o value by applying the signal uniformly to all responses instead of applying it to one response only. This is precisely the reason why in radiometry receivers the RF input is left wide open to allow the "signal," which is now broadband noise, to cover all responses of the receiver. An important improvement in SNR_o is realized in this manner.

10.3.   NOISE FACTOR IN MULTI-RESPONSE SYSTEMS

Many receiver systems have multiple responses due to superheterodyne stages. Most common is a response at the image frequency on the other side of the local oscillator frequency. It is apparent that additional responses will add noise to the system. To account for this, the Friis' definition of noise factor (9-3) takes the input SNR to include input noise *only in the band(s) occupied by the signal*. The IRE definition (9-2) accounts for it by referring to "that portion of the output noise engendered by the input termination which appears in the output via the *principal-frequency* transformation(s) of the system." Thus the unused images are excluded.

This leads to the situation where the noise factor of a multi-response receiver depends on its usage, or specifically, on the type of signal being received. For example, when a multi-response receiver is used in a radar or a communication system, the signal occupies *one* response* only and there is no signal in the image response. Therefore, in accordance with the above definitions, the available input noise power term in the denominator of (9-2) or (9-3) must be taken for *one* response only. We thus obtain what is called the single-sideband (SSB) noise factor, sometimes also called the radar noise factor. Using (10-5) we can write

$$F_{\text{SSB}} = \frac{N_o}{G_1 k T_0 B_1} = \frac{k(T_0 + T_e)\sum_{r=1}^{R} G_r B_r}{G_1 k T_0 B_1}$$

$$= \left(1 + \frac{T_e}{T_0}\right)\left(1 + \frac{G_2 B_2}{G_1 B_1} + \frac{G_3 B_3}{G_1 B_1} + \cdots + \frac{G_r B_r}{G_1 B_1} + \cdots + \frac{G_R B_R}{G_1 B_1}\right)$$

$$\tag{10-23}$$

$$= \left(1 + \frac{T_e}{T_0}\right)\mathscr{R} \tag{10-24}$$

---

*It is important to understand clearly the distinction between the terms multi-response and multi-sideband. The term "multi-response" indicates the total number of input/output responses (denoted by R) in a given transducer, regardless of how many are actually used in signal transmission. The term "multi-sideband" (denoted by i) describes how many responses contain the desired signal.

where T_e, G_r, B_r, and $\mathscr{R}$ are as defined in Section 10.1, and R includes all responses of the receiver. Note carefully that the definition of noise factor requires the use of $T_0 = 290°\text{K}$ as the source temperature for all responses.

The term single-sideband indicates that the desired signal is present in one response or sideband only. However, N_o in the numerator of (10-23) always stands for the *total noise*, including contributions from all image responses, used or unused.

It is instructive to consider the special case of single-sideband operation using a dual-response system. This case is frequently encountered in practice where a broadband downconverter or mixer, passing both the signal and the image, is used as the front end of a receiver. Let the signal and image gains G_1 and G_2 be equal and assume that all higher-order image responses are negligible, that is, their gains G_3, G_4, and so on, are very small compared to G_1 and G_2. Assume further that $B_1 = B_2$. Then from (10-23) we obtain

$$F_{\text{SSB}} = 2\left(1 + \frac{T_e}{T_0}\right) \tag{10-25}$$

which is applicable to radar or communication system work where the desired signal resides in the so-called signal band only.

If a multi-response receiver is used for radiometry work, the legitimate signal might be broadband galactic noise, for example, which would occupy several responses, certainly the first image band, and perhaps even a number of higher-order ones if the receiver has any (Fig. 10.1). In accordance with the IRE definition of noise factor, this fact must be taken into account by expanding the denominator of (9-2) to include all those (and *only* those) responses which do contain the desired signal. Thus we get what might in general be called the multi-sideband (MSB) noise factor:

$$F_{\text{MSB}} = \frac{N_o}{G_1 k T_0 B_1 + G_2 k T_0 B_2 + \cdots + G_i k T_0 B_i}$$

$$= \frac{k G_1 B_1 (T_0 + T_e)(1 + G_2 B_2/G_1 B_1 + \cdots + G_R B_R/G_1 B_1)}{k G_1 B_1 T_0 (1 + G_2 B_2/G_1 B_1 + \cdots + G_i B_i/G_1 B_1)}$$

$$= \left(1 + \frac{T_e}{T_0}\right)\left(\frac{\mathscr{R}}{1 + G_2 B_2/G_1 B_1 + \cdots + G_i B_i/G_1 B_1}\right) \tag{10-26}$$

Equation (10-26) is a very general one and allows for the case where the multi-response transducer has a total of R responses, but the desired signal occupies only a portion of the R, namely, i.

Looking at (10-23) and (10-26), we note that in either case—single-sideband or multi-sideband operation—the earlier simple relationship derived for single-response transducers

the single-response SNR_o value by applying the signal uniformly to all responses instead of applying it to one response only. This is precisely the reason why in radiometry receivers the RF input is left wide open to allow the "signal," which is now broadband noise, to cover all responses of the receiver. An important improvement in SNR_o is realized in this manner.

10.3. NOISE FACTOR IN MULTI-RESPONSE SYSTEMS

Many receiver systems have multiple responses due to superheterodyne stages. Most common is a response at the image frequency on the other side of the local oscillator frequency. It is apparent that additional responses will add noise to the system. To account for this, the Friis' definition of noise factor (9-3) takes the input SNR to include input noise *only in the band(s) occupied by the signal*. The IRE definition (9-2) accounts for it by referring to "that portion of the output noise engendered by the input termination which appears in the output via the *principal-frequency* transformation(s) of the system." Thus the unused images are excluded.

This leads to the situation where the noise factor of a multi-response receiver depends on its usage, or specifically, on the type of signal being received. For example, when a multi-response receiver is used in a radar or a communication system, the signal occupies *one* response* only and there is no signal in the image response. Therefore, in accordance with the above definitions, the available input noise power term in the denominator of (9-2) or (9-3) must be taken for *one* response only. We thus obtain what is called the single-sideband (SSB) noise factor, sometimes also called the radar noise factor. Using (10-5) we can write

$$F_{\text{SSB}} = \frac{N_o}{G_1 k T_0 B_1} = \frac{k(T_0 + T_e) \sum_{r=1}^{R} G_r B_r}{G_1 k T_0 B_1}$$

$$= \left(1 + \frac{T_e}{T_0}\right)\left(1 + \frac{G_2 B_2}{G_1 B_1} + \frac{G_3 B_3}{G_1 B_1} + \cdots + \frac{G_r B_r}{G_1 B_1} + \cdots + \frac{G_R B_R}{G_1 B_1}\right)$$

(10-23)

$$= \left(1 + \frac{T_e}{T_0}\right)\mathscr{R}$$

(10-24)

*It is important to understand clearly the distinction between the terms multi-response and multi-sideband. The term "multi-response" indicates the total number of input/output responses (denoted by R) in a given transducer, regardless of how many are actually used in signal transmission. The term "multi-sideband" (denoted by i) describes how many responses contain the desired signal.

where T_e, G_r, B_r, and $\mathscr{R}$ are as defined in Section 10.1, and R includes all responses of the receiver. Note carefully that the definition of noise factor requires the use of $T_0 = 290°\text{K}$ as the source temperature for all responses.

The term single-sideband indicates that the desired signal is present in one response or sideband only. However, N_o in the numerator of (10-23) always stands for the *total noise*, including contributions from all image responses, used or unused.

It is instructive to consider the special case of single-sideband operation using a dual-response system. This case is frequently encountered in practice where a broadband downconverter or mixer, passing both the signal and the image, is used as the front end of a receiver. Let the signal and image gains G_1 and G_2 be equal and assume that all higher-order image responses are negligible, that is, their gains G_3, G_4, and so on, are very small compared to G_1 and G_2. Assume further that $B_1 = B_2$. Then from (10-23) we obtain

$$F_{\text{SSB}} = 2\left(1 + \frac{T_e}{T_0}\right)$$

(10-25)

which is applicable to radar or communication system work where the desired signal resides in the so-called signal band only.

If a multi-response receiver is used for radiometry work, the legitimate signal might be broadband galactic noise, for example, which would occupy several responses, certainly the first image band, and perhaps even a number of higher-order ones if the receiver has any (Fig. 10.1). In accordance with the IRE definition of noise factor, this fact must be taken into account by expanding the denominator of (9-2) to include all those (and *only* those) responses which do contain the desired signal. Thus we get what might in general be called the multi-sideband (MSB) noise factor:

$$F_{\text{MSB}} = \frac{N_o}{G_1 k T_0 B_1 + G_2 k T_0 B_2 + \cdots + G_i k T_0 B_i}$$

$$= \frac{k G_1 B_1 (T_0 + T_e)(1 + G_2 B_2/G_1 B_1 + \cdots + G_R B_R/G_1 B_1)}{k G_1 B_1 T_0 (1 + G_2 B_2/G_1 B_1 + \cdots + G_i B_i/G_1 B_1)}$$

$$= \left(1 + \frac{T_e}{T_0}\right)\left(\frac{\mathscr{R}}{1 + G_2 B_2/G_1 B_1 + \cdots + G_i B_i/G_1 B_1}\right)$$

(10-26)

Equation (10-26) is a very general one and allows for the case where the multi-response transducer has a total of R responses, but the desired signal occupies only a portion of the R, namely, i.

Looking at (10-23) and (10-26), we note that in either case—single-sideband or multi-sideband operation—the earlier simple relationship derived for single-response transducers, $F = 1 + (T_e/T_0)$, does not hold. It is necessary to

have additional information on the various responses before either the F_{SSB} or the F_{MSB} can be calculated. We can, however, draw one conclusion. If (10-23) is compared to (10-26), it follows that

$$F_{MSB} = \frac{F_{SSB}}{1 + G_2 B_2 / G_1 B_1 + \cdots + G_i B_i / G_1 B_1} \tag{10-27}$$

Equations (10-23), (10-24), (10-26), and (10-27) remain our basic tools in analyzing noise performance of multi-response transducers and systems. They are, however, too broad and will not serve us efficiently in bringing to light specific relevant points. Let us therefore make certain assumptions which are generally true in practice. First, the use of subscripts R and i, particularly $R > i$, may be questioned. What does $R > i$ mean? Granted, we could envision receivers having R responses with finite (perhaps even equal) gains, but with input signals occupying only some of them, $i < R$. Equation (10-26) would then apply as is. But how practical is this case? A typical broadband receiver is usually designed to have no more responses than required, and those responses would then all carry the desired signals. Clearly, then, we could in almost all practical situations limit ourselves to $R = i$. This results in a drastic simplification of (10-26) and (10-27):

$$F_{MSB} = 1 + \frac{T_e}{T_0} = \frac{F_{SSB}}{\mathscr{R}} \tag{10-28}$$

Be sure to note that we have as yet said nothing about the *value* of R. We could have a system with, say, four responses. As long as *all four* responses are used and carry legitimate signals (again, a hypothetical radiometry receiver comes to mind), the noise factor would be given by (10-28).

Let us go one step farther. A very common case is $R = 2$. That is, the transducer responds only to the signal and to the first-order image (Fig. 10.1). The response factor $\mathscr{R}$ in (10-24) is now truncated to

$$F_{SSB} = \left(1 + \frac{T_e}{T_0}\right)\left(1 + \frac{G_2 B_2}{G_1 B_1}\right) \tag{10-29}$$

With $R = i = 2$, it makes more sense to call (10-26), (10-27), and (10-28) the double-sideband (DSB)* noise factor, that is,

$$F_{DSB} = 1 + \frac{T_e}{T_0} \tag{10-30}$$

*A warning is in order. The terms single sideband and double sideband have occasionally been used to designate the number of *responses* in the system. Thus according to this convention an amplifier would have a single-sideband noise factor, whereas the mixer is said to have a double-sideband noise factor. The prevailing usage is, however, as given above.

Now, for the important special case of $G_1 = G_2$ and $B_1 = B_2$, we obtain from (10-29) and (10-30):

$$F_{SSB} = 2 F_{DSB} \tag{10-31}$$

This, then, is the basis for the much-quoted 3-dB difference between the single- and double-sideband noise factors. As we already noted in connection with (10-16), this ratio is not unique, being only a special case. It is easy to see that if the signal and the image gains G_1 and G_2 are *unequal*, which in practice they could well be, or if the noise bandwidths B_1 and B_2 differ, then the ratio F_{SSB}/F_{DSB} will not be 2 (or 3 dB) either.

As an illustrative example consider what happens when $G_2 = \frac{1}{3} G_1$, that is, the two gains are different by 4.8 dB. If we keep $B_1 = B_2$, then from (10-27) we have

$$F_{DSB} = \frac{F_{SSB}}{1 + 1/3} = \frac{F_{SSB}}{1.33} \tag{10-32}$$

or a ratio of only 1.25 dB. We could be off in the other direction also. Suppose, for example, our receiver had four active responses such that $G_1 = G_2 = 2G_3 = 2G_4$. From (10-28) the multi-sideband noise factor would be

$$F_{MSB} = 1 + \frac{T_e}{T_0} \tag{10-33}$$

as before. The F_{SSB}, on the other hand, from (10-23) becomes

$$F_{SSB} = \left(1 + \frac{T_e}{T_0}\right)\left(1 + 1 + \frac{1}{2} + \frac{1}{2}\right) = 3 F_{DSB} \tag{10-34}$$

These examples serve as a fair warning that a multi-response transducer can indeed have two significantly different noise factors, depending on its use. More importantly, the above results are not some kind of mathematical quirks, but have a real physical basis which is not difficult to understand. Consider a receiver, about to process an incoming narrowband signal residing in one response only—a typical single-sideband situation. Let the input SNR of this signal be SNR_i. Assume further that the receiver is perfect and contributes no noise of its own. Had this receiver been a *single-response* transducer, then clearly

$$SNR_i = SNR_o$$

and its noise factor would be unity. If, however, the receiver has an image response and if the two gains G_1 and G_2 are equal, the output SNR, SNR_o, would immediately be halved on account of unwanted noise in the unused but

unavoidable image band (we assume here that the image is not open or short circuited at the input of the receiver), which is also converted into the output band (IF frequency) whether we like it or not. The end result is a *nonunity* noise factor, specifically,

$$F_{SSB} = \frac{SNR_i}{\frac{1}{2}SNR_i} = 2 \quad or \quad 3 \text{ dB} \tag{10-35}$$

despite the absence of any internally generated noise in our ideal receiver.

Let us now return to (10-28) for a moment. We saw that if there is a broadband signal which covers *all* responses, then the F_{MSB} becomes

$$F_{MSB} = 1 + \frac{T_e}{T_0}$$

regardless of the number of responses. But this expression is identical to (9-7). It may seem strange that (9-7), originally derived for single-response transducers, should also hold for the multi-response case. This is not hard to understand if one considers the definition of noise factor as expanded in the beginning of this section. Heuristically, we have i signals at the output, one from each response. But we also have i noise powers, one from each response. The SNR_o is therefore

$$SNR_o = \frac{iS_o}{iN_o} = \frac{S_o}{N_o} \tag{10-36}$$

If we had only a single response, the ratio would still be the same even though i is now equal to unity. The result in (10-28) is therefore quite general, and holds even if $G_1 \neq G_2 \neq G_3 \cdots$ and $B_1 \neq B_2 \neq B_3 \cdots$ as can be seen from (10-26) when $R = i$.

10.4.   OUTPUT NOISE AND NOISE FACTOR

Section 10.1 explored the relationship between the total available output noise N_o and the input noise temperature T_e. This section will consider similar relationships between N_o and the various noise factors. As we have stated before, the noise factor as a parameter by itself is not the end goal of the analysis, but merely a means to an end, although an important one. The real end goal is the output signal-to-noise ratio SNR_o. We are therefore ultimately interested in the total output noise power N_o from which the SNR_o can be calculated. For *single-response* transducers, N_o is given either by

$$N_o = kGB(T_g + T_e) \tag{10-37}$$

or

$$N_o = kGB[T_g + T_0(F - 1)] = kGBT_0\left(\frac{T_g}{T_0} + F - 1\right) \tag{10-38}$$

These follow from (8-3) and (9-8). The noise models corresponding to the given expressions appeared in Figs. 8.2 and 9.2.

For *multi-response* transducers the situation is more complicated. Not only must one take into account the added noise converted to the output from image responses, but if the transducer was characterized by its noise factor it is necessary to know what kind of noise factor it is.

We already saw in (10-5) how the relationship between N_{no} and T_e was modified by the presence of multiple responses. Now suppose that instead of T_e we had been given the double-sideband noise factor F_{DSB}. How do we calculate the output noise of the transducer from this parameter? Could we start with (10-26) and simply solve for N_o? The answer is that this would indeed be a valid procedure as long as the source (generator) temperature for the particular setup, for which the N_o is sought, is equal to $T_0 = 290°K$. Quite often the transducer is not connected to a source having its internal termination at T_0, but faces instead an antenna such that $T_a \neq T_0$. We must then fall back on (10-7). The procedure is as follows. First, assume for simplicity that all source temperatures of the individual responses are equal:

$$T_{g1} = T_{g2} = \cdots = T_{gR} \equiv T_g \tag{10-39}$$

Then, after rearranging, we get from (10-7) that

$$N_o = kG_1B_1(T_g + T_e) + kG_2B_2(T_g + T_e) + \cdots + kG_RB_R(T_g + T_e)$$

$$= kG_1B_1\left[T_g + T_0\left(\frac{T_e}{T_0} + 1 - 1\right)\right]\left(1 + \frac{G_2B_2}{G_1B_1} + \cdots + \frac{G_RB_R}{G_1B_1}\right)$$

$$= kG_1B_1T_0\left(\frac{T_g}{T_0} + (F_{DSB} - 1)\right)\left(1 + \frac{G_2B_2}{G_1B_1}\right) \tag{10-40}$$

Here we used (10-30) and, more importantly, abbreviated the last term in parentheses to just G_2B_2 because of the argument advanced earlier in connection with (10-28), (10-30), and the case where $R = i = 2$. If the given noise factor had been an F_{MSB}, that is, $R = i > 2$, then the last expression in parentheses would be the complete response factor $\mathscr{R}$. If this is confusing, it is suggested that Section 10.3, and particularly the first footnote, be carefully reviewed.

Equation (10-40) is the multi-response equivalent of (10-38) and gives the desired answer in terms of the transducer parameters and the F_{DSB}. We could,

of course, be given the single-sideband noise factor F_{SSB} instead. Even so, N_o can still be calculated from (10-40) except that we must first convert the given F_{SSB} into the F_{DSB} using either (10-27) or (10-31), depending on the number of responses as explained earlier.

As in the case of T_e and Fig. 10.2, we can develop an analogous noise model using F_{MSB} (or F_{DSB}). It follows directly from (10-40) when the latter is rearranged:

$$N_o = k\left[T_g + T_0(F_{MSB} - 1)\right](G_1 B_1 + G_2 B_2 + \cdots + G_R B_R) \quad (10\text{-}41)$$

The resultant model is shown in Fig. 10.3.

One may wonder why the foregoing discussion appears to be biased in favor of the F_{DSB} rather than F_{SSB}. The reason is that the most prevalent method of measuring noise factors—the broadband noise-source method—inherently gives the F_{MSB} as we shall see in Chapter 11.

In conclusion we note that linear *single-response* receiver systems require the following three pieces of information for system noise evaluation (see also Section 9.9):

(a) Noise factor (or the input noise temperature).

(b) Noise bandwidth.

(c) Noise temperature of the source.

In *multi-response* systems one must also know the relative gains and noise bandwidths of the other responses. This is enough if the analysis is in terms of a system noise temperature. However, if the noise factor is required, it is also necessary to know the distribution of the input signal within the various responses, so that the correct expression for noise factor is used. Be careful with quoted noise factors. Always make sure you understand which noise factor is meant.
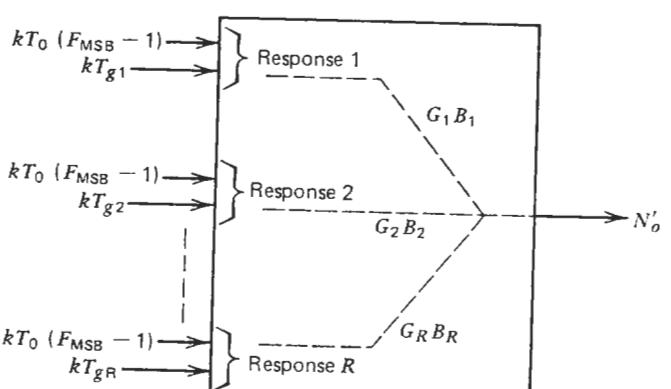
FIGURE 10.3. Noise model of a multi-response transducer using the noise factor.

10.5. CASCADING OF MULTI-RESPONSE TRANSDUCERS

The two cascading formulas, (8-27) and (9-11), were derived in a straightforward manner by summing all noise contributions at the final output of the cascaded chain of transducers. The most important result was that the gain of a given stage, if sufficiently high, $G \gg 1$, could render the noise contribution of the succeeding stages negligible. This enables us to realize an overall system noise temperature or noise factor which is essentially equal to that of the first stage. We need only to provide a high-quality first stage of sufficient gain, and the overall noise factor would follow suit.

When the system contains multi-response transducers, the situation is more complicated. Equations (8-27) and (9-11) no longer apply and must be modified. This is not surprising for the noise from image responses is expected to have some effect on the cascaded noise performance. We shall now investigate several cases where single- and multi-response transducers are connected in tandem. Although in practice a multi-response transducer is commonly a frequency converter or mixer, the present discussion will be in terms of general "black boxes". The specific case of mixers will be treated in Section 10.6. To simplify the discussion, we shall assume that there are only two responses, signal (denoted by a subscript 1) and image (denoted by a 2). That is, we again assume as we did earlier that G_3, G_4, and so on, are very small compared to G_1 and G_2. This situation is often encountered in practice. The extension of the analysis to an arbitrary number of responses is not difficult, but the equations become unwieldly and little additional insight is gained. As a further convenience, let the source temperature T_g be the same for both responses, that is, $T_{g1} = T_{g2} \equiv T_g$. Last, as a shorthand notation, denote a general dual-response transducer by the symbol DRT throughout this section.

Broadband Amplifier and the DRT

Consider the circuit shown in Fig. 10.4, where a broadband amplifier precedes a DRT. Let the bandwidth of the amplifier be wide enough to include both responses of the DRT, but allow initially the two amplifier gains G_{a1} and G_{a2} to be unequal. Assume further that the DRT contains an output filter which sets the operating bandwidth of the entire group, that is, $B_1 = B_2 = B_b \equiv B$, and that the signal and image gains of the DRT are equal,*

$$G_{b1} = G_{b2} \equiv G_b \quad (10\text{-}42)$$

The total available noise power at the output terminals a–a' is given by

$$N_o = \underbrace{kB\left[(T_g + T_{ea})G_{a1}G_b + T_{eb}G_b\right]}_{\text{total noise in signal response}} + \underbrace{kB\left[(T_g + T_{ea})G_{a2}G_b + T_{eb}G_b\right]}_{\text{total noise in image response}}$$

$$(10\text{-}43)$$

*In a practical mixer this would mean that the conversion loss is the same for signal and image.
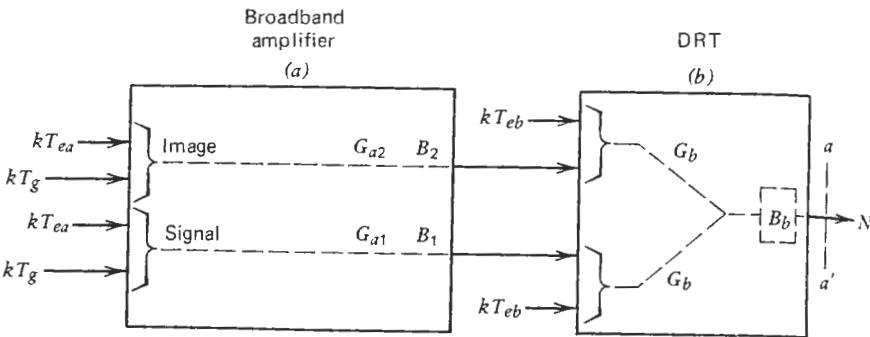
FIGURE 10.4.   Noise relationships in a cascaded combination of a broadband amplifier and a dual-response transducer.

where T_{ea} is the input noise temperature of the amplifier; T_{eb} the input noise temperature of the DRT; G_{a1} and G_{a2} the gain of the amplifier at the signal and image frequencies, respectively; G_b the gain of the DRT (assumed to be the same at signal and image frequencies); B the system noise bandwidth; and T_g the noise temperature of source termination (same for signal and image).

The overall noise factor F_T can now be calculated from (9-2). All that is needed is to set T_g in (10-43) equal to T_0 in accordance with the definition of noise factor, and to use the correct expression in the denominator of (9-2). The latter is simply the source noise as measured at the output $a-a'$, but its formulation, as we saw in Section 10.3, depends on how the amplifier/DRT combination is used. If there is usable information in the signal response only, then

$$(N_i)_o = kT_0 G_{a1} G_b B \tag{10-44}$$

In case of a broadband dispersed signal which covers both responses,

$$(N_i)_o = kT_0 (G_{a1} + G_{a2}) G_b B \tag{10-45}$$

The overall single-sideband noise factor is now calculated in accordance with (9-2) as the ratio of (10-43) to (10-44):

$$(F_T)_{\text{SSB}} = \frac{kG_b B[(T_0 + T_{ea})(G_{a1} + G_{a2}) + 2T_{eb}]}{kBG_b T_0 G_{a1}}$$

$$= F_a \left(1 + \frac{G_{a2}}{G_{a1}}\right) + \frac{2(T_{eb}/T_0) + 2 - 2}{G_{a1}}$$

$$= F_a \left(1 + \frac{G_{a2}}{G_{a1}}\right) + \frac{2[(F_b)_{\text{DSB}} - 1]}{G_{a1}} \tag{10-46}$$

where (9-7) and (10-30) were also used.

If the gain of the amplifier is the same at signal and image, then

$$G_{a1} = G_{a2} \equiv G_a \tag{10-47}$$

and (10-46) can be simplified to

$$(F_T)_{\text{SSB}} = 2F_a + \frac{2[(F_b)_{\text{DSB}} - 1]}{G_a}$$

$$= 2\left(F_a + \frac{(F_b)_{\text{DSB}} - 1}{G_a}\right) \tag{10-48}$$

$$= 2F_a + \frac{(F_b)_{\text{SSB}} - 2}{G_a} \tag{10-49}$$

Equation (10-49) follows from (10-31). The meaning of (10-48) and (10-49) should be fully appreciated. Equation (10-48) states that if one were given the *double-sideband* noise factor of a DRT, that is, the $(F_b)_{\text{DSB}}$, and if the amplifier/DRT combination were used in a situation where usable signal is present in one response only, then the SNR_o calculations must be based on (10-48). If, on the other hand, the *single-sideband* noise factor of the DRT is available, while the overall application of the system remains as before, then (10-49) applies.

The overall double-sideband noise factor $(F_T)_{\text{DSB}}$ applies when usable signals are present in both responses as in radiometry work. We again use (9-2) and (10-43), but the denominator is now (10-45). The result becomes

$$(F_T)_{\text{DSB}} = \frac{kG_b B[(T_0 + T_{ea})(G_{a1} + G_{a2}) + 2T_{eb}]}{kBG_b T_0 (G_{a1} + G_{a2})}$$

$$= \left(1 + \frac{T_{ea}}{T_0}\right) + \frac{2T_{eb} + 2 - 2}{T_0(G_{a1} + G_{a2})}$$

$$= F_a + \frac{2(F_b)_{\text{DSB}} - 1}{G_{a1} + G_{a2}} \tag{10-50}$$

which for $G_{a1} = G_{a2} = G$ becomes

$$(F_T)_{\text{DSB}} = F_a + \frac{(F_b)_{\text{DSB}} - 1}{G_a} \tag{10-51}$$

$$= F_a + \frac{(F_b)_{\text{SSB}} - 2}{2G_a} \tag{10-52}$$

Equations (10-48) and (10-52) are quite interesting. Note particularly the following:

(a) As we surmised, the cascade equation can under some conditions differ appreciably from (9-11).

(b) Contrary to expectations, the cascading, when multi-response circuits are involved, may not reduce the overall single-sideband noise factor below 2, no matter what the first-stage gain or noise factor are.

(c) If (10-49) and (10-52) seem unfamiliar, keep in mind that they could be stranger yet if the signal and the image gains had been unequal and situations such as (10-32) and (10-34) had prevailed. As we saw earlier, the validity of (10-31) is by no means always guaranteed. This should be remembered throughout the following discussion.

Note carefully that we are at liberty to use *either* the F_{DSB} or the F_{SSB} of a DRT, whichever is available, to calculate *either* the overall $(F_T)_{DSB}$ or the overall $(F_T)_{SSB}$, whichever is required. One does not imply the other. The particular measurement technique or available data sheet determines the available noise factor for the DRT. The ultimate system application determines which overall noise factor should be calculated.

Narrowband Amplifier and the DRT

Now consider a narrowband amplifier such that the image response falls outside the amplifier passband (Fig. 10.5). In other words, the amplifier gain at the image frequency, G_{a2}, is very much less than that at signal frequency, G_{a1}. However, assume the signal and image gains of the DRT to be equal as before. The term $kB(T_g + T_{ea})G_{a2}G_b$ in (10-43) is now negligible, but we have to be careful before proceeding. It is true that the image contribution $kB(T_0 + T_{ea})$ through the amplifier has now been greatly reduced. However, the input of the DRT still looks at something at the image frequency. What does it see? The question is not trivial, because the answer defines the status of the image channel and its noise contribution in this new configuration. In addition, we must make sure that the value of the noise factor assigned to the DRT, that is, F_b, is consistent with the actual output impedance of the amplifier seen when in series with the latter. It will be remembered that the rules governing noise calculations in cascaded networks are rather strict—both the available power gain G and the noise factor F of the n-th transducer must be based on the output impedance of the $(n-1)$th transducer, which now acts as the generator to the n-th stage. In view of Section 9.9, where the dependence of the transducer's noise performance on the generator impedance was discussed, it is clear that a detailed answer to our question requires more than a casual analysis. To simplify matters and to illustrate the main point, let us therefore assume that a broadband match covering both the signal and the image
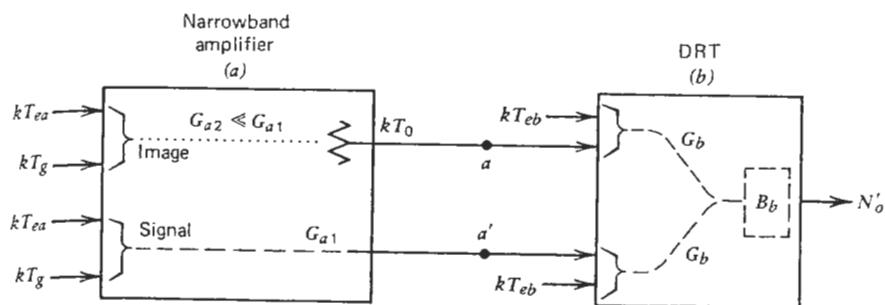
FIGURE 10.5.   Noise relationships in a cascaded combination of a narrowband amplifier and a dual-response transducer.

responses exists between the amplifier output and the DRT input at a–a'. Moreover, assume that the noise factor F_b of the DRT is referenced to this same impedance. We can then rewrite (10-46) for the narrowband case in hand as follows:

$$(F_T)_{SSB} = \frac{kBG_b[(T_0 + T_{ea})G_{a1} + T_0 + 2T_{eb}]}{kBG_bT_0G_{a1}} \tag{10-53}$$

$$= F_a + \frac{2[(T_{eb}/T_0) + 1] - 1}{G_{a1}} = F_a + \frac{(F_b)_{SSB} - 1}{G_{a1}} \tag{10-54}$$

Note that the second term T_0 in the numerator of (10-53) is the noise contributed by the matched output impedance of the amplifier (assuming it to be at T_0, of course). This is what the image response of the DRT sees looking back toward the amplifier.

The result (10-54) is interesting because it states that a high G_{a1} leads to a low system noise factor despite the fact that the system is used in a single-sideband mode. A little reflection will show that this is not surprising, for, after all, we have eliminated the amplifier contribution of the image response.

But how do we realize this situation in practice? There are two possibilities. The first one is to build just such a narrowband amplifier with a matched output stage. The second one is to use a broadband amplifier, but to insert a filter at its *output* to block the amplified image channel. Such a filter has a highly reactive impedance at the rejected image frequency. In order to avoid problems in defining F_b for the DRT with a reactively terminated image (see Section 10.6), a ferrite isolator can be placed after the filter as shown in Fig. 10.6. Our equations will then be consistent.

Equation (10-54) should thus be compared to (10-49) of the previous section to appreciate the difference. There is no expression analogous to (10-51) or (10-52) for the narrowband case because the use of a narrowband amplifier already implies that only the single-sideband operation is of interest.
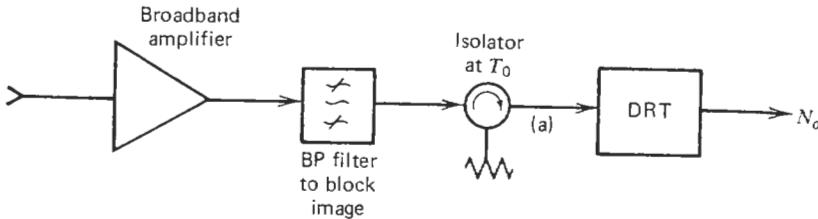
FIGURE 10.6. The use of an image filter in a cascaded system.

Several Amplifiers and the DRT

Frequently, the DRT is preceded by several amplifiers or other single-response components (Fig. 10.7). The extension of the cascade formula to this case is simple. We merely sum all (uncorrelated) contributions at the output, then divide the result by either (10-44) or (10-45). The details are left as an exercise, but the results are:

$$(F_T)_{SSB} = 2\left(F_a + \frac{F_b - 1}{G_a} + \frac{(F_c)_{DSB} - 1}{G_a G_b}\right) \quad (10\text{-}55)$$

and

$$(F_T)_{DSB} = F_a + \frac{F_b - 1}{G_a} + \frac{(F_c)_{DSB} - 1}{G_a G_b} \quad (10\text{-}56)$$

These are seen to be analogous to (10-48) and (10-51).

The DRT Followed by Amplifier

Figure 10.8 depicts the reverse of Fig. 10.4—a dual-response transducer preceding a single-response component. In practice, this configuration corresponds to a mixer/IF amplifier combination. The calculation of the total
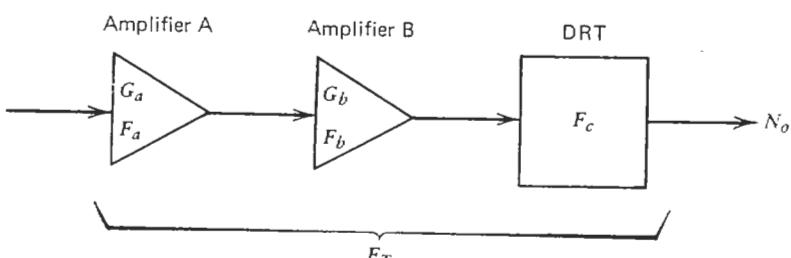
FIGURE 10.7. Dual-response transducer preceded by several amplifier stages.
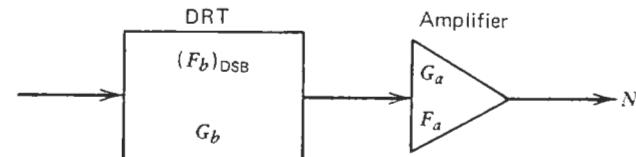
FIGURE 10.8. Amplifier preceded by a dual-response transducer.

cascaded noise factor starts as before with the summation of total available noise N_o at the final output. Thus, from Fig. 10.8,

$$N_o = 2k(T_{eb} + T_0)G_a G_b B + T_{ea} k G_a B \quad (10\text{-}57)$$

Using (10-57) in conjunction with either (10-44) or (10-45), and assuming the signal and image gains of the DRT are again equal ($G_{b1} = G_{b2} = G_b$), we can readily get the $(F_T)_{SSB}$ and $(F_T)_{DSB}$:

$$(F_T)_{SSB} = 2\left(1 + \frac{T_{eb}}{T_0}\right) + \frac{T_{ea}/T_0}{G_b} = 2(F_b)_{DSB} + \frac{F_a - 1}{G_b}$$

$$= (F_b)_{SSB} + \frac{F_a - 1}{G_b} \quad (10\text{-}58)$$

and

$$(F_T)_{DSB} = \left(1 + \frac{T_{eb}}{T_0}\right) + \frac{T_{ea}/T_0}{2G_b}$$

$$= (F_b)_{DSB} + \frac{F_a - 1}{2G_b} \quad (10\text{-}59)$$

As seen from these formulas, the procedure of deriving them is quite simple once it is properly understood. The key in each case is the correct summation of the final output noise N_o. The rest is a straightforward application of (9-2).

Conclusions

We could think of increasingly exotic combinations of single- and dual-response transducers, but this would be hardly worthwhile. The basic technique of summing all noise contributions at a reference point, commonly the output, and then calculating the required noise parameters from it, has already been amply illustrated. It remains to summarize the findings.

First, it should be clear by now that the classical cascade formula (9-11) cannot be applied blindly when multi-response transducers are involved. Noise introduced through spurious image channels may appear at the output in full strength, and may result in an unexpected increase in the total noise factor, (10-48) and (10-49). Second, we must be careful which noise factor to use when dealing with multi-response transducers. Depending on the choice, the cascade formula will change. Third, if the input amplifier provides proper filtering at the image frequency, the effect of image channel noise can be suppressed, (10-54). Fourth, remember that most equations in this section were based on assumptions which may not always hold. In particular, the validity of (10-42) and (10-47) must be examined in each case. If these expressions do not hold, then (10-31) cannot be used either and subsequent equations starting with (10-43) will have to be rederived.

10.6.   CONSIDERATIONS OF PHYSICAL FREQUENCY CONVERTERS

Previous sections were devoted to the general properties of multi-response transducers. It was shown how the concepts of input noise temperature, operating noise temperature, noise factor, noise modeling, and cascading are applied to such systems where multiple responses couple different input frequencies to a common output frequency or band of frequencies. Useful relationships were derived using only a generalized "black box" formulation of the transducers. Several assumptions were then made to illustrate specific points, but these assumptions were in line with conditions frequently encountered in practice, and did not restrict the practical value of the analysis unduly. The emphasis was on the clarity of presentation. At times this was done at the expense of rigor, to be justified by the stated end goal of the book—to describe the fundamentals of noise analysis as applied to engineering practice.

The same philosophy will also apply to this section where we shall study the physical aspects of multi-response transducers, how they are realized in practice, and what their peculiarities are. After all, the discussion of T_e and G in an abstract, theoretical case takes us only so far. It is of much greater interest to learn how these quantities relate to measurable parameters of a physical mixer, which one may be asked to evaluate or to consider for possible system use. Similarly, it is important to know what a given assumption would mean in practical terms. The emphasis in this section will, however, be on an overview rather than on a detailed analysis. The reason for this is that our goal is not a detailed design of mixers, but rather an analysis of terminal characteristics of such devices, and how they affect the operation of the system. It is assumed that the designer has done the job and we are given a piece of hardware. What would we expect? True, there will be a set of specifications, but they may not tell the whole story. Or perhaps further interpretation is necessary before we can really tell whether a given mixer is usable in the given application. Either way, it is of great value to know something about practical mixers and of the

degree to which they would approach the ideal. What we are therefore attempting to do is threefold:

(a) To relate theory to practice.
(b) To familiarize the reader with the various concepts and terms associated with frequency converters.
(c) To develop the required background and some of the mathematical tools with which serious students could tackle the rigorous and more advanced treatments of this topic.

General Description

When one refers to multi-response transducers, it is the common downconverting mixer that comes to mind. The simplest version of such a mixer is a nonlinear conductance, typically a Schottky-barrier diode, either in series or in shunt with the line. If a large-amplitude ac voltage $v_p = V_p \cos \omega_p t$ (called the local oscillator or pump) is impressed across this device, the latter is switched alternately between a very low-resistance forward-bias state and a very high-resistance reverse-bias state. The pumped diode thus resembles a switch, opening and closing at the local oscillator rate ω_p. The first consequence is the generation of a large number of harmonics of ω_p, $m\omega_p$, where m is any positive integer. Second, if another ac signal $v_s = V_s \cos \omega_s t$ is impressed across the same diode, then a large number of mixing products will also appear. The principal ones are the difference frequency $\omega_p - \omega_s$, usually called the IF frequency; the sum frequency $\omega_p + \omega_s$; and the image frequency $2\omega_p - \omega_s$ (see Fig. 10.1). If $V_p \gg V_s$, it can be shown that the relationship between the input v_s and the output v_0 becomes a linear one despite the highly nonlinear mechanism involved in the frequency conversion. As a consequence, the local oscillator does not enter the analysis explicitly, having become an implicit part of the frequency conversion process. For the same reason, the local oscillator terminals will not be shown as a separate port in the illustrations to follow.

There are various ways of numbering these mixing products, each tailored to facilitate a particular analysis. From now on we shall adopt the concise notation of Saleh [6], also used by Held and Kerr [1] and Kerr [2, 3], and others. Thus the frequency components are designated by

$$\omega_m = \omega_0 + m\omega_p \qquad (10\text{-}60)$$

where $m = \pm 1, \pm 2$, and so on. Thus the upper and lower sidebands around the local oscillator (LO) frequency, which are the signal and the principal image (or depending on the usage, in the reverse order) frequencies, become ω_1 and ω_{-1}, respectively, where ω_{-1} is simply folded about $\omega = 0$ to yield the lower sideband as shown in Fig. 10.1. One advantage of such a notation is that the $\pm m$ are related to the Fourier coefficients of the periodic LO current waveform at the mixer diode.

Clearly, the mixer just described, is quite crude. It has no provisions for separating the various mixing products, the signal or the local oscillator, nor has the question of terminating impedances seen by all these frequency components been addressed. A practical mixer, even the simplest kind, must therefore be designed with quite a few additional considerations and requirements in mind. In particular, the designer must, as a minimum, consider the following:

(a) Filters to separate ω_1, ω_p, and ω_0.
(b) Terminations to be provided for $\pm m\omega_p$ and all higher-order images.
(c) Termination seen by the principal image ω_{-1}.
(d) Input match at the signal and local oscillator frequencies.
(e) The characteristics of the particular diode.
(f) Interface between the mixer output ω_0 and the IF amplifier input.

If in addition it is kept in mind that physical mixer diodes are neither ideal switches with infinite and zero conductances at the two extreme bias states, nor are they rarely pumped to switch instantaneously from one state to the other, then the complexity of the problem becomes apparent. Figure 10.9 shows the generally accepted equivalent circuit for a Schottky-barrier mixer diode. There is, of course, the nonlinear conductance $g_d(v)$, but we have additionally a series resistance R_s, called the spreading resistance, and a nonlinear, voltage-
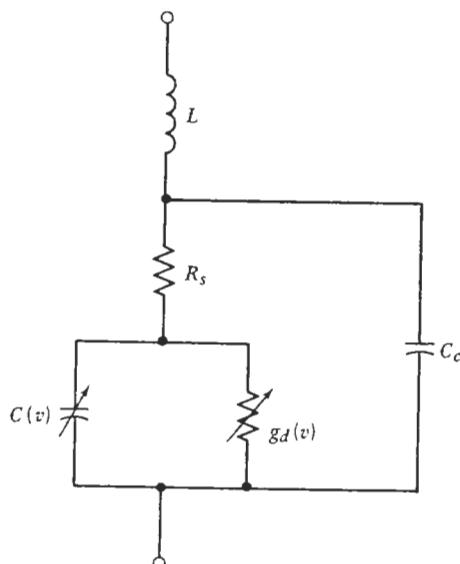
FIGURE 10.9. Equivalent circuit of a solid-state diode.

variable barrier or junction capacitance $C(v)$ given by

$$C(v) = C_0\left(1 - \frac{v}{\phi}\right)^{-\gamma} \tag{10-61}$$

where γ is $\frac{1}{2}$ or $\frac{1}{3}$ depending on the doping profile, and ϕ is the contact potential. Although R_s is usually assumed to be constant with frequency and γ is assumed to be constant with the applied pumping voltage v_p, neither of these assumptions holds in a refined analysis [1]. The diode package contributes additional spurious elements—lead inductance L and the case capacitance C_c.

The nonlinear conductance $g_d(v)$ of a Schottky diode has the following exponential voltage-current characteristic:

$$i_d = I_s\left[\exp\left(\frac{qv}{\eta kT}\right) - 1\right] = I_s(e^{\alpha v} - 1) \tag{10-62}$$

where I_s is the saturation current and $\alpha = q/\eta kT$. η is the diode ideality factor, typically 1.05–1.4 depending on the particular diode [7, 8]. For an ideal Schottky-barrier diode, η approaches unity. For point-contact diodes, $\eta > 1$ [9]. T is the physical temperature of the diode. Other terms have been defined earlier. At room temperature, $\alpha \approx 35v^{-1}$. From (10-62) we get the incremental diode conductance

$$g_d = \frac{di_d}{dv} = \alpha(i_d + I_s) \tag{10-63}$$

Clearly, for high forward-bias values, g_d is quite high while in the reverse direction the opposite is true. This is as far into the device physics as we need to go at this time. The attention will now be turned to important parameters of a typical mixer.

Conversion Loss

The most important attribute of any mixer is its conversion loss, usually denoted by L. The conversion loss is simply the reciprocal of gain G used in the previous sections. Since it is recognized that for passive mixer circuits* the conversion loss is always greater than unity, $L > 1$, it makes sense to use L rather than G in dealing with mixers, to avoid fractional numbers.

Thus, given an available input power at some input frequency ω_1, the available output power at some translated frequency is always less than the input power. This is true for both types of frequency converters: upconverters, which convert a low (IF) frequency to a higher (RF) frequency, or downconverters, which achieve the opposite. It is not entirely trivial to ask the question: why is there conversion loss in the first place? Granted, even the best physical

---

*If the nonlinear capacitance of the diode (Fig. 10.9) is taken into account [10], it can be shown that the mixer could have conversion gain.

mixer has resistive losses in its structure and in the diode itself (R_s). A portion of the input power is therefore expected to be lost. Suppose, however, that we had a perfect mixer with an ideal switch-like diode and no resistive losses anywhere. What then? Would the conversion loss of this mixer now be unity? Unfortunately no, or perhaps better, not automatically. It does not require much reflection to understand qualitatively why this is so. As we saw earlier, a basic mixer generates not just the desired output frequency, but a very large number of other mixing products of the type $\omega_m = \omega_0 + m\omega_p$ as well. The given input power will thus be divided among the many spurious and largely unwanted frequency components. If these frequencies are able to see a resistive load, as would be the case in a wide-open broadband mixer, then the power associated with them will be irretrievably lost, and the desired output frequency ω_0 inherits that much less of the input power.

It is not difficult to express this quantitatively. Kelly [11] has analyzed this using an idealized diode with no parasitics, and with infinite and zero conductances in the forward and reverse directions, respectively. He further assumes an ultrabroadband mixer with the signal and all image responses resistively terminated (matched condition). His reasoning is repeated here to illustrate the point. Remembering that the role of the local oscillator is to provide switching action in the mixer diode, the latter can be modeled ideally as a switch that opens and closes instantaneously at the LO rate ω_p (Fig. 10.10a). The signal current $(V_s/R)\cos\omega_1 t$ then becomes as shown in Fig. 10.10b. This waveform can be considered as a product of the original input current and a train of rectangular pulses of unit amplitude, $f(t)$, shown in Fig. 10.10c. The procedure is to expand $f(t)$ in a Fourier series and then to separate out the term corresponding to the output frequency ω_0, that is, $\omega_0 = \omega_1 \pm \omega_p$. Thus from Kelly's work the diode current at ω_0 for minimum conversion loss is

$$i_d = -\left(\frac{2V_s}{\pi R}\right)\left(\frac{1}{2}\cos(\omega_p - \omega_1)t + \cos(\omega_p + \omega_1)t\right) \qquad (10\text{-}64)$$

The conversion loss for this broadband case is now given by the ratio of the available power from the signal source,

$$\frac{\left(V_s/\sqrt{2}\right)^2}{4R} = \frac{V_s^2}{8R} \qquad (10\text{-}65)$$

to the power dissipated in R by the ω_0 term, that is, either the $(\omega_p - \omega_1)t$ or the $(\omega_p + \omega_1)t$ term of the current:

$$\left(\frac{2V_s}{R2\sqrt{2}\,\pi}\right)^2 R = \frac{V_s^2}{2\pi^2 R} \qquad (10\text{-}66)$$

This ratio, (10-65) to (10-66), is

$$L = \left(\tfrac{1}{2}\pi\right)^2 \quad \text{or} \quad 3.92 \text{ dB} \qquad (10\text{-}67)$$
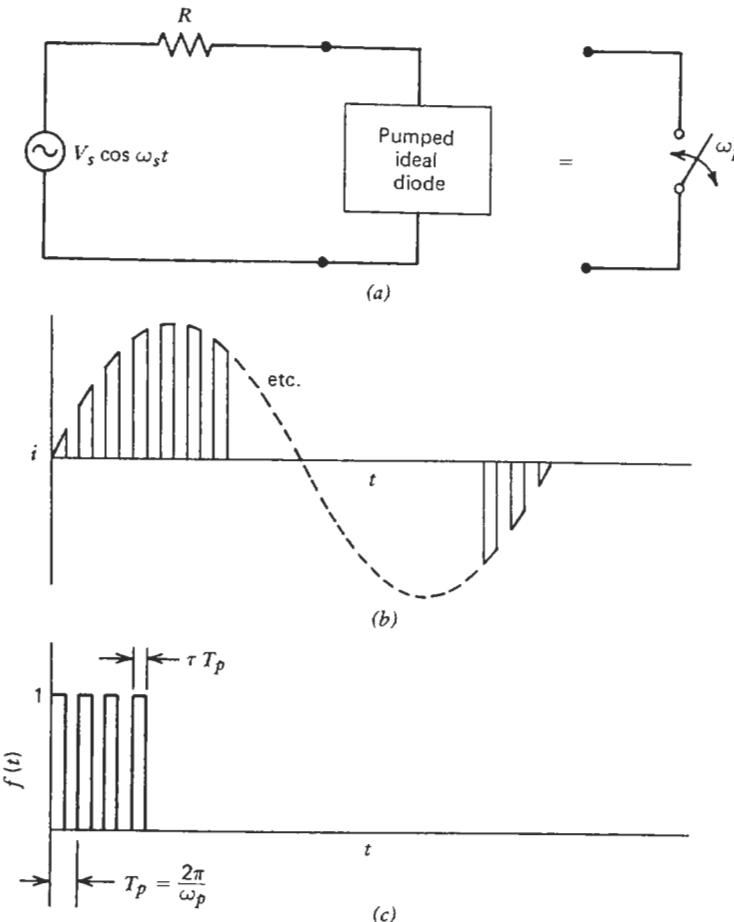
FIGURE 10.10.   Waveforms in an ideal pumped diode.

Incidentally, this minimum occurs when the diode is driven such that its conductance "duty ratio" τ (see Fig. 10.10c) is 50%. This conversion loss represents the fundamental limit for a broadband mixer with all idlers matched. Kelly's paper also shows that for an ideal double-sideband mixer, which has all frequencies other than the signal ω_1, the image ω_{-1}, and the output ω_0 reactively terminated, the theoretical conversion loss is 2 (or 3 dB). The power lost is equally divided between the mixer-generated image ω_{-1} and the power reflected at ω_1 by the mixer input. We can therefore say that the excess 0.92 dB in (10-67) represents the loading effect of all the idler frequencies when these are allowed to dissipate power in external resistances.* It should be noted that

*Kelly's analysis can be extended to a mixer with an open-circuited image by setting the second Y_{0t} in his Eq. (7) equal to zero. The minimum theoretical conversion loss for such a case then becomes unity, that is, 0 dB.

the word "external" denotes *any* resistive loss external to the idealized nonlinear conductance. These include not only the actual load resistances, but spurious losses in the diode mount as well.

Thus the conversion loss of a mixer is not some inherently fixed quantity, but remains a sensitive function of the image, idler, and output terminations [2], even for an ideal mixer. We saw how through judicious choices of the image and idler terminations the conversion loss could be minimized. Conversely, it is not difficult to go in the other direction and end up with a high conversion loss if the various terminations and/or circuit tuning are not correct.

The practical aspects of all this will be better appreciated when the application of image rejection filters in practical systems is discussed. In summary, it is safe to say that whenever the termination of a particular response is altered, be it the image ω_{-1}, a harmonic of the local oscillator, $m\omega_p$, or a higher idler, $\omega_m = \omega_0 + m\omega_p$, the principal conversion loss L_1 (from ω_1 to ω_0) and, consequently, the noise characteristics of the mixer will be affected. This is easily understood if we keep in mind that a mixer is not a passive device which merely converts an incoming frequency to some other frequency. It also generates a large number of other mixing products and harmonics, all of which interact by virtue of their terminating impedances.

A practical question therefore arises: Given the most common type of mixer —a broadband double-sideband mixer—what should its image termination be to minimize the conversion loss from signal ω_1 to the IF ω_0? This has been investigated by numerous authors [12–16]. Theoretically, the lowest L and F are obtained when the image ω_{-1} is open-circuited, although the improvement appears to be less than 1 dB [12–15], and the penalty is that the signal and the IF impedances may become high [16]. In practice it has been found [15, 16] that a short-circuited image may yield a lower noise factor at higher microwave frequencies (X-Band) because of difficulties of realizing an ideal open-circuited termination at ω_{-1}. It has also been reported [15] that the conversion loss of a mixer with a short-circuited image appears to be less sensitive to harmonic terminations. It should be emphasized that when we talk about a particular image termination, the latter is meant to be applied right at the nonlinear element. Trying to realize an open or short circuit at image frequency by means of a reactive filter placed some distance away from the diode represents a different situation due to the transmission line which has been introduced (Section 10.7).

A judicious termination of images in reactive loads, including open and short circuits, forms the basis of the so-called image-enhancement techniques which allow the power, normally lost in resistive image terminations, to be recovered. This effectively lowers the conversion loss. As attractive as this idea is, it had to await the advent of low-loss Schottky-barrier diodes before becoming practical. Clearly, if the diode itself is already lossy (e.g., high R_s) as in the old point-contact-type mixer diodes, eliminating the effect of the

external load at ω_{-1} by means of reactive filters would not reduce substantially the loading seen by the image to make the exercise worthwhile.

There are other practical considerations worth mentioning when working with or characterizing physical mixers:

(a)   It has been assumed that the local oscillator contributes no noise of its own. Since the LO is a large-signal waveform and basically drives the diode conductance between two extreme states, any small-amplitude noise on the LO signal will not affect the mixer. However, if the LO is derived from an internally leveled sweep generator which uses a backward-wave oscillator tube, there is a possibility of higher harmonics of the leveling signal causing sidebands to appear around the LO signal. These could be converted to the IF band if their spacing happens to be right, and they then appear as additional noise. If this is suspected to happen, turning the leveler off would quickly diagnose the problem.

(b)   Mixer specifications usually list the minimum conversion loss and the minimum noise factor realizable with the unit. Care is required in interpreting these data, for F_{min} and L_{min} may not occur at the same LO level [17]. This is because a high LO amplitude enhances the nonlinearity of the diode characteristic and, consequently, improves the conversion loss. At the same time the increased forward current during the forward swings of the LO voltage means an increase in shot noise, and therefore a degradation in the noise factor.

(c)   At times it may be necessary to include a matching or transforming network at the input port of the mixer. Although we keep referring to a separate "image port," physically, both ω_1 and ω_{-1} use the same input connector. The distinction is in electrical terms only. Thus, unless the image is well contained within the mixer structure by means of reactive terminations, the transforming network would affect the image as well. The consequence is yet another effect on the conversion loss L_1.

Noise Model of Mixer

The preceding section discussed the conversion loss L of a generalized, ideal mixer, and emphasized the dependence of L on terminations seen by the various mixing products. Since the noise properties of a mixer are intimately related to the conversion loss, it is not surprising that the noise factor of a mixer also depends on how these sidebands are terminated. Another subtlety not always appreciated is the characteristics of the interface between the mixer output and what follows, typically an IF amplifier. This can be understood by recalling the discussion on the overall noise factor of cascaded systems. In mixer/IF combinations the usual procedure is first to measure the noise factor of the amplifier by itself using some type of noise generator. The F_{IF} is thereby referred to that particular source impedance. Since F is a function of source impedance, the cascaded noise factor as calculated would not be correct unless

the amplifier sees the same source impedance looking into the output port of the mixer. Remember now that in a mixer/IF combination the first stage—the mixer—has loss instead of gain. Therefore, whatever happens to the second-stage noise factor will not be covered up by the gain of the first stage as would be if we had an amplifier as the first stage. Under these conditions the cascade noise factor is very much at the mercy of the mixer, not only because of the vagaries of the conversion loss, but also because any change in the output impedance of the mixer may affect the noise contribution of the F_{IF} substantially [16]. The anticipated improvement in the total noise performance could well turn into a degradation instead if our attempts to improve the mixer circuitry have unwittingly upset the mixer/IF amplifier interface.

Before discussing the noise properties of mixers, it is necessary to understand a fundamental relationship pertaining to passive linear multi-port networks. Consider a multi-port as shown in Fig. 10.11. Let the network be at a physical temperature T_x and let the physical temperatures of the terminations $Z_1 \cdots Z_R$ be $T_{g1} \cdots T_{gR}$, respectively. The port 0 is designated as the output, terminated in a conjugate load $Z_{\text{out}}^* = Z_L$. It can then be shown [2] that the total available noise power at the output port in a narrow bandwidth df is given by

$$N_o = N_{no} + (N_i)_o = k\,df\left[T_x\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) + \sum_{i=1}^{R}\frac{T_{gi}}{L_i}\right] \tag{10-68}$$

where L_i is the loss from port i to the output port 0. The L_i is defined as the ratio of the available power at port i to the power delivered to Z_L. Equation (10-68) yields the effective noise temperature looking into the output port:

$$T_n = T_x\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) + \sum_{i=1}^{R}\frac{T_{gi}}{L_i} \tag{10-69}$$
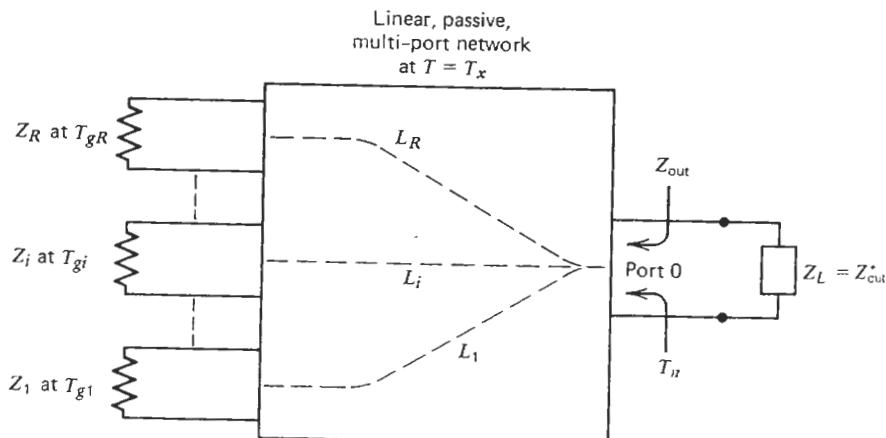
FIGURE 10.11.   Noise temperature of a linear, passive, multi-port network.

A similar expression has also been derived by Wait [18]. It should be noted that (10-69) could be obtained in a somewhat less rigorous manner directly from Pierce's rule (Section 5.2). If a unit of power is fed *into* the output port 0, then each termination Z_i must receive a portion $\alpha_i = 1/L_i$ of the input power. But since the total power lost in the multi-port network and its terminations must be unity, we obtain the sum

$$\sum_{i=1}^{R}\left(\frac{1}{L_i}\right) + \alpha_x = 1 \tag{10-70}$$

where α_x is the fraction of input power dissipated in the network. Hence

$$\alpha_x = 1 - \sum_{i=1}^{R}\frac{1}{L_i} \tag{10-71}$$

By Pierce's rule $T_{\text{eff}} = \alpha_1 T_{g1} + \alpha_2 T_{g2} + \cdots + \alpha_R T_{gR} + \alpha_x T_x$, and therefore

$$T_{\text{eff}} = T_n = T_x\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) + \sum_{i=1}^{R}\frac{T_{gi}}{L_i} \tag{10-72}$$

which is the same as (10-69). It will be remembered that (10-69) was already encountered in Section 5.2 where the output noise temperature of a matched attenuator was evaluated. There, because only one response was involved, $R = 1$ and (10-69) reduced to (5-29).

We are now ready to consider the attenuator noise model of a mixer. A noise model of a multi-response transducer was developed in Fig. 10.2, but this model is too general and does not relate to specific parameters of a mixer. For example, an input noise temperature T_e was assigned to each response in accordance with the IRE definition [4], but it is not clear how this T_e is related to the mixer diode and to the conversion losses of individual responses. A rigorous analysis of conversion loss and noise in mixers is not a simple matter. As Held and Kerr have shown [1], a good agreement between theory and experimental results requires that such aspects of practical mixers as the nonlinear junction capacitance $C(v)$, the effect of the series resistance R_s, the correlation in the downconverted shot noise, terminations of higher-order sidebands, and so on, be carefully considered. It is, however, possible to derive an attenuator model of noise which, to be sure, represents an idealized and somewhat oversimplified case, but which is nevertheless useful in analyzing the noise characteristics of mixers. Consider the following assumptions:

(a)   The diode characteristic is exponential (10-62) and ideal in the sense that its series resistance R_s and nonlinear capacitance $C(v)$ are negligible (static junction capacitance C_0 is allowed).

(b)   As a consequence of (a), the diode generates no thermal noise. It has full shot noise only, $\overline{i^2} = 2qI\,df$, and this is the only noise mechanism in the diode.

(c) The model has ports for all sidebands at which real power can flow between the diode and the loading external to the diode.

(d) Correlation between shot-noise components [1] is neglected.

(e) All LO harmonics and idlers above R are reactively terminated.

Kerr [2] has shown that if these assumptions hold then the pumped diode is equivalent to a lossy multi-port (or multi-response) network (Fig. 10.11) which is at a physical temperature T_x such that

$$N_{no} = k\,df\,T_x\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) = k\,df\left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) \qquad (10\text{-}73)$$

where N_{no} is the excess-noise power at the output; T_d the physical temperature of the diode; η the ideality factor of the diode [see (10-62)]; and other terms are as in (10-68). The *total* available output noise power, *including* the source noise, is, as in (10-68), given by

$$N_o = k\,df\left[\left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) + \sum_{i=1}^{R}\frac{T_{gi}}{L_i}\right] \qquad (10\text{-}74)$$

The last term represents the usual thermal noise from the source terminations, $k\,df\,T_{gi}$, referred to the output by dividing each component by its conversion loss L_i. We have allowed each source impedance Z_{gi} to have for the sake of generality a different temperature T_{gi}. In practice these are often equal. Thus

$$N_o = k\,df\left[\left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right) + T_g\sum_{i=1}^{R}\frac{1}{L_i}\right] \qquad (10\text{-}75)$$

In view of the assumptions listed earlier, this noise model applies where Schottky diodes with low R_s and $C(v)$ are available, primarily at lower microwave frequencies. It breaks down when $C(v)$ causes parametric effects to appear, and also does not hold at cryogenic temperatures where the assumption (d) is no longer valid as explained in [1].

Equation (10-75) allows us to derive a number of useful noise relationships for mixers, and to relate these to the more abstract terms T_e and G used in earlier sections. If (10-73) is treated in the manner of (10-69), we obtain the output noise temperature of the mixer including the effect of the source (Fig. 10.12):

$$T_n = \underbrace{\left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R}\frac{1}{L_i}\right)}_{\text{mixer}} + \underbrace{T_g\sum_{i=1}^{R}\frac{1}{L_i}}_{\text{source}} \qquad (10\text{-}76)$$
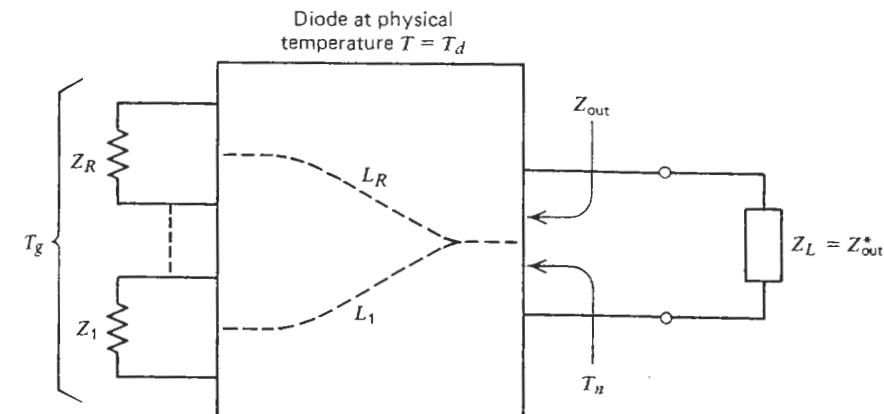
FIGURE 10.12. Output noise temperature of a physical mixer.

Equations (10-75) and (10-76) will now be applied to a few special cases. In particular, a dual-response mixer will be investigated, first with its image response ω_{-1} reactively terminated (sometimes also referred to as the narrowband mixer), and then with the image response matched to a resistive termination (broadband mixer).

Single-Sideband (Narrowband) Mixer

As a first example, consider a mixer in which the idlers and the image response ω_{-1} are all reactively terminated, that is, $L_i \to \infty$ for all i except $i = +1$ (Fig. 10.13). Note carefully that in this section the term "single-sideband mixer" means a frequency converter with its image response wholly contained within the mixer structure and reactively terminated, that is, there is no resistive loading at ω_{-1}. In accordance with (10-76), the output noise temperature becomes

$$T_n = \left(\frac{\eta T_d}{2}\right)\left(1 - \frac{1}{L_1}\right) + \frac{T_g}{L_1} \qquad (10\text{-}77)$$

If this is referred to the input, then

$$T_{ni} = \left(\frac{\eta T_d}{2}\right)(L_1 - 1) + T_g \qquad (10\text{-}78)$$

from which we can identify

$$T_e = \left(\frac{\eta T_d}{2}\right)(L_1 - 1) \qquad (10\text{-}79)$$
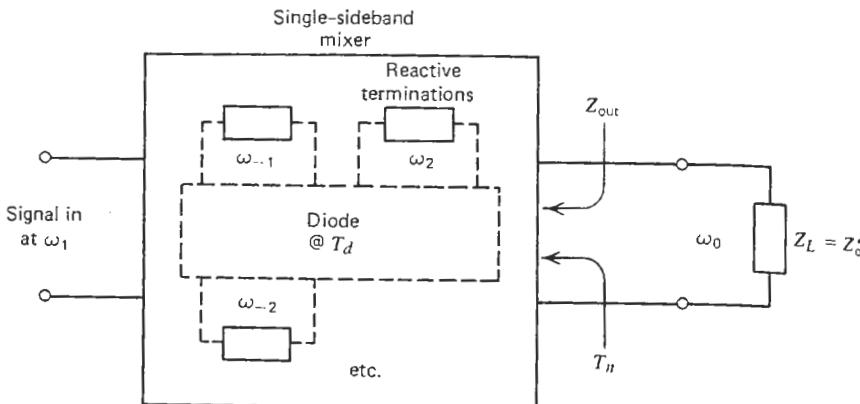
FIGURE 10.13.   Equivalent circuit of a single-sideband mixer.

which is similar to (8-30), the equivalent input noise temperature of an attenuator. This is not surprising, because all other responses were eliminated by reactive terminations and the mixer is effectively just a single-response, two-port, black box as far as its input (ω_1) and output (ω_0) are concerned.

The noise factor for this mixer follows in a straightforward manner:

$$ F = \frac{N_o}{kT_0\,df(1/L_1)} = \frac{k\,df\left[(\eta T_d/2)(1 - 1/L_1) + (T_0/L_1)\right]}{kT_0\,df(1/L_1)} $$

$$ = 1 + \left(\frac{\eta T_d}{2T_0}\right)(L_1 - 1) = 1 + \frac{T_e}{T_0} \qquad (10\text{-}80) $$

as expected. Remember again that the definition of noise factor requires $T_g = T_0$.

Double-Sideband (Broadband) Mixer

Consider next a more typical broadband mixer with three electrical ports, signal ω_1, image ω_{-1}, and IF or output ω_0 resistively terminated (Fig. 10.14). From (10-75) the total noise at output now becomes*

$$ N_o = \underbrace{k\,df\left[\left(\frac{\eta T_d}{2}\right)\left(1 - \frac{1}{L_1} - \frac{1}{L_{-1}}\right)\right.}_{\text{mixer excess noise } N_{no}} + \underbrace{\left. T_g\left(\frac{1}{L_1} + \frac{1}{L_{-1}}\right)\right]}_{\text{source noise}} \qquad (10\text{-}81) $$

*The conversion loss L_1 in (10-81) is not necessarily the same as that in (10-77)–(10-80). The change in image termination, from reactive to resistive, does affect L_1.
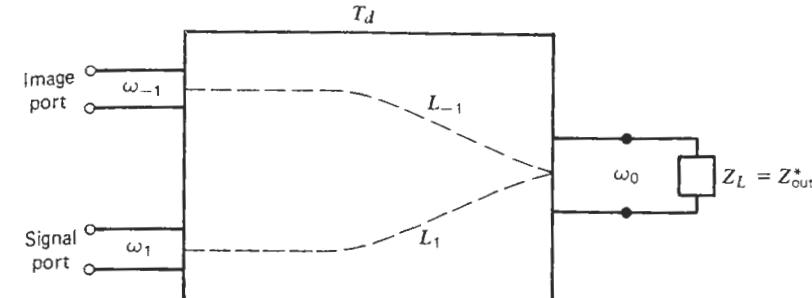
FIGURE 10.14.   Equivalent circuit of a broadband mixer.

Note that we have not specified whether there is signal in the signal response ω_1 *only* (single-sideband operation) or whether the signal occupies both responses—signal *and* image (double-sideband operation). The term "double-sideband mixer" is here again used *only* to designate the number of "open" sidebands, not the type of operation.

As in the preceding section, the next step is to find the input noise temperature, that is, that temperature to which the input termination must be raised (with the mixer held at $0°K$) to obtain the same output noise N_o as in the actual case. At this point there are two choices. In both cases the signal occupies only the "signal" channel ω_1, but there is a difference in how the input noise temperature is assigned.

Option A.   One could argue that the excess noise originating *in the mixer* at the image frequency ω_{-1}, as well as that at the signal frequency ω_1, should *both* be charged against the input (signal) port (see Fig. 10.15). This is particularly logical for a single-sideband mode of operation as in a radar or communication receiver where the signal is contained in the ω_1 response only, and hence only that response is of principal interest. In other words, we consider the total excess noise N_{no} observed at the output as having been contributed by a single fictitious noise generator of temperature T_e' connected
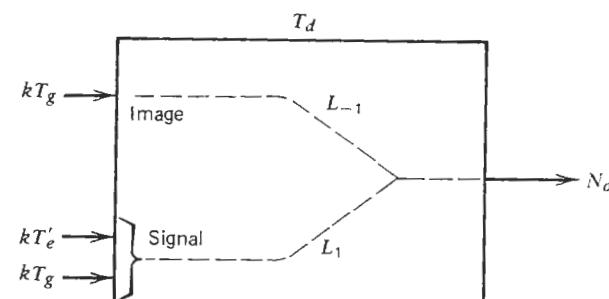
FIGURE 10.15.   Illustration of the single-sideband noise temperature.

to the signal input port. Thus from (10-81)

$$N_{no} = k\,df\left(\frac{\eta T_d}{2}\right)\left(1 - \frac{1}{L_1} - \frac{1}{L_{-1}}\right) \equiv \frac{T_e'k\,df}{L_1} \qquad (10\text{-}82)$$

or

$$T_e' = \frac{\eta T_d}{2}\left(L_1 - 1 - \frac{L_1}{L_{-1}}\right) \qquad (10\text{-}83)$$

This is sometimes called the single-sideband noise temperature. If $L_1 = L_{-1} = L$, then

$$T_e' = \tfrac{1}{2}\eta T_d(L - 2) \qquad (10\text{-}84)$$

The noise factor corresponding to this situation can be obtained either from (10-81) in the conventional manner, or from the T_e'. We shall illustrate both methods. Starting with (10-81), and noting that what we are deriving here is the single-sideband noise factor F_{SSB} (because the signal appears only at ω_1), we get

$$
\begin{aligned}
F_{SSB} &= \frac{N_o}{k\,df\,T_0(1/L_1)} \\
&= \frac{k\,df\left[\tfrac{1}{2}\eta T_d(1 - 1/L_1 - 1/L_{-1}) + (T_0/L_1)(1 + L_1/L_{-1})\right]}{k\,df\,T_0(1/L_1)} \\
&= \frac{\eta T_d}{2T_0}\left(L_1 - 1 - \frac{L_1}{L_{-1}}\right) + \left(1 + \frac{L_1}{L_{-1}}\right) \qquad (10\text{-}85)
\end{aligned}
$$

Again, if $L_1 = L_{-1} = L$, then

$$F_{SSB} = 2 + \left(\frac{\eta T_d}{2T_0}\right)(L - 2) \qquad (10\text{-}86)$$

Doing it the other way, notice again from Fig. 10.15 how the equivalent excess noise is lumped at the input of the *signal* port. Therefore evaluating the N_o yields

$$N_o = k\,df\left(T_0 + T_e'\right)\left(\frac{1}{L_1}\right) + k\,df\,T_0\left(\frac{1}{L_{-1}}\right) \qquad (10\text{-}87)$$

and hence

$$F_{SSB} = \frac{N_o}{k\,df\,T_0(1/L_1)} = \left(1 + \frac{T_e'}{T_0}\right) + \frac{L_1}{L_{-1}} \qquad (10\text{-}88)$$

If this expression seems confusing to you, have patience because the next section will show not only how T_e' is related to the T_e of earlier sections and Figs. 10.2, 10.4, and so on, but that (10-86) and (10-88) are indeed in agreement with the more general (10-29).

Option B. In this option we again start with the total excess noise at the output, but this time it is assumed that N_{no} is contributed by *two* fictitious noise generators applied as shown in Fig. 10.16 and more generally in Fig. 10.2. In accordance with the IRE definition (see Section 10.1), both generators have the same noise temperature T_e, although the conversion losses L_1 and L_{-1} could be different. In analogy to (10-82):

$$N_{no} = k\,df\left(\frac{\eta T_d}{2}\right)\left(1 - \frac{1}{L_1} - \frac{1}{L_{-1}}\right) \equiv \left(\frac{T_e}{L_1} + \frac{T_e}{L_{-1}}\right)k\,df \qquad (10\text{-}89)$$

or

$$T_e = \frac{\eta T_d}{2}\left(\frac{L_1 - 1 - L_1/L_{-1}}{1 + L_1/L_{-1}}\right) \qquad (10\text{-}90)$$

which could be called the broadband noise temperature.
For $L_1 = L_{-1} = L$, Eq. (10-90) becomes

$$T_e = \left(\frac{\eta T_d}{2}\right)\left(\frac{L - 2}{2}\right) \qquad (10\text{-}91)$$

Equations (10-90) and (10-91) at last give the desired relationships between the general input noise temperature T_e introduced earlier and the noise parameters η and T_d of a physical mixer. Furthermore, by comparing (10-91) to (10-84), we have for the equal-conversion-loss case,

$$T_e = \tfrac{1}{2}T_e' \qquad (10\text{-}92)$$
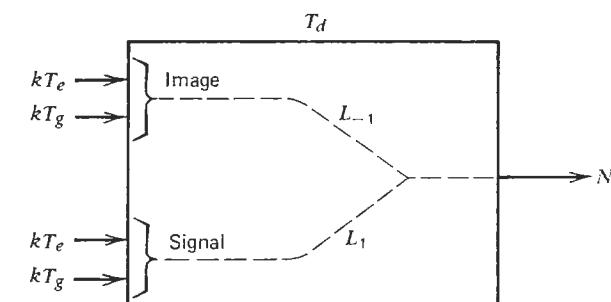
FIGURE 10.16. Illustration of the broadband noise temperature.

or more generally from (10-90) and (10-83),

$$T_e = \frac{T_e'}{1 + L_1/L_{-1}} \tag{10-93}$$

Notice, however, that (10-93) still represents a special (although a very common) case, because we had earlier in connection with (10-77) restricted the number of responses to just 2, that is, L_1 and L_{-1} only.

Finally, by substituting (10-93) into (10-88), or (10-90) into (10-85), we get

$$F_{SSB} = \frac{T_e}{T_0}\left(1 + \frac{L_1}{L_{-1}}\right) + \left(1 + \frac{L_1}{L_{-1}}\right)$$

$$= \left(1 + \frac{T_e}{T_0}\right)\left(1 + \frac{L_1}{L_{-1}}\right) \tag{10-94}$$

which is precisely the same as (10-29), because we have assumed the same bandwidth for both responses [$B_1 = B_2$ in (10-29), or using the present notation, $B_1 = B_{-1}$].

As the last illustration, consider the double-sideband operation, that is, one in which the signal resides in both responses. In accordance with the definition of noise factor, only the denominator of (10-85) must be modified to include the effect of the image band:

$$F_{DSB} = \frac{k\,df\left[(\frac{1}{2}\eta T_d)(1 - 1/L_1 - 1/L_{-1}) + T_0(1/L_1 + 1/L_{-1})\right]}{k\,df\,T_0(1/L_1 + 1/L_{-1})}$$

$$= \left(\frac{\eta T_d}{2T_0}\right)\left(\frac{L_1 - 1 - L_1/L_{-1}}{1 + L_1/L_{-1}}\right) + 1 \tag{10-95}$$

or using (10-90),

$$F_{DSB} = 1 + \frac{T_e}{T_0} \tag{10-96}$$

Again, a reference to Section 10.3 shows that (10-96) is the same as (10-30). If (10-95) and (10-85) are compared, it is seen that

$$F_{SSB} = F_{DSB}\left(1 + \frac{L_1}{L_{-1}}\right) \tag{10-97}$$

which is in agreement with (10-29)–(10-31).

Mixer/IF Amplifier Combination

The noise factor of a mixer alone is not a very significant number because in practice the mixer is nearly always followed by an IF amplifier and the two are characterized and delivered as a package. We saw in the beginning of this section that the mixer-amplifier interface can have a sizable effect on the overall noise factor. It is therefore of interest to review the cascaded noise factor of the mixer/amplifier combination and to see to what degree each component contributes to the total.

The cascaded noise factor of a mixer/amplifier package (Fig. 10.17) is easily obtained from Section 10.5. Starting with the cascade formula (10-58) and using (10-85), it follows that

$$(F_T)_{SSB} = \left(\frac{\eta T_d}{2T_0}\right)\left(L_1 - 1 - \frac{L_1}{L_{-1}}\right) + \left(1 + \frac{L_1}{L_{-1}}\right) + (F_{IF} - 1)L_1 \tag{10-98}$$

which for $L_1 = L_{-1} = L$ becomes

$$(F_T)_{SSB} = \left(\frac{\eta T_d}{2T_0}\right)(L - 2) + 2 + L_1(F_{IF} - 1) \tag{10-99}$$

This formula is seen to agree with (10-58) if (10-91) is used. However, in practice a different form of (10-99) is used. The latter form uses the noise-temperature ratio t_m, defined as

$$t_m = \frac{\text{available output (IF) noise power in } df}{kT\,df} \tag{10-100}$$

which is analogous to (8-9), but defined specifically for mixer use. Here T is the ambient temperature which is usually taken as $T_0 = 290\,°K$. In other words, t_m compares the actual noise measured at the mixer output, when the mixer and its terminations are held at a physical temperature T, to the noise available from a pure resistor at the same temperature T.
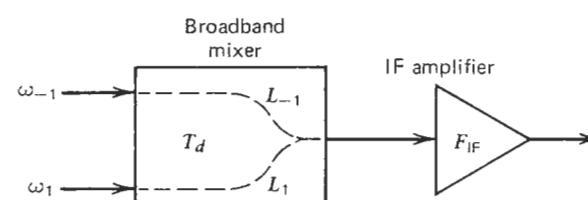
FIGURE 10.17.   Equivalent circuit of a broadband mixer and IF amplifier combination.

The noise-temperature ratio for a dual-response system with a matched image response follows directly from (10-81):

$$t_m = \frac{k\,df\left[\left(\tfrac{1}{2}\eta T_d\right)(1 - 1/L_1 - 1/L_{-1}) + T(1/L_1 + 1/L_{-1})\right]}{kT\,df}$$

$$= \left[\left(\frac{\eta T_d}{2T}\right)\left(L_1 - 1 - \frac{L_1}{L_{-1}}\right) + \left(1 + \frac{L_1}{L_{-1}}\right)\right]\left(\frac{1}{L_1}\right) \qquad (10\text{-}101)$$

Comparing (10-101) to (10-98) yields the more familiar expression for a mixer/IF amplifier combination:

$$(F_T)_{\text{SSB}} = L_1 t_m + L_1(F_{\text{IF}} - 1) = L_1(t_m + F_{\text{IF}} - 1) \qquad (10\text{-}102)$$

Note that (10-101) and (10-102) represent a single-sideband noise factor, that is, a broadband mixer, but with signal only in one response, ω_1.

If the mixer/amplifier package is used in a double-sideband (radiometry) application, then for the mixer it is (10-95), not (10-85), that applies. Assume as before that the conversion losses L_1 and L_{-1} are equal; then (10-59) can be used. Applying (10-95) and (10-101) yields

$$(F_T)_{\text{DSB}} = \left(\frac{\eta T_d}{2T_0}\right)\left(\frac{L-2}{2}\right) + 1 + \frac{L(F_{\text{IF}} - 1)}{2}$$

$$= \frac{L}{2}\left(t_m - \frac{2}{L}\right) + 1 + \frac{L(F_{\text{IF}} - 1)}{2}$$

$$= \frac{L}{2}(t_m + F_{\text{IF}} - 1) \qquad (10\text{-}103)$$

If the losses are not equal, then (10-59) must be rederived starting with (10-27) and using $G_{b1} \neq G_{b2}$.

The important conclusion to be drawn from (10-102) and (10-103) is that the overall noise factor is *not* equal to the product of the conversion loss and the IF noise factor as is frequently assumed. Partly for this reason and partly because of the rather stringent mixer/IF amplifier interface requirements, care is required in calculating the overall noise factor, given the individual quantities L and F_{IF}. When cascading a mixer with an IF amplifier, it is advisable to ensure that the amplifier would see its optimum source impedance looking into the mixer output. This will result in the lowest noise factor even when there is a mismatch [17]. It is also worth noting that in mixers using 180° hybrids, where any mismatch at ω_1 is reflected back to the source, the IF amplifier impedance will affect the RF input VSWR (voltage standing-wave ratio). Consequently, the minimum overall noise factor may not coincide with the minimum input VSWR at the mixer input port.

Concluding Remarks

The purpose of this section was to present an overview of the operating parameters and characteristics of physical mixers and frequency converters.

The emphasis was on engineering aspects rather than on rigorous theory. It should therefore be remembered that the noise models and their mathematical formulations as outlined are based on assumptions which may not always hold. Still, the value of this treatment is that the otherwise very difficult problem can be rendered tractable. In addition, it becomes possible to determine what the ultimate performance limits of a given mixer are. A standard is thereby provided against which the actual performance could be measured.

Finally, it is hoped that the preceding discussion has clarified at least some of the aspects of the mixer noise theory—an area which seems to be plagued by recurring confusion.

10.7. POSTSELECTORS AND PRESELECTORS

The use of various types of filters is commonplace in electronic and microwave systems. The filters serve a wide variety of functions, such as protecting the system against strong external signals, or limiting spurious emission from one's own system. The list of possibilities is endless.

The application of filters in single-response systems, cascaded amplifiers, for example, is from the standpoint of thermal noise straightforward. The filter usually has some resistive loss within its passband, but this loss can be modeled as an attenuator and treated in accordance with Figs. 8.8 and 9.2 and Eq. (9-22) as far as noise analysis is concerned. There could, of course, be problems of a different nature such as instability when broadband, high-gain, active transducers are preceded or followed by reactive (and hence highly mismatched) filters. The resultant unstable operation could affect the characteristics of active devices and could lead to an increase in overall system noise. However, these problems are related to the design of circuits and remain therefore outside the scope of this book.

In this section we shall consider the fundamentals of using filters in multi-response systems containing frequency converters or mixers. Typical applications of filters in such cases are the blocking of spurious signals at image frequencies from entering the receiver, or the use of image filters to counteract the effect of image noise on the final SNR.

It was shown in Section 10.6 that the analysis of mixers with arbitrary image and higher idler terminations is not a trivial task. The input terminals of individual responses are not isolated entities, but are coupled together through the action of the nonlinear conductance, that is, the diode. Hence the magnitude and phase of the impedance terminating a particular response affects the transmission between any other two responses. To express this in simpler terms, the conversion loss of a downconverting mixer and hence its noise properties do depend on what is connected to the image terminals. Although a rigorous analysis of any specific configuration may be quite involved, useful conclusions regarding general characteristics are still possible. We shall consider a few examples.

A topic of considerable practical interest is the noise contribution from the principal image response ω_{-1}. As shown in earlier sections and particularly in Eqs. (10-7), (10-12), and (10-49), this contribution can be appreciable if the given mixer is a broadband unit with equal gains at both signal and image frequencies, and with a wide-open input which receives source noise through both responses—signal and image. In the single-sideband operating mode, the extra image noise is quite undesirable in view of its first-order effect on the SNR_o. While broadband radiometry systems overcome this problem by admitting noise and signal through both responses, communication and radar systems operate with signals in one response only and any image noise is simply unwanted ballast.

The straightforward way of solving the problem would be to use a single-sideband mixer, that is, one in which the image frequency has been internally terminated in an optimal manner. Such a termination is equivalent to an image filter. We already mentioned that an appropriate reactive termination of the image does improve conversion loss and hence the noise factor. However, the drawbacks of this method are a restricted bandwidth of the mixer and, once the unit is designed and fabricated, that its operating frequencies are fixed. The flexibility of a broadband mixer has been lost.

In order to remedy this situation, the first obvious idea would be to place a filter, either a bandpass filter tuned to ω_1, or a band reject filter tuned to ω_{-1}, in front of the system. This is shown in Fig. 10.18. The image noise originating in the source resistance has now been blocked and the problem appears to be solved. Or is it?

Remembering the discussion in the previous section, one may already suspect that the situation is not that simple. We have indeed eliminated the noise at the image frequency coming from the *external* source, but this is not all. Since the amplifier input is usually matched, there must be noise generated at ω_{-1} in the input circuit. This noise is normally radiated out toward the source in accordance with the requirements of thermal equilibrium. However, with the image filter in place, the image noise is reflected back into the amplifier. We have therefore gained little, if anything.

Now, there could be an observable change in N_o at the system output, but this change is a function of several variables. If the system is connected to an antenna having a low noise temperature T_a, then the addition of an image filter could worsen the output noise, because the initial low-temperature image noise
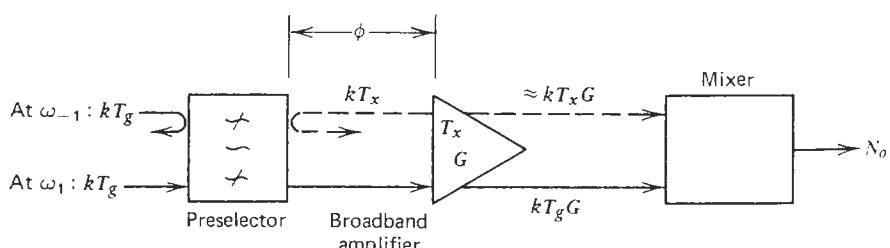
originating in the cold sky has been replaced by reflected image noise from the much warmer input circuit of the amplifier. Also, the position ϕ, that is, the phase of the reflected noise component, and the bandwidth of the reflecting filter will have an effect. This stems from possible correlation effects between the noise and its reflected replica, shifted in time phase. It will be remembered from Section 3.1 that although *white* noise (unlimited bandwidth) has an autocorrelation function equal to an impulse function, the same noise reflected from a band-limited filter is no longer white. The autocorrelation function $R(\tau)$ now extends over a finite time interval (3-20), and there could be correlation between the noise and its reflection depending on the round-trip delay. A detailed analysis remains outside the scope of this book. One should note, however, that the use of a preselector does not guarantee the elimination of image noise. The outcome could well be better, the same, or worse than what we had initially.

The futility of inserting an isolator (Fig. 10.19) between the preselector filter and the amplifier (or in its absence, the mixer) input should be equally clear. True, the image noise from the amplifier or mixer would now be dissipated in the isolator termination, but in its place the noise originating in the termination would appear.

Is there any way at all to eliminate the effect of the image noise? Yes, there is, and we already saw this in Section 10.5 in connection with Figs. 10.5 and 10.6. The difference between Fig. 10.6 and 10.18 is that in the former case the filter was used as a *postselector*. Equation (10-54) demonstrated that the image noise could then indeed be suppressed. We can also see this qualitatively. The source noise in Fig. 10.6 experiences gain at *both* frequencies, signal and image, when passing through the broadband amplifier. When it reaches the postselector, the noise at ω_{-1} is blocked. To be sure, it is replaced by new image noise at point (a), this time contributed by the isolator or in its absence by the reflected mixer noise. However, this noise has not undergone amplification and is much weaker than the source noise in the amplified signal response. As a result, it can be neglected, and we are left essentially with noise corresponding to one response only. Thus, as shown in (10-54), by providing enough gain in the preceding amplifier, the effect of the second stage can be eliminated.

Another way of eliminating the image noise is to use an image rejection mixer after the first-stage amplifier, that is, in place of the postselector filter. This method will be described in the next section on imageless mixers.

FIGURE 10.18.   The use of a preselector with a broadband amplifier/mixer combination.
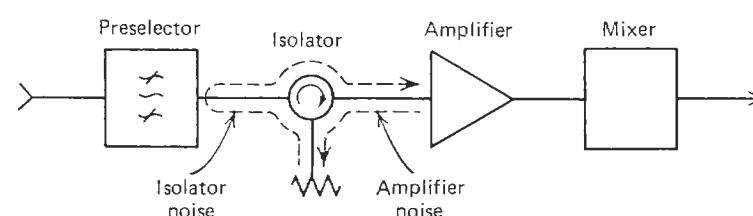
FIGURE 10.19.   The noise effect cannot be eliminated by using an input isola
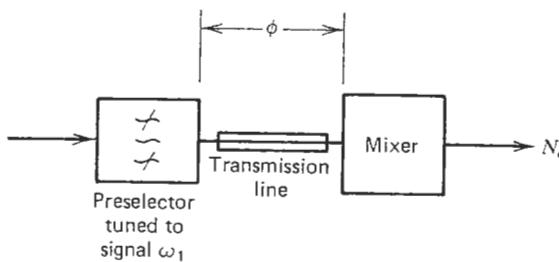
FIGURE 10.20. Noise properties and the conversion loss of the mixer depend on the location of the preselector.

In the absence of a preamplifier, the filter would precede the mixer as the first component in the chain (Fig. 10.20). The outcome of this configuration is not easily predictable as we saw earlier, because the reactive termination at the image frequency, *as seen at the diode*, could be anything from an open to a short, depending on the electrical line length ϕ and the specific design of the given mixer. The variation in the noise factor as a function of ϕ could be several dB [12, 13, 15, 16], because such a termination affects both the conversion loss L_1 at ω_1, and the noise properties of the mixer. In addition, any physical line between the filter and the mixer diode implies added resistive losses which tend to blur the distinction between the broadband and the idealized reactive termination cases. The advice therefore is to adjust the position of the filter until the best noise performance is obtained.

Finally, it should be always kept in mind that the isolation between the three physical ports of the mixer, signal, local oscillator, and IF is not infinite. A change at one port could affect the conditions at others simply through direct leakage. For example, in Fig. 10.20 LO leakage into the input line would also be reflected by the filter. Depending on the magnitude and phase of this reflection, the pumping of the diode could change, with obvious consequences on the mixer performance.

In summary, the primary purpose of a *preselector* is to protect the receiver against spurious *signals* outside the desired signal band, including those at the various image frequencies. *Postselectors*, on the other hand, are used to eliminate *noise* at the image frequency. Preselectors may accomplish this also, but the situation is more involved and success is not automatically guaranteed.

10.8. IMAGELESS MIXER

In order to discriminate against unwanted signals at the image frequency, an imageless or image rejection mixer is often used. Figure 10.21 shows the circuit, and Appendix C shows that the IF output due to signals at the image frequency are absorbed in the termination $R2$, while those due to the desired
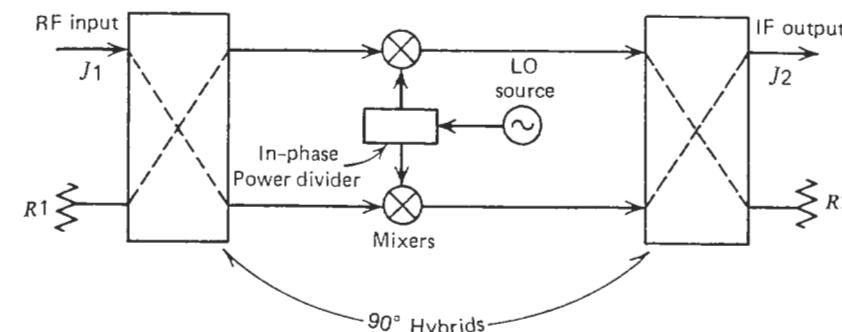
FIGURE 10.21. The image rejection or imageless mixer.

signal at ω_1 are sent to the output port $J2$. An image rejection of 20–30 dB is easily obtained in this manner.

The image noise entering the mixer through $J1$ will follow the image signal, and its contribution will therefore also end up in $R2$. Does this mean that the image noise has been eliminated? Unfortunately, no. Image noise generated *internally* within $R1$ will simply replace the external image noise and the noise factor remains unaffected (unless, of course, we cool $R1$ to cryogenic temperatures).

Although the imageless mixer, if used by itself as the front end in a superheterodyne receiver, gives no relief from image noise, it does a fine job if preceded by a high-gain broadband amplifier. The reason is the same as given earlier in connection with the postselector. Again, we cannot eliminate the image noise per se, but we can trade the amplified image noise against the much weaker image noise originating in the termination of the first hybrid. The image noise contribution is thus rendered negligible. This method is a popular one since the postselector and the mixer functions are combined in one device.

10.9. SUMMARY

(a)   Total output noise for multi-response transducer:

$$N_o = kG_1B_1\left(T_g + T_e\right)\mathscr{R} = kG_1B_1T_0\left(\frac{T_g}{T_0} + \left(F_{\mathrm{MSB}} - 1\right)\right)\mathscr{R}$$

(b)   Response factor:

$$\mathscr{R} = 1 + \frac{G_2B_2}{G_1B_1} + \cdots + \frac{G_RB_R}{G_1B_1}$$

where R denotes the number of responses.

(c) Equivalent input noise temperature of a response (same for all responses):

$$T_e = \frac{N_{no}}{kG_1 B_1 \mathscr{R}}.$$

(d) Operating noise temperature (assuming all sources at the same temperature):

$$T_{op} = (T_g + T_e)\mathscr{R}.$$

(e) Single-sideband noise factor:

$$F_{SSB} = \left(1 + \frac{T_e}{T_0}\right)\mathscr{R}.$$

(f) Multi-sideband noise factor:

$$F_{MSB} = \left(1 + \frac{T_e}{T_0}\right)\left(\frac{\mathscr{R}}{1 + G_2 B_2/G_1 B_1 + \cdots + G_i B_i/G_1 B_1}\right)$$

$$= \left(1 + \frac{T_e}{T_0}\right)$$

which is valid for an arbitrary number of responses if a signal is present in all.

(g) Special case of two responses having equal gains and bandwidths:

$$F_{SSB} = 2\left(1 + \frac{T_e}{T_0}\right) = 2F_{DSB}.$$

(h) Conventional cascade formula does not necessarily hold for multi-response systems:

(1) Broadband amplifier plus dual-response transducer:

$$(F_T)_{SSB} = 2\left(F_a + \frac{(F_b)_{DSB} - 1}{G_a}\right)$$

$$= 2F_a + \frac{(F_b)_{SSB} - 2}{G_a} = 2(F_T)_{DSB}.$$

(2) Narrowband amplifier plus dual-response transducer:

$$(F_T)_{SSB} = F_a + \frac{(F_b)_{SSB} - 1}{G_{a1}}.$$

(i) Output noise temperature of a passive, linear, multi-port network:

$$T_n = T_x\left(1 - \sum_{i=1}^{R} \frac{1}{L_i}\right) + \sum_{i=1}^{R} \frac{T_{gi}}{L_i}.$$

(j) Output noise temperature of ideal mixer:

$$T_n = \left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R} \frac{1}{L_i}\right) + \sum_{i=1}^{R} \frac{T_{gi}}{L_i}$$

where T_d is the physical temperature of the mixer diode.

(k) Single-sideband noise temperature of mixer (at the input of signal port):

$$T_e' = \left(\frac{\eta T_d}{2}\right)\left(1 - \sum_{i=1}^{R} \frac{1}{L_i}\right).$$

(l) Single-sideband noise factor of mixer/IF amplifier combination:

$$F_{SSB} = L_1(t_m + F_{IF} - 1)$$

where t_m is the noise-temperature ratio of mixer.

(m) Overall mixer/IF amplifier noise factor is not equal to the product of mixer conversion loss and IF noise factor.

(n) Lowest noise factor of mixer/IF amplifier combination does not occur when mixer/amplifier interface is matched.

REFERENCES

1. D. N. Held and A. R. Kerr, "Conversion Loss and Noise in Microwave and Millimeter-Wave Mixers, Parts 1 and 2," *IEEE Trans. MTT*, vol. MTT-26, pp. 49–61, February 1978.
2. A. R. Kerr, "Shot-Noise in Resistive-Diode Mixers and the Attenuator Noise Model," *IEEE Trans. MTT*, vol. MTT-27, pp. 135–140, February 1979.
3. A. R. Kerr, "Noise and Loss in Balanced and Subharmonically Pumped Mixers, Parts I and II." *IEEE Trans. MTT*, vol. MTT-27, pp. 938–950. December 1979.
4. "IRE Standards on Electron Tubes: Definition of Terms, 1962," *Proc. IEEE*, vol. 51, pp. 434–435, March 1963.
5. E. G. Nielsen, "Amplifier Noise," in R. F. Shea (Ed.), *Amplifier Handbook*, McGraw-Hill, New York, 1966, Chapter 7.
6. A. A. M. Saleh, *Theory of Resistive Mixers*, M.I.T. Press, Cambridge, 1971.
7. W. F. Stacey, Raytheon SMDO, private communication.
8. J. W. Gewartowski, "Noise Figure for a Mixer Diode," *IEEE Trans. MTT*, vol. MTT-19, p. 481, May 1971.
9. A. M. Cowley and H. O. Sorensen, "Quantitative Comparison of Solid-State Microwave Detectors," *IEEE Trans. MTT*, vol. MTT-14, pp. 588–602, December 1966.

10. C. Dragone, "Performance and Stability of Schottky Barrier Mixers." *Bell Syst. Tech. J.*, vol. 51, pp. 2169–2196, December 1972.

11. A. J. Kelly, "Fundamental Limits on Conversion Loss of Double Sideband Resistive Mixers." *IEEE Trans. MTT.* vol. MTT-25, pp. 867–869, November 1977.

12. P. D. Strum, "Some Aspects of Mixer Crystal Performance," *Proc. IRE*, vol. 41, pp. 875–889, July 1953.

13. R. J. Mohr and S. Okwit, "A Note on the Optimum Source Conductance of Crystal Mixers." *IRE Trans. MTT*, vol. MTT-8, pp. 622–627, November 1960.

14. M. R. Barber, "Noise Figure and Conversion Loss of the Schottky Barrier Mixer Diode." *IEEE Trans. MTT*, vol. MTT-15, pp. 629–635, November 1967.

15. K. M. Johnson, "X-Band Integrated Circuit Mixer with Reactively Terminated Image." *IEEE Trans. MTT*, vol. MTT-16, pp. 388–397, July 1968.

16. M. Katoh and Y. Akaiwa, "4-GHz Integrated-Circuit Mixer," *IEEE Trans. MTT*, vol. MTT-19, pp. 634–637, July 1971.

17. D. Neuf and D. Brown, "What To Look For In Mixer Specs," *Microwaves*, pp. 48–60, November 1974.

18. D. F. Wait, "Thermal Noise from a Passive Linear Multiport," *IEEE Trans. MTT*, vol. MTT-16, pp. 687–691, September 1968.

11 MEASUREMENT OF NOISE PARAMETERS

This chapter will review the techniques and instrumentation used to measure noise parameters of linear transducers, such as the input noise temperature T_e and the noise factor F. The basic method of measuring T_e was already outlined in Section 8.4. The idea there and in most other methods is to apply two different, but precisely known, power levels of broadband white noise from a "hot" and a "cold" noise source to the input of the transducer under test, and to measure the resulting change in the output noise level. The change is expressed as the Y-factor, which is simply the ratio of the two output noise levels corresponding to the two different noise inputs. From this information, the input noise temperature T_e can be calculated directly as in (8-25), although the noise factor F may require additional information as we saw in Chapter 10. It may be necessary to know something about the other responses if the transducer is a multi-response device.

Admittedly, the foregoing is a bit oversimplified, because the actual measurement procedure and the required test equipment cover a wide territory, the extent of which depends on the frequency of operation, the precision required, the magnitudes involved, and the specific goal of the measurement. The sections to follow will therefore discuss these aspects in greater detail, but the reader is encouraged to consult the indicated references and the general bibliography at the end of the book for an in-depth treatment of specific topics, which the limited scope of this book must necessarily exclude. Thus our emphasis will be largely on high-frequency and microwave measurements

where one deals typically with noise power and noise density rather than with noise voltages and currents, as is the case at low frequencies. For measurements of noise in low-frequency lumped-element circuits, amplifiers, and so on, refs. [1–5] should prove helpful.

This chapter is organized in five major sections. Section 11.1 will discuss general aspects of noise measurement. Sections 11.2 and 11.3 will describe the noise standards and sources, and the various measurement procedures and their pitfalls, respectively. Section 11.4 describes the techniques of measuring the noise parameters of two-port transducers, which are based on the equivalent noise resistance and conductance model. Finally, Section 11.5 will consider some auxiliary test equipment needed to complete practical laboratory test setups.

11.1.   BASIC CONSIDERATIONS

It was stated above that the first prerequisite in measuring noise parameters is the availability of two accurately calibrated reference noise power levels. Either white noise in both cases, or white noise and a single-frequency CW signal could be used. Accordingly, we refer to the dispersed-signal technique and to the signal-generator technique, respectively. To these we should add a third, the so-called direct noise measurement technique which, as will be shown, is useful only in special cases.

The methods of noise measurement can also be classified according to whether a primary or a secondary noise standard is used. A primary standard is one whose output power is precisely calculable (in principle at least) from basic physics. The foremost in this group is the heated resistor at some temperature T. Its available output noise power density is given by $kT\,df$, or more precisely by (3-30):

$$P_n = \frac{hf\,df}{e^{hf/kT} - 1} \tag{11-1}$$

The noise from a resistor is therefore the best primary standard we have. In contrast, the widely used gas-discharge or solid-state noise sources are secondary standards since their effective noise temperature T_h, and hence the available noise power density kT_h, cannot be calculated with the required precision, owing to the many variables involved. However, when such secondary, but stable, noise sources are calibrated against a primary standard, they become accurate and versatile measuring tools, and are considered the workhorses in routine noise measurements.

The one important parameter in noise measurements is the Y-factor, which is the ratio N_{oh}/N_{oc} where the two output noise powers correspond to the two inputs $kT_h\,df$ and $kT_c\,df$ as in (8-22) and (8-23). From these data, the T_c is

calculated in accordance with

$$T_e = \frac{T_h - YT_c}{Y - 1} \tag{11-2}$$

We have so far said nothing about the numerical values of T_h and T_c, nor about their ratio T_h/T_c, but a consideration of (11-2) shows that given the proper combination of T_h, T_c, and T_e, the Y-factor could come uncomfortably close to unity. Under this condition, $Y - 1$ is very sensitive to errors in Y and the accuracy of the entire test becomes questionable. Let us therefore rearrange (11-2) to

$$Y = \frac{T_e + T_h}{T_e + T_c} \tag{11-3}$$

It is now evident that when T_e is low, as in the case of low-noise amplifiers ($T_e < 200\,°$K or so), a $T_h = 400\,°$K and $T_c = 80\,°$K (or $T_h/T_c = 5$) would be quite adequate. On the other hand, a $T_e = 2610\,°$K, corresponding to a not uncommon noise figure of 10 dB, would require a $T_h = 3150\,°$K or a $T_h/T_c = 39.4$, assuming the same T_c, to achieve the same accuracy in $Y - 1$. The user must therefore select input noise reference levels accordingly.

The foregoing is not meant to imply that it is possible in practice to select whatever T_h or T_c one may desire by a simple adjustment. As we shall see shortly, nature sets definite limits in this area. There are, however, a number of alternatives available, and hence it is important to know the rationale of how to choose the best one.

One may surmise from this remark that there is also an upper limit for T_h. Indeed, the upper limit is set by linearity considerations which, in turn, stem from the concept of noise factor itself. We mentioned earlier in Chapter 9 that the noise factor F is defined only for linear systems. That is, the transducer under test *and* the test instrumentation must both be linear over the applicable dynamic range. But what is the "applicable range"? The first thought might be that the maximum-required dynamic range need not be much larger than the Y-factor itself. After all, this is the range of the two output powers. However, recall from Section 4.6 that while a sinusoid has a peak factor of $\sqrt{2}$, the gaussian noise has no theoretical upper limit. The peak factor for gaussian noise has therefore been set on the basis of practical considerations at $V_{peak}/V_{rms} = 4$, because the probability of exceeding this peak amplitude is only 0.01%—a fairly safe choice. This means that in order to accommodate 99.99% of noise peaks without limiting and without losing linearity, we must add another $20\log 4 = 12$ dB to the dynamic-range requirement imposed on the transducer and the measuring setup. This is a rather significant increase. Clearly, an excessive T_h could result in compression which leads to a decreased Y-factor and hence to a pessimistic T_e and noise factor. As a general rule, T_h should be somewhat greater than the T_e to be measured [6], but a great

disparity is not recommended. Mumford and Scheibe [7] give an interesting analysis of the optimum value of Y, based on the accuracy of reading a linear power meter. It is shown there that the optimum occurs when the noise figure to be measured is 1.5 dB lower than the excess-noise ratio N_R of the noise source, defined as

$$N_R = 10 \log\left(\frac{T_h}{T_0} - 1\right) \tag{11-4}$$

This criterion is, however, quite broad and noncritical.

The second important consideration is the source impedance Z_g. It has by now been repeatedly emphasized that both T_e and F are functions of Z_g, and must be measured with the same source impedance as that encountered in the actual system where the transducer is ultimately used. Otherwise, the measured values of T_e and F do not apply, because the transducer would be operated under different terminal conditions. Typically, the T_e or F measurement is carried out with well-defined source impedances (50 ohms for coaxial systems, waveguide impedance for waveguide sources). In practice, however, the transducer, such as a low-noise amplifier, may well face an unknown antenna with other than a perfectly matched source impedance. This creates problems, particularly when T_e is low. A discussion of this problem and numerical data on resultant uncertainties appear in Wait [8]. Precise correction curves cannot be given because the error depends on both the magnitude and phase of the reflection coefficients, and, as will be remembered from Section 9.9, also on the internal configuration of the transducer.

The third point to be noted is that the source impedance presented to the transducer by the input noise source must remain constant while the noise power is changed from $kT_h\,df$ to $kT_c\,df$ in the course of the measurement. The necessity for this requirement should be obvious, but the reason for concern will become apparent when practical noise sources are discussed. We are not talking about major differences here. Even minor changes in the VSWR at the noise-source/transducer interface could be significant. Pastori [9] gives the following estimate for the error in measured noise figure (in dB) caused by this mechanism:

$$d(NF) = \pm\frac{4.34(dM/M)Y}{Y-1} \tag{11-5}$$

where

$$\frac{dM}{M} = \frac{(1 + |\Gamma_c|\,|\Gamma_r|)^2}{(1 - |\Gamma_h|\,|\Gamma_r|)^2} - 1 \tag{11-6}$$

and Γ_c is the reflection coefficient of the cold source; Γ_h the reflection

coefficient of the hot source; and Γ_r the input reflection coefficient of the transducer under test.

As an example of (11-5), consider the following case:

$$\text{VSWR of the hot source} = 1.20{:}1, \qquad |\Gamma_h| = 0.091$$

$$\text{VSWR of the cold source} = 1.50{:}1, \qquad |\Gamma_c| = 0.20$$

$$\text{Input VSWR of the amplifier under test} = 1.6{:}1, \qquad |\Gamma_r| = 0.23$$

Equation (11-6) yields

$$\frac{dM}{M} = 0.142$$

and therefore

$$d(NF) = \pm\frac{0.62Y}{Y-1}\ \text{dB} \tag{11-7}$$

Assume further that a noise figure of 4 dB was measured using this source, and that the excess-noise ratio N_R of the source was 15.7 dB. The Y-factor corresponding to these numbers is 12 dB or $Y = 15.85$. Thus from (11-7) the uncertainty would be ± 0.7 dB.

Note that the change in the source impedance also affects the available power gain of the transducer under test which, in turn, will introduce an error in the measured Y-factor. We shall discuss this in Section 11.3.

The fourth point has to do with the flatness of the noise power delivered by the source over the frequency range of interest. The power density must be uniform over *all* significant responses of the transducer. Included in this requirement are not just the responses used for information transmission, such as the signal and in some applications also the first image ω_{-1}, but all other spurious ones as well.

The principal source of error in noise measurement can thus be summarized as follows:

(a)   Uncertainty in the values of T_h and T_c.

(b)   Error in the reading of Y.

(c)   Difference between the output impedances of the hot and cold noise sources, or the difference in the output impedance of one source when fired and unfired.

(d)   Difference between the noise-source impedance and the generator impedance used in actual operation.

(e)   Inadequate dynamic range of either the transducer under test or of the test instrumentation.

Each of these areas will be discussed in some detail in subsequent sections.

11.2. NOISE SOURCES AND NOISE STANDARDS

This section will describe a number of commonly used white-noise generators used in the laboratory to measure T_e and F. This field has in the past 10–15 years undergone a significant change. It was not so long ago when various types of gas-discharge tubes, either coaxial or waveguide, ruled supreme in noise test setups. At UHF and VHF frequencies the temperature-limited vacuum diode was a widely used device. The advent of the solid-state diode noise source, which can in one unit conveniently cover the entire frequency range up to 26 GHz, has resulted in a rapid demise of first the coaxial noise tube and the VHF/UHF noise diode, and in time also the bulky waveguide noise tube. Now it is only at frequencies above 20 GHz where the waveguide noise source is still holding its own. However, in view of the many gas-discharge noise sources still in use throughout the industry, a brief description of these devices will be given in this section.

Hot/Cold Noise Sources

In theory the most accurate and most simple noise generator would be a heated resistor. Its available noise power is given by (11-1). The application of such a device is, conceptually, equally simple. The resistor is first heated to T_h, yielding a noise power of $kT_h \, df$. It is then cooled to a low temperature T_c to generate a noise power of $kT_c \, df$. The rest is represented by Eqs. (8-25) and (9-7).

As easy as this sounds in theory, it is remarkably difficult to realize properly in practice. The first practical problem lies in constructing such a resistor source physically, because the temperature range over which a physical resistor could operate is obviously restricted. Although vacuum-enclosed tungsten filaments can reach 2000–3000 °C, practical heated-resistor noise sources barely reach 1000 °C, with typical values in mere hundreds of degrees [10]. Second, no physical resistor would have the same resistance value at both temperature extremes. Yet the constancy of the source resistance is mandatory. This implies that practical hot/cold noise sources must use two resistor elements—one held at a high temperature and the other one at a cold temperature. The user must somehow switch between the two, which is a drawback. However, the resistances of the two elements can now be precisely trimmed to be the same at their respective operating temperatures.

It should be appreciated that the limited temperature range achievable with this type of noise source is not always a disadvantage. True, according to (11-3) the hot/cold source is best suited only for low-noise, that is, low T_e, measurements. Fortuitously, this also happens to be the region where maximum accuracy is required—something that a properly constructed hot/cold source can provide.

Heated-resistor-type noise sources have been constructed in a variety of ways in coaxial and waveguide configurations, the latter all the way to

millimeter wavelengths (WR12). The principal problems, aside from the complexities of construction, are: (a) determining the exact temperature of the element; (b) achieving a uniform temperature throughout the element; and (c) evaluating the effect of the temperature gradient along the connecting transmission line. Since the supporting structure will also heat up, the output noise temperature is therefore an integrated effect, differing somewhat from the temperature of the resistor element itself.

The other extreme—a cooled-resistor noise source—is not easy to construct either. The basic technique is to immerse an appropriate resistor in some cryogenic fluid—liquid oxygen, nitrogen, or helium—but the problems are similar to those of the heated resistor. A good summary of the design and construction details of hot/cold noise sources can be found in refs. [10–12].

In view of the difficulties in maintaining and measuring arbitrary temperatures at either end of the temperature range, it becomes logical to utilize temperatures provided by natural processes, such as the boiling points of liquids. Thus a popular form of hot/cold noise source, usable to 9 GHz (Fig. 11.1), uses two precision 50-ohm terminations—one held at the boiling point of water, $T_h = 373.2°\,$K, and the other at the boiling point of liquid nitrogen, $T_c = 77.3°\,$K (at standard pressure). The typical measurement procedure using such a source involves physical shifting of the unit under test from the "hot" output connector to the "cold" output connector, unless the source
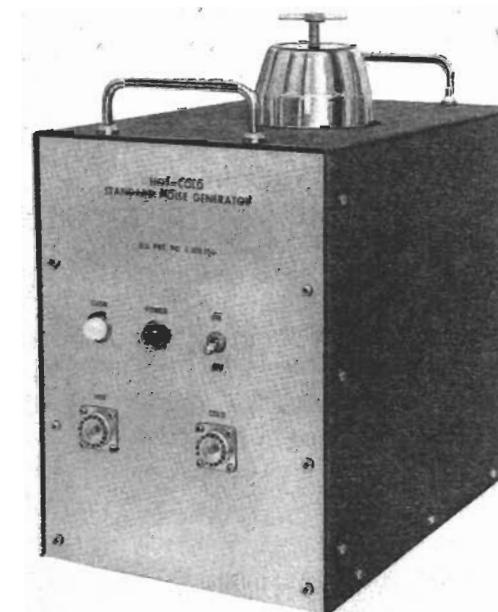

FIGURE 11.1.   AILTECH 7009 hot/cold noise source. (Courtesy of Eaton Corporation, Electronic Instrumentation Division.)

has a built-in switch. Neither method is without potential problems. The connect/disconnect process causes wear of the critical connector interface, and there is always the problem of repeatability unless care is used in mating the connectors. The microwave switch, on the other hand, adds loss and could also contribute VSWR problems.

Additional points regarding the use of hot/cold noise sources in general and liquid-N_2 type in particular are:

(a) The precise cold temperature T_c depends on the boiling point of nitrogen which, in turn, depends on the atmospheric pressure and altitude (0.3°K per 1000 feet), and on the purity of the liquid nitrogen at hand. Appropriate correction factors are usually provided by the manufacturer of the source.

(b) When shifting from one connector or waveguide port to the other, great care must be exercised to guarantee repeatability. As always, it is a good idea to take at least two readings and average the results. Coaxial hot/cold sources use precision connectors which must never be forced. Losses and irregularities at this interface contribute the bulk of errors in hot/cold measurements. The condition of the connector should be inspected before each test. Periodic cleaning of the connector interface in accordance with manufacturer's recommendations is also advised.

(c) Any coaxial or waveguide adapter, added between the output connector of the source and the unit under test, contributes loss and therefore modifies the reference temperature of the noise in accordance with

$$T_i' = \frac{T_i}{L} + T_x\left(\frac{L-1}{L}\right) \qquad (11\text{-}8)$$

where T_i is the rated output noise temperature, either T_h or T_c, of the source; L is the added loss; and T_x is the physical temperature of the lossy adapter. The effect of a matched resistive loss can thus be corrected. The effect of a spurious VSWR on the other hand can be quite insidious. This is because the noise waveform has a random phase and the power loss due to such a mismatch cannot be easily calculated. More will be said about this problem in Section 11.3.

(d) If the noise temperature of the two resistors, T_h and T_c, differ from those assumed, an error is introduced in the calculated F in accordance with (11-14). The nomogram in Fig. 11.2 gives the error in noise figure NF as a function of known errors in T_h and T_c. This is easily derived from (11-3) and (11-14).

The hot/cold noise source is potentially the most accurate noise standard in practical use. In view of its low excess-noise ratio, it is best suited for low-noise measurements. Since the hot and the cold resistors have been carefully matched to have the same value at their respective temperatures, the source mismatch

Reference values:
T_h = 373.2°K
T_c = 77.3°K
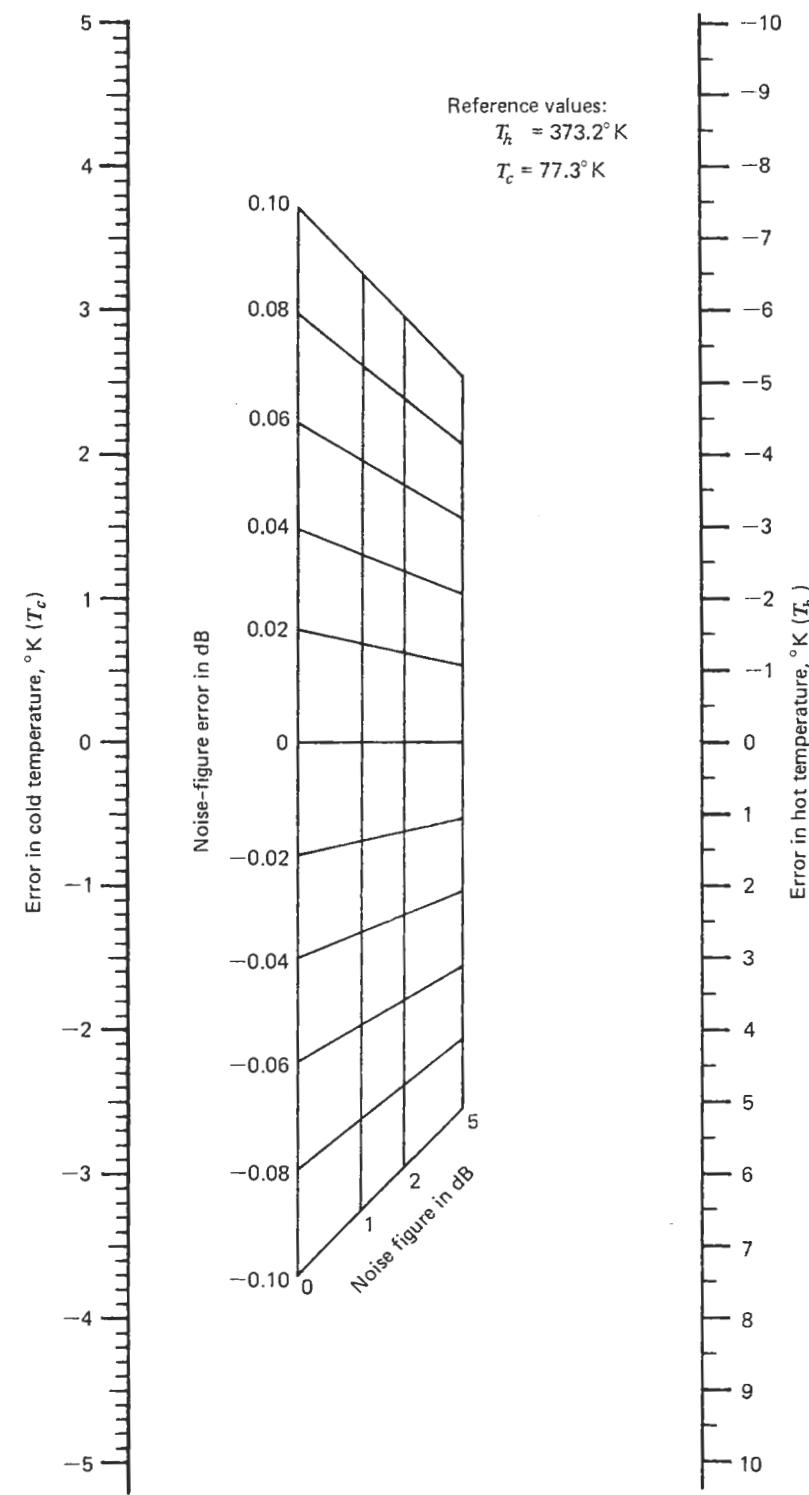
FIGURE 11.2. Error in noise figure as a function of errors in T_h and T_c.

problem between hot and cold states is significantly reduced. The drawback of the hot/cold source is that the technique is slow and cannot be easily automated. Also, the need for liquid nitrogen or other cryogenic fluid complicates the use of this source in the field. It is therefore used mainly in stationary applications. With these reservations, however, the hot/cold noise source is the best noise reference available.

Gas-Discharge Noise Sources

The gas-discharge noise source was demonstrated by W. W. Mumford in 1949 [13]. He showed that the positive column of a low-pressure gaseous discharge emits noise of uniform spectral density over a wide band of frequencies. This noise behaves as thermal noise from a fictitious resistance at a very high temperature, typically in excess of $10,000\,°K$. The apparent temperature is due to electrons of high average energy contained in the plasma column of the tube. The gas-discharge tube is not a primary noise standard because its noise temperature, and hence the actual noise output, depend on variables such as the gas pressure, the type of gas used, and the diameter of the plasma column. Its noise output is not readily calculable. However, once the tube is calibrated against a primary standard, its noise output remains remarkably independent of the operating current and the temperature of the glass envelope.

The principal asset of gas-discharge noise sources is their high noise output, corresponding to an N_R in the neighborhood of 16 dB. A further advantage is the operational simplicity since the switching from hot to cold reference is accomplished easily by turning the high voltage supply on and off. There are no cryogenic liquids to worry about. This capability is a prerequisite for automated noise measurements.

As a result of the high effective T_h, there is no need for a cryogenic T_c. For gas-tube noise sources, the T_c is therefore the ambient temperature or, more accurately, $T_c = T_0 = 290\,°K$. The important parameter in the gas tube is the method of coupling between the plasma column and the transmission line. In waveguide sources the tube is mounted through the broad walls of the waveguide at an angle of 7–10° (Fig. 11.3). In coaxial structures the coupling is achieved by wrapping a helical transmission line around the envelope of the tube. In either configuration one end of the transmission line is terminated in a matched load, ideally at T_0, which provides the reference noise T_c when the tube is turned off. It should be noted that the noise output of the gas-discharge noise sources when fired is independent of the terminating resistance. The gas used in these sources is typically argon. Helium and neon would yield even higher noise levels, but helium has a short life time and generates excessive heat in operation. Neon is similar to argon, but harder to fire.

The difference in the input VSWR as the tube is fired and extinguished is quite small for waveguide noise sources. Typically, both VSWRs are below 1.2:1. The coaxial sources, used primarily below 5 GHz, could have a VSWR = 1.15:1 when fired, but 1.5:1 when turned off. This is one reason why the coaxial noise tube could offer so little competition to the solid-state source.
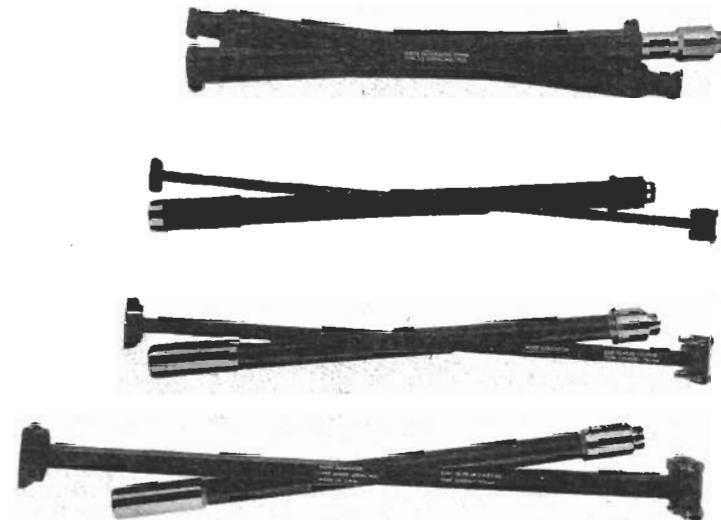
FIGURE 11.3.   Typical gas-discharge waveguide noise sources. (Courtesy of Eaton Corporation, Electronic Instrumentation Division.)

Additional drawbacks for the gas-discharge noise tubes in general are their fragility and the need for a high-voltage power supply.

At present, the solid-state diode noise source has almost completely superseded the gas tube at frequencies below 18 GHz. The commercial production of gas-tube noise sources has therefore been largely discontinued except at millimeter wavelengths. A large number of these sources is still in daily use, and some familiarity with them is therefore desirable. Section 11.3 will discuss some operational peculiarities worth noting.

Solid-State Diode Noise Source

The most popular noise source in RF work today is the solid-state diode source. Its advantages are such that it has completely taken over the 10 MHz to 18 GHz frequency range. The solid-state source started out as a zener diode, but special diodes were soon developed to yield the very broadband, high noise output that characterizes this device at present. The diode is basically a *PIN* structure, reverse biased (typically 13 volts at 50 mA) at the breakdown [14]. The noise output of one single unit comfortably covers the full 10 MHz to 18 GHz range, and at least one commercial unit, the Eaton Model 7626 (Fig. 11.4) extends this range to 26 GHz. The flatness of the N_R over the full range depends on the particular unit, but it can be as low as ± 0.4 dB. In practice the manufacturer provides a calibration table with each unit, with the measured N_R given at each GHz, plus several extra points below 1 GHz. The worst-case calibration uncertainty is typically ± 0.3 dB.
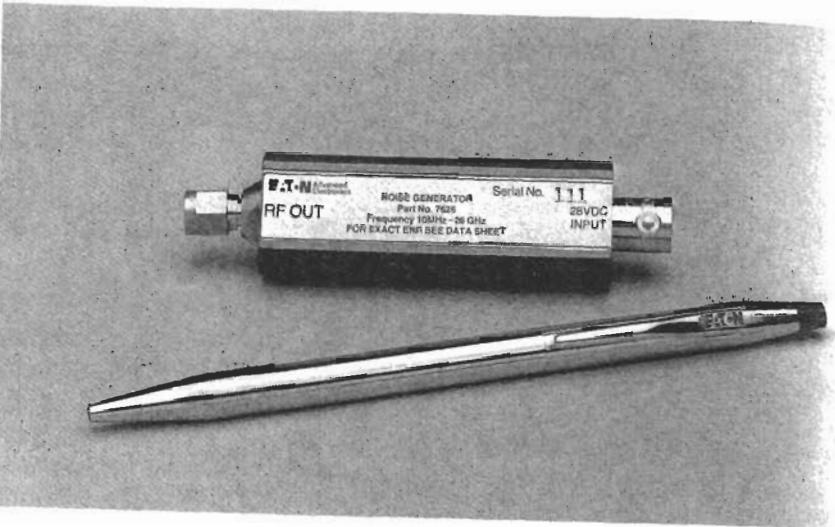
FIGURE 11.4. AILTECH 7626 solid-state diode noise source. (Courtesy of Eaton Corporation, Electronic Instrumentation Division.)

The basic diode can provide excess-noise ratios of 32–38 dB, but this cannot always be utilized directly because the impedance of the diode varies from approximately 20 ohms when turned on, to perhaps 400 ohms or so when turned off. In order to provide a constant output impedance, the unit must be heavily padded. This keeps the VSWR typically at 1.3:1 maximum, and invariant from one diode state to another. The final N_R is thereby lowered to the customary value of around 15 dB. However, this inherent high-noise capability is of crucial importance in applications such as the measurement of the radar noise figure, where the noise source is permanently coupled to the system through a high-value directional coupler so that automatic noise measurements could be performed under operating conditions. In this case the directional coupler acts as the pad, but this still leaves a solid 15 dB or so of excess-noise ratio to provide adequate noise for testing. Recall from the discussion at the beginning of this chapter and from (11-3) that for high noise figures (common to radar systems) the T_h must have a correspondingly high value.

Additional advantages of the solid-state noise source are its small size and light weight, and the fact that only a $+28$ VDC ± 1 volt power supply is required for biasing. In view of these features, it is not surprising that the solid-state noise source has found such a wide acceptance.

Vacuum-Diode Noise Source

Another white-noise source is the temperature-limited vacuum-tube diode. A typical example is the (since discontinued) HP Model 343A. These units are

rarely used nowadays, but some may still be around. Hence a brief description is warranted.

The noise output of temperature-limited diodes is due to a randomly fluctuating current component, superimposed on the steady-state saturation current flowing through the tube, from the heated cathode to the anode. Schottky showed that the mean-square value of this fluctuating current is

$$\overline{i^2} = 2qI_s \, df \tag{11-9}$$

where q is the electron charge, 1.59×10^{-19} coulombs and I_s the dc saturation current in amperes. The latter depends on the temperature of the cathode. As the anode voltage e_b is increased, the dc current i_b increases until a saturation value I_s for the given cathode temperature is reached. At this point all electrons emitted by the cathode are collected by the anode and no further increase in current is observed unless the cathode temperature is raised. The temperature-limited diode source can therefore be considered a primary standard because its output is calculable from (11-9) and (11-10). The N_R of the HP 343A, for example, is 5.2 ± 0.5 dB. The vacuum diode is usable only in the 10–600 MHz VHF/UHF region because at higher frequencies circuit parasitics, and, more importantly, the electron transit time, cause the noise output to drop.

A practical temperature-limited noise source consists of the diode connected in parallel with a resistor R. It can be shown that the equivalent noise temperature of such a source, when the diode is turned on, is

$$T_{eq} = \left(\frac{q}{2k}\right)I_s R + T_x = 5761 I_s R + T_x \tag{11-10}$$

where T_x is the physical temperature of the resistor R. The first observation is that the noise output is a function of R. This is an advantage in that R could be selected to correspond to the source resistance that the transducer under test would eventually see. In fact, some commercial units have such an adjustment, permitting a selection of several discrete source resistances. Also, there is virtually no change in the source resistance as the diode is turned on and off. This is due to the extremely high plate resistance of the diode in the temperature-limited region (theoretically infinite). As long as the added resistance R remains below 1000 ohms or so, the effect of the shunting plate resistance is entirely negligible. On the debit side, the value of R must be precisely known to avoid substantial errors.

11.3. METHODS OF NOISE MEASUREMENT, F AND T_e

The measurement procedure for T_e and F is in principle the same for all dispersed, that is, white-noise, sources, although there are slight differences in

the handling of the individual sources. These will be outlined shortly. It is more important to realize that the correct interpretation and the use of measured results requires a deeper understanding of the underlying principles than is commonly assumed. This is particularly true for cases where the measuring setup contains a mixer, or better yet, if the mixer itself is the component under test. It is therefore recommended that the reader review Chapter 10 at this point in order to have the subtleties of multi-response circuits fresh in the mind.

Fundamentals of Using Broadband (White) Noise Sources

This section will describe the application of the broadband noise source, such as the hot/cold source, to the measurement of T_e and F. A number of important general aspects of using broadband noise sources will also be covered here. It is therefore suggested that this section be carefully reviewed even if the intent is to work only with gas-discharge or solid-state noise sources.

Figure 11.5 outlines the basic setup, which consists of the noise source, the unit under test (UUT), and an output indicator which reads the output power. The details of the UUT will be left arbitrary at this point, except that the UUT must be linear and capable of being characterized by an input noise temperature T_e and an available power gain (or loss) G. The linearity is a prerequisite for the definition of noise factor to apply.

The first step is to connect the input terminal at (a) of the UUT to the cold output terminal of the hot/cold source. The resultant output power at (b) is as in (8-23),

$$N_c = k\,df(T_c + T_e)G \tag{11-11}$$

The unit is next shifted to the hot terminal which results in an output power

$$N_h = k\,df(T_h + T_e)G \tag{11-12}$$

The ratio of N_h/N_c is the Y-factor. Therefore as before in (8-25):

$$T_e = \frac{T_h - YT_c}{Y - 1} \tag{11-13}$$
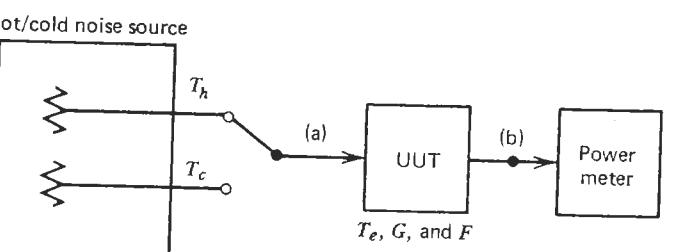
FIGURE 11.5. Basic noise-measurement setup using hot/cold noise source.

Since from (9-7), $F = 1 + T_e/T_0$, the noise factor becomes

$$F = \frac{(T_h/T_0 - 1) - Y(T_c/T_0 - 1)}{Y - 1} \tag{11-14}$$

A more typical setup is shown in Fig. 11.6. The UUT could be a low-noise amplifier, followed by a postamplifier to increase the output power level to measurable values. In addition, a precisely calibrated variable attenuator is connected just before the output indicator. This attenuator is adjusted to keep the power level at the indicator constant. The Y-factor in dB is thus read as the difference of the two attenuator settings at N_c and N_h. The attenuator is clearly not mandatory, but it is recommended because the attenuator settings can, in general, be read with greater accuracy than the scale of a power meter. More importantly, since the power meter now sees a constant level, all errors due to possible nonlinearities in the meter reading are eliminated. Note also that this technique permits the use of any output-indicating device, however nonlinear.

If the attenuator is used, the preceding gain (LNA plus the postamplifier) must be high enough so that the effect of the attenuator change is negligible at point (a). If this cannot be guaranteed, then the Y-factor must be measured directly on the power meter and hence a linear instrument is a must.

It is clear from Fig. 11.6 that what we get from (11-13) and (11-14) are the T_e and the F of the combined system. In order to find the noise parameters of the low-noise amplifier alone, we must subtract out the noise contribution of the second stage—the postamplifier. Again, it is assumed here that the components after the postamplifier no longer matter. For the configuration in Fig. 11.6 the task is trivial since both amplifiers have only one response. Thus

$$T_e = T_e' - \frac{T_e''}{G} \tag{11-15}$$
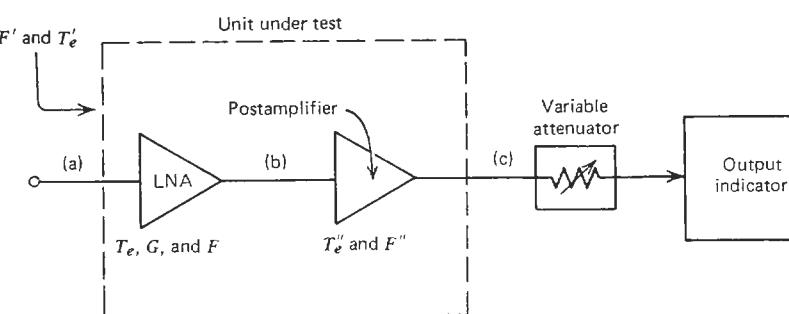
FIGURE 11.6. The use of a variable attenuator to measure the Y-factor.
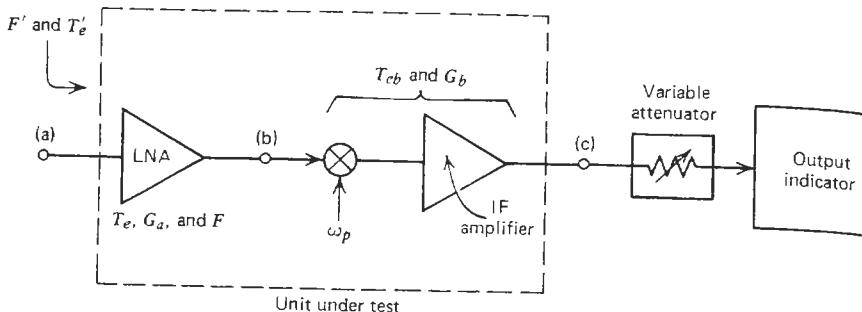
FIGURE 11.7. Typical noise-measurement setup using a downconverting mixer.

and

$$F = F' - \left(\frac{F'' - 1}{G}\right) \tag{11-16}$$

If G is sufficiently high or T_e'' and F'' are sufficiently low, then $T_e \approx T_e'$ and $F \approx F'$, and a correction may not even be necessary. The values of T_e'' and F'' must, of course, be known or be measured separately.

The method shown in Fig. 11.6 is a perfectly good one and is at times used. In practice, however, high-frequency postamplifiers may not always be available. Furthermore, a new postamplifier and variable attenuator may be required whenever the test frequency changes. The more prevalent technique is to downconvert the amplifier output to some convenient IF frequency, where high-gain amplification is readily available. Typically, an IF frequency below 100 MHz is used. This setup is illustrated in Fig. 11.7.

When a mixer is used as the second stage, the subtraction of the second-stage effect is more complicated. We know from Chapter 10 that a mixer, being a dual-response* transducer, can have two noise factors, F_{SSB} and F_{DSB}, which are to be used in accordance with the particular application of the unit. Several questions therefore arise. When a broadband noise source is connected to a system as in Fig. 11.7 and Eq. (11-14) is used, which noise factor do we get? And once we have F', what should the F'' be in (11-16) in order to yield the correct values for F and T_e?

In order to answer these questions, consider the diagram in Fig. 11.8. The LNA is a broadband unit, covering both signal and image bands, although the two gains, G_{a1} and G_{a2}, may be unequal. We assume, however, that the input noise temperature T_e of the LNA is the same at both frequencies ω_1 and ω_{-1}, and that the gain of the mixer/IF amplifier combination is the same in both responses, that is, $G_{b1} = G_{b2} = G_b$.

Apply a broadband noise source to the input of the LNA so that equal amounts of noise power are injected into the signal and the image bands.

---

*As discussed in Chapter 10, a mixer can have higher-order image responses. These are generally much weaker than the two principal ones—signal and image. Hence only a dual-response case is considered here. that is, we shall use the F_{DSB}, not the F_{MSB}.
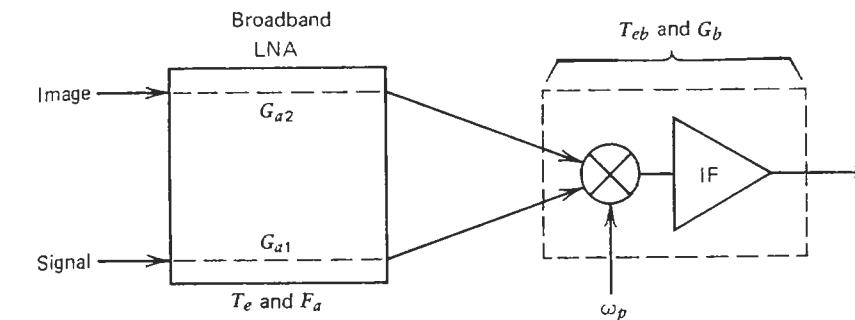
FIGURE 11.8. Illustration of noise relationships in measuring noise parameters with a downconverting mixer.

Repeating the earlier derivation of N_h and N_c yields

$$N_h = k\,df\left[\underbrace{(T_h + T_e)G_{a1} + T_{eb}}_{\text{signal}} + \underbrace{(T_h + T_e)G_{a2} + T_{eb}}_{\text{image}}\right]G_b \tag{11-17}$$

and

$$N_c = k\,df\left[(T_c + T_e)G_{a1} + T_{eb} + (T_c + T_e)G_{a2} + T_{eb}\right]G_b \tag{11-18}$$

Using $N_c = YN_h$, and solving for T_e yields ·

$$T_e = \frac{T_h - YT_c}{Y - 1} - \frac{2T_{eb}}{G_{a1}(1 + G_{a2}/G_{a1})} \tag{11-19}$$

Since according to (9-7) for any single-response unit (such as the LNA)

$$F = 1 + \frac{T_e}{T_0}$$

we obtain from (11-19)

$$F_a = \frac{(T_h/T_0 - 1) - Y(T_c/T_0 - 1)}{Y - 1} - \frac{2(T_{eb}/T_0) + 2 - 2}{G_{a1}(1 + G_{a2}/G_{a1})}$$

$$= \frac{(T_h/T_0 - 1) - Y(T_c/T_0 - 1)}{Y - 1} - \frac{2[(F_b)_{\text{DSB}} - 1]}{G_{a1}(1 + G_{a2}/G_{a1})} \tag{11-20}$$

But we have encountered this expression before as (10-50). Now, by comparing (11-20) to (10-50), it follows that the first term in (11-20), which is the calculated result from measured data, is the double-sideband noise factor of the assembly.

As a general rule, remember therefore that whenever a broadband noise source is used on a setup containing a multi-response component, the multi-sideband noise factor is obtained. This can also be seen qualitatively since in the process of measurement the "signal," that is, $kT_h\,df$ when the noise source is fired, appears in all responses. The same happens at the other extreme when $kT_c\,df$ is applied. This answers the first question.

The second question concerning the subtraction of the second-stage effect is also answered by (11-20). The second term there shows that what we need is the F_{DSB} of the mixer/IF amplifier combination (easily measured with the same broadband noise source, but applied at point (b) in Fig. 11.7. Given this information and the two gains G_{a1} and G_{a2}, the noise factor of the LNA, F_a, follows directly from (11-20).

Let us now carefully recapitulate what we have done. Starting with the double-sideband system noise factor obtained from the hot/cold experiment [first term in (11-20)], we isolated the F_a of the LNA. This by itself is obviously an unambiguous number and nontroublesome. Now we reverse the situation. Let this broadband amplifier be used in a system which includes a broadband mixer as in Fig. 11.8, and has an IF frequency low enough so that both signal and image channels are within the passband of the LNA. Assume further that the system has signals only in the "signal" response. The danger is now that we may, out of habit, write for the overall noise factor:

$$F_T = F_a + \frac{F_{mixer} - 1}{G_a} \tag{11-21}$$

without realizing that the effect of the image response has thereby been neglected. True, we would use for F_{mixer} the double-sideband value obtained from a typical measurement using a broadband noise source, but this is not enough. Compare (11-21) to (10-50) and it becomes apparent that what we have in (11-21) is the $(F_T)_{DSB}$ instead of the correct $(F_T)_{SSB}$. The latter requires the use of (10-46), and our initial result, hastily calculated from (11-21), could be off by 3 dB. The situation may get even more involved if (10-42) and (10-47) do not hold. Hence caution is in order when dealing with noise in dual-response systems.

The noise factor of the broadband mixer or the mixer/IF amplifier combination is easy to measure. The pertinent power levels are:

$$N_h = k\,df(T_h + T_{eb})(G_{b1} + G_{b2}) \tag{11-22}$$

and

$$N_c = k\,df(T_c + T_{eb})(G_{b1} + G_{b2}) \tag{11-23}$$

where the second term in parentheses expresses the effect of the two responses.

From (11-22) and (11-23):

$$T_{eb} = \frac{T_h - YT_c}{Y - 1} \tag{11-24}$$

and in view of the general rule covering multi-response transducers after (11-20),

$$(F_b)_{DSB} = 1 + \frac{T_{eb}}{T_0} = \frac{(T_h/T_0 - 1) - Y(T_c/T_0 - 1)}{Y - 1} \tag{11-25}$$

where T_{eb} is the input noise temperature of each response (same for all responses in accordance with the IRE definition); G_{b1} the total gain in the *signal* response, equal to the product of the conversion loss $(1/L_1)$ and the IF amplifier gain); and G_{b2} the total gain in the *image* response, defined the same way as for G_{b1}.

Note that although we have allowed for possibly unequal gains in the signal and the image responses, the final result is independent of this.

If the mixer has its image internally terminated, then, as far as the input and output terminals of the mixer are concerned, we have a single-response system and the usual procedure of measuring noise parameters of such systems applies. The effect of the image is implicitly contained in the measured conversion loss and noise factor of the only accessible (signal) response, and no further allowances need be made.

This comment also applies to the case shown in Fig. 10.6 if the image filter and the isolator are considered as part of the image termination system and hence as part of the mixer. A measurement of noise factor performed at the input of the image filter yields a unique F for the entire combination and the F_{SSB}/F_{DSB} distinction becomes irrelevant.

At times one may wish to measure the single-sideband noise factor of the mixer directly. It should be clear from Sections 10.6 and 10.7 that simply placing a reactive, lossless filter ahead of the mixer to block the image noise will not do. The image termination realized in this manner can affect the operation of the mixer and no useful information is gained. The only way to perform this measurement meaningfully is as shown in Figs. 10.6 and 11.9. The image filter prevents the noise from the noise source from reaching the mixer. The derivation follows the pattern used earlier in Eqs. (11-22) and (11-23), except that the image always sees (looking out from the mixer) the noise temperature of the isolator, which is T_0. The details are given in ref. 14 and will not be repeated here. The result is

$$(F_{SSB} - 1)L_i L_f = \frac{T_h/T_0 - Y(T_c/T_0)}{Y - 1} \tag{11-26}$$

where L_i and L_f are the resistive insertion losses of the isolator and the filter,
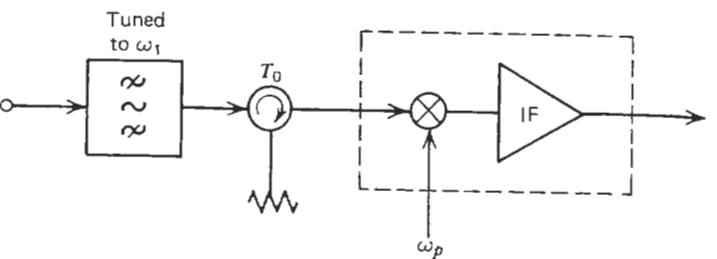
FIGURE 11.9.   Measurement of single-sideband noise factors.

respectively, at the signal frequency. If both losses are negligible ($L \approx 1$), then (11-26) reduces to the familiar

$$F_{SSB} = \frac{(T_h/T_0 - 1) - Y(T_c/T_0 - 1)}{Y - 1} \qquad (11\text{-}27)$$

verifying that the hot/cold measurement data obtained in this manner do indeed lead to F_{SSB}.

Application of Physical Noise Sources

The basic theory of using a broadband noise source, detailed in the previous section, applies equally to all white noise sources—the hot/cold noise source, the gas tube, the solid-state diode source, and the temperature-limited diode. We shall now turn to those operational aspects which are peculiar to one or the other noise source.

The use of the hot/cold source is straightforward, and much of it was covered in Section 11.2. Aside from the precautions listed there, little need be added. As is true with any precision equipment—and the hot/cold source certainly qualifies as one—a careful review of the operating manual is recommended before proceeding.

Since the gas-discharge tubes and the solid-state sources have a much higher noise temperature T_h, they are particularly suited for measuring medium and high noise factors in accordance with (11-3). The formulas for T_e and F follow directly from (11-13) and (11-14) by letting $T_c = T_0$. Thus

$$T_e = \frac{T_h - YT_0}{Y - 1} \qquad (11\text{-}28)$$

and

$$F = \frac{T_h/T_0 - 1}{Y - 1} \qquad (11\text{-}29)$$

As seen, the noise figure NF becomes directly proportional to the excess-noise ratio N_R in dB:

$$NF = 10 \log(T_h/T_0 - 1) - 10 \log(Y - 1) \qquad (11\text{-}30)$$

This shows why it is important to know N_R as accurately as possible.

The measurement procedure using these noise sources is very simple. The noise source is turned on and off, and the corresponding N_h and N_c are determined. The same pitfalls and precautions apply here as were outlined in Section 11.2, but there are a few additional ones:

(a)   The source VSWR of a coaxial gas tube varies appreciably between the fired and cold states. This change may have an effect on the available gain G of the unit being tested. This effect can be measured by injecting a CW test signal, some 15–20 dB greater than the maximum output of the noise source, into the input of the unit under test (through a 20–30 dB directional coupler). As the noise tube is turned on and off, the change in the output level of the CW signal, which is a measure of the change in gain, ΔG, is noted. The measured Y-factor should then be corrected as follows [15]:

$$\text{Corrected } Y \text{ (dB)} = \text{measured } Y \text{ (dB)} - \Delta G \text{ (dB)} \qquad (11\text{-}31)$$

where $\Delta G = G_{\text{tube on}} - G_{\text{tube off}}$. A positive ΔG thus calls for a decrease in the measured Y-factor.

(b)   Continued use of the gas tube causes a heating of the tube structure since the voltage drop across the tube is approximately 100 volts and the tube current is in the range of 50–250 mA. Some 20 W could be dissipated in the tube. This, in turn, means that when the tube is turned off, the "cold" temperature of the termination could be appreciably higher than $290°$ K. Equation (11-11) thus changes to

$$N_c'' = k\, df\, G(T_0 + \Delta T + T_e) \qquad (11\text{-}32)$$

and the measured Y-factor becomes

$$\frac{N_h}{N_c''} = Y'' = \frac{T_h + T_e}{T_0 + \Delta T + T_e} \qquad (11\text{-}33)$$

which is for positive ΔT smaller than it would be for $\Delta T = 0$. If this smaller Y factor is unknowingly used in (11-29), the calculated F would be too high. Figure 11.10 shows* the correction to be applied [16]. One way to avoid this problem is shown in Fig. 11.11. Instead of relying on the gas tube to provide the cold reference T_0, a matched variable attenuator, kept at T_0, could be used.

*If the T_h used is not equal to 10,000 ° K, then Fig. 11.10 must be recalculated using (11-33) and (11-29).
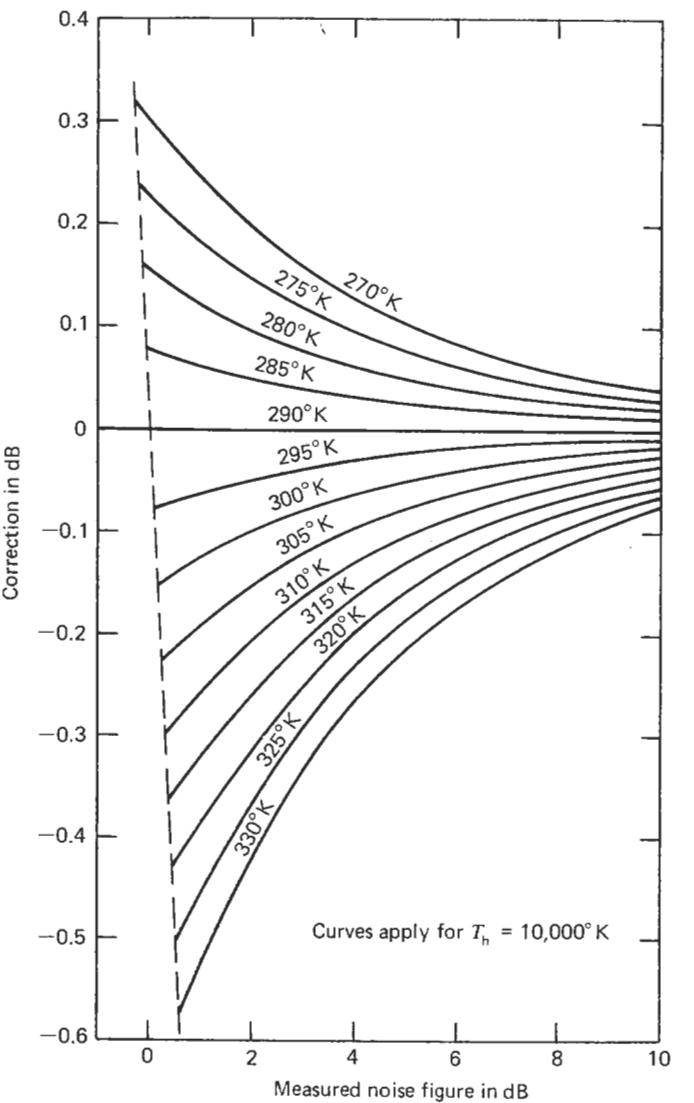
FIGURE 11.10. Error in noise figure due to temperature of the terminating load unequal to $T_0 = 290°$K. (Courtesy of Eaton Corporation, Electronic Instrumentation Divison.)
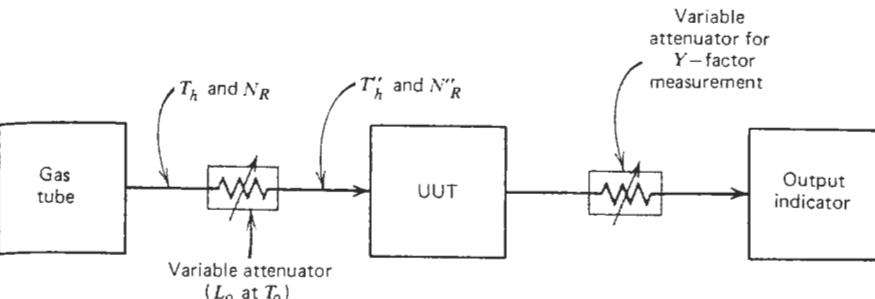
FIGURE 11.11. A method of eliminating termination problems of gas-discharge noise sources.

For N_h the attenuator is set to zero. For N_c the attenuator is set at a high value. Needless to say, the insertion loss of a practical attenuator at zero setting, L_0, must be taken into account, and the new T_h'', N_R, and NF now become:

$$T_h'' = \frac{T_h}{L_0} + \left(\frac{L_0 - 1}{L_0}\right) T_0 \tag{11-34}$$

$$N_R'' = 10\log\left(\frac{T_h''}{T_0} - 1\right) = 10\log\left(\frac{T_h}{T_0} - 1\right) - L_0 \text{ (dB)}$$

$$= N_R - L_0 \text{ (dB)} \tag{11-35}$$

$$NF = N_R'' - 10\log(Y - 1) \tag{11-36}$$

(c)   When coaxial gas tubes are used, ignition spikes may couple from the high-voltage leads to the helical RF transmission line. This was no problem with the relatively robust front ends of yesterday's receivers. The sensitive solid-state devices of today can be easily damaged, however. A high-pass filter may therefore be required to block this leakage. The waveguide gas tube is free of this problem, because the waveguide is below the cutoff for frequencies in the spike.

The advantage of all rapidly switchable noise sources is that the measurement can be automated. This eliminates the tedium of manual measurements, but even more importantly, allows a continuous indication of noise figure as the circuit is being adjusted. In this technique the noise source is turned on and off at a rapid rate, typically 500–1000 Hz, and the resultant pulse train of alternately N_h and N_c from the output of the unit under test is processed as usual. It is downconverted and sent to a special noise-figure meter where it is amplified at the IF frequency. After detection and appropriate processing to obtain the N_h/N_c and the F, the end result is displayed directly on a meter. The AILTECH Models 7514 (Fig. 11.12) and 2075 (Fig. 11.13) are examples of such instruments. Another example is Hewlett-Packard 8970A (Fig. 11.14). These instruments have built-in calibration circuits and adjustments to scale
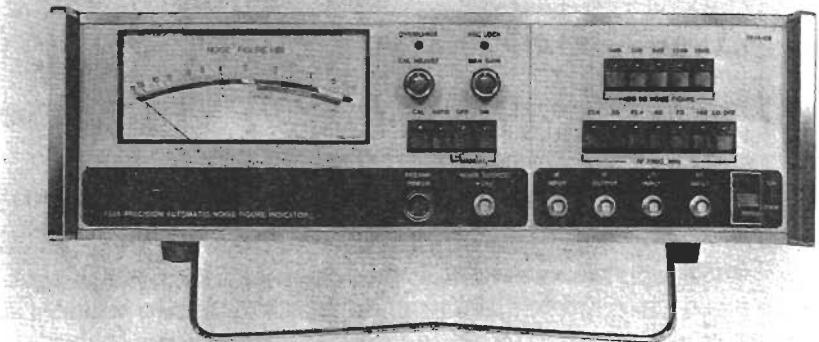
FIGURE 11.12. AILTECH 7514 precision automatic noise-figure indicator. (Courtesy of Eaton Corporation, Electronic Instrumentation Division.)

FIGURE 11.14. Hewlett-Packard 8970A noise-figure meter. (Courtesy of Hewlett-Packard Company.)

the output reading properly for a range of excess-noise ratios. The 8970A is controlled by a microprocessor which, among other things, automatically corrects for the second-stage effect.

Signal-Generator Method

The principle of the signal-generator method is illustrated in Fig. 11.15. This technique relies on the calibrated CW output of a signal generator. The other reference level is the noise from the internal termination of the signal generator, which is assumed to be at T_0. The measurement procedure and derivation (for a single-response system) follows below:

(a) Observe the output power level N_o on the output indicator when the generator is turned off and emitting only white thermal noise at T_0:

$$N_o = kBG(T_0 + T_e) \qquad (11\text{-}37)$$

FIGURE 11.13. AILTECH 2075 noise-gain analyzer. (Courtesy of Eaton Corporation, Electronic Instrumentation Division.)
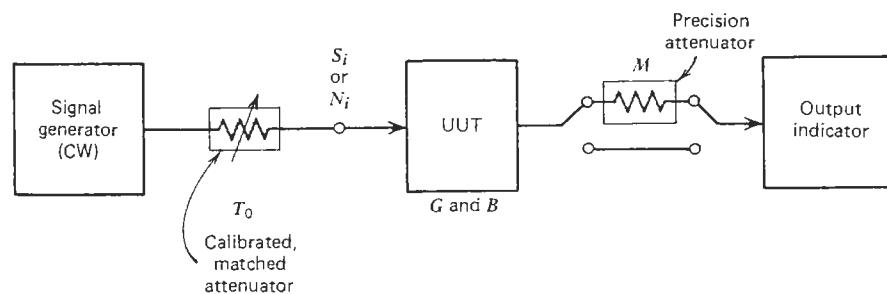
FIGURE 11.15. Measurement setup for the signal generator method.

(b) Increase the setting of the IF attenuator by 3.01 dB. This corresponds to a factor $M = 2$. The new output becomes

$$N_o'' = N_o/2 \qquad (11\text{-}38)$$

(c) Turn on the signal generator and increase its output S_i until the level observed under (a) is restored:

$$N_o = N_o'' + S_i(G/2) \qquad (11\text{-}39)$$

because the overall gain G has also been halved.

(d) From (11-37) to (11-39),

$$T_e = \frac{S_i}{kB} - T_0 \qquad (11\text{-}40)$$

and*

$$F = 1 + \frac{T_e}{T_0} = \frac{S_i}{kBT_0} \qquad (11\text{-}41)$$

Notice that whereas the dispersed-signal measurement did not require the measurement of the transducer noise bandwidth B, this method does. The reason is that when the *broadband* noise was applied first as a high-level "signal" (kT_hB) and then as a low-level noise (kT_cB or kT_0B), the bandwidth information was included in both and cancelled out. This, as we saw earlier, is also the reason why the broadband noise source yields a multi-sideband, or in most practical cases, the double-sideband noise factor.

With the signal-generator method it is necessary to evaluate the noise bandwidth B accurately because its value has a direct bearing on the calculated noise factor. Furthermore, the CW power S_i must be at the reference frequency f_0 used in calculating the noise bandwidth B_0 (see Fig. 3.9). In determining the noise bandwidth, the *entire* bandwidth through which the initial noise kT_0B in step (a) may enter is to be included. For example, some amplifiers have spurious responses outside their rated passband. These, too, must be considered because the initial noise reading kT_0B includes their effect.

Multi-response transducers must be treated more carefully. Here it is necessary to account for all significant image bands. In practice, it is usually only the principal image ω_{-1} which is of importance. Still, the general expression for this case, which is analogous to (11-37), is (10-8). Using (10-8) in (11-39) we obtain

$$T_e = \frac{S_i}{kB_1\mathcal{R}} - T_0 \qquad (11\text{-}42)$$

*The factor M in step (b) need not be 2. If another value is used, then (11-41) becomes $F = S_i/kBT_0(M-1)$.

and from (10-24),

$$F_{\text{SSB}} = \frac{S_i}{kT_0B_1} \qquad (11\text{-}43)$$

Note that (11-43) is independent of the response factor $\mathcal{R}$, and the latter need not be evaluated. This is not surprising since in contrast to the broadband-noise-source technique where the test noise fills all responses, the signal-generator technique applies the CW signal to one response only while still allowing noise to enter through all responses. The signal-generator method thus gives inherently the single-sideband noise factor, and the absence of $\mathcal{R}$ in (11-43) reflects this.

By the definition of noise factor, the reference temperature, and hence the assumed generator temperature, is $T_0 = 290\,°\text{K}$. If it happens to be T_x, then F_{SSB} is calculated from

$$F_{\text{SSB}} = \frac{S_i}{kT_0B_1} + \left(1 - \frac{T_x}{T_0}\right)\mathcal{R} \qquad (11\text{-}44)$$

which is a simple extension of (11-43). Note, however, that now the response factor $\mathcal{R}$ *does* matter and must be determined.

When instead of F_{SSB} the multi-sideband noise factor F_{MSB} is desired, but data from the signal-generator measurement are given, then the response factor is required. From (10-28) and (11-42),

$$F_{\text{MSB}} = \frac{S_i}{kT_0B_1\mathcal{R}} \qquad (11\text{-}45)$$

It was already mentioned that S_i must be applied at the reference frequency of the noise bandwidth. When measuring a dual-response transducer, the noise bandwidth of the image, B_{-1}, must use the reference frequency such that

$$\omega_{-1} = \omega_p \pm \omega_1 \qquad (11\text{-}46)$$

Since a typical dual-response receiver almost invariably uses a narrowband IF amplifier somewhere in the chain, both noise bandwidths B_1 and B_{-1} become equal to the noise bandwidth of the narrowband IF section. They thus cancel out of $\mathcal{R}$ which simplifies the procedure.

There are a few precautions to be observed with this method. A typical signal generator has an output power control, but this may use a waveguide below cutoff to achieve accurate calibration. Clearly, when such an attenuator is turned to its maximum setting in step (a), the transducer under test will not see T_0. The safest way therefore is to insert a separate, calibrated and matched attenuator (Fig. 11.15) which can be guaranteed to be at T_0.

As in all noise measurements, the source impedance must at all times equal the actual source impedance with which the transducer is to be used. This is another reason why the attenuator technique is recommended. When using the internal power adjustment of an unknown signal generator, the exact source impedance remains in doubt.

The signal-generator method is suited for high noise factors where ultimate accuracy may not be required. In this sense we have another fortuitous case because the accuracy of the noise factor determined in this manner is at the mercy of the noise bandwidth. For simplicity, the 3-dB bandwidth instead of the noise bandwidth is often used. The resultant error is then of limited consequence. Note, however, that for multi-response cases the response factor $\mathscr{R}$ must still be correctly applied. Compared to the difference between the 3-dB and the noise bandwidth, or even a minor out-of-band spurious amplifier response, the multi-response effect is a gross one and cannot be overlooked.

Direct Method

This method follows from the basic definition of noise factor (9-2). Theoretically, given the measured total output noise N_o over some bandwidth B, and the total gain between the input and the output, the noise factor is calculated directly from $F = N_o/kT_oGB$. In logarithmic terms,

$$NF = 10 \log N_o - 10 \log B - 10 \log G + 204 \qquad (11\text{-}47)$$

or

$$NF = 10 \log N_o' - 10 \log G - 204 \qquad (11\text{-}48)$$

if noise density is used.

It is readily seen that this method is usable primarily for measuring high noise figures. One example would be the noise figure of an upconverting exciter, or the exciter/transmitter combination. Since by the definition of noise factor the source noise must be at T_0, high gain is required to yield a conveniently measurable output. Unless F is a high number, (11-48) means subtracting two large, nearly equal numbers. This means problems with accuracy. Also, the noise bandwidth B must be known.

The output power N_o could be measured with a power meter if the output level is adequate, but more typically the direct noise-factor measurement uses a spectrum analyzer (Fig. 11.16). This technique has a side benefit because spurious signals, such as hum or other undesirable RF pickup, can be spotted on the spectrum analyzer while the power meter would simply indicate the total power.

It is clear from Fig. 11.16 that what is being measured is not just the amplifier under test, F_a, but the total noise factor F_T, including the effect of the postamplifier and the spectrum analyzer. This is because N_o includes all
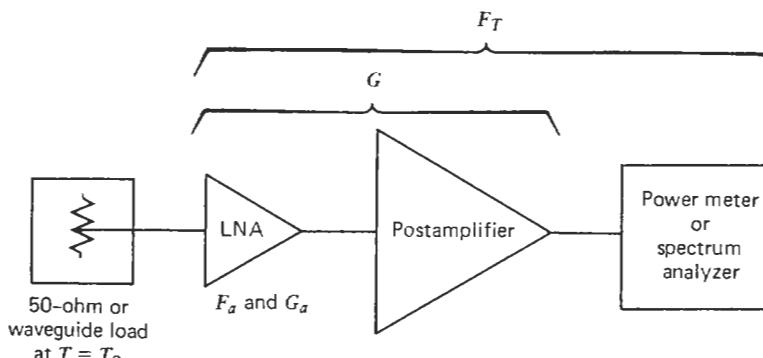
FIGURE 11.16. Measurement setup for the direct-noise-measuring technique.

noise contributions. If the first amplifier has sufficient gain, then, as before, $F_a \approx F_T$. If not, then the usual cascade formula must be applied to obtain F_a.

A procedure to measure F by this technique is described in a Hewlett-Packard application note [17]. The spectrum analyzer is switched to linear scale to increase resolution and the noise level is read in volts. Equation (11-48) then becomes

$$NF = 10 \log \frac{\overline{v^2}}{R} - 10 \log B - 10 \log G + 204 \qquad (11\text{-}49)$$

where v is the noise voltage read on the spectrum analyzer and R the system impedance level, typically 50 ohms.

The bandwidth is the noise bandwidth of the spectrum analyzer (usually the narrowest bandwidth in the chain). For a so-called gaussian filter shape the noise bandwidth B is 1.2 times the 3-dB bandwidth listed on the bandwidth setting. We must also include the 1.05-dB correction factor from Section 11.5 for the internal detector of the analyzer. Equation (11-49) thus becomes for a 50-ohm system:

$$NF = 10 \log \overline{v^2} - 10 \log B_{3\,dB} - G \text{ (dB)} - 10 \log(1.2)(50)$$

$$- 10 \log(kT_0) + 1.05$$

$$= 10 \log \overline{v^2} - 10 \log B_{3\,dB} - G \text{ (dB)} + 187.3 \qquad (11\text{-}50)$$

Modern spectrum analyzers such as the HP Model 8566A accomplish much of this automatically, and yield the final noise density in dBm/Hz as a direct readout, including all correction factors. We can then use

$$NF = 10 \log N_o' - G \text{ (dB)} + 174 \qquad (11\text{-}51)$$

where N_o' is the noise density determined in this manner.

11.4. MEASUREMENT OF TWO-PORT NOISE PARAMETERS

The significance of the noise parameters F_0, R_n, G_{g0}, and B_{g0} (and their duals G_n, R_{g0}, and X_{g0}) of a linear two-port transducer was explained in Sections 6.2 and 9.9. We shall now show how these quantities can be determined from a series of noise-factor measurements.* The starting point is Eq. (9-65), repeated here for convenience:

$$F = F_0 + \frac{R_n}{G_g}\left[(G_g - G_{g0})^2 + (B_g - B_{g0})^2\right] \qquad (11\text{-}52)$$

The classical method of measuring these parameters is straightforward. The spot noise factor F (i.e., using a bandwidth df as narrow as possible) of the transducer at the given test frequency is measured as a function of the source susceptance B_g. The latter is suitably varied, while the real part of Y_g is kept constant (Fig. 11.17a). When the tabulated results are plotted, a curve similar to Fig. 11.17b is obtained. The B_g corresponding to F_{min} is the B_{g0}, that is, the value which satisfies (9-61).

The procedure is now repeated except that G_g is varied (Fig. 11.17a) and B_g is kept constant. The result is Fig. 11.18 and hence we can get G_{g0} [see (9-63)].

Now return to (11-52) and rewrite it as

$$F = F_0 + R_n\left(\frac{|Y_g - Y_{g0}|^2}{G_g}\right)$$

$$\equiv F_0 + R_n Y'' \qquad (11\text{-}53)$$

But (11-53) is the equation of a straight line with R_n as the slope and F_0 as the offset. Thus, using the earlier tabulated values of $F(B_g)$ and $F(G_g)$, and calculating additional values of Y'' as needed, Fig. 11.19 can be plotted. The slope R_n and the offset F_0 can now be easily determined.

The process is self-checking in the sense that if the measured data are correct, the plot in Fig. 11.19 will be a straight line. Any deviation is an indication of experimental error. The two most common sources of error are nonlinearities in the transducer or in the test equipment (the output indicator, perhaps), and the possibility that the variations in G_g and B_g were not entirely independent.

The difficult part of this procedure is varying the real and imaginary parts of Y_g independently of each other while trying to measure F at each point. Adamian and Uhlir [20–22] have developed a different method which is based on measuring the output noise power rather than the noise factor F as a

*Only the admittance representation is treated here. The dual follows in an analogous manner. Details are in refs. [18] and [19].

(a)

(b)
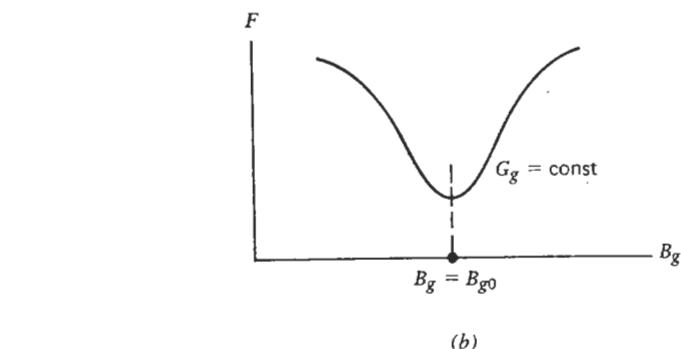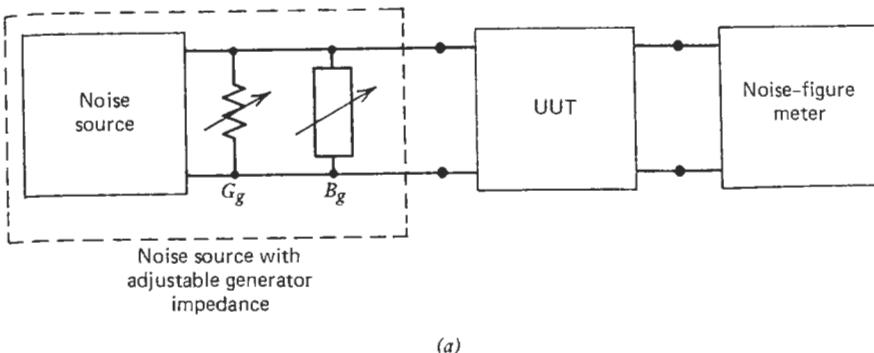
FIGURE 11.17. Measurement of two-port noise parameters: (a) basic setup and (b) variation of F with varying source susceptance.
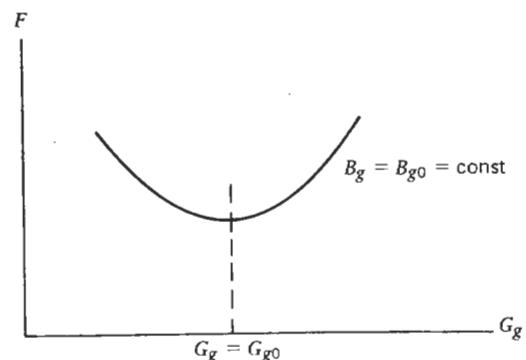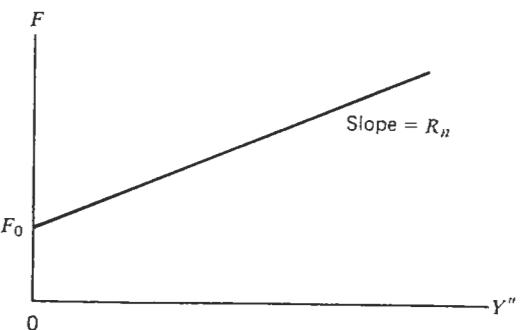
FIGURE 11.18. Variation of F with varying source conductance.

FIGURE 11.19. Determination of F_0 and R_n from measured Y_{g0}.

function of the source admittance Y_g. Their technique requires knowing the input admittance of the transducer under test, and measuring the noise factor at one known source admittance. The rest of the data are obtained by measuring the resultant output noise power as various passive one-ports are connected to the input of the transducer.

11.5. CONSIDERATIONS OF THE OUTPUT SENSOR

An important component in any noise-measuring setup is the output sensor which gives the output noise values N_h and N_c. It determines the value and the ultimate accuracy of the measured Y-factor. The fundamental requirement for this instrument is that its reading be proportional to either the power, $\overline{v^2}$ or $\overline{i^2}$ of the output noise. There are three broad classes of instruments that would do this:

(a)   Those activated by magnetic forces.
(b)   Quadratic devices.
(c)   Heat-sensing devices.

The first type requires high driving powers and is therefore not suitable for our purpose. The second type is represented by a diode operating in its square-law region. Typically, low-barrier Schottky diodes are used in commercially available diode power sensors such as the HP 8484A [23, 24]. It must be emphasized at the start that diodes act as quadratic devices only for low RF power levels, less than -20 dBm or so, depending on the specific diode type. In this region the dc output voltage or current is approximately proportional to the input RF power, that is, to the RF voltage squared. Hence the name square-law detector. For large signals the diode acts as a conventional half-wave rectifier, and its dc output voltage now is proportional to the *average* voltage of the rectified RF input. This by itself does not necessarily disqualify the diode detector as a power sensing device, provided the user understands clearly

what is involved. In case of *sine* waves all that is required is to multiply this measured average dc voltage by 2.22 [see Eqs. (4-33) and (4-34)] to get the rms value. We must remember, however, that for *gaussian noise* the ratio of rms to average is 2.51 instead [from (4-31) and (4-32)]. Hence, if a diode sensor and a dc voltmeter were used to measure high-level RF or microwave noise, the result should be multiplied by 2.51 instead to get the rms values of the noise voltage. Many ac instruments are in reality average-reading meters, with built-in rectifiers and with scales calibrated to read rms. It may be possible to use such a meter directly on low-frequency noise, but the indicated rms value will be too low by $20 \log(2.51/2.22) = 1.05$ dB. As long as this correction is taken into account, rms values of noise can be measured in this manner. One warning, however, is in order. Meters with rectifiers have sometimes large capacitors across their meter movements to increase the time constant. This makes them act more like peak-reading instruments. In view of the fundamental difference between the irregular noise peaks and the regular peak value of a sine wave (Section 4.6), it is clear that peak-reading meters cannot be used for measuring power in noise waveforms.

Diode detector power sensors have the outstanding advantage that they can measure down to -70 dBm, which is a good 35–40 dB lower than typical thermistor-type heat-sensing devices. The principal limiting mechanism is thermal—drift caused by thermal voltage generated by temperature gradients that appear across the diode. Commercial units have carefully designed sensor mounts to overcome this problem as much as possible.

The heat-sensing power sensors convert the incident RF energy to heat which is then converted into an appropriate electrical signal to activate the output meter. These sensors come in two principal categories—bolometers and thermocouple sensors. The former, in turn, include the thermistors, which are semiconductors with negative temperature coefficients, and barretters, which rely on a measurable change in resistance of a very thin piece of wire (positive coefficient). The barretters are by their very nature more easily destroyed by accidental overloads, and consequently it is the more rugged thermistor which is typically used in measuring average RF and microwave power. The HP 8478A and the HP 486A series are examples of typical thermistor mounts.

The thermocouple sensors (HP 8481A, Narda 8420, etc.) utilize the voltage generated by the thermocouple as a result of a temperature gradient along the device. These sensors have a broader dynamic range than the thermistor types. Obviously, all heat-sensing sensors measure true power and are therefore suitable for noise measurements. The choice between the different types is based on the required dynamic range, required accuracy, cost, and so on.

In practical noise tests it may not even be necessary to worry about these considerations because the modern noise figure meters, such as the ones shown in Figs. 11.12 to 11.14 have their own integral sensors, and the final reading includes all necessary corrections. However, at times the need may arise to select a power sensor separately. The preceding overview should then prove helpful. For further information, ref. [24] should also be consulted.

We pointed out earlier in Section 11.3 and Fig. 11.6 that it is possible to circumvent the power-sensor problem entirely if a precision attenuator is included in the setup. This keeps the power level at the sensor constant and any nonlinearities no longer matter. The prerequisite clearly is that the gain in the preceding stages of the system be high enough to mask any possible effect of varying the attenuator.

11.6.   SUMMARY

(a)   Primary noise standards (or sources) include:
   (1)   Heated resistor sources (hot/cold sources).
   (2)   Temperature-limited vacuum diodes.

(b)   Secondary noise standards include:
   (1)   Gas-discharge tubes.
   (2)   Solid-state diode noise sources.
   These must be calibrated against a primary standard.

(c)   The "hot" reference temperature T_h of a noise source should be somewhat greater than T_e to be measured for best Y-factor accuracy.

(d)   Y-factor: N_h/N_c, usually expressed in dB.

(e)   Excess-noise ratio of noise source:

$$N_R = 10 \log\left(\frac{T_h}{T_0} - 1 \right)$$

(f)   Generator impedance of noise source must remain constant when fired or unfired, and must equal the impedance seen by the device under test when in actual operation.

(g)   For any broadband (white) noise source:

$$T_e = \frac{T_h - YT_c}{Y - 1}.$$

(h)   For single-response transducers

$$F = 1 + \frac{T_e}{T_0} = \frac{[(T_h/T_0) - 1] - Y[(T_c/T_0) - 1]}{Y - 1}$$

and

$$NF = N_R - 10 \log(Y - 1).$$

(i)   For transducers with multiple responses, the broadband noise source measurement always yields the multi-sideband noise factor.

(j)   Commonly used excess-noise ratios for gas tubes and solid-state noise sources are 15–16 dB.

(k)   Gas-tube noise sources can have a significant change in internal impedance when fired and unfired. Also, terminating resistance temperature may deviate from T_0 due to the heating of tube structure.

(l)   Signal-generator method of measuring F:

$$F = \frac{S_i}{kBT_0(M - 1)}$$

where M is the ratio of reference output powers.

(m)   Direct method uses actual measured output noise to calculate the noise factor from the basic definition.

(n)   Noise parameters of two-port transducers (F_0, G_{g0}, and B_{g0}) can be determined from noise-factor measurements using different generator impedances.

REFERENCES

1.   S. Letzter and N. Webster, "Noise in Amplifiers," *IEEE Spectrum*, pp. 67–75, August 1970.
2.   A. van der Ziel, *Noise*, Prentice-Hall, Englewood Cliffs, NJ, 1954.
3.   C. D. Motchenbacher and F. C. Fitchen, *Low-Noise Electronic Design*, Wiley, New York, 1973.
4.   E. G. Nielsen, "Amplifier Noise," in R. F. Shea (Ed.), *Amplifier Handbook*, McGraw-Hill, New York, 1966, Chapter 7.
5.   A. P. G. Peterson and E. E. Gross, Jr., *Handbook of Noise Measurement*, General Radio Company, Concord, Mass., 1972.
6.   W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960, Chapter 8.
7.   W. W. Mumford and E. H. Scheibe, *Noise Performance Factors in Communication Systems*, Horizon House, Dedham, 1968.
8.   D. F. Wait, "Measurement of Amplifier Noise," *Microwave J.*, pp. 25–29, January 1973.
9.   W. E. Pastori, "New Equation Analyzes Mismatch Errors in Noise Measurements," *Microwaves*, pp. 58–63, April 1968.
10.   C. K. S. Miller, W. C. Daywitt, and M. G. Arthur, "Noise Standards, Measurements, and Receiver Noise Definitions," *Proc. IEEE*, vol. 55, pp. 865–877, June 1967.
11.   C. T. Stelzried, "Microwave Thermal Noise Standards," *IEEE Trans. MTT*, vol. MTT-16, pp. 646–655, September 1968.
12.   P. A. H. Hart, "Standard Noise Sources," *Philips Tech. Rev.*, vol. 23, pp. 293–309, 1961/1962.
13.   W. W. Mumford, "A Broadband Microwave Noise Source," *Bell Syst. Tech. J.*, vol. 28, pp. 608–618, October 1949.
14.   W. E. Pastori and S. Van Sleteren, Eaton Corporation, private communication.
15.   R. L. Steven, "A Guide to Accurate Noise Measurement," *Microwaves*, pp. 14–21, July 1964.
16.   W. E. Pastori, *Measuring Microwave Noise, A Series of Notes*, Airborne Instruments Laboratory, Deer Park, New York.

17. *Spectrum Analysis...Noise Measurements*, Hewlett-Packard Application Note 150-4, April 1974.

18. E. G. Nielsen, ref. [4].

19. H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE*, vol. 44, pp. 811–818, June 1956.

20. V. Adamian and A. Uhlir, Jr., "A Novel Procedure for Receiver Noise Characterization," *IEEE Trans. Instr. Meas.*, vol. IM-22, pp. 181–182, June 1973.

21. V. Adamian and A. Uhlir, Jr., "Input Reflection and Noise Figure Dependence on Source Reflection: Determination Using Noise Generator," presented at Electro-82, Boston, Mass., May 1982.

22. V. Adamian and A. Uhlir, Jr., "Simplified Noise Evaluation of Microwave Receivers," submitted to the *IEEE Trans. Instr. Meas.*, 1983.

23. S. Wetenkamp, "Comparing Single-Diode and Balanced Power Detectors," *Microwaves & RF*, pp. 120–122, August 1983.

24. "Fundamentals of RF and Microwave Power Measurements," Hewlett-Packard Application Note 64-1, August 1977.

12 | SIGNAL AND NOISE IN CRYSTAL DETECTORS

The complete theory of solid-state crystal detectors is a broad topic because these devices come in many forms and designs [1] and are used in a wide variety of applications—crystal-video receivers, modulators/demodulators, power detectors, and so on. The terminology itself varies with usage. One hears terms such as video detector, video crystal, quadratic detector, crystal detector, square-law detector, and so on—some derived from a particular application and others from a functional viewpoint. The purpose of this chapter is not to present a detailed analysis of such detectors because in-depth treatments of the subject are readily available elsewhere (e.g., [1, 2–4] to list a few). Instead, we shall limit ourselves to an overview of the more commonly used methods of characterizing the signal and noise properties of crystal detectors when used in RF circuits. The distinction between this and the foregoing chapters is that crystal detectors are inherently nonlinear devices, and it is this very property which is exploited in their application. The nonlinearity is not just a hidden mechanism which yields in the end a *linear* output as in mixers, but it is utilized to obtain a *nonlinear* input/output relationship. The present chapter is therefore a follow-up to the preceding material which was devoted to noise in linear transducers. It is hoped that bringing both subjects under one roof will help the RF engineer who constantly encounters both situations in daily work.

12.1.   GENERAL PARAMETERS

When an amplitude-modulated RF signal is applied to a crystal detector as in Fig. 12.1, the modulation is recovered. This action is due to the rectifying
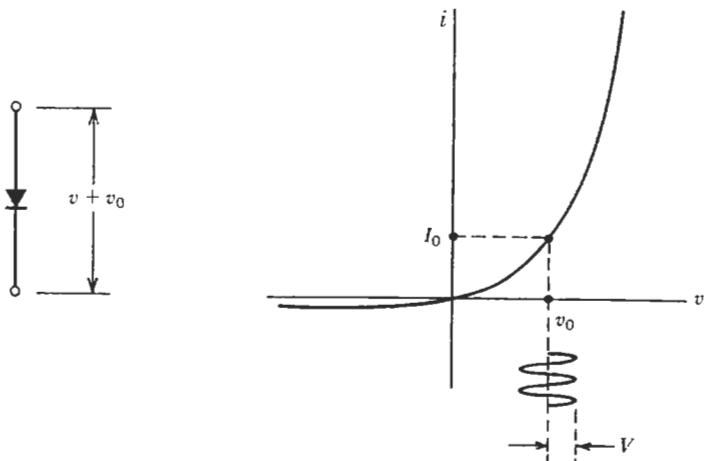
FIGURE 12.1.   Square-law region of a diode.

nonlinearity in the current/voltage characteristic of the diode which is given by

$$i = I_s\left[\exp\left(\frac{qv}{\eta kT}\right) - 1\right] \tag{12-1}$$

where the terms are as previously defined for (10-62). The nonlinearity can be expressed in terms of a Taylor expansion about an arbitrary bias point v_0 (which could be zero):

$$i = f(v_0 + v) = f(v_0) + f'(v_0) + \frac{f''(v_0)}{2!}v^2 + \cdots \tag{12-2}$$

Since microwave detectors are typically operated with a dc bias, the combined signal becomes $v = v_0 + V\cos\omega t$. The amplitude could be a constant as in the CW case, or it could contain amplitude modulation in the form $V(t) = V_c(1 + mf(t))$, where m is the modulation index, $f(t)$ an arbitrary time function, and V_c the amplitude of the unmodulated carrier.

When this expression is substituted into (12-2) and the latter is expanded, we obtain [1]:

$$i = I_0 + \left(\frac{V^2}{4}f''(v_0) + \frac{V^4}{64}f''''(v_0)\right) + \left(Vf'(v_0) + \frac{V^3}{8}f'''(v_0)\right)\cos\omega t \tag{12-3}$$

where only the dc (or video) and the carrier-frequency terms have been

retained. The term I_0 is merely the dc current at the bias point and of no interest in this discussion. Since the physical RF detector (Fig. 12.2) contains appropriate chokes and bypass capacitors to separate the high-frequency RF input ωt and its harmonics $n\omega t$ from the detected low-frequency components contained in $V(t)$, the last term can also be ignored as far as detected output is concerned. The middle term in (12-3) is the incremental increase in output current, δi, which is due to the applied RF signal. Basically, for small RF signals, that is, for low-level detection, it is the term $V^2 f''(v_0)/4$ which contributes nearly all of the detected amplitude modulation, or as is usually said, the video output.

We note from (12-3) that the detected output is essentially proportional to: (a) the *curvature* $f''(v_0)$, or the degree of nonlinearity of the current/voltage characteristic; and (b) the square of the RF input amplitude V^2. Since V^2 is, in turn, proportional to the RF input power, the familiar square-law condition is obtained and the reason for the term "quadratic detector" becomes clear.

Another term in (12-3) which deserves attention is $f'(v_0)$. This is the slope of the current/voltage characteristic at v_0:

$$f'(v_0) = \frac{di}{dv}\bigg|_{v=v_0} \equiv \frac{1}{R_b} \tag{12-4}$$

which is also called the barrier resistance, and is calculated from (12-1) and (12-4) as

$$R_b = \frac{\eta kT}{q(I_s + I_0)} \tag{12-5}$$

where I_0 is the bias current. In an ideal diode, free of any resistive or reactive parasitics, this would also be the video output resistance or video resistance R_v. Actual diodes have a parasitic series resistance R_s, contributed by the bulk
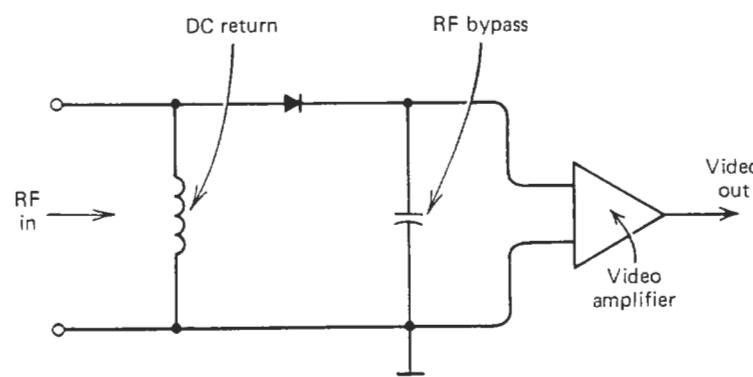
FIGURE 12.2.   Equivalent circuit of an RF detector and video amplifier.

effect of the crystal and other losses in the diode package. Hence the actual video resistance is (ignoring the transforming effect of reactive parasitics)

$$R_v = R_s + R_b \qquad (12\text{-}6)$$

As shown by Uhlir [3], and Cowley and Sorensen [1], a useful parameter is the short-circuit current sensitivity, sometimes called the current responsivity β of the detector, defined as

$$\beta = \frac{\delta i}{P} \qquad (12\text{-}7)$$

where P is the RF input power and δi the short-circuit output current for the unbiased detector, or the change in output current as a result of P if a dc bias current is used. β is expressed in mA/mW or μA/μW. As shown in [1], β consists of two terms:

$$\beta = \beta_0(1 + \Delta) \qquad (12\text{-}8)$$

where

$$\beta_0 = \frac{f''(v_0)}{2f'(v_0)} \qquad (12\text{-}9)$$

is the low-level value of β when the operation is perfectly square law. The quantity Δ given by [1] as

$$\Delta = \frac{P}{8f'(v_0)}\left(\frac{f''''(v_0)}{f''(v_0)} - 2\frac{f'''(v_0)}{f'(v_0)}\right) \qquad (12\text{-}10)$$

expresses the deviation from the square-law condition, which becomes increasingly significant as the RF input power P increases.

An analogous term, the open-circuit voltage sensitivity can also be defined as

$$\gamma = \frac{\delta v}{P} \qquad (12\text{-}11)$$

where δv is defined analogously to δi. The term γ is expressed in either mV/mW or μV/μW.

The β and the γ are related through

$$\beta = \frac{\gamma}{R_v} \qquad (12\text{-}12)$$

It should be stated that in a practical detector β and γ are dependent on detector bias, load resistance, and operating frequency. The dependence on the input power P is particularly critical as shown by (12-8) and (12-10).

Using (12-4), (12-7), and ref. [3], we can draw an equivalent circuit for the output of the detector, as shown in Fig. 12.3. Let us next calculate the video power dissipated in the load conductance G_L:

$$P_L = V_L^2 G_L = \frac{(\beta P)^2 G_L}{(1/R_v + G_L)^2} = \frac{(\beta P R_v)^2 R_L}{(R_v + R_L)^2} \qquad (12\text{-}13)$$
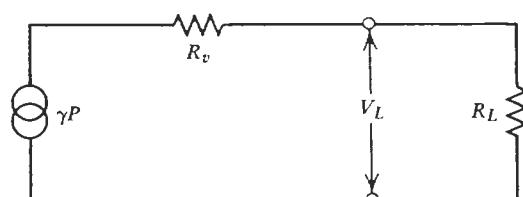
From (12-13) we can define the conversion loss of the detector as

$$L_c = \frac{\text{input RF power}}{\text{output (video) power}} = \frac{P}{P_L} = \frac{(R_v + R_L)^2}{(\beta R_v)^2 R_L P} \qquad (12\text{-}14)$$

where it was assumed (as is the case throughout this chapter) that the RF input of the detector is matched to the input frequency so that no part of the incident power is lost through reflections.

Note from (12-14) that L_c is a function of P. While the dependence of the detector performance on β and R_v is well understood, the dependence on P is often overlooked. Consider L_c for two different input powers, P_1 and P_2. From (12-14),

$$\frac{L_{c1}}{L_{c2}} = \frac{P_2}{P_1} \qquad (12\text{-}15)$$

Or
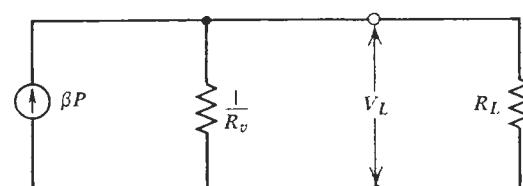
FIGURE 12.3.   Equivalent circuit of an RF detector at video output.

and from the definition of L_c the ratio of output powers is

$$\frac{P_{L2}}{P_{L1}} = \left(\frac{P_2}{P_1}\right)^2 \tag{12-16}$$

Thus the common practice of inserting a pad of L dB ahead of the detector for better VSWR will increase the conversion loss of the detector by *twice* the attenuation in dB, that is, by $2L$.

12.2.   SENSITIVITY AND NOISE IN DETECTORS

The concept of noise factor is not applicable to nonlinear transducers because the definition of F, either (9-2) or (9-3), assumes that the transducer gain G is independent of power level. The gain terms then cancel out, rendering F a unique number, invariant with applied signal levels. As we saw in (12-15), this does not hold for square-law detectors.

The noisiness of detectors must therefore be expressed differently. Before an overview of the various methods is given, it is worth noting that detectors are normally used in conjunction with a video amplifier, and the two are characterized as a combined video receiver (Fig. 12.4), where the gain, bandwidth, and noise factor of the video amplifier, G_v, B_v, and F_v, respectively, also appear in the final equations. It must be again remembered that F_v is to be based on a source resistance equal to the detector output resistance R_v.

Minimum Detectable Signal (MDS)

One of the earliest attempts to define the sensitivity of detectors is the minimum detectable or discernible signal (MDS), which was defined as the smallest RF input signal which could be observed against the given background noise level at the output. Typically, the calibrated RF signal was pulse modulated and its peak value was slowly increased until the pulses became just noticeable in the noise. The RF peak value at this instant was then regarded as the MDS in dBm. This technique is highly subjective and the results can vary by several dB, particularly among inexperienced observers. It should be noted

RF in → P → [Detector] → [Video amplifier] → Video out

γ, β, R_v $F_v, G_v,$ and B_v
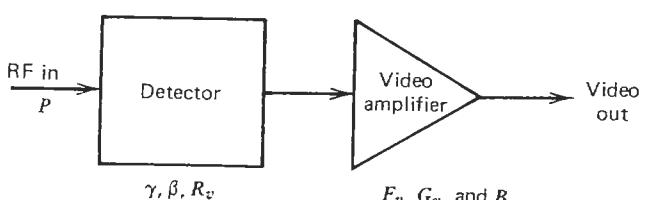
FIGURE 12.4.   Electrical parameters of a crystal video receiver

that the MDS corresponds roughly to an output signal-to-noise ratio of $SNR_o \approx 1$.

Nominal Detectable Signal (NDS)

In view of the uncertainties associated with the MDS, Uhlir [3] recommended that the MDS definition be refined and formalized by using a $SNR_o = 1$ (i.e., *exactly* unity) as the basis. He called it the nominal detectable signal (NDS). This step permitted an analytical derivation of NDS in terms of the various diode and circuit parameters. The derivation is readily available elsewhere [3, 5] and will not be repeated here. The final expression is

$$NDS = \frac{2}{\gamma}\sqrt{kTR_vB_v\left(t_w + \frac{T_0}{T}(F_v - 1) + \frac{f_N}{B_v}\ln\frac{f_h}{f_s}\right)} \tag{12-17}$$

where γ is the open-circuit voltage sensitivity; T the physical temperature of the diode; and B_v the video bandwidth equal to $f_h - f_s$, where f_h and f_s are the upper and lower band edge, respectively. f_N is the noise corner of the $1/f$ noise. This quantity varies widely with the type of diode and the applied bias. It could be anywhere from 0.1 to 2 kHz for hot-carrier diodes, to perhaps greater than 5 MHz for point-contact diodes [1]. Finally, t_w is the "white-noise" temperature ratio, which is approximately 0.8–0.9 for hot-carrier diodes.

If the flicker-noise corner is at a low frequency such that $f_N \ll B_v$, and if we can consider $t_w \approx 1$ and $T = T_0$, then (12-17) reduces to

$$NDS \approx \frac{2}{\gamma}\sqrt{kT_0R_vB_vF_v} \tag{12-18}$$

The important observation is that the NDS is proportional to the square root of the video-bandwidth and the video-amplifier noise factor.

The formal basis of NDS, $SNR_o = 1$, removes the element of arbitrariness that plagues the MDS. It is now possible to measure the sensitivity more accurately. All that is required is a true rms meter. With the RF signal turned off, the background noise level is noted at the output of the video receiver. The input signal level is then increased until the combined, signal plus noise, output increases 3 dB above the initial level. This signal level is the NDS.

Noise Equivalent Power (NEP)

The NDS is still dependent on the bandwidth B_v of the video amplifier or the measuring setup in general because the total background noise power is clearly determined by the bandwidth used in test. The definition of sensitivity was therefore further revised by Cowley and Sorensen [1] who retained the criterion $SNR_o = 1$ but based it on a bandwidth of 1 Hz. This definition is called the

noise equivalent power (NEP). Its main advantage is that it permits an analysis of sensitivity from the fine-grain viewpoint. It will be remembered that in addition to thermal noise (which is flat with frequency), there will be an increasing contribution from the $1/f$ noise at low video frequencies. It is therefore difficult to ascertain the outcome in a narrowband application if the sensitivity of the detector was originally specified on the basis of a large video bandwidth. The NEP, being a sensitivity measure per unit bandwidth, will take this variation into account. The analytical expression for the NEP, for the detector alone, is [1]:

$$ \text{NEP} = \frac{2\eta kT}{q} \left(\frac{4kTt_w}{R_b} \right)^{1/2} \left(1 + \frac{R_s}{R_b} \right)^{1/2} \left(1 + \frac{f}{f_c} \right)^{1/2} \left(1 + \frac{f_N}{f_v} \right)^{1/2} \quad (12\text{-}19) $$

where f_c is the cutoff frequency of the diode {Eq. (21) in ref. [1]}; f the RF input frequency; f_N the noise corner of the $1/f$ noise; and f_v the video output frequency. Other terms have been defined previously.

This expression, as derived by Cowley and Sorensen, also takes into account the effect of the junction capacitance of the diode.

Tangential Signal Sensitivity (TSS)

The fourth, quite popular measure of detector sensitivity is the tangential signal sensitivity (TSS). The TSS is another outgrowth of practical laboratory work as was the MDS. It is measured with a pulse-modulated RF input signal and is defined as that peak value of the RF input pulse (see Fig. 12.5) at which
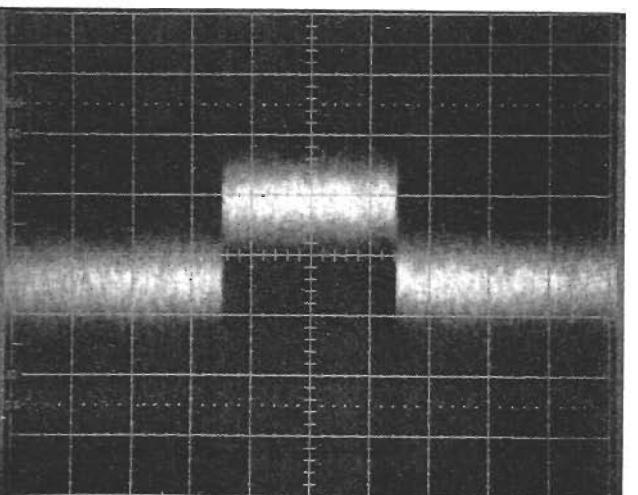
FIGURE 12.5. Illustration of tangential signal sensitivity (TSS) condition. (Courtesy of Microphase Corporation.)

the highest noise peaks observed on the oscilloscope in the *absence* of signal pulse are at the same level as the lowest noise peaks in the *presence* of the signal pulse [6]. The TSS is thus also a subjective measurement, but the results tend to be more consistent from observer to observer than those of the MDS. It has been determined that for the TSS condition, the signal-to-rms noise voltage ratio at the video output is close to 2.5:1. As a result, this number has been adopted as the "official" definition for the TSS. In logarithmic terms the ratio becomes

$$ \text{SNR}_o = 20 \log 2.5 = 8 \text{ dB} \quad (12\text{-}20) $$

Note also that since we are dealing with a square-law device, for which the output voltage is proportional to input power, the SNR, when transferred to the input, becomes

$$ \text{SNR}_i = 10 \log 2.5 = 4 \text{ dB} \quad (12\text{-}21) $$

Now recall that the NDS definition was based on a condition $\text{SNR}_i = \text{SNR}_o = 1$. From this it follows that

$$ \text{TSS (dBm)} = \text{NDS (dBm)} + 4 \text{ dB} \quad (12\text{-}22) $$

It further follows that the TSS is also proportional to the square root of the video bandwidth and the noise factor of the video amplifier:

$$ \text{TSS} \sim \sqrt{B_v F_v} \quad (12\text{-}23) $$

Thus the reference value of B_v and F_v must be given if a quoted value of the TSS is to be meaningful. The commonly used reference bandwidths are 1 and 2 MHz, but other values are used also. The typical value for F_v is 1.5 dB. In any case, a quantitative comparison of two detectors is easily possible through (12-23), for example, given a TSS $= -50$ dBm, based on a $B_{v1} = 2$ MHz and a $F_{v1} = 1.5$ dB. If the new reference conditions are $B_{v2} = 100$ kHz and $F_{v2} = 2$ dB, the TSS value becomes

$$ \text{TSS}_2 \text{ (dBm)} = \text{TSS}_1 \text{ (dBm)} + 5 \left(\log \frac{B_{v2}}{B_{v1}} + \log \frac{F_{v2}}{F_{v1}} \right) $$

$$ = -50 + 5(\log 0.05 + \log 1.33) = -55.9 \text{ dBm} $$

The TSS of microwave detectors depends on the type of diode used, the value of the applied dc bias, if any, and the operating frequency. For the common microwave region below 12 GHz, biased Schottky detectors (100–300 μA bias) range from -51 to -54 dBm ($B_v = 2$ MHz and $F_v = 1.5$ dB). Zero-bias Schottkys are in the same class. Back diode detectors (no bias) are

perhaps 2–3 dB less sensitive, and point-contact units (no bias) somewhat worse than biased Schottkys, largely because of inferior $1/f$ noise characteristics. It should be emphasized that this comparison is only for illustrating the numbers involved. There are other variables that must be considered. Thus a normally biased Schottky detector could quickly lose 20–30 dB in TSS if the bias were removed. Similarly, the point-contact detector can be improved by 2–3 dB if a modest bias of perhaps 20 μA is used. The system designer must also scrutinize and compare performance characteristics such as the RF and video impedances, power rating of the diode, dynamic range, temperature stability, and so on. A selected listing of source material on this subject appears in the general bibliography section at the end of the book.

It should also be pointed out that the bandwidth B_v is that of the entire video circuit, which is not necessarily equal to the bandwidth of the video amplifier if other components of narrower bandwidth are included in the chain.

The relationship between the TSS and the NEP is given by

$$\text{TSS (dBm)} = \text{NEP (dBm)} + 4 + 5\log B_v \qquad (12\text{-}24)$$

since the NEP was based on a unit bandwidth.

The upper limit f_h of the bandwidth B_v is usually much higher than f_s, the lower band edge. Thus f_s is often neglected and f_h is used as the bandwidth B_v. In actual practice B_v seldom extends to dc, for that would include flicker noise, especially if the f_N (corner frequency of the $1/f$ noise) for the particular diode is high (> 50 kHz). If a detector is intended for an application where f_s is low, then this fact must be kept in mind. Otherwise the TSS obtained with a test setup, with the low-frequency end safely cut off, would be noticeably better than that observed in the actual system where $1/f$ noise may come in strongly.

Measurement of the TSS

The usual method of measuring the TSS is shown in Fig. 12.6. The result will obviously be somewhat subjective, depending on one's perception of the "tangential" condition. It is, however, possible to measure [7] the TSS more accurately if a true rms voltmeter, such as the HP 3400A or HP 3403C, is used in place of the oscilloscope. First, recall from (12-20) that since the TSS represents a SNR$_o$ of 2.5 (8 dB),

$$\frac{V_p}{(e_n)_{\text{rms}}} = 2.5 \qquad (12\text{-}25)$$

where the signal is as defined in Fig. 12.7. Now note in Fig. 12.7 that the ac coupling of the voltmeter causes the input pulse waveform to shift such that the dc component is eliminated. If one defines a duty ratio $w = \tau/T$, then by
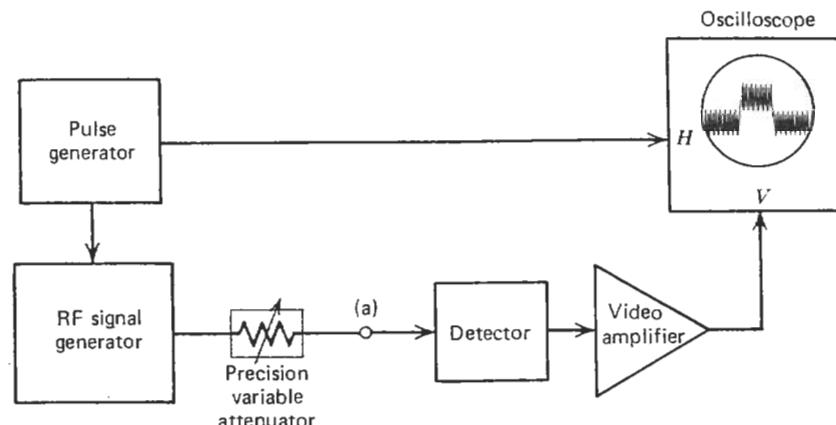
FIGURE 12.6.   Typical measurement setup for the tangential signal sensitivity.

the usual integration,

$$\overline{v_s^2} = V_p^2 w(1-w) \qquad (12\text{-}26)$$

The *total* rms voltage read by the meter is

$$\overline{v_T^2} = \overline{e_n^2} + \overline{v_s^2} = \overline{e_n^2} + V_p^2 w(1-w) \qquad (12\text{-}27)$$

By the TSS condition,

$$V_p = 2.5\sqrt{\overline{e_n^2}} \qquad (12\text{-}28)$$

and therefore

$$\overline{v_T^2} = \overline{e_n^2}[1 + 6.25w(1-w)] \qquad (12\text{-}29)$$

or

$$10\log\left(\frac{\overline{v_T^2}}{\overline{e_n^2}}\right) \doteq 10\log[1 + 6.25w(1-w)] \qquad (12\text{-}30)$$
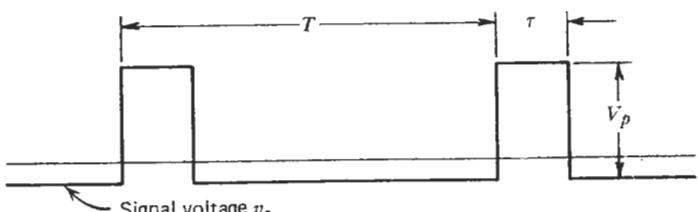
FIGURE 12.7.   Definition of parameters of pulsed RF waveform for TSS measurement.

For example, for a 10% duty ratio, $w = 0.1$, and the ratio of $\overline{v_T^2}/\overline{e_n^2}$ becomes 1.94 dB. Thus, to measure the TSS:

(a) Select and set the duty ratio accurately.
(b) Calibrate the input CW power at (a) in Fig. 12.6.
(c) Turn the signal generator off and note the noise level on the rms voltmeter.
(d) Turn the signal generator on and pulse it. Slowly increase the pulse amplitude until the voltmeter indication increases by the amount calculated from (12-30).
(e) The RF level at (a) is now equal to the TSS value sought.

For a square-wave modulation, $w = 0.5$ and (12-30) yields a value of 4.08 dB.

12.3. SUMMARY

(a) The concept of noise factor does not apply to detectors because their conversion loss is a function of the applied signal level:

$$\frac{L_{c1}}{L_{c2}} = \frac{P_2}{P_1}.$$

(b) Open-circuit voltage sensitivity:

$$\gamma = \frac{\delta v}{P_{in}}.$$

(c) Short-circuit current sensitivity:

$$\beta = \frac{\delta i}{P_{in}}.$$

(d) Relationship between γ and β is $R_v = \gamma\beta$, where R_v is the video resistance.
(e) Minimum detectable signal (MDS) is that input signal level at which the signal is just discernible in the noise (SNR$_o \approx 1$).
(f) Nominal detectable signal (NDS) is the input level at which the SNR$_o$ is exactly unity.
(g) Noise equivalent power (NEP) is defined the same as the NDS, except that it is based on a bandwidth of 1 Hz.
(h) Tangential signal sensitivity (TSS) is observed in terms of the noise level raised by the RF input pulse. The commonly accepted definition

is

$$SNR_o = 8 \text{ dB}$$

(i) The TSS is proportional to the square root of the video bandwidth and the noise factor of the video amplifier used in the test:

$$TSS \sim \sqrt{B_v F_v}$$

REFERENCES

1. A. M. Cowley and H. O. Sorensen, "Quantitative Comparison of Solid-State Microwave Detectors," *IEEE Trans. MTT*, vol. MTT-14, pp. 588–602, December 1966.
2. H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, McGraw-Hill, New York, 1948.
3. A. Uhlir, Jr., "Characterization of Crystal Diodes for Low-Level Microwave Detection," *Microwave J.*, July 1963, pp. 3–11.
4. A. van der Ziel, *Noise*, Prentice-Hall, Englewood Cliffs, NJ, 1954.
5. "The Video Detector, A Technical Discussion," monograph by Sage Laboratories, Natick, Mass., 1968.
6. Y. Anand and W. J. Moroney, "Microwave Mixer and Detector Diodes," *Proc. IEEE*, vol. 59, pp. 1182–1190, August 1971.
7. "The Criterion for the Tangential Sensitivity Measurement," Hewlett-Packard Application Note 956-1, October 1973.

APPENDIX
A | THE COMPLEMENTARY ERROR FUNCTION

APPENDIX
B | THE *ABCD* MATRIX

The complementary error function is defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} \, du$$

x	$\text{erfc}(x)$	x	$\text{erfc}(x)$
0.0	1.000	2.2	1.86×10^{-3}
0.2	0.777	2.4	6.9×10^{-4}
0.4	0.572	2.6	2.4×10^{-4}
0.6	0.396	2.8	7.9×10^{-5}
0.8	0.258	3.0	2.3×10^{-5}
1.0	0.157	3.2	6.0×10^{-6}
1.2	0.0897	3.4	1.5×10^{-6}
1.4	0.0487	3.6	3.6×10^{-7}
1.6	0.0237	3.8	7.7×10^{-8}
1.8	0.0109	4.0	1.5×10^{-8}
2.0	7.21×10^{-3}	5.0	1.5×10^{-12}

The *ABCD* matrix of a two-port network relates the input voltage and current, v_1 and i_1, respectively, to the output voltage and current, v_2 and i_2, respectively, in accordance with Fig. B.1 and Eqs. (B-1) and (B-2):

$$v_1 = Av_2 + Bi_2 \tag{B-1}$$

and

$$i_1 = Cv_2 + Di_2 \tag{B-2}$$

Note the direction of i_2 in this case as compared to that in Fig. 6.1.

The principal advantage of the *ABCD* matrix is that it can be cascaded very simply by the usual rules of matrix multiplication, and the result is the overall *ABCD* matrix of the cascaded package:

$$\begin{bmatrix} A_T & B_T \\ C_T & D_T \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \times \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \tag{B-3}$$

It is therefore possible to obtain the *ABCD* matrix for fairly complex networks by combining the matrices of the individual components. The *ABCD* matrices of common elementary networks are shown in Table B.1. Thus, for example, we have for an L-attenuator (consisting of a series and a shunt

resistor):

$$\begin{bmatrix} 1 & R_1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1/R_2 & 1 \end{bmatrix} = \begin{bmatrix} 1 + R_1/R_2 & R_1 \\ 1/R_2 & 1 \end{bmatrix}$$

Some general rules pertaining to $ABCD$ matrices are:

(a) For all $ABCD$ matrices: $AD - BC = 1$.
(b) For symmetrical networks: $A = D$.
(c) For lossless networks: A and D are real while B and C are imaginary.

The various circuit parameters in Chapter 7 can be expressed in terms of the $ABCD$ matrix. By dividing (B-1) by (B-2) we obtain the input impedance of the network:

$$Z_{\text{in}} = \frac{AR_L + B}{CR_L + D} \tag{B-4}$$

because $v_2 = i_2 R_L$. By inverting this,

$$Z_{\text{out}} = \frac{DR_g + B}{CR_g + A} \tag{B-5}$$

The $ABCD$ matrix can also be used for the gain definitions in Table 7.1. We note that the first three cases, G_p, G_i, and G_t, all involve $(v_2/V_g)^2$. Thus from Fig. B.1:

$$v_1 = Av_2 + Bi_2 + V_g\left(\frac{Z_{\text{in}}}{R_g + Z_{\text{in}}}\right) \tag{B-6}$$
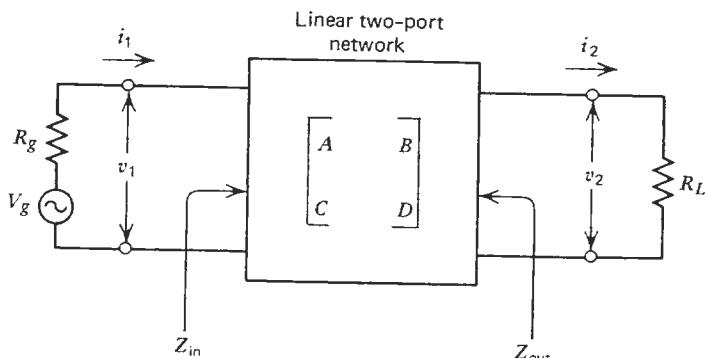
FIGURE B.1. $ABCD$ matrix of a linear two-port network.

and because of $i_2 = v_2/R_L$,

$$\frac{v_2}{V_g} = \frac{Z_{\text{in}}}{(A + B/R_L)(R_g + Z_{\text{in}})} \tag{B-7}$$

Using (B-4) yields

$$\frac{v_2}{V_g} = \frac{1}{A + B/R_L + CR_g + D(R_g/R_L)} = \frac{1}{(A + CR_g)(1 + Z_{\text{out}}/R_L)} \tag{B-8}$$

TABLE B.1. $ABCD$ Matrices of Common Networks

Network	Representation	$ABCD$ Matrix
Series impedance		$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$
Shunt admittance		$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
Lossless transmission line		$\begin{bmatrix} \cos\beta x & jZ_0\sin\beta x \\ (j\sin\beta x)/Z_0 & \cos\beta x \end{bmatrix}$
Ideal transformer		$\begin{bmatrix} n_1/n_2 & 0 \\ 0 & n_2/n_1 \end{bmatrix}$

Either form of (B-8) can be used in the expressions of Table 7.1.

The available gain G_a is developed somewhat differently. From (7-30),

$$G_a = \frac{|V_{\text{out}}|^2 R_g}{V_g^2 \text{Re}(Z_{\text{out}})} \tag{B-9}$$

Since the G_a is defined independently of the load, that is, on the basis of the open-circuit voltage of the generator, (B-1) becomes:

$$v_1 = Av_2 = AV_{\text{out}} \tag{B-10}$$

and (B-4) becomes

$$Z_{\text{in}} = \frac{A}{C} \tag{B-11}$$

Combining (B-6) and (B-9)–(B-11) yields

$$G_a = \frac{R_g}{\text{Re}\left[(R_g D + B)(R_g C + A)^*\right]} \tag{B-12}$$

As a check, consider the shunt resistor shown in Fig. 7.6. From Table B.1, $A = D = 1$, $B = 0$, and $C = 1/R$. From (B-12),

$$G_a = \frac{R_g}{R_g(R_g/R + 1)} = \frac{R}{R_g + R}$$

which is the same as (7-29).

Using (B-12) one can show that for lossless networks, G_a is always unity. We saw earlier that for this case, A and D are real, and $B = jB'$ and $C = jC'$ are imaginary.

Equation (B-12) now becomes (note the complex conjugate)

$$G_a = \frac{R_g}{\text{Re}\left[(R_g D + jB')(-jR_g C' + A)\right]}$$

$$= \frac{1}{AD + B'C'} = 1$$

APPENDIX C

THE IMAGELESS MIXER

The imageless or image rejection mixer discriminates between the signal and the image by diverting the IF outputs derived from each input to different output ports. Figure C.1 shows a typical layout of imageless mixer. Assume an upper sideband $\omega_1 = \omega_p + \omega_{\text{IF}}$ and a lower sideband $\omega_2 = \omega_p - \omega_{\text{IF}}$ signal present at $J1$. Assume further that the ω_1 is the desired signal and ω_2 is the undesirable image.

Thus for the *upper-sideband* signal ω_1 we have these frequency components at the referenced points:

At (1): $\cos(\omega_p + \omega_{\text{IF}})t$.

At (3): $\cos \omega_p t$.

At (5): $[\cos(\omega_p + \omega_{\text{IF}})t](\cos \omega_p t) \Rightarrow \cos \omega_{\text{IF}}t$ (after retaining only the IF term and ignoring all amplitude factors).

At (2): $\cos[(\omega_p + \omega_{\text{IF}})t - 90°] = \sin(\omega_p t + \omega_{\text{IF}})t$.

At (4): same as for (3).

At (6): The IF term becomes $\sin \omega_{\text{IF}}t$.

After passing through the quadrature output hybrid, we have at $J2$:

$$\underbrace{\sin \omega_{\text{IF}}t}_{\text{from (6)}} + \underbrace{\cos(\omega_{\text{IF}}t - 90°)}_{\text{from (5)}} = \sin \omega_{\text{IF}}t + \sin \omega_{\text{IF}}t$$

Thus the IF signals due to the upper sideband ω_1 combine at $J2$.

A similar analysis shows that at $R2$:

$$\underbrace{\cos\omega_{IF}t}_{\text{from (5)}} + \underbrace{\sin\left(\omega_{IF}t - 90°\right)}_{\text{from (6)}} = \sin\left(\omega_{IF}t + 90°\right) + \sin\left(\omega_{IF}t - 90°\right)$$

that is, the signals cancel. There is no upper-sideband output at $R2$.

For the *lower-sideband* signal ω_2 (the unwanted image) we obtain:

At (1): $\cos(\omega_p - \omega_{IF})t$.

At (2): $\cos[(\omega_p - \omega_{IF})t - 90°] = \sin(\omega_p - \omega_{IF})t$

Again, multiplying both signals by $\cos\omega_p t$, retaining only the IF terms, and combining the IF outputs at $J2$ and $R2$ yields:

At $J2$: $\underbrace{\sin\left(\omega_{IF}t - 180°\right)}_{\text{from (6)}} + \underbrace{\sin\omega_{IF}t}_{\text{from (5)}}$

Hence signals cancel at $J2$.

At $R2$: $\underbrace{\sin\left(\omega_{IF}t + 90°\right)}_{\text{from (5)}} + \underbrace{\sin\left(\omega_{IF}t - 270°\right)}_{\text{from (6)}}$

and the signals add.

Thus the desired IF output, generated by the upper-sideband signal ω_1, combines at $J2$ where it becomes available. The undesired IF from the lower-sideband image combines at $R2$ where it is absorbed.
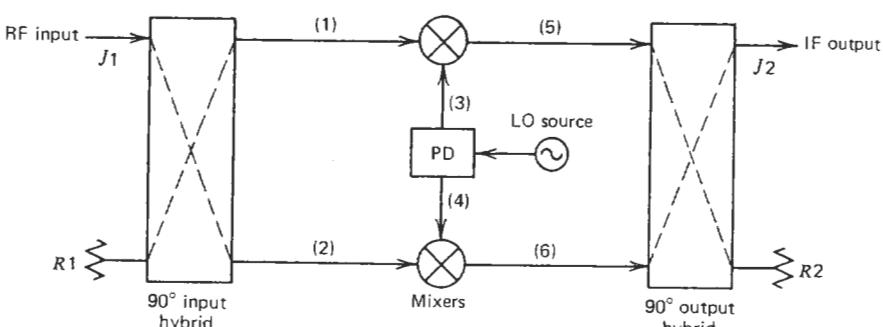
FIGURE C.I.   Layout of an imageless mixer.

GENERAL BIBLIOGRAPHY

Adamy, D. L., "Calculate Receiver Sensitivity," *Electron. Design*, pp. 118–121, December 6, 1973.

Ashley, J. R., T. A. Barley, and G. J. Rast, "The Measurement of Noise in Microwave Transmitters," *IEEE Trans. MTT*, vol. MTT-25, pp. 294–318, April 1977.

Barber, M. R., "Ultimate Noise Figure and Conversion Loss of the Schottky Barrier Mixer Diode," *IEEE Int. Microwave Symp. Digest*, pp. 13–17, May 1966.

Bedrosian, E., "On the Noise Temperature of Coupling Networks," *IRE Trans. MTT*, vol. MTT-8, p. 463, July 1960.

Bennett, W. R., "Methods of Solving Noise Problems," *Proc. IRE*, vol. 44, pp. 609–638, May 1956.

Blake, L. V., "Antenna and Receiving System Noise-Temperature Calculation," *Proc. IRE*, vol. 49, pp. 1568–1569, October 1961.

Brady, M. M., "The Influence of Mismatch Error in Noise Performance Measurements," *Proc. IEEE*, vol. 52, pp. 1075–1076, September 1964.

Cheadle, D. L., "Measure Mixer Noise with Your Power Meter," *Microwaves*, pp. 42–49, March 1975.

Cohn, S. B., "The Noise Figure Muddle," *Microwave J.*, pp. 7–12, March 1959.

Crowe, W. J., "My Adventures with Tunnel Diode Amplifiers," *Microwaves*, pp. 24–31, July 1965.

Dickens, L. E., "Low Conversion Loss Millimeter Wave Mixers," *1973 G-MTT Symp. Digest*, pp. 66–68, June 1973.

Dickens, L. E., and D. W. Maki, "An Integrated-Circuit Balanced Mixer, Image and Sum Enhanced," *IEEE Trans. MTT*, vol. MTT-23, pp. 276–281, March 1975.

Edden, F., "What's All This Noise in Satellite Communications Systems?" *Microwave Syst. News*, pp. 43–48A, June/July 1976.

Edrich, J., "Low Noise and Permanent Josephson Mixers for MM-Waves," *1976 G-MTT Symp. Digest*, June 1976.

Eng, S. T., "Low-Noise Properties of Microwave Backward Diodes," *IRE Trans. MTT*, vol. MTT-9, pp. 419–425, September 1961.

Freeman, R. L., *Telecommunication Transmission Handbook*, Wiley, New York, 1981.

Fukui, H., "Available Power Gain, Noise Figure, and Noise Measure of Two-Port and Their Graphical Representation," *IEEE Trans. Circuit Theory*, vol. CT-15, pp. 137–142, June 1966.

Gabriel, W. F., "Tunnel Diode Low Level Detection," *IEEE Trans. MTT*, vol. MTT-15, pp. 538–553, October 1967.

Grimm, H. H., "Noise Temperature in Passive Circuits," *Microwave J.*, pp. 52–54, February 1960.

Gupta, M. S., "Applications of Electrical Noise," *Proc. IEEE*, vol. 63, pp. 996–1010, July 1975.

Haus, H. A., and R. B. Adler, *Circuit Theory of Linear Noisy Networks*, Wiley, New York, 1959.

Hewlett-Packard Application Notes Nos. 57 ("Noise Figure Primer") 1962, 907 ("The Hot Carrier Diode, Theory, Design and Application") and 1967, 923 ("Hot Carrier Diode Video Detectors").

Hooge, G. N., "Discussion of Recent Experiments on $1/f$ Noise," *Physica*, vol. 60, pp. 130–144, July 1972.

"Hot Carrier Diodes as Detectors and Mixers," *Hewlett-Packard J.*, vol. 17, December 1965.

Howson, D. P., and R. B. Smith, *Parametric Amplifiers*, McGraw-Hill, New York, 1970.

IEEE Trans. MTT, Special Issue on Noise, vol. MTT-16, September 1968.

Kerr, A. R., "Low-Noise Room-Temperature and Cryogenic Mixers for 80–120 GHz," *IEEE Trans. MTT*, vol. MTT-23, pp. 781–787, October 1975.

Lance, A., W. D. Seal, and P. M. Fujimoto, "A New Way to Measure Noise," *Microwaves*, pp. 32–41, March 1975.

Lane, R. Q., "The Determination of Device Noise Parameters," *Proc. IEEE*, vol. 57, pp. 1461–1462, August 1969.

Larock, K. D., and R. P. Meys, "Automatic Noise Temperature Measurement Through Frequency Variation," *IEEE Trans. MTT*, vol. MTT-30, pp. 1286–1288, August 1982.

Lawson, J. L., and G. E. Uhlenbeck, *Threshold Signals*, McGraw-Hill, New York, 1950, Chapter 13.

Martines, G., and M. Sannino, "Determination of Microwave Transistor Noise and Gain Parameters Through Noise Figure Measurements Only," *IEEE Trans. MTT*, vol. MTT-30, pp. 1255–1259, August 1982.

Messenger, G. C., and C. T. McCoy, "Theory and Operation of Crystal Diodes as Mixers," *Proc. IRE*, vol. 45, pp. 1269–1283, September 1957.

Meys, R. P., "A Wave Approach to the Noise Properties of Linear Microwave Devices," *IEEE Trans. MTT*, vol. MTT-26, pp. 34–37, January 1978.

Mouw, R. B., and F. M. Schumacher, "Tunnel Diode Detectors," *Microwave J.*, pp. 27–36, January 1966.

Mukaihata, T., B. L. Walsh, M. F. Bottjer, and E. B. Roberts, "Subtle Differences in System Noise Measurements and Calibration of Noise Standards," *IEEE Trans. MTT*, vol. MTT-10, pp. 506–516, November 1962.

Netzer, Y., "The Design of Low-Noise Amplifiers," *Proc. IEEE*, vol. 69, pp. 728–741, June 1981.

Neuf, D., "Mixer and Preselector System Noise Figure," *Microwave J.*, pp. 45–58, June 1972.

Okean, H. C., and A. J. Kelly, "Low-Noise Receiver Design Trends Using State-of-the-Art Building Blocks," *IEEE Trans. MTT*, vol. MTT-25, pp. 254–267, April 1977.

Okean, H. C., and P. P. Lombardo, "Noise Performance of M/W and MM-Wave Receivers," *Microwave J.*, pp. 41–50, January 1973.

Pastori, W. E., "Image and Second-Stage Corrections Resolve Noise Figure Measurement Confusion," *Microwave Syst. News*, pp. 67–86, May 1983.

Pritchard, W. L., "Notes on a Crystal Mixer Performance," *IRE Trans. MTT*, vol. MTT-3, pp. 37–39, January 1955.

Ragazzini, J. R., and S. S. L. Chang, "Noise and Random Processes," *Proc. IRE*, vol. 50, pp. 1146–1151, May 1962.

Rice, S. O., "Mathematical Analysis of Random Noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, July 1944; vol. 24, pp. 46–156, January 1945.

Siegal, B., and C. Lai, "Multidiode Approach Offers Improved Microwave Detector Performance," *Microwave Syst. News*, pp. 101–110, August/September 1975.

Siegal, B., and E. Pendleton, "Zero-Bias Schottky Diodes as Microwave Detectors," *Microwave J.*, pp. 40–43, September 1975."

"Standards on Electron Devices: Methods of Measuring Noise," *Proc. IRE*, vol. 41, pp. 890–896, July 1953.

Stelzried, C. T., "Operating Noise-Temperature Calibrations of Low-Noise Receiving Systems," *Microwave J.*, pp. 41–47, June 1971.

Stelzried, C. T., "Post-Amplifier Noise Temperature Contribution in a Low-Noise Receiving System," *Proc. IEEE*, vol. 52, pp. 76–77, January 1964.

Stracca, G. B., "On Frequency Converters Using Non-Linear Resistors," *Alta Frequenza*, vol. 38, pp. 318–331, May 1969.

Strum, P. D., "A Note on Noise Temperature," *IRE Trans. MTT*, vol. MTT-4, pp. 145–151, July 1956.

Sucher, M., and J. Fox (Eds.), *Handbook of Microwave Measurements*, vol. III, pp. 865–887, Polytechnic Press, Polytechnic Institute of Brooklyn, New York, 1963.

Thompson, F. H., "A Minimum Detectable Signal Nomograph," *Microwave J.*, pp. 52–56, March 1967.

Uhlir, A., Jr., "Shot Noise in P-N Junction Frequency Converters," *Bell Syst. Tech. J.*, vol. 37, pp. 951–987, July 1958.

Vernon, F. L., M. F. Millea, M. F. Bottjer, A. H. Silver, R. J. Pedersen, and M. McColl, "The Super-Schottky Diode," *IEEE Trans. MTT*, vol. MTT-25, pp. 286–294, April 1977.

Vigneri, R., G. G. Gulbenkian, and N. Diepeveen, "A Graphical Method for the Determination of Equivalent Noise Bandwidth," *Microwave J.*, pp. 49–52, June 1968.

Weinreb, S., and A. R. Kerr, "Cryogenic Cooling of Mixers for Millimeter and Centimeter Wavelengths," *IEEE J. Solid-State Circuits*, vol. SC-8, pp. 58–63, February 1973.

Whelehan, J. J., "Low-Noise Millimeter-Wave Receivers," *IEEE Trans. MTT*, vol. MTT-25, pp. 268–280, April 1977.

van der Ziel, A., "Noise in Solid-State Devices and Lasers," *Proc. IEEE*, vol. 58, pp. 1178–1206, August 1970.

INDEX