NOISE IN MEASUREMENTS

ALDERT VAN DER ZIEL

Electrical Engineering Department
University of Minnesota

Electrical Engineering Department
University of Florida

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, New York • London • Toronto • Sydney

Library of Congress Cataloging in Publication Data:

Van der Ziel, Aldert, 1910–
Noise in measurements.

"A Wiley-Interscience publications."
Includes bibliographical references.
1. Electric measurements. 2. Electronic measurements. 3. Random noise theory. 4. Electric noise. 5. Electronic noise. I. Title.
QC535.V35 621.37 76-12108
ISBN 0-471-89895-3

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

NOISE IN MEASUREMENTS

PREFACE

This book discusses the effect of noise on the accuracy of measurements. It brings together material that was spread throughout many textbooks and a vast amount of past and current literature. The text was developed from a series of lectures given at the Universities of Minnesota and Florida.

Chapters 1–7 provide the necessary background. After a short introduction, the method of distribution functions for calculating averages, autocorrelation, and cross-correlation functions is developed in Chapter 2. Chapter 3 considers a few simple applications, and Chapter 4 examines binomial, Poisson, and normal distribution functions and develops the variance theorem. Chapter 5 discusses Fourier analysis methods and shows how spectral intensities can be calculated. Chapter 6 takes up noise characterization in two-terminal and four-terminal devices, and Chapter 7 examines flicker noise and generation-recombination noise.

Chapters 8–17 deal with applications. Chapter 8 treats measurements of small currents, voltages, and charges; Chapter 9 studies thermal radiation detectors like thermocouples and bolometers. Chapter 10 investigates photodetectors of the photoemissive, photodiode, and the classical detector types. Chapter 11 deals with photoconductive detectors, and Chapter 12 considers pyroelectric detectors and capacitive bolometers, then Chapter 13 examines noise in television pick-up tubes. Chapter 14 investigates photomixing, after which Chapter 15 deals with light amplification with electroluminescence. Chapter 16 gives a discussion of Josephson junction devices. Chapter 17 briefly examines high-energy quantum and particle detectors. The appendix derives a few formulas of the theory of ferroelectrics used in Chapter 12.

I am indebted to my graduate students at the Universities of Minnesota and Florida who helped shape the manuscript and to Mrs. van der Ziel who helped prepare it.

A. van der Ziel

Minneapolis, Minnesota
April 1976

CONTENTS

1

INTRODUCTION

In physics and electrical engineering one often encounters fluctuating signals generated in electrical circuits, electrical devices, or other measuring systems. Such fluctuating signals are generally called *noise*.

The name "noise" requires explanation. If the fluctuating voltage or current generated in a circuit component or electronic device is amplified by a low-frequency amplifier and the amplified signal is fed into a loudspeaker, the loudspeaker produces a hissing sound; hence the name "noise." The name "noise" now refers to any spontaneous fluctuation, independent of whether an audible sound is produced.

Noise sets a lower limit to the signals that can be processed electronically. In the same way it sets lower limits to practically all types of measurement. It is important to minimize the noise-to-signal ratio in any such measurement and so determine the limit of accuracy of these measurements. It is the aim of this book to familiarize the reader with these problems in the measurement of currents, voltages, charges, and amounts of radiation.

The important sources of noise that will be encountered are thermal noise, shot noise, generation–recombination noise, temperature-fluctuation noise, and flicker noise. We now discuss these noise sources in somewhat greater detail.

Thermal noise is due to the random motion of carriers in any conductor; as a consequence of this random motion a fluctuating electromotive force (e.m.f.) $V(t)$ is developed across the terminals of the conductor. The same phenomenon occurs in the conducting channel of field-effect transistors (FET). It is the dominant noise source in any device that is electrical in nature and in thermal equilibrium with a temperature bath kept at a fixed temperature T.

Shot noise occurs whenever a noise phenomenon can be considered as a series of independent events occurring at random. For example, in the case of emission of electrons by a thermionic cathode or by a photocathode, the emission of electrons consists of a series of independent random events; hence the emission currents show shot noise. In p–n junctions and transistors the crossing of a junction by electrical carriers (electrons or holes) constitutes a series of independent random events, hence the currents in such devices show shot noise. It equally holds when transitions occur between two energy levels, such as in the generation and recombination of carriers in a semiconductor, or when photons are emitted by a laser. In each case one must ask what entities make up the series of independent random events that produce shot noise.

Generation–recombination noise occurs whenever free carriers are generated or recombine in a semiconductor material. The fluctuating rates of generation and recombination can be considered as a series of independent events occurring at random, and hence the process can be considered as a shot-noise process. However, it is also useful to consider the fluctuation δn in the carrier density n as giving rise to a fluctuation δR in the resistance R of the device. This resistance fluctuation δR can be detected by passing a d.c. current I through the sample; the current I develops a fluctuating e.m.f. $V(t)=I\,\delta R(t)$ across its terminals, and this e.m.f. can be amplified and measured by standard techniques.

Temperature-fluctuation noise of a small body occurs because of the fluctuating heat exchange between the body and its environment due to fluctuations in the emitted and received radiation and to fluctuations in the heat conduction. The first can be described by fluctuations in the rate of emission and absorption of quanta by the small body. The fluctuations in the heat conduction are always present, since the small body must always have some heat-conducting path (wires, connections, etc.) to its environment. When air is blown over the small body or liquid is flowing past the small body, there is also a fluctuating *heat convection*; it is not essential, however, since it can be eliminated by proper techniques.

Flicker noise can be due to various causes and is characterized by its spectral intensity (see Chapter 5). Most noise sources have a spectral intensity that is constant at low frequencies and decreases more or less rapidly above a certain "turnover" frequency that is characteristic for the noise source in question. The various forms of flicker noise have in common the condition that their spectral intensity is of the form const$/f^{\alpha}$ with α close to unity, so that their effect is most pronounced at low frequencies.

Fluctuating quantities like currents, voltages, temperatures, or numbers of carriers are called *random variables*. One speaks of a *continuous* random variable when the fluctuating quantity can assume a continuous range of values and of a *discrete* variable when the fluctuating quantity can only assume discrete values. The fluctuating number of carriers in a semiconductor sample is a discrete random variable.

2

DISTRIBUTION FUNCTIONS, AVERAGES, AUTOCORRELATION, AND CROSSCORRELATION FUNCTIONS

In calculations about noise in electrical measuring systems one must often calculate the averages of a function $g(X)$ of the random variable $X(t)$ in question. It is denoted by $\overline{g(X)}$ and is calculated with the help of the *probability density function* or distribution function of the variable $X(t)$.

This function, in turn, is introduced by considering probabilities in an *ensemble*, namely a very large assembly of systems subjected to independent fluctuations. To make the discussion more precise, the number of systems should go to infinity.* We discuss this in Section 2.1a for a single random variable and in Section 2.1b for multiple random variables.

In the case of two random variables $X(t)$ and $Y(t)$ with $\overline{X}=\overline{Y}=0$, the average $\overline{XY}$ may not be zero. The quantities $X(t)$ and $Y(t)$ are then said to be *correlated*. A particular case of correlation occurs if we consider a random variable $X(u)$ at the instants t and $(t+s)$, the function $\overline{X(t)X(t+s)}$ is called the autocorrelation *function*. Extension to several random variables leads to *autocorrelation* and *crosscorrelation functions*. (Section 2.2).

2.1 DISTRIBUTION FUNCTIONS AND AVERAGES

2.1a Single Random Variable

We consider an ensemble of N systems in which the fluctuations are described by the random variable $X(t)$ and let N go to infinity. Let ΔN elements of the ensemble have a value of $X(t)$ between X and $(X+\Delta X)$ at

*In an ensemble with N elements the relative accuracy of the averages is $N^{-1/2}$, so that it corresponds to 0.01 for $N=10^4$.

the instant t_1. One then calls $\Delta P = (\Delta N/N)$ the probability that the random variable $X(t)$ had a value between X and $(X+\Delta X)$ at the instant t_1. Obviously ΔN is proportional to ΔX as long as ΔX is sufficiently small, so that $(\Delta P/\Delta X)$ is independent of ΔX. More precisely, we may write in differential form

$$\frac{dP}{dX} = f(X,t_1), \quad \text{or} \quad dP = f(X,t_1)\,dX \tag{2.1}$$

The function $f(X,t_1)$ is called the *probability density function* of X at the instant t_1. When $f(X,t_1+t)$ is independent of t, that is,

$$f(X,t_1+t) = f(X,t_1) = f(X) \tag{2.1a}$$

the variable is said to be *stationary*. The noise processes encountered in physics and engineering are nearly always stationary.

Since the variable X must certainly lie within the range of allowed values, we have, if the integration is extended over all allowed values of X,

$$\int f(X)\,dX = 1 \tag{2.2}$$

Such a function $f(X)$ is said to *normalized*. If $f(X)$ is not normalized, it can be multiplied by a normalizing factor C so that $Cf(X)$ is normalized, that is,

$$\int Cf(X)\,dX = 1 \quad \text{or} \quad C = \left[\int f(X)\,dX\right]^{-1} \tag{2.2a}$$

We may thus assume without lack of generality that $f(X)$ is normalized.

We can now define *ensemble averages* as follows: The ensemble average of X^m, denoted by $\overline{X^m}$, is defined as

$$\overline{X^m} = \int X^m f(X)\,dX \tag{2.3}$$

and the average of a function $g(X)$ of X is defined as

$$\overline{g(X)} = \int g(X) f(X)\,dX \tag{2.3a}$$

where the integration is extended over all values of X. If $f(X)$ is symmetrical in X, that is, if $f(X) = f(-X)$, and X can vary between $-X_0$ and X_0, then the averages of all odd powers of X are zero.

The most important averages are $\overline{X}$ and $\overline{X^2}$. If $\overline{X}$ is not zero, one should introduce $\Delta X = (X - \overline{X})$ as a new random variable. The most important average is then $\overline{\Delta X^2}$, which is denoted by the symbols $\operatorname{var} X$ or σ_x^2,

$$\operatorname{var} X = \sigma_x^2 = \overline{(X-\overline{X})^2} = \overline{X^2} - \overline{2X\overline{X}} + (\overline{X})^2 = \overline{X^2} - (\overline{X})^2 \tag{2.4}$$

If we look at a single element of the ensemble for the time interval $0 \leqslant t \leqslant T$, we can make up the time average $\langle g(X)\rangle$ of a function $g(X)$ of X by the definition

$$\langle g(X)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T g(X)\,dt \tag{2.5}$$

If this time average approaches the ensemble average (2.3a) in the limit when T goes to infinity, the noise processes under investigation are said to be *ergodic*. The noise processes encountered in physics and engineering are practically always ergodic.

In the case of discrete random variables the definitions must be properly modified and all integrations must be replaced by summations. Let $P(n)$ be the probability that the discrete variable has the value n, then the normalization condition becomes

$$\sum_n P(n) = 1 \tag{2.6}$$

and the ensemble average of n^m must be defined as

$$\overline{n^m} = \sum_n n^m P(n) \tag{2.7}$$

$(m = 1, 2\ldots)$. The variance of n is again defined as

$$\operatorname{var} n = \overline{(n-\overline{n})^2} = \overline{n^2} - (\overline{n})^2 \tag{2.8}$$

2.1b Multivariate Distributions and Averages

For two continuous variables $X_1(t)$ and $X_2(t)$ one can evaluate the probability that $X_1(t)$ has a value between X_1 and $(X_1 + dX_1)$ and that simultaneously $X_2(t)$ has a value between X_2 and $(X_2 + dX_2)$ at the instant t_1. In analogy with (2.1) the *joint probability* dP may be then written

$$dP = f(X_1, X_2, t_1)\,dX_1\,dX_2 \tag{2.9}$$

and $f(X_1, X_2, t_1)$ is called the *joint probability density function* for the variables X_1 and X_2 at the instant t_1. Usually

$$f(X_1,X_2,t_1+t)=f(X_1,X_2,t_1)=f(X_1,X_2) \tag{2.9a}$$

for all values of t; the noise process is then said to be *stationary*.

The normalization condition is now

$$\int\int f(X_1,X_2)\,dX_1\,dX_2=1 \tag{2.10}$$

and averages are defined in the same manner as for single variables, that is,

$$\overline{X_1^n X_2^m}=\int\int X_1^n X_2^m f(X_1,X_2)\,dX_1\,dX_2 \tag{2.11}$$

where the integration extends over all values of X_1 and X_2.

Usually $\overline{X_1}=\overline{X_2}=0$; the most important averages are then $\overline{X_1^2}$, $\overline{X_2^2}$, and $\overline{X_1X_2}$. If $\overline{X_1X_2}=0$, the quantities are said to be *uncorrelated*; if $\overline{X_1X_2}\neq 0$, the quantities are said to correlated; the parameter

$$c=\frac{\overline{X_1X_2}}{\left(\overline{X_1^2}\cdot\overline{X_2^2}\right)^{1/2}} \tag{2.12}$$

is called the *correlation coefficient*. Applying the fact that $\overline{(aX_1+bX_2)^2}\geqslant 0$ for all values of a and b, it can be shown that $-1\leqslant c\leqslant 1$. The case $|c|=1$ is called *full* correlation; the case $|c|<1$ is called *partial* correlation.

The considerations are easily extended to m variables $X_1,X_2\ldots X_m$ or to discrete variables $n_1,n_2\ldots n_m$. The definitions are similar to the two-variable case.

If two random variables $X(t)$ and $Y(t)$ are partly correlated, namely if $|c|<1$, then one can split Y into a part aX that is fully correlated with X and a part Z that is uncorrelated with X. That is, we may write

$$Y=aX+Z \tag{2.13}$$

where $\overline{X}=\overline{Y}=\overline{Z}=0$ and $\overline{XZ}=0$. Since $\overline{XY}=a\overline{X^2}$ and $\overline{Y^2}=(a^2\overline{X^2}+\overline{Z^2})$, we have from the definition of c

$$a=c\left[\frac{\overline{Y^2}}{\overline{X^2}}\right]^{1/2} \qquad \overline{Z^2}=\overline{Y^2}(1-c^2) \tag{2.13a}$$

These formulas are useful in the discussion of noise in bipolar transistors and FETs.

2.2 AUTOCORRELATION AND CROSSCORRELATION FUNCTIONS

A particularly useful case of two partly correlated variables occurs when $X_1(t) = X(t)$ and $X_2(t) = X(t+s)$. Then the joint probability density function $f(X_1, X_2)$ can be introduced and averages can be defined in the usual manner. The average $\overline{X_1X_2} = \overline{X(t)X(t+s)}$ is called the *autocorrelation function*; it measures how long a given fluctuation persists at later times.

The autocorrelation function has the following properties

1. $\overline{X(t)X(t+s)}$ is independent of t if $X(t)$ is stationary.
2. $\overline{X(t)X(t+s)}$ is either continuous or a δ function in s. If $\overline{X(t)X(t+s)}$ is not a δ function in s, then any discontinuities in $X(t)$ and $X(t+s)$ occur at different instants for different elements of the ensemble so that they are averaged out in the averaging process. As a consequence $\overline{X(t)X(t+s)} = \overline{X^2(t)}$ for $s=0$, unless the autocorrelation function is a δ function in s.
3. $\overline{X(t)X(t+s)}$ is symmetrical in s, if $X(t)$ is stationary. The reason is as follows:

$$\overline{X(t)X(t+s)} = \overline{X(u-s)X(u)} = \overline{X(u)X(u-s)} = \overline{X(t)X(t-s)}$$

The first step comes about by putting $u = (t+s)$. The second step is an interchange of terms. The third step involves replacing u by t, which is allowed since $X(t)$ is stationary.

4. For $s \to \infty$, $\overline{X(t)X(t+s)}$ goes to zero sufficiently fast, so that

$$\int_{-\infty}^{\infty} |\overline{X(t)X(t+s)}|\, ds \tag{2.14}$$

exists. This is the case for all practical noise sources, except perhaps flicker noise.

5. The correlation coefficient

$$c(s) = \frac{\overline{X_1X_2}}{\left[\overline{X_1^2} \cdot \overline{X_2^2}\right]^{1/2}} = \frac{\overline{X(t)X(t+s)}}{\overline{X^2}} \tag{2.15}$$

is called the *normalized* autocorrelation function; it exists if $\overline{X(t)X(t+s)}$ is not a δ function in s. Here we have made use of the fact that $X(t)$ is stationary, so that $\overline{X_1^2} = \overline{X_2^2} = \overline{X^2}$.

In the particular case of two partially correlated quantities $X(t)$ and $Y(t)$ one can introduce the *autocorrelation* functions $\overline{X(t)X(t+s)}$ and $\overline{Y(t)Y(t+s)}$ and the *crosscorrelation* functions $\overline{X(t)Y(t+s)}$ and

$\overline{X(t+s)Y(t)}$. The first are symmetrical in s, whereas the latter are usually not. Moreover, the crosscorrelation functions, although related, are not identical. We have in analogy with the case of a single variable

$$\overline{X(t)Y(t+s)} = \overline{X(u-s)Y(u)} = \overline{X(t-s)Y(t)} \tag{2.16}$$

$$\overline{X(t+s)Y(t)} = \overline{X(u)Y(u-s)} = \overline{X(t)Y(t-s)} \tag{2.16a}$$

In each case the first step replaces $(t+s)$ by u and the second replaces u by t. The latter is allowed if the noise processes are stationary.