

by

NORBERT WIENER

Professor of Mathematics

Massachusetts Institute of Technology

Nonlinear Problems in Random Theory

Published jointly by

The Technology Press of

The Massachusetts Institute of Technology

and

John Wiley & Sons, Inc., New York

Chapman & Hall, Limited, London

Copyright © 1958
by
The Massachusetts Institute of Technology

All Rights Reserved
This book or any part thereof must not
be reproduced in any form without the
written permission of the publisher.

Library of Congress Catalog Card Number: 58-59720
Printed in the United States of America

Foreword

There has long been a need in science and engineering for systematic publication of research studies larger in scope than a journal article but less ambitious than a finished book. Much valuable work of this kind is now published only in a semiprivate way, perhaps as a laboratory report, and so may not find its proper place in the literature of the field. Professor Wiener's contribution is the first of the Technology Press Research Monographs, which we hope will make selected timely and important research studies readily accessible to libraries and to the independent worker.

J. A. STRATTON

Preface

For some time I have been interested in a group of phenomena depending upon random processes. On the one hand, I have recorded the random shot effect as a suitable input for testing nonlinear circuits. On the other hand, for some of the work that Professor W. A. Rosenblith and I have been doing concerning the nature of the electroencephalogram, and in particular of the alpha rhythm, it has occurred to me to use the model of a system of random nonlinear oscillators excited by a random input. For many years I have been discussing my ideas with Professor Y. W. Lee and Professor A. G. Bose of the Electrical Engineering Department and of the Research Laboratory of Electronics of the Massachusetts Institute of Technology. About the beginning of 1958 they suggested to me that it might be useful to give a few lectures on the subject in a seminar for a chosen group of graduate students in electrical engineering. They offered me every possible help in the technique of recording these lectures and of working them up in a semipermanent or permanent form. They made use of tape recording of my spoken words and of the photography of formulae that I wrote down on the blackboard. From these they prepared a set of hectographed notes.

At the beginning we had contemplated a sequence of only four or five lectures. My ideas developed *pari passu* with the course, and by the end of the term we found ourselves with a set of fifteen lectures. The last few of these were devoted to the application of my ideas to problems in the statistical mechanics of gases. This work is both new and tentative, and I found that I had to supplement my course by the writing over of these lectures with the help of Professor Y. W. Lee.

I here wish to express my deep gratitude to Professors Lee and Bose, and to the members of the course whose names I give here: D. A. Chesler, D. A. George, I. M. Jacobs, A. H. Nuttall, C. E. Wernlein, Jr. (members of the Research Laboratory of Electronics and graduate students in the Department of Electrical Engineering, M.I.T.). Without their active help and sympathy and without the relief and freedom that they gave me by assuming the hard work of writing up the notes, this book would either never have come to exist or at best would have waited one or more years. They also freely advised me on the type of content that would be interesting to electrical engineers and to workers in related fields. I also wish to express my appreciation to Mrs. Ruth E. Rosensweig for undertaking the difficult task of typing the lectures from tape recording, and for typing the manuscript that was prepared by members of the course.

This book is written in an informal style, very largely as I spoke the material in the lectures. This has both advantages and disadvantages, but I hope that the informality will be pardoned in view of the life thrown into the material by presenting it just as it was given. I hope that my results will be of interest to workers in several fields. Among these are: electrical engineers working in communication theory; students of non-linear electrical networks; those interested in the stability of electrical generating systems; biophysicists, and, in particular, those concerned with all types of rhythmical phenomena; atomic physicists; and students of all branches of statistical mechanics. Among the last I have put students of hydrodynamics as well as meteorologists and oceanographers. My book represents a transcription of the present state of a growing subject, and it is my sincere hope that it will soon be superseded by the future work provoked by it.

Massachusetts Institute of Technology
Cambridge, Massachusetts
June 2, 1958

Contents

FOREWORD	v
PREFACE	vii
LECTURE 1 The Random Function of Time and Phase	1
LECTURE 2 Homogeneous Polynomial Functionals and Their Averages	16
LECTURE 3 Orthogonal Functions	28
LECTURE 4 Orthogonal Functions and Autocorrelation Functions	39
LECTURE 5 Application to Frequency-Modulation Problems—I	49
LECTURE 6 Application to Frequency-Modulation Problems—II	56
LECTURE 7 Application to Frequency-Modulation Problems—III	63
LECTURE 8 Application to the Study of Brain Waves, Random Time, and Coupled Oscillators	67
LECTURE 9 Some Thoughts on Quantum Theory	78
LECTURE 10 Nonlinear Systems—I	88
LECTURE 11 Nonlinear Systems—II	97
LECTURE 12 Coding	101
LECTURE 13 Decoding	110
LECTURE 14 A New Approach to Statistical Mechanics—I	118
LECTURE 15 A New Approach to Statistical Mechanics—II	123
INDEX	129

LECTURE 1

The Random Function of Time and Phase

I wish to discuss methods of handling a random process and, particularly, methods of handling the spectrum of the process, although these methods have a much more general application. In order to do this, I wish to bring in a discussion of the Brownian motion. By this I mean a Brownian motion in one dimension. Furthermore, for this hour perhaps, I shall discuss the Brownian motion over an interval of time 0 to 1 (Figure 1.1).

Let us consider the wandering of a particle. As time goes on, I want this particle to wander in a random way, so that the amount that it has

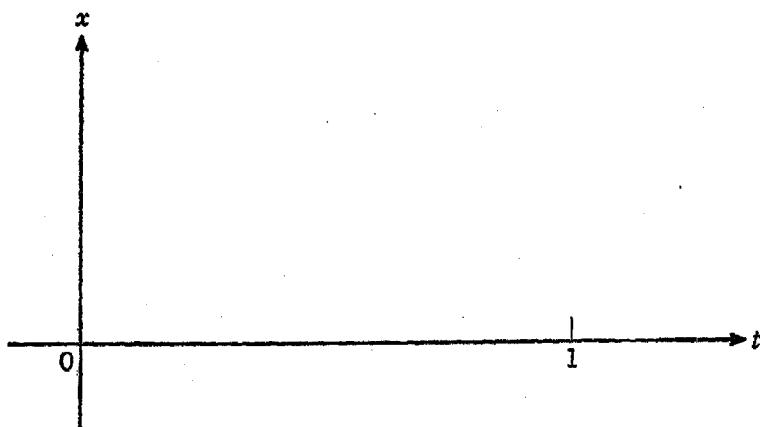


Figure 1.1

departed from a given position at a given time has a Gaussian distribution at another time, and so that in nonoverlapping intervals of time these Gaussian distributions are independent. I shall first discuss these motions at fixed times, and then, on the basis of the discussion of this motion at fixed times, I shall discuss the distribution of curves of this motion—first, over a finite time, and then over an infinite time. Then I am going to consider processes that depend upon this process in linear and in nonlinear ways.

A formula that we want is the formula of composition of Gaussian distributions that are independent over independent intervals of time. Consider a quantity x that has a Gaussian distribution. The probability that this quantity is between x_1 and x_2 is given by

$$P = \int_{x_1}^{x_2} \frac{1}{(2\pi a)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2a}\right) dx \quad (1.1)$$

I have not yet said how a depends upon time, but in order to get a reasonable way of doing this, I shall first discuss the composition of two such motions. That is, I want a particle that at the end of a certain time has a Gaussian distribution. Starting from where it is, the new departure has a Gaussian distribution. What is the distribution at the end of the two times?

So, I start with the joint distribution function

$$\left[\frac{1}{(2\pi a)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2a}\right) \right] \left[\frac{1}{(2\pi b)^{\frac{1}{2}}} \exp\left(-\frac{(y-x)^2}{2b}\right) \right] \quad (1.2)$$

Note that it is $y - x$ rather than y which is the parameter of the second probability density in Expression 1.2. Now, I want the probability distribution for y when x goes to all possible values. What I am interested in is

$$dy \int_{-\infty}^{\infty} \frac{1}{(2\pi a)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2a}\right) \frac{1}{(2\pi b)^{\frac{1}{2}}} \exp\left(-\frac{(y-x)^2}{2b}\right) dx \quad (1.3)$$

It is easy to compute the integral of Expression 1.3. It will not be necessary for me to go through it here. It's quite trivial. The answer is given by Equation 1.4:

$$\begin{aligned} dy \int_{-\infty}^{\infty} & \frac{1}{(2\pi a)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2a}\right) \frac{1}{(2\pi b)^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2b}\right] dx \\ &= dy \frac{1}{[2(a+b)]^{\frac{1}{2}}} \exp\left(-\frac{y^2}{2(a+b)}\right) \end{aligned} \quad (1.4)$$

That is the law of composition of Gaussian distributions. Notice that this parameter a adds up when we compound two Gaussian distributions.

If we then consider that the wandering in nonoverlapping intervals is Gaussian and independent, and that the amount of wandering is dependent only on the time interval and not on the original time, then we see that, because a and b add, the distribution starting from a certain time and ending at a certain other time is given by

$$\frac{1}{(2\pi k)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2k}\right) \quad (1.5)$$

where k depends linearly on the time difference t ; and I shall normalize it so that it is t itself; thus

$$\frac{1}{(2\pi t)^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2t}\right) \quad (1.6)$$

Notice that if I compound Gaussian distributions, then I add the time parameters for the new Gaussian distributions.

For the moment, I am going to consider trying to map all paths of particles on the variable α when α goes from 0 to 1. I want to set up this mapping in detail. The first thing is that all random motions of particles end up somewhere; that is, they will all map on this interval 0 to 1 of α . At the present stage, I have not separated the different ranges of α ; I have taken the whole range from 0 to 1 and assigned it to all of the curves.

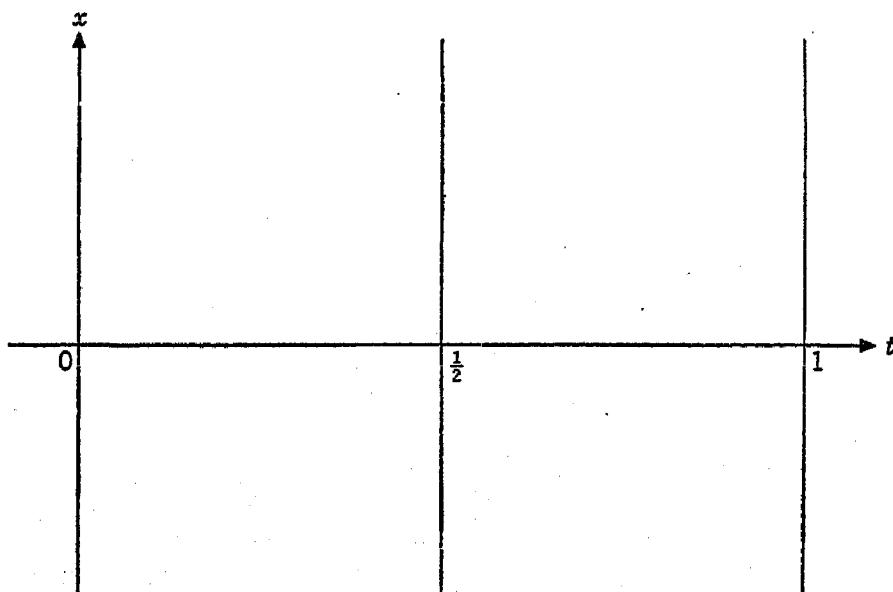


Figure 1.2

Let us consider the Brownian motion at two times, $\frac{1}{2}$ and 1 (Figure 1.2). I shall introduce four classes of Brownian motion. Notice that at time $\frac{1}{2}$ we have two possibilities: below the axis and above the axis. The probability of it landing on the axis is 0. It is like tossing a coin, where

the probability of standing on end is 0. We toss the coin twice: once at time $\frac{1}{2}$ and once at time 1. Given that it is either heads or tails, there is a possibility of heads-heads, heads-tails, tails-heads, and tails-tails. What are these possibilities of heads-heads, tails-tails, and so forth? The two values of a and b in the case of Equation 1.4 corresponding to the two tossings are both $\frac{1}{2}$. The distribution is given by

$$\frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2(\frac{1}{2})}\right] \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2(\frac{1}{2})}\right] dx dy \quad (1.7)$$

Expression 1.7 is the probability that the particle at time $\frac{1}{2}$ lies between x and $x + dx$, and at time 1 lies between y and $y + dy$. On this basis, the probability that the particle lies, say, below the axis at time $\frac{1}{2}$ (that x be negative at time $\frac{1}{2}$) and x be negative at time 1 is given by

$$\int_{-\infty}^0 \int_{-\infty}^0 \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2(\frac{1}{2})}\right] \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2(\frac{1}{2})}\right] dx dy \quad (1.8)$$

This expression is a quantity that has a definite value. Now let us start from 0 and lay out an interval AB of that length on the line of α (Figure 1.3). We shall say that AB corresponds to curves ending below both



Figure 1.3

axes. Then we shall take x between 0 and ∞ , and take y between $-\infty$ and 0; and that probability will be given by

$$\int_{x=0}^{\infty} \int_{y=-\infty}^0 \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2(\frac{1}{2})}\right] \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2(\frac{1}{2})}\right] dx dy \quad (1.9)$$

Let us then lay off another interval, BC , of α , that will correspond to Expression 1.9 (Figure 1.4). (Notice that the first two probabilities will

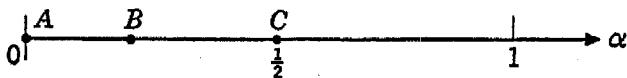


Figure 1.4

add up to the probabilities of all the curves that end below the axis, and that probability will be $\frac{1}{2}$.) Then we integrate Expression 1.7 as follows:

$$\int_{x=-\infty}^0 \int_{y=0}^{\infty} \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2(\frac{1}{2})}\right] \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2(\frac{1}{2})}\right] dx dy \quad (1.10)$$

That will give us a third segment CD of the line (Figure 1.5). Finally, Expression 1.11 gives the fourth segment DE (Figure 1.6).

$$\int_{x=0}^{\infty} \int_{y=0}^{\infty} \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2(\frac{1}{2})}\right] \frac{1}{[2\pi(\frac{1}{2})]^{\frac{1}{2}}} \exp\left[-\frac{(y-x)^2}{2(\frac{1}{2})}\right] dx dy_1 \quad (1.11)$$

We now have four segments of the line α , that is, four probabilities that add up to 1.

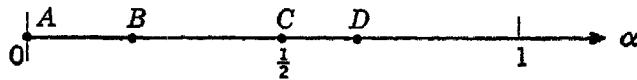


Figure 1.5

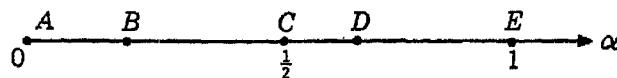


Figure 1.6

Let us take a finer subdivision of the Brownian motions; we shall take the subdivisions where, as you will easily see, every possibility lies in one and only one of the previous possibilities, and where a certain fixed number of these possibilities will add up to give the previous complete probability. We shall get a finer subdivision of the α line; it will correspond to regions of wandering of the Brownian motion. The first thing that I shall do, is to divide the time finer. That is, I shall work by quarter

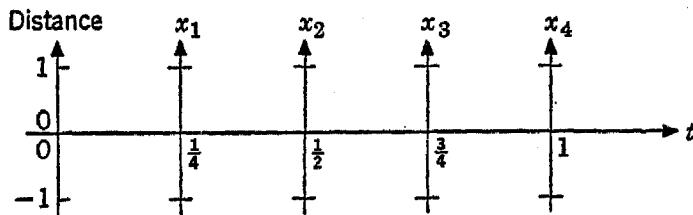


Figure 1.7

times (Figure 1.7). The probability that a particle lies in a small region specified at each of the four times is given by the integral of

$$\begin{aligned} & \left\{ \frac{1}{[2\pi(\frac{1}{4})]^{\frac{1}{2}}} \right\}^4 \exp\left[-\frac{x^2}{2(\frac{1}{4})}\right] \exp\left[-\frac{(x_2 - x_1)^2}{2(\frac{1}{4})}\right] \\ & \times \exp\left[-\frac{(x_3 - x_2)^2}{2(\frac{1}{4})}\right] \exp\left[-\frac{(x_4 - x_3)^2}{2(\frac{1}{4})}\right] dx_1 dx_2 dx_3 dx_4 \quad (1.12) \end{aligned}$$

over the appropriate region.

I now consider the probability that the particle moves through a hole (region) specified at each of the four times. I divide the time into regions

$\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and 1, and the regions will now be 4 for each time. At each one of these times, not only do I distinguish positive and negative values but I distinguish values from $-\infty$ to -1 , from -1 to 0, from 0 to 1, and from 1 to ∞ . How many types of curves will I have? I will have four possible regions for each time, but there will be four successive times. That is 4^4 regions, which is 256. How do I get the probability for each one of those regions? I simply integrate Expression 1.12 with respect to the different x 's over that region. I will get 256 quantities that add up to 1; moreover, there will be 64 of these quantities adding up to each of the previous intervals. They will not be the same length, but I will be able to get, with a definite arrangement, 256 subregions, each adding up by 64's to the previous subregions, and I will have a mapping of ranges of my random curve on smaller ranges of α .

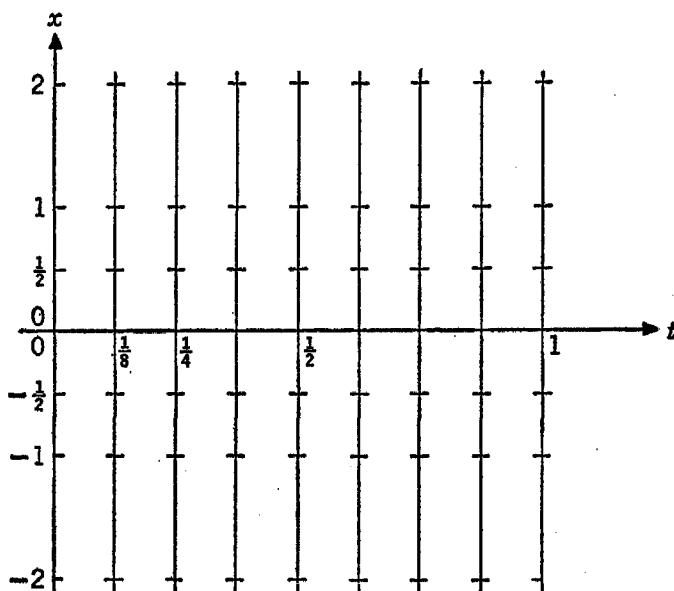


Figure 1.8

Having done this, I take a finer subdivision (Figure 1.8). In the first place, the times will now go by eights, so that for the probability density we shall have

$$\left\{ \frac{1}{[2\pi(\frac{1}{8})]^{\frac{1}{2}}} \right\}^8 \exp \left[-\frac{x^2}{2(\frac{1}{8})} \right] \cdots \exp \left[-\frac{(x_8 - x_7)^2}{2(\frac{1}{8})} \right] \quad (1.13)$$

I shall make the integration over regions in the following way (and this trick I shall continue): I have now eight times. I take all of the previous subdivisions, not only for the times I had before but also for these other intermediate times. Then, for any of these intervals that have limits at both ends, I introduce a new subdivision halfway up. There is no virtue to a half, but using a half is a perfectly good way of

doing it. For the regions that run to $+\infty$, the new subdivision will be one unit up; for the regions that run to $-\infty$, the new subdivision will be one unit down.

Now we have not 4^4 regions but 8^8 of these curves. Each one of these will have a length; each one of these will fit into one of the regions of the previous subdivisions, and have finer divisions of the 0-to-1 line of α . Thus, as we continue cutting the motion finer and finer, both in time and space, we shall have a larger number of steps of α . With this trick, or with a much more general trick of subdivision, it is easy to show that these intervals of α will all go to 0 in length. If we consider any value of α that is not one of this denumerable set of subdivisions, this value of α will lie in one of the four regions that we got at the first stage, and in one of the 256 regions that we got at the second stage, and so on. It will be uniquely determined (except for the boundaries that do not count) by where it is at the different stages of subdivisions. In other words, I am able to box in these wanderings of our point more and more finely.

It will not be true, in general, that as I do this I close down on a particular curve that is continuous; but I want to show you that I can get around this difficulty. I can introduce a certain quantity, which I shall call the straightness of a series of subdivisions. There is no virtue in the 2^n at present, so that I just say that I have a series of holes—one hole at each of an arbitrary number of times (Figure 1.9).

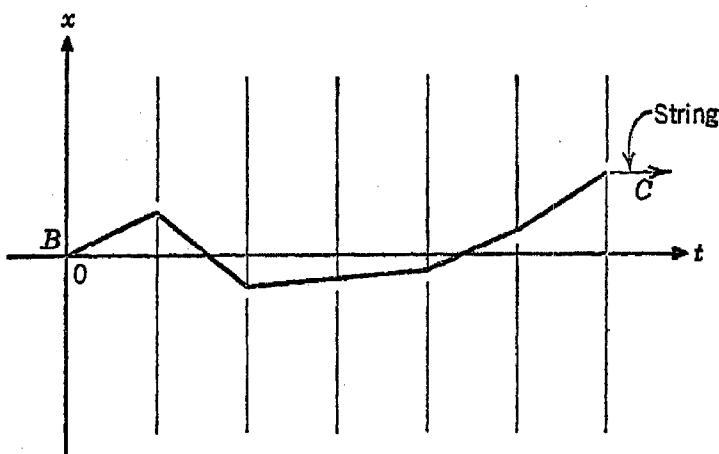


Figure 1.9

Suppose that we take a string BC fastened at 0 and thread it through the holes. Let me pull that string taut. Then there will be at least one part of the curve with maximum slope. This maximum slope for some taut string will be less than the maximum slope of any other curve passing through the same series of holes. We can call this taut string the straightest string that lies in the series of holes. This string will give us the smoothness of the series of holes in the following manner. I call this

straightest string through a series of holes, $x = s(t)$, for that series of holes; a particular s will belong to each series of holes. Now consider

$$\max \frac{|s(t + \tau) - s(t)|}{|\tau|^{\frac{1}{4}}} \quad (1.14)$$

That maximum value I call the roughness of the series of holes. There is no virtue in one fourth, except this: It is a definite number less than one half. Now I have a series of holes at each stage of subdivision, and at each stage of subdivision I throw away all of the series of holes for which the roughness is greater than A ; that is,

$$\max \frac{|s(t + \tau) - s(t)|}{|\tau|^{\frac{1}{4}}} > A \quad (1.15)$$

I can now prove the following: The sum of the lengths of the α mappings of the series of holes that I have thrown away at each stage will be finite. At each stage when I throw away this series of holes, I will have an expression $l(A)$, the sum of the lengths that I have thrown away. There will be only a denumerable number of holes in the series (a denumerable number of intervals); the sum of these lengths will converge and will form what we call a measurable set. The measure of that set will be less than $l(A)$.

Although I shall not spend the time to do it here, I can prove that

$$\lim_{A \rightarrow \infty} l(A) = 0 \quad (1.16)$$

That is, the sum of the lengths of all of the series of holes at any stage which are less smooth than a certain amount is finite; and if A increases, this sum goes to 0. Let me discard the series of holes that are less than a certain smoothness. That means that I discard a certain set of values of α of measure less than some small amount. Let us take the remaining series of holes. The remaining series of holes will contain at each stage a continuous curve; moreover, all of these continuous curves will satisfy the same condition of equicontinuity. That is, Δx goes to 0 faster than a certain function of Δt .

We invoke the following mathematical theorem: Suppose that I take a set of curves that pass through a series of holes and satisfy a certain condition of equicontinuity. Suppose, also, that I make the holes narrow down to zero, and that I increase the density of the time instants at which the holes are defined. Our condition of equicontinuity binds my curves tighter and tighter at more and more points. The curves then tend uniformly to a limit that satisfies the same condition of equicontinuity.

That is, if I throw away a certain set of values of α and s such that the measure of s is less than ϵ , then to all of the remaining values of α there will be assigned a limit curve for all the series of holes that correspond successively to this value of α . The limit curve will be continuous; not only that, it will also satisfy the same condition of equicontinuity. Thus, by this process, I have assigned to all values of α (except for a set of zero measure) a curve, which I call $x(t, \alpha)$, that satisfies some condition of equicontinuity; and what is more, to all of these curves except a set of measure zero I have assigned a value of α . These limit curves will prove to be unique. Therefore I will have assigned, except for a set of zero measure, to every value of α , one and only one continuous curve. This I call $x(t, \alpha)$, as I have said. It is a well-defined function of t for almost all values of α . It can easily be shown to be a measurable, bounded function of t and α , and a continuous function of t for almost all α . Furthermore, if the process at each stage of arrangement of the holes is given definitely, which can be done, this is a well-defined function of t and α —as well-defined as any mathematical function. So, I have now introduced what we call the stochastic function $x(t, \alpha)$.

Are there any other things that I can say about $x(t, \alpha)$? I have said that for almost all values of α this is continuous. Is it differentiable? I shall not go into the proof—it is shown by merely taking the contrary cases, adding them, and counting—but I can say that the following thing can be proved.

Consider the set of curves for which

$$\frac{\Delta_t x(t, \alpha)}{\Delta t} \quad (1.17)$$

has a limit as Δt tends to 0 for at least one value of t . This set of curves has zero measure. That is, almost all of the curves $x(t, \alpha)$ are nowhere differentiable. This is important. We shall have to use nondifferentiable continuous curves in the work that we are doing. Not only that; the limit of Expression 1.18 will exist for no t for almost all α if λ is greater than $\frac{1}{2}$, and will exist for every t for almost all α uniformly if λ is less than $\frac{1}{2}$.

$$\frac{\Delta_t x(t, \alpha)}{(\Delta t)^\lambda} \quad (1.18)$$

(I am leaving out the $\frac{1}{2}$ case. I am stating the facts here rather than proving.)

The function that I want to use in our further work is $x(t, \alpha)$. I assume that we have established $x(t, \alpha)$. We shall call it the stochastic function. Now I want to build up some of the integral properties of $x(t, \alpha)$.

Let us consider the following integral:

$$I = \int_0^1 [x(t_2, \alpha) - x(t_1, \alpha)]^n d\alpha \quad [t_1 < t_2] \quad (1.19)$$

Now, $[x(t_2, \alpha) - x(t_1, \alpha)]$ has a Gaussian distribution. That can be proved very easily from our definition. With $[t_2 - t_1]$ as the parameter of the Gaussian distribution, Equation 1.20 follows.

$$I = \frac{1}{[2\pi(t_2 - t_1)]^{1/2}} \int_{-\infty}^{\infty} u^n \exp\left[-\frac{u^2}{2(t_2 - t_1)}\right] du \quad [t_1 < t_2] \quad (1.20)$$

This equation is certainly true for binary intervals, and by continuity we can extend it easily to nonbinary intervals. That is, these distributions are Gaussian. Equation 1.20 can be computed as follows. Let

$$v = \frac{u}{(t_2 - t_1)^{1/2}} \quad (1.21)$$

Then

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} v^n \exp\left(-\frac{v^2}{2}\right) dv \quad (1.22)$$

Note that $\exp(-v^2/2)$ is an even function. If n is odd, then

$$(t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} v^n \exp\left(-\frac{v^2}{2}\right) dv = 0 \quad (1.23)$$

because the integral from $-\infty$ to $+\infty$ of the product of an odd and an even function is zero.

If n is even, we can compute Equation 1.22 by an integration by parts. Note that

$$-v \exp\left(-\frac{v^2}{2}\right) dv = d \exp\left(-\frac{v^2}{2}\right) \quad (1.24)$$

Then we have

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} (-1)v^{n-1} d \exp\left(-\frac{v^2}{2}\right) \quad (1.25)$$

Integrating by parts, we get

$$\begin{aligned} I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} & \left\{ \left[-v^{n-1} \exp\left(-\frac{v^2}{2}\right) \right]_{-\infty}^{\infty} \right. \\ & \left. + \int_{-\infty}^{\infty} (n-1)v^{n-2} \exp\left(-\frac{v^2}{2}\right) dv \right\} \end{aligned} \quad (1.26)$$

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} (n-1) \int_{-\infty}^{\infty} v^{n-2} \exp\left(-\frac{v^2}{2}\right) dv \quad (1.27)$$

Continuing this method, we get

$$I = (t_2 - t_1)^{n/2} (n - 1)(n - 3) \cdots (1) \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \quad (1.28)$$

The series $\{n - 1 - 2k\}$ goes down to 1 because n is even. At the 1 stage, we can evaluate

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv = 1 \quad (1.29)$$

Hence

$$\begin{aligned} & \int_0^1 [x(t_2, \alpha) - x(t_1, \alpha)]^n d\alpha \\ &= \begin{cases} 0, & n \text{ odd} \\ (t_2 - t_1)^{n/2} (n - 1)(n - 3) \cdots (1), & n \text{ even} \end{cases} \end{aligned} \quad (1.30)$$

The right-hand side of Equation 1.30 is interesting. Suppose that I have n terms and that I want to consider the number of ways in which I can divide n terms into pairs. If n is odd, there is no way of separating n terms into pairs. If n is even, let us see how many ways there are. We take any term. There are $n - 1$ possible terms we can match it with. When we have done that, there are $n - 2$ terms. Then we take any term, and we can match that with $n - 3$ possibilities. Hence, in all cases, odd or even, the following equation is true:

$$\begin{aligned} & \int_0^1 [x(t_2, \alpha) - x(t_1, \alpha)]^n d\alpha = (t_2 - t_1)^{n/2} \\ & \times (\text{the number of ways of separating } n \text{ terms into pairs}) \end{aligned} \quad (1.31)$$

This is the beginning of the calculus of random functions.

Now let us consider the following integral:

$$\int_0^1 f(t) dx(t, \alpha) \quad (1.32)$$

Notice that as an ordinary Stieltjes integral this integral will not exist because $x(t, \alpha)$ is almost never differentiable. We can get around this. For the moment, let us suppose that $f(t)$ is differentiable and that $f'(t)$ is bounded. (I confine myself at present to that case.) First, we shall define Expression 1.32 by integration by parts.

$$\int_0^1 f(t) dx(t, \alpha) = f(1) x(1, \alpha) - f(0) x(0, \alpha) - \int_0^1 f'(t) x(t, \alpha) dt \quad (1.33)$$

Now $x(0, \alpha)$ is 0. Remember that the strings that I pick pass through the origin. Therefore

$$\int_0^1 f(t) dx(t, \alpha) = f(1) x(1, \alpha) - \int_0^1 f'(t) x(t, \alpha) dt \quad (1.34)$$

Furthermore, the right-hand side of Equation 1.34 is a well-defined expression for almost all values of α . The function $x(t, \alpha)$ is bounded, f' is bounded, and hence the integral of Equation 1.34 exists. Moreover, if I change $x(t, \alpha)$ into $-x(t, \alpha)$, you will see that the distribution of the Brownian motions is not changed at all. If I just simply turn $x(t, \alpha)$ into $-x(t, \alpha)$, not only will Equation 1.34 remain valid, but its integral with respect to α will be 0.

The next expression that I want to evaluate is

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 \quad (1.35)$$

Integrating by parts and expanding, we have

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 = \int_0^1 d\alpha \left[f(1) x(1, \alpha) - \int_0^1 f'(t) x(t, \alpha) dt \right]^2 \quad (1.36)$$

This yields

$$\begin{aligned} \int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 &= \int_0^1 d\alpha f^2(1) x^2(1, \alpha) \\ &\quad - 2 \int_0^1 d\alpha f(1) x(1, \alpha) \int_0^1 f'(t) x(t, \alpha) dt \\ &\quad + \int_0^1 d\alpha \int_0^1 f'(t) x(t, \alpha) dt \int_0^1 f'(s) x(s, \alpha) ds \end{aligned} \quad (1.37)$$

We now consider the following expression:

$$\int_0^1 x(t_1, \alpha) x(t_2, \alpha) d\alpha = \int_0^1 x(t_1, \alpha) [x(t_2, \alpha) - x(t_1, \alpha)] d\alpha \quad (1.38)$$

I assume that t_1 is less than t_2 . By adding and subtracting the same thing, we are able to check the right-hand side of Equation 1.38. Note that $x(t_1, \alpha)$ and $[x(t_2, \alpha) - x(t_1, \alpha)]$ are independent in distribution. So, when one is plus, the other is equally likely to be plus or minus; and the average of the product will be zero. Thus, the only quantity that survives is

$$\int_0^1 x(t_1, \alpha) x(t_2, \alpha) d\alpha = \int_0^1 x^2(t_1, \alpha) d\alpha = t_1 \quad (1.39)$$

Now consider the three terms of Equation 1.37. The first term is given by

$$\int_0^1 d\alpha f^2(1) x^2(1, \alpha) = f^2(1) \quad (1.40)$$

Let us suppose that we can interchange the order of integration for the remaining two terms of Equation 1.37. (It can be fairly easily proved that this interchange is justified.) Hence, for the second term, we have

$$-2 \int_0^1 d\alpha f(1) x(1, \alpha) \int_0^1 f'(t) x(t, \alpha) dt = -2f(1) \int_0^1 t f'(t) dt \quad (1.41)$$

Equation 1.41 is true, because t is less than 1.

Now consider the remaining term of Equation 1.37.

$$\begin{aligned} & \int_0^1 d\alpha \int_0^1 f'(t) x(t, \alpha) dt \int_0^1 f'(s) x(s, \alpha) ds \\ &= \int_0^1 f'(t) dt \int_0^1 f'(s) ds \int_0^1 x(t, \alpha) x(s, \alpha) d\alpha \end{aligned} \quad (1.42)$$

Recall that

$$\int_0^1 x(t, \alpha) x(s, \alpha) d\alpha = \begin{cases} s, & s \leq t \\ t, & t \leq s \end{cases} \quad (1.43)$$

Equation 1.43 means that we can divide the integral of Equation 1.42 into two ranges; in one of them, s is smaller; in one, t is smaller. But that means that the two integrals are the same integral, because s and t are completely interchangeable. I substitute Equation 1.43 in Equation 1.42, keeping s less than t , and I multiply by 2 to account for the two cases. The right-hand side of Equation 1.42 becomes

$$\begin{aligned} & \int_0^1 f'(t) dt \int_0^1 f'(s) ds \int_0^1 x(t, \alpha) x(s, \alpha) d\alpha \\ &= 2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds \end{aligned} \quad (1.44)$$

Substitution of Equations 1.40, 1.41, and 1.44 in Equation 1.37 gives

$$\begin{aligned} \int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 &= f^2(1) - 2f(1) \int_0^1 t f'(t) dt \\ &+ 2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds \end{aligned} \quad (1.45)$$

Now

$$2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds = 2 \left[\int_0^t s f'(s) ds f(t) \right]_0^1 - 2 \int_0^1 t f(t) f'(t) dt \quad (1.46)$$

$$2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds = 2f(1) \int_0^1 s f'(s) ds - 2 \int_0^1 t f(t) f'(t) dt \quad (1.47)$$

Therefore

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 = f^2(1) - 2 \int_0^1 t f(t) f'(t) dt \quad (1.48)$$

Now

$$\frac{d}{dt} [t f^2(t)] = f^2(t) + 2t f(t) f'(t) \quad (1.49)$$

Therefore

$$2 \int_0^1 t f(t) f'(t) dt = [t f^2(t)]_0^1 - \int_0^1 f^2(t) dt \quad (1.50)$$

$$2 \int_0^1 t f(t) f'(t) dt = f^2(1) - \int_0^1 f^2(t) dt \quad (1.51)$$

Thus

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 = \int_0^1 f^2(t) dt \quad (1.52)$$

Notice what we have if we start with the assumption that $f(t)$ is differentiable and belongs to L^2 and we go from that to the function of α given by Equation 1.52. We have a unitary transformation. This allows us to extend the definition of the integral to any function $F(\alpha)$ that belongs to the Lebesgue class L^2 by the following trick.

Suppose that we have a sequence $f_n(t)$ of real functions which belong to L^2 such that Condition 1.53 is satisfied:

$$\int_0^1 [f_n(t) - f(t)]^2 dt \rightarrow 0 \quad (1.53)$$

Clearly, given any function f in L^2 , I can find such a sequence. (Every function of L^2 , Lebesgue-measurable and Lebesgue-integrable-square, can be approximated by functions of bounded derivatives.) There is no problem about that.

Then I form $F_n(\alpha)$ defined by Equation 1.54.

$$F_n(\alpha) = \int_0^1 f_n(t) dx(t, \alpha) \quad (1.54)$$

It follows at once that

$$\int_0^1 [F_n(\alpha) - F_m(\alpha)]^2 d\alpha = \int_0^1 [f_n(t) - f_m(t)]^2 dt \quad (1.55)$$

and that

$$\int_0^1 [f_n(t) - f_m(t)]^2 dt \rightarrow 0 \quad (1.56)$$

as m and n tend to ∞ independently.

Now we use the Riesz-Fischer theorem. If we have a sequence $\{F_n(\alpha)\}$ belonging to L^2 such that

$$\int_0^1 [F_n(\alpha) - F_m(\alpha)]^2 d\alpha \rightarrow 0 \quad (1.57)$$

then there is a function of α to which they converge in the mean.

$$F(\alpha) = \lim_{n \rightarrow \infty} F_n(\alpha) \quad (1.58)$$

It can be proved that $F(\alpha)$ does not depend on the sequence $\{F_n(\alpha)\}$ that approximates it but that it will be the same for any such sequence.

I now define

$$F(\alpha) = \int_0^1 f(t) dx(t, \alpha) \quad (1.59)$$

Equation 1.59 applies to almost all values of α . $F(\alpha)$ is a function of L^2 . There is no problem in verifying that

$$\int_0^1 F^2(\alpha) d\alpha = \int_0^1 f^2(t) dt \quad (1.60)$$

So, we have extended our integral to all functions belonging to the Lebesgue class L^2 .

REFERENCES

- N. Wiener, "Generalized Harmonic Analysis," *Acta Mat.* 55, 117-258 (1930).
- R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, Colloquium Publication No. 19, American Mathematical Society, New York, 1934.

LECTURE 2

Homogeneous Polynomial Functionals and Their Averages

Let us suppose that

$$\phi \in L^2 \quad \text{over } [0, 1] \quad (2.1)$$

I have previously defined

$$\int_0^1 \phi(t) dx(t, \alpha) \quad (2.2)$$

As you remember, $x(t, \alpha)$ is the Brownian motion function, which I defined and which is a perfectly explicit function. Now, I proved first that

$$\int_0^1 d\alpha \int_0^1 \phi(t) dx(t, \alpha) = 0 \quad (2.3)$$

and that

$$\int_0^1 d\alpha \left[\int_0^1 \phi(t) dx(t, \alpha) \right]^2 = \int_0^1 \phi^2(t) dt \quad (2.4)$$

Here, $\phi(t)$ is real.

From Lecture 1, it follows that:

Given $x(t_1, \alpha) x(t_2, \alpha) \cdots x(t_n, \alpha)$, where the t 's and α 's all lie between 0 and 1, then for n odd,

$$\int_0^1 x(t_1, \alpha) x(t_2, \alpha) \cdots x(t_n, \alpha) d\alpha = 0 \quad (2.5)$$

and for n even,

$$\int_0^1 x(t_1, \alpha) x(t_2, \alpha) \cdots x(t_n, \alpha) d\alpha = \sum \prod \int_0^1 x(t_j, \alpha) x(t_k, \alpha) d\alpha \quad (2.6)$$

where the sum is over all ways of dividing n terms t_1, \dots, t_n into pairs, and the product is over all pairs in this way of dividing.

It follows at once that

$$\int_0^1 d\alpha \int_0^1 \phi_1(t_1) dx(t_1, \alpha) \int_0^1 \phi_2(t_2) dx(t_2, \alpha) \cdots \int_0^1 \phi_n(t_n) dx(t_n, \alpha) \quad (2.7)$$

will be a sum of terms similar to the right-hand side of Equation 2.6. Evaluating, we get

$$\begin{aligned} \int_0^1 d\alpha \int_0^1 \phi_1(t_1) dx(t_1, \alpha) \int_0^1 \phi_2(t_2) dx(t_2, \alpha) \cdots \int_0^1 \phi_n(t_n) dx(t_n, \alpha) \\ = \sum \prod \int_0^1 \phi_j(t) \phi_k(t) dt \end{aligned} \quad (2.8)$$

Let us assume

$$K_n(\tau_1, \dots, \tau_n) = \phi_1(\tau_1) \cdots \phi_n(\tau_n) \quad (2.9)$$

where the ϕ 's all belong to L^2 . More generally, I will use a sum of such products of terms. That is,

$$K_n(\tau_1, \dots, \tau_n) = \sum \phi_1(\tau_1) \cdots \phi_n(\tau_n) \quad (2.10)$$

Then, for n odd,

$$\int_0^1 d\alpha \int_0^1 \cdots \int_0^1 dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) K_n(\tau_1, \dots, \tau_n) = 0 \quad (2.11)$$

and for n even,

$$\begin{aligned} \int_0^1 d\alpha \int_0^1 \cdots \int_0^1 dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) K_n(\tau_1, \dots, \tau_n) \\ = \sum \int_0^1 d\tau_1 \cdots \int_0^1 d\tau_m K_n(\tau_1, \tau_1, \tau_2, \tau_2, \dots, \tau_m, \tau_m) \end{aligned} \quad (2.12)$$

where $2m = n$. That is, I simply multiply out the factors of K_n , and the sum is over all ways of dividing in pairs.

I am going to assume that K_n is symmetrical, so that I shall not have to distinguish the order in which I integrate. I think you can see that if K_n is not symmetrical I can make a symmetrical K_n in an expression of this sort by simply taking all the permutations of the τ 's, adding them, and dividing by the number of permutations. The functional that I get

by that method will not be changed; I am adding the same functional to itself with a different labeling. Then

$$\begin{aligned} & \int_0^1 d\alpha \int_0^1 \cdots \int_0^1 dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) K_n(\tau_1, \dots, \tau_n) \\ &= (2m - 1)(2m - 3) \cdots (1) \int_0^1 d\tau_m K_n(\tau_1, \tau_1, \dots, \tau_m, \tau_m) \quad (2.13) \end{aligned}$$

In other words, for a symmetrical K_n , we divide the τ 's into pairs, identify the τ 's in each pair, integrate over them, and add up all the ways of doing that. How many ways are there of doing that? Well, all of these operations will be the same if the K_n 's are symmetrical. How many ways are there of dividing any $2n$ things into pairs? I showed you that at the last lecture: $(2n - 1)(2n - 3) \cdots (1)$. So, we have an integral not merely for linear functions but also for a certain very important class of nonlinear functions.

Let me now take another set of functions $\phi_n(t)$. These are to be normal and orthogonal over $[0, 1]$, and real, for the moment. Let us consider expressions of the form

$$\int_0^1 d\alpha \left[\int_0^1 \phi_n(t) dx(t, \alpha) \right]^\nu \quad (2.14)$$

where the ϕ_n 's belong to L^2 . The ϕ_n 's must be a normal, orthogonal set, and

$$\int_0^1 d\alpha \left[\int_0^1 \phi_n(t) dx(t, \alpha) \right]^\nu = 0, \quad \text{for } \nu \text{ odd} \quad (2.15)$$

$$\int_0^1 d\alpha \left[\int_0^1 \phi_n(t) dx(t, \alpha) \right]^\nu = (\nu - 1)(\nu - 3) \cdots (1) \int_0^1 \phi_n^2(t) dt, \quad \text{for } \nu \text{ even} \quad (2.16)$$

But

$$\int_0^1 \phi_n^2(t) dt = 1 \quad (2.17)$$

And so, for ν even,

$$\int_0^1 d\alpha \left[\int_0^1 \phi_n(t) dx(t, \alpha) \right]^\nu = (\nu - 1)(\nu - 3) \cdots (1) \quad (2.18)$$

Now, the right-hand side of Equation 2.18 represents again, simply, without any other factor, the number of ways of dividing ν things into pairs; and if I take

$$\frac{1}{(2\pi)^{\frac{\nu}{2}}} \int_{-\infty}^{\infty} u^\nu \exp\left(-\frac{u^2}{2}\right) du \quad (2.19)$$

that will be the same thing as Expression 2.18, as I pointed out last time. In other words, if I consider the moments of the expressions

$$\int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.20)$$

then these moments will be exactly the same as Expression 2.19. What I have said is equivalent to the following statement: The moments will completely determine the distribution (that, we know), and the distribution of the quantity in Expression 2.20 is Gaussian. That is, the probability that

$$u \leq \int_0^1 \phi_n(t) dx(t, \alpha) \leq u + du \quad (2.21)$$

is given by

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{u^2}{2}\right) du \quad (2.22)$$

So, the first thing that I have is that each of these expressions has a Gaussian distribution. Now, I want to consider the simultaneous distribution of two or more expressions of the form of Expression 2.20. Let us see what we get. Let us take

$$\int_0^1 \left[\int_0^1 \phi_m(t) dx(t, \alpha) \right]^{\nu_1} \left[\int_0^1 \phi_n(t) dx(t, \alpha) \right]^{\nu_2} \left[\int_0^1 \phi_p(t) dx(t, \alpha) \right]^{\nu_3} d\alpha \quad (2.23)$$

where ϕ_m , ϕ_n , and ϕ_p are distinct. We take three terms as representative; we are not confining ourselves to two.

Now we want to get the integral of Expression 2.23. When we do that, remember what we do. We divide the ϕ 's into pairs in all possible ways, integrate each pair, and add. But since the ϕ 's are orthogonal, the integral of the product of two distinct ϕ 's is 0. So, any pair in which the ϕ 's are distinct disappears, and the only ones that remain are those in which the ϕ 's are the same. And so, we find that Expression 2.23 is equal to the product of the moments of the individual terms. In other words, the moment of the product is the product of the moments. This allows us to prove that the distributions are independent. That is, we can merely multiply the moments; and we can easily show that we then simply multiply the distributions. In other words, the quantities

$$\int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.24)$$

all have the same distribution (which is Gaussian, with the same parameter) and are independent of one another. That is the first thing. To

repeat—the quantities

$$\int_0^1 \phi_n(t) dx(t, \alpha)$$

are distributed Gaussianly and independently.

The next thing is this. If the ϕ 's are a closed set, it is easy to show that the polynomials in these integrals are a closed set of functions of α and that every function of α can be represented in terms of them. I am not going into the details. The result is that, given any closed normal and orthogonal set of functions of α , we have reduced the problem of the distribution of the functions to the distribution of these integrals, and these integrals all have the same distribution. They are all Gaussian and all independent, as I just explained. Now, this is extremely interesting, and you will see that there is something here that is suggestive of Fourier coefficients. Formally,

$$\int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.25)$$

will be the Fourier coefficients of

$$\frac{d}{dt} x(t, \alpha) \quad (2.26)$$

with respect to the functional ϕ . However, this does not exist in the ordinary sense. It is not a differentiable function. If we have a differentiable function here, say $F(t)$, the integral

$$\int_0^1 F(t) \phi_n(t) dt \quad (2.27)$$

is the Fourier coefficient of $F(t)$. Call this a_n , and it follows that

$$\sum_0^\infty |a_n|^2 = \int_0^1 F^2(t) dt \quad (2.28)$$

Everything here is real, and the space of this set of coefficients is called Hilbert space. This is real Hilbert space. (Later on, I shall go on to complex Hilbert space. But for the moment we are confining ourselves to real Hilbert space.) However, these coefficients

$$a_n(\alpha) = \int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.29)$$

are not a set that is such that the sum of the squares converges. However, they are a set with independent Gaussian distributions. If I take the

sum of the squares, the probability that this sum is finite is 0. So here I have something that is quite analogous to Hilbert space but is not Hilbert space. In Hilbert space the sum of the squares of the coefficients is finite. Here the coefficients all have independent Gaussian distributions. This space I call differential space.

There are many ways to go ahead. First, suppose that I consider another normal and orthogonal set $\{\psi_n\}$ which is also closed, and I consider

$$\int_0^1 \psi_n(t) dx(t, \alpha) \quad (2.30)$$

and compare these expressions with

$$\int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.31)$$

These two random variables will have the same distribution for the ϕ 's as they had for the ψ 's because the Gaussian properties are not changed. Now, I make a linear transformation that transforms ϕ_n into ψ_n and any series in one into a series in the other. That is what we call a unitary transformation. It is the general sort of transformation that does not affect the integral of the square of the function. I am talking of real functions now. So, I can write

$$\psi_n(t) = U \phi_n(t) \quad (2.32)$$

which is a unitary transformation, a real unitary transformation. Now what do I do here to the α ? I can work backwards from my distribution of

$$\int_0^1 \phi_n(t) dx(t, \alpha) \quad (2.33)$$

to the values of α . We shall now confine our attention to a particular set of functions of $\phi_n(t)$ known as Walsh functions. These functions have the property of being linear combinations of all functions that are 1 over an interval $(k/2^n, (k+1)/2^n)$, where $0 \leq k < 2^n$, and 0 elsewhere, and the characteristic functions of such sets are linear combinations of the Walsh functions. The first Walsh function is 1 over the line 0 to 1. The next of the functions is 1 over the first half of the line and -1 over the second half. It is orthogonal to the first. The next function is +1 over the first quarter, -1 over the second quarter, 1 over the third quarter, and -1 over the fourth quarter. That function is orthogonal to both the others. The next function is 1 over the first quarter, -1 over the second

quarter, -1 over the third, 1 over the fourth. These functions are known as Walsh functions, and I shall give you the general rule for them. I take a function which is 1 over the first 2^{-n} th of the line, -1 over the second 2^{-n} th, 1 over the third, -1 over the fourth, 1 over the fifth, and so forth. This is the n th Haar function. Now, I take the product of the Haar functions two at a time, three at a time, and so on. The resultant functions ϕ_n are the Walsh functions. Any one of these Walsh functions, over a large interval, will always range between 1 and -1 while any other is 1 or -1 . The result is that these Walsh functions will be orthogonal to one another, and will be normal.

Then in terms of these Walsh functions of ϕ we take

$$\int_0^1 \omega(t) dx(t, \alpha) \quad (2.34)$$

where $\omega(t)$ is one of the Walsh functions of t . From that I can obtain α because I have the integrals and I have the differences between the x 's over every binary interval. From that I can go over the process by which I have obtained the α . Now, if I start with another set of orthogonal functions than the Walsh functions, I can say that the integrals of Expression 2.34 are to be the corresponding integrals of the Walsh functions of another variable, and I get the α of the other variable. When I have done that, I close down on α , and I get a one-one point-mapping of the new α 's on the old α 's; and this mapping preserves measure because it preserves the integral over any interval. In other words, if I make this transformation, I get the following result:

$$\int_0^1 U[\phi_n(t)] dx(t, \alpha) = \int_0^1 \psi_n(t) dx(t, T\alpha) \quad (2.35)$$

where $T\alpha$ is a measure-preserving transformation. In other words, I am able to refer unitary transformations on the ϕ space to measure-preserving transformations on the α space. This is very important.

There are several things of a fairly simple character that I can do to generalize this material. There are two generalizations that I will have time to do now: (a) the generalization for a range of t , not 0 to 1 , but $-\infty$ to ∞ (which is quite important when we consider spectral problems), and (b) the generalization from a real differential space to a complex differential space. Now for the first step. Suppose that we take the line of t and divide it from 0 to 1 , 1 to 2 , 2 to 3 , -1 to 0 , -2 to -1 , and so on. Suppose we have a Brownian motion curve in each region (Figure 2.1).

It does not matter which way I start the Brownian motion curve in each interval, so I shall start it at the beginning of each interval. Remember, Brownian motion really is symmetrical because the differences

are symmetrical whether we go from the beginning or the end of an interval. The Brownian motion curve is defined by a set of functions $x(t, \alpha_n)$. Now α_n is a variable that runs between 0 and 1. I want to show you that I can represent this entire sequence of functions in terms of one α that goes between 0 and 1, such that this single α will give us a measure distribution that will generate all of the measure distributions of all the

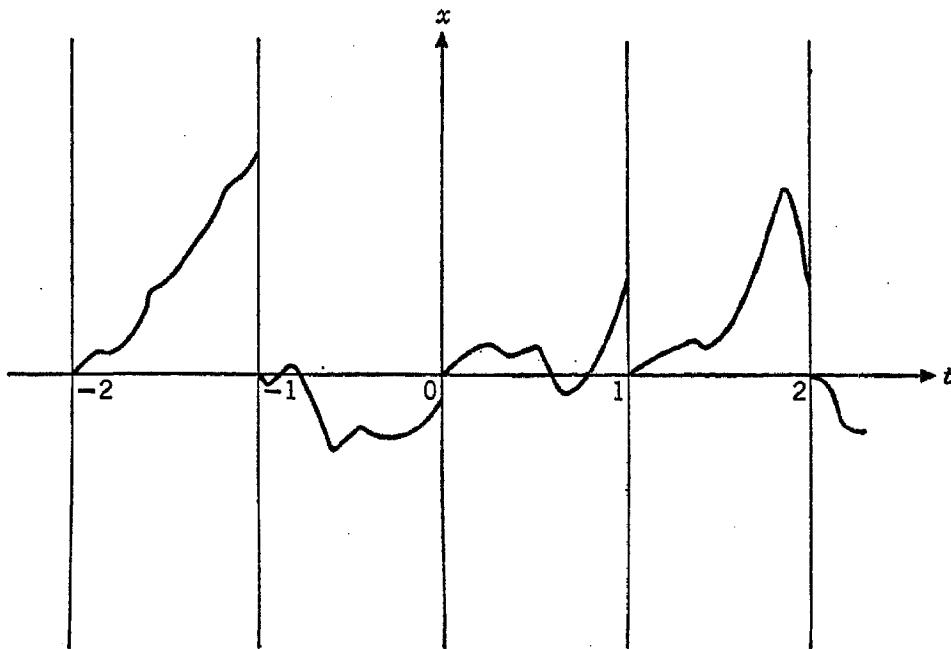


Figure 2.1

other α 's. (Here, by the way, n goes from $-\infty$ to ∞ .) I shall tell you how I get that. I take α_0 , and α_0 can be represented uniquely, except for a set of 0 measure, by a binary fraction. The binary fraction is

$$\alpha_0 = .\alpha_{01} \alpha_{02} \alpha_{03} \dots \quad (2.36)$$

and for α_1 we have

$$\alpha_1 = .\alpha_{11} \alpha_{12} \alpha_{13} \dots \quad (2.37)$$

and so on. In this way, I have a denumerable set of values arranged in a double sequence. I can rearrange this in a single sequence, and there are lots of ways in which I can do this. I can, for example, run along the dotted line

$$\begin{aligned} \alpha_0 &= \rightarrow .\alpha_{01} \alpha_{02} \alpha_{03} \alpha_{04} \dots \\ \alpha_1 &= .\alpha_{11} \alpha_{12} \alpha_{13} \alpha_{14} \dots \\ \alpha_2 &= .\alpha_{21} \alpha_{22} \alpha_{23} \alpha_{24} \dots \end{aligned} \quad (2.38)$$

to give

$$\dots \alpha_{01} \alpha_{02} \alpha_{11} \alpha_{21} \alpha_{12} \alpha_{03} \alpha_{04} \alpha_{13} \alpha_{22} \dots$$

This sequence will again give me a single binary number between 0 and 1, and this mapping is almost everywhere a one-one mapping. Furthermore, given any set of intervals in the different α 's, the measure for the new α of the set of intervals that I just sketched is simply the same as the measure of the single interval I got before. So, this is a measurement. This is a transformation that preserves measure from an infinite number of single dimensions to a single dimension. I shall call this new variable α , without a label. Then I have a sequence of expressions $x_n(t, \alpha)$, where there is a different x_n for each interval. Here I have carried out the reduction of the α_n to a single α for the range from 0 to ∞ . The process of doing so from $-\infty$ to ∞ is not essentially different.

Now I do the following thing: From Figure 2.1 I construct Figure 2.2. I just slide the curve in a region up or down till the starting point in that region is the same as the end-point of the preceding region for regions with

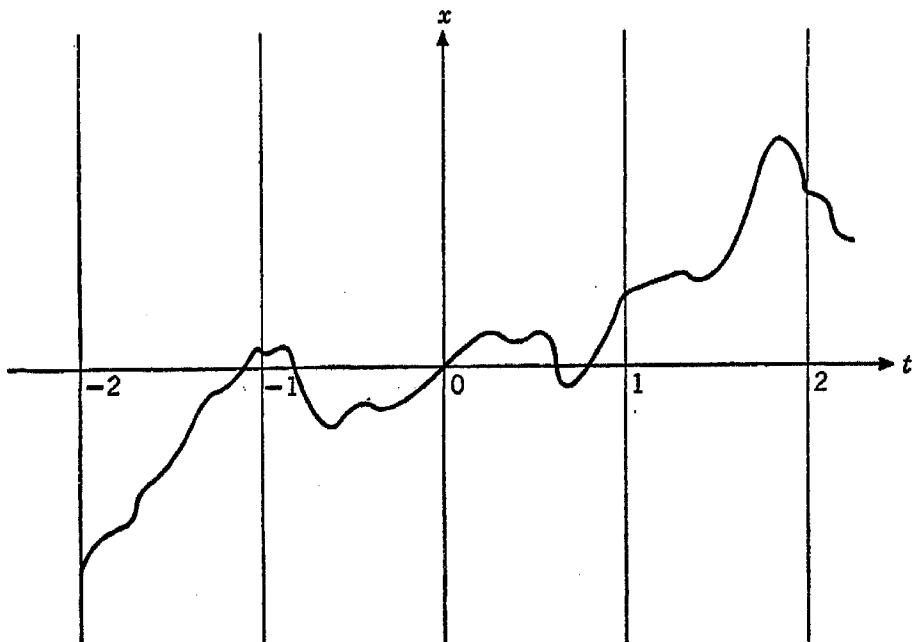


Figure 2.2

n positive, and vice versa for n negative. The first region is left unchanged. Now, when I have put all these together I call that function $X(t, \alpha)$, and $X(t, \alpha)$ is defined in terms of the old $x_n(t, \alpha)$. The properties of $X(t, \alpha)$ are similar to the properties of $x(t, \alpha)$. It begins at 0 at the origin, and the distribution of the difference in value over any interval is Gaussian. The integral of the square of the difference over an interval is proportional to the length of the interval, and the difference is independent for nonoverlapping intervals.

Suppose that we have a function $\phi(t)$ belonging to L^2 over the infinite interval. Now, $\phi(t)$ can be cut out into a denumerable set of ϕ_n . That is, we can take

$$\phi_1(t) = \begin{cases} \phi(t), & t \in [0, 1] \\ 0, & \text{otherwise} \end{cases} \quad (2.39)$$

$$\phi_2(t) = \begin{cases} \phi(t), & t \in [1, 2] \\ 0, & \text{otherwise} \end{cases} \quad (2.40)$$

and so on. I can get an infinite set. Then I form

$$\sum \int \phi_n(t) dx_n(t, \alpha) \quad (2.41)$$

and define it as equal to

$$\int_{-\infty}^{\infty} \phi(t) dX(t, \alpha) \quad (2.42)$$

There is no problem about that. The manipulation is elementary. Having done that, I have defined this for every function which is L^2 over the whole infinite interval. Now, I have the following result:

$$\int_0^1 \left[\int_{-\infty}^{\infty} \phi(t) dX(t, \alpha) \right]^2 d\alpha = \int_{-\infty}^{\infty} \phi^2(t) dt \quad (2.43)$$

where the α is the new α , and the right-hand side is the sum of the parts. Not only that; we also have that Expression 2.42 will have a Gaussian distribution. It will be the sum of a large number of terms with Gaussian distributions, so the sum of the squares of these terms converges to a Gaussian distribution. Also, the terms of this type will be independent for orthogonal ϕ 's over the infinite interval, as they were before; and everything that we have done can be carried over to the infinite interval. In other words, there is no virtue in our working from 0 to 1 with ϕ . We could just as well work from $-\infty$ to ∞ .

Now, having got this behind us, let us, either for the finite or the infinite case—it does not matter which—go over to the complex theory. By the way, in the engineering cases that I am going to deal with, the theory that I want is the real theory. In quantum theory, the theory that I want is the complex theory. Let me put

$$\phi(t) = \phi_1(t) + i \phi_2(t) \quad (2.44)$$

where ϕ_1, ϕ_2 are real and belong to L^2 . This is the most general ϕ which

belongs to L^2 in the complex case. Now let me construct

$$\frac{1}{(2)^{\frac{1}{2}}} \left\{ \left[\int_0^1 \phi_1(t) dx(t, \alpha) - \int_0^1 \phi_2(t) dx(t, \beta) \right] + i \left[\int_0^1 \phi_1(t) dx(t, \beta) + \int_0^1 \phi_2(t) dx(t, \alpha) \right] \right\} \quad (2.45)$$

When I discuss this complex expression, I am discussing nothing but the real and the imaginary parts separately. In order to get the renormalization, because I am dealing with two functions rather than one, I have put a $1/\sqrt{2}$ outside. The $1/\sqrt{2}$ is there merely because we now have two separate functions at right angles which are normalized. I shall define Expression 2.45 to be

$$\int_0^1 \phi(t) dy(t, \alpha, \beta) \quad (2.46)$$

That is the first step. With very easy computation, it follows that

$$\int d\alpha \int d\beta \left| \int \phi(t) dy(t, \alpha, \beta) \right|^2 = \int |\phi(t)|^2 dt \quad (2.47)$$

Notice that there is an absolute value in the square. That is, we add the square of the real part and the square of the imaginary part. These integrals may be over either the finite interval or the infinite interval. That is merely a matter of manipulation. Separate the parts, work with them separately, add them, and everything comes out.

This suggests that we do the following thing. We take $x(t, \alpha) + ix(t, \beta)$ —remember that that is $y(t, \alpha, \beta)$ —and we map it. I have given you the infinite mapping, and this is only a “two” mapping. We can write α as a binary fraction

$$\alpha = . \alpha_1 \alpha_2 \alpha_3 \dots \quad (2.48)$$

and also

$$\beta = . \beta_1 \beta_2 \beta_3 \dots \quad (2.49)$$

and form

$$\gamma = . \alpha_1 \beta_1 \alpha_2 \beta_2 \alpha_3 \beta_3 \dots \quad (2.50)$$

This mapping gives us the mapping of an $\alpha\beta$ square onto γ from 0 to 1. I now define $y(t, \alpha, \beta)$ as $y(t, \gamma)$, which gives me a complex Brownian motion function. You will notice that at each stage, in going from the finite to the infinite and going from the real to the complex, I have gone through a specific process, and the functions that I get are perfectly

definite in each case. Now, I get the result that

$$\int_0^1 d\gamma \left| \int \phi(t) dy(t, \gamma) \right|^2 = \int |\phi(t)|^2 dt \quad (2.51)$$

over a finite or infinite range.

Now let us consider

$$\int \phi(t) dx(t, \gamma) \quad (2.52)$$

What is the distribution of this expression? Well, it is easy to show that if we take the real and the imaginary parts of this expression, both of these parts have a Gaussian distribution—the same Gaussian distribution but independent. In other words, we have now reduced our complex Brownian motion theory, not to the real independent Gaussian distribution, but to the complex one. The fact that a unitary transformation of functions of t generates the measure-preserving transformation of γ over 0 to 1 is still true. But now we are not confined in our unitary transformations to real unitary transformations. That is our first result. In the next lecture I shall go back to the theory of orthogonal functions defined in terms of $x(t, \alpha)$.

I am going to give you an analysis of functions of α in terms of functions of varying degrees in $x(t, \alpha)$, and this is the appropriate analysis for nonlinear problems in electrical engineering and similar fields, just as harmonic analysis is the analysis for linear problems. Notice that although I have not given the details of the particular order in which I take the subdivisions of the line 0 to 1, or just which particular mapping I take, I could have made them specific in every case; and $x(t, \alpha)$ or $y(t, \gamma)$ are well-defined functions that can be used quite as readily as sine functions and cosine functions in analyzing spectra; and that is what I propose to do. Then I shall use this theory in studying the spectra of the response of nonlinear oscillators to random inputs, and I shall have that in a closed form.

Orthogonal Functions

I am going to discuss the hierarchy of orthogonal functionals. Suppose that we have a second-degree function $K(\tau_1, \tau_2)$. In order that we have sufficient hypotheses to work with (although we shall remove some of these hypotheses later), let us take a finite sum for $K(\tau_1, \tau_2)$ as

$$K(\tau_1, \tau_2) = \sum_n a_n \phi_n(\tau_1) \phi_n(\tau_2) \quad (3.1)$$

Thus we avoid all troubles of rigor. Now, I shall be working with an expression of the form

$$\iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \quad (3.2)$$

The function $K(\tau_1, \tau_2)$ is assumed to be symmetrical, by the way, although there is no restriction in considering a function of this sort to be symmetrical. If it is not symmetrical, I interchange the τ_1, τ_2 , add it to itself, and divide by 2. This does not change the final quantity in Expression 3.2. Similarly, given a functional such as

$$\iiint K(\tau_1, \tau_2, \tau_3) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \quad (3.3)$$

I can symmetrize it, and so on.

I now notice the following: The sum of any two expressions of the form of Expression 3.2 is an expression of the same sort, and similarly, the sum of any two expressions of the form of Expression 3.3 is an expression of the same sort. The functionals in Expressions 3.2 and 3.3 are homogene-

ous polynomial functionals of the Brownian motion. The homogeneous functional of zero order is K_0 . The homogeneous functional of first order is given by

$$\int K(\tau) dx(\tau, \alpha) \quad (3.4)$$

the homogeneous functional of the second order is given by Expression 3.2, and so on. I can now get nonhomogeneous functionals of the Brownian motion of any degree I desire. For example, I shall call

$$\int K_1(\tau) dx(\tau, \alpha) + K_0 \quad (3.5)$$

a nonhomogeneous functional of the first degree. These functionals will be functions of α belonging to L^2 . We have no trouble proving that, if the kernels themselves are functions in L^2 .

I now want to do the following things: First, to take the constant and normalize it; then, to take a first-degree homogeneous functional plus a constant, make it orthogonal to all constants, and normalize. Next I shall take the homogeneous expression of the second degree plus a homogeneous expression of the first degree plus a constant one. I make that orthogonal to all constants and functionals of the first degree; and so on. In this way, we get a hierarchy of functionals of different degrees, each of them orthogonal to all functionals of lower degree. This is important because it enables us to express a given function of α in terms of orthogonal functionals of different degrees.

Let us first start with the zeroth degree, K_0 , which is a constant. The mean of the square of K_0 is K_0^2 , and its absolute value is 1 (I am dealing here with reals); that is, K_0 is plus or minus 1. We then have a zeroth degree normalized functional.

Now let us consider the first-degree expression as in Expression 3.5. I want this to be orthogonal to all zero-degree functionals. But notice that changing Brownian motion into its negative does not change the distribution of Brownian motion. Therefore, it follows that if we multiply the first term of Expression 3.5 by a constant and average, the average will be zero. Then multiplying Expression 3.5 by a constant C and averaging, we get CK_0 . It must be zero if Expression 3.5 is to be orthogonal to all constants, which is the case only if $K_0 = 0$. Therefore, all first-degree homogeneous functionals are orthogonal to all zero-degree homogeneous functionals. The constant term will have to be zero to make Expression 3.5 orthogonal. I now normalize the first-degree orthogonal functional:

$$\int K_1(\tau) dx(\tau, \alpha) \quad (3.6)$$

Now

$$\int_0^1 d\alpha \left[\int K_1(\tau) dx(\tau, \alpha) \right]^2 = \int K_1^2(\tau) d\tau \quad (3.7)$$

Thus we have our category of (homogeneous) first-degree functionals, orthogonal with respect to α to all homogeneous zero-degree functionals, and these will be represented by

$$\int K_1(\tau) dx(\tau, \alpha) \quad (3.8)$$

where

$$\int K_1^2(\tau) d\tau = 1 \quad (3.9)$$

So now we have two categories: the zero-degree functionals that are normalized, and the first-degree functionals that are orthogonal to all zero-degree functionals and are normalized.

The computations now become a little more complicated. I consider a second-degree functional like

$$\iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) + \int K_1(\tau) dx(\tau, \alpha) + K_0 \quad (3.10)$$

where K_2 is symmetrical. I am assuming that K_2 can be represented as the sum of products, as in Equation 3.1. Now I want Expression 3.10 to be orthogonal to every constant, so it is enough to say that it is orthogonal to 1. To say that Expression 3.10 is orthogonal to 1 is to say that the average of Expression 3.10 multiplied by 1 is 0. This yields

$$\int K_2(\tau, \tau) d\tau + K_0 = 0 \quad (3.11)$$

We also want to make Expression 3.10 orthogonal to any expression

$$\int C(\tau) dx(\tau, \alpha) \quad (3.12)$$

If we do that, the first term in the product of Expressions 3.10 and 3.12 is third degree and has a zero average. The second term is second degree; the last term is first degree and has a zero average. We then get

$$\int K_1(\tau) C(\tau) d\tau = 0 \quad (3.13)$$

That means, since Equation 3.13 is true for any $C(\tau)$, that $K_1(\tau) = 0$. Also, from Equation 3.11, we have

$$K_0 = - \int K_2(\tau, \tau) d\tau \quad (3.14)$$

Therefore, the expression that is orthogonal to every zero-degree and first-degree expression is

$$\iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) - \int K_2(\tau, \tau) d\tau \quad (3.15)$$

We have orthogonalized Expression 3.15 to every zero-degree and first-degree expression, and we now normalize it. I square Expression 3.15 and integrate with respect to α from 0 to 1:

$$\int_0^1 d\alpha \left[\iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) - \int K_2(\tau, \tau) d\tau \right]^2 \quad (3.16)$$

This becomes

$$\begin{aligned} \int_0^1 d\alpha & \left\{ \iiint K_2(\tau_1, \tau_2) K_3(\tau_3, \tau_4) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) dx(\tau_4, \alpha) \right. \\ & - 2 \int K_2(\tau, \tau) d\tau \iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \\ & \left. + \left[\int K_2(\tau, \tau) d\tau \right]^2 \right\} \quad (3.17) \end{aligned}$$

Now remember our rule: We identify the variables by pairs and integrate. There are three ways of identifying the variables τ_1 , τ_2 , τ_3 , and τ_4 in the first term of Expression 3.17 by pairs: τ_1 and τ_2 , τ_3 and τ_4 ; τ_1 and τ_3 , τ_2 and τ_4 ; and τ_1 and τ_4 , τ_2 and τ_3 . Remember that K_2 can be, and is chosen, symmetrical. From Expression 3.17 we get

$$\begin{aligned} & \left[\int K_2(\tau, \tau) d\tau \right]^2 + 2 \iint K_2^2(\tau_1, \tau_2) d\tau_1 d\tau_2 - 2 \left[\int K_2(\tau, \tau) d\tau \right]^2 \\ & + \left[\int K_2(\tau, \tau) d\tau \right]^2 \quad (3.18) \end{aligned}$$

where the first two terms result from identification of the variables τ_1 , τ_2 , τ_3 , and τ_4 in various combinations. Summing, we get

$$2 \iint K_2^2(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (3.19)$$

So we now have the second category of orthogonal functionals. In this category we have

$$\iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) - \int K_2(\tau, \tau) d\tau \quad (3.20)$$

where

$$2 \iint K_2^2(\tau_1, \tau_2) d\tau_1 d\tau_2 = 1 \quad (3.21)$$

I shall construct the third category in order to give you the feel of this. Then I will go over to the general theory. We take

$$\begin{aligned} & \iiint K_3(\tau_1, \tau_2, \tau_3) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \\ & + \iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) + \int K_1(\tau) dx(\tau, \alpha) + K \end{aligned} \quad (3.22)$$

where K_3 is symmetrical. This is the third-degree nonhomogeneous functional of $x(t, \alpha)$. This is to be orthogonalized to a constant, to a first-degree functional, and to a second-degree functional. First consider the constant. The average of the first term of Expression 3.22 is zero because it is third degree; the second term of Expression 3.22 yields

$$\int K_2(\tau, \tau) d\tau \quad (3.23)$$

The average of the third term is zero, and the last term gives K_0 . Thus we have

$$\int K_2(\tau, \tau) d\tau + K_0 = 0 \quad (3.24)$$

Also, Expression 3.22 is supposed to be orthogonal to any expression

$$\int C(\tau) dx(\tau, \alpha) \quad (3.25)$$

So, we multiply Expression 3.22 by Expression 3.25 and average. Since K_3 is symmetrical, all of the ways of dividing τ_1, τ_2, τ_3 in pairs are alike, and we obtain

$$3 \iint K_3(\tau_1, \tau_2, \tau_3) C(\tau) d\tau d\tau_1 + \int K_1(\tau) C(\tau) d\tau = 0 \quad (3.26)$$

for all $C(\tau)$. And since $C(\tau)$ is arbitrary, the function that is orthogonal

to every function $C(\tau)$ is zero. Therefore, we obtain

$$3 \int K_3(\tau_1, \tau_1, \tau) d\tau_1 + K_1(\tau) = 0, \quad \text{for all } \tau \quad (3.27)$$

That is the necessary and sufficient condition for orthogonality with linear functionals. Next, we multiply Expression 3.22 by

$$\iint C(\sigma_1, \sigma_2) dx(\sigma_1, \alpha) dx(\sigma_2, \alpha) \quad (3.28)$$

where $C(\sigma_1, \sigma_2)$ is symmetric, and then we average. That gives us a fifth-degree expression. The first and third terms of the fifth-degree expression have zero averages. We then get

$$\begin{aligned} \int C(\sigma, \sigma) d\sigma \int K_2(\tau, \tau) d\tau + 2 \iint C(\tau_1, \tau_2) K_2(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ + K_0 \int C(\sigma, \sigma) d\sigma = 0 \end{aligned} \quad (3.29)$$

where the first term results from identifying σ 's and τ 's among themselves, and the second term results from identifying σ 's with τ 's. Then if we use Equation 3.24, we see that the first term plus the third term in Equation 3.29 is zero. Now, whatever $C(\sigma_1, \sigma_2)$ is, Equation 3.29 must be zero. Also, any symmetric function orthogonal to all symmetric functions is identically zero. Therefore, any K_2 is identically zero, and from Equation 3.27 we have

$$K_1(\tau) = -3 \int K_3(\tau_1, \tau_1, \tau) d\tau_1, \quad \text{for all } \tau \quad (3.30)$$

Using Equation 3.30 in Expression 3.22, we find our third-degree functional to be

$$\begin{aligned} \iiint K_3(\tau_1, \tau_2, \tau_3) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \\ - 3 \iint K_3(\tau_1, \tau_1, \tau) d\tau_1 dx(\tau, \alpha) \end{aligned} \quad (3.31)$$

This has been orthogonalized to every constant, linear, and quadratic functional, but it has not yet been normalized. So we want

$$\begin{aligned} 1 = \int_0^1 d\alpha \left[\iiint K_3(\tau_1, \tau_2, \tau_3) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \right. \\ \left. - 3 \iint K_3(\tau_1, \tau_1, \tau) d\tau_1 dx(\tau, \alpha) \right]^2 \end{aligned} \quad (3.32)$$

Expanding this, we have

$$\begin{aligned}
 1 = & \int_0^1 d\alpha \left[\iiint K_3(\tau_1, \tau_2, \tau_3) K_3(\tau_4, \tau_5, \tau_6) dx(\tau_1, \alpha) \right. \\
 & \times dx(\tau_2, \alpha) dx(\tau_3, \alpha) dx(\tau_4, \alpha) dx(\tau_5, \alpha) dx(\tau_6, \alpha) \\
 & - 6 \iiint K_3(\tau_1, \tau_2, \tau_3) K_3(\sigma_1, \sigma_1, \tau) d\sigma_1 dx(\tau, \alpha) \\
 & \times dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \\
 & \left. + 9 \iiint K_3(\tau_1, \tau_1, \tau_2) K_3(\tau_3, \tau_3, \tau_4) d\tau_1 d\tau_3 dx(\tau_2, \alpha) dx(\tau_4, \alpha) \right] \tag{3.33}
 \end{aligned}$$

The first term on the right-hand side of Equation 3.33 is a sixfold integral, and we can identify the variables in a number of different ways. One of the things that we can do is to identify every τ in the first parentheses with some τ in the second parentheses. There are three ways of dealing with the first; and after that, there are only two ways of picking the second, and one way for the third. So we get

$$3! \iiint K_3^2(\tau_1, \tau_2, \tau_3) d\tau_1 d\tau_2 d\tau_3 \tag{3.34}$$

for the first combination. The second combination is obtained by identifying one τ in the first parentheses of Equation 3.33 with another τ in the first parentheses. That means that we have one τ left over which will be identified with one τ in the second parentheses and there will be two τ 's left over from that. Then we have

$$\iiint K_3(\tau_1, \tau_1, \tau) K_3(\tau_2, \tau_2, \tau) d\tau_1 d\tau_2 d\tau \tag{3.35}$$

as a typical term. How many ways are there of obtaining that term? The single term in the first parentheses of Equation 3.33 can be picked out in three ways. The single term in the other parentheses can also be picked out in three ways. So we have nine ways of doing it. Thus, the first term of Equation 3.33 yields nine times Expression 3.35 plus Expression 3.34. Now we have to take

$$\begin{aligned}
 -6 \int_0^1 d\alpha \iiint K_3(\tau_1, \tau_2, \tau_3) K_3(\sigma_1, \sigma_1, \tau) d\sigma_1 dx(\tau, \alpha) dx(\tau_1, \alpha) \\
 \times dx(\tau_2, \alpha) dx(\tau_3, \alpha) \tag{3.36}
 \end{aligned}$$

which is the second term in Equation 3.32b. Now, τ will have to be identified with one of the τ_1, τ_2, τ_3 , for there is no other way to do it. That

means that the two remaining τ 's are to be identified with one another, and we get exactly the same thing as Expression 3.35 except for a scale factor of -6×3 , or -18 . There is left the last term of Equation 3.33. There we identify τ_2 and τ_4 and obtain

$$9 \int \left[\int K_3(\tau_1, \tau_1, \tau) d\tau_1 \right]^2 d\tau \quad (3.37)$$

Equation 3.33 becomes

$$\begin{aligned} & 3! \iiint K_3^2(\tau_1, \tau_2, \tau_3) d\tau_1 d\tau_2 d\tau_3 \\ & + 9 \iiint K_3(\tau_1, \tau_1, \tau) K_3(\tau_2, \tau_2, \tau) d\tau_1 d\tau_2 d\tau \\ & - 18 \iiint K_3(\tau_1, \tau_1, \tau) K_3(\tau_2, \tau_2, \tau) d\tau_1 d\tau_2 d\tau \\ & + 9 \int \left[\int K_3(\tau_1, \tau_1, \tau) d\tau_1 \right]^2 d\tau \end{aligned} \quad (3.38)$$

The last three integrals in Expression 3.38 are the same—just check out the variables. So we get

$$\begin{aligned} & \iiint K_3(\tau_1, \tau_2, \tau_3) dx(\tau_1, \alpha) dx(\tau_2, \alpha) dx(\tau_3, \alpha) \\ & - 3 \iint K_3(\tau_1, \tau_1, \tau) d\tau_1 dx(\tau, \alpha) \end{aligned} \quad (3.39)$$

where

$$3! \iiint K_3^2(\tau_1, \tau_2, \tau_3) d\tau_1 d\tau_2 d\tau_3 = 1 \quad (3.40)$$

I have given you the first three orthogonal functionals. Actually, I could construct the fourth. The fourth-degree functional will contain the fourth-order, second-order, and zeroth-order terms, and so on. In any case, the leading term (the term of highest order) will determine the other terms; and the integral of the square of the orthogonal functional, if it is not normalized, will be the integral of the square of the highest symmetric term, with respect to the three variables multiplied by $3!$, or, more generally, with respect to the n variables multiplied by $n!$.

Consider, once more, the second-degree-functional case (the third-order case will be similar to this case). Notice that this function of α , Expression 3.15, is related to a normalized symmetric function of τ_1, τ_2

in such a way that, if we divide K_2 in Expression 3.15 by $\sqrt{2}$, the integral of the square, Expression 3.16, will become the integral of the square (Equation 3.21) and will be 1. I now employ the Riesz-Fischer theorem. Suppose that we have a sequence of symmetric second-order functionals, $\{K_{2n}(\tau_1, \tau_2)\}$. Then if

$$\lim_{m, n \rightarrow \infty} \iint [K_{2m}(\tau_1, \tau_2) - K_{2n}(\tau_1, \tau_2)]^2 d\tau_1 d\tau_2 = 0 \quad (3.41)$$

then the functions K_{2n} converge in the mean to a function $K_2(\tau_1, \tau_2)$. It follows that the expressions of the form of Expression 3.2 converge in the mean in α to a limit that is independent of the sequence chosen. This means that, although I have started with rather special functions $K(\tau_1, \tau_2)$, as in Equation 3.1, which are finite sums of products of functions in L^2 , convergence in the mean of these functions will define an operator on α that will be the limit in the mean of the approximation, and I shall define this operator as

$$G_2(K_2, \alpha) = \iint K_2(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) - \int K_2(\tau, \tau) d\tau \quad (3.42)$$

Notice, then, I shall define for every function K_2 which is symmetrical, or can be made symmetrical, and which is L^2 , a functional $G_2(K_2, \alpha)$. I shall have to be a little careful here. It does not follow that I can immediately use the representation in Equation 3.42 for the following reason: The second term of Equation 3.42 may not exist even if K_2 is in the class L^2 in the two variables together. Therefore

$$\int K_2(\tau, \tau) d\tau < \infty \quad (3.43)$$

is an added requirement. However, we can still define $G_2(K_2, \alpha)$ as a well-defined function of α belonging to L^2 when K_2 is any symmetrical function of τ_1, τ_2 belonging to L^2 . Similarly, I can define $G_3(K_3, \alpha)$, $G_1(K_1, \alpha)$, and $G_0(K_0, \alpha)$; the last does not actually depend on α .

I am now going ahead of the game and assuming that I proved this for general n , which I shall in fact do later. Thus, I can define a hierarchy of functionals of α . Furthermore, it is easy to see that all of the functionals at each level are a closed set in themselves; that is, the sum of any two functionals at each level will be a limit in the mean of functions at this level because of the convergence in the mean property. In other words, we obtain a series of classes of functionals of α which are L^2 . Every functional in each class is orthogonal to every functional in every other class. If in each class we are given a definite K_0, K_1, K_2, \dots , we

have a hierarchy of functionals of α . Furthermore, I say that it can be proved that this is a closed set of functionals, that every function of α belonging to L^2 can be approximated by a polynomial functional of this set, and that, since the G 's are orthogonal, these approximations represent the projections of the function of α on successive classes or spaces—spaces of all functions of α . In other words, we can represent the function of α completely in this way. We shall go over that in the next lecture.

Then, given any function $F(\alpha)$ belonging to L^2 , $F(\alpha)$ can be represented in a unique way as

$$F(\alpha) = \lim_{N \rightarrow \infty} [G_0(K_0, \alpha) + G_1(K_1, \alpha) + \cdots + G_N(K_N, \alpha)] \quad (3.44)$$

Furthermore, previously determined terms are not changed by taking more terms in the approximation to $F(\alpha)$. This is just like ordinary orthogonal development, except that here we have an orthogonal development in whole classes of functionals—something that will give us a canonical development of any function of α that belongs to L^2 . We are going to find that this canonical development in a functional of a random function is extremely useful, and we shall go into that in more detail in the next lecture.

Now, we actually have more than the mere knowledge that we can get a development of the sort obtained in Equation 3.44. Indeed, we have a technique by which, given a function of α , we can pick out the respective components. This means picking out K_0, K_1, K_2, \dots . This is the next thing that I am going to talk about.

I want to call your attention to the fact that this development is analogous to the Fourier integral theory, which is a development of a particular sort in which we get the coefficients of the “expansion” of the function by an integration process. Here we have a development of another sort that depends upon $x(t, \alpha)$. But $x(t, \alpha)$ is a well-defined function—we have already defined it. The present development is different from the ordinary Fourier integral development in the following way. We have a denumerable set of functionals, but nevertheless, the general problem is the same: namely, given a function of α , get the development.

In the next lecture I shall show you how to develop explicitly any function of α in L^2 in a canonical development. We shall then use that development for various problems arising in connection with random functions. In particular, we shall use it for the following. Suppose that we have a function not of α alone, but of $T^t \alpha$, where $T^t \alpha$ is obtained by taking $x(\tau, \alpha)$ and changing it to $x(t + \tau, \alpha)$. We have already seen that this is a measure-preserving transformation of α . I am going to show you that when we make this transformation, each term in the development of $F(T^t \alpha)$ comes from one term in the development of $F(\alpha)$, and they do

not mix. I shall then be able to compute

$$\int_0^1 dx F(T^t \alpha) F(\alpha) \quad (3.45)$$

using the fact that orthogonality and normality in α for the G 's goes over to orthogonality and normality for each one of the G 's, with shifted arguments of the kernels in their respective set of variables. Under these conditions, we shall be able to get the autocorrelation function of nonlinear operators on Brownian motion and their spectra.

REFERENCES

- N. Wiener, *Response of a Nonlinear Device to Noise*, Report No. 129, Radiation Laboratory, M.I.T., 1942.
- R. H. Cameron and W. T. Martin, "The Orthogonal Development of Nonlinear Functionals in Series of Fourier-Hermite Functionals," *Annals of Mathematics* 48, 1947, pp. 385-389.
- N. Wiener, *Seminar in Nonlinear Networks*, Research Laboratory of Electronics February, 1949, Unpublished.
- N. Wiener, *Mathematical Problems of Communication Theory*, Summer Session Lecture Notes, 1953-54, M.I.T. (unpublished).
- A. G. Bose, *A Theory of Nonlinear Systems*, Report No. 309, Research Laboratory of Electronics, M.I.T., May, 1956.
- J. F. Barrett, *The Use of Functionals in the Analysis of Nonlinear Physical Systems*, Statistical Advisory Unit Report No. 1/57, Ministry of Supply, G.B. (no date given).
- K. O. Friedrichs, *Integration of Functionals*. Notes from a course given at New York University, 1957 (unpublished).

LECTURE 4

Orthogonal Functions

and Autocorrelation Functions

Today I am going to discuss the general theory of the development that I gave at the last lecture. We have

$$\begin{aligned}
 F_n(\alpha) = & \int \cdots \int K_n(\tau_1, \dots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \\
 & + \int \cdots \int K_{n-1}(\tau_1, \dots, \tau_{n-1}) dx(\tau_1, \alpha) \cdots dx(\tau_{n-1}, \alpha) \\
 & + \cdots + K_0
 \end{aligned} \tag{4.1}$$

I am assuming, by the way, that K_n has all smooth properties, since I am dealing with a particular case. I shall take K_n as given. The other K 's are to be determined so that $F_n(\alpha)$ is orthogonal, as a function of α , to every integral

$$\int \cdots \int Q(\tau_1, \dots, \tau_m) dx(\tau_1, \alpha) \cdots dx(\tau_m, \alpha) \tag{4.2}$$

for which $m < n$.

That being the case, I am interested in obtaining

$$\int_0^1 F_n^2(\alpha) d\alpha \tag{4.3}$$

From my assumptions, it follows at once that

$$\begin{aligned} \int_0^1 F_n^2(\alpha) d\alpha &= \int_0^1 \left[\int \cdots \int K_n(\sigma_1, \dots, \sigma_n) dx(\sigma_1, \alpha) \cdots dx(\sigma_n, \alpha) \right] \\ &\quad \times \left[\int \cdots \int K_n(\tau_1, \dots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \right. \\ &\quad \left. + \text{lower terms} \right] \end{aligned} \quad (4.4)$$

since all of the lower-order terms will be orthogonal to $F_n(\alpha)$ over $[0, 1]$. In order to perform the integration, we take all of the σ variables and all of the τ variables, make identification of variables by pairs in all possible ways, integrate, and add. Notice that if we identify two σ variables, then we get an expression of lower degree. The integral of such a term multiplied by $F_n(\alpha)$ will be 0. Therefore, the only things that are left are those identifications where I identify each σ variable with one of the τ variables.

We can identify each σ variable with a τ variable only in the first term of $F_n(\alpha)$. There are $n!$ ways of doing this. Thus, I get

$$\int_0^1 F_n^2(\alpha) d\alpha = n! \int \cdots \int K_n^2(\tau_1, \dots, \tau_n) d\tau_1 \cdots d\tau_n \quad (4.5)$$

In other words, without actually computing the expression, I have obtained the formula for the integral of the square of the n th-degree orthogonalized F , at least in the case in which K_n is the finite sum of products of L^2 functions.

I can now use the argument that I gave before. Given any limit in the mean of expressions of the form of Equation 4.1, then a necessary and sufficient condition for the limit in the mean to tend in the mean to a limit will be that the K_n 's tend in the mean to a limit. Under those conditions, we shall still get the relation of Equation 4.5.

The general function that I shall get from Equation 4.1, I shall designate as $G_n(K_n, \alpha)$, where K_n is a symmetric function of n variables. I now extend this by the limit in the mean to all K_n 's that are L^2 in the different variables. Any $G_n(K_n, \alpha)$ is orthogonal to any $G_m(K_m, \alpha)$ for which $m \neq n$.

The next question is the closure of these functions, but we have really settled that question. The set of all the G_n 's is closed, as will be shown. Do you remember how we got our α 's? We took $x(t, \alpha)$, and we defined it in terms of t 's that are terminating binary numbers. Suppose that we have any function

$$\phi[x(\tau_1, \alpha) x(\tau_2, \alpha) \cdots x(\tau_n, \alpha)] \quad (4.6)$$

where the t_n 's are binary numbers of a certain order. Remember that when I defined the curve going through a set of "holes" (Figure 1.9), I defined it in terms of the x 's only. In other words, I defined the measure of α in terms of the functions that depended on $x(t, \alpha)$ at a larger and larger binary number of points. It is easy to use that to show closure of functions of α at these points.

However, because of the closure of $x^n e^{-x^2}$, it is easy to show that ϕ can be approximated by polynomials passing through the given points. Therefore, any function of α belonging to L^2 can be approximated in the mean by polynomials depending on x 's, or differences between x 's, at a finite number of points. Since the approximating polynomials can be approximated by orthogonal polynomials, we have at least one approximation of this sort. Therefore, the G_n 's form a closed set if K_n runs through all functions that are symmetrical and L^2 in n variables.

In summary, any function of α can be approximated by a sum of orthogonal G functions. Furthermore, if we take the sum of two G 's of n th degree, we get a term of n th degree. If we take the projection of $F(\alpha)$ on all terms of lower degree than n , where n is the degree of the highest term in an approximation to F , this projection does not change as we increase n . The projection of a projection will still be the same projection.

From the preceding discussion we can come to the conclusion that, if $F(\alpha)$ belongs to L^2 , we can write, uniquely,

$$F(\alpha) = \lim_{N \rightarrow \infty} \sum_{\nu=0}^N G_\nu(K_\nu, \alpha) \quad (4.7)$$

where the best representation for each choice of N does not change the lower-order G 's but only introduces new ones.

Now comes the problem of determining the G_ν 's, which means determining the K 's. Suppose that I take

$$G_\mu(Q_\mu, \alpha) \quad (4.8)$$

where G_μ is a given $G_\mu(Q_\mu, \alpha)$. Any expression of that sort will be orthogonal to all of the terms in Equation 4.7 except the μ th, and we have

$$\int_0^1 d\alpha F(\alpha) G_\mu(Q_\mu, \alpha) = \int_0^1 G_\mu(K_\mu, \alpha) G_\mu(Q_\mu, \alpha) d\alpha \quad (4.9)$$

All other terms will vanish. We have already obtained our formula for this integral. It is

$$\int_0^1 d\alpha F(\alpha) G_\mu(Q_\mu, \alpha) = \mu! \int d\tau_1 \cdots \int d\tau_\mu K_\mu(\tau_1, \dots, \tau_\mu) Q_\mu(\tau_1, \dots, \tau_\mu) \quad (4.10)$$

In order to show this, I first work with $(Q_\mu + K_\mu)$ in Equation 4.5, then with $(Q_\mu - K_\mu)$, subtract the results, and divide by 4. Dividing by $\mu!$, I obtain

$$\begin{aligned} \int d\tau_1 \cdots \int d\tau_\mu K_\mu(\tau_1, \dots, \tau_\mu) Q_\mu(\tau_1, \dots, \tau_\mu) \\ = \frac{1}{\mu!} \int_0^1 d\alpha F(\alpha) G_\mu(Q_\mu, \alpha) \end{aligned} \quad (4.11)$$

Equation 4.11 can be used for many purposes. One of the simplest is the following.

Let us consider $Q_\mu(\alpha)$ to be a particular Q_μ , which we define as

$$Q_\mu = \begin{cases} 1, & \text{for every } \tau_k (1 \leq k \leq \mu), t_k < \tau_k < t_k + \epsilon \\ 0, & \text{elsewhere} \end{cases} \quad (4.12)$$

That is a symmetric function. If we take this particular Q_μ , Equation 4.11 reduces to

$$\frac{1}{\mu!} \int_0^1 d\alpha F(\alpha) G_\mu(Q_\mu, \alpha) = \int_{t_1}^{t_1+\epsilon} \cdots \int_{t_\mu}^{t_\mu+\epsilon} K_\mu(\tau_1, \dots, \tau_\mu) d\tau_1 \cdots d\tau_\mu \quad (4.13)$$

From Equation 4.13 we obtain

$$\frac{1}{\epsilon^\mu \mu!} \int_0^1 d\alpha F(\alpha) G_\mu(Q_\mu, \alpha) \quad (4.14)$$

as the multiple average of K_μ over the ϵ interval. If K_μ is a function belonging to L^2 , it can be proved that the average will converge in the mean to K_μ . Therefore, we have a procedure by which we can explicitly determine the K_μ 's and the expansion of $F(\alpha)$.

To summarize: We have here an expansion in terms of which, given F , we can, by integrating F times the known G 's and going to the limit in the mean, determine the different K 's and the different terms in the development. This, then, is a formulation analogous to the Fourier integral development in that, given any function belonging to L^2 , I can not only say that there is a development like Equation 4.7 but I can determine, explicitly, the coefficients by a limit in the mean of certain integrals. This is the expansion that is important for nonlinear work, just as the Fourier integral is important for linear work.

There are other particular cases that I can work out. One important case is the one in which we have

$$G_\mu(\tau_1, \dots, \tau_\mu) = \phi_1(\tau_1) \phi_2(\tau_2) \cdots \phi_\mu(\tau_\mu) \quad (4.15)$$

where the ϕ 's are a normal and orthogonal set of functions.

By the way, it is not necessary for G_μ to be symmetrical. It can always be symmetrized. Some of the ϕ 's may be the same, and some of them may be different. I can still get the development of the K 's in terms of the ϕ 's. I am going to use that later.

Now, suppose that we start with

$$F(\alpha) = \text{l.i.m. } \sum_{n=0}^{\infty} G_n(K_n, \alpha) \quad (4.16)$$

and suppose that

$$K_n = K_n(\tau_1, \dots, \tau_n) \quad (4.17)$$

Let me replace Equation 4.17 by

$$K_n = K_n(t + \tau_1, \dots, t + \tau_n) \quad (4.18)$$

I think that it is clear that this replacement is generated by a measure-preserving transformation of α . I think that it is also clear, from the way in which the different terms are formed, that each K_n in Equation 4.17 will go into the corresponding K_n in Equation 4.18 under the same transformation, and that the G_n 's will go into the G_n 's under the same transformation.

If we designate the transformation of α as T^t , then

$$F(T^t\alpha) = \sum_{n=0}^{\infty} G_n(K_n^*, \alpha) \quad (4.19)$$

where

$$K_n^*(\tau_1, \dots, \tau_n) = K_n(t + \tau_1, \dots, t + \tau_n). \quad (4.20)$$

First, I shall work with a real case, then say a word or two about the complex case. It follows that

$$\begin{aligned} & \int_0^1 F(T^t\alpha) F(\alpha) d\alpha \\ &= \sum_{n=0}^{\infty} n! \int \cdots \int K_n(\tau_1, \dots, \tau_n) K_n(t + \tau_1, \dots, t + \tau_n) \\ & \quad \times d\tau_1 \cdots d\tau_n \end{aligned} \quad (4.21)$$

If the K 's are complex, we can separate real and imaginary parts, and the final results will be:

$$\begin{aligned} & \int_0^1 F(T^t\alpha) \overline{F(\alpha)} d\alpha = \sum_{n=0}^{\infty} n! \int \cdots \int \overline{K_n(\tau_1, \dots, \tau_n)} \\ & \quad \times K_n(t + \tau_1, \dots, t + \tau_n) d\tau_1 \cdots d\tau_n \end{aligned} \quad (4.22)$$

There is no problem in proving that.

Equation 4.22 is an autocorrelation obtained by averaging on α , but we now go back to a certain ergodic argument. If we change $x(t, \alpha)$ to

$x(t + \sigma, \alpha) - x(\sigma, \alpha)$, we change the Brownian motion in the following way (remember that t runs from $-\infty$ to ∞): Instead of referring the motion to the origin, we refer it to the point P , as shown in Figure 4.1.

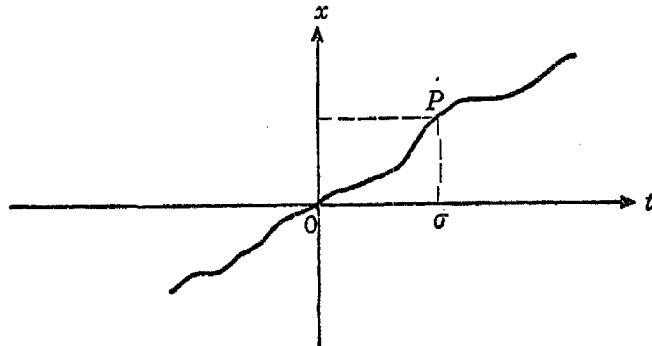


Figure 4.1

We still get a Brownian motion with the same distribution. I think I showed you the other day that, when we obtain our new α by "cutting down," the measures will be the same. This transformation can be written as

$$x(t + \sigma, \alpha) - x(\sigma, \alpha) = x(t, T^\sigma \alpha) \quad (4.23)$$

and is a measure-preserving transformation of α generated by translation. It is such a measure-preserving transformation that is a condition for what is known as metric transitivity.

I shall first state the ergodic theorem in general, then I shall apply it to this case. Suppose that we have a variable α and a measure-preserving transformation $T\alpha$. As a matter of fact, I am going to consider a group $T^t\alpha$ which is such that

$$T^{t_1}\alpha T^{t_2}\alpha = T^{t_1+t_2}\alpha \quad (4.24)$$

is a measure-preserving transformation of α . If I assume that I have a function $F(T^t\alpha)$ which belongs to L^2 and is measurable in the pair of variables t and α , then the Birkhoff ergodic theorem says that

$$\lim_{A \rightarrow \infty} \frac{1}{A} \int_0^A f(T^t\alpha) dt \text{ exists for almost all } \alpha \quad (4.25)$$

It can be proved that the set of values of α for which this integral lies between 0 and ∞ is a measurable set, and it can be proved that this set is invariant under every transformation T^t .

If the transformation T^t has the property of having no invariant measurable set of measure other than 0 or 1, then for almost all α the limit of Expression 4.25 is the same. This is called metrical transitivity. It is also known as the ergodic hypothesis. Then, under the ergodic hypothe-

sis, it follows that for any metrically transitive transformation

$$\lim_{A \rightarrow \infty} \frac{1}{A} \int_0^A f(T^t \alpha) dt = C \quad (4.26)$$

for almost all values of α .

What is C ? Let us integrate with respect to α .

$$\int_0^1 C d\alpha = C = \lim_{A \rightarrow \infty} \frac{1}{A} \int_0^A dt \int_0^1 d\alpha f(T^t \alpha) \quad (4.27)$$

Then, $f(\alpha)$ is absolutely integrable, and we can interchange orders of integration. Integrating first with respect to t , we get

$$C = \lim_{A \rightarrow \infty} \frac{1}{A} \int_0^A f(T^t \alpha) dt = \int_0^1 f(\alpha) d\alpha \quad (4.28)$$

Hence, if we have a metrically transitive transformation, the time average for almost all values of α will be the α average.

From this we deduce that, if the metrically transitive hypothesis is fulfilled,

$$\int_0^1 F(T^t \alpha) \overline{F(\alpha)} d\alpha = \lim_{A \rightarrow \infty} \frac{1}{A} \int_0^A F(T^t \alpha) \overline{F(\alpha)} dt \quad (4.29)$$

since

$$F(T^t \alpha) \overline{F(\alpha)} \quad (4.30)$$

is obviously a function which will be L^2 , being the product of two functions in L^2 . Therefore, if we can establish metric transitivity for this particular transformation, we shall have established the time autocorrelation of $F(\alpha)$ for almost all values of α .

Let us go back to the question of metric transitivity. My statement is that the transformation of α which is generated by a translation in time of the Brownian motion is, in fact, metrically transitive; that is, if a set S , of values of α , is invariant under all transformations $T^t \alpha$, its measure is either 1 or 0. I shall sketch the proof.

Consider a set S of values of α , and assume that S is measurable. There is a theorem for infinite dimensional space which states that, if we have a measurable set of values of α , it can be approximated by a finite number of intervals on the α line. A finite number of intervals can, furthermore, be approximated by a finite set of intervals corresponding to the binary subdivisions of the α line. Accordingly, given the set S , there is a set S_f , depending only upon a finite number of points, which is such that

$$m(S_f S + S_f \bar{S}) < \epsilon \quad (4.31)$$

If we take $T^t S_f$, and $|\tau|$ is large enough, then $T^t S_f$ is independent of S_f . This means that if we take the logical products $(ST^t S_f)$ or $(T^t S_f S)$ the measure of the logical product is as close to the product of the measures as we wish. Also, since $m(S)$ is invariant under all transformations T^t , $m(S_f)$ will be nearly invariant. Using these facts in Expression 4.31, we obtain, in the limit, since $m(S_f) \approx m(S)$,

$$m(S) m(T^t \bar{S}) + m(T^t S) m(\bar{S}) = 2\{m(S) - [m(S)]^2\} < \epsilon \quad (4.32)$$

for any ϵ . Therefore,

$$m^2(S) - m(S) = 0 \quad (4.33)$$

which means that the measure of S is either 0 or 1. Therefore, the translation operators are really a metrically transitive or ergodic group of transformations.

To summarize: We can use the Birkhoff ergodic theorem, and we can show that for almost all values of α (and that is all we have to do) the time autocorrelation of $F(\alpha)$ is what we have obtained by our integration formulae as the autocorrelation in phase.

We're now in a position to go ahead and obtain actual spectra of certain time series that depend nonlinearly on $x(t, \alpha)$. There are two cases of nonlinear spectra that I am going to discuss in later lectures. There are many others that this method will apply to, but I shall discuss these two cases in some detail.

One of these cases is the following:

$$\exp \left[i \int K(t - \tau) dx(\tau, \alpha) \right] \quad (4.34)$$

That is essentially what we get, physically, when we have a clock with a rotating hand, but the hand is loose and subject to Brownian perturbations. As a matter of fact, Expression 4.34 represents a linear Brownian motion of rotation that depends on the parameter x . We are going to take that up at the next lecture, when I shall discuss the spectrum of the motion described by Expression 4.34.

This is a very important thing, physically; for example, in the sort of problem that Professor Jerrold Zacharias is dealing with—that is, highly perfect clocks. He does not have a perfectly accurate time measurement. The time measurement depends on the clock, as it were. If you have an inaccurate time measurement and a highly perfect clock, how will the time inaccuracies affect the spectrum? We are going to show that the spectrum is as illustrated in Figure 4.2. Although I have drawn this spectrum centered about zero frequency, I can have it centered at any

frequency. The effect of inaccuracies is to take the spectral line and spread it into a band. Actually, the character of this band is not Gaussian, as we shall see.

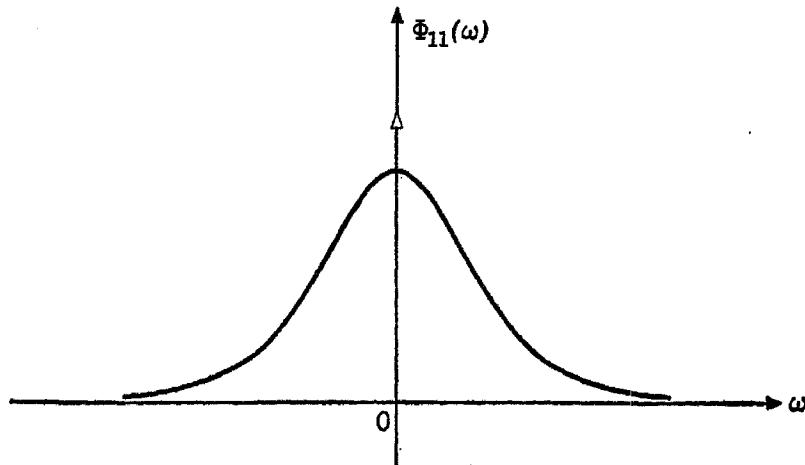


Figure 4.2

The next question is similar but one order higher. We have a Brownian motion affecting the speed of a clock in a quadratic way.

$$\exp \left[i \iint K_2(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (4.35)$$

What is the spectrum going to be? I shall tell you in general terms, and we shall compute it at a later date.

Expression 4.35 will give us a line at zero frequency in the spectrum. In addition, there will be side bands. Their profile will sometimes appear like the one shown in Figure 4.3. However, I am going to show that in

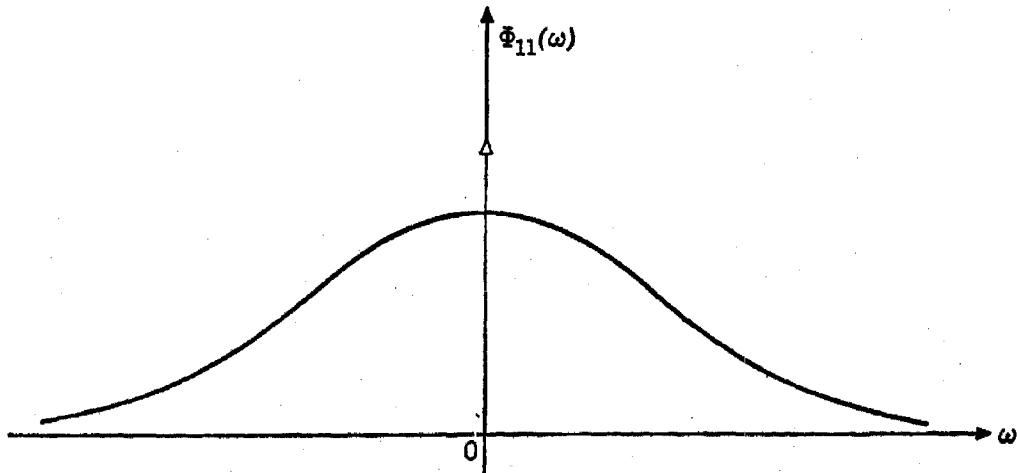


Figure 4.3

certain specified cases the profile of the side bands will be a little different. It will have a dip at the origin, as shown in Figure 4.4.

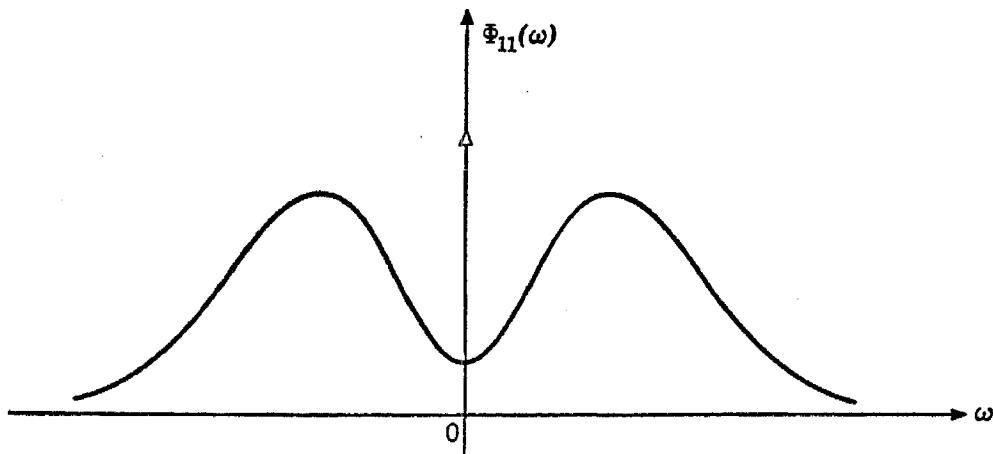


Figure 4.4

The reason that I started this work was that I ran into the case of a spectrum in which certainly must have been a random element, with side bands that dip at the center. This came about in my discussion of brain waves. As you know, the electroencephogram gives a voltage as a function of time. We can get the spectrum of such a function by a method that is familiar to us, and we can get it to a high degree of accuracy. I did this experimentally. I found this dip phenomenon occurring. I tried to think of a possible explanation of it and to get a working model of a system that would exhibit it. The thing that has led me into this work was the attempt to get a working model.

The fact is that we do get something of this sort. We are going to begin the evaluation of it at the next lecture.

Application to Frequency-Modulation Problems—I

Today, I want us to consider the function $F_n(\alpha)$ as given by Equation 5.1:

$$F_n(\alpha) = \int \cdots \int K_n(\tau_1, \dots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \quad (5.1)$$

I want to represent $F_n(\alpha)$ in what we call canonical form,

$$F_n(\alpha) = \sum_{k=0}^n G_k[L_k(\tau_1, \dots, \tau_k), \alpha] \quad (5.2)$$

where the $\{G_k\}$ are the orthogonal functionals. I want to determine the $L_k(\tau_1, \dots, \tau_k)$ explicitly. In order to do that, I use a certain lemma.

Suppose that I have any expression,

$$\sum_{k=0}^n \int \cdots \int R_k(\tau_1, \dots, \tau_k) dx(\tau_1, \alpha) \cdots dx(\tau_k, \alpha) \quad (5.3)$$

and I want to multiply this expression by

$$\int \cdots \int K_n(\sigma_1, \dots, \sigma_n) dx(\sigma_1, \alpha) \cdots dx(\sigma_n, \alpha) \quad (5.4)$$

and integrate over α from 0 to 1. If I bring the integration on $K_n(\sigma_1, \dots, \sigma_n)$ inside the summation, the expression to be evaluated

becomes:

$$\int_0^1 d\alpha \sum_{k=0}^n \int \cdots \int K_n(\sigma_1, \dots, \sigma_n) R_k(\tau_1, \dots, \tau_k) \\ \times dx(\sigma_1, \alpha) \cdots dx(\sigma_n, \alpha) dx(\tau_1, \alpha) \cdots dx(\tau_k, \alpha) \quad (5.5)$$

Remember how we integrate this term by term. We have here $(k + n)$ variables in each term. We divide them into pairs in all possible ways, identify the variables of each pair, and integrate. Since n is greater than or equal to k , it is quite clear that when we identify variables by pairs we must identify variables by pairs among the σ 's alone for a certain distance, before we begin identifying variables between the σ 's and the τ 's, because there are more σ 's than τ 's. I think it is clear that when I do this, and integrate over only the σ 's that are paired with themselves, I shall obtain in each term of the sum an expression,

$$\int \cdots \int K_n(\sigma_1, \dots, \sigma_k, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, \lambda_\nu, \lambda_\nu) d\lambda_1 \cdots d\lambda_\nu \quad (5.6)$$

where $n = k + 2\nu$. Then I have to begin identifying the σ 's with the τ 's and carry on the rest of the integrations. It is quite clear that I shall get exactly the same results in this two-step integration as I should if I had integrated over all the pairings of σ 's and τ 's in one step. How many times does Expression 5.6 enter into each term of the sum? Well, in the first place, how many ways are there for choosing (2ν) variables out of n variables? I think it is clear that there are $\binom{n}{2\nu}$ ways, where $\binom{n}{2\nu}$ is the binomial coefficient,

$$\binom{n}{2\nu} = \frac{n!}{(2\nu)!(n - 2\nu)!} \quad (5.7)$$

In addition to choosing the 2ν variables, I have to identify them in pairs. This gives me $(2\nu - 1)(2\nu - 3) \cdots (1)$ additional terms. Hence, in the integration of Expression 5.5, in each term I can replace

$$\int \cdots \int K_n(\sigma_1, \dots, \sigma_n) dx(\sigma_1, \alpha) \cdots dx(\sigma_n, \alpha) \quad (5.8)$$

by

$$\int \cdots \int \binom{n}{2\nu} (2\nu - 1)(2\nu - 3) \cdots (1) \int \cdots \int K_n(\sigma_1, \dots, \sigma_k, \\ \lambda_1, \lambda_1, \dots, \lambda_\nu, \lambda_\nu) d\lambda_1 \cdots d\lambda_\nu dx(\sigma_1, \alpha) \cdots dx(\sigma_k, \alpha) \quad (5.9)$$

I can next simplify the coefficient as follows:

$$\begin{aligned} \binom{n}{2\nu} (2\nu - 1)(2\nu - 3) \cdots (1) &= \frac{n!}{(2\nu)!(n - 2\nu)!} \frac{(2\nu)!}{(2\nu)(2\nu - 2) \cdots (2)} \\ &= \frac{n!}{(n - 2\nu)!} \frac{1}{2^\nu \nu!} \end{aligned} \quad (5.10)$$

From the results of this lemma, it follows at once that if Expression 5.3 represents a G -function, I can apply the lemma to the following integration:

$$\int_0^1 d\alpha \int \cdots \int K_n(\tau_1, \dots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \times G_k[M_k(\sigma_1, \dots, \sigma_k), \alpha] \quad (5.11)$$

We shall get exactly what we should get if we replace K_n in Expression 5.11 by Expression 5.9. I know that this will be

$$\begin{aligned} K! \int \cdots \int \frac{n!}{(n - 2\nu)! 2^\nu \nu!} K_n(\tau_1, \dots, \tau_k, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, \lambda_\nu, \lambda_\nu) \\ \times d\lambda_1 \cdots d\lambda_\nu M_k(\tau_1, \dots, \tau_k) d\tau_1 \cdots d\tau_k \end{aligned} \quad (5.12)$$

Remember, Expression 5.11 is equal to Expression 5.12 only when k is of the same parity as n . When k is of different parity from n , Expression 5.11 is identically 0.

From this, it is easy to come to the conclusion that, for the L_k of Equation 5.2, the L_k 's in which k is of different parity from n do not come into the expression; that is, $L_{n-2\nu-1}$ is identically 0 in all cases; and the $L_{n-2\nu}$ are

$$\begin{aligned} L_{n-2\nu}(\tau_1, \dots, \tau_{n-2\nu}) &= \frac{n!}{(n - 2\nu)! 2^\nu \nu!} \\ &\times \int \cdots \int K_n(\tau_1, \dots, \tau_{n-2\nu}, \sigma_1, \sigma_1, \dots, \sigma_\nu, \sigma_\nu) d\sigma_1 \cdots d\sigma_\nu \end{aligned} \quad (5.13)$$

Equations 5.1, 5.2, and 5.13 can be combined and written as

$$\begin{aligned} &\int \cdots \int K_n(\tau_1, \dots, \tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \\ &= \sum_{\nu=0}^{\left[\frac{n}{2}\right]} \frac{n!}{(n - 2\nu)! 2^\nu \nu!} G_{n-2\nu} \left[\int \cdots \int K_n(\tau_1, \dots, \tau_{n-2\nu}, \sigma_1, \sigma_1, \dots, \sigma_\nu, \sigma_\nu) \right. \\ &\quad \left. \times d\sigma_1 \cdots d\sigma_\nu, \alpha \right] \end{aligned} \quad (5.14)$$

where $[n/2] = n/2$ if n is even and equals $(n - 1)/2$ if n is odd. From this obtained I have the homogeneous polynomial development in terms of the orthogonal functionals that belong to it. These orthogonal functionals are all of the same parity as the K_n , and, for example, in the case of G_ν , are obtained by taking the variables of K_n beyond ν , identifying by pairs, and integrating.

Now, I shall go directly to the use of this development for the study of frequency modulation (FM). I want to study the spectrum of an expression like

$$\exp \left[i \int f(t + \tau) dx(\tau, \alpha) \right] \quad (5.15)$$

Notice, of course, that we shall get two terms here, the cosine and the sine, but they can be discussed in terms of Expression 5.15. Thus, we have a message that is the response of a linear resonator to a Brownian input, and we are looking for its spectrum. To get the spectrum, we have seen that it is enough to express this sort of thing in terms of the fundamental orthogonal functionals. Now, I am going to make this a little more general. We shall introduce a complex number a into the exponent, which, for convenience, will include the factor i ; f can now be considered normalized. Hence, Expression 5.15 becomes

$$\exp \left[a \int \phi(t + \tau) dx(\tau, \alpha) \right] \quad (5.16)$$

where

$$\int \phi^2(\tau) d\tau = 1 \quad (5.17)$$

I want the development of Expression 5.16 in terms of orthogonal functionals. That is easy. The expression can be represented as a sum of homogeneous polynomial functionals by simply using the exponential development

$$\begin{aligned} & \exp \left[a \int \phi(\tau) dx(\tau, \alpha) \right] \\ &= \sum_{n=0}^{\infty} \frac{a^n}{n!} \int \cdots \int \phi(\tau_1) \cdots \phi(\tau_n) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha) \end{aligned} \quad (5.18)$$

where I have let ϕ be a function of τ alone. Later on, I shall reintroduce t . Now, I use Equation 5.14 for the expression in terms of the orthogonal

functional of each term of the sum of Equation 5.18, and I obtain

$$\begin{aligned} & \exp \left[a \int \phi(\tau) dx(\tau, \alpha) \right] \\ &= \sum_{n=0}^{\infty} a^n \sum_{\nu=0}^{\left[\frac{n}{2}\right]} \frac{1}{(n-2\nu)! 2^\nu \nu!} G_{n-2\nu}[\phi(\tau_1) \cdots \phi(\tau_{n-2\nu}), \alpha] \quad (5.19) \end{aligned}$$

where we have paired ϕ 's, integrated, and applied Equation 5.17. Rearranging Equation 5.19, and letting $\mu = n - 2\nu$, we obtain,

$$\sum_{\mu=0}^{\infty} \frac{1}{\mu!} G_\mu[\phi(\tau_1) \cdots \phi(\tau_\mu), \alpha] \sum_{\nu=0}^{\infty} \frac{a^{\mu+2\nu}}{2^\nu \nu!} \quad (5.20)$$

Expression 5.20 can be simplified a great deal. Summing on ν , it becomes

$$\exp \left[a \int \phi(\tau) dx(\tau, \alpha) \right] = \sum_{\mu=0}^{\infty} \frac{a^\mu}{\mu!} \exp \left(\frac{a^2}{2} \right) G_\mu[\phi(\tau_1) \cdots \phi(\tau_\mu), \alpha] \quad (5.21)$$

I have now given, for the FM case, the development in orthogonal polynomial functionals. I am going to do two things with this development. I am going to apply it immediately to the FM spectrum, and I am going to use it as a general tool for handling more complicated cases. First, replace a by ai . We obtain the following function of α :

$$\begin{aligned} f(\alpha) &= \exp \left[ai \int \phi(\tau) dx(\tau, \alpha) \right] \\ &= \sum_{\mu=0}^{\infty} \frac{(ai)^\mu}{\mu!} \exp \left(-\frac{a^2}{2} \right) G_\mu[\phi(\tau_1) \cdots \phi(\tau_\mu), \alpha] \quad (5.22) \end{aligned}$$

If we now replace τ by $t + \tau$, we obtain a function of both t and α :

$$\begin{aligned} f(t, \alpha) &= \exp \left[ai \int \phi(t + \tau) dx(\tau, \alpha) \right] \\ &= \sum_{\mu=0}^{\infty} \frac{(ai)^\mu}{\mu!} \exp \left(-\frac{a^2}{2} \right) G_\mu[\phi(t + \tau_1) \cdots \phi(t + \tau_\mu), \alpha] \quad (5.23) \end{aligned}$$

Next, we perform the following integration, using the formula that we

have for the integral of the product of two G 's:

$$\begin{aligned} \int_0^1 f(t+s, \alpha) \overline{f(s, \alpha)} d\alpha &= \sum_{\mu=0}^{\infty} \frac{\alpha^{2\mu}}{(\mu!)^2} \exp(-\alpha^2) \mu! \left[\int \phi(\tau) \phi(t+\tau) d\tau \right]^{\mu} \\ &= \exp(-\alpha^2) \exp \left[\alpha^2 \int \phi(t+\tau) \phi(\tau) d\tau \right] \\ &= \exp \left\{ \alpha^2 \left[\int \phi(t+\tau) \phi(\tau) d\tau - 1 \right] \right\} \end{aligned} \quad (5.24)$$

where $\overline{f(t, \alpha)}$ is the conjugate of $f(t, \alpha)$. Equation 5.24 is, by the way, the autocorrelation taken on the α -scale, and we have seen, by the ergodic theorem, that this is almost always the autocorrelation taken on the t -scale. Hence, Equation 5.24 gives the autocorrelation explicitly in one of the simple FM cases.

We now want to make a harmonic analysis of this FM case. When we expand Equation 5.24, we obtain a series of powers of the autocorrelation of ϕ . If we take the Fourier transform, we obtain the power spectrum as a sum of terms that involve the spectrum of ϕ and its respective convolutions. Letting $h(t)$ represent the autocorrelation of ϕ and $H(\omega)$ its power spectrum,

$$h(t) = \int \phi(t+\tau) \phi(\tau) d\tau \quad (5.25)$$

and

$$H(\omega) = \frac{1}{(2\pi)^{1/2}} \int h(t) e^{i\omega t} dt \quad (5.26)$$

I take repetitions of $H(\omega)$; $H^1(\omega), \dots, H^n(\omega)$, where

$$H^1(\omega) = H(\omega) \quad (5.27)$$

and

$$H^n(\omega) = \int_{-\infty}^{\infty} H^{n-1}(\omega + \mu) H(\mu) d\mu \quad (5.28)$$

Then I obtain the spectrum as a series in the repeated $H(\omega)$'s. This gives us the explicit form for the spectrum of frequency modulation.

I think that the next thing for me to do is to state in ordinary language what I have done. I have treated a message that is the response of a linear resonator to a Brownian input. (The actual messages that I send are not like that, but their distribution will not be too far from that.) Suppose that I could send this message by frequency modulation. What would be the spectrum of the transmission? This is the problem that I have solved, and it is a large part of the FM problem. I can go much further than this. Suppose that I have a known message and to it I add a random error distributed in frequency. I want to determine how the

$$= \int H^{n-1}(\omega - \mu) H(\mu) d\mu$$

random error in the accuracy of the message will affect the frequency modulation. This problem can be solved by similar methods. I am going to leave this now, but I can use the results already obtained as a tool.

Now I want to discuss a more complicated case:

$$\exp \left[i \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (5.29)$$

We may call this case “quadratic FM.” The first thing to do is to realize that, in a large number of cases, $K(\tau_1, \tau_2)$, a symmetric function, can be written as

$$K(\tau_1, \tau_2) = \sum a_n \phi_n(\tau_1) \phi_n(\tau_2) \quad (5.30)$$

where the ϕ_n 's are the characteristic functions of the kernel, $K(\tau_1, \tau_2)$, and the a_n 's are the characteristic numbers. This is the bilinear formula. In these cases, the ϕ_n 's are real and orthogonal, and the a_n 's, under certain restrictions that I am going to assume to begin with, are real and positive. Then, Expression 5.29 becomes, at least formally,

$$\begin{aligned} & \exp \left[i \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ &= \prod \exp \left\{ ia_n \left[\int \phi_n(t + \tau) dx(\tau, \alpha) \right]^2 \right\} \end{aligned} \quad (5.31)$$

The first case to handle, then, is simply

$$\exp \left\{ ia_n \left[\int \phi_n(t + \tau) dx(\tau, \alpha) \right]^2 \right\} \quad (5.32)$$

I shall obtain this spectrum by means of the method used in the linear FM case. Then we multiply Functions 5.32. When we multiply, note one very important thing. If we have K 's composed of different ϕ 's, the product of orthogonal functionals of our hierarchy also belongs to our hierarchy.

Let us notice that

$$\sum_{\mu=0}^{\infty} u^\mu H_\mu(u) = \exp \left(-\frac{u^2}{2} + u \right) \quad (5.33)$$

where the $H_\mu(u)$ are the Hermite polynomials. With Equation 5.21, this establishes the identity

$$H_\mu \left[\int \phi(\tau) dx(\tau, \alpha) \right] = G_\mu \left[[\phi(\tau_1) \cdots \phi(\tau_\mu), \alpha] \right] \quad (5.34)$$

Application to Frequency-Modulation Problems—II

In the previous lecture I proved that

$$\exp \left[a \int \phi(t) dx(t, \alpha) \right] = \sum_{\nu=0}^{\infty} \frac{a^\nu}{\nu!} \exp \left(\frac{a^2}{2} \right) G_\nu[\phi(\tau_1), \dots, \phi(\tau_\nu), \alpha] \quad (6.1)$$

where

$$\int |\phi^2(t)| dt = 1 \quad (6.2)$$

and

$$\operatorname{Re}(a) < 1 \quad (6.3)$$

This time I want to study

$$\exp \left\{ b \left[\int \phi(t) dx(t, \alpha) \right]^2 \right\} \quad (6.4)$$

which equals

$$\sum_{n=0}^{\infty} \frac{b^n}{n!} \int \phi(\tau_1) dx(\tau_1, \alpha) \cdots \int \phi(\tau_{2n}) dx(\tau_{2n}, \alpha) \quad (6.5)$$

Then we see that, when we go to the G -functions, Expression 6.5 equals

$$\sum_{n=0}^{\infty} C_{2n} G_{2n}[\phi(\tau_1), \dots, \phi(\tau_{2n}), \alpha] \quad (6.6)$$

The problem here is the determination of the C_{2n} 's. This could be done directly, but I am going to use a method of generating functions. Multiplying Equation 6.1 by Expression 6.4 on the left-hand side and by

Expression 6.6 on the right, and integrating over α , we obtain

$$\begin{aligned} \int_0^1 d\alpha \exp \left[a \int \phi(t) dx(t, \alpha) \right] \exp \left\{ b \left[\int \phi(t) dx(t, \alpha) \right]^2 \right\} \\ = \sum_{n=0}^{\infty} a^{2n} \exp \left(\frac{a^2}{2} \right) C_{2n} \quad (6.7) \end{aligned}$$

The right-hand side of Equation 6.7 follows because

$$\begin{aligned} \int_0^1 d\alpha G_\nu[\phi(\tau_1), \dots, \phi(\tau_\nu), \alpha] G_{2n}[\phi(\tau_1), \dots, \phi(\tau_n), \alpha] \\ = \begin{cases} \nu!, & \text{for } 2n = \nu \\ 0, & \text{otherwise} \end{cases} \quad (6.8) \end{aligned}$$

Now let

$$u = \int \phi(t) dx(t, \alpha) \quad (6.9)$$

and remember that u has a Gaussian distribution. Also,

$$\int_0^1 u d\alpha = 0 \quad (6.10)$$

and

$$\int_0^1 u^2 d\alpha = 1 \quad (6.11)$$

Therefore

$$\begin{aligned} \int_0^1 d\alpha \exp \left[a \int \phi(t) dx(t, \alpha) \right] \exp \left\{ b \left[\int \phi(t) dx(t, \alpha) \right]^2 \right\} \\ = \int_0^1 d\alpha \exp(au) \exp(bu^2) \\ = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp \left(-\frac{u^2}{2} \right) \exp(au + bu^2) du \quad (6.12) \end{aligned}$$

where we have made use of our knowledge of the distribution of u and have obtained an expression that we can integrate. Let

$$v = u(1 - 2b)^{\frac{1}{2}} \quad (6.13)$$

where we must work with $b < \frac{1}{2}$. This is no difficulty, since we want to work with small values of b . Notice that if b is imaginary we can work with large values of b . Now, Equation 6.12 becomes

$$(2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp \left[-\frac{v^2}{2} + \frac{av}{(1 - 2b)^{\frac{1}{2}}} \right] \frac{dv}{(1 - 2b)^{\frac{1}{2}}} \quad (6.14)$$

and completing the square, we obtain

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp \left[-\frac{v - \frac{a}{(1-2b)^{\frac{1}{2}}}}{2} \right] \exp \left[\frac{a^2}{2(1-2b)} \right] \frac{dv}{(1-2b)^{\frac{1}{2}}} \quad (6.15)$$

Now we have

$$(2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp \left[-\frac{v - \frac{a}{(1-2b)^{\frac{1}{2}}}}{2} \right] dv = 1 \quad (6.16)$$

because even in a complex plane, moving v up or down as well as right or left does not change that integral. In order to show this, we use the Cauchy theorem. We are integrating an expression along a line, and it makes no difference that we integrate it along a parallel line because the integrand goes down to 0 very rapidly as we go to ∞ . Also, there are no singularities inside. So, we use the Cauchy theorem, and the integral is 1. Hence, Expression 6.15 equals

$$\frac{1}{(1-2b)^{\frac{1}{2}}} \exp \left[\frac{a^2}{2(1-2b)} \right] \quad (6.17)$$

Substituting Expression 6.17 in Equation 6.7, we obtain

$$\begin{aligned} \sum_{n=0}^{\infty} a^{2n} C_{2n} &= \left[\frac{1}{(1-2b)^{\frac{1}{2}}} \right] \exp \left(-\frac{a^2}{2} \right) \exp \left[\frac{a^2}{2(1-2b)} \right] \\ &= \frac{1}{(1-2b)^{\frac{1}{2}}} \exp \left(\frac{ba^2}{1-2b} \right) \end{aligned} \quad (6.18)$$

Remember, by the way, that

$$C_{2n+1} = 0 \quad (6.19)$$

Expanding the right-hand side of Equation 6.18, we have

$$\sum_{n=0}^{\infty} a^{2n} C_{2n} = \frac{1}{(1-2b)^{\frac{1}{2}}} \sum_{n=0}^{\infty} \frac{a^{2n}}{n!} \left(\frac{b}{1-2b} \right)^n \quad (6.20)$$

Equating coefficients of a^{2n} gives

$$C_{2n} = \frac{1}{(1-2b)^{\frac{1}{2}}} \left(\frac{b}{1-2b} \right)^n \frac{1}{n!} \quad (6.21)$$

See how we have saved the trouble of adding up a lot of series by using the generating function.

We can now write

$$\begin{aligned} \exp \left\{ b \left[\int \phi(t) dx(t_1, \alpha) \right]^2 \right\} \\ = \sum_{n=0}^{\infty} \frac{b^n}{n!} \left(\frac{1}{1 - 2b} \right)^{n+\frac{1}{2}} G_{2n}[\phi(\tau_1), \dots, \phi(\tau_{2n}), \alpha] \quad (6.22) \end{aligned}$$

Next, I shall replace b by ib , and there is nothing in the previous work that prevents this. The imaginary case is even more advantageous than the real, because of the fact that the denominator of Equation 6.21 becomes infinite as b becomes infinite, and now we shall not have to worry about coming to a place where the series will fail to converge. So, we have

$$\begin{aligned} \exp \left\{ ib \left[\int \phi(t) dx(t, \alpha) \right]^2 \right\} \\ = \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} \left(\frac{1}{1 - 2ib} \right)^{n+\frac{1}{2}} G_{2n}[\phi(\tau_1), \dots, \phi(\tau_{2n}), \alpha] \quad (6.23) \end{aligned}$$

Let us consider

$$\exp \left[ib \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (6.24)$$

where $K(\tau_1, \tau_2)$ is symmetric. For certain cases (cases that represent a generalization of the ordinary Fredholm situation), we shall find

$$K(\tau_1, \tau_2) = \sum_{\nu=0}^{\infty} b_{\nu} \phi_{\nu}(\tau_1) \phi_{\nu}(\tau_2) \quad (6.25)$$

where the $\phi_{\nu}(\tau)$ are a set of orthogonal real functions called "characteristic functions," and the b_{ν} 's are the "characteristic numbers" of the problem. Then Expression 6.24 becomes

$$\prod_{\nu} \sum_{n=0}^{\infty} \frac{(ib_{\nu})^n}{n!} \left(\frac{1}{1 - 2ib_{\nu}} \right)^{n+\frac{1}{2}} G_{2n}[\phi_{\nu}(\tau_1), \dots, \phi_{\nu}(\tau_{2n}), \alpha] \quad (6.26)$$

by substitution of Equation 6.25 into Expression 6.24 and use of Equation 6.23.

Next, we want to rearrange this series in a series of orthogonal functionals of different orders, and we go to certain properties of the G 's that I have not taken up yet. I want to say that

$$G_n[\phi_j(\tau_1), \dots, \phi_j(\tau_n), \alpha] G_m[\phi_k(\tau_{n+1}), \dots, \phi_k(\tau_{n+m}), \alpha] \quad (6.27)$$

is a polynomial of degree $(n + m)$ that is orthogonal to all polynomials of lower degree. This is clear because of the independence—any function in one set of variables is independent of any function in another set of vari-

ables. Also, we know that an expression of lower degree will have at least one part of lower degree than the corresponding part of Expression 6.27. Hence, when we integrate with respect to α , we get 0. This tells us at once how we can rearrange Expression 6.26 in orthogonal functionals of lower degrees. Remember that the ϕ_ν 's are orthogonal to one another. There is an even easier way of doing this.

When we multiply two orthogonal polynomials together, we get exactly the same result as though we had multiplied their leading terms together and completed the expression. The leading term will be the leading term, so we consider that there is related to Expression 6.26 a function consisting of the leading terms of Expression 6.26—a similar series. We multiply the series of leading terms together, and then we get the terms with which we are going to operate to get the G 's of the desired expansion. The expression

$$\prod_\nu \sum_{n=0}^{\infty} \frac{(ib_\nu)^n}{n!} \left(\frac{1}{1 - 2ib_\nu} \right)^{n+\frac{1}{2}} \left[\int \phi_\nu(\tau) dx(\tau, \alpha) \right]^{2n} \quad (6.28)$$

corresponds to the leading terms of Expression 6.26. When we multiply out, the terms that we get will be the leading terms of the G 's in the desired expansion. Let us explore what this is. Expression 6.28 is

$$\prod_\nu \frac{1}{(1 - 2ib_\nu)^{\frac{1}{2}}} \exp \left\{ \left[\int \phi_\nu(\tau) dx(\tau, \alpha) \right]^2 \frac{ib_\nu}{1 - 2ib_\nu} \right\} \quad (6.29)$$

which represents the generating operator. Now I am going to expand this operator and get a sum of homogeneous operators of the different orders. Then I shall replace these homogeneous operators by the G 's of the same leading terms and obtain the desired series. But this leads to a discussion of the functions

$$\sum_\nu \frac{ib_\nu}{1 - 2ib_\nu} \left[\int \phi_\nu(\tau) dx(\tau, \alpha) \right]^2 \quad (6.30)$$

This expression can be written as a homogeneous, second-degree expression:

$$i \iint R(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \quad (6.31)$$

just as we wrote Equation 6.25. Then Expression 6.29 becomes

$$\prod_\nu \frac{1}{(1 - 2ib_\nu)^{\frac{1}{2}}} \exp \left[i \iint R(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (6.32)$$

and this is the function that gives the desired leading coefficients.

The next thing is to investigate

$$\prod_{\nu} \frac{1}{(1 - 2ib_{\nu})^{\frac{1}{2}}} \quad (6.33)$$

At least formally, this is

$$\exp \left[-\frac{1}{2} \sum_{\nu} \ln (1 - 2ib_{\nu}) \right] \quad (6.34)$$

To begin with, take

$$|b_{\nu}| < \frac{1}{2} \quad (6.35)$$

Then

$$\frac{1}{2} \sum_{\nu} \ln (1 - 2ib_{\nu}) = \frac{1}{2} \sum_{\nu} \left[(2ib_{\nu}) + \frac{(2ib_{\nu})^2}{2} + \frac{(2ib_{\nu})^3}{3} + \dots \right] \quad (6.36)$$

so

$$\frac{1}{2} \sum_{\nu} \ln (1 - 2ib_{\nu}) = \frac{1}{2} \left[2i \sum_{\nu} b_{\nu} + \frac{(2i)^2}{2} \sum_{\nu} b_{\nu}^2 + \frac{(2i)^3}{3} \sum_{\nu} b_{\nu}^3 + \dots \right] \quad (6.37)$$

Now, remember that

$$K(\tau_1, \tau_2) = \sum_{\nu} b_{\nu} \phi_{\nu}(\tau_1) \phi_{\nu}(\tau_2) \quad (6.38)$$

where the ϕ_{ν} 's are a normal and orthogonal set, and the b_{ν} 's real and positive. Then, at least formally, we have

$$\sum_{\nu} b_{\nu} = \int K(\tau, \tau) d\tau \quad (6.39)$$

$$\sum_{\nu} b_{\nu}^2 = \iint K^2(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (6.40)$$

$$\sum_{\nu} b_{\nu}^3 = \iiint K(\tau_1, \tau_2) K(\tau_2, \tau_3) K(\tau_3, \tau_1) d\tau_1 d\tau_2 d\tau_3 \quad (6.41)$$

and so on. Therefore, Equation 6.37 can be expressed in terms of the K 's, at least when the b_{ν} 's are small enough; and there is no problem if the b_{ν} 's are bigger, for the absolute value of

$$\frac{1}{(1 - 2ib_{\nu})^{\frac{1}{2}}} \quad (6.42)$$

is

$$\frac{1}{(1 + 4b_{\nu}^2)^{\frac{1}{4}}} \quad (6.43)$$

and the bigger the b_n 's, the better the convergence. Hence, Expression 6.33 is a parameter depending on the K , and this parameter is not going to be of serious importance to us in discussing the spectrum. It will merely give a constant factor for the autocorrelation and, therefore, a constant factor for the spectrum that will be the same for all terms. Therefore, we can confine our attention to

$$\begin{aligned} & \exp \left[i \iint R(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ &= 1 + i \iint R(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \\ & - \frac{1}{2!} \iiint \iint R(\tau_1, \tau_2) R(\tau_3, \tau_4) dx(\tau_1, \alpha) \cdots dx(\tau_4, \alpha) + \cdots \quad (6.44) \end{aligned}$$

These integrals are homogeneous functionals and can be reduced to the form in which the kernels are symmetrical simply by taking all of the permutations and adding. Having done this, we can go and build the corresponding G 's by the prescription given previously, and we have the development of the expression

$$\exp \left[i \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (6.45)$$

in terms of the orthogonal polynomial.

You see how I save trouble by using generating functions. Now I did not have to use them; I could have added up the series directly. But why waste time when there is an easy way?

LECTURE 7

Application to Frequency-Modulation Problems—III

In the previous lecture we established that

$$\begin{aligned} \exp \left\{ a \left[\int \phi(t) dx(t, \alpha) \right]^2 \right\} \\ = \frac{1}{(1 - 2a)^{\frac{1}{2}}} \sum_{n=0}^{\infty} \left(\frac{a}{1 - 2a} \right)^n G_{2n}[\phi(\tau_1), \dots, \phi(\tau_{2n}), \alpha] \quad (7.1) \end{aligned}$$

where a is either real or imaginary. If I consider

$$K(\tau_1, \tau_2) = \sum_{\nu} a_{\nu} \phi_{\nu}(\tau_1) \phi_{\nu}(\tau_2) \quad (7.2)$$

where the ϕ 's are orthogonal, it follows that

$$\begin{aligned} \exp \left[\iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ = \prod_{\nu} \frac{1}{(1 - 2a_{\nu})^{\frac{1}{2}}} \sum_{n=0}^{\infty} \left(\frac{a_{\nu}}{1 - 2a_{\nu}} \right)^n G_{2n}[\phi_{\nu}(\tau_1), \dots, \phi_{\nu}(\tau_{2n}), \alpha] \quad (7.3) \end{aligned}$$

Remember that if we are dealing with orthogonal functions, ϕ , the product of Hermitian polynomials is a Hermitian polynomial of the

degree of the sum. Therefore,

$$\begin{aligned} & \exp \left[\iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ &= \left(\prod_{\nu} \frac{1}{(1 - 2a_{\nu})^{\frac{1}{2}}} \right) \sum_{n=0}^{\infty} \sum_{k_i} G_{2n} [\phi_{k_1}(\tau_1), \phi_{k_1}(\tau_2), \phi_{k_2}(\tau_3), \phi_{k_2}(\tau_4), \dots, \right. \\ & \quad \left. \phi_{k_n}(\tau_{2n-1}), \phi_{k_n}(\tau_{2n}), \alpha] \prod_{k_i} \frac{a_{k_i}}{1 - 2a_{k_i}} \quad (7.4) \end{aligned}$$

in which the terms in the sum over k_i are obtained by making identifications of k_i 's with ν 's in all possible ways, and the product over k_i is over the same set of $2nk_i$'s used in the argument of G_{2n} .

There is one bit of notation I want to introduce. I have given you the G_n for symmetric functions. The G_n for unsymmetric functions, that is

$$G_n[F(\tau_1, \dots, \tau_n), \alpha] \quad (7.5)$$

where F is not a symmetric function of τ_1, \dots, τ_n , will be interpreted as

$$G_n[G(\tau_1, \dots, \tau_n), \alpha] \quad (7.6)$$

The quantity G is obtained from F by making all possible permutations of the τ 's, summing, and dividing by the number of permutations. This symmetric rearrangement is convenient for our purposes.

The expression for

$$\exp [i \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha)] \quad (7.7)$$

is Expression 7.4 with a_{ν} replaced by ia_{ν} .

We next symmetrize the kernels in Equations 7.4. We take the sum over k_i , rearrange terms in all possible ways, add, and divide by the number of ways. This will give our formula. However, I am chiefly interested in the lowest-order terms, G_0 and G_2 . Since those terms are already symmetric, we need not perform that operation here.

If I now introduce an ϵ , which multiplies all the a_{ν} , we have

$$\begin{aligned} & \exp \left[\epsilon \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ &= \prod_{\nu} \frac{1}{(1 - 2\epsilon a_{\nu})^{\frac{1}{2}}} \left[1 + G_2 \left(\sum_{\nu} \frac{a_{\nu} \epsilon}{1 - 2a_{\nu} \epsilon} \phi_{\nu}(\tau_1) \phi_{\nu}(\tau_2), \alpha \right) \right. \\ & \quad \left. + \text{higher powers of } \epsilon \right] \quad (7.8) \end{aligned}$$

If we assume that ϵ is small,

$$1 - 2a_v\epsilon \cong 1, \quad (7.9)$$

and that higher-order terms in ϵ are unimportant,

$$\exp \left[\epsilon \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \cong 1 + \epsilon G_2[K(\tau_1, \tau_2), \alpha] \quad (7.10)$$

For the case in which we have $i\epsilon$, the expression for the leading terms will be

$$\exp \left[i\epsilon \iint K(\tau_1, \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \cong 1 + i\epsilon G_2[K(\tau_1, \tau_2), \alpha] \quad (7.11)$$

The terms in Equation 7.10 or Equation 7.11 represent, respectively, the lowest-degree term in the development and the second-degree term in the development. We shall now consider the spectra they represent.

I am interested in the function of t ,

$$\exp \left[\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (7.12)$$

The leading terms of Expression 7.12 are given by

$$\begin{aligned} \exp \left[\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ \cong 1 + \epsilon G_2[K(t + \tau_1, t + \tau_2), \alpha] \end{aligned} \quad (7.13)$$

In case we are interested in $i\epsilon$, the expression for the leading terms is then

$$1 + i\epsilon G_2[K(t + \tau_1, t + \tau_2), \alpha] \quad (7.14)$$

Now, I notice the following: When I have developed a function of t like Expression 7.12 in terms of the orthogonal polynomials, its spectrum is the sum of the spectra of the terms. Remember that I have shown this before, because every n th-degree symmetric functional is orthogonal to every one of any other degree. Hence, the cross terms go out.

If we consider only the leading terms given by Equation 7.13 or Expression 7.14, the spectrum of Expression 7.12 will then be the spectrum of 1 plus the spectrum of $\epsilon G_2[K(t + \tau_1, t + \tau_2), \alpha]$. The spectrum of 1 is simply a line at $\omega = 0$. The spectrum of the second term is ϵ^2 multiplied by the spectrum of G_2 , whether we deal with Equation 7.13 or Expression 7.14. This follows from the fact that we multiply conjugates in computing correlation functions. We can, similarly, derive higher-degree terms in the spectrum.

What we wish to determine next is the spectrum of G_2 . The spectrum of Equation 7.13 is the transform of its autocorrelation function. The

autocorrelation function is given by

$$1 + \epsilon^2 \iint K(t + \tau_1, t + \tau_2) K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (7.15)$$

The first term merely gives a line in the spectrum at zero frequency.

If we define $Q(\omega_1, \omega_2)$ as the double Fourier transform of $K(\tau_1, \tau_2)$, then

$$K(\tau_1, \tau_2) = \frac{1}{2\pi} \iint Q(\omega_1, \omega_2) e^{i(\omega_1\tau_1 + \omega_2\tau_2)} d\omega_1 d\omega_2 \quad (7.16)$$

The Fourier transform of $K(t + \tau_1, t + \tau_2)$ is given by

$$K(t + \tau_1, t + \tau_2) = \frac{1}{2\pi} \iint Q(\omega_1, \omega_2) e^{i(\omega_1\tau_1 + \omega_2\tau_2)} e^{i(\omega_1 + \omega_2)t} d\omega_1 d\omega_2 \quad (7.17)$$

Therefore, the Fourier transform of $K(t + \tau_1, t + \tau_2)$ is, instead of $Q(\omega_1, \omega_2)$,

$$Q(\omega_1, \omega_2) e^{i(\omega_1 + \omega_2)t} \quad (7.18)$$

When we consider the integral of a product of a function by its conjugate and remember that a real function is its own conjugate, the integral of the product is the same as the integral of the product of the Fourier transform by its conjugate. Therefore,

$$\begin{aligned} \epsilon^2 \iint K(t + \tau_1, t + \tau_2) K(\tau_1, \tau_2) d\tau_1 d\tau_2 \\ = \epsilon^2 \iint |Q(\omega_1, \omega_2)|^2 e^{-i(\omega_1 + \omega_2)t} d\omega_1 d\omega_2 \end{aligned} \quad (7.19)$$

Making the substitutions

$$\omega_1 + \omega_2 = \omega, \quad \omega_1 = u \quad (7.20)$$

we obtain, for the right-hand side of Equation 7.19

$$\epsilon^2 \iint |Q(u, \omega - u)|^2 e^{-i\omega t} du d\omega \quad (7.21)$$

Therefore, the Fourier transform of Equation 7.19 is

$$(2\pi)^{\frac{1}{2}} \epsilon^2 \int_{-\infty}^{\infty} |Q(u, \omega - u)|^2 du \quad (7.22)$$

which is the spectrum of the second term in Equation 7.13.

LECTURE 8

Application to the Study of Brain Waves, Random Time, and Coupled Oscillators

I want to point out how I came to go into the question of the dip in frequency of a power-density spectrum. I first observed it in a specific problem. I was interested in the spectrum of the brain waves. As you know, if you put a pair of electrodes on the scalp, you get fluctuating potentials that can be amplified. Records can be taken, and a spectrum can be obtained from the autocorrelation coefficients. I do not need to say very much about the instrumentation. The trick that we use is the following. We take the record of the voltage of the brain waves, but, instead of recording it on paper, we record it on magnetic tape, and we record in frequency modulation. Frequency modulation is very important for the following reasons. With repeated playing, a certain amount of erasure of the message occurs on the magnetic tape. The effect of this erasure is of much less importance for frequency modulation than for amplitude modulation. In fact, for frequency modulation the erasure is of little importance until the erasure is almost complete. Thus FM avoids distorting our message in replay.

By means of two reading heads separated in distance we can, on playback, simultaneously obtain the original record and the record delayed. The voltages of the two reading heads correspond to $f(t)$ and $f(t - \tau)$. We can add these voltages. We can very easily study not the voltages but the mean powers, which are proportional to the mean voltages squared. The mean power is given by Equation 8.1:

$$\overline{[f(t) + f(t - \tau)]^2} = \overline{f^2(t)} + \overline{f^2(t - \tau)} + 2\overline{f(t)f(t - \tau)} \quad (8.1)$$

where the line indicates time average.

In the first place, it is easy to square a voltage by a square-law rectifier. It is easy to integrate a voltage too, by charging a condenser. That is, the quantity of electricity is proportional to the voltage across it, but the quantity is proportional to the integral of the current over time.

Consider the terms of Equation 8.1:

$$\overline{f^2(t)} = \overline{f^2(t - \tau)} = \text{constant} \quad (8.2)$$

$$\text{Autocorrelation function} = \overline{f(t)f(t - \tau)} \quad (8.3)$$

In this way we get the autocorrelation of the brain wave, and that autocorrelation is subject to a harmonic analysis. This is not a difficult problem at all. I shall not go into the details; the subject belongs to the familiar general harmonic analysis.

By the way, one of the easiest ways of handling the harmonic analysis is to heterodyne the brain wave with some frequency near its most important frequency and then make a schedule analysis. When we do that, we get a very-fine-spectrum analysis of the brain wave.

It has been known for a long time that the spectrum of the brain wave has the general shape shown in Figure 8.1. There is some activity around

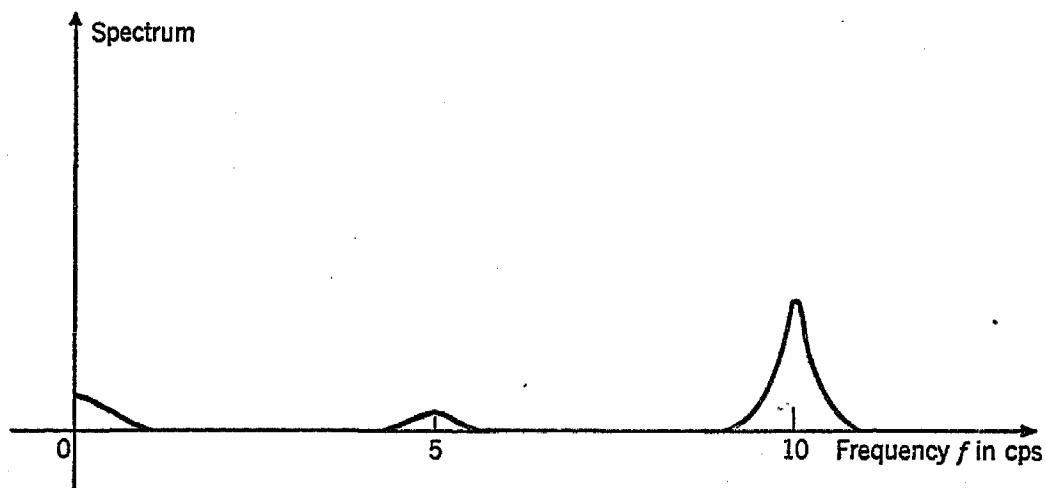


Figure 8.1

0 cycle per second, maybe a little around 5 cycles per second, but a very conspicuous activity around 10 cycles per second. The activity around 10 cycles per second is the so-called alpha rhythm.

Notice, by the way, that the autocorrelation method has one advantage in studying waves of this sort. We are studying not merely the power around a certain frequency, but we are studying all the power, and if it happens to be around that frequency, we get it. We are not looking for the power at a specific frequency, we are looking for an integral of the power up to a certain point.

Notice too that when we make a harmonic analysis, a certain question of resolving power comes in. If we want to read this to a tenth of a cycle, we must have a hundred successive oscillations to work with, that is, at least 10 seconds. We must have the autocorrelation for 10 seconds. We have a theory here that is exactly like optical theory. The great advantage of this method of doing harmonic analysis is that it makes it easy to use a large time base for the harmonic time analysis. The principle is exactly the same as that of a Michelson interferometer in which we do the same thing for light.

In the Michelson interferometer, we take a light beam, collimate it, divide it into two beams by means of a half-silvered mirror, reflect the two beams together, and let them interfere with one another. We then measure the amount of light as a function of the difference in paths between the two beams. This process gives an extremely high resolving power.

The large time base for a harmonic time analysis is entirely equivalent to using a high resolving power.

We tried a high resolving power on some actual brain-wave cases. The people were awake but at rest with their eyes closed. Figure 8.2 gives the magnified spectrum around 10 cycles per second.

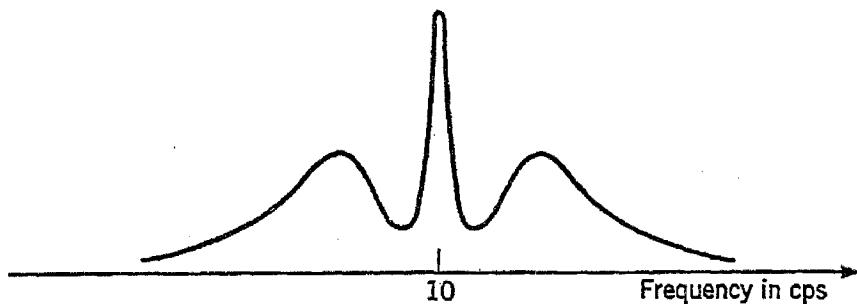


Figure 8.2

It would have been quite conceivable that there would not be a phenomenon of that sort; i.e., it might have been that all our work on the fine structure of the brain waves was wasted. However, once we find this effect, we are under an obligation to try to give an explanation of it. There are two things that are striking here: One is the very narrow line at the center of the spectrum, and the other is that this line arises from a dip. Then the problem arose of giving a reasonable explanation of what caused this spectrum. As a matter of fact, that is what started me on a good deal of this more recent work that I have told you about in the last few lectures.

In the first place, there is a suggestion that in the brain we have some sort of oscillators, and that these oscillators in some sense constitute a

more accurate oscillator *en masse* than they do singly. This suggestion led me to wonder whether this phenomenon occurs anywhere in electrical engineering. The answer is that it does occur. I shall give you one example.

Consider a number of generators feeding a load through bus bars, as illustrated in Figure 8.3. Not only do we have a load, but we have a fluctuating load. People are turning their lights off and on in a random

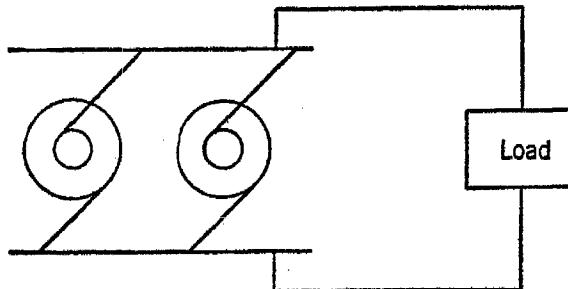


Figure 8.3

way. We can represent this by something rather closely associated with the shot-effect or Brownian motion sort of expression. Do not take me too seriously here, but we have something of that sort. What maintains the frequency, let us say, in a 60-cycle system? Well, there are several stages of considering it. One is that the individual generator is driven by a prime mover which has a governor on it, so that if it is going fast the steam, or whatever other power there is, is cut off. When it is going slowly, more power is emitted. This tends to keep the apparatus going at a constant rate. Yet these governors, while quite good, are by no means as good as the regulation of the entire system. The regulation of our electric light system is something within a matter of three minutes a day, or something of that sort, even if we do not fake it. We do fake it in two ways. One of the things we do is that we deliberately speed up or slow down the system. There is another thing that we do; we have one particular generator in the system equipped with a rather broad range governor to compensate as much as possible the wider fluctuations of the load. Even when that is done, the finer fluctuations of the load are compensated by the mass of systems and by narrow-range governors. I may say that we can neglect the broad-range governor for our purposes because the narrow-range governors, so long as the system is primed by the wider-range governor, do actually constitute a single governor of a statistical nature for the system. How does such a system of governors act? We shall forget the distinction between narrow and broad governors. I shall tell you that if the generator is leading or going fast it takes more of the load than it would normally take. The fact that it is going too fast means that the governor will act; the governor will combine with the

increased load to slow it up. If it is going too slowly, it will take less load and tend to speed up. In other words, the frequencies of these different oscillators (and we can regard the generators as nonlinear oscillators) are not independent of one another but are pulled together. The important thing here is the coupling of the frequencies of the generators, since we know, as a matter of fact, that the coupling is responsible for the good regulation that we have of frequency.

Several questions come up here. Question 1: Can we expect that this method of regulation of frequency gives us a sharp central frequency, which would here be 60 cycles per second, say, with a 60-cps generator? Question 2: Does this frequency regulation tend to produce bands outside of the central frequency but near to it? Question 3: Does there tend to be a dip between these bands and the central frequency? Question 4: Are these ideas applicable to the brain waves? I am going to answer all these questions in the affirmative. I am going to connect them with the sort of thing that I was talking about in the previous lecture.

In the case of the brain wave, it is known that if we put a sudden electric impulse on the brain, then this is followed by an active discharge at the rate of about 10 cycles per second. There are some things like local oscillators. This phenomenon is largely quite local in the brain, or at least it begins this way. Thus there are oscillators in the brain at about 10 cycles per second, as demanded. Moreover, these are nonlinear oscillators.

The proof of that is very simple. Their frequency of discharge depends upon the intensity of the stimulus. A linear oscillator has its frequency independent of this intensity. This suggests that the frequencies of these oscillators in the brain can interact, and now we ask: How do they interact? Can there be a pulling-together phenomenon just such as we had with the generator? Is that a reasonable assumption? The answer is that that is a reasonable assumption.

We can flicker a lamp into the eye at about 10 cycles per second. When we do that, it tends to pull the α rhythm in the brain into frequency and phase with itself. In other words, the different frequencies tend to pull together. While you may say that this is a visual stimulation, of course it is also an electric stimulation because there are action potentials that you get by this stimulation.

Furthermore, some work has been done about a direct electrical driving of the brain, and it has been done by putting wires into the brains of animals. It has been done with man in a rather interesting way.

We suspend a sheet of tin from the ceiling of a room and connect the sheet to one terminal of a 400-volt 10-cps generator. This apparatus can produce electrostatic induction in anything in the room. It can

actually drive the brain, causing a decidedly unpleasant sensation. In other words, that gives us a way of putting a signal in, not through a sense organ, but directly into the brain.

The questions are then: What can we say about the mathematics of systems that pull together in frequency, and does a mechanism of that sort produce the dip characteristic? That is what I want to connect with what we were doing in the previous lecture. We have been speaking here about something like, say, a 10-cycle or a 60-cycle frequency, according to whether it was the electric circuit of the brain or a power system. We need to have the pulls by frequencies on the opposite sides, equal and opposite, so that if frequencies above pull the brain wave or the electric light wave one way, frequencies below by the same amount pull it the same amount the other way. That is what we mean by pulling together or pulling apart. Otherwise, they would pull in the same direction, and that is not the case that we want. We want a sort of action and reaction, equal or nearly equal.

We are not going to modify our theory very much by working about a central frequency other than 0. If we heterodyne our wave with a central frequency, say, about 10 cycles per second for the brain wave or 60 cycles per second for the power line, then instead of bigger and smaller frequencies, we have positive and negative frequencies. In other words, we want two frequencies to have a pull on each other opposite to the pull that the corresponding negative frequencies have. Note that a negative frequency is obtained by a reversal of time:

$$e^{i(-\omega)t} = e^{i\omega(-t)} \quad (8.4)$$

Therefore, we are asking the following question: If we have some pull between our frequencies that is equal and opposite for frequencies of opposite sign, or times of opposite sign, can we, out of such a mechanism,

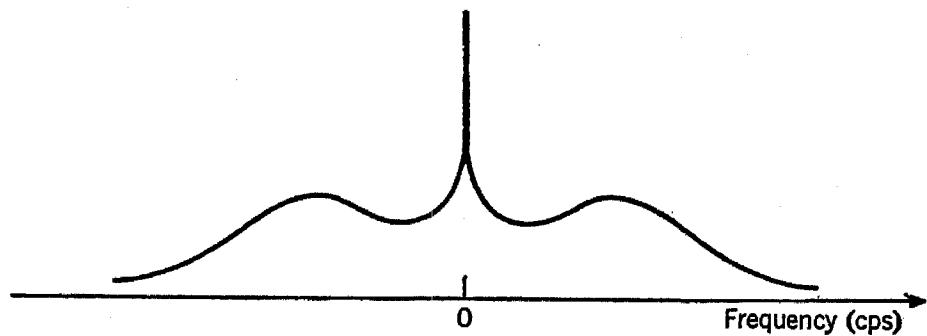


Figure 8.4

get a spectrum like Figure 8.4, with a peak at 0 frequency and with the dips around 0? That is the question that we are going to discuss in this lecture.

Let us first take up the case where the central frequency is 0. We have discussed the spectrum represented by

$$\exp \left[i\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \quad (8.5)$$

Expression 8.5 means that we have a random input, and we want this random input to have the exactly opposite effect, if we reverse either all of our frequencies or all of our times, from that which it would have if we keep all our frequencies and times the same. The suggestion is fairly clear that, if Equation 8.6 is true, the pull is turned around in direction when we change larger frequencies to smaller frequencies or smaller times to larger times:

$$K(-\tau_1, -\tau_2) = -K(\tau_1, \tau_2) \quad (8.6)$$

We are going to make the assumption that Equation 8.6 is true. This is essentially the assumption, if we examine Expression 8.5 from the frequency standpoint, that the action of positive or negative frequencies is the opposite in sign of the action of negative or positive frequencies, respectively; i.e., the effects are equal and opposite. In the brain wave case, the assumption is that if for a brain wave we are a little above the 10 cycles per second the effect is the opposite of being the same amount below 10 cycles per second.

We want to discuss the spectrum of Expression 8.5, a spectrum which we obtained in the previous lecture. There were certain leading terms in it and certain terms of less importance. Recall that

$$\begin{aligned} & \exp \left[i\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right] \\ &= 1 + i\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) + \dots \end{aligned} \quad (8.7)$$

where the remaining terms are of second or higher degree in ϵ and are small for small ϵ . The first two terms on the right-hand side of Equation 8.7 can be written as

$$\begin{aligned} & 1 + i\epsilon \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \\ &= 1 + i\epsilon \int K(\tau, \tau) d\tau + i\epsilon G_2[K(t + \tau_1, t + \tau_2), \alpha] \end{aligned} \quad (8.8)$$

The first two terms on the right-hand side of Equation 8.8 are constants.

The autocorrelation of the right-hand side of Equation 8.8 is given by Expression 8.9:

$$1 + \epsilon^2 \left[\int K(\tau, \tau) d\tau \right]^2 + 2\epsilon^2 \iint K(t + \tau_1, t + \tau_2) \overline{K(\tau_1, \tau_2)} d\tau_1 d\tau_2 \quad (8.9)$$

where the bar indicates the complex conjugate. The first two terms of Expression 8.9 are constants. They contain no t . If we make the harmonic analysis of a constant, we see that it is just a spectrum line at 0 cycle per second, so that the spectrum we get from the first two terms of Expression 8.9 is just a line at 0, as shown in Figure 8.5. Then we superimpose on that the spectrum that we get from the last term of Expression 8.9.

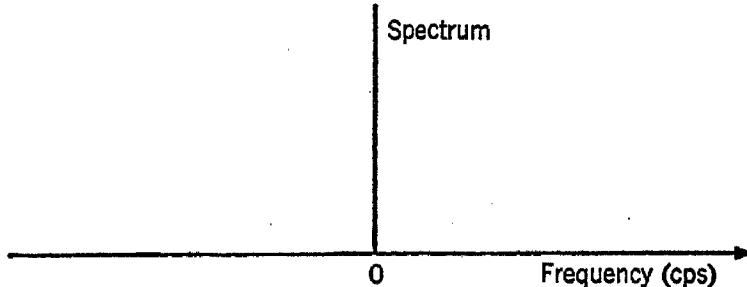


Figure 8.5

Now I shall get the part of the spectrum due to the last term of Expression 8.9. That is we want:

$$\frac{1}{2\pi} \int e^{-i\omega t} \left[\epsilon^2 \iint K(t + \tau_1, t + \tau_2) \overline{K(\tau_1, \tau_2)} d\tau_1 d\tau_2 \right] dt \quad (8.10)$$

We define a Fourier transform in two variables as

$$Q(\omega_1, \omega_2) = \frac{1}{2\pi} \iint K(\tau_1, \tau_2) e^{-i\omega_1 \tau_1} e^{-i\omega_2 \tau_2} d\tau_1 d\tau_2 \quad (8.11)$$

By writing out the Fourier transform formally, we get

$$e^{i(\omega_1 + \omega_2)t} Q(\omega_1, \omega_2) = \frac{1}{2\pi} \iint K(t + \tau_1, t + \tau_2) e^{-i\omega_1 \tau_1} e^{-i\omega_2 \tau_2} d\tau_1 d\tau_2 \quad (8.12)$$

Now it follows that

$$\begin{aligned} \epsilon^2 \iint K(t + \tau_1, t + \tau_2) \overline{K(\tau_1, \tau_2)} d\tau_1 d\tau_2 \\ = \epsilon^2 \iint |Q(\omega_1, \omega_2)|^2 e^{i(\omega_1 + \omega_2)t} d\omega_1 d\omega_2 \end{aligned} \quad (8.13)$$

Change variables now,

$$\omega = \omega_1 + \omega_2 \quad (8.14)$$

$$u = \omega_1 \quad (8.15)$$

Then

$$\begin{aligned} \epsilon^2 \iint |Q(\omega_1, \omega_2)|^2 e^{i(\omega_1+\omega_2)t} d\omega_1 d\omega_2 \\ = \epsilon^2 \iint |Q(u, \omega - u)|^2 e^{i\omega t} du d\omega \end{aligned} \quad (8.16)$$

But the right-hand side of Equation 8.16 is just the Fourier transform of

$$\epsilon^2 \int |Q(u, \omega - u)|^2 du \quad (8.17)$$

Then, the part of the spectrum that we were trying to evaluate in Expression 8.10 is given by Expression 8.17.

We see that the spectrum given by Expression 8.17 is continuous, unlike the spectrum of the line at 0. It is essentially positive because of the absolute value of the square. All spectra are essentially positive.

The question is: How does the spectrum given by Expression 8.17 behave when ω is 0? At $\omega = 0$, the value of the spectrum is

$$\epsilon^2 \int |Q(u, -u)|^2 du \quad (8.18)$$

But if $Q(u, -u)$ is identically 0, then Expression 8.17, which is positive, must go down to 0 at $\omega = 0$ and $\omega = \infty$. In other words the spectrum has the dip at that point which we want.

Now, the fact that $Q(u, -u)$ is 0 follows at once from the definition of Q if Equation 8.19 is satisfied:

$$K(\tau_1, \tau_2) = -K(-\tau_1, -\tau_2) \quad (8.19)$$

The proof is as follows:

$$Q(\omega_1, \omega_2) = \frac{1}{2\pi} \iint e^{-i(\omega_1\tau_1 + \omega_2\tau_2)} K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (8.20)$$

$$Q(\omega_1, \omega_2) = \frac{1}{2\pi} \iint e^{i(\omega_1\tau_1 + \omega_2\tau_2)} K(-\tau_1, -\tau_2) d\tau_1 d\tau_2 \quad (8.21)$$

$$Q(\omega_1, \omega_2) = \frac{1}{2\pi} \iint e^{i(\omega_1\tau_1 + \omega_2\tau_2)} [-K(\tau_1, \tau_2)] d\tau_1 d\tau_2 \quad (8.22)$$

Let

$$\omega_1 = u \quad (8.23)$$

$$\omega_2 = -u \quad (8.24)$$

Then Equations 8.20 and 8.22 become

$$Q(u, -u) = \frac{1}{2\pi} \iint e^{-iur_1} e^{iur_2} K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (8.25)$$

$$Q(u, -u) = -\frac{1}{2\pi} \iint e^{iur_1} e^{-iur_2} K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (8.26)$$

where K is symmetric. Hence,

$$Q(u, -u) = \frac{1}{2\pi} \iint e^{iur_1} e^{-iur_2} K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (8.27)$$

$$Q(u, -u) = -\frac{1}{2\pi} \iint e^{iur_1} e^{-iur_2} K(\tau_1, \tau_2) d\tau_1 d\tau_2 \quad (8.28)$$

But a quantity that is equal to its own negative must be 0.

In other words, we have established quite rigorously that at 0 there is a dip in the spectrum down to 0.

We now consider

$$\exp i \iint K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \quad (8.29)$$

It is quite clear that there will be other terms superimposed on the spectrum. The leading part of the spectrum is shown as a in Figure 8.6.

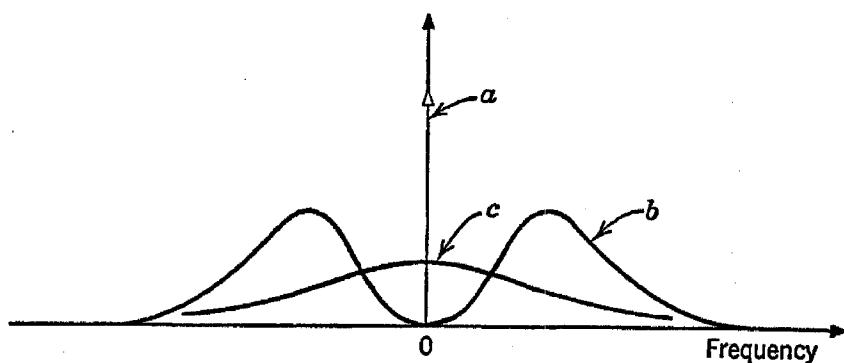


Figure 8.6

The next part is b . The remaining parts are shown as c . Some of the parts of c actually have a peak at the origin, but if ϵ is small the effects of these are small. Still further investigation is necessary for those cases, but at least we see that the dip is not unnatural. So then I have

given at least a preliminary justification of my assumption that the pulling together of frequencies is responsible for the dip phenomenon.

I suspect that this dip phenomenon will occur in a great many other physical cases. We shall probably find in molecular spectra a similar nonlinear binding of frequencies. There is at least one interesting case where it is known to occur. After all, the planets are oscillators of a nonlinear characteristic, and after all, their perturbations represent nonlinear coupling. As we know, besides the great planets, we have the asteroids, which are small planets. There are many of them, enough to give a small population, and their periods are somewhat different. Now we should expect that their periods would be continuously distributed. They are not. There are certain ranges of the periods of the asteroids that are missing, and these ranges are about simple, rational multiples of the frequency of Jupiter. We have a perturbation of the system by this powerful frequency of Jupiter, and we are going to have something like a cyclotron effect with a frequency rationally related to that of Jupiter. We shall have a cumulative effect over a great many periods of the perturbations due to Jupiter. They will come in the same place, and they may knock the orbit into something quite different. We should expect much lower stability near the frequencies of Jupiter. I suspect that this is just the same thing we find with brain waves, which causes a dip about a spectrum line. I suspect that if we were to make a tight mathematical analysis of the frequencies occurring in a cyclotron, we should find that frequencies near the driving frequency are less represented than frequencies a little bit away, i.e., that too near the driving frequency is unstable in a cyclotron.

Some Thoughts on Quantum Theory

Here I want to start something quite different that has to do with quantum theory. For quantum theory, we want to take up not real but complex differential space. I have already said something about complex differential space in Lecture 2.

Consider

$$\int_{-\infty}^{\infty} \phi_n(\xi) dY(\xi, \alpha) \quad (9.1)$$

where Y is a complex expression representing complex Brownian motion. I use ξ instead of t because time suggests something a little more special than what I am doing.

If the ϕ_n 's are a normal and orthogonal set of functions (not necessarily real) and if Y has a derivative, then Expression 9.1 represents the complex Fourier coefficient of the derivative of Y . There is no derivative of Y in general. The Brownian motions, real or complex, are nondifferentiable except for a set of measures 0. Nevertheless, Expression 9.1 is a formal Fourier coefficient.

We have seen before that Expression 9.1 can be written as

$$\int_{-\infty}^{\infty} \phi_n(\xi) dY(\xi, \alpha) = A_n(\alpha) + iB_n(\alpha) \quad (9.2)$$

Both A and B have Gaussian distributions of standard form. If the ϕ_n 's are orthogonal for different n 's, then the distributions are independent. We have seen that result before.

The next thing I want to consider is the distribution of the absolute value of these Fourier coefficients. That is, we are interested in the distribution of the following expression for different n 's:

$$\left| \int_{-\infty}^{\infty} \phi_n(\xi) dY(\xi, \alpha) \right| \quad (9.3)$$

Suppose then that we consider that we have two independent variables x and y . The distribution of them is given by

$$p_1(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{y^2}{2}\right) \quad (9.4)$$

Equation 9.4 gives us the distribution of the point (x, y) in the complex plane. Now let us consider $p_1(x, y)$ not as a Cartesian distribution but as a polar distribution. The distribution then becomes

$$p_2(r, \theta) dr d\theta = \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right) r dr d\theta \quad (9.5)$$

Integrating on θ , we get

$$p_2(r) dr = dr \int_0^{2\pi} p_2(r, \theta) d\theta = r \exp\left(-\frac{r^2}{2}\right) dr \quad (9.6)$$

We see that the distribution of the absolute value of the Fourier coefficient of Expression 9.3 is given by Equation 9.6.

I now have two expressions,

$$\int \phi_1(\xi) dY(\xi, \alpha) = A_1(\alpha) + iB_1(\alpha) = r_1 \quad (9.7)$$

and

$$\int \phi_2(\xi) dY(\xi, \alpha) = A_2(\alpha) + iB_2(\alpha) = r_2 \quad (9.8)$$

I am going to ask for the probability

$$P\{|a|r_1 > |b|r_2\} \quad (9.9)$$

The question is perfectly definite because I have just derived the distribution of the two quantities r_1 and r_2 . This is a good probability question in the classical sense of measure theory.

Probability 9.9 is given by

$$P\{|a|r_1 > |b|r_2\} = \iint_{|a|r_1 > |b|r_2} r_1 \exp\left(-\frac{r_1^2}{2}\right) r_2 \exp\left(-\frac{r_2^2}{2}\right) dr_1 dr_2 \quad (9.10)$$

Equation 9.10 is true because r_1 and r_2 have independent distributions. The right-hand side of Equation 9.10 can be evaluated by changing variables.

Let

$$u = \frac{r_1^2}{2} \quad (9.11)$$

$$v = \frac{r_2^2}{2} \quad (9.12)$$

Then

$$\begin{aligned} P\{|a|r_1 > |b|r_2\} &= \iint_{|a|^2 u > |b|^2 v} e^{-u} e^{-v} du dv \\ &= \int_0^\infty e^{-u} du \int_0^{\left|\frac{a}{b}\right|^2 u} e^{-v} dv \\ &= \int_0^\infty e^{-u} \left[1 - \exp\left(-\left|\frac{a}{b}\right|^2 u\right) \right] du \end{aligned} \quad (9.13)$$

$$P\{|a|r_1 > |b|r_2\} = 1 - \frac{1}{1 + \left|\frac{a}{b}\right|^2} \quad (9.14)$$

Therefore

$$P\{|a|r_1 > |b|r_2\} = \frac{|a|^2}{|a|^2 + |b|^2} \quad (9.15)$$

We wish to extend similar considerations to randomness in several dimensions. Here we have to introduce a fundamental change in our viewpoint. Up to the present, we have generally associated randomness with a function $x(t, \alpha)$, or some complex analogue, which represents randomness in time. We can equally well interpret t to be a space variable, but it seems as if it must be a one-dimensional variable. This is, however, not as essential as it may seem. The trick of applying random theory to n -dimensional space is simple: We study, not the analogue of $x(t, \alpha)$, which is indeed tied to a one-dimensional t , but a certain additive functional of sets of points associated with α , which generalizes equally well to any number of dimensions.

Suppose we have a measurable set S of values of s . I take $S(s)$ as a function, defined by S , which is 1 on S and 0 elsewhere. I am working now in one dimension, but I am going to extend this. The function $S(s)$ is L^2 , obviously, since S has a finite measure. I now consider

$$\int_{-\infty}^{\infty} S(s) dx(s, \alpha) \quad (9.16)$$

that represents mass in some sense or other. It is an additive, but not absolute-additive, functional of S . This integral, Expression 9.16, is clearly defined from what we have done before. Furthermore, if I have another set $T(s)$, measurable and not overlapping $S(s)$, then I have

$$\int_{-\infty}^{\infty} T(s) dx(s, \alpha) \quad (9.17)$$

Expression 9.17 is orthogonal to Expression 9.16 and, by what we have said, has an independent Gaussian distribution, due to the orthogonality of $T(s)$ and $S(s)$. The mean-square of Expression 9.17 depends on the measure of T on s . These properties are true not only for 2 but also for any number. In other words, $x(s, \alpha)$ defines an additive, but not absolute-additive, function of measurable sets S . It is always Gaussian in distribution, and in each case the modulus of the Gaussian distribution depends on the measure of S and is independent for all orthogonal regions. In other words, I can replace the function $x(s, \alpha)$ by this functional of sets that will completely define $x(s, \alpha)$ and that is completely defined by it. In a one-dimensional case there is no advantage, but in the n -dimensional case there is an advantage, since, if I take the finite region of the plane, I can extend this to the infinite region. There is an area-preserving mapping of a square of the plane on the line, or a cube in space on the line, and so on. That is the way we do it; we take the square and divide it into 4 parts and represent them on 4 equal segments of $[0, 1]$. Then we divide each of the quarter squares into 4 parts and represent them on 16 equal segments of $[0, 1]$, and so on. That gives us a measure-preserving transformation between square measure on the plane and linear measure on the line. This technique will go over to three or more dimensions. That is, if we have an additive functional in one dimension, as in Expression 9.16, we have it in n dimensions. That is the important thing.

Thus, while the concept of $x(s, \alpha)$ is not too good a concept for space, the concept of the additive functional of sets of points generated by it is just as good in n dimensions as in 1. I am bringing this in to show the naturalness of using the differential-space idea in larger dimensions than 1. It is a perfectly good idea. Instead of taking $x(s, \alpha)$, we must take the additive functional that it generates. In other words, we have in any number of dimensions a random additive functional, and we have this not only in the real but in the complex case, that is, in the case in which we have the sum of two random ones, one real and one purely imaginary, independent of one another. That idea has nothing to do with dimensionality.

The reason for the preceding discussion is that we are going on to quantum theory, where we are going to deal with randomness in space,

and we do not want to be confined too much in this to the idea of one dimension because it is not relevant to what we are doing. Nevertheless, in order to avoid excessive notations, there is no reason why I may not refer $x(s^*, \alpha)$ in n dimensions to $x(s, \alpha)$ in one dimension by a measure-preserving mapping and keep up the same notation that I had before. This is a preliminary remark that is very important for our present purposes: because we are going to quantum theory; because we are dealing with Brownian motion, not in time but in space; and because we want to get rid of any idea that this is tied up with one dimension. I want to come to quantum theory. There is no need to go into a great deal of detail on Hamiltonians now, for we are not interested in the derivation of the equation of quantum theory from classical mechanics, but the sort of description that it gives of the world. I am going to take the discrete spectrum case; it will give us our main ideas for the present lecture. I do not need to go beyond that for the moment.

The state of the world is represented in terms of a certain distribution of characteristic functions, modes of oscillations, of the world as a whole. At one particular time, the state of the world is represented by

$$\sum a_n \phi_n(s) \quad (9.18)$$

where s is the space variable in n dimensions, s and a_n are complex in general, and we have the assumption,

$$\sum |a_n|^2 = 1 \quad (9.19)$$

That is the way of representing the world at a given time. In this world, the quantum-theory assumption is that the world is either completely in one quantum state or completely in another. Suppose that it is completely in the n th state. The quantum hypothesis is that the quantity $|a_n|^2$ in some sense or other represents the probability of its being the n th state. From Equation 9.19, the sum of these probabilities is 1. The coefficient a_n is called the probability amplitude. Being a complex number, a_n can not be a probability. However, $|a_n|^2$ is a probability. Quantum theory has this remarkable interplay of complex quantities determining probability and real quantities that are probabilities. The large part of the mystery of quantum theory is why, from this whole complex apparatus, we should get a real probability. The interplay of real and complex has been made very clear by postulates, but why the square of a complex number should be the percentage of cases that are in a certain state has never been made clear.

The Schrödinger equation is a way of connecting this distribution with the time. It says that, if the ϕ_n are eigenfunctions, the proper sort of normal and orthogonal functions that belong to a certain problem,

then the representation of Expression 9.18 at time t is

$$\sum a_n \exp(i\lambda_n t) \phi_n(s) \quad (9.20)$$

In other words, if we are referring to the proper characteristic functions of the system, the absolute value of the amplitude of these characteristic functions that you select does not change with time, but the phase changes. The λ_n are called the characteristic frequencies, and the a_n are called the characteristic amplitudes.

When we have referred the system to the characteristic functions ϕ , then

$$|a_n \exp(i\lambda_n t)|^2 = |a_n|^2 \quad (9.21)$$

where t is real, and thus the probability is not variable with time. You have a stationary system. It is not necessary for me to go into the derivation of the Schrödinger equation, which has this solution from the dynamical considerations, because what I am considering now is not the mechanics but the phenomenology of quantum theory, the sort of thing that quantum theory is looking for. Now we shall see that something here is very queer; that we start with a complex theory and end up with a real probability. The probability of the n th state is

$$|a_n|^2 = \frac{|a_n|^2}{\sum |a_\nu|^2} \quad (9.22)$$

since, from Equation 9.19,

$$\sum |a_\nu|^2 = 1 \quad (9.23)$$

I am not trying today to answer any philosophical problem of basic quantum theory, but what I am trying to do is to ask if there is any actual situation in which we get a probability as the quotient of the square of the absolute value of a complex number by the sum of the squares of the absolute value of similar complex numbers. The answer is yes, there is such a situation. This is where the Brownian motion idea comes in again. The simplest case to discuss is the one in which there are two eigenfunctions, that is, where you have only two alternative states. As a matter of fact, we could throw all the rest together as one sort of eigenfunction and handle the general case as that. The probability with two eigenfunctions that only one of two is realized is

$$\frac{|a_1|^2}{|a_1|^2 + |a_2|^2} \quad \text{or} \quad \frac{|a_2|^2}{|a_1|^2 + |a_2|^2} \quad (9.24)$$

Both probabilities in Expression 9.24 are real, and their sum is 1.

Here, let us confine ourselves to a discussion of a situation like this because all the real points will come out. Can I give a situation in which

probabilities come out naturally in a form of this sort with the square of an absolute value coming out as divided by the sum of the square of two absolute values as the probability? The answer I gave when I deduced Equation 9.15 was, "Yes, there is such a case." I do not, for the present, say this is the actual information of quantum theory. What I am showing is that there is an intellectual model of probabilities, tied up with the Brownian motion, in which exactly this situation occurs. What I have said previously would indicate that there is no fundamental distinction between one and more dimensions in space. If we have a set of orthogonal functions in two dimensions, we can map them by a measure-preserving change into orthogonal functions in one dimension, and the whole theory goes over with merely formal changes. Having made that remark, I am then legitimately entitled to work in one dimension, because I am not really working in one dimension but in n dimensions. Suppose that I have $a_1 \phi_1(s)$ and $a_2 \phi_2(s)$. These generate

$$A = a_1 \int \phi_1(s) dy(s, \alpha) \quad (9.25)$$

$$B = a_2 \int \phi_2(s) dy(s, \alpha) \quad (9.26)$$

Equations 9.25 and 9.26 are complex quantities with distributions depending on α . I ask the probability that $|A| \geq |B|$, that is, the measure of the set of values of α for which the inequality is true. This is a perfectly good probability question with regard to Lebesgue measure. The answer at which I have arrived in Equation 9.15 is that

$$P(|A| \geq |B|) = \frac{|a_1|^2}{|a_1|^2 + |a_2|^2} \quad (9.27)$$

which is just the sort of probability that occurs in quantum theory. In other words, what I have done is to set up a valid process that gives me a genuine probability which would be numerically equal to the sort of probability that we get in quantum theory and may give us a working model of what happens in quantum theory.

In other words, we have set up an abstract mathematical problem where the solution comes out as a real probability. We have an actual Lebesgue measure of a set of points which has exactly the same form as the probability that occurs in quantum theory in the choice between two eigenfunctions. The quantum-theory expression has been explained physically as a probability, but the mathematical reason why it should come out as a probability has not been made clear. In other words, without metaphysics, we have a working model of the situation that

occurs in quantum theory for the choice between two eigenfunctions. For the present I am not going any further. I want it to be clear that I have done this in space and that I have tied this up with the Schrödinger equation.

There is one other point that I want to make here. The Schrödinger equation transforms the expression

$$\sum a_n \phi_n(s) \quad (9.28)$$

into

$$\sum a_n \exp(i\lambda_n t) \phi_n(s) \quad (9.29)$$

Furthermore, this is a unitary transformation. What I mean is that

$$\int |\sum a_n \phi_n(s)|^2 ds = \int |\sum a_n \exp(i\lambda_n t) \phi_n(s)|^2 ds \quad (9.30)$$

because the eigenfunctions are normal and orthogonal functions. In fact, both sides of Equation 9.30 are equal to the sum of the square-magnitude a_n 's. In other words, the Schrödinger equation can be considered as determining a group of measure-preserving transformations in the s space.

Now, however, let me consider

$$\sum a_n \exp(i\lambda_n t) \phi_n(s) = \Phi(s, t) \quad (9.31)$$

a function of space and time. Equation 9.31 may be written as

$$\Phi(s, t) = U^t \Phi(s) \quad (9.32)$$

since this is a unitary transformation, a complex transformation that is linear and preserves

$$\int |\phi_n(s)|^2 ds \quad (9.33)$$

Let us consider

$$D = \int \Phi(s, t) dy(s, \alpha) = \int U^t \Phi(s) dy(s, \alpha) \quad (9.34)$$

We have already seen that a unitary transformation on the s -scale generates a measure-preserving transformation on the α -scale. We have seen that before. In other words, Equation 9.34 becomes

$$D = \int \Phi(s) dy(s, T^t \alpha) \quad (9.35)$$

where $T^t \alpha$ is a pointwise measure-preserving transformation. The group of unitary transformations that generates the solution of the Schrödinger

equation at any time from the solution at one time can be reduced to a group of measure-preserving transformations on the α space or the phase space. This is most important because groups of measure-preserving transformations are nothing new in physics. They come up in Gibbs' statistical mechanics. What I am saying is that we here have a form of quantum theory that is intimately related to the Gibbsian statistical mechanics.

I have just time enough to state this, and that will be the last thing in the lecture today. If you remember, we have a number of particles, and they have a certain set of coordinates in classical physics: a set of coordinates q_n in position and p_n in momentum. I shall not go now into the particular way that these are connected to the Hamiltonian. Actually, when we start in mechanics, we have position and velocity coordinates, but, in general, the velocity coordinates are not the momentum coordinates. The momentum coordinates are obtained from the position coordinates and the velocity coordinates by a certain set of transformations given by Lagrangian coordinates. What we have is this. The transformation has been proved to have a measure invariant; that is, the product of dp_n and dq_n at a given time equals that product at a later time. Consider a phase space. If we have one momentum and one position coordinate, we have a two-dimensional phase space. If we have three momentum and three position coordinates, we have six-dimensional phase space. Let us take the two-dimensional case. Here is a certain region of p 's and q 's, not positions and velocities, but positions and momenta. At the end of a little time, this region will have become wobbly. The important thing in Gibbsian statistics is that, although this transformation is not an identity, it is a volume-preserving transformation. There is an invariance of volume in phase. So the Gibbsian statistical theory is based on the fact that a system in time will undergo a series of such transformations infinitesimally. A group of such transformations, if the system is in equilibrium statistically, will preserve the measure. We have exactly the parallel situation here. We have on the α line, between 0 and 1, a transformation of such a kind that any set of points will go into a set of the same measure, and we have a group of such transformations when we consider the solution of the Schrödinger equation in time. In other words, there is a highly exact analogy between this way of stating quantum theory and the traditional way of stating Gibbsian statistical mechanics. Gibbs is considering this sort of flow of a system into itself. That is, if you take a little time later, you have a mixture of this sort. If you take twice the time, this mixture appears twice, but the volume is still conserved; and so on. I want to emphasize that referring the Schrödinger equation and the transformation generated by it to something very similar in Gibbsian

statistical mechanics gives an important parallelism and one that I think we shall have more to speak of later.

This will finish what I shall say about quantum theory at this time. There is a lot more that has to be done here, but I want this suggestion to be understood in this lecture. Later on in the lectures I want to discuss turbulence and some of the work that I am doing on that subject. That work will be rather incomplete, but that work will tie up with quantum theory, and I have a good deal to say about it. In order to get the more definite work done earlier, I think that in the next lecture I shall go to the work that Professor Amar Bose and I have been doing on the application of random functions to the study of nonlinear electrical systems, both as to their synthesis and analysis. Some of the methods we have given are directly applicable there.

Nonlinear Systems—I

Now I am going to discuss the analysis and synthesis of networks that are nonlinear. I am going to bring out that we can discuss the behavior of nonlinear networks, under certain very general conditions, by the response to a Brownian or shot-effect input. One fact about the shot effect is that shot effect can, in a certain sense, imitate any input. That is, suppose, as in Figure 10.1, that we have a finite number of gaps. There is a finite,

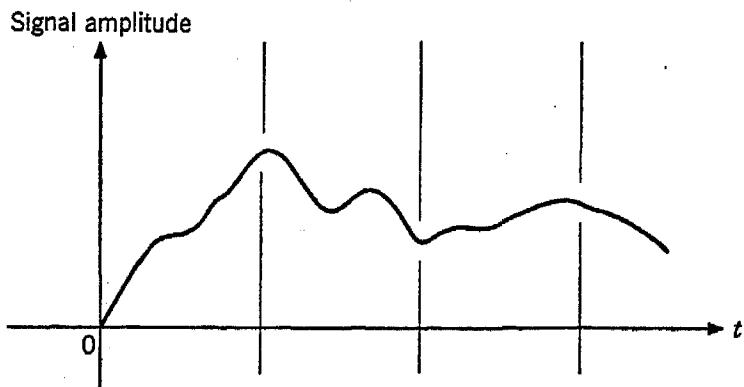


Figure 10.1

nonzero probability that a Brownian motion will go through these gaps. If we move these gaps closer together and make them narrow enough, the set of gaps can “approximate” as closely as we desire any continuous motion, and still there is a finite probability that a Brownian motion will pass through them. So, unless a network is so dividing time that this approximation is no longer good, the response of the network to Brownian motions gives us essentially the response to any motion.

For the present, we shall take the network to be a black box with two inputs and two outputs. I do not know why boxes are black, but boxes are *ex officio* black. Now, it is easy to make a shot-effect generator in which the total quantity of electricity that passes in a certain time will be rather close to a Brownian motion. Hence, we can study the output of the black box with a shot-effect generator on the input. Assuming that things do not blow up, and so on—these specifications must be made tighter—let us see how the output for shot-effect input can be represented.

We have seen that if we have $x(t, \alpha)$, then any L^2 function of α can be represented by Expression 10.1:

$$\sum G_n[K_n(\tau_1, \dots, \tau_n), \alpha] \quad (10.1)$$

If I want an effect that is carried along with the input, we have also seen that we are simply making a shift in all of the τ 's. Hence, Expression 10.2 is a general representation for a function of t , determined by the whole history of α , which moves along as t moves along, provided that the function is L^2 in α for each t :

$$\sum G_n[K_n(\tau_1 + t, \dots, \tau_n + t), \alpha] \quad (10.2)$$

Now, we want to give the characteristic of an electric circuit, i.e. the output in terms of the input: that is, of a four-terminal network in general, which includes two-terminal networks as a special case. Notice, I am not assuming conservation of energy. Amplifiers are allowable networks provided the output does not blow up, and provided that the output is an L^2 function of the input. Assumptions that prevent explosiveness are important. We can take some networks and represent their characteristics in the form of Expression 10.2. Explosive networks, which include networks that are not explosive in the strict sense but that have intrinsic oscillations of their own, will not go this way. We are assuming some sort of a deadbeat nonlinear network which does not go into spontaneous oscillation, so that it does not depend on the infinite past. If the response of this apparatus depends on the remote past, then the Brownian motion is not a good approximation because we shall always have to consider the remote past.

So we are considering nonlinear networks of a certain deadbeat character, in which the output is asymptotically independent of the remote-past input, and in which it is L^2 . Another thing, Expression 10.2 is a little too general for the characteristics of a nonlinear network, since the output of a nonlinear network depends not on the future of the input but only on the past of the input. Nevertheless, it depends on the input; hence by restricting Expression 10.2, we can get the sort of output we can realize. I am going to do this a little later.

I want to contrast what we are doing with what is done in linear networks. In linear networks, the standard input has been a trigonometric input. That comes from the fact that if we put two trigonometric inputs into a linear network, the outputs add. We can study each separately, and they do not mix. In a nonlinear network, they do mix, and we get no great advantage by bringing in trigonometric functions. In other words, trigonometric functions are functions that have a certain type of invariance in time: that is, except for the phase, it makes no difference where we start. Trigonometric functions have a linear invariance with respect to the translation group. The Brownian motion has an invariance that is much more than linear with respect to the translation group. In other words, we are working now with complete independence at different times.

In Expression 10.1, if the kernels, $K_n(\tau_1, \dots, \tau_n)$, depend only on the past of their arguments, then we have the sort of a characteristic that can actually be realized in a network where the output depends only on the past of the input. What I can do is work with a closed set of functions,

$$\{\phi_n(t)\} \quad (10.3)$$

depending only on the past. Since these are a closed set of functions, I can represent anything depending on the past in terms of the coefficients of these functions. Notice the sort of expressions that I can get. The constant presents no problem. For the linear term, when I have Expression 10.4, for all n , I know everything that is to be known linearly, if the ϕ_n 's are a closed set on the past.

$$\int \phi_n(t) dx(t, \alpha), \quad n = 1, 2, \dots \quad (10.4)$$

Therefore, $G_1[K_1(\tau), \alpha]$ in Expression 10.1 can be obtained by a linear combination of Expressions 10.4. Suppose that I go to the next order. I am using the generalization that, if a $K_n(\tau_1, \dots, \tau_n)$ is not symmetrical, I add on all permutations and make it symmetrical. Now, if I know

$$G_2[\phi_m(\tau_1) \phi_n(\tau_2), \alpha], \quad m, n = 1, 2, \dots \quad (10.5)$$

then I know everything about my quadratic terms, since I can express all the G_2 's as a linear combination of terms in Expression 10.5 due to the orthogonality development. That is, if the ϕ_n 's are a closed set, then the functions in Expression 10.5 or their symmetrizations are a closed set in two variables. Similarly, if I consider

$$G_3[\phi_{m_1}(\tau_1) \phi_{m_2}(\tau_2) \phi_{m_3}(\tau_3), \alpha], \quad m_1, m_2, m_3 = 1, 2, 3, \dots \quad (10.6)$$

then I have a development for the cubic term. I can get any cubic expression from Expression 10.6.

So the question comes of getting a suitable set of ϕ 's for the expansion of the K 's. If I can get such a suitable set of ϕ 's for the K 's, then I have a way of analyzing the output in terms of the input, and, as we shall see later, I have a way of synthesizing that output in terms of that input. So, we shall now discuss—and this goes back to some old work of Professor Y. W. Lee and myself—a particular set of orthogonal functions of the past in terms of which we can do this development. These are the so-called Laguerre functions. I want to discuss the Laguerre functions in two ways, mathematically and from an engineering viewpoint, i.e. how the Laguerre functions of the past of the input can be attained by instruments from the input. I shall then go into the matter of synthesis and analysis of electric networks.

The first thing that I want to do is to bring out some simple properties of the one-stage lattice depicted in Figure 10.2. Note that there is no difficulty working on open circuit, since there are instruments that are

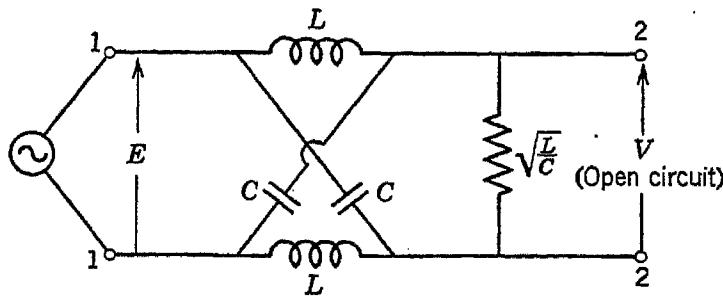


Figure 10.2

actuated by voltage without taking appreciable current, in particular, those employing cathode followers. The question is: What is the output of the simple lattice in terms of the input? We can determine this output in terms of frequency or in terms of time. It is quite clear that the output cannot depend on the future input; it cannot in any electric circuit. Now, if ω is the frequency,

$$\frac{V(\omega)}{E(\omega)} = \frac{1 - i\omega(LC)^{\frac{1}{2}}}{1 + i\omega(LC)^{\frac{1}{2}}} \quad (10.7)$$

The important thing to notice about Equation 10.7 is that the simple lattice is a phase shifter without being an amplitude shifter. Furthermore, the input impedance of the lattice is a pure resistance equal to $(L/C)^{\frac{1}{2}}$.

By the way, the compression of the apparatus that is about to be shown originates with Professor Lee. I think it came about in this way. I saw that electric circuits could realize impedances of this sort, and Professor Lee pointed out that these circuits would dovetail into one another,

so that the elements of one circuit could also be used for the next, allowing a much more compact and economical way of making this type of circuit. I think that is a fair account.

Now, consider two stages on open circuit, as in Figure 10.3.

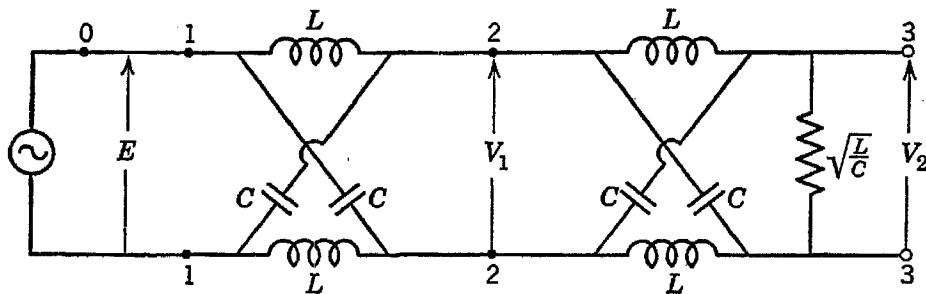


Figure 10.3

Note that for the first stage nothing has changed, since the second stage appears to the first stage to be identical to the original resistance, $(L/C)^{1/2}$. Thus,

$$\frac{V_1(\omega)}{E(\omega)} = \frac{1 - i\omega(LC)^{1/2}}{1 + i\omega(LC)^{1/2}} \quad (10.8)$$

$$\frac{V_2(\omega)}{E(\omega)} = \left[\frac{1 - i\omega(LC)^{1/2}}{1 + i\omega(LC)^{1/2}} \right]^2 \quad (10.9)$$

This process can be repeated as often as we want. We continue to replace the resistance by an additional stage, so that we can obtain a transfer characteristic of the form

$$\frac{V_n(\omega)}{E(\omega)} = \left[\frac{1 - i\omega(LC)^{1/2}}{1 + i\omega(LC)^{1/2}} \right]^n \quad (10.10)$$

It is possible to multiply all of the transfer characteristics by an additional factor, $1/[1 + i\omega(LC)^{1/2}]$, by simply placing an inductance of value L between the source and the first stage, i.e. between the terminals 0 and 1 in Figure 10.3. I now have a network that gives me a set of functions:

$$\frac{[1 - i\omega(LC)^{1/2}]^n}{[1 + i\omega(LC)^{1/2}]^{n+1}} \quad (10.11)$$

The important point that I am making is that we can simultaneously get any number of voltage transfer ratios of the form of Expression 10.11. In order to avoid carrying this confounded $(L/C)^{1/2}$ along, we shall just choose our units so that it is 1. This is essentially a choice of time units. Hence we get

$$\frac{(1 - i\omega)^n}{(1 + i\omega)^{n+1}}, \quad n \geq 0 \quad (10.12)$$

The question comes: What is the operator in time? In order to get that operator, we take the Fourier transform of Expression 10.13. Since the Fourier transform of an operator in frequency gives us an operator in time, we take

$$\int_{-\infty}^{\infty} \frac{(1 - i\omega)^n}{(1 + i\omega)^{n+1}} e^{i\omega t} d\omega \quad (10.13)$$

Instead of working this out, I want to give another approach, and then we shall compare them. I am turning things around and working with t positive rather than negative, but this can be turned around again equally easily. Now I say that the functions

$$t^n e^{-t}, \quad 0 \leq t < \infty, \quad n = 0, 1, 2, \dots \quad (10.14)$$

are all functions of L^2 , and it is not difficult to prove that they are a closed set of functions of L^2 . I may go into that later. Whether they are closed or not, being of L^2 , they can be normalized. Consider

$$\int_0^{\infty} (e^{-t})^2 dt = \frac{1}{2} \quad (10.15)$$

So, if I take $(2)^{\frac{1}{2}} e^{-t}$, I have a normalized function. Then I go through the orthogonalization process. I take

$$(at + b)e^{-t} \quad (10.16)$$

and submit it to two conditions: that it should be orthogonal to e^{-t} and that it should be normal. First, I obtain

$$\int_0^{\infty} (at + b)e^{-2t} dt = 0, \quad \text{giving } b = -\frac{a}{2} \quad (10.17)$$

Next, I obtain

$$a^2 \int_0^{\infty} (t - \frac{1}{2})^2 e^{-2t} dt = 1, \quad \text{giving } a = (\frac{4}{5})^{\frac{1}{2}} \quad (10.18)$$

Thus, $(\frac{4}{5})^{\frac{1}{2}}(t - \frac{1}{2})e^{-t}$ is normal and orthogonal to the first function. I keep this process up indefinitely. These functions will be closed over 0 to ∞ since they are obtained by a normalization of a closed set of functions. Therefore, any function can be represented in terms of them. They give me a complete representation of the present and past. Now, I say that these normal and orthogonal functions are just the functions given by Expression 10.13 except for a multiplicative constant that is different for each n . This can be shown. In other words, I get from the apparatus consisting of the cascade of simple lattices plus an initial series

inductance outputs of the form of Expression 10.4 when the input is shot noise. The ϕ_n 's are the Laguerre functions.

I obtain the appropriate G 's in the following way. First note that the G 's are functions only of the outputs of the Laguerre network and hence are operators depending only on the past. Consider

$$G_{\sum n_k} \left\{ \prod_k \phi_k^{n_k}, \alpha \right\} \quad (10.19)$$

Expression 10.19 gives me a closed basis for the G 's of any expression whatever, because I have a closed set of symmetrical K 's. Now, I have stated previously, and it is easy to verify, that, since the ϕ 's are orthogonal, Expression 10.19 can be rewritten as

$$G_{\sum n_k} \left\{ \prod_k \phi_k^{n_k}, \alpha \right\} = \prod_k G_{n_k} \{ \phi_k^{n_k}, \alpha \} \quad (10.20)$$

These are the so-called Hermite polynomials in the coefficients of the Laguerre functions of the past. I can obtain all terms in Equation 10.20 if I can do the following things: if I can obtain at each time the Laguerre coefficient of the past, if then I can obtain a polynomial in this which will be a Hermite polynomial, and if then I can multiply these polynomials with one another.

From the standpoint of electrical engineering, what must I do to accomplish this by apparatus? Well, one thing I must be able to do is to add voltages, but there is no problem about that. Since I am also obtaining polynomials, I must be able to multiply voltages. Then if I can add and multiply voltages, there is no problem in getting all these terms. How can I multiply voltages? That is easy. Suppose that I have one particular instrument which is a square-law rectifier. If I have a voltage $V(t)$, I want an apparatus that will give a voltage $V^2(t)$. That is, I want a nonlinear apparatus with that kind of characteristic. There are various ways of getting it, but the general principle involves using a push-pull. Suppose that I get the sort of characteristic shown in Figure 10.4, but its exact shape does not matter too much. Then, I can also get the characteristic

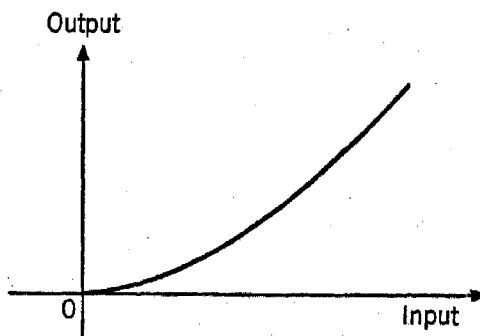


Figure 10.4

of Figure 10.5, which is the same characteristic as Figure 10.4 but reflected about the 0-input line. There is no problem in getting the rectifier to work the other way. So, a vacuum tube or a transistor can very easily

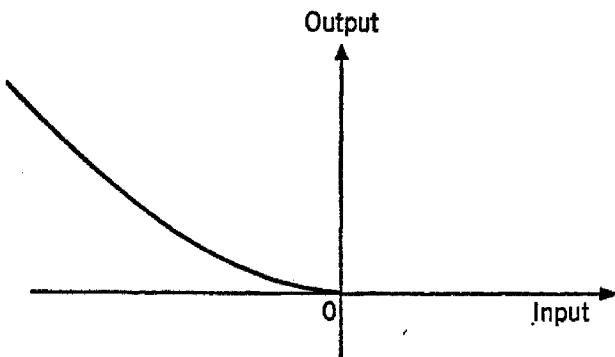


Figure 10.5

be made, by push-pull, to give the sort of characteristic shown in Figure 10.6. If I work on a narrow part of this curve, the narrower the better, and then amplify, I shall approach a square characteristic. So, square-

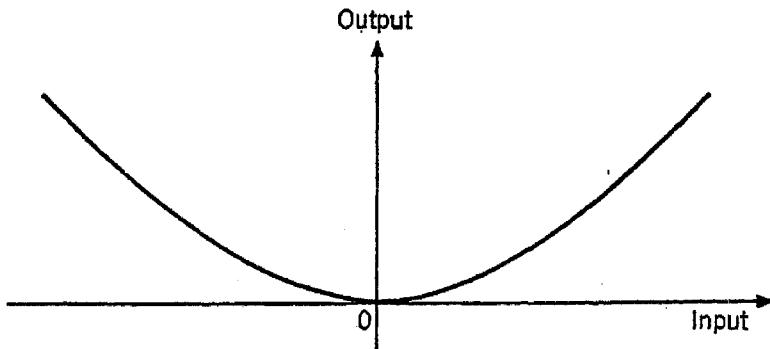


Figure 10.6

law rectifiers are technically not difficult to obtain. In order to multiply $V_1(t)$ and $V_2(t)$, I form $V_1(t) + V_2(t)$ and $V_1(t) - V_2(t)$ electrically. Then with the square-law device, I get $[V_1(t) + V_2(t)]^2$ and $[V_1(t) - V_2(t)]^2$. There is no trouble subtracting voltages, so I subtract and obtain $4V_1(t)V_2(t)$. Therefore, I can multiply voltages with existing apparatus. Then having multiplied voltages, I can multiply again. I not only can square but I can cube. I can get fourth powers. I can get polynomials of finite degree. Therefore, I can get Hermite polynomials. So there is no problem in getting Hermite polynomials and products of Hermite polynomials, since I can multiply the different Laguerre coefficients of the past. Since I can add up the different Hermite polynomials, I can synthesize anything that can be synthesized out of my symmetric functions; i.e., I can approximate as closely as I

want any function of the past that moves along with a given function. In other words, I can synthesize any nonlinear characteristic that is independent of the remote past and is L^2 . Therefore, I have here a universal method for synthesizing.

Now the question is: Can I analyze? Knowing what a function is, how can I determine the coefficients in my development? How can I express the thing in the Hermite polynomials in the coefficients of the Laguerre polynomials, and write down the characteristics of the system? The important point is that if I can do this for nonlinear circuits, I have done the equivalent of what I do regularly by impedances for linear circuits. That is, if I express how the circuit will respond to any Brownian input, I know how it will respond to any input. Later I shall go into that more.

Now, at least, I want to state what I am doing. I have given a complete characterization of a circuit. I have analyzed it. The important thing to realize is that impedance or voltage ratio is not an adequate way of characterizing nonlinear circuits. Many years ago, nearly forty years, I was working with Professor Vannevar Bush on these problems, and we tried to put the question: How can we define the impedance of a nonlinear circuit? The answer is that we do not want the impedance of a nonlinear circuit. The impedance is not an adequate characterization of the behavior of the circuit; that is, the behavior of a circuit cannot be given by considering merely a trigonometric output. What we want is the behavior of the circuit with a shot-effect input, and then we shall have a theory that is equivalent, parallel but not identical, to impedance theory.

I shall be able to get my coefficients in terms of averages. These averages will be averages involving α . What I can get easily enough in a circuit is a time average. I am going to go into the details of that in the next lecture. The time average is, in general, not the same as the α average, but under the ergodic hypothesis time averages will almost always be α averages. Therefore, if we have ergodicity, this will work. Furthermore, the ergodic hypothesis is satisfied by the Brownian motion. That is, if we translate a Brownian motion in time, then that transformation gives a measure-preserving transformation of α , and it can be shown that this measure-preserving transformation has no invariant sets that have a measure if this measure is other than 0 or 1. In other words, by taking a particular shot-effect input and obtaining time averages, we can obtain α averages that will be what we need for obtaining coefficients. There are several questions here: There is the question of how to relate the α averages to the coefficients, and there is the question of how actually to get α averages. I shall discuss this in the next lecture. We have nearly finished the synthesis problem for nonlinear circuits, and we shall take up the analysis problem in the next lecture.

Nonlinear Systems—II

We now come to the problem of analyzing a system. Consider two black boxes, one of which has known characteristics, the other being the unknown. One of the things that I can do is to interconnect these boxes in the way shown in Figure 11.1, with their inputs paralleled across a

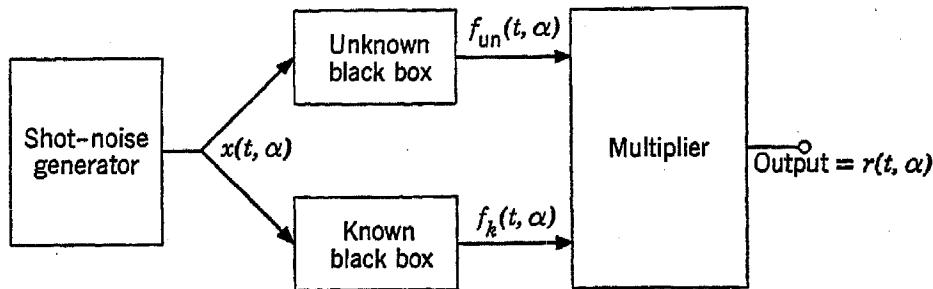


Figure 11.1

shot-effect generator. We then multiply their outputs—we suppose that we have multipliers—to obtain the system output $r(t, \alpha)$.

The unknown black box will have an output

$$f_{un}(t, \alpha) = \sum_n G_n[K_n(t + \tau_1, \dots, t + \tau_n), \alpha] \quad (11.1)$$

The other box will have an output

$$f_k(t, \alpha) = \sum_n G_n[H_n(t + \tau_1, \dots, t + \tau_n), \alpha] \quad (11.2)$$

Then, the product output of the system will be

$$\begin{aligned} r(t, \alpha) = & \sum_m \sum_n G_m[K_m(t + \tau_1, \dots, t + \tau_m), \alpha] \\ & \times G_n[H_n(t + \tau_1, \dots, t + \tau_n), \alpha] \end{aligned} \quad (11.3)$$

If I consider the time average of this product, it will be

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_T^T r(t, \alpha) dt, \quad \text{for almost all } \alpha \quad (11.4)$$

But we know that the transformation we get by adding t to all of the Brownian motions is a measure-preserving transformation of the Brownian motions, and this measure-preserving transformation is an ergodic transformation. We have already seen that. Therefore, Expression 11.4 is equal to

$$\int_0^1 d\alpha \sum_m \sum_n G_m[K_m(\tau_1, \dots, \tau_m), \alpha] G_n[H_n(\tau_1, \dots, \tau_n), \alpha] \quad (11.5)$$

That is the result of the ergodic theorem. For almost all α , the phase average is the same as the time average.

We now bring up material that we have already covered. When the m 's and the n 's are different, the integration with respect to α gives 0. When the m 's and n 's are the same, we get

$$n! \int \cdots \int K_n(\tau_1, \dots, \tau_n) H_n(\tau_1, \dots, \tau_n) d\tau_1 \cdots d\tau_n \quad (11.6)$$

Thus, if we read the average value of $r(t, \alpha)$, we obtain

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{2T} \int_T^T r(t, \alpha) dt \\ &= \sum_n n! \int \cdots \int K_n(\tau_1, \dots, \tau_n) H_n(\tau_1, \dots, \tau_n) d\tau_1 \cdots d\tau_n \end{aligned} \quad (11.7)$$

Since the known box can be made anything we want, we can, in particular, choose it to be a "one-term" box. We can make that particular H

$$H_n(\tau_1, \dots, \tau_n) = \prod_{i=1}^n \phi_k(\tau_i) \quad (11.8)$$

where the ϕ_k 's are the Laguerre functions. If we have a set of H 's of this sort, which correspond to the product of Laguerre functions, the average output will be, except for the $n!$, the K_n times the known Laguerre function. In that way we can get the development of K_n in terms of products

of Laguerre functions, and we have the material to resynthesize the unknown box. That means we have an analysis of the system and a synthesis procedure.

There are one or two little tricks that we can do in getting our average. I shall not go into the details here, but it turns out that if we take some random quantity of the sort shown in Figure 11.2 and average it over

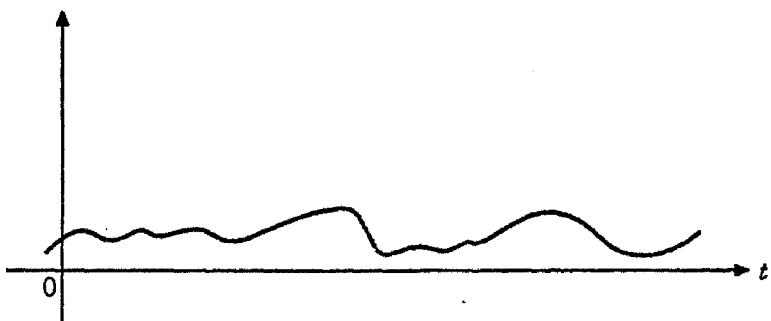


Figure 11.2

a long period with the weighting of Figure 11.3, or a weighting that falls off exponentially, as shown in Figure 11.4, the averages will almost always be the same. I can go through the computations, but that can



Figure 11.3

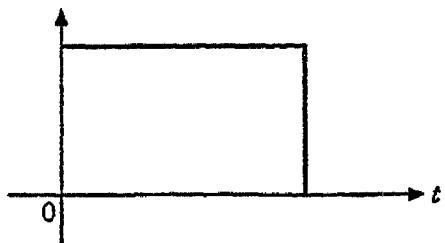


Figure 11.4

be easily checked. In other words, we may compute the time average in Equation 11.7, using an exponential damping of the past.

We may do the multiplication and averaging of $f_{un}(t, \alpha)$ and $f_k(t, \alpha)$ electrically, as we do all the others, but one of the ways of obtaining the average product with a weighting function is to use an electrodynamometer. Such an electrodynamometer is shown in Figure 11.5. The electrical torque on the dynamometer is proportional to the product of the input voltages V_{un} and V_k .

But now suppose that we mount on the shaft of the electrodynamometer a copper disc, and put an electromagnet across that disc, as shown in Figure 11.6. The electromagnet is supplied from an external source. This will give us an exponentially damped average of the past. The heavier the damping, the farther the averaging extends into the past.

We see what will happen. The damping will prevent the shaft from turning easily. The long-term effect of this is to average, with a weighting function, the previous shaft positions. Then, as we increase the damping by letting the current in the electromagnet approach infinity, the shaft position asymptotically approaches the average of the product of V_{un} and V_k .

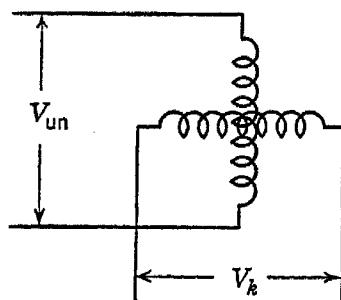


Figure 11.5

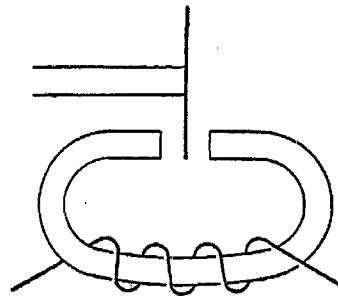


Figure 11.6

Using this instrumentation, we obtain a reading on the electrodynamometer of the average of the product for a given test box, and that gives us the integral in Equation 11.7 with probability 1, which is as good a value as we can get in engineering. If we have a Hermitian polynomial in the Laguerre function box and an unknown box, connect them in this manner, read the stationary position for the product, and divide by $n!$, we obtain the integral of the product of the Hermitian polynomial of degree n and K_n . This gives us the development of K_n in terms of sums of products of n th-order Hermitian polynomials.

In other words, this procedure allows us to use the synthesis apparatus, of which we spoke in the previous lecture, as an analysis apparatus. All we need for the test boxes are Laguerre function boxes, which I have told how to get, since I have told how to make polynomials from them and how to multiply these polynomials. The necessary test boxes are obtained from an assembly of Laguerre function boxes and multipliers. We put in a given unknown box, put in a Brownian input, read the deflection under damping, and thereby obtain the particular coefficients for it. With those coefficients we can assemble the terms and reconstruct the apparatus.

So, what I have given is both an analysis and a synthesis theory of nonlinear apparatus. I gave the synthesis first, but, given the synthesis, I have shown how to get an analysis and remake the apparatus synthetically of these Laguerre and Hermite expressions.

Coding

Now I am going on to another problem, and that problem is discrete coding. I am going to consider a time series that, unlike the Brownian motion series, is given only at discrete times. I am going to consider a function f which is real and Lebesgue integrable. It does not, at this point, have to be L^2 , although I shall restrict it to being L^2 later in the discussion.

I consider $f(\alpha)$, but I also consider a measure-preserving transformation T and $f(T^n\alpha)$. Now I consider the sequence $f(\alpha)$ and its past, $f(T^{-1}\alpha)$, $f(T^{-2}\alpha)$, and so forth. I shall assume that we do not have complete dependency on the past. As a matter of fact, we permit much less than complete dependency on the past.

By the way, it will be convenient for me to change from my original α to an α that is completely determined by the sequence of values $f(T^{-n}\alpha)$. There is no problem in doing this. It is the sort of thing that we have done before with the Brownian motion. It is a similar mapping procedure. I introduce a new variable α .

I take $f(\alpha)$, and I first take the entire line of, say, α_1 . There are, first, two possibilities: $f(\alpha) > 0$ or $f(\alpha) < 0$, drawn in Figures 12.1 and 12.2, respectively. Each of these will determine a certain contingency. I shall map these on the two parts of the α_1 line, as shown in Figure 12.3. Then, I shall do the same sort of thing as I did before. I divide the range of $f(\alpha)$ into 4 parts, and I also divide $f(T^{-1}\alpha)$ into 4 parts. This gives me 16 contingencies, as sketched in Figure 12.4. I am assuming that these will be measurable sets, so I get 16 probabilities. Then, I divide my ranges finer and finer, while adding $f(T^{-2}\alpha)$, $f(T^{-3}\alpha)$, exactly as I did

in the Brownian motion case. In this way I close down on the α_1 that represents f and its past at a discrete number of points. Then I can represent the distribution of $f(\alpha), f(T^{-1}\alpha), \dots$ in terms of α_1 .

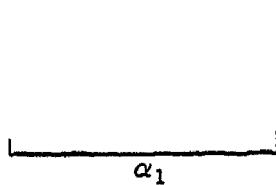


Figure 12.1

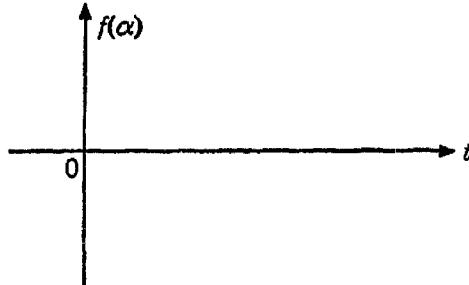


Figure 12.2

This is a good representation because nothing else interests us in α but this discrete set of values. An α , in general, may not be completely determined by the f 's, but I have a new α uniquely determined by them.

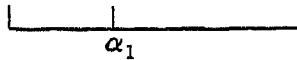


Figure 12.3

So then, I have now remapped the discrete time sequence as $\phi(\alpha_1), \phi(T^{-1}\alpha_1), \dots$. Not only do I obtain that sequence, but I also obtain a conditional distribution of $\phi(\alpha_1)$ when the past is known, as

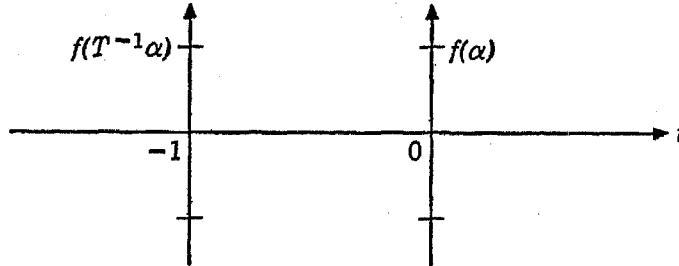


Figure 12.4

can be seen in the following manner. I put $f(\alpha)$ in a box that is on some measurable set of points. In addition, I now close down on $f(T^{-1}\alpha), f(T^{-2}\alpha), \dots$, just as I did before. Well, to say that $f(T^{-1}\alpha), f(T^{-2}\alpha), \dots$ lie in a region and that $f(\alpha)$ lies in a region S gives me a region common to all of these events, which will have a measure. This will give me the conditional distribution when I close down in terms of my conditions on the past of $f(\alpha)$. It can be shown by mathematical argument that for almost all points in the past there will be a definite conditional distribution of f in terms of the past.

So, I have a conditional distribution for $f(\alpha)$, when the past is given, for almost all pasts of $f(\alpha)$. That will be an increasing function. The probability that, given the past, $f(\alpha)$ is less than a certain number will increase as the number increases, as shown in Figure 12.5.

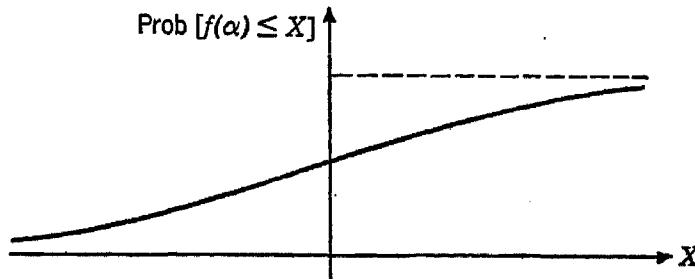


Figure 12.5

When we consider an increasing function, it is well known that it may be made up of three parts. One part has a denumerable number of jumps. That is called the discontinuous part of the increasing function, as illustrated in Figure 12.6. Another part of the function can be made

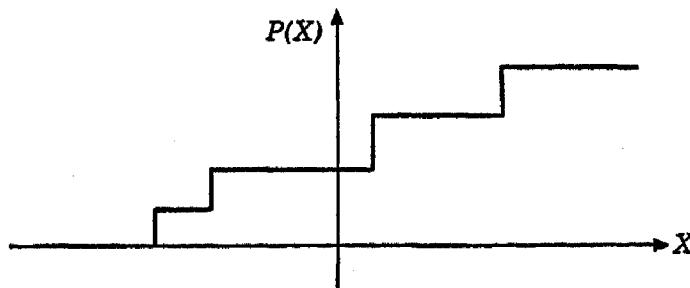


Figure 12.6

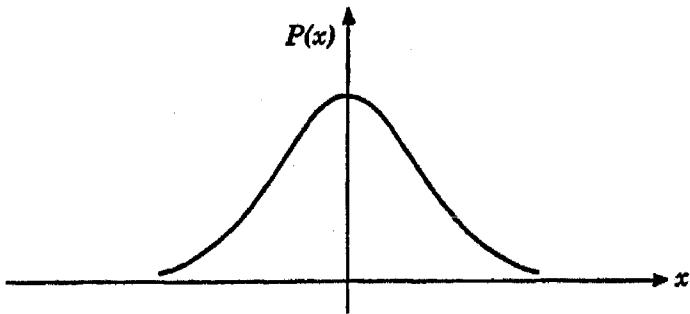


Figure 12.7

up of the integral of a positive expression, such as shown in Figure 12.7. Such a function, an example of which is shown in Figure 12.8, is called the absolutely increasing part of the increasing function.

However, the general increasing function contains a third part, and I shall give you an example that is continuous but that has its climb

over a set of 0 measure. I give you the example so that you can see what the possible cases are. It will be a continuous function that does not possess a derivative which can represent a density function.

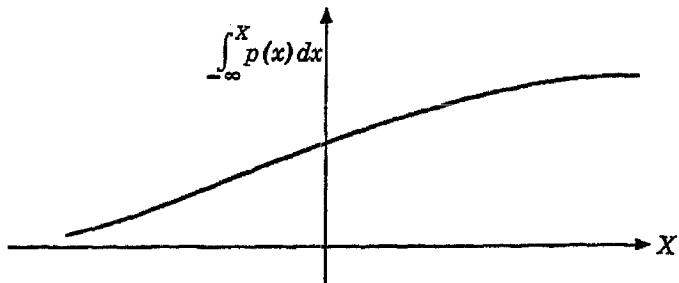


Figure 12.8

In order to construct such a function, suppose that I take the set from 0 to 1. I take the middle third out of that, from $\frac{1}{3}$ to $\frac{2}{3}$, and I assume that I have a function which is $\frac{1}{2}$ over that region. I consider the first third, take the middle third of it, and call the function $\frac{1}{4}$ over that region. I take the middle third of the last third, and make the function $\frac{3}{4}$ over it. I then take the middle thirds of the remaining intervals and make the function $\frac{1}{8}, \frac{3}{8}, \frac{5}{8}$, and $\frac{7}{8}$ over them, as shown in Figure 12.9.

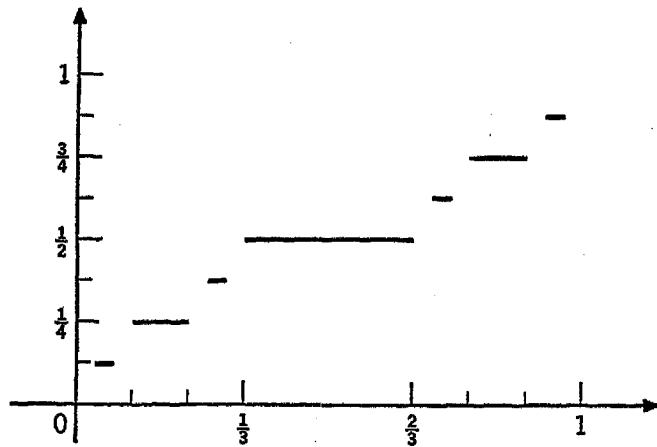


Figure 12.9

I can continue filling in these lines. We notice that any vertical gap is going to be filled in, sooner or later. The function will, therefore, be continuous and increasing. However, the total climb is over a set of zero measure. Notice that there is no climb over the middle third. There is no climb over the middle thirds of the next two thirds. When I add these regions of zero climb, I obtain

$$\frac{1}{3} + 2\left(\frac{1}{9}\right) + 4\left(\frac{1}{27}\right) + \cdots = 1 \quad (12.1)$$

which is the entire line length. So, this is a function that has no climbing over a set of points of the density of the line. This is a nondifferentiable, continuous function.

Next I am going to consider a continuous function that extends from $-\infty$ to ∞ , and which may or may not be differentiable. I shall require one extra condition on it, which will prevent the existence of a past with this behavior. I shall assume that the function is always *properly* increasing (i.e., the "middle-third" function is not properly increasing, nor is the function shown in Figure 12.10). I am saying nothing about the

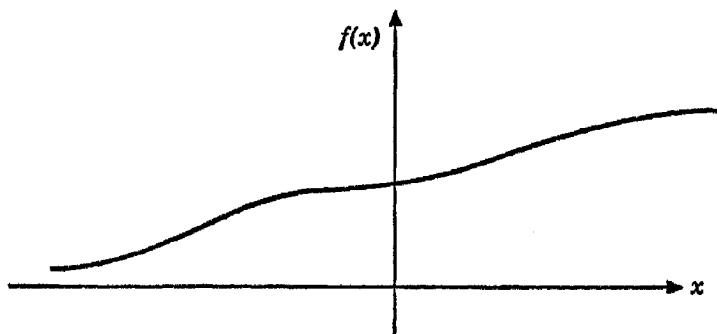


Figure 12.10

derivative. The derivative may be 0 for a while; that does not bother me. The function may have the same height at two different points. Notice that, even if I have a function like the one in Figure 12.9, when I add a properly increasing continuous function to it, I get a properly increasing continuous function.

I am assuming that the conditional distribution function for $f(\alpha)$, given $f(T^{-1}\alpha), f(T^{-2}\alpha), \dots$, is, for almost all pasts of $f(\alpha)$, properly increasing. Under that condition, given the past, I get an increasing function that starts at 0 for $f(\alpha) = -\infty$ and increases to 1 for $f(\alpha) = \infty$. In addition, this function, shown in Figure 12.11, depends upon the past.

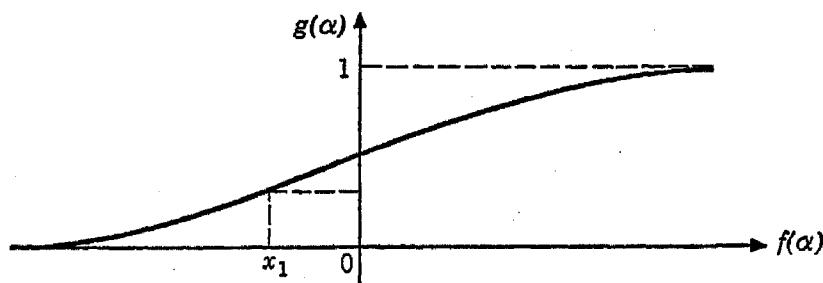


Figure 12.11

Now I can rescale the function separately for each past. If I plot the conditional distribution against the area A under the conditional density

curve, Figure 12.12, rather than against $f(\alpha)$, I shall obtain a function that goes from 0 to 1 at 45° , as shown in Figure 12.13. The distribution function depends upon the past of $f(\alpha)$, and hence, depends upon the value of α for the particular distribution. I shall call it $g(\alpha)$.

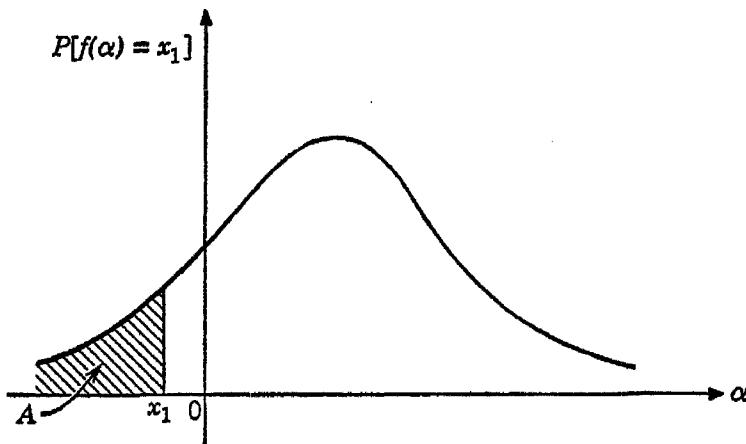


Figure 12.12

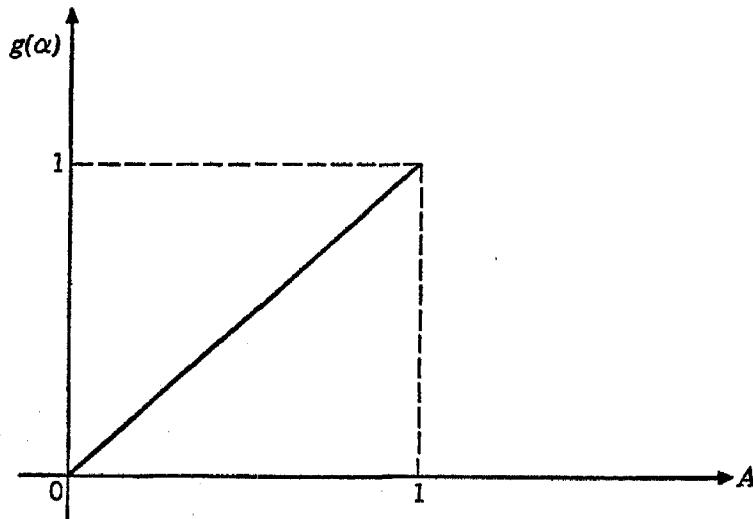


Figure 12.13

By the way in which I have obtained it, $g(\alpha)$ is uniformly distributed between 0 and 1, no matter what the past is. This is because I have made the normalization dependent upon the past. The distribution is as shown in Figure 12.14, independent of the past. Now consider $g(T^{-1}\alpha)$. The function $g(\alpha)$ is dependent upon $f(\alpha)$ and its past. The function $g(T^{-1}\alpha)$ is dependent upon $f(T^{-1}\alpha)$ and its past; $g(T^{-1}\alpha)$ is also uniformly distributed between 0 and 1. No matter what you can say about it, no matter what value it takes on, it does not affect the distribution of $g(\alpha)$. Therefore, $g(\alpha)$ and $g(T^{-1}\alpha)$ are, under this assumption, independent of one another. No matter what range we give the one, the other is still uniformly distributed between 0 and 1.

Similarly, by an easy extension the functions $g(\alpha), g(T^{-1}\alpha), g(T^{-2}\alpha), \dots$ are uniformly distributed and independent of one another. Here we get a set of variables that vary independently from 0 to 1. These represent, in a nonlinear sense, what is new about the f 's.

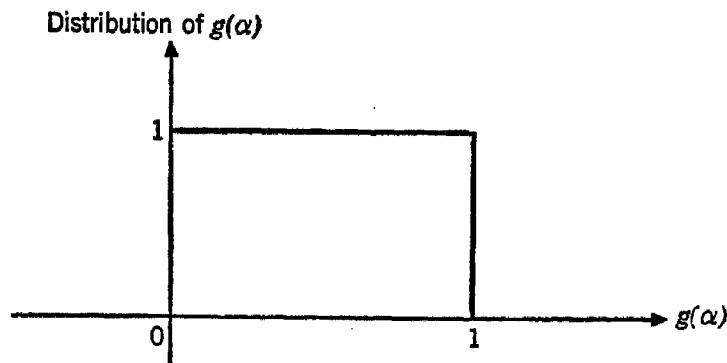


Figure 12.14

Now, the g 's can be determined in terms of the f 's of the past. Under what condition can f be determined in terms of the g 's of the past? Let me say this; the g 's have been coded in terms of f in the past. Can we decide the g 's to obtain the f 's? If we can do that, we have given a way to make the discrete set of messages, satisfying the conditions we have given, depend upon a discrete set of completely independent messages that lie in $[0, 1]$. This is the essential coding problem.

Notice the following thing. Because the distribution functions are properly increasing, when I know $f(T^{-1}\alpha), f(T^{-2}\alpha), \dots$, and $g(\alpha)$, I can express $f(\alpha)$ in terms of them. Knowing $f(T^{-1}\alpha), f(T^{-2}\alpha), \dots$ determines the distribution for $g(\alpha)$, and then the particular value of $g(\alpha)$ determines the value of $f(\alpha)$. So I have a conditional coding of f , not yet in terms of the g 's, but in terms of the first g and the rest of the f 's.

That being the case, I can code $f(T^{-1}\alpha)$ in terms of $g(T^{-1}\alpha)$ and the earlier f 's. It follows that I can code $f(\alpha)$ in terms of $g(\alpha), g(T^{-1}\alpha), g(T^{-2}\alpha), f(T^{-3}\alpha), \dots$. I keep this up, and I have a coding of $f(\alpha)$ in terms of $g(\alpha), g(T^{-1}\alpha), g(T^{-2}\alpha), \dots, g(T^{-n}\alpha), f[T^{-(n+1)}\alpha], f[T^{-(n+2)}\alpha], \dots$.

I want an assumption that will allow this coding, which we have made in terms of the g 's and the past beyond a certain point, to be made in terms of the g 's alone. The first thing that I shall do is to state that assumption. Later on we shall discuss whether or not this assumption is the natural one.

Let me consider α . This is essentially α_1 , because I want α to depend only on the f 's. I define measurable sets of points, belonging to a certain class, that depend only on $f(T^{-n}\alpha), f[T^{-(n+1)}\alpha], f[T^{-(n+2)}\alpha], \dots$. These are measurable sets of points that can be given by statements

about $f(T^{-n}\alpha)$, $f(T^{-(n+1)}\alpha)$, \dots . I shall say they belong to a class C_n . Any measurable set of points belonging to C_{n+1} will automatically belong to C_n because a statement involving terms only beyond the $(n+1)$ th also involves a statement about terms only beyond the $(n+1)$ th. If I consider the class of sets that can be measured in this way, they are, essentially, the Borel fields defined by the various classes. These are the

$$C_n \supset C_{n+1} \supset C_{n+2} \supset C_{n+3} \dots \quad (12.2)$$

classes of sets that can be defined by the remote past.

There will be certain sets S that belong to all the C_n 's. They can be defined by as remote a past as I want. Those sets will belong to C_∞ . C_n represents those sets of points definable by statements involving only the f 's from the n th onward. C_∞ can be defined this way no matter how large n is.

Now, the assumption that I make is concerned with C_∞ , and it is closely related to the ergodic assumption. We cannot say that there are no classes in C_∞ . The whole universe involves no restrictions, and, therefore, can be defined by the whole universe in one state, the whole universe in an earlier state, and so forth. The assumption that I am making is this:

If S is measurable and belongs to C_∞ , then the measure of S is either 0 or 1.

We are saying that the only measurable sets that can be defined by the arbitrarily remote past are either of the measure of the whole universe or of 0 measure. Notice how closely this is related to the ergodic hypothesis. The measure 1 or 0 assumption appears again.

If this hypothesis is true, then $f(\alpha)$ is determined completely by $g(\alpha)$, $g(T^{-1}\alpha)$, $g(T^{-2}\alpha)$, \dots , without any remote past, except over a set of cases of probability 0. What we are assuming is that the remote past is ineffective in determining $f(\alpha)$, that any statement based on the remote past is almost always or almost never true for a measurable set, or, alternatively, that any measurable set defined by the remote past is of measure either 1 or 0.

Then we can prove that when we know the g 's we have our $f(\alpha)$, in such a way that, knowing the distribution of the g 's, we have the distribution of the f 's. We can code the f 's in terms of g 's of the present and the past. I shall take this up in the next lecture, but the result we get is that we have not only coded our message into a set of independently varying quantities, uniformly varying between 0 and 1, but we can decode our message, a result that is, of course, important.

There is no use in coding a message if you cannot decode it, but under our assumption decoding is possible. That is the subject on which I am going to spend part of the next lecture. There is one thing, though, that

is simple enough for me to say today, as a lead. We have coded f into a message that is uniformly distributed over the range $[0, 1]$, as in Figure 12.14. Let us take this distribution about the middle and change it by a change in horizontal scale, as shown in Figure 12.15. Lo and

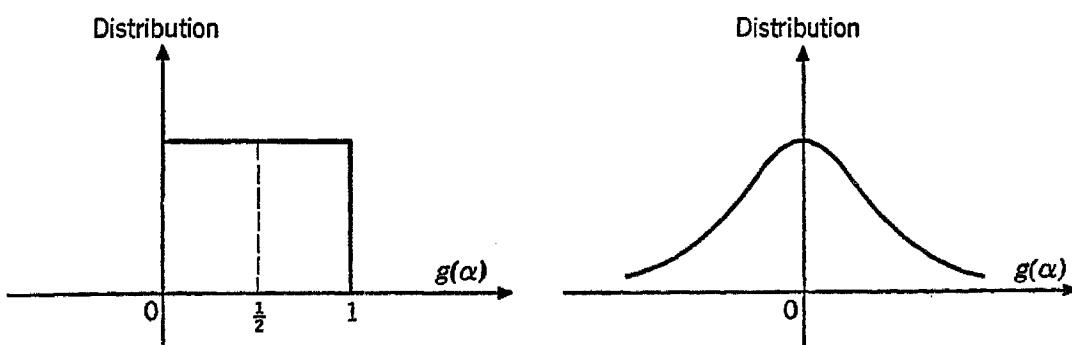


Figure 12.15

behold, I can obtain the Gaussian distribution by a proper scale change. Therefore, if we can prove that we can code the f 's into the g 's with distributions which are uniform, we can code them into independent Gaussian distributions. This we shall discuss in the next lecture, and this allows us, in the coding problem, to get a discrete analogue of the Brownian motion theory that we have already presented.

In the next lecture I shall go into decoding, tighten up the argument, then go on to the Gaussian coding. I shall then show that the orthogonal development theory which we have given in the Brownian case will also work here, and this gives us a way of expressing the message in terms of fundamental, independently distributed Gaussian quantities as a series in Hermite polynomials.

Decoding

I wish to amplify the previous lecture a little. We started with

$$f(\alpha), f(T^{-1}\alpha), f(T^{-2}\alpha), \dots \quad (13.1)$$

and from this we obtained an expression

$$g(\alpha), g(T^{-1}\alpha), g(T^{-2}\alpha), \dots \quad (13.2)$$

where $g(\alpha)$ depended on the past and present of $f(\alpha)$. These g 's were independent and distributed uniformly between 0 and 1. The question is now: Can we go back from the distribution of the g 's to the distribution of the f 's?

We can express $f(\alpha)$ in terms of

$$g(\alpha), g(T^{-1}\alpha), \dots, g(T^{-n}\alpha), f(T^{-n-1}\alpha), f(T^{-n-2}\alpha), \dots \quad (13.3)$$

We want to consider some statements about the f 's. That is, I have a statement about

$$f(\alpha), f(T^{-1}\alpha), \dots, f(T^{-k}\alpha) \quad (13.4)$$

These occupy a certain region of k space. Let its characteristic function, measurable in α , be $S(\alpha)$. We wish to know if we can express the measure of this function of α by a statement concerning

$$g(\alpha), g(T^{-1}\alpha), g(T^{-2}\alpha), \dots \quad (13.5)$$

We can certainly make it by a statement concerning all of Expression 13.3, because every one of the members of Expression 13.4 can be expressed

in terms of all of Expression 13.3. Remember that we are choosing the representation of α completely determined by the f 's. Thus, a statement about $S(\alpha)$ can be made by statements involving a finite number of g 's and the f 's beyond a certain point. The latter set can be approximated by a set determined by all of the g 's and the f 's beyond any point. That is, we can represent each of the finite number of f 's in Expression 13.4 in terms of $g(\alpha)$, $g(T^{-1}\alpha)$, and so on, and the infinite past of f .

Any statement, even a conditional statement, made about the infinite past of f , and giving a measurable set of f is a statement that is almost always true or almost always false. Therefore, if I leave out statements that are almost always true or almost always false, I can say anything I want to about the distribution of any number of the f 's in terms of the g 's alone. The infinite past will not be effective because any class given to that infinite past will have measure either 1 or 0. Remember that sets of measure either 1 or 0 have no effect on distributions. Therefore, I can give the distribution of any number of f 's completely in terms of the g 's. That brings us to where we stopped in the previous lecture.

What I shall do then is to go from the set $g(\alpha)$ to the set $h(\alpha)$, which is obtained merely by a remapping of $g(\alpha)$, as shown in Figure 13.1. The $\frac{1}{2}$ point of the $g(\alpha)$ curve goes into the center of the $h(\alpha)$ curve. We map

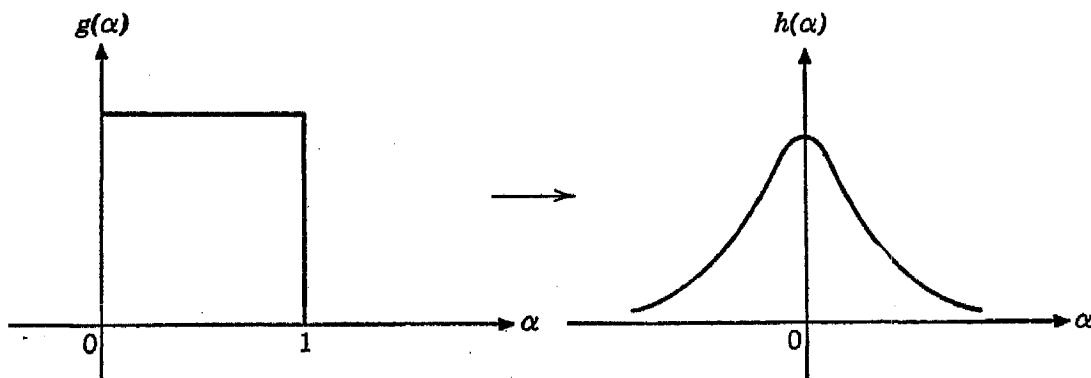


Figure 13.1

equal areas under the Gaussian curve $h(\alpha)$ into equal areas under the curve $g(\alpha)$. In this way, we have an infinite number of independent, Gaussianly distributed quantities:

$$h(\alpha), h(T^{-1}\alpha), h(T^{-2}\alpha), \dots \quad (13.6)$$

Then we know everything that we want to know about the distribution of the f 's, providing $f(\alpha)$ is a function of the infinitely many variables of Expression 13.6. I have expressed the original time series in terms of the Gaussianly distributed time series. Let me make the assumption that $f(\alpha)$ belongs to L^2 . Then I can express $f(\alpha)$ in terms of a set of

independent, Gaussianly distributed variables:

$$f(\alpha) = F[h(\alpha), h(T^{-1}\alpha), \dots] \quad (13.7)$$

In other words, I have done something exactly analogous in the discrete case to what I have done in the continuous case. In the continuous case, I describe the time series in terms of a Brownian motion. Here I am describing the discrete time series in what corresponds to the rise and fall of the Brownian motion in the unit interval. The amplitudes of these rises and falls will also be Gaussianly distributed and independent for independent intervals. This being the case, I want to point out that the entire theory that we have already developed for orthogonal polynomial functionals will apply quite as well in the discrete case as in the continuous case.

Let us use the following variables:

$$u_0(\alpha), u_1(\alpha), u_2(\alpha), \dots \quad (13.8)$$

These variables all depend on α . Now, as we see, they are first degree in the Brownian motion; that is, they correspond in degree to the Brownian motion. What I am saying is that I can develop any function of α in terms of constants, first-degree expressions, second-degree expressions, third-degree expressions, and so on, in these expressions exactly as I could in the continuous case. It goes just the same way. First, the function K_0 is a constant that obviously belongs to L^2 . Suppose that I consider

$$\sum_0^\infty a_n u_n(\alpha) \quad (13.9)$$

If we consider something depending only on the past, n will go from 0 to ∞ . If it is to depend on the past and future, n goes from $-\infty$ to ∞ . The mean of Expression 13.9 is

$$\int_0^1 d\alpha \sum_0^\infty a_n u_n(\alpha) = 0 \quad (13.10)$$

Expression 13.9 will be orthogonal to any constant.

Consider the average value of the square of Expression 13.9:

$$\begin{aligned} & \int_0^1 d\alpha \left[\sum_0^\infty a_n u_n(\alpha) \right]^2 \\ &= \sum_0^\infty a_n^2 \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^\infty \exp\left(-\frac{u_n^2}{2}\right) u_n^2 du_n = \sum_0^\infty a_n^2 \end{aligned} \quad (13.11)$$

The cross terms are zero because they contain the product of two independent, zero-mean, Gaussian distributions.

We first start with a finite sum

$$\sum_0^N a_n u_n(\alpha) \quad (13.12)$$

We then have the same sort of an argument as we had in the continuous case. By the Riesz-Fischer theorem, if

$$\sum_0^\infty a_n^2 \text{ converges} \quad (13.13)$$

then Expression 13.12 converges in the mean. That is, the integral of the square of the difference between the n th partial sum and the m th partial sum goes to 0. Therefore, the limit in the mean of f exists, and I shall call

$$\lim_{N \rightarrow \infty} \sum_0^N a_n u_n(\alpha) = \sum_0^\infty a_n u_n(\alpha) \quad (13.14)$$

where the right-hand expression is now well defined. This works not only for 0 to ∞ but also for $-\infty$ to ∞ . I define the function of α in L^2 as in Equation 13.14. We see that this is exactly analogous to what we have done in the continuous case. We shall go ahead with this idea. I think that this is a good exercise to see what we were doing in the continuous case, because in the discrete case it is a little easier to see.

Let us now take a quadratic expression

$$\sum \sum a_{mn} u_m(\alpha) u_n(\alpha) \quad (13.15)$$

That is a homogeneous quadratic expression. Take Expression 13.15 as a finite sum to begin with. We want to discuss the average of Expression 13.15 and the average of the square of Expression 13.15. There is no restriction in taking a_{mn} symmetrical, because we would get exactly the same thing if we replaced a_{mn} by

$$\frac{a_{mn}}{2} + \frac{a_{nm}}{2} \quad (13.16)$$

If we consider Expression 13.15 as a finite sum, its average is then given by

$$\int_0^1 d\alpha \sum \sum a_{mn} u_m(\alpha) u_n(\alpha) = \sum a_{mm} \quad (13.17)$$

The cross terms again come out. Therefore, Expression 13.18 is orthogonal to any constant:

$$\sum_m \sum_n a_{mn} u_m(\alpha) u_n(\alpha) - \sum_m a_{mm} \quad (13.18)$$

Expression 13.18 is also orthogonal to every linear expression, for the following reason. The product of a linear term with Expression 13.8 will contain first-degree and third-degree terms. But terms of odd degree cannot be grouped as products of pairs, and hence, the average value of terms of odd degree is 0. Therefore, Expression 13.18 is orthogonal to every first-degree and second-degree expression exactly as before.

Now let us compute the average value of the square of Expression 13.18:

$$\begin{aligned} & \int_0^1 d\alpha \left[\sum_m \sum_n a_{mn} u_m(\alpha) u_n(\alpha) - \sum_m a_{mm} \right]^2 \\ &= \int_0^1 d\alpha \left[\sum_m \sum_n \sum_p \sum_q a_{mn} a_{pq} u_m(\alpha) u_n(\alpha) u_p(\alpha) u_q(\alpha) \right. \\ &\quad \left. - 2 \sum_m \sum_n \sum_p a_{mn} a_{pp} u_m(\alpha) u_n(\alpha) + \sum_m \sum_p a_{mm} a_{pp} \right] \quad (13.19) \end{aligned}$$

Consider the first term on the right-hand side of Equation 13.19. The only nonzero cases occur when (1) all of the subscripts are the same, or (2) when some two are the same and the other two are the same. First we shall take the case when all are the same. Since we have the fourth power of $u_m(\alpha)$, we get

$$3 \sum_m a_{mm}^2 \quad (13.20)$$

The 3 comes from the number of ways of grouping 4 elements in pairs. Another way of getting the number is to evaluate the integral directly:

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \int u^4 \exp\left(-\frac{u^2}{2}\right) du = 3 \quad (13.21)$$

Now we consider the possibility that $m = n$, $p = q$, but $m \neq p$. Note that there is only one way of doing this pairing. The prime on the summation sign indicates that the variables of summation are never equal,

$$\sum'_{m,p} a_{mm} a_{pp} \quad (13.22)$$

The next two possibilities are (1) $m = p$, $n = q$, but $m \neq n$; or (2) $m = q$, $n = p$, but $m \neq n$. The two cases are exactly the same in structure. Hence, we use twice the value that we would get from one of the possibilities

$$2 \sum'_{m,n} a_{mn}^2 \quad (13.23)$$

We have now evaluated the first term of the right-hand side of Equation 13.19. The second term of the right-hand side of Equation 13.19

has nonzero values only when $m = n$. That second term then becomes

$$-2 \sum_m \sum_p a_{mm} a_{pp} \quad (13.24)$$

The complete right-hand side of Equation 13.19 becomes

$$\begin{aligned} 3 \sum_m a_{mm}^2 + \sum'_{m,p} a_{mm} a_{pp} + 2 \sum'_{m,n} a_{mn}^2 \\ - 2 \sum_{m,p} a_{mm} a_{pp} + \sum_{m,p} a_{mm} a_{pp} \end{aligned} \quad (13.25)$$

The last two terms can be combined. We can cancel more terms by noting that

$$\sum_m a_{mm}^2 + \sum'_{m,p} a_{mm} a_{pp} = \sum_m \sum_p a_{mm} a_{pp} \quad (13.26)$$

Expression 13.25 becomes

$$2 \sum_m a_{mm}^2 + 2 \sum'_{m,n} a_{mn}^2 \quad (13.27)$$

The first term of Expression 13.27 is the sum on m and n when m and n are equal; the second term is the sum when m and n are different. Expression 13.27 becomes

$$2 \sum_m \sum_n a_{mn}^2 \quad (13.28)$$

Therefore

$$\int_0^1 d\alpha \left[\sum_m \sum_n a_{mn} u_m(\alpha) u_n(\alpha) - \sum_m a_{mm} \right]^2 = 2 \sum_m \sum_n a_{mn}^2 \quad (13.29)$$

This result is similar to the result in the continuous case.

We have done this for a finite set of a_{mn} . But now let us use the Riesz-Fischer theorem. Suppose that we have an infinite set of a_{mn} such that Expression 13.30 is satisfied.

$$\sum_m \sum_n a_{mn}^2 \text{ converges} \quad (13.30)$$

That is, the difference of the sum of the squares of the coefficients of the finite sum goes to 0. Then again we use the Riesz-Fischer theorem, and we show the following:

$$\lim_{N \rightarrow \infty} \left[\sum_{m=0}^N \sum_{n=0}^N a_{mn} u_m(\alpha) u_n(\alpha) - \sum_{m=0}^N a_{mm} \right] = \phi(\alpha) \quad (13.31)$$

This was done merely on the assumption that the double sum of a_{mn}^2 converges in the mean. Therefore, we can define $\phi(\alpha)$ as

$$\phi(\alpha) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} u_m(\alpha) u_n(\alpha) - \sum_{m=0}^{\infty} a_{mm} \quad (13.32)$$

In other words, we have now a hierarchy of second-degree nonhomogeneous functionals, orthogonal to all terms of first and zero degree.

I do not believe that it is necessary for me to go into more detail. This work will go exactly the same way as it went with the continuous functions for all levels. It is clear then that I have a hierarchy of functions of α , orthogonal at each step to every other step, involving zeroth, first, second, and third degree, and so on. Remember that α is determined by the j 's, or by the g 's, or by the h 's. I say that it is easy to show that the polynomials in this set of independent variables with Gaussian distribution are closed. We can approximate to any function of L^2 by means of them. We can approximate to any function over an interval by means of them. These polynomials are the Hermitian polynomials and products of Hermitian polynomials of independent u 's. We have here a complete development of any function of L^2 . Notice, this will work whether we are working with a one-sided infinite set of the u 's or with a two-sided infinite set. If we keep the one-sided case, we shall get a one-sided development here. With a two-sided development, we shall get a similar two-sided result. In other words, given any function $\phi(\alpha)$ that belongs to L^2 , then $\phi(\alpha)$ can be written as

$$\phi(\alpha) \sim \sum_0^\infty G_n(K_{\nu_1 \dots \nu_n}, \alpha) \quad (13.33)$$

This notation is fairly obvious and is entirely analogous to the notation we used in the Brownian motion case. The summation can be from $-\infty$ to ∞ , but in the case in which we happen to know only the present and past it is from 0 to ∞ . The function G_n is obtained by

$$G_n = \sum K_{\nu_1 \dots \nu_n} u_{\nu_1}(\alpha) \cdots u_{\nu_n}(\alpha) + \text{lower terms} \quad (13.34)$$

where the terms of lower degree are necessary to make G_n orthogonal to all expressions of still lower degree. Expression 13.33 gives a universal description for all functions of α in terms of the u_n 's, where

$$u_n = h(T^{-n}\alpha) \quad (13.35)$$

This is exactly analogous to what we have done before.

If I replace Expression 13.36 by Expression 13.37,

$$K_{\nu_1 \dots \nu_n} \quad (13.36)$$

$$K_{\nu_1 \pm \nu \dots \nu_n \pm \nu} \quad (13.37)$$

then I am essentially replacing α by

$$T^{\pm \nu} \alpha \quad (13.38)$$

(The minus sign represents moving backwards in time.) Therefore, if I represent $f(\alpha)$ in a series like Expression 13.33, then I also have a representation of

$$f(T^{\pm\nu}\alpha) \quad (13.39)$$

in terms of the same type of series. This gives me then a form of coding the f 's in terms of the present and past of the h 's. In other words, I have now a standard form of expression in which I can represent my original time series if it is an L^2 series. If the distribution of h in terms of the past is properly monotone, and if the effect of the remote past (the Borel past) contains sets of only measure 0 or 1, then I have a standard form of coding or decoding a message in terms of independent Gaussian distributions. In other words, the theory here is entirely analogous to the Brownian motion theory that I have discussed, and we have really made some progress in coding.

I want to say that in the Brownian motion case there is a lot of work still to be done on this reconstruction, in particular in obtaining the g 's and h 's from the f 's in a continuous time series. We are going to need a great many auxiliary assumptions here, for the simple reason that if we take a brief interval of time, the present contains no part independent of the past; i.e., we do not have the freedom of a completely new choice to work with, as in the discrete case. Work should be done on this. I shall leave that question for the present. So much for what I am going to say about coding and decoding.

A New Approach to Statistical Mechanics—I

I want to start today with a series of ideas on statistical mechanics. This topic will occupy me to the end of the lectures. I want to give a model for a gas or a liquid. We might even say for a solid, but it does not fit so well for a solid, for the reason that in a solid we can, to a first extent, consider the molecules and assume that the forces are only between adjacent molecules. The long-distant forces are not as important as in a gas or as even in a liquid. In a solid, we do have forces extending beyond the range of a single particle. For the moment, I shall talk about gas without the assumption that I am excluding liquids from my discussion, for I am not. Let us use the usual picture of a gas: There are particles and these particles have positions given by three coordinates x , y , and z , and they have velocities given by three components v_x , v_y , and v_z . We want to describe the gas not so much by saying for each particle what its coordinates are, but by saying how many particles are in a certain region. The logical way of doing this for the particles of a gas is to define the six-dimensional box. How many particles lie in a certain range of position and velocity? That is, I am specifying not where the particle is as an individual particle but—given a certain region of position and velocity—how many particles are in it.

In statistical mechanics my results are rather tentative compared with what I have done before in these lectures, because there are many problems in it that I have not yet solved. I think their field is important. I believe that my approach is new: How fruitful it may be is for the future to determine. It gives us a lead in the kinetic theory of gases (including the theory of fluctuations), the theory of liquids, the theory of turbulence,

and much more. I have hopes that the present approach will ultimately lead to applications to quantum theory. My point of view is connected with the study of random functions that we have developed in these lectures.

Let us discuss how we may set up a field theory of gases in random motion. I am going to consider the case of a gas of particles that are acted on only by central forces which depend on the distances between the particles. I shall confine myself to the isotropic case, and I shall simplify everything. The methods that I introduce are really more general, but I shall discuss only the simplest formulation of the method. My discussion is devoted rather to the development of a methodology than to the obtaining of concrete results.

Let us consider a three-dimensional gas, from the point of view of Willard Gibbs. The gas consists of an assembly of particles, which we shall label by the numbers $1, 2, \dots, n, \dots$. The n th particle will possess the position coordinates x_n, y_n, z_n and the momentum coordinates u_n, v_n, w_n . The entire assembly $x_1, x_2, \dots, w_1, w_2, \dots$ will determine a point in what Gibbs calls phase space. This is a space in which an assembly of complete gases is distributed, in a manner analogous to that in which the expressions $\sum_0^\infty G_n[\phi_n(\tau_1, \dots, \tau_n), \alpha]$ are distributed in a space determined by the parameter α varying between 0 and 1. As in the case that we have already discussed, we shall wish to integrate over this space, as we have integrated over the variable α , to obtain the average values and the distributions of the quantities characterizing the gas.

However, this description of a gas does not fit the methodology employed in the previous lectures. Therefore, I suggest an alternative description of the gas. Let R be a region in phase space. Then R will contain a number, ν , of particles lying within R both in regard to position and momentum. This number will be an additive functional of the set R and will characterize the distribution properties of the gas.

Let us note that such a characterization has one definite advantage over the Gibbsian characterization. It will be invariant under any permutation of the particle. From the observational standpoint, such a permutation does not change the gas in any way.

We have already been accustomed to additive functions of sets of values of a single variable, depending on a parameter of distribution α . We have represented these by a function

$$f(x, \alpha) \tag{14.1}$$

where, if $x_2 < x_1$, the function $[f(x_2, \alpha) - f(x_1, \alpha)]$ gives us the mass, in the case labeled α , lying between x_2 and x_1 . Here, if S is a set of values of x and $S(x)$ is the characteristic function of this set, the mass over S in

case α is

$$\int_S df(x, \alpha) \quad (14.2)$$

In the case of an n -dimensional distribution, there is no difficulty in extending the definition of the mass in case α over S to a similar integral, now remembering that the integration is not with respect to a function $f(\alpha)$ but with respect to an additive functional of sets of points.

We now wish to give a standard representation of such a functional in the n -dimensional case. We wish to confine ourselves to additive functionals that are not changed under an identical translation of the position coordinates of all points, though they may be changed, and generally are changed, by a similar translation of the momentum coordinates.

We have already seen that any function of α belonging to L^2 may be written as

$$\sum_{n=0}^{\infty} G_n[\phi_n(\tau_1, \dots, \tau_n), \alpha] \quad (14.3)$$

This representation has nothing to do with the dimensionality of the space of the τ_k 's, as every such space can be mapped on a space of one dimension with preservation of measure, and as such a mapping will preserve the Gaussian distribution of the mass on a given measurable region and will preserve the independence of the distribution of masses on nonoverlapping regions. Thus, any function of α belonging to L^2 may be written in the form

$$\sum_{n=0}^{\infty} G_n[\phi_n(x_1, y_1, z_1, u_1, v_1, w_1; x_2, y_2, z_2, u_2, v_2, w_2; \dots \\ x_n, y_n, z_n, u_n, v_n, w_n), \alpha] \quad (14.4)$$

where we have written our τ_k 's in an expanded form as sets of coordinates in phase space.

The transformation

$$x^* = x + \xi, y^* = y + \eta, z^* = z + \zeta, u^* = u, v^* = v, w^* = w \quad (14.5)$$

is a unitary transformation. Such a transformation is generated by a measure-preserving transformation

$$T_x^\xi T_y^\eta T_z^\zeta \alpha \quad (14.6)$$

where $T_x^\xi, T_y^\eta, T_z^\zeta$ are permutable.

Therefore, any function $F(x, y, z, u, v, w, \alpha)$ that is L^2 in α for every x, y, z, u, v, w , and that does not change its statistical nature when we

add ξ to x , η to y , ζ to z , and leave u , v , w unchanged, can be written as

$$\sum_0^{\infty} G_n [\phi_n(x_1 + x, y_1 + y, z_1 + z, u_1, v_1, w_1; \dots; \\ x_n + x, y_n + y, z_n + z, u_n, v_n, w_n), \alpha] \quad (14.7)$$

From this expression we can form a linear functional of a set S of values of x, y, z, u, v, w by taking

$$\iint_S F(x, y, z, u, v, w) dx dy dz du dv dw \quad (14.8)$$

This can be taken to represent a distribution of some quantity in phase space and is invariant under translations in position. It has a close similarity therefore to the sort of quantity that we need to describe a gas in phase space. Notice, however, that, unlike the classical distribution of a gas, it does not assume a mass distribution in a denumerable set of discrete points in phase space. As a matter of fact, it cannot describe such a gas.

Nevertheless, this description may be useful to us in gas theory. On the one hand, as for example in turbulence theory, there are problems in which we do not come down to molecular dimensions. Furthermore, even when our sort of theory is not applicable directly, it can yield a statistical theory highly analogous to ordinary gas theory, and there is at least a possibility that, where it is not correct, the errors in the theory due to the granular structure of a true gas may be corrected for specific purposes by the introduction of adjustment terms specifically introduced for the purpose.

Let us then introduce a description of this type for a gas consisting of spherically symmetric particles bound to one another by forces depending on their distances and along the line connecting one particle with another. If the n th particle has coordinates x_n, y_n, z_n , the components of force acting on it will be

$$\sum_m F\{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}\} \\ \times \frac{x_n - x_m}{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}} \quad (14.9)$$

$$\sum_m F\{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}\} \\ \times \frac{y_n - y_m}{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}} \quad (14.10)$$

$$\sum_m F\{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}\} \times \frac{z_n - z_m}{[(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2]^{\frac{1}{2}}} \quad (14.11)$$

If we replace these coordinates by a quantity $\rho(x, y, z, u, v, \alpha)$ representing density in phase space, we shall have as the components of the force acting on a unit mass at the point (x, y, z)

$$\begin{aligned} & \int \int \int \int \int \int F\{[(x - \xi)^2 + (y - \eta)^2 + (z - \varsigma)^2]^{\frac{1}{2}}\} \rho(\xi, \eta, \varsigma, \lambda, \mu, \nu, \alpha) \\ & \text{All phase space} \\ & \times \frac{x - \xi}{[(x - \xi)^2 + (y - \eta)^2 + (z - \varsigma)^2]^{\frac{1}{2}}} d\xi d\eta d\varsigma d\lambda d\mu d\nu, \text{ etc.} \end{aligned} \quad (14.12)$$

where $\rho(\xi, \eta, \varsigma, \lambda, \mu, \nu, \alpha)$ is the density of particles in phase space at positions ξ, η, ς and with momenta λ, μ, ν for the value α of the phase parameter.

For our present purposes, we are concerned with the development of methods rather than the obtaining of results. For these purposes, the one-dimensional gas is as satisfactory as the three-dimensional gas. Our force between particles has now only one coordinate and will be of the form

$$\int \int F(x - \xi) \rho(\xi, \lambda, \alpha) d\xi d\lambda \quad (14.13)$$

Let it be noted that $\rho(x, y, z, u, v, w, \alpha)$ constitutes a full description of a gas *at one time*. If we are to investigate the dynamics of the gas, which means a description of the behavior of the gas in time, we must introduce the extra parameter t and write the density in phase space at time t as $\rho(t, x, y, z, u, v, w, \alpha)$.

In order to develop the dynamics of ρ , it is desirable to write ρ in some sort of canonical form. We shall put

$$\rho(t, x, u, \alpha) = \sum_0^\infty G_n[\rho_n(t, u; x + x_1, u_1, \dots, x + x_n, u_n), \alpha] \quad (14.14)$$

Implicit in this representation is a fundamental additive Brownian functional corresponding to $X(x, u, \alpha)$. However, this functional merits further discussion. As time goes on, the relation of parameters of the changing system to a fixed X becomes more and more remote. To describe the dynamics of the system in these terms is to refer its randomness to an epoch that is continually escaping us in the past. Such a description is not too well suited to the employment of differential-equation theory, where the discussion of each moment in the system is complete in itself and we are not perpetually hovering between two instants.

A New Approach to Statistical Mechanics—II

Let us now discuss the dynamics of a particle gas under central force. Here we shall consider only the one-dimensional case explicitly. There are no difficulties of principle in discussing the three-dimensional gas directly, but we add nothing new that is relevant to the methodology of the situation, and we complicate the notation considerably.

The velocity at point (x, u) of phase space will be u . Thus, after time dt , the position coordinate x will be changed into

$$x + u dt \quad (15.1)$$

Here let us notice that we take u to be the velocity, corresponding to a momentum for a mass m at this point of mu . The new velocity coordinate will be

$$u + dt \iint F(x - \xi) \rho(\xi, \lambda) d\xi d\lambda \quad (15.2)$$

where the integral is to be taken over the whole of the phase space (ξ, λ) and ρ is the density of mass in this phase space. We may write this infinitesimal transformation as

$$dx = u dt \quad (15.3)$$

$$du = dt \iint F(x - \xi) \rho(\xi, \lambda) d\xi d\lambda \quad (15.4)$$

Let us notice that the Jacobian of this infinitesimal transformation is

$$\begin{vmatrix} 1 & dt \\ dt \iint F'(x - \xi) \rho(\xi, \lambda) d\xi d\lambda & 1 \end{vmatrix} = 1 - (dt)^2 \iint F'(x - \xi) \rho(\xi, \lambda) d\xi d\lambda \quad (15.5)$$

Notice that dt enters quadratically. Therefore, to the first order in dt , this Jacobian is 1. In other words, the infinitesimal transformation that we perform on (x, u) is measure-preserving.

We know that in one dimension, if we form

$$x(Tt, \alpha) \quad (15.6)$$

where T is a measure-preserving transformation, then we shall find that

$$\int \phi(t) dx(Tt, \alpha) = \int \phi(T^{-1}t) dx(t, \alpha) \quad (15.7)$$

where $\phi(T^{-1}t)$ differs from $\phi(t)$ only by a unitary transformation. Hence, we may write

$$\int \phi(t) dx(t, \alpha) = \int \phi(t) dx(t, S\alpha) \quad (15.8)$$

where S is a measure-preserving transformation. Since the properties of $x(t, \alpha)$, which we shall write as $X(t, \alpha)$ to avoid confusion with the variable x , are not dependent on dimensionality, we see that the infinitesimal change of x and u by dx and du , respectively, generates an infinitesimal measure transformation on α . It is true that in this discussion of infinitesimal transformations we are talking heuristically, but in this presentation of the basis of statistical mechanics our whole procedure is heuristic.

We shall now consider the development of $\rho(t, x, u, \alpha)$, the density of our distribution of x and u at time t , and with the parameter of distribution α . Here α does not indicate our random variable at time 0 but at time t , and α differs from this random variable at time 0 by the resultant of an infinite sequence of infinitesimal unitary transformations, which will be a unitary transformation. This transformation will, in fact, depend on the history of the variable ρ , and this must be discussed in any thorough justification of our procedure, but we shall not go into this at present.

Now we wish to discuss the infinitesimal change in an expression

$$\int \cdots \int \phi(x_1, u_1, \dots, x_n, u_n) dX(x_1, u_1, \alpha) \cdots dX(x_n, u_n, \alpha) \quad (15.9)$$

caused by the change in X with the time. This change will be the sum of the infinitesimal changes, each caused by the change in a single X . Thus, we need to discuss only the change in

$$\int \phi(x, u) dX(x, u, \alpha) \quad (15.10)$$

caused by the change in time. To identify the time at which X is taken, we shall write Expression 15.10 as

$$\int \phi(x, u) dX(t, x, u, \alpha) \quad (15.11)$$

Now,

$$\int \phi(x, u) dX(t + dt, x, u, \alpha) = \int \phi(x, u) dX(t, x - u dt, u - F dt, \alpha) \quad (15.12)$$

where F represents the acceleration of the forces at (x, u) . Equation 15.12 may be written to a first approximation in dt :

$$\begin{aligned} & \int \phi(x + u dt, u + F dt) dX(t, x, u, \alpha) \\ &= \int \phi(x, u) dX(t, u, \alpha) + dt \int \left(u \frac{\partial \phi}{\partial x} + F \frac{\partial \phi}{\partial u} \right) dX(t, u, \alpha) \end{aligned} \quad (15.13)$$

Now let us return to our discussion of the time course of the density function $\rho(t, x, u, \alpha)$. We have already given for this the development

$$\sum G_n[\rho_n(t, x, u, x_1, u_1, \dots, x_n, u_n), \alpha] \quad (15.14)$$

Here the variable α is the correct variable at time t . For a really rigorous discussion, the expansion of ρ is the desirable one. However, for our present heuristic consideration, it may be better to consider the expansion

$$\begin{aligned} \rho(t, x, u, \alpha) &= \sum_{n=0}^{\infty} \int \cdots \int P_n(t, u; x + x_1, u_1; \dots; x + x_n, u_n) \\ &\quad \times dX(t, x, u_1, \alpha) \cdots dX(t, x_n, u_n, \alpha) \end{aligned} \quad (15.15)$$

Equation 15.15 bears the same relation to the development in terms of the ρ_n 's that a Taylor series development does to a development in Hermite polynomials, and leads us to a set of annoying and difficult questions concerning convergence and summability. Here we shall

merely take cognizance that such questions exist but shall not discuss them.

Let us now examine $\rho(t + dt, x, u, \alpha)$, and let us search for a set of differential equations connecting the various derivatives of

$$P_n(t, u; x + x_1, u_1; \dots; x + x_n, u_n) \quad (15.16)$$

Before we actually undertake writing down the equations, not to mention their solution, it will be well to consider what results we expect to get from them. Our equations will yield us the entire statistical history of a gas, not necessarily in equilibrium. From this we may hope to get quantities giving at each time the distribution of kinetic and potential energy (more or less corresponding to problems of temperature), the rate of transfer of potential energy across a point i or, in the three-dimensional case, a unit area, which will be the mathematical equivalent of pressure, and various other questions as well concerning the fluctuations of these and other quantities. The mean density will, in general, be given by the mean density at time 0, as our particles will not move rapidly enough to make a substantial change in the mass in a great volume. All of this is quite practical and opens new fields in nonequilibrium statistical mechanics.

As to equilibrium statistical mechanics, we have two possible procedures. One is to put each $\partial P_n / \partial t$ equal to zero. This leads to a set of differential equations purely in the x and u variables. The other is to formulate and solve the nonequilibrium problem and to investigate the asymptotic behavior of the solution after a long time.

Let it be noted that the equilibrium problem, on the hypotheses set forth here, will have no significant solution. A system of many degrees of freedom in equilibrium will, under fairly general assumptions, show an equipartition of energy over the various degrees of freedom. Thus, a continuous gas will have no temperature, strictly speaking. If such a gas is allowed to develop itself under its own forces for a long time, all of the energy will run away into the structure of the gas of very small dimensions, and the asymptotic state will be a static one of perfect rest.

All this does not mean, however, that methods of our type are hopeless, but merely that they must be modified to take account in some other ways of the properties due to the granular structure of the gas. There must be superimposed on the processes of which we have already taken account some process that continually reimposes on the gas a new fine randomness like that due to the separate identity of the molecules. Temperature and similar notions must be redefined in such a way that they will not be rendered infinite by the fine randomness. Related notions have already been suggested by Kolmogoroff and merit much further discussion.

Let us consider the expression

$$\begin{aligned} \rho(t + dt, x, u, \alpha) \\ = \sum_{n=0}^{\infty} \int \cdots \int P_n(t + dt, u; x + x_1, u_1; \cdots; x + x_n, u_n) \\ \times dX(t + dt, x_1, u_1, \alpha) \cdots dX(t + dt, x_n, u_n, \alpha) \end{aligned} \quad (15.17)$$

In evaluating Equation 15.17, we must take account both of the substantive change in P_n due to the dynamical change in the gas and of the change due to the change in the variable α with time. This will give us

$$\begin{aligned} \frac{\partial \rho(t, x, u, \alpha)}{\partial t} = \sum_{n=0}^{\infty} \int \cdots \int & \left[u \frac{\partial P_n}{\partial x} + u_1 \frac{\partial P_n}{\partial x_1} + \cdots \right. \\ & + u_n \frac{\partial P_n}{\partial x_n} + \frac{\partial P_n}{\partial u} \iint K(x - \xi) \rho(t, \xi, \lambda, \alpha) d\xi d\lambda \\ & + \frac{\partial P_n}{\partial u_1} \iint K(x_1 - \xi) \rho(t, \xi, \lambda, \alpha) d\xi d\lambda \Big] + \cdots \\ & \left. + \frac{\partial P_n}{\partial u_n} \iint K(x_n - \xi) \rho(t, \xi, \lambda, \alpha) d\xi d\lambda \right] \\ & \times dX(t, x_1, u_1, \alpha) \cdots dX(t, x_n, u_n, \alpha) \end{aligned} \quad (15.18)$$

However, the quantities $\rho(t, \xi, \lambda, \alpha)$ have themselves known random developments. We have

$$\begin{aligned} \rho(t, \xi, \lambda, \alpha) = \sum_{n=0}^{\infty} \int \cdots \int P_n(t, \lambda; \xi - \xi_1, \lambda_1; \cdots; \xi - \xi_n, \lambda_n) \\ \times dX(t, \xi_1, \lambda_1, \alpha) \cdots dX(t, \xi_n, \lambda_n, \alpha) \end{aligned} \quad (15.19)$$

Thus, at least formally, we have represented $\partial \rho(t, x, u, \alpha)/\partial t$ as a sum of terms

$$\begin{aligned} \int \cdots \int Q(t, x, u, x_1, u_1, \cdots, x_n, u_n, \xi_1, \lambda_1, \cdots, \xi_k, \lambda_k) \\ \times dX(t, x_1, u_1, \alpha) \cdots dX(t, x_n, u_n, \alpha) \\ \times dX(t, \xi_1, \lambda_1, \alpha) \cdots dX(t, \xi_k, \lambda_k, \alpha) \end{aligned} \quad (15.20)$$

where the $Q(t, x, u, x_1, u_1, \cdots, x_n, u_n, \xi_1, \lambda_1, \cdots, \xi_n, x_n)$ are finitely representable integral expressions involving K , the P_n 's up to a certain stage, and their derivatives. We can take Q and permute all the pairs $x_1, u; x_2, u_2; \cdots; x_n, u_n; \xi_1, \lambda_1; \cdots; \xi_k, \lambda_k$ in all possible ways, add them, and divide by the number of permutations. If we note that the sym-

metrical K_n in the expression

$$\int \cdots \int K_n(S_1, \dots, S_n) dX(S_1, \alpha) \cdots dX(S_n, \alpha) \quad (15.21)$$

is essentially unique, and that any sum of such expressions of differing degrees can be broken down into its terms in a unique way, and if we remember that the differentiation of such an expression with respect to a parameter t will lead, at least formally, term by term to a similar development, we see that we shall have led, at least formally, to a series of differential integral equations in which each $\partial P_n / \partial t$ will be expressed as a differential integral expression involving the P_k 's ($k \leq n$) and their partial derivatives with respect to the various x and u variables. There is at least the possibility of starting with P_0 and P_1 and of solving these equations in sequence.

What I have done in the last two lectures is to lay out a program for further research rather than to state definitive results. Accordingly, even though I have written out explicitly for my own benefit some of the earlier integrodifferential equations of the sequence that I have indicated, it would serve very little purpose to go into details here. However, a few general comments are in order. We can expect for the equations not a single solution but a system of solutions depending on various numerical and functional parameters. Some of these will correspond to quantities necessary to fix the physical situation beyond the point where it can be fixed by the force F between particles. Among these are such quantities as the density and the pressure of the gas, or any two equivalent quantities. We cannot expect, however, that when these quantities are fixed the problem will become specific. Our representation of a random system in terms of the G_n 's or their equivalents need not be unique, from the point of view of the correlations and higher-order quantities of the same nature defining the statistical state of a gas. All we can hope to do by our methods is to find one or a number of statistical representations of a gas that may have enough free parameters to be of practical use. For this practical work, we are free to impose on the solutions of our differential equations such conditions as will facilitate the obtaining of solutions in usable form but will not be so drastic as to exclude the cases in which we are interested.

Index

- Alpha rhythm, 68
Amplifiers, 89
Analysis of systems, 96, 97, 100
Asteroids, 77
Autocorrelation, 43, 65–69
- Black box, 89, 97
Borel fields, 108
Bose, A. G., 87
Brain waves, 70, 71
Brownian motion, 5, 12, 16, 22, 23, 29, 31, 44, 46, 52, 70, 78, 82, 83, 88–90, 96, 98, 112, 117
Bush, Vannevar, 96
- Calculus of random functions, 16
Canonical development, 37, 42
Cascade of simple lattices, 83
Cathode followers, 91
Central frequency, 72, 73
Characteristic amplitudes, 83
Characteristic frequencies, 83
Characteristic functions, 81–83
Characteristics of an electric circuit, 89
Closure, 40, 41
Coding, 107–109, 117
- Complex differential space, 22, 25, 27, 59, 63, 78, 80, 81, 83
Condenser, 68
Conditional distribution, 102, 103, 105
Continuous time series, 117
Cyclotron, 77
- Damping, 99, 160
Decoding, 108, 110, 128
Differential equations, 122
Differential space, 21, 27, 78
Dip, 48, 67, 69, 72, 75, 76
Driving, 71
- Eigenfunctions, 81, 82, 85
Electrical engineering, 27
Electrodynamometer, 99
Electrostatic induction, 71
Equicontinuity, 8
Equilibrium statistical mechanics, 126
Equipartition of energy, 126
Ergodic hypothesis and theory, 44–46, 96, 98, 108
Explosive networks, 89
Exponential damping, 99
- Field theory of gases, 119

- Fine structure of spectra, 69
 Flicker, 71
 Fluctuating load, 70
 Fluctuations, 70, 118
 Fourier analysis, 26, 42, 54
 coefficients, 54
 transform, 74, 93
 Fredholm, E. I., 59
 Frequency, 70–72
 Frequency modulation, 49, 54–56, 63, 67
 G_n for unsymmetric functions, 64
 Gases, 119, 121, 122, 126, 127
 Gaussian distributions, 3, 10, 19–21, 24,
 25, 54, 57, 75, 111–112, 116, 120
 Generating functions, 56
 Generators, 70
 Gibbs, J. W., 86, 119
 Governor, 70
 Granular structure of gas, 126
 Haar functions, 22
 Hamiltonian, 82, 86
 Harmonic analysis, 42, 54, 68, 69
 Hermite polynomials, 63, 94, 96, 100,
 109, 116, 125
 Hilbert space, 20
 Homogeneous polynomial functionals, 28,
 29, 52, 60
 Impedance, 91
 Independence, 59, 107, 108, 110
 Inductance, 92
 Infinite past, 111
 Infinitesimal transformation, 124
 Integrodifferential equations, 127
 Kinetic energy, 126
 Kinetic theory of gases, 118
 Kolmogoroff, A. N., 126
 L^2 , 17, 117
 Lagrangian coordinates, 86
 Laguerre functions, 91, 94, 95, 98, 100
 Lattice, 91
 Lebesgue integrable functions, 107
 measure, 79, 80, 84, 101
 Lee, Y. W., 91
 Linear networks, 90
 Linear oscillators, 71
 Linear resonators, 52
 Load, 70
 Magnetic tape, 67
 Measure-preserving transformations, 43,
 44, 85, 86, 98, 124
 Measure theory, 79, 80, 84
 Metrically transitive transformation, 45
 Michelson interferometer, 69
 Narrow-range governor, 90
 Nondifferentiable functions, 9, 26
 Nonequilibrium statistical mechanics, 126
 Nonhomogeneous polynomial functionals,
 29
 Nonlinear networks, 88–90, 96
 Nonlinear oscillators, 27
 Normalization, 93
 One-dimensional gas, 122, 123
 Optical theory, 69
 Orthogonal functions, 28–33, 35, 37, 40,
 42, 49, 52, 53, 59–61, 63, 65, 78, 82, 84,
 93, 112
 Oscillators, 69, 71
 Particle gas, 123
 Phase shifter, 91
 Phase space, 119, 121, 123
 Polynomial functionals, 112
 Potential energy, 126
 Pressure, 128
 Pull of frequencies, 72, 73
 Quantum theory, 78, 82–84, 86, 89, 119
 Random additive functional, 81
 Random functions, 1, 27, 119, 122, 124
 Random input, 73
 Random output, 27
 Randomness in space, 80, 81
 Random process, 1, 11, 74
 Random processes in several dimensions,
 79

- Regulation, 70, 71
Resolving power, 69
Riesz-Fischer theorem, 15, 36, 113,
 115
Roughness, 8

Schedule analysis, 68
Schrödinger equation, 82, 83, 85, 86
Shot effect, 70, 88, 89, 96, 97
 generator, 89, 96, 97
Side bands, 48
Smoothness, 38
Spectral line, 47
Spectrum, 1, 22, 27, 40, 47, 50, 52, 54,
 65–69, 74–76
Square-law rectifier, 95
Statistical mechanics, 118, 123, 126
Stochastic function, 9

Synthesis of systems, 95, 96, 100
Taylor series, 125
Temperature, 126
Three-dimensional gas, 122
Time sequence, discrete, 102
Transfer characteristics, 92
Translation invariance, 127
Turbulence, 118, 121

Uniformly distributed message, 109, 110
Unitary transformation, 85, 124

Voltage multipliers, 95, 97, 99, 100
Walsh functions, 21, 22

Zacharias, J. R., 46