Am. Jour. Phys. 29, 276 (1961)

Book Reviews

Statistical Theory of Communication. Y. W. Lee. Pp. 509, John Wiley & Sons, Inc., New York, 1960. Price $16.75.

In a field whose growth is rapid, and where a premium seems to be put on extreme specialization, it is hard to reach common agreement on terminology. Only in this way is it possible to understand the appearance in 1960 of a book with the title, *Statistical Theory of Communication,* which contains not a single reference to Shannon, and devotes exactly two sentences (pp. 2 and 3) to the body of concepts and theorems which, to this reviewer, *is* the field of "communication theory."

If this book had been entitled, "Introduction to the theory of Wiener filters," there would be no confusion as to content; for that is exactly what it is. It is intended as a textbook, based on the author's teaching experience since 1947, for a one-semester course given to first-year graduate students in the M.I.T. Department of Electrical Engineering.

As might be expected from its origin, the book *is* written in a beautifully clear manner, starting with the most elementary notions and leading us very gently, with a great deal of discussion and examples, through the theory of autocorrelation functions and power spectra, random functions, ensembles, ergodic hypotheses, the formulation and solution of the Wiener-Hopf integral equation for optimal filtering and prediction by the mean-square error criterion. The author considers several applications, error analysis of optimal and near-optimal systems, practical aspects of sampling techniques, and the use of orthonormal functions, particularly Laguerre functions, for vstem synthesis.

Particularly noteworthy is the fact that examples are not all abstract calculations; numerous oscillograms give actual results obtained by correlation filtering in the well-known experimental work of Professor Lee and collaborators, which had appeared previously in several technical reports of the M.I.T. Research Laboratory of Electronics.

The use of explanatory remarks, examples, and other digressions is so lavish as to border on too much of a good thing. The superior student, who can grasp a concept at first reading, may find that they sometimes interrupt, rather than clarify, the line of reasoning. Of course, good pedagogy requires a certain amount of repetitiousness; too often, students complain because a book errs in the opposite direction.

In the last half of the book, one can find several useful results which have not, as far as I am aware, appeared previously in print. Unfortunately, however, the book cannot be regarded as an easy-to-read substitute for Norbert Wiener's own exposition. It is only an elementary introduction to the subject, and some important material given in Wiener's *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (John Wiley & Sons, New York, 1949) has been left out. For example, the author avoids having to mention the Paley-Wiener criterion for predictability and realizability by the device of considering only power spectra of such simple analytical form that the fundamental factorization operation can be performed by inspection. But a person who tries to use this theory in his own problems will quickly find, to dismay, that there exist random functions which are ...irely reasonable physically, but for which the factorization is impossible. This is the case for what the physicist would probably regard as the most fundamental of all random functions, the one whose power spectrum is given by the Planck blackbody radiation law. Here the autocorrelation function $\varphi(t)$ turns out to be analytic for all real t, and the Wiener prediction filter does not exist.

For practical purposes, this kind of difficulty can be avoided by approximating the power spectrum by a finite number of Laguerre functions, for each of which the factorization is done by inspection; but the conscientious student will see that there is a fundamental unresolved question of principle here, for which the present book gives him no help. This is, of course, the singular case where the Wiener theory gives the formally correct, but physically entirely misleading, result that the future of the random function is perfectly predictable from knowledge of its past. Norbert Wiener's original work, popularly known as "The yellow peril," was outstandingly difficult reading for persons not previously familiar with such things as functional integration over Wiener measure; but at least this kind of material was in it, and could be extracted with a little honest sweat.

The criticisms made above should be regarded as extremely mild, and indeed one can argue that the features mentioned actually enhance, rather than detract from, the intended purpose of the book. But now I want to point out two disturbing features which, in the interest of fairness and accuracy, should be corrected in future editions.

The author's failure to make any reference to the work of Shannon concerning information sources, entropy, channel capacity, etc., while regretted by this reviewer, cannot be the basis for criticism; because, after all, an author has the right to choose his own topic. But no such charitable view is possible with regard to Chap. 14, where the problem of the optimum linear filter is formulated and solved. Here the reader can find not a single hint of the fact that, in 1950, a paper by Bode and Shannon appeared which gave a simpler, and much more physically motivated, derivation of these same results.

To many workers in this field, the Bode-Shannon paper was the key which, for the first time, made it possible to understand what we were doing in the Wiener theory. In depriving the reader of this very important contribution to the subject, the author does violence to his own stated purpose.

The final criticism is of particular concern to physicists. Every electrical engineer who writes on communication theory feels the need to insert a few remarks concerning its historical basis in statistical mechanics. My theory is that someone used the word "Gibbs" and a little imagination, to compose the first of these incantations, and that subsequent writers have simply appropriated the result, with embellishments. Whatever the cause, there is today a growing body of misinformation about the history of statistical mechanics, propagating through the ranks of communication engineers.

A book on information theory which appeared in 1953 contained a photograph of Gibbs, with this amazing caption: "J. Willard Gibbs (1839–1903), whose ergodic hypothesis is the forerunner of fundamental ideas in information theory." In the book presently under review, this is amplified as follows. On p. 207, we are told that "Gibbs originated the assumption that, in a closed system where the total energy remains constant, a time average over the motions of a system of particles has an equivalent average obtained by integration over a surface in phase space called the ergodic surface." On the following page, the author continues, "In the attempt to justify Gibbs' assumption as questions concerning its validity arose, J. C. Maxwell proposed the hypothesis that '. . . the system, if left to itself in its actual state of motion, will sooner or later, pass through every phase which is consistent with the equation of energy.'"

To this, several comments can be made. In the first place, Maxwell died in 1879, 23 years before Gibbs' work on statistical mechanics appeared. It was, of course, not Gibbs but Boltzmann whose work Maxwell was referring to in the passage quoted.

In the second place, Gibbs not only did not *formulate* any ergodic hypothesis, he did not *employ* it or the word "ergodic" at any place in his book. Indeed, to do so would have run counter to his whole purpose, as stated in the preface. Gibbs was probably the first person to see the possibility of founding statistical mechanics directly on prior probability assignments, independently of any such unproved physical assumptions.

Nowhere does Gibbs state or imply that the probability $p(R)$ assigned to a certain region R of phase space is equal to the long-run relative *frequency* $f(R)$ with which a physical system occupies R, which is what an ergodic hypothesis would amount to. $p(R)$ is a frequency only in an imaginary "ensemble," which is a conceptual device for describing a certain *state of knowledge*. This distinction is pointed out clearly by Tolman [*The Principles of Statistical Mechanics* (Oxford University Press, New York, 1938) pp. 63–70].

The reliability of the predictions then comes, not from any hypothesis about the long-run behavior of an individual system, but from the fact, which Gibbs stresses over and over, that when we pass from a probability distribution in phase space to a probability distribution for the particular phase functions $f(q,p)$ which are measured in thermodynamic experiments, we obtain distributions with enormously sharp peaks, practically all of the probability being concentrated in an interval much smaller than the experimental error. This being the case, it is clear that the success of statistical mechanics (equality of ensemble averages and experimental values) can be understood without recourse to any ergodic hypothesis.

By using the "global" viewpoint from the start, and by introducing the canonical and grand canonical ensembles, Gibbs brought to this subject unity, formal elegance, and a technique of calculation which the labors of another sixty years have not been able to improve on. Gibbs' own contributions are great enough. We do not need to credit him with things which were not only done several years earlier by Clausius, Boltzmann, and Maxwell, but were actually foreign to his viewpoint.

Gibbs' work was incomplete in that he gives no principle for choosing the ensemble; the canonical and grand canonical ensembles are introduced in an apparently arbitrary manner. For this reason, many physicists did not at first accept the Gibbs methods. The Ehrenfests, in their famous review article of 1912, dismiss them as "analytical tricks" made solely for ease of calculation, and stress the physical superiority of Boltzmann's viewpoint. Since then, the mathematical superiority of Gibbs' method has become so clear that today most physicists would consider the main results of Boltzmann (collision equation, H theorem, etc.) as valid only to the extent that they can be derived from the equations of Gibbs. But the arbitrariness in choosing the initial ensemble has remained, and debate over the justification of Gibbs' methods has continued. Recently, I have discussed in some detail the possibility that the concepts introduced by Shannon may prove to be the missing link which removes this arbitrariness and gives the final justification for Gibbs' methods.

In summary, the prospective reader should be forewarned that this book is not about communication theory as the term is usually understood, but takes up only one specialized topic; the Wiener filter and necessary preliminaries thereto. For the person who wants an easy introduction to this part of communication theory, it is by far the best source available. It is recommended that Chap. 14 be read in conjuction with the paper, "A Simplified derivation of linear least-square smoothing and prediction theory," by H. W. Bode and C. E. Shannon [Proc. I.R.E., 38, 417 (1950)].

E. T. Jaynes
Washington University