Introduction to “Communication in the Presence of Noise” by C. E. Shannon

AARON D. WYNER, FELLOW, IEEE, AND SHLOMO SHAMAI (SHITZ), FELLOW, IEEE

I. PERSONAL PREFACE

Dr. Aaron D. Wyner, who passed away after a severe, long illness on September 29, 1997, was the most natural choice to author an introduction to this milestone paper of Shannon’s. Aaron’s work in information theory, among many other fundamental contributions, extended and put on sound mathematical footing the basic ideas, notions, and arguments first introduced in this and other papers of Shannon’s. Aaron was one of the most accomplished members of the IEEE Information Theory Society, and for his profound contributions to the discipline and his faithful service to the society for more than three decades, he was the recipient of many prestigious recognitions, among them the society’s highest honor: the Claude E. Shannon Award.

I (the second author) was honored to be asked by Aaron to join him in writing this introduction to Shannon’s paper. I feel fortunate and privileged to have known Aaron and his unique family in person and to have worked with him at Bell Laboratories for many years. It was just a short while ago, during my summer visit to Bell Laboratories, when we were working together daily on this introduction, and Aaron was, as always, productive, original, and inspiring. His vitality, optimism, enthusiasm, and continuing scientific productivity constitute the main reason why we, his friends, though aware of his terminal illness, reacted with a deep sense of shock to his sudden passing.

Aaron’s ultimate departure is a great loss not only to his family and friends but to the whole Information Theory Society. This is strongly reflected in the decision by the society’s Board of Governors in the September 29, 1997, meeting, expressing a profound sense of loss, to name Aaron as one of the society’s most accomplished and valued members.

Manuscript received October 26, 1997.
A. D. Wyner, deceased, was with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA.
S. Shamai (Shitz) was with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA. He is now with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000 Israel (e-mail: sshlomo@ee.technion.ac.il).
Publisher Item Identifier S 0018–9219(98)01297-3.
rate-distortion theory (some of the insights and results appear already in [1]) developed later [3], with focus on continuous-time signals (waveforms) and a mean square error distortion measure.

In this paper, we give a summary of the contents of Shannon’s paper, along with a retrospective view of subsequent developments and refinements of the fundamental ideas first presented there. We focus in particular on the sampling theorem and the geometric representation of a communications system, along with a geometric proof of the Gaussian capacity formula and the channel capacity for nonwhite (colored) Gaussian channels. We also briefly mention other ideas, which in a historical perspective we find to be significant. Last, we summarize the significance of the ideas of Shannon as appears in this paper.

III. SAMPLING THEOREM

After presenting the standard block diagram of a communications system, which first appeared in [1], and repeating important notions and definitions such as the reliably transmitted rate, Shannon focuses on continuous-time systems, which manipulate and process waveforms. He introduces the sampling theorem in Section II as a central instrument that can deal with such a framework. Via this tool, the original time-continuous problem essentially translates into a discrete time problem. The central idea is phrased simply: a continuous function of time possessing an approximate bandwidth $W$ and of approximate time duration $T$ has approximately $2WT$ degrees of freedom and therefore can be represented in a space of dimension approximately $2WT$. The standard sampling theorem is proved, that is, any function $f(t)$ (which belongs to either $L_1$ or $L_2$—a technical condition not mentioned in the paper) with vanishing Fourier transform outside the band $[-W, W]$ can be exactly (limit in mean convergence) represented by its Nyquist samples $f(n/2W)$ $n$-integer (a terminology coined in the paper) at rate $2W$, where the reconstruction is achieved by the sinc $(\sin \pi x/\pi x)$ kernel [(7) in Shannon’s paper]. Shannon cites Whittaker [4], who eventually reported on similar results, as did Kotelnikov. See the appropriate references and historical notes in [5].

The application of the sampling theorem to represent essentially band- and time-limited signals, however, is somewhat problematic. Shannon explicitly mentioned that no (nonzero) signal can be simultaneously time and band limited. Therefore, he speaks of signals that are strictly band limited but almost time limited. Nonetheless, Shannon then defines signals of finite duration $T$ such that their Nyquist samples outside this time interval vanish. That is, those signals being strictly band limited to a single side bandwidth $W$ are represented by exactly $2WT$ samples. The underlying assumption is that if the Nyquist samples of a strictly band-limited signal disappear outside an interval of duration $T$, then the energy of that signal outside this time interval must be small. This is not necessarily the case, as has been shown by Pollak [6]. Shannon also points out the existence of irregularly spaced sampling techniques and refers to specific schemes that involve sampling the signal along with its derivative (or derivatives) at a reduced rate. Shannon also emphasizes what was later referred to as the “stability” of the sampling technique [5], that is, small errors in the sampling leading to small errors in reconstruction; the more irregular the sampling process is, the higher the accuracy needed in the representation of each sample to keep the reconstruction error bounded within a limited time interval. A variety of sampling methods originating in Shannon’s classical exposition and their extensions appear in [5].

Ignoring those stability arguments gives way to so-called paradoxes [7] arising essentially from the fact that a strictly band-limited signal can be exactly reproduced from any time interval, as small as desired [5]. All controversies and paradoxes have been settled in the important work of Slepian et al. (see [8] for a tutorial account). The definition of an almost band-limited (or equivalently time-limited) function considers those functions $s(t)$, which vanish outside the time interval $s(t) = 0$, $t \notin [0, T]$ and are of finite power $(1/T) \int_0^T |s(t)|^2 dt \leq R_0$, and the energy of which outside the frequency band $[-W, W]$ is no larger than $\varepsilon > 0$ [8 and references therein]. This notion is the key to a rigorous treatment of the capacity problem, as reported in Wyner [9], [10], to be further discussed in the following.

IV. THE GEOMETRIC APPROACH

One of Shannon’s most important ideas, which set the stage for the profound developments in information and communications theory in the last 50 years, was the geo-
metric interpretation of the communications systems in general and messages, code words, and the encoding and decoding procedures in particular. This important approach is summarized in Table 1 of Shannon’s paper, and its essence is the representation of a signal as a point in the $N$-dimensional Euclidean space. The Parseval equality (10) is used to relate the power of the continuous waveform to the Euclidean norm in the vector space of reals $\mathbb{R}^N$, where again $N \approx 2W^2T$. Source messages are also represented as points in the space of an appropriate dimension $2DT$, where $D$ represents—in Shannon’s terminology—the effective dimensionality of the message space. Thus, the encoding procedure: within this view, the transmitter is a mapping from the message space to the coded signal space, while the decoding is viewed as a mapping from the received signal space to the message space.

This geometric representation not only facilitated the adoption and use of the terminology, results, reasoning, and techniques of vector spaces and their geometry (as explicitly mentioned by Shannon and used to derive the capacity formula) but in fact has been crucial in the understanding and designing of sophisticated codes [11]. These state-of-the-art modern coding methods, about 50 years after Shannon’s paper, come close in performance to the ultimate limits of Shannon’s capacity formula for the band-limited Gaussian channel.

Before using this geometric interpretation to prove the classical Gaussian capacity formula, Shannon in Section VI of his paper gave a qualitative explanation for the threshold effect in bandwidth-expanding analog modulation methods (such as classical wide-band FM). The explanation is somewhat problematic since, in principle, the raw signal can also be transmitted as part of the coded messages in a bandwidth-expanding modulation (systematic coding when the message is discrete valued), hence alleviating to a large extent the threshold effect when an optimal demodulation procedure is used. Shannon went on to present the information compression idea qualitatively, where redundancy and irrelevant information in the message space should be ignored so that only the so-called effective message coordinates are isolated, giving rise to the reduction of dimensionality of the message space. The possibility to reduce the bandwidth of a signal at the expense of increasing the signal resolution, or deteriorating the signal-to-noise-ratio, has also been addressed.

In Section VII of his paper, Shannon exploits the geometric representation to prove Theorem 2, the celebrated capacity formula for the white Gaussian channel $C = W \log(1 + S)$ where $S = P/N$ and where $P$ is the signal power and $N$ is the noise power at the channel output. This result, as Shannon explicitly points out, was counterintuitive at the time. The theory and practice of that time supported the belief that reduction of the error probability must come at the cost of gradual monotonic decrease of the corresponding information rate. This belief was derived from the common practice of that time in analog systems and uncoded digital systems, where the increase of output signal-to-noise ratio (or the decrease of bit/symbol error rate) with a given input power implied the reduction of the information transmitted bandwidth (or rate, in the digital communications case). In Shannon’s words, his capacity formula

$$C = W \log(1 + S)$$

is a rather surprising result, since one would expect that reducing the frequency of errors would require reducing the rate of transmissions and that the rate must approach zero as the error frequency does. Actually, we can send at the rate $C$ but reduce errors by using more involved encodings and longer delays at the transmitter and receiver.

Using the observation that the noise turns the coded signal point into a cloud sphere of an essentially fixed radius (for transmission time $T \rightarrow \infty$), and adhering basically to what is currently known as sphere packing bounding, Shannon argued that no reliable transmission at rates larger than capacity $C$ is possible. On the other hand, using the fundamental idea of random selection of code books along with the geometrical interpretation, Shannon translated the determination of the probability of error to the determination of the probability of not having any of the other randomly and independently chosen code words closer to the received point than the correct (transmitted) code word. This probability is estimated in a rather straightforward fashion to yield the capacity $C$. The random coding idea first used in [1] circumvents the need to devise optimal or close to optimal codes—a goal that has not yet been fully achieved, even on the band-limited Gaussian channel. This method of random coding has been a standard essential working tool for information theorists since. The outcome that the performance averaged over all codes is in fact close to optimal is somewhat surprising, and in Shannon’s words:

It turns out, rather surprisingly, that it is possible to choose our $M$ signal functions at random from the points inside the sphere of radius $\sqrt{2WT}$, and achieve the most that is possible. Physically, this corresponds very nearly to using $M$ different samples of band-limited white noise with power $P$ as signal functions.

Wyner in [9] and [10] has rigorized the capacity derivation for the Gaussian band-limited channel, relying on the definition of approximately band-limited signals [8 and references therein]. In [10], a derivation is presented that also accounts for the physical assumption of finite accuracy, prohibiting the possibility to distinguish close enough signals. The classical Shannon capacity formula is proved for coded continuous-time waveforms that are strictly time limited to the interval $[0, T]$ (where $T \rightarrow \infty$), with an energy (norm) constraint of $PT$, and band limited to the band $[-W, W]$ at degree $\varepsilon$ (that is, the energy outside this band is not larger than $\varepsilon$ [8]). The decoder accuracy to degree $\varepsilon'$ is formulated in terms of norms such that if $y_1(t), y_2(t), 0 \leq t \leq T$ are mapped by the decoder to different code words, then $\| y_1(t) - y_2(t) \|^2 > \varepsilon'$ where $\varepsilon' > 4\varepsilon$. 

\[ 444 \]  
PROCEEDINGS OF THE IEEE, VOL. 86, NO. 2, FEBRUARY 1998
V. COLORED GAUSSIAN CHANNELS, RATE-DISTORTION THEORY, AND OTHER RESULTS

Shannon went on in Section VIII to discuss his results and provide the expression for the capacity of the colored Gaussian channel, which is now known as the “water pouring” technique [12]. This result has been rigorously derived in [12] using a Karhunen–Loève decomposition. Shannon’s result hinges on a sound engineering interpretation of splitting the available band into small subbands over which the noise spectral density is essentially constant. This conclusion that capacity can be achieved by independent narrow-band ideal communications systems, operating on disjoint frequency fragments, propelled the recent use of multicarrier systems for a variety of applications [13]. Shannon also shows that among all colored noises with a given power, the Gaussian white noise is the worst in terms of minimizing capacity. Shannon points out explicitly the considerable potential gain (8 dB in power) as predicted by information theory compared to the state-of-the-art practice of that time for digital communications, i.e., the digital pulse code modulated system. He also indicates the statistical structure of the coded transmitted waveform of a good communications system that approaches capacity. Those waveforms, in fact, should mimic the white band-limited Gaussian noise statistical properties. This intuitive prediction was rigorized and generalized only recently [14]. Some other important observations of Shannon’s, part of which appeared already in his first paper [1], are the relations (bounds) between the capacity of a Gaussian channel and that of a non-Gaussian channel having the same noise spectral density. The proof of the lower bound relies on the entropy power inequality, an important relation that is also due to Shannon. See [15] for a rigorous proof of this inequality and historical notes.

In the last section in Shannon’s paper dealing with continuous sources, Shannon presents the basic ideas underlying his later development of the rate-distortion theory [3], that is, the required rate of source coding where the information is to be regenerated with respect to a criterion of fidelity. Hinging on the geometric approach, not only does he provide the rate-distortion expression for the Gaussian memoryless source with respect to the mean square error distortion measure but, in Theorem 5, he gives upper and lower bounds on this rate-distortion function for band-limited non-Gaussian sources as well. In fact, Theorem 5 (which appears essentially already in [1]), as phrased, also implies the celebrated “Shannon’s separation theorem” [16], where with no complexity constraints, source coding (with a fidelity criterion) and channel coding can (for a single user channel) be carried out separately, with no loss of optimality.

VI. CONCLUSIONS

We conclude our introduction with a short summary of some of the more important implications and the significance of the ideas presented in Shannon’s paper.

One of the main reasons that Shannon’s paper received immediate recognition is that it is written from an intuitive engineering point of view rather than as a rigorous mathematical exposition. This paved the way to a rather immediate understanding of the revolutionary ideas of the paper, which dramatically affected the way communications theory is understood and practiced. In this sense, the paper can be viewed as a restatement from an engineering point of view of the fundamental elements of information theory first created in [1].

- The geometric approach to the communications systems design and analysis problem is fundamental, and today it is a common, almost self-evident tool in the arsenal of specialists in the field. Not only does this approach facilitate the application of the wealth of relevant mathematical terminology and results to analyze properties of communications systems quantitatively (as has indeed been demonstrated in the paper by proving the channel capacity formula) but also it is central in devising state-of-the-art codes, which are known as lattice and trellis codes [11]. Shannon used this approach in a subsequent work [17] to markedly enhance the insight into the performance and limitations of optimal communications systems operating on a Gaussian channel.

- The sampling theorem first introduced to the engineering community in the paper is probably the most useful tool to the analysis of signals and their processing and has had an enormous spectrum of applications. (See [5] for a tutorial review.) This theorem, and the notion of essentially time- and band-limited signals, motivated many future studies [8]–[10], which were able to put the fundamental results of Shannon’s paper (as the channel capacity of a band-limited Gaussian channel) on rigorous footing [9], [10], [18].

- The random coding argument, which appeared already in [1], is probably one of the strongest tools to date in proving coding theorems (direct parts), while sphere packing arguments, as demonstrated in the paper, constitute the main ingredients in proving converses to coding theorems. See standard textbooks [12], [15] for a variety of applications.

- The engineering intuition in the solution of the capacity of the colored Gaussian channel by means of frequency band partitioning, though not completely rigorous, not only led to a classical theoretical result but also set the groundwork for multicarrier communications systems, which currently approach, in a variety of applications [13] (such as the asymmetrical digital subscriber line), the ultimate limits determined by Shannon in his paper.

One of the most profound ideas is coding waveforms with respect to a reconstruction fidelity criterion. These ideas, which later matured as the rate-distortion theory [3], [19], provide the theoretical basis to quantization of analog signals (for example, speech coding, vector quantization, and the like [20]), which now are ubiquitous in our everyday life (cellular phones, for example). Shannon’s ideas, described in this paper in a lucid engineering fashion and complementing his celebrated work [1], which established
the discipline of information theory, affected in a profound fashion the very thinking on the structure, components, design, and analysis of communications systems in general. As such, and through the implementation of state-of-the-art communications systems and their implications on modern communications technology, which plays an ever increasing role in modern society, these ideas of Shannon’s had an influence well beyond the technical professional world of electronics engineers and mathematicians.

ACKNOWLEDGMENT

The authors wish to thank E. Telatar and S. Verdú for a careful reading of the manuscript and helpful comments.

REFERENCES


Aaron D. Wyner (Fellow, IEEE) received the B.S. degree from Queen’s College, City University of New York, NY, in 1960 and the B.S.E.E., M.S., and Ph.D. degrees from Columbia University, New York, NY, in 1960, 1961, and 1963, respectively.

From 1963 to 1974, he conducted research in various aspects of information and communication theory and related mechanical areas at AT&T Bell Laboratories, Murray Hill, NJ.

From 1974 to 1993, he was Head of the Communications Analysis Research Department at Bell Laboratories. During 1969–1970, he was with the Department of Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel, and the Faculty of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, on a Guggenheim Foundation fellowship. He was a full- or part-time Member of the Faculty of Columbia University, Princeton University, Princeton, NJ, and the Polytechnic Institute of Brooklyn, Brooklyn, NY.

Dr. Wyner was active in the IEEE Information Theory Society. He was on the society’s Board of Governors, Chairman of its Metropolitan New York chapter, Associate Editor for Shannon Theory (1970–1972) and Editor-in-Chief of IEEE TRANSACTIONS ON INFORMATION THEORY, and Cochairman of two international symposia (1969 and 1972) and one workshop (1984). In 1976, he was President of the society. In 1984, he received the IEEE Centennial Medal. He was the 1994 Shannon Lecturer at the International Symposium on Information Theory in Trondheim, Norway. He was a member of the Union Radio Scientifique Internationale, where he was Chairman of U.S. Commission C from 1988 to 1990 and Vice Chairman of International Commission C from 1990 to 1993. In 1994, he was elected to the U.S. National Academy of Engineering.

Shlomo Shamai (Shitz) (Fellow, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Technion—Israel Institute of Technology, Haifa, in 1975, 1981, and 1986, respectively.

During 1975–1985, he was with the Signal Corps Research Labs (Israel Defense Forces) as a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Professor. His research interests include topics in information theory and digital and analog communications. He is especially interested in theoretical limits in communication with practical constraints, digital communication in optical channels, information-theoretic models for multiuser cellular radio systems and magnetic recording, channel coding, combined modulation and coding, turbo coding, and digital spectrally efficient modulation methods employing coherent and noncoherent detection.

Dr. Shamai (Shitz) is a member of the Union Radio Scientifique Internationale. He is Associate Editor for Shannon Theory of IEEE TRANSACTIONS ON INFORMATION THEORY. He has been on the Board of Governors of the IEEE Information Theory Society since 1995.