References

[1] B. S. Bosik, "The spectral density of a coded digital signal," *Bell Syst. Tech. J.*, vol. 51, pp. 921–932, Apr. 1972.

[2] R. C. Houts and T. A. Green, "Comparing bandwidth requirements for binary baseband signals," *IEEE Trans. Commun.* (Corresp.), vol. COM-21, pp. 776–781, June 1973.

[3] P. D. Shaft, "Bandwidth compaction codes for communications," *IEEE Trans. Commun.*, vol. COM-21, pp. 687–695, June 1973.

[4] C. V. Freiman and A. D. Wyner, "Optimum block codes for noiseless input-restricted channels," *Inform. Contr.*, vol. 7, pp. 398–415, Sept. 1964.

[5] W. H. Kautz, "Fibonacci codes for synchronization control," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 284–292, Apr. 1965.

[6] D. T. Tang and L. R. Bahl, "Block codes for a class of constrained noiseless channels," *Inform. Contr.*, vol. 17, pp. 435–461, Dec. 1970.

[7] P. A. Franaszek, "Sequence-state methods for run-length-limited coding," *IBM J. Res. Develop.*, vol. 14, pp. 376–383, July 1970.

[8] E. Gorog, "Redundant alphabets with desirable frequency spectrum properties," *IBM J. Res. Develop.*, vol. 12, pp. 234–241, May 1968.

[9] H. Kobayashi, "Schemes for reduction of intersymbol interference in data transmission systems," *IBM J. Res. Develop.*, vol. 14, pp. 343–353, July 1970.

An Upper Bound Associated with Errors in Gray Code

STEPHAN R. CAVIOR

Abstract—**Suppose $0 \le i,j \le 2^n - 1$. We prove that, if i,j are encoded as binary Gray codewords whose Hamming distance is $m \ge 1$, then $|i - j| < 2^n - 2^m/3$.**

In [1], Yuen finds a lower bound on the signal error that produces an m-bit error in its Gray codeword. Denoting the Gray codeword for i by $g(i)$, he proves that if the Hamming distance between $g(i)$ and $g(j)$ is $m \ge 1$, then $|i - j| > 2^m/3$. The object of this correspondence is to establish the related upper bound.

Theorem: Suppose $0 \le i,\ j \le 2^n - 1$. If the Hamming distance between $g(i)$ and $g(j)$ is m, $m \ge 1$, then

$$|i - j| < 2^n - 2^m/3.$$

Proof: Suppose that, in binary notation, $i = (i_{n-1} \cdots i_0)_2$, $j = (j_{n-1} \cdots j_0)_2$, and that we define $l = (l_{n-1} \cdots l_0)_2$, where $l_k \equiv i_k + j_k \pmod 2$, $k = 0, 1, \cdots, n-1$. In [1], Yuen proves that l is an integer whose Gray codeword $g(l)$ has weight m; moreover, he notes that if the m ones in $g(l)$ occur in positions $k_1 < k_2 < \cdots < k_m$, then $l_k = 1$ for $k_{m-1} < k \le k_m$, $l_k = 0$ for $k_{m-2} < k \le k_{m-1}$, and so forth. This is a consequence of the fact that for any integer $j = (j_{n-1} \cdots j_0)_2$ encoded in Gray code as $g(j) = (g_{n-1}^j \cdots g_0^j)_2$,

$$j_s \equiv \sum_{i=s}^{n-1} g_i^j \pmod 2, \qquad s = 0, 1, \cdots, n-1.$$

See [2] for a proof of this formula.

Without loss of generality, assume $i > j$. By the definition of l, $i - j$ is maximized, under the constraint of a distance m between $g(i)$ and $g(j)$, if $i = l$ and $j = 0$. Furthermore, l is maximized if $l_k = 1$, for $m - 1 \le k \le n - 1$, and if the bits $l_{m-2}, l_{m-3}, \cdots, l_0$ alternate between 0 and 1 beginning with 0. For example, when $m = 5$, $l = (1 \cdots 10101)$; when $m = 6$, $l = (1 \cdots 101010)$.

In general, if $m = 2t + 1$, we have $l = 2^n - 1 - A$, where

$$A = 2^1 + 2^3 + \cdots + 2^{2t-1},$$

Manuscript received October 24, 1974.

The author is with the Department of Mathematics, State University of New York at Buffalo, Amherst, N.Y. 14226.

and if $m = 2t + 2$, we have $l = 2^n - 1 - B$, where

$$B = 2^0 + 2^2 + \cdots + 2^{2t}.$$

Note that $A + B = 2^{2t+1} - 1$ and $b = 1 + 2A$. Thus,

$$A = \frac{2^{2t+1} - 2}{3} = \left[\frac{2^{2t+1}}{3}\right]$$

and it follows that

$$B = \frac{2^{2t+2} - 1}{3} = \left[\frac{2^{2t+2}}{3}\right].$$

($[x]$ denotes the greatest integer not exceeding x). Consequently, for all $m \ge 1$, we have

$$l = 2^n - 1 - [2^m/3] < 2^n - 2^m/3.$$

References

[1] C. K. Yuen, "The separability of Gray code," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-20, p. 668, Sept. 1974.

[2] ——, "Comments on 'correction of errors in multilevel Gray coded data'," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 283–284, Mar. 1974.

A Sequential Approach to Heart-Beat Interval Classification

ERNEST T. TSUI AND EUGENE WONG, FELLOW, IEEE

Abstract—**Application of sequential testing to a Markovian model of cardiac rhythm intervals is investigated. An approximate expression for the expected number of observations is obtained for Wald's sequential test under dependent sampling. The interclass separability of three selected cardiac rhythms is then determined, and the results are used to evaluate the feasibility of an on-line implementation of a sequential classification procedure in a coronary care ward.**

I. Introduction

Because of developments in preventive therapy in cardiac intensive care wards, the problem of obtaining reliable detection of certain specific types of abnormalities in rhythm that frequently prelude serious arrhythmias has recently received much attention. Romhilt *et al.* in a recent survey [1] have indicated that the present methods of using high and low rate alarms, one minute electrocardiogram (ECG) printouts every hour, and continuous multibed supervision by skilled personnel, though very reliable in the detection of serious fatal arrhythmias, are unreliable in the detection of the premonitory rhythm changes. They cite delays of several hours in the detection of premonitory rhythms such as premature atrial contractions, premature ventricular contractions, and various ventricular arrhythmias. Thus there appears to be a need for an economical on-line automated system or subsystem that can detect certain rhythm changes reliably. Such a system can be realized only if the number of features extracted as well as the number of pattern classes considered can be minimized. Recently, a hardware monitoring system with artifact rejection and physician-controlled parameters has been used by Dell'osso [2].

Several investigators [3]–[5] have proposed (see Fig. 1) that the extraction of only the R-wave interval feature not only can

Manuscript received June 14, 1974; revised October 11, 1974. This work was supported by the Army Research Office, Durham, under Contract DAHC04-74-G0087.

The authors are with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, Calif. 94720.

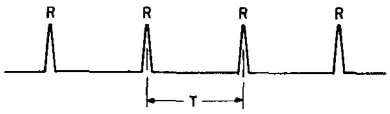

Fig. 1. R-wave interval sequence.

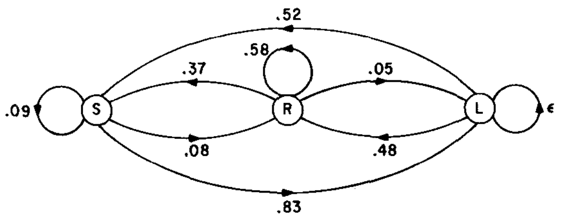

Fig. 2. Markov transition graph model of PAVC.

satisfactorily separate certain types of premonitory and serious rhythms but also allows a relatively simple feature extraction algorithm to be implemented. Additional data reduction has been obtained by transforming the R-wave intervals into three states: short, regular, and long. Certain dependencies among the states observed for some rhythms have motivated Gersch *et al.* [3] to model the ECG as a three-state first-order Markov chain (see Fig. 2). This interesting model suggests a method of classification based on contextual information.

Previous procedures for classification using only R-wave intervals have almost exclusively used fixed sample tests (an exception was the use of a finite state deterministic acceptor by Hristov [4]). Among the problems with a fixed sample test is the possibility that if the sample size is too large then a transient phenomenon such as a short string of anomalous beats may go undetected due to the relatively large number of normal beats used in the average. This situation also illustrates the difficulty in assigning *a priori* probabilities to the rhythm classes. The difficulty is similar to the radar detection problem where normal beats can be likened to having an absence of a target and abnormal beats to having a target. Thus if the observation of an abnormal rhythm or target is rare, there is difficulty in assigning meaningful *a priori* rhythm probabilities.

Recently, Patrick *et al.* [14] have made a survey of pattern recognition applications in medical diagnosis that include references to several sequential classification applications. Fu [6] has demonstrated the feasibility of using nonparametric sequential analysis in order to minimize the cost of obtaining features (clinical tests, etc.) and of misclassification by ordering the "best" features first. In this study, the fact that only intervals are extracted, the difficulty of specifying *a priori* rhythm probabilities, and the availability of the rhythm class distributions, have suggested the use of Wald's sequential probability ratio test. Its provisions for controlling error rates rather than sample size and the computationally efficient recursive form seem particularly attractive. The purpose of this study is to investigate the feasibility of utilizing Wald's sequential test in cardiac rhythm classification.

II. Markov Model of Cardiac R-Wave Rhythm

Let T_0 be the scanned average interval length. The various interval classes are $S = \{T: T \leq T_0 - \delta\}$, $L = \{T: T \geq T_0 + \delta'\}$, and $R = \{T: T_0 - \delta < T < T_0 + \delta'\}$ (Gersch *et al.* have specified $\delta = 0.1$ s and $\delta' = 0.15$ s). The sequence of observations $X_1, X_2, \cdots$ are the RR intervals reduced to the S,R,L interval classes.

Denote the probability corresponding to class i by $\mathscr{P}_i$ and let $\{X_j\}$ form a discrete parameter Markov chain with stationary transition probabilities,

$$P_i(k \mid l) = \mathscr{P}_i(X_{n+1} = k \mid X_n = l),$$
$$i = 1,\cdots,C,\ k,l = 1,2,\cdots,Q$$

where Q is the number of states and C is the number of rhythm classes (hypotheses) considered. The transition probabilities can be expressed as a matrix P_i with $P_i(k \mid l)$ as its element in the kth column and lth row.

III. Sequential Hypothesis Testing

Let $\{X_i\}$ be a sequence that is Markov under either of two hypotheses. Define the log likelihood ratio for sample size n as,

$$S_n(X) = \log \frac{P_2(X_1)}{P_1(X_1)} + \sum_{j=1}^{n-1} \log \frac{P_2(X_{j+1} \mid X_j)}{P_1(X_{j+1} \mid X_j)}.$$

The sequential likelihood ratio test is given as follows: if

$$S_n(X) \geq B \text{ choose } P_2;\text{ if}$$

$$S_n(X) \leq A \text{ choose } P_1;\text{ and if}$$

$A < S_n < B$, continue testing. Wald [7] has derived a well-known sequential test in which the absorbing or decision boundaries depend only on the two kinds of errors. If excess over boundaries is neglected, the approximate expression for the two boundaries are $A \simeq \log(\varepsilon_{12}/1 - \varepsilon_{21})$ and $B \simeq \log(1 - \varepsilon_{12}/\varepsilon_{21})$, where ε_{ij} is the error of choosing hypothesis i given that hypothesis j is true. If we let two and one denote abnormal and normal rhythms, respectively, ε_{12} is the false negative error rate and ε_{21} is the false alarm error rate.

IV. Performance of Wald Sequential Test

A measure of performance in the sequential test is given by the average sample sizes needed to reach a decision for each hypothesis, i.e., $E_i(n)$. The derivation of an expression for the average number of observations for a first-order Markov sequence was first suggested by Bellman [9], and similar results have been derived subsequently by a number of authors [10]–[12].

Let us establish some preliminary notation before presenting Bellman's result and a simplified expression under the assumption of stationarity. Let $H(t)$ denote the stationary modified Markov matrix with elements

$$P_{kl} e^{tL(k,l)}, \qquad l,k = 1,\cdots,Q$$

where

$$L(k,l) = \log \frac{P_2(X_{n+1} = k \mid X_n = l)}{P_1(X_{n+1} = k \mid X_n = l)}$$

t is the complex variable associated with the characteristic equation of S_n, and P_{kl} is the transition probabilities of the Markov matrix defined earlier. Observe that P_{kl} and hence $H(t)$ depends on the hypothesis, and there is one for each hypothesis.

Let $\nu_1(t)$ be the row eigenvector of $H(t)$ corresponding to the dominant eigenvalue $\lambda_1(t)$, i.e., $\nu_1(t)H(t) = \lambda_1(t)\nu_1(t)$. Bellman, following the proof of Wald's fundamental identity (from which Wald's equation can also be derived for the independent identically distributed (i.i.d.) case), obtained

$$E\left\{\frac{e^{tS_n}[\lambda_1{}^{-n}(t)]\nu_1(t)(p_n \mid X_n)}{\nu_1(t)p_0}\right\} = 1$$

where p_0 is the initial probability distribution vector of $X_1(Q \times 1)$ and

$$(p_n \mid X_n = j) = j\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}.$$

By differentiating with respect to t (denoted by a prime) and setting $t = 0$, the expression for the expected number of observations for the decision is found to be

$$E\{n\} = \frac{E\{S_n\} + E\{\nu_1'(0)(p_n \mid X_n)\} - \nu_1'(0)p_0}{\lambda_1'(0)}.$$

Note that the preceding expression depends on knowledge of the distribution of the state at the decision time. Depending on the application, this knowledge may or may not be readily derivable. However, if the Markov source is assumed stationary, i.e., $p_n = p_0$, for all $n \geq 0$, a simplified expression can be derived as follows. Note

$$\begin{aligned} E\{\nu_1'(0)(p_n \mid X_n = j)\} &= E\{j\text{th element of } \nu_1'(0)\} \\ &= \sum_{j=1}^{Q} v_{1j}'(0)P_n(X_n = j) \\ &= \sum_{j=1}^{Q} v_{1j}'(0)P_0(X_0 = j) \\ &= \nu_1'(0)p_0 \end{aligned}$$

where we used the stationarity property in going from the second line to the third. Thus we obtain the simplified formula

$$E\{n\} = \frac{E\{S_n\}}{\lambda_1'(0)}.$$

Observe that the expectation operator E, in fact, depends on the hypothesis, and we shall indicate this by writing E_i for expectation with respect to the distribution corresponding to hypothesis i.

Calculation of $\lambda_1'(0)$

It can be seen that, in the Markov case, $\lambda_1'(0)$ corresponds to the expectation of the log likelihood ratio (i.i.d. case) or the average step size in a random walk with absorbing boundaries. Note that $\lambda_1'(0)$ can be obtained from the characteristic equation of $H(t)$, $\det(H(t) - \lambda_1(t)I) = 0$. Expanding the characteristic equation, we have $\lambda_1^Q(t) + C_1(t)\lambda_1^{Q-1}(t) + \cdots + C_Q(t) = 0$. Differentiating with respect to t, at $t = 0$, and noting $\lambda_1(0) = 1$ yields

$$\lambda_1'(0) = \frac{\sum_1^Q C_i'(0)}{Q + \sum_1^Q iC_{Q-i}(0)}.$$

Example: For the case $Q = 3$, the characteristic equation is

$$\lambda_1^3(t) + p(t)\lambda_1^2(t) + q(t)\lambda_1(t) + r(t) = 0$$

and

$$\lambda_1'(0) = \frac{-p'(0) - q'(0) - r'(0)}{3 + 2p(0) + q(0)}.$$

Thus

$$E_i\{n\} = \frac{+E_i\{S_n\}(3 + 2p(0) + q(0))}{-p'(0) - q'(0) - r'(0)}$$

which we will take to be the interclass separability measure for sequential testing.

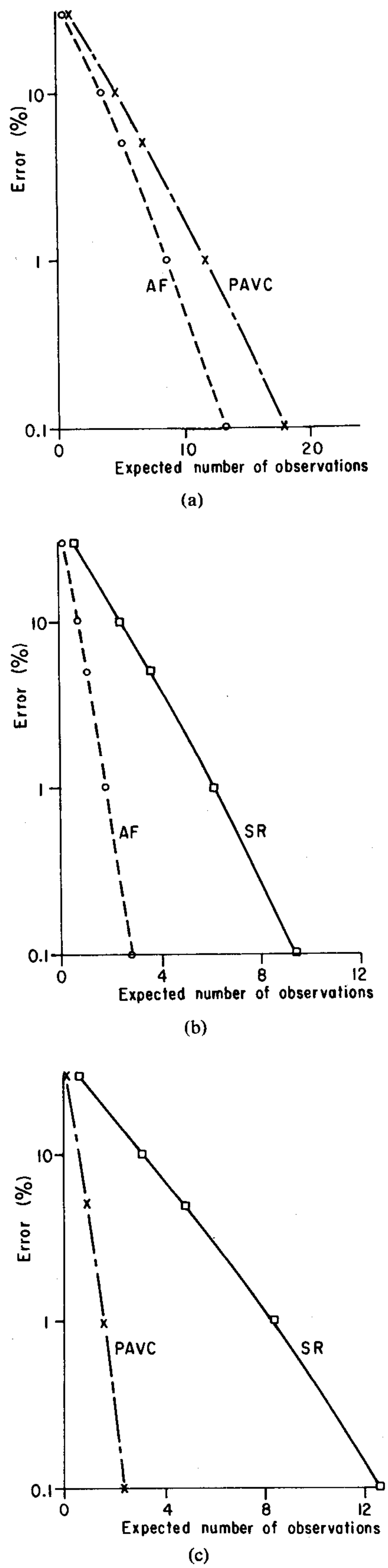

Fig. 3. Expected number of observations as function of error rate ($\varepsilon_{12} = \varepsilon_{21}$). (a) AF: PAVC separability. (b) AF: SR separability. (c) PAVC: SR separability.

V. Numerical Results

The graphs of the expected number of observations under pairwise testing of three selected rhythm classes as a function of error rate (the two kinds of errors were equated) are shown in Fig. 3(a)–(c). The rhythm classes are atrial fibrillation (AF), which is a serious rhythm; normal sinus rhythm (SR); and premature atrial and ventricular contractions (PAVC) in the presence of SR, which are important premonitory rhythms.

Small ε transition probabilities were included in the rhythm transition graphs where the experimentally estimated probabilities were zero in order to facilitate the computation of the expression for the expected number of observations. As a result, the expression for the expected number of observations can be regarded as conservative. Note that the expression is only approximate in any case because excess over the boundaries has been neglected and because of the assumption in Bellman's result that $\ln [P_2(X_1)/P_1(X_1)]$ can be neglected for large sample size. The latter assumption gives a conservative result for small sample sizes.

VI. Discussion of Results

The graphs in Fig. 3 are self-explanatory and indicate that, for all error rates considered, the average number of observations for classification will not exceed 20 observations (roughly 20 s for SR). Note that, in all the pairwise comparisons, classification of the serious rhythm consistently required fewer observations than the premonitory and normal rhythms. The expected number of observations given that PAVC is true at an error rate of 0.1 percent is approximately two observations that, on the average, will detect any short sequence of anomalous beats. Any sequence of observed samples that requires excessive time for decision, would presumably also cause difficulties with conventional monitoring. Thus the sequential system expresses its "confusion" by delaying classification rather than arbitrarily selecting a class as in a fixed sample test. In these cases a monitoring subsystem can be used to give an alarm to an operator, or to call a more elaborate program for a complete analysis. Fu [13] has proposed the use of time-varying truncated stopping boundaries as a compromise between excessive length and specifiable error rates.

Acknowledgment

The authors are grateful to Eugene Dong, Jr., M.D., for supplying the transition probabilities used in this correspondence.

References

[1] D. Romhilt, S. Bloomfield, T. Chou, and N. Fowler, "Unreliability of conventional electrocardiographic monitoring for arrhythmia detection in coronary care units," *Amer. J. Cardio.*, vol. 31, pp. 457–461, 1973.
[2] L. Dell'osso, "An arrhythmia-anomalous beat monitoring system," *IEEE Trans. Biomed. Eng.*, vol. BME-20, pp. 43–50, Jan. 1973.
[3] W. Gersch, E. Dong, Jr., and D. Eddy, "Cardiac arrhythmia classification: A heart-beat interval Markov chain approach," *Comput. Biomed. Res.*, vol. 4, pp. 385–392, 1970.
[4] H. Hristov, G. Astardjian, and C. Nachev, "An algorithm for the recognition of heart rate disturbances," *Med. Biol. Eng.*, vol. 9, pp. 221–228, 1971.
[5] W. Haisty, C. Batchlor, J. Cornfield, and H. Pipberger, "Discriminant function analysis of *RR* intervals: An algorithm for on-line arrhythmia diagnosis," *Comput. Biomed. Res.*, vol. 5, pp. 247–255, 1972.
[6] K. S. Fu and M. Loew, "Automatic medical diagnosis using nonparametric sequential classification procedures," in *Proc. Advisory Group for Aerospace Research and Development Conf.*, no. 94, 1971.
[7] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
[8] E. Lehmann, *Testing Statistical Hypotheses*. New York: Wiley, 1959.
[9] R. Bellman, "On a generalization of the fundamental identity of Wald," *Proc. Cambridge Phil. Soc.*, vol. 53, pp. 257–259, 1957.
[10] R. Phatarfod, "Sequential analysis of dependent observations," *Biometrika*, vol. 52, pp. 157–165, 1965.
[11] M. Tweedie, "Generalization of Wald's fundamental identity of sequential analysis to Markov chains," *Proc. Cambridge Phil. Soc.*, vol. 56, pp. 205–214, 1960.
[12] H. Miller, "Absorption probabilities for sums of random variables defined on a finite Markov chain," *Proc. Cambridge Phil. Soc.*, vol. 58, pp. 286–298, 1962.
[13] K. S. Fu and Y. T. Chien, "A modified sequential recognition machine using time-varying stopping boundaries," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 206–214, Apr. 1966.
[14] E. Patrick, F. Stelmack, and L. Shen, "Review of pattern recognition in medical diagnosis and consulting relative to a new system model," *IEEE Trans. Syst., Man, and Cybern.*, vol. SMC-4, pp. 1–16, Jan. 1974

Contributors

Richard R. Anderson was born in Marshall, Minn., on September 5, 1920. He received the B.S. degree in mechanical engineering from Northwestern University, Evanston, Ill., and the M.S. degree in electrical engineering from Stevens Institute of Technology, Hoboken, N.J., in 1949 and 1959, respectively.

He joined Bell Telephone Laboratories, Holmdel, N.J., in 1949 as a Member of the Technical Staff. His fields of interest have included telephone switching theory, magnetic tape recorder design, data-transmission theory, and the design of computer simulation models.

Mr. Anderson is a member of Sigma Xi, Tau Beta Pi, Pi Tau Sigma, and the American Association for the Advancement of Science.

Toby Berger (S'60–M'66–SM'74) was born in New York, N.Y., on September 4, 1940. He received the B.E. degree in electrical engineering from Yale University, New Haven, Conn., in 1962, and the M.S. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, Mass., in 1964 and 1966, respectively. He was the recipient of a National Honor Society Boeing Aircraft Scholarship and a Grumman Scholarship at Yale and held a Raytheon Company Advanced Study Program Fellowship at Harvard.

From 1962 to 1968 he was a Senior Scientist at Raytheon Company, Wayland, Mass., specializing in statistical communication theory and information theory and directing research on long-term coherent signal processing. In 1968 he joined the faculty of Cornell University, Ithaca, N.Y., where is is presently an Associate Professor of Electrical Engineering and a Consultant to the Raytheon Company. His current research interests include source and channel coding theory, ergodic theory, radar, and biochemical data processing. He is the author of *Rate Distortion Theory: A Mathematical Basis for Data Compression*, a textbook in the Prentice-Hall series in Information and System Sciences.

Dr. Berger is currently Associate Editor for the Shannon Theory of this Transactions and a member of the Board of Governors of the IEEE Information Theory Group. He was selected as a delegate to the first IEEE–USSR information theory workshop and has been awarded a 1975–76 Guggenheim Fellowship for studies in ergodic theory and information theory. Dr. Berger has served as Chairman of the IEEE Ithaca Section and is a member of the American Association for the Advancement of Science, Tau Beta Pi, and Sigma Xi.