

be maximal, its initial index must be $a_0 - \delta$ and its final index b_0 . This proves the direct assertion.

For $\delta = 1$, $K = 5$ the sequence $s = (B; AAABA)$ has δ -inner periodicities $(1, 5; 4)$ and $(2, 3; 1)$ and thus does not satisfy the conclusion. Corresponding counter examples can be constructed immediately for any K and $\delta < K - 3$.

Proof of Corollary: Suppose first that $\alpha \geq 2$ and form the periodic segment by choosing $x_a = A$, $x_{a+1} = A, \dots, x_{a+\alpha-2} = A$, $x_{a+\alpha-1} = b$ and continuing periodically. If there is an x_{a-1} , choose it to be A , thus violating the α periodicity. If there is an x_{b+1} , choose it to be A or B as necessary to violate α periodicity. If there are remaining x_i , they can be chosen to be either A or B . A sequence so constructed has the δ -inner periodicity $(a, b; \alpha)$. By the theorem it can have no other δ -inner periodicity.

If $\alpha = 1$, the construction is the same except that $x_a = \dots = x_b = A$ and $x_{a-1} = B$. The conclusion follows in the same way.

For the cases $K = 1$, $\delta \geq 0$ and $K = 2$, $\delta \geq 0$, it is trivial to construct a sequence with no δ -inner periodicities using only two letters. For the case $K = 3$, $\delta = 0$ it obviously cannot be done. For all other cases for which $\delta \geq K - 3$, it can be done as follows: let $s = (x_1, \dots, x_L)$ be given by $x_1 = \dots = x_{\delta+1} = A$, $x_{\delta+2} = \dots = x_{L-1} = B$, $x_L = A$. In order for a segment to yield a δ -inner periodicity with period one, its length must be at least $\delta + 2$. Since $L - 1 - (\delta + 2) + 1 + K - 2 \leq \delta + 1$, neither the string of the A nor the string of the B is long enough to yield a δ -inner periodicity. The segments (x_1, \dots, x_L) ,

$(x_2, \dots, x_L), \dots, (x_{\delta+1}, \dots, x_L)$ are periodic with periods $L - 1$, $L - 2, \dots, L - \delta - 1$, respectively, but none of these meet the length condition of $\delta + \alpha + 1$.

REFERENCES

- [1] N. J. Fine and H. S. Wilf, "Uniqueness theorems for periodic functions," *Proc. Amer. Math. Soc.*, vol. 16, no. 1, pp. 109-114, Feb. 1965.
- [2] S. B. Boyer and J. S. Moore, "A fast string searching algorithm," *Comm. A.C.M.*, vol. 30, pp. 762-771, July 1977.
- [3] L. J. Guibas and A. M. Odlyzko, "A new proof of the Boyer-Moore string searching algorithm," in *18th Annu. Symp. on Foundations of Comp. Science Proc.*, 1977.
- [4] W. L. Root, "Some coding results for nonprobabilistic channels," *Proc. 1979 Conf. Inform. Sciences and Systems*, The Johns Hopkins Univ., pp. 372-376, 1979.
- [5] —, "Primitive coding theory for certain classes of discrete channels," *Proc. 14th Annu. Allerton Conf. Circuit and System Theory*, Univ. of Illinois, pp. 684-693, 1976.
- [6] —, "Nonprobabilistic and partly probabilistic channel coding and time-varying channels," in *Communication Systems and Random Process Theory*, J. Skwirzynski, Ed. Alphen aan den Rijn: Sijthoff and Noordhoff, 1978, pp. 223-247.
- [7] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, no. 3, pp. 8-19, 1956.
- [8] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 1, pp. 1-7, 1979.
- [9] C. Berge, *Principles of Combinatorics*. New York and London: Academic, 1971.

Information-Singularity and Recoverability of Random Processes

BRUCE HAJEK, MEMBER, IEEE

Abstract—Two notions of information-singular and strong information-singular random processes were proposed by Berger as processes which are deterministic or negligible in a physically meaningful, information theoretic sense. This paper serves two purposes. First, it shows that strong information-singularity of a random process is equivalent to information-singularity plus a quite different property called recoverability. Secondly, it shows that these properties can be completely characterized in the case where the processes of interest are (jointly) stationary and satisfy a mild integrability condition.

Manuscript received February 10, 1980; revised June 25, 1981. This work was supported in part by the National Science Foundation under Grant NSF ENG75-20864 and in part by the Joint Services Electronics Program under Contract N00014-79-C-0424. The results of this paper were presented at the Fifteenth Annual Allerton Conference, Monticello, IL, September 1977.

The author is with the Department of Electrical Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.

I. INTRODUCTION

THE BASIC definitions and notation of the paper are presented first. The main theorems are then stated followed by remarks concerning the theorems and motivation of the definitions. For more information regarding the basic motivation and examples the reader is referred to the pioneering work of Berger [1] and to Hajek [5]. The three main theorems are proved in Sections II-IV, respectively.

A. Preliminaries

Throughout this paper (B, d) will denote a complete, separable metric space with metric d . For some fixed $\alpha > 0$ define a distortion function ρ on $B \times B$ by $\rho(a, b) = d^\alpha(a, b)$ for $a, b \in B$. Then for positive integers m , define

a distortion function ρ_m on $B^m \times B^m$ by $\rho_m(a, b) = (1/m)\sum_{i=1}^m \rho(a_i, b_i)$. It will be frequently assumed that random processes $X = \{X_k\}$ with state space (B, d) satisfy the following condition.

Integrability Condition (IC): There exists a $b^* \in B$ such that, for all k , $E\rho(X_k, b^*) < +\infty$.

A code $C = (m, e_m, d_m)$ for (B, d) consists of a positive integer m , a measurable mapping $e_m: B^m \rightarrow S$, where S is some set called the set of codewords, and a mapping $d_m: S \rightarrow B^m$. Let M denote the cardinality of S ; then C is a *fixed-length code* with rate $(\log_2 M)/m$ bits per symbol. If $S \subset \{0, 1\}^*$, where $\{0, 1\}^*$ is the set of all finite length sequences of zeros and ones, then C is also a *variable-length code*. Let $l(s)$ denote the length of $s \in \{0, 1\}^*$. If $C = (m, e_m, d_m)$ is a variable-length code for (B, d) and if $X^m = (X_1, \dots, X_m)$ is a random vector with values in B^m , then the average rate of C when applied to X^m is $El(e_m(X^m))/m$ bits per symbol. For any code $C = (m, e_m, d_m)$, the average per-letter distortion when applied to X^m is $E\rho_m(X^m, d_m \circ e_m(X^m))$, where $d_m \circ e_m$ denotes the composition of the functions d_m and e_m .

Given a random process $Z = \{Z_k\}$, Z_p^m will denote the random vector $(Z_p, Z_{p+1}, \dots, Z_m)$ for $p \leq m$, and Z^n will denote $Z_1^n = (Z_1, \dots, Z_n)$.

Let $X = \{X_k\}$ be a random process with state space (B, d) . X is *fixed-length information-singular* (resp. *variable-length information-singular*) if there exists a sequence of fixed-length (variable-length) codes $C_n = (m = 2n + 1, e_m, d_m)$ such that when C_n is applied to X_{-n}^n , the rate (average rate) and average per-letter distortion both converge to zero as n tends to infinity.

For the next definition it is natural to assume that (B, d) is a Banach space with norm $\|\cdot\|$, so that $d(a, b) = \|b - a\|$. Let $X = \{X_k\}$ and $N = \{N_k\}$ be a pair of possibly dependent random processes with common state space (B, d) . Then X is *strongly fixed-length* (resp. *variable-length*) information-singular with respect to N if there exists a sequence of fixed-length (variable-length) codes $C_n = (m = 2n + 1, e_m, d_m)$ such that when applied to $X_{-n}^n + N_{-n}^n$ the rate (average rate) tends to zero and $E\rho_m(X_{-n}^n, \hat{X}_{-n}^n) \rightarrow 0$ as $n \rightarrow \infty$, where $\hat{X}_{-n}^n = d_m \circ e_m(X_{-n}^n + N_{-n}^n)$.

If $X = \{X_k\}$ and $Z = \{Z_k\}$ are each random processes with the state space (B, d) , then X is *recoverable from Z* if there is a sequence of measurable mappings $g_n: B^m \rightarrow B^m$, $m = 2n + 1$, such that $E\rho_m(X_{-n}^n, g_n(Z_{-n}^n)) \rightarrow 0$ as $n \rightarrow +\infty$.

A notion similar to recoverability is subordination. Define a random variable X to be *subordinate* to a random variable Z if X is measurable with respect to the σ -algebra $\sigma(Z)$ generated by Z and the events of probability zero. (This definition is equivalent to that of Pinsker [9].) In case X is distributed on a separable metric space, X is subordinate to Z if and only if $P(X = f(Z)) = 1$, for some measurable function f [3, p. 605]. A random process $X = \{X_k\}$ is defined to be subordinate to a random process $Z = \{Z_k\}$ if when viewed as random variables with values in spaces consisting of infinite sequences, X is subordinate to Z .

B. The Main Theorems

The first theorem shows that strong information-singularity is equivalent to two simpler properties.

Theorem 1: Suppose that $X = \{X_k\}$ and $N = \{N_k\}$ are possibly dependent random processes each with the separable Banach space (B, d) as state space. Suppose that X satisfies the integrability condition. Then X is strongly fixed-length (resp. variable-length) information-singular with respect to N if and only if X is fixed-length (variable-length) information-singular and X is recoverable from $X + N$.

The following two theorems characterize information-singularity, recoverability, and hence strong information-singularity in case the processes of interest are stationary. See [1], [2], or [9] for the definition of entropy for abstract state space stationary random processes.

Theorem 2: If X is stationary with a complete separable state space (B, d) and if X satisfies IC, then the following are equivalent:

- a) X is fixed-length information-singular,
- b) X is variable-length information-singular,
- c) X has zero entropy.

Theorem 3: Let $X = \{X_k\}$ and $Z = \{Z_k\}$ be jointly stationary random processes. Suppose that X has a complete, separable metric state space and that X satisfies IC. Then X is recoverable from Z if and only if X is subordinate to Z .

C. Remarks

(1) One of the implications of strong information-singularity of X with respect to N is that for the purpose of transmitting blocks from the "source" process $\{X_k + N_k\}$ with small average per-letter distortion, the component X is negligible. From another point of view, we could be given a "signal" process X and a "noise" process N . The strong information-singularity of X with respect to N implies that the process X is deterministic in the following sense. Even when a noisy version of it is observed (the noise N may be measurement inaccuracy, for example), namely $\{X_k + N_k\}$, the process X can be estimated with arbitrarily small per-letter distortion and then can be conveyed over any channel of positive capacity.

If X is a "signal" process and N is "noise", recoverability of X from $X + N$ implies that if one considers long enough blocks, one can estimate blocks of the signal process from a noisy version of the signal process with high accuracy. It is easy to see that recoverability of X from $X + N$ is equivalent to recoverability of N from $X + N$. Theorem 1 demonstrates that strong information singularity is equivalent to two quite different conditions—information-singularity and recoverability.

(2) Parts of Theorem 2 were established by Berger [1]. He proved the equivalence of variable-length information-singularity and the zero entropy property for stationary processes in case the state space (B, d) is finite. He also

stated without proof that fixed-length information-singularity is also equivalent to the zero entropy property under the additional (and, as we see, unnecessary) assumption that X is ergodic.

(3) General conditions for the equivalence of fixed-rate and variable-rate information-singularity are not known for nonstationary processes.

II. CHARACTERIZATION OF STRONG INFORMATION-SINGULARITY

The purpose of this section is to prove Theorem 1; the conditions of that theorem will be assumed throughout.

A. Strong Singularity Implies Singularity Plus Recoverability

Suppose that X is strongly fixed-length (variable-length) information-singular with respect to N . Thus there exists a sequence of fixed-length (variable-length) codes $C_n = (m = 2n + 1, e_m, d_m)$ which have rate (average rate when applied to $X_{-n}^n + N_{-n}^n$) converging to zero as $n \rightarrow \infty$ and such that

$$\lim_{n \rightarrow \infty} E\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)) = 0.$$

That X is recoverable from $X + N$ is immediate—simply set $g_n = d_m \circ e_m$ to satisfy the definition of recoverability. It is proven next that X is fixed-length (variable-length) information-singular.

For convenience, assume that the codewords of each of the codes C_n are indexed or ranked in some arbitrary fashion. Define

$$s(X_{-n}^n) = E[\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)) | X_{-n}^n] \quad (1)$$

and, in the variable-length case, let

$$r(X_{-n}^n) = E[l(e_m(X_{-n}^n + N_{-n}^n)) | X_{-n}^n]. \quad (2)$$

That is, s and r are Borel measurable mappings from B^m to the reals such that $s(X_{-n}^n)$ and $r(X_{-n}^n)$ are versions of the conditional expectations in (1) and (2).

Consider now only the fixed-length case. That is, suppose X is strongly fixed-length information-singular with respect to N . Define a code $\hat{C}_n = (m = 2n + 1, \hat{e}_m, d_m)$ by defining $\hat{e}_m(c)$, for $c \in B^m$, to be that codeword for C_n of the lowest index which minimizes $\rho_m(c, d_m \circ \hat{e}_m(c))$. It is easy to check that (1) implies that

$P(\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)) \leq s(X_{-n}^n) | X_{-n}^n) > 0$, almost everywhere (a.e.). Since

$\rho_m(X_{-n}^n, d_m \circ \hat{e}_m(X_{-n}^n)) \leq \rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n))$, it follows that with probability one that

$$\rho_m(X_{-n}^n, d_m \circ \hat{e}_m(X_{-n}^n)) \leq s(X_{-n}^n).$$

Hence, the average per-letter distortion of \hat{C}_m when applied to X_{-n}^n is not greater than

$$E\delta(X_{-n}^n) = E\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)),$$

which converges to zero as $2n + 1 = m \rightarrow \infty$. Also, the rate of the fixed-length code \hat{C}_m is at most equal to that of

C_m (which converges to zero as $m \rightarrow \infty$) since the codewords of \hat{C}_m are also codewords of C_m . Hence X is fixed-length information-singular.

We will now give the corresponding proof for variable-length information-singularity. The same idea is used but now the code rate as well as distortion must be controlled. Assume that X is strongly variable-length information-singular with respect to N . Define for each m a code $\tilde{C}_n = (m = 2n + 1, \tilde{e}_m, d_m)$ by defining a coding function \tilde{e}_m as follows. For $c \in B^m$, let $\tilde{e}_m(c)$ be the codeword of C_n with lowest rank which minimizes $\rho_m(c, d_m \circ \tilde{e}_m(c))$ subject to the constraint that $l(\tilde{e}_m(c)) \leq 2r(c)$. Such codeword exists since $r(c)$ is at least as large as the length of the shortest codeword. We will show that the average rate and distortion of \tilde{C}_n when applied to X_{-n}^n are at most twice the average rate and distortion (with X_{-n}^n) of C_n when applied to $X_{-n}^n + N_{-n}^n$. This will establish the variable-length information-singularity of X .

It follows from (1) by an easily verified conditional version of Chebyshev's inequality that

$$P(\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)) > 2s(X_{-n}^n) | X_{-n}^n) < 1/2, \text{ a.e., } (3)$$

and similarly from (2) that

$$P(l(e_m(X_{-n}^n + N_{-n}^n)) > 2r(X_{-n}^n) | X_{-n}^n) < 1/2, \text{ a.e.} \quad (4)$$

Note that (3) and (4) combine to imply that with probability one there is for X_{-n}^n an $\eta \in B^m$ such that if $c = e_m(X_{-n}^n + \eta)$, then $\rho_m(X_{-n}^n, d_m(c)) \leq 2s(X_{-n}^n)$ and $l(c) \leq 2r(X_{-n}^n)$. Thus with probability one $\rho_m(X_{-n}^n, d_m \circ \tilde{e}_m(X_{-n}^n)) \leq 2s(X_{-n}^n)$ and $l(\tilde{e}_m(X_{-n}^n)) \leq 2r(X_{-n}^n)$. It follows that when \tilde{C}_n is applied to X_{-n}^n , the average distortion is at most

$$2Es(X_{-n}^n) = 2E\rho_m(X_{-n}^n, d_m \circ e_m(X_{-n}^n + N_{-n}^n)),$$

and the average rate is at most

$$2Er(X_{-n}^n)/m = 2El(e_m(X_{-n}^n))/m$$

as was to be proved.

B. Singularity Plus Recoverability Imply Strong Singularity

The key to the proof of the second half of Theorem 1 is the following lemma. The lemma states that a good estimate of a low entropy random variable may be modified to also have low entropy. Recall that $0 < \alpha < \infty$ and define $r(\alpha) = \max(1, 2^{\alpha-1})$.

Lemma 1: Let ξ and η be random variables with values in a complete, separable metric space (M, δ) . Assume that there is an $m^* \in M$ such that $E\delta^\alpha(\xi, m^*)$ is finite. Suppose that for some $D, \epsilon > 0$

$$H(\xi) < \epsilon \quad (5)$$

and

$$E\delta^\alpha(\xi, \eta) < D. \quad (6)$$

Then there exists a (Borel) measurable mapping f^* map-

ping M to a finite subset of M such that

$$H(f^*(\eta)) < 2\epsilon \quad (7)$$

and

$$E\delta^\alpha(\xi, f^*(\eta)) < 3Dr(\alpha). \quad (8)$$

Proof: To avoid problems of measurability, we can assume without loss of generality that both ξ and η are distributed on countable subsets of M . Indeed, $H(\xi) < \infty$ implies that if we ignore an event of zero probability, ξ already satisfies this condition. On the other hand, there exists a Borel measurable map $g: M \rightarrow M$ with countable range so that $E\delta^\alpha(\xi, g(\eta)) < D$. To prove this, note first that since $E\delta^\alpha(\xi, m^*) < \infty$, the variable η may be assumed bounded, for if c is large enough and η is replaced by m^* whenever $\delta(\eta, m^*) < c$, then (6) still holds. If $g_n: M \rightarrow M$ is a uniformly bounded sequence of functions, each with finite range, such that $g_m(\eta) \rightarrow \eta$, then the dominated convergence theorem yields that $E\delta^\alpha(\xi, g_n(\eta)) \rightarrow E\delta^\alpha(\xi, \eta) < D$. Simply choose $g = g_n$ for some sufficiently large n . Now if the lemma were already established for discrete random variables, it could be applied to the pair $\xi, g(\eta)$. This would imply the existence of a Borel measurable function $f: M \rightarrow M$ such that $f(M)$ is a finite set and such that (7) and (8) are satisfied with $f^* = f \circ g$. Thus, Lemma 1 is true in general if it is true when ξ and η have a discrete distribution.

We shall now assume that ξ and η are distributed on $R = \{r_1, r_2, \dots\} \subset M$ and $S = \{s_1, s_2, \dots\} \subset M$ respectively. Letting $f^*(s) = m^*$ if $s \in M - S$, it remains to specify f^* on S . Let $q(i|j) = P(\xi = r_i | \eta = s_j)$. Construct a random vector $\mathbf{W} = (W_1, W_2, \dots)$ so that $(\xi, \eta), W_1, W_2, \dots$ are mutually independent and $P(W_j = r_i) = q(i|j)$, all i, j . Associated with a given possible value w of \mathbf{W} is a function $f_w: S \rightarrow R$ defined by $f_w(s_j) = w_j$. Note that f_w is a random function mapping S to R and that $f_w(\eta)$ and $f_w(\eta)$ for w fixed are each R -valued random variables. We will show that for some choice w^* of w that the function f_w satisfies the requirements (7) and (8) for f^* .

Note that

$$\begin{aligned} P(f_{\mathbf{W}}(\eta) = r_i | \eta = s_j) &= P(f_{\mathbf{W}}(s_j) = r_i) \\ &= P(W_j = r_i) = q(i|j) \end{aligned}$$

so that $(f_{\mathbf{W}}(\eta), \eta)$ and (ξ, η) have the same distribution. It follows that

$$\begin{aligned} D > E\delta^\alpha(\xi, \eta) &= E\delta^\alpha(f_{\mathbf{W}}(\eta), \eta) \\ &= \int E\delta^\alpha(f_w(\eta), \eta) dP(\mathbf{W} = w) \end{aligned} \quad (9)$$

and that $H(f_{\mathbf{W}}(\eta)) = H(\xi) < \epsilon$. By the fact that conditional entropy is less than entropy,

$$\begin{aligned} \epsilon > H(f_{\mathbf{W}}(\eta)) &\geq H(f_w(\eta) | \mathbf{W}) \\ &= \int H(f_w(\eta)) dP(\mathbf{W} = w). \end{aligned} \quad (10)$$

Inequalities (9) and (10) yield that $P\{w: E\delta^\alpha(\eta, f_w(\eta)) \geq 2D\} < 1/2$ and $P\{w: H(f_w(\eta)) \geq 2\epsilon\} \leq 1/2$, respectively.

Thus there exists w^* so that, with $f^* = f_{w^*}$,

$$E\delta^\alpha(\eta, f^*(\eta)) < 2D \quad (11)$$

and (7) is satisfied.

By the concavity (convexity) of s^α for $\alpha < 1$ ($\alpha \geq 1$) on $s \geq 0$, for any $x, y, z \in M$,

$$\begin{aligned} \delta^\alpha(x, z) &\leq (\delta(x, y) + \delta(y, z))^\alpha \\ &\leq r(\alpha)(\delta^\alpha(x, y) + \delta^\alpha(y, z)). \end{aligned} \quad (12)$$

Using this inequality (8) follows from (6) and (11). Finally, f^* has range $\{r_1, r_2, \dots\}$ but if K is chosen large enough and $f^*(\eta)$ is changed to equal m^* unless $f^*(\eta) \in \{r_1, \dots, r_K\}$, then by the dominated convergence theorem (8) still holds. Thus, f^* satisfying Lemma 1 has been constructed.

The proof of Theorem 1 will now be completed, first for the fixed-length case and then for the variable-length case. Our proof for the fixed-length case is considerably simpler than for the variable-length case and does not depend on Lemma 1.

Suppose X is fixed-length information-singular and that X is recoverable from $X + N$. Choose any $R > 0$ and $\gamma > 0$. Let $C = (m, e_m, d_m)$ be a fixed-length code with rate at most R and with average per-letter distortion at most γ when applied to X_{-n}^n , and let $g: B^m \rightarrow B^m$ be a Borel measurable function such that $E\rho_m(X_{-n}^n, g(X_{-n}^n + N_{-n}^n)) \leq \gamma$. Such codes C and functions g exist for all large $m = 2n + 1$ by the definition of information-singularity and recoverability. We can and do assume that for $c \in B^m$, $e_m(c)$ is chosen to be a codeword of C which minimizes $\rho_m(c, d_m \circ e_m(c))$.

We shall need the fact that $\delta \triangleq (\rho_m)^{1/\alpha}$ is a metric on B^m so that (12) is valid. Indeed, by Minkowski's inequality, for $a, b, c \in B^m$,

$$\begin{aligned} \delta(a, c) &= \left(\frac{1}{m} \sum_i d^\alpha(a_i, c_i) \right)^{1/\alpha} \\ &\leq \left(\frac{1}{m} \sum_i (d(a_i, b_i) + d(b_i, c_i))^\alpha \right)^{1/\alpha} \\ &\leq \left(\frac{1}{m} \sum_i d(a_i, b_i)^\alpha \right)^{1/\alpha} + \left(\frac{1}{m} \sum_i d(b_i, c_i)^\alpha \right)^{1/\alpha} \\ &= \delta(a, b) + \delta(b, c). \end{aligned}$$

Now, let $\hat{C} = (m, \hat{e}_m, d_m)$, where $\hat{e}_m = e_m \circ g$. Then \hat{C} is a fixed-length code with rate less than or equal to the rate of C and, using (12) twice,

$$\begin{aligned} &\rho_m(X_{-n}^n, d_m \circ \hat{e}_m(X_{-n}^n + N_{-n}^n)) \\ &= \delta^\alpha(X_{-n}^n, d_m \circ e_m \circ g(X_{-n}^n + N_{-n}^n)) \\ &\leq r(\alpha)\delta^\alpha(X_{-n}^n, g(X_{-n}^n + N_{-n}^n)) \\ &\quad + r(\alpha)\delta^\alpha(g(X_{-n}^n + N_{-n}^n), d_m \circ e_m \circ g(X_{-n}^n + N_{-n}^n)) \\ &\leq r(\alpha)\delta^\alpha(X_{-n}^n, g(X_{-n}^n + N_{-n}^n)) \\ &\quad + r(\alpha)\delta^\alpha(g(X_{-n}^n + N_{-n}^n), d_m \circ e_m(X_{-n}^n)) \\ &\leq (r(\alpha) + r(\alpha)^2)\delta^\alpha(X_{-n}^n, g(X_{-n}^n + N_{-n}^n)) \\ &\quad + r(\alpha)^2\delta^\alpha(X_{-n}^n, d_m \circ e_m(X_{-n}^n)). \end{aligned}$$

Hence,

$$E\rho_m(X_{-n}^n, d_m \circ \hat{e}_m(X_{-n}^n + N_{-n}^n)) \leq \gamma(r(\alpha) + 2r(\alpha)^2).$$

Since γ and R were arbitrary, the codes \hat{C} demonstrate the strong fixed-length information-singularity of X with respect to N .

Suppose now that X is variable-length information-singular and recoverable from N , and that IC is satisfied for $b^* \in B$. Choose any $R > 0$ and $D > 0$. Let $C = (m = 2n + 1, e_m, d_m)$ be a variable-length code such that when applied to X_{-n}^n , it has average rate less than R and average per-letter distortion less than $D/2r(\alpha)$. Let $g: B^m \rightarrow B^n$ be a Borel measurable mapping such that the per-letter distortion is bounded as in the following:

$$E\rho_m(X_{-n}^n, g(X_{-n}^n + N_{-n}^n)) \leq D/2r(\alpha).$$

Such codes C and mappings g exist for all large m by the definitions of variable-length information-singularity and recoverability.

Let $M = B^m$, and define the metric $\delta = (\rho_m)^{1/\alpha}$ on M as before. Let $m^* = (b^*, \dots, b^*) \in M$ and set $\epsilon = mR$, $\eta = g(X_{-n}^n + N_{-n}^n)$ and $\xi = d_m \circ e_m(X_{-n}^n)$. Since C has average rate less than R , it follows that $H(\xi) < \epsilon$. Using (12) we obtain that

$$\begin{aligned} E\delta^\alpha(\xi, \eta) &\leq r(\alpha)(E\delta^\alpha(\xi, X_{-n}^n) + E\delta^\alpha(X_{-n}^n, \eta)) \\ &< r(\alpha)(D/2r(\alpha) + D/2r(\alpha)) = D. \end{aligned}$$

Hence, the conditions of Lemma 1 are satisfied, so that there exists a Borel measurable function $f^*: M \rightarrow M$ with finite range such that

$$\frac{1}{m} H(f^*(\eta)) < 2R \quad (13)$$

and

$$E\delta^\alpha(\xi, f^*(\eta)) < 3Dr(\alpha). \quad (14)$$

By condition (13), $f^*(\eta)$ may be noiselessly encoded with a variable-length code $(m, \hat{e}'_m, \hat{d}_m)$ with average rate at most $2R + m^{-1}$ [2]. Finally, let $\hat{C} = (m, \hat{e}_m, \hat{d}_m)$ where $\hat{e}_m = \hat{e}'_m \circ f^* \circ g$. Then \hat{C} is a variable-length code with rate at most $2R + m^{-1}$ when applied to X_{-n}^n , and

$$\begin{aligned} E\rho_m(X_{-n}^n, \hat{d}_m \circ \hat{e}_m(X_{-n}^n + N_{-n}^n)) \\ = E\delta^\alpha(X_{-n}^n, f^*(\eta)) \\ \leq r(\alpha)(E\delta^\alpha(X_{-n}^n, \xi) + E\delta^\alpha(\xi, f^*(\eta))) \\ \leq r(\alpha)(D/2r(\alpha) + 3D) \\ = D(1/2 + 3r(\alpha)). \end{aligned}$$

Since R and D were arbitrary, the strong variable-length information-singularity of X with respect to N is proven. Theorem 1 is completely proved.

III. INFORMATION-SINGULARITY OF STATIONARY PROCESSES

A. Equivalence of Fixed-Length and Variable-Length Information-Singularity

In this subsection some recent results from rate distortion theory are applied to prove the equivalence of a) and

b) in Theorem 2. The proof of Theorem 2 is completed in the next subsection.

The canonical realization of stationary random processes with state space (B, d) will be used. Coordinate functions X_n are defined on $B^\infty = \{(\dots, \theta_0, \theta_1, \dots) : \theta_i \in B\}$ by $X_n(\theta) = \theta_n$ and the shift transformation T mapping B^∞ onto itself (bijectively and measurably relative to Borel sets $\mathcal{B}(B^\infty)$) is defined by $X_n(T\theta) = X_{n+1}(\theta)$. If μ is a T invariant probability measure on $(B^\infty, \mathcal{B}(B^\infty))$, then $\{X_k\}$ is a stationary random process on the probability space $(B^\infty, \mathcal{B}(B^\infty), \mu)$. By a change of probability space, any stationary process can be thus realized. Without confusion, $(B^\infty, \mathcal{B}(B^\infty), \mu)$ itself is called a stationary random process (or source).

Let $(B^\infty, \mathcal{B}(B^\infty), \mu)$ be a stationary random process. A string $\theta \in B^\infty$ is regular if there exists a unique stationary ergodic measure μ_θ on $(B^\infty, \mathcal{B}(B^\infty))$ such that for all bounded, continuous functions $f: B^\infty \rightarrow \mathbb{R}$, $n^{-1} \sum_{i=1}^n f(T^i \theta) \rightarrow \int f d\mu_\theta$. Assuming (as we do) that (B, d) is a separable, complete metric space, a version of the ergodic decomposition theorem states that the set Λ of regular elements of B^∞ satisfies $\Lambda \in \mathcal{B}(B^\infty)$ and $\mu(\Lambda) = 1$, the function $\theta \mapsto \int f d\mu_\theta$ is measurable and shift invariant, and

$$\int_{B^\infty} f d\mu = \int_{\Lambda} \left\{ \int_{B^\infty} f d\mu_\theta \right\} d\mu$$

for $(B^\infty, \mathcal{B}(B^\infty))$ -measurable, positive functions f .

For $\theta \in \Lambda$, let $D_\theta(R)$ be the distortion-rate function for the ergodic source $(B^\infty, \mathcal{B}(B^\infty), \mu_\theta)$ when B is also used as the reproduction alphabet and $\rho = d^\alpha$ is used as distortion measure (see [2] for definitions). By well-known coding theorems, for each $R \geq 0$, $D_\theta(R)$ is the smallest number such that for each $\epsilon > 0$ there exist fixed-length codes (resp. variable-length codes) of rate (average rate) at most $R + \epsilon$ and average per-letter distortion at most $D_\theta(R) + \epsilon$ when applied to blocks from $(B^\infty, \mathcal{B}(B^\infty), \mu_\theta)$. By this characterization it is seen that $D_\theta(R)$ is right continuous in R and that $D_\theta(R)$ is a nonnegative, measurable function of θ on Λ . It also is a consequence of this characterization that $D_\theta(0) = 0$ if and only if $(B^\infty, \mathcal{B}(B^\infty), \mu_\theta)$ is fixed-length (resp. variable-length) information-singular.

Define two distortion-rate functions for a stationary random process $(B^\infty, \mathcal{B}(B^\infty), \mu)$ as follows:

$$\delta(R) = \int_{\Lambda} D_\theta(R) \mu(d\theta) \quad (15)$$

and

$$D(R) = \inf \int_{\Lambda} D_\theta(R(\theta)) \mu(d\theta), \quad (16)$$

where the infimum in (16) is over the collection of non-negative Borel-measurable functions $R(\theta)$ on Λ such that

$$\int_{\Lambda} R(\theta) \mu(d\theta) \leq R.$$

Lemma 2: Let $(B^\infty, \mathcal{B}(B^\infty), \mu)$ be a stationary random process satisfying IC and let $R \geq 0$. Then $\delta(R)$ (resp. $D(R)$) is the smallest number such that for any $\epsilon > 0$ there

exist fixed-length codes (resp. variable-length codes) with rates (average rates) at most $R + \epsilon$ and average per-letter distortion at most $\delta(R) + \epsilon$ (resp. $D(R) + \epsilon$).

Proof: Lemma 2 for fixed-length codes is a slight generalization of the combination of two theorems in Neuhoff *et al.* [8, theorems 4.2 and 5.4]. Their Theorem 4.2 states that “weakly minimax (fixed-rate) universal codes for sources satisfying IC are weighted universal as well.” By their Theorem 5.4, for the case of metric distortion ($\alpha = 1$) the source $(B^\infty, \mathcal{B}(B^\infty)\mu)$ is “weakly minimax universal.” Theorem 5.4 can easily be modified to hold for distortion d^α for any $\alpha > 0$, and then [8, theorem 4.2] implies Lemma 2 above. To modify [8, theorem 5.4], one need only establish [8, lemma A.1] for distortion d^α , which is easily done using (11). Lemma 2 for variable-length codes is proved in [10] under the condition that B is finite and can be deduced in the general case from [7, theorem 4, corollary 2].

It is now easy to prove the equivalence of a and b of Theorem 2. By Lemma 2, a stationary random process $(B^\infty, \mathcal{B}(B^\infty), \mu)$ satisfying condition IC is fixed-length (resp. variable-length) information-singular if and only if $\delta(0) = 0$ (resp. $D(0) = 0$). By (15) and (16) it is clear that $D(0) = 0$ (resp. $\delta(0) = 0$) if and only if $D_\theta(0) = 0$ for μ a.e. θ . Hence, a and b of Theorem 2 are each equivalent to the information-singularity of almost all the ergodic sub-sources of μ . We emphasize that $D(R) = \delta(R)$ is not true in general, but $D(R) = 0$ if and only if $\delta(R) = 0$.

B. Quantization and Entropy of Stationary Information-Singular Processes

To complete the proof of Theorem 2, it must be shown that a (or, equivalently b) is equivalently to c . This was done in [1] under the assumption that the state space B is finite. As noted in [1], the equivalence in the general case is then easily implied by the following proposition. The proposition was asserted without proof in [1] and is false if stationarity is relaxed to wide-sense stationarity [6].

Proposition 1: Let $X = \{X_k\}$ be a stationary process, on the complete separable metric space (B, d) , which is information-singular when distortion d^α is used. Suppose that $f: B \rightarrow F$ is Borel measurable, where F is some finite set. Then $\hat{X} = \{\hat{X}_k = f(X_k)\}$ is also information-singular.

Note that no metric on the set F , the state space of \hat{X} , is mentioned in the above proposition. However, since F is finite, any finite distortion function on F which is strictly positive off the diagonal of $F \times F$ leads to the same class of information-singular processes. Hence, we will prove the above using the Hamming metric on F , given by $d_H(a, b) = 1$ if $a \neq b$. The proof of Proposition 1 will be preceded by two short lemmas regarding quantizations.

Lemma 3: Let $\mathcal{P} = (P_1, \dots, P_N)$ be a measurable partition in a complete separable metric space (B, d) , and let ν be a finite measure on $\mathcal{B}(B)$. Then given $\epsilon > 0$ there exist disjoint, open sets G_1, \dots, G_N such that $\nu(P_i \Delta G_i) < \epsilon$ for $i = 1, \dots, N$. (Where $A \Delta B = (A - B) \cup (B - A)$ for sets A, B .)

Proof: Since finite Borel measures in metric spaces are inner-regular with respect to compact sets, there exists necessarily disjoint compact subsets K_1, \dots, K_N of B such that $K_i \subset P_i$ and $\nu(P_i - K_i) < \epsilon/N$. Let, for $1 \leq i \leq N$,

$$G_i = \{a \in B \mid d(a, K_i) < d(a, K_j) \text{ } j \neq i\},$$

where the distance between a point and set is defined as usual. The G_i are clearly disjoint, open, and $K_i \subset G_i \subset B - \bigcup_{j \neq i} K_j$. Therefore

$$\begin{aligned} \nu(P_i \Delta G_i) &= \nu(P_i - G_i) + \nu(G_i - P_i) \\ &\leq \nu(P_i - K_i) + \nu\left(B - \bigcup_{j \neq i} K_j - P_i\right) \\ &= \nu(P_i - K_i) + \nu\left(\bigcup_{j \neq i} (P_j - K_j)\right) < \epsilon. \end{aligned}$$

Lemma 3 will be used to prove the following lemma.

Lemma 4: Let X_0 be distributed on a metric space (B, d) and let $\hat{X}_0 = f(X_0)$ where f is as in Proposition 1. Then given $0 < \alpha < \infty$ and $\epsilon > 0$, there exists a $\delta > 0$ and a Borel measurable function $g: B \rightarrow F$ so that any random variable Y_0 satisfying

$$Ed^\alpha(X_0, Y_0) < \delta \quad (17)$$

also satisfies

$$P(f(X_0) \neq g(Y_0)) < \epsilon. \quad (18)$$

Proof: Let P_1, \dots, P_N be the sets of constancy of f , forming a partition of B . Define a measure ν on $(B, \mathcal{B}(B))$ by $\nu(C) = P(X_0 \in C)$. Using Lemma 3, obtain disjoint open sets G_1, \dots, G_N such that

$$\nu(G_i \Delta P_i) = P(X_0 \in G_i \Delta P_i) < \epsilon/3N. \quad (19)$$

Define $g: B \rightarrow F$ as follows. Let $g(a) = f(P_i)$ for $a \in G_i$ and let $g(a) = f(P_1)$, say, for $a \in B - G$, where $G = \bigcup_i G_i$. It follows from (19) that

$$P(f(X_0) \neq g(X_0)) \leq \sum_{i=1}^N P(X_0 \in G_i \Delta P_i) < \epsilon/3 \quad (20)$$

and that

$$P(X_0 \notin G) \leq \sum_{i=1}^N P(X_0 \in P_i - G_i) < \epsilon/3. \quad (21)$$

For $\eta > 0$, let G_i^η denote the set of points a in G_i such that the closed ball in (B, d) of radius η centered at a is contained in G_i . In view of (21) and the fact that the sets G_i are open, if η is sufficiently small then

$$P\left(X_0 \notin \bigcup_{i=1}^N G_i^\eta\right) < \epsilon/3 \quad (22)$$

(indeed, as η tends to zero the left-hand side of (22) tends to the left-most quantity in (21)). Finally, by Chebyshev's inequality, δ can be chosen so small that if (17) holds then

$$P(d(X_0, Y_0) > \eta) < \epsilon/3. \quad (23)$$

Since $f(X_0) = g(Y_0)$ on the event

$$\{f(X_0) = g(X_0)\} \cap \left\{X_0 \in \bigcup_{i=1}^N G_i^\eta\right\} \cap \{d(X_0, Y_0) \leq \eta\},$$

(18) follows from (20), (22), and (23). The proof of Lemma 4 is complete.

Proof of Proposition 1: Let $C_m = (m, e_m, d_m)$ be a sequence of fixed-length codes with rate tending to zero and with average per-letter distortion, when applied to $\{X_k\}$, also tending to zero as $m \rightarrow \infty$, (i.e., C_m is a sequence of codes demonstrating information-singularity of $\{X_k\}$). Later we will construct mappings $\hat{d}_m: S \rightarrow F^m$, where S is the set of codewords for C_m , such that

$$Ed_{H,m}(\hat{X}^m, \hat{d}_m \circ e_m(X^m)) \quad (24)$$

is arbitrarily small for large m . Here $d_{H,m}$ is the normalized Hamming metric on $F^m \times F^m$ given by

$$d_{H,m}(a, b) = \frac{1}{m} \sum_{i=1}^m d_H(a_i, b_i), \quad a, b \in F^m.$$

By the trivial fact that a (real) random variable is less than or equal to its mean on a nonempty set, for each $\hat{x}^m \in F^m$ with $P(\hat{X}^m = \hat{x}^m) > 0$ there is a $\varphi(\hat{x}^m) \in B^m$ such that

$$\begin{aligned} E[d_{H,m}(\hat{x}^m, \hat{d}_m \circ e_m(X^m)) | \hat{X}^m = \hat{x}^m] \\ \geq d_{H,m}(\hat{x}^m, \hat{d}_m \circ e_m \circ \varphi(\hat{x}^m)). \end{aligned} \quad (25)$$

Denote $\hat{e}_m = e_m \circ \varphi$ and average each side of (25) over $\hat{x} \in F^m$ using the distribution of \hat{X}^m to obtain that (24) is greater than or equal to

$$Ed_{H,m}(\hat{X}^m, \hat{d}_m \circ \hat{e}_m(\hat{X}^m)). \quad (26)$$

Define the codes $\hat{C}_m = (m, \hat{e}_m, \hat{d}_m)$. The average distortion of \hat{C}_m applied to \hat{X}^m is given by (26), and hence converges to zero as m tends to infinity since, by the construction below, the same is true of the quantity (24). Furthermore, the rate of \hat{C}_m is at most equal to the rate of C_m and thus the rate of \hat{C}_m also tends to zero as m tends to infinity. Thus, the existence of the codes \hat{C}_m implies that \hat{X} is indeed information-singular.

It remains to construct the sequence of functions \hat{d}_m so that the expectation (24) tends to zero as m tends to infinity. Let $\epsilon > 0$ and choose g and δ as in Lemma 4. Note that since X is stationary, if X_0 were replaced by X_k in both (17) and (18), then (17) would still imply (18). Choose m so large that the average per-letter distortion of C_m , when applied to X , is less than $\delta\epsilon$. That is, so that with $Y^m = (Y_1, \dots, Y_m) = d_m \circ e_m(X^m)$,

$$\frac{1}{m} \sum_{k=1}^m Ed^\alpha(X_k, Y_k) \leq \delta\epsilon.$$

It follows that if J is the set of k such that $Ed^\alpha(X_k, Y_k) \geq \delta$, then J has at most $m\epsilon$ elements. Now, if we define \hat{d}_m (with values in F^m) so that its k th component is the composition

of g with the k th component of d_m , then

$$\begin{aligned} Ed_{H,m}(\hat{X}^m, \hat{d}_m \circ e_m(X^m)) &= \frac{1}{m} \sum_{k=1}^m P(\hat{X}_k \neq g(Y_k)) \\ &= \frac{1}{m} \sum_{k \in J} P(\hat{X}_k \neq g(Y_k)) \\ &\quad + \frac{1}{m} \sum_{k \notin J} P(\hat{X}_k \neq g(Y_k)) \\ &\leq \frac{1}{m} \sum_{k \in J} 1 + \frac{1}{m} \sum_{k \notin J} \epsilon \leq 2\epsilon. \end{aligned}$$

Thus, this construction of \hat{d}_m completes the proof of Proposition 1.

Remark: The cause of difficulty in the above proof is the fact that even if one random variable converges to another, the corresponding quantized variables need not converge. In fact, Proposition 1 fails if stationarity is replaced by wide-sense stationarity [6, example].

IV. RECOVERABILITY OF JOINTLY STATIONARY RANDOM PROCESSES

Theorem 3 is a direct consequence of the equivalence of 1) and 3) in the following proposition.

Proposition 2: Let $X = \{X_k\}$ and $Z = \{Z_k\}$ be jointly stationary random processes with respective state spaces (B, d) , a complete, separable metric space, and (G, \mathcal{G}) , an arbitrary measure space. Suppose X satisfies the integrability condition IC for $b^* \in B$. Then the following are equivalent.

1) There exists a sequence of measurable functions $g_n: G^m \rightarrow B^m$, where $m = 2n + 1$, such that

$$E\rho_m(X_{-n}^n, g_n(Z_{-n}^n)) \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

2) There exists a sequence of measurable functions $f_i: G^j \rightarrow B$, where $j = 2i + 1$, such that

$$E\rho(X_0, f_i(Z_{-i}^i)) \rightarrow 0 \quad \text{as } i \rightarrow +\infty.$$

3) X is subordinate to Z .

Proof: We shall prove that 1) and 3) are each equivalent to 2).

1 \Rightarrow 2: Assume 1) is true. Write $g_n = (g_n^{(-n)}, \dots, g_n^{(n)})$. Clearly for each $n \geq 1$,

$$\min_{|k| \leq n} E\rho(X_k, g_n^{(k)}(Z_{-n}^n)) \leq E\rho_m(X_{-n}^n, g_n(Z_{-n}^n)), \quad (27)$$

which converges to zero as $n \rightarrow \infty$. Let k^* be a value of k (depending on n) which achieves the minimum on the left side of (27), and define $f_i: G^j \rightarrow B$, where $i = 2n$ and $j = 2i + 1$, by

$$f_i(a_{-i}^i) = g_n^{(k^*)}(a_{-n-k^*}^{n-k^*}).$$

By the joint stationarity of the processes X and Z ,

$$\begin{aligned} E\rho(X_0, f_i(Z_{-i}^i)) &= E\rho(X_0, g_n^{(k^*)}(Z_{-n-k^*}^n)) \\ &= E\rho(X_{k^*}, g_n^{(k^*)}(Z_{-n}^n)), \end{aligned}$$

which converges to zero as n tends to infinity. Hence 2) is established.

$2 \Rightarrow 1$: Assume that 2) is true and let $\epsilon > 0$. Let i be so large that there exists a measurable function $f_i: G^j \rightarrow B$, where $j = 2i + 1$, so that

$$E\rho(X_0, f_i(Z_{-i}^i)) \leq \epsilon/2,$$

and then choose any n so large that $2i\Gamma m^{-1} < \epsilon$, where $m = 2n + 1$ and $\Gamma = (X_0, b^*) < +\infty$. Define $g_n = (g_n^{(-n)}, \dots, g_n^{(n)}): G^n \rightarrow B$ by $g_n^{(k)}(\alpha_{-n}^n) = f_i(\alpha_{k-i}^{k+i})$ if $|k| \leq n - i$ and $g_n^{(k)}(\alpha_{-n}^n) = b^*$ otherwise. Then

$$\begin{aligned} E\rho_m(X_{-n}^n, g_n(Z_{-n}^n)) &= \frac{1}{m} \sum_{n-i < |k| \leq n} E\rho(X_k, b^*) \\ &\quad + \frac{1}{m} \sum_{|k| \leq n-i} E\rho(X_k, f_i(Z_{k-i}^{k+i})) \\ &\leq 2i\Gamma m^{-1} + \epsilon/2 < \epsilon. \end{aligned}$$

Since ϵ was arbitrary, 1) is true.

$2 \Rightarrow 3$: Assume 2) holds. Associate with each function $f_i: G^j \rightarrow B$ a function $\bar{f}_i: G^\infty \rightarrow B$ by $\bar{f}_i(\alpha) = f_i(\alpha_{-i}^i)$. The functions \bar{f}_i are measurable with respect to the infinite product σ -algebra \mathcal{G}^∞ and $E\rho(X_0, \bar{f}_i(Z))$ tends to zero as i tends to infinity. Hence, for some subsequence of integers i_k , $\bar{f}_{i_k}(Z)$ converges to X_0 with probability one. Thus, if $f: G^\infty \rightarrow B$ is defined by $f(z) = \lim_{k \rightarrow \infty} \bar{f}_{i_k}(z)$ or $f(z) = b^*$ if the limit does not exist, then $P(f(Z) = X_0) = 1$. Hence, X is subordinate to $Z = \{Z_k\}$.

$3 \Rightarrow 2$: Assume that X is subordinate to Z . Condition 3) implies there is a measurable function $f: G^\infty \rightarrow B$ such that $P(X_0 = f(Z)) = 1$. Since the σ -algebra \mathcal{G}^∞ is generated by cylinder sets and B is separable, there exists a sequence of functions $f_i: G^j \rightarrow B$, where $j = 2i + 1$, such

that the sequence $\bar{f}_i(Z) \stackrel{\Delta}{=} f_i(Z_{-i}^i)$ converges in probability to $f(Z) = X_0$. Finally, since $E\rho(X_0, b^*)$ is finite, it is easy to modify the functions f_i so that $E\rho(f_i(Z_{-i}^i), X_0)$ tends to zero.

ACKNOWLEDGMENT

I owe thanks to Professor M. B. Pursley, to Dr. Ken Mackenthun, and to an anonymous reviewer for helpful suggestions. This paper was extracted from the thesis [6], which was completed under the direction of Professor M. B. Pursley.

REFERENCES

- [1] T. Berger, "Information singular processes," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 502-511, Sept. 1975.
- [2] ———, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [3] J. Doob, *Stochastic Processes*. New York: Wiley, 1953.
- [4] R. M. Gray and J. C. Kieffer, "Mutual information rate, distortion, and quantization in metric spaces," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 412-422, July 1980.
- [5] B. Hajek, "On the strong information-singularity of certain stationary processes," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 605-609, Sept. 1979.
- [6] ———, "Information-singularity and recoverability of random processes," M.S. thesis, University of Illinois, 1977, and Coordinated Science Lab. Tech. Rep. T-51, Oct. 1977.
- [7] J. Kieffer, "A unified approach to weak universal source coding," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 674-682, Nov. 1978.
- [8] D. I. Neuhoff, R. M. Gray, and L. D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 511-523, Sept. 1975.
- [9] M. S. Pinsker, *Information and Information Stability of Random Variables*. Moscow: Akad. Nauk., 1960. (English translation by A. Feinstein, San Francisco: Holden Day, 1964.)
- [10] P. Shields *et al.*, "The distortion-rate function for non-ergodic sources," *Annals of Probability*, vol. 6, no. 1, pp. 138-143, 1978.