ON THE CAPACITY OF CHANNELS WITH
GAUSSIAN AND NON-GAUSSIAN NOISE

by

Ian W. McKeague

FSU Statistics Report M-570
February, 1981

Department of Statistics
Florida State University
Tallahassee, Florida 32306

This research was supported by the Office of Naval Research under Grant
N00014-75-C-0491
Abstract

We evaluate the information capacity of channels for which the noise process is a Gaussian measure on a quasi-complete locally convex space. The coding capacity is calculated in this setting and for time-continuous Gaussian channels using the information capacity result. The coding capacity of channels with non-Gaussian noise having finite entropy with respect to Gaussian noise of the same covariance is shown not to exceed the coding capacity of the Gaussian channel. The sensitivity of the information capacity to deviations from normality in the noise process is also investigated.
1. Introduction

In this paper we extend some of the classical results of information theory to the setting of topological spaces. The capacity of a channel is currently defined in two closely connected ways. In the information capacity sense it is the supremum of the mutual information between an allowable input and the output of the channel. In the coding capacity sense it is the highest rate at which coded messages can be sent with arbitrarily small probability of error. We shall evaluate the information capacity for the additive Gaussian channel where the Gaussian noise is defined on any quasi-complete locally convex space. The coding capacity is calculated in this setting and also for time-continuous channels. For the case of non-Gaussian noise, an upper bound on the coding capacity is given. The question of robustness of the information capacity is also investigated.

Shannon (1948) determined the information capacity of the white noise Gaussian channel with bandlimited input signals. Kadota, Zakai and Ziv (1971) rigorously treated the case with causal feedback and the Wiener process as noise. Hitsuda and Ihara (1975) extended these results to a large class of Gaussian channels with causal feedback by using the Cramér-Hida representation of the noise. Baker (1978a) obtained the information capacity of the general Gaussian channel without feedback, assuming that signal and noise belong to separable Banach spaces. Baker proved his results for separable Hilbert spaces and extended them to separable Banach spaces using the Banach-Mazur Theorem. Here we consider the problem of extending these results to some class of topological spaces beyond Banach spaces and to a case that is important for applications, the product space.
$\mathbb{R}^T$ where $T$ is an arbitrary index set.

The evaluation of the coding capacity of additive Gaussian channels has been carried out by Shannon (1948) and Wyner (1966) for bandlimited channels with stationary Gaussian noise having a flat spectral density, by Shannon (1959) for the time-discrete Gaussian channel and by Fortet (1961), Bethoux (1962) and Ash (1965) for a channel with stationary Gaussian noise and signals satisfying an average energy constraint. We shall evaluate the coding capacity for a generalization of the time-discrete Gaussian channel and use this to find the capacity of the time-continuous channel with arbitrary Gaussian noise. The coding capacity of channels with non-Gaussian noise satisfying an entropy condition is shown not to exceed the coding capacity of the Gaussian channel.
2. Preliminary Results

Let $E$ be a quasi-complete locally convex space with topological dual $E^\prime$, cylindrical $\sigma$-algebra $\sigma(E^\prime)$ and Borel $\sigma$-algebra $\mathcal{B}(E)$. Let $\mu_X, \mu_Y$ be two Gaussian probability measures on $(E, \sigma(E^\prime))$ with covariance operators $R_X, R_Y$ and means $m_X, m_Y$ respectively. (For definitions see Vakhania and Tarieladze (1978).) Assume that $R_X$ and $R_Y$ map $E^\prime$ into $E$ and $m_X, m_Y$ belong to $E$. Schwartz (1964) has shown that there exists a unique Hilbert space $H_X$ which is a vector subspace of $E$, such that the injection $j_X$ of $H_X$ into $E$ is continuous and $R = j_X j_X^\ast$. The space $H_X$ is called the reproducing kernel Hilbert space (RKHS) of $R_X$. Assume $m_X = 0$. The linear space of $\mu_X$, denoted $L_X$, is the closed subspace of $L^2(E, \mu_X)$ spanned by $E^\prime$. The space $L_X$ is a Hilbert space under the inner product induced from $L^2(E, \mu_X)$. It is well known that there exists a unitary operator $U_X: L_X \rightarrow H_X$ such that $U_X f = j_X f$, for all $f \in E^\prime$.

It is known that $\mu_X$ and $\mu_Y$ are either mutually absolutely continuous $(\mu_X \bowtie \mu_Y)$ or orthogonal $(\mu_X \perp \mu_Y)$ in the cases that $E$ is a product space IR$^T$ (see Feldman (1958), Pan Yi-Min (1966) for instance), a separable Hilbert space (Rao and Varadarajan (1963)) and a separable Banach space. (Kuelbs (1970), Baker (1978b).) A unified statement of these results, valid for any quasi-complete locally convex space, follows from the product space case and the following facts. $E$ may be identified with a subspace of $E^a$, the algebraic dual of $E^\prime$ and $\sigma(E^\prime)$ is the $\sigma$-algebra induced by the cylindrical $\sigma$-algebra on $E^a$. Endowed with the $\sigma(E^a, E^\prime)$ topology, $E^a$ is isomorphic to the product space IR$^A$, where $A$ is an algebraic basis of $E^\prime$. (See Robertson and Robertson (1964, p. 96).)
Theorem 2.1. \( \mu_X \mu Y \) if and only if the following conditions are satisfied:

(i) \( H_X = H_Y \) as subsets of \( E \);

(ii) \( m_Y \in H_X \);

(iii) \( I_Y = J J^* \) is Hilbert-Schmidt, where \( J : H_X \rightarrow H_Y \) is the injection of \( H_X \) into \( H_Y \) and \( I_Y \) is the identity on \( H_Y \).

Moreover, if conditions (i), (ii) and (iii) hold, and if \( \{ \lambda_n \} \) denotes the set of eigenvalues of \( J J^* \) which are different from 1, while \( \{ v_n \} \) denotes the corresponding sequence of normalized eigenvectors, then

\[
\frac{d \mu_Y}{d \mu_X}(x) = \exp \left[ (J J^*)^{-1/2} m_Y(x) - 1/2 \langle m_Y (J J^*)^{-1} m_Y, H_X \rangle ight] - 1/2 \sum_{n=1}^{\infty} \left[ \left( \frac{1}{v_n^2} (\frac{1}{\lambda_n} - 1) \right) \log \lambda_n \right].
\]

Furthermore, if one or more of the conditions (i), (ii), or (iii) does not hold, then \( \mu_X \perp \mu_Y \).

The following result was given by Baker (1973) for the case that \( B \) and \( E \) are separable Banach spaces and \( B \) is reflexive.

Theorem 2.2. Let \( E \) be a locally convex space, \( \mu \) a probability measure on \( (E, \sigma(E^*)) \). Suppose \( B \) is a separable or reflexive Banach space and \( j : B \rightarrow E \) is a continuous linear injection. Then the following are equivalent:

(i) \( \mu^*(j(B)) = 1 \);

(ii) \( \mu = \nu \circ j^{-1} \), where \( \nu \) is a (unique) probability measure on \( (B, \sigma(B^*)) \).

Moreover, if (i) or (ii) holds then \( \mu \) is Gaussian if and only if \( \nu \) is Gaussian. If \( B \) is both separable and reflexive then \( j(B) \in \sigma(E^*)^u \), so that condition (i) may be written \( \overline{\nu}(j(B)) = 1 \).

Proof. Write \( M = j(B) \). Then it is easy to show that \( M \cap \sigma(E^*) = j[\sigma(j^*(E^*))] \).

Also \( j^*(E) \) is total in \( B^* \) from the fact that \( E \) is locally convex and \( j \) is
injective. By Perlman (1972, Theorem 8) it follows that \( \sigma(j^*(E^*)) = \sigma(B^*) \), since it is assumed that \( B \) is separable or reflexive. Thus \( M \cap \sigma(E^*) = j[\sigma(B^*)] \).

Suppose that (i) holds. By Theorem 1.1 of Doob (1937) we may define a probability measure \( \tilde{\mu} \) on \( M \cap \sigma(E^*) \) by \( \tilde{\mu}(M \cap \Lambda) = \mu(\Lambda) \), for \( \Lambda \in \sigma(E^*) \). Define

\[ \nu(\Lambda) = \tilde{\mu}(j(\Lambda)) \],

for \( \Lambda \in \sigma(B^*) \). Then \( \nu \) is a probability measure on \( \sigma(B^*) \) and \( \nu \circ j^{-1}(\Lambda) = \tilde{\mu}(\Lambda \cap M) = \mu(\Lambda) \), for \( \Lambda \in \sigma(E^*) \). Each \( \Lambda \in \sigma(B^*) \) is of the form \( j^{-1}(\Lambda) \) for some \( \Lambda \in \sigma(E^*) \), so that \( \nu \) is unique.

Conversely, suppose (ii) holds. Let \( \Lambda \in \sigma(E^*) \) and \( \Lambda \supset M \). Then

\[ \mu(\Lambda) = \nu \circ j^{-1}(\Lambda) = \nu(B) = 1 \].

Thus \( \mu^*(j(B)) = 1 \) and (i) holds. Now let \( f \in E^* \). Then \( \nu(x) = \langle j^*f, x \rangle \leq k \rangle = \tilde{\mu}(\langle f, jx \rangle \leq k) = \tilde{\mu}(y \in M: \langle f, y \rangle \leq k) = \tilde{\mu}(M \cap \{y \in E: \langle f, y \rangle \leq k\}) = \mu(y \in E: \langle f, y \rangle \leq k) \), and by Perlman (1972, Theorem 8) it follows that \( \nu \) is Gaussian if and only if \( \nu \) is Gaussian.

Finally suppose that \( B \) is both separable and reflexive. The separability implies that \( \nu \) is a Radon measure so that \( \mu = \nu \circ j^{-1} \) is a Radon measure.

Now \( B = \bigcup_{n=1}^{\infty} nU \), where \( U \) is the unit ball of \( B \). Thus, since \( B \) is reflexive, \( U \) is weakly compact, and since \( j \) is weakly continuous (see Schaefer (1966, p. 158)) we have that \( j(B) \) is a Borel subset of \( E \). Hence \( j(B) \in \overline{\sigma(E^*)}^U \), since by a result of Badrikian and Chevet (1974, p. 347) \( \overline{\sigma(E^*)}^U = \overline{\sigma(E^*)}^U \) for Radon measures \( \mu \).

Let \((\Omega, B), (\Lambda, F)\) be measurable spaces, \( \mu_{XY} \) a probability measure on the product space \((\Omega \times \Lambda, B \times F)\) with marginals \( \mu_X \) and \( \mu_Y \). The average mutual information \( I(\mu_{XY}) \) of the measure \( \mu_{XY} \) is defined by

\[ I(\mu_{XY}) = \sup \sum_{i,j} \mu_{XY}(A_i \times B_j) \log \frac{\mu_{XY}(A_i \times B_j)}{\mu_X(A_i) \mu_Y(B_j)} \],

where the upper bound is taken over all possible finite measurable partitions \( \{A_i\}, \{B_j\} \) of \( \Omega \) and \( \Lambda \) respectively.
Let $E_X$ and $E_Y$ be quasi-complete locally convex spaces, and let $\mu_{XY}$ be a zero mean Gaussian measure on $(E_X \times E_Y, \sigma(E_X) \times \sigma(E_Y))$ with covariance operator $R: E_X^* \times E_Y^* \to E_X \times E_Y$. The marginal and cross-covariance operators of $R$ are denoted $R_X, R_Y$ and $R_{XY}$. It is known (see Baker (1973), McKeague (1981)) that $R_{XY} = j_X V_{XY} j_Y^*$, where $V_{XY}: H_Y \to H_X$ is a unique bounded linear operator. The following result was proved by Baker (1978a) for $E_X, E_Y$ Hilbert spaces and the same proof works for arbitrary $E_X, E_Y$ by using Theorem 2.1.

**Proposition 2.3.** (a) $\mu_{XY} \sim \mu_X \otimes \mu_Y$ if and only if $V_{XY}$ is Hilbert-Schmidt and $\|V_{XY}\| < 1$.

(b) If $\mu_{XY} \sim \mu_X \otimes \mu_Y$ then $I(\mu_{XY}) < \infty$ and $I(\mu_{XY}) = -\frac{1}{2} \sum \log (1 - \gamma_n)$ where $\{\gamma_n, n \geq 1\}$ are the eigenvalues of $V_{XY}^* V_{XY}$. 
3. The Information Capacity of Gaussian Channels

Suppose that messages are selected according to a probability distribution \( \mu_X \) on \((E_X', \sigma(E_X'))\) and encoded through a measurable map \( A: E_X \to E_X'. \) A noise process \( \mu_N \) on \((E_Y, \sigma(E_Y))\) is assumed to be known. The joint distribution of the transmitted message and received signal is given by

\[
\mu_{XY}(D) = \mu_X \otimes \mu_N \{(x, y): (x, Ax + y) \in D, D \in \sigma(E_X') \times \sigma(E_Y')\}.
\]

The information capacity, subject to some constraints \( Q \) on \( \mu_X \) and \( A \), is defined to be \( \sup_Q I(\mu_{XY}) \). For the remainder of this section we shall assume that \( \mu_N \) is Gaussian with zero mean and covariance operator \( R_N: E_Y \to E_Y \).

The RKHS of \( R_N \) will be denoted \( H_N \) and the injection into \( E_Y \) by \( j_N'. \) Let the measure \( \mu_X \circ A^{-1} \) be denoted \( \mu_{AX} \). The constraints \( Q \) that will be put on \( \mu_X \) and \( A \) are the minimal ones to ensure that the information capacity is finite. One such constraint is \( \mu_{AX}^*(H_N) = 1 \) (see Baker (1979a)). By Theorem 2.2, if \( H_N \) is separable then this constraint may be written \( \overline{\mu}_{AX}(H_N) = 1 \). The following result was obtained by Baker (1978a) for the Hilbert space case and the proof for general \( E_X, E_Y \) is along similar lines.

**Proposition 3.1** Suppose that \( \mu_X \) is Gaussian with zero mean and covariance operator \( R_X: E_X \to E_X' \), \( A \) is a continuous linear map, \( H_N \) is separable and \( \overline{\mu}_{AX}(H_N) = 1 \). Then there exists a unique trace-class covariance operator \( T \) on \( H_N \) such that \( AR_XA^* = j_N' T j_N^* \) and \( I(\mu_{XY}) = \frac{1}{2} \sum_n \log(1 + \gamma_n^2) \) where \( \{ \gamma_n, n \geq 1 \} \) are the eigenvalues of \( T \).

**Proof.** Since \( A \) is linear and \( \mu_X \) is Gaussian, \( \mu_{AX} \) is Gaussian with covariance operator \( AR_XA^* \). Since \( \overline{\mu}_{AX}(H_N) = 1 \) and \( H_N \) is separable, by Theorem 2.2 there exists a unique Gaussian measure \( \nu \) on \( H_N \) such that \( \mu_{AX} = \nu \circ j_N^{-1} \). Let \( T \) be the covariance operator of \( \nu \). \( T \) is trace-class and the covariance operator of \( \nu \circ j_N^{-1} \) is \( j_N' T j_N^* \). Thus \( AR_XA^* = j_N' T j_N^* \).
Let \( \{ \tau_n, n \geq 1 \} \) be the eigenvalues of \( T \). We will show that the eigenvalues of \( V^*_Y V_{XY} \) are \( \{ \tau_n (1+\tau_n)^{-1}, n \geq 1 \} \) and the result will follow from Proposition 2.3. We have \( R_Y = AR_X A^* + R_N = j_N T j_N^* + j_N j_N^* \).
Thus \( j_Y j_Y^* = j_N (I+T) j_N^* \), where \( I \) is the identity on \( H_N \). This implies that \( H_Y \) equals the RKHS of \( I+T \). Let \( \varrho \) denote the injection of \( H_Y \) into \( H_N \), so that \( j_Y = j_N \varrho \). Next we have that \( R_{XY} = j_X V j_Y^* = R_X A^* = j_X j_X^* A^* \), so that \( V j_Y^* = j_X^* A^* \) and \( A j_X = j_Y V^* \). Hence \( AR_X A^* = Aj_X j_X^* A^* = j_Y V^* V j_Y^* = j_N \varrho V^* V \varrho j_N^* \).
But we also have \( AR_X A^* = j_N T j_N^* \). Therefore \( T = \varrho V^* V \varrho \). This last equation uniquely determines \( V^* V \). Let \( \{ e_n, n \geq 1 \} \) be a CONS in \( H_N \) consisting of eigenvalues of \( T \) so that \( T = \sum_n \tau_n e_n \otimes H_N e_n \). It is now easily checked that \( V^* V = \sum_n \tau_n e_n \otimes H_Y e_n \). Moreover \( u_n = (1+\tau_n)^{1/2} e_n \), \( n \geq 1 \), is a CONS for \( H_Y \) and \( V^* V = \sum_n \tau_n (1+\tau_n)^{-1} u_n \otimes H_Y u_n \). Therefore, \( \{ \tau_n (1+\tau_n)^{-1}, n \geq 1 \} \) is the point spectrum of \( V^* V \), as required.

Remark. The hypotheses in this Proposition 3.1 may be weakened slightly. Instead of requiring that \( H_N \) be separable and \( \mu_{AX}(H_N) = 1 \), it suffices to have a separable subspace \( H_N^1 \) of \( H_N \) such that \( \mu_{AX}(H_N^1) = 1 \).

The next two results extend Theorems 1 and 2 of Baker (1978a) to quasi-complete locally convex spaces under the restriction that \( A \) be one-to-one. This restriction allows us to use a result of Dobrushin (1959) in order to reduce to the case of Gaussian messages and avoid the martingale arguments of Baker (1979b).

**Theorem 3.2.** Suppose that \( \dim(H_N) \geq M \) \((M < \infty \text{ and fixed})\), \( \dim(E_X) \geq M \) and the following constraints \( Q \) are imposed on \( \mu_X \) and \( A \):

(a) \( A \) is \( \sigma(E_X)/\sigma(E_Y) \) measurable;

(b) \( \mu_X \circ A^{-1} \) is concentrated on an \( M \)-dimensional subspace of \( H_N \).
(c) \( A \) is 1-1 and \( \frac{\mu_X}{\sigma(E_X)} / \frac{\mu_{AX}}{\sigma(E_Y)} \) bimeasurable on a measurable subset of \( E_X \) which supports \( \bar{\mu}_X \);

(d) \( \int_{E_X} ||Ax||^2 d\bar{\nu}_X(x) \leq P_0 \), where \( P_0 < \infty \) is fixed.

Then

(1) \( \sup_{Q} I(\mu_{XY}) = (M/2) \log(1 + P_0/M) \);

(2) the supremum is attained when \( \mu_X \circ A^{-1} \) is Gaussian with zero mean and covariance operator \( (P_0/M) \sum_{n=1}^{M} e_n \otimes e_n \), where \( \{e_n, n = 1, \ldots, M\} \) is an orthonormal subset of \( H_N \). The maximizing \( \mu_X \circ A^{-1} \) can be obtained with a Gaussian \( \mu_X \) and a continuous linear \( A \).

**Proof.** Let \( Q \) denote the set of \( A, \mu_X \) satisfying constraints (a)-(d), \( Q^c \) the set of probability measures \( \mu_W \) on \( E_Y \) such that \( \mu_W \) is concentrated on an \( M \)-dimensional subspace \( L_W \) of \( H_N \) and \( \int_{E_Y} ||x||^2 d\bar{\nu}_W(x) \leq P_0 \). Let \( \mu_W \in Q^c \) and let \( L_X \) denote an \( M \)-dimensional subspace of \( E_X \), \( A_0 \) an isomorphism from \( L_X \) onto \( L_W \). By a consequence of the Hahn-Banach Theorem is locally convex spaces we may extend \( A_0 \) to a continuous linear map \( A : E_X \rightarrow L_W \). \( A \) is \( \sigma(E_X)/ (E_Y) \) measurable. Since \( \mu_W \) extends to a Radon measure on \( E_Y \) we have \( \mathcal{B}(E_Y) = \sigma(E_Y) \) (see Badrikian and Chevet (1974, p. 374)), and since \( \mu_W \) is concentrated on \( L_W \) we may define a Radon measure \( \bar{\mu}_X \) on \( E_X \) by \( \mu_X(D) = \bar{\mu}_W[A(D \cap L_X)] \) for \( D \in \mathcal{B}(E_X) \). Note that \( L_X = \mathcal{B}(E_X) \subset \sigma(E_X) \) and \( \mu_X \) is concentrated on \( L_X \). Let \( C \in \sigma(E_Y) \). Then \( \mu_X(C) = \bar{\mu}_W(C \cap L_W) = \bar{\mu}_W[A^{-1}(C \cap L_W)] = \bar{\mu}_W[A^{-1}(C \cap L_X)] = \mu_X[A^{-1}(C)] \), so that \( \mu_W = \mu_X \circ A^{-1} \). Therefore \( \bar{\mu}_W = \mu_X \circ A^{-1} \) bimeasurable on \( L_X \) so that \( \mu_X \) and \( A \) satisfy constraints (b) and (d). \( A \) is 1-1 and \( \frac{\sigma(E_X)}{\sigma(E_Y)} \) bimeasurable on \( L_X \) so that \( \mu_X \) and \( A \) satisfy constraint (c). Hence \( A, \mu_X \) belong to \( Q \), and as in Baker (1978a, Lemma 4) we have \( I(\mu_X, AX+N) = I(\mu_{AX}, AX+N) \).

Conversely, given \( A, \mu_X \in Q \) it is clear that \( \mu_W \equiv \mu_X \circ A^{-1} \) belongs to \( Q^c \), and as before \( I(\mu_X, AX+N) = I(\mu_W, W+N) \). Therefore \( \sup_{A, \mu_X \in Q} I(\mu_X, AX+N) = \sup_{\mu_W \in Q^c} I(\mu_X, AX+N) \), and to prove the theorem it suffices to consider the case.
where $E_X = E_Y = E$ and $A$ is the identity on $E$.

Let $F$ be a finite dimensional subspace of $E^\prime$, $f_1, \ldots, f_n$ a basis of $F$ and define $\pi_n : E \rightarrow \mathbb{R}^n$ by

$$\pi_n(x) = (\langle f_1, x \rangle, \ldots, \langle f_n, x \rangle).$$

A subset of $C$ of $E$ is said to be a cylinder set based on $F$ if it is of the form $C = \pi_n^{-1}(D)$, where $D \in \mathcal{B}(\mathbb{R}^n)$. Let $\mathcal{C}_F$ denote the $\sigma$-algebra of all cylinder sets based on $F$, and let $\mathcal{C} = \cup \mathcal{C}_F$ where $F$ runs over all finite dimensional subspaces of $E^\prime$. It is clear that $\mathcal{C}$ is a subalgebra of $\sigma(E^\prime)$ and generates $\sigma(E^\prime)$. Therefore, by Dobrushin (1959), we have

$$I(\mu_{XY}) = \sup_F \sup_{\{C_i\}, \{D_j\}} \sum_{i,j} \log \left[ \frac{\mu_{XY}(C_i \times D_j)}{\mu_X(C_i) \mu_Y(D_j)} \right],$$

where the outside supremum runs over all finite dimensional subspaces $F$ of $E^\prime$ and the inside supremum over all possible finite $\mathcal{C}_F$-measurable partitions $\{C_i\}, \{D_j\}$ of $E$. Now, since $C_i, D_j \in \mathcal{C}$, there exist $A_i, B_j \in \mathcal{B}(\mathbb{R}^n)$ such that $C_i = \pi_n^{-1}(A_i), D_j = \pi_n^{-1}(B_j)$ and

$$\mu_{XY}(C_i \times D_j) = \mu_X \otimes \mu_Y((x,y) : (x+y) \in \pi_n^{-1}(A_i) \times \pi_n^{-1}(B_j))$$

$$= \mu_X \otimes \mu_N((x,y) : (\pi_n x, \pi_n x + \pi_n y) \in A_i \times B_j)$$

$$= \mu_X \otimes \mu_N \{ (u,v) \in \mathbb{R}^n : (u, u+v) \in A_i \times B_j \},$$

where $\mu_X^n = \mu_X \circ \pi_n^{-1}$ and $\mu_N^n = \mu_N \circ \pi_n^{-1}$. Similarly $\mu_Y(D_j) = \mu_X \otimes \mu_N^n (u,v) \in \mathbb{R}^n : u + v \in B_j$. Therefore, the inside supremum in the expression for $I(\mu_{XY})$ is equal to the mutual information for a channel with message $\mu_X^n$ and noise $\mu_N^n$ on $\mathbb{R}^n$. For such a channel it is known (see Baker (1978a, Lemma 6)) that, when the covariance of $\mu_X^n$ is fixed and its distribution allowed to vary, the mutual information is maximized when $\mu_X^n$ is Gaussian. We conclude that in order to evaluate $\sup_Q I(\mu_{XY})$ it suffices to consider Gaussian $\mu_X$ belonging to $Q'$. We already know an expression for $I(\mu_{XY})$ in this case. By Proposition 3.1 and the remark following its proof,
\[ I(\mu_{XY}) = \frac{1}{2} \sum_{n=1}^{M} \log(1 + \tau_n) \text{ where } \tau_n \geq 0; \quad R_X = \sum_{n=1}^{M} \tau_n e_n \otimes e_n \text{ and} \]
\[ \{e_n, n = 1, \ldots, M\} \text{ are orthonormal in } H_N. \]
Constraint (d) implies that \[ \sum_{n=1}^{M} \tau_n \leq P_0. \]
It follows (see Baker (1978a, p. 84)) that \( I(\mu_{XY}) \) is maximized when \( \tau_n = P_0 / M \) for \( n = 1, \ldots, M \) and we conclude that
\[ \sup_Q I(\mu_{XY}) = (M/2) \log(1 + P_0 / M). \]
\[ \square \]

**Theorem 3.3** Suppose that \( E_X \) and \( H_N \) are infinite dimensional and \( H_N \) is separable. Impose the following constraints \( Q \) on \( \mu_X \) and \( A \):
(a), (c), (d) of Theorem 3.2 and (b) \( \mu_{AX}(H_N) = 1 \). Then \[ \sup_Q I(\mu_{XY}) = P_0 / 2, \]
and the supremum cannot be attained.

**Proof.** First constrain \( \mu_X \circ A^{-1} \) to have \( M \)-dimensional support.
From Theorem 3.2, we can then find \( \mu_{AX}(H_N) \) satisfying \( Q \) and such that
\[ I(\mu_{XY}) = (M/2) \log(1 + P_0 / M). \]
Thus, we obtain \[ \lim_{M \to \infty} I(\mu_{XY}) = P_0 / 2, \]
showing that \[ \sup_Q I(\mu_{XY}) \geq P_0 / 2. \]
To show the converse inequality use Baker (1978a, Lemma 4) to reduce to the case where \( E_X = E_Y = E \), \( A = \text{identity of } E \). As in the proof of
Theorem 3.2, \( \mu_X \) may be assumed Gaussian. Then by Proposition 3.1 and (c) it follows that \[ \sup_Q I(\mu_{XY}) \leq \frac{1}{2} \sum_{n=1}^{M} \tau_n \leq \frac{1}{2} P_0. \]
If the supremum is attained it must be attained by a Gaussian \( \mu_X \) on \( E \). Then, using the notation of Proposition 3.1, we have \[ \sum_{n=1}^{M} \log(1 + \tau_n) = P_0. \]
But from Theorem 2.2 and the constraints \( \mu_{AX}(H_N) = 1 \),
\[ \int_E ||x||_{H_N}^2 d\mu_X(x) \leq P_0, \] it follows that \( \sum_{n=1}^{M} \tau_n \leq P_0. \) These two equations can hold simultaneously only if \( P_0 = 0 \). Hence \( \sup_Q I(\mu_{XY}) \) cannot be attained. \( \square \)
4. **Robustness of the Information Capacity**

In this section we investigate the robustness and sensitivity of the information capacity to small deviations from normality in the noise process $v_N$. Gualtierotti (1979, 1980) introduced a class of contaminated Gaussian laws, called QN-laws, and studied the information capacity problem for channels having QN-laws as noise. Here we make use of an inequality of Ihara (1978) to put bounds on the information capacity of these contaminated Gaussian channels and give conditions under which their information capacity tends to the information capacity of the corresponding Gaussian channel.

Let $P_1$ and $P_2$ be two probability measures defined on the same measurable space $(\Omega, \mathcal{F})$. The entropy $H_{P_2}(P_1)$ of $P_1$ with respect to $P_2$ is defined by $H_{P_2}(P_1) = \sup \sum_i P_1(C_i) \log \frac{P_1(C_i)}{P_2(C_i)}$, where the supremum is taken over all finite measurable partitions $\{C_i\}$ of $\Omega$. Let $E$ be a locally convex space and let $v_N$ be a noise process on $(E, \sigma(E'))$ which is not necessarily Gaussian. Assume however that $v_N$ is of weak second order. In the notation of Section 3 we shall here only be considering channels with $E_X = E_Y = E$ and $\Lambda = $ the identity on $E$. Let $K$ be a class of Gaussian covariance operators on $E$. Let $X$ be the class of weak second-order measures $\mu_X$ on $(E, \sigma(E'))$ having a covariance operator belonging to $K$. We assume that $X$ gives the class of all messages $\mu_X$ which can be input to the channel. The constraints of Theorems 3.2 and 3.3 are of this form. Let $C(\mu_N; X)$ denote the information capacity $\sup \{I(\mu_{XY}^*: \mu_X \in X)\}$. The following result is an extension of Ihara's inequality to the case of locally convex spaces.
Theorem 4.1. Suppose that there exists a zero mean Gaussian measure \( \mu^e_N \) on \( (E, \sigma(E')) \) having the same covariance operator as \( \mu^e_N \). Then

\[
C(\mu^e_N; X) \leq C(\mu^e_N; X) \leq C(\mu^e_N; X) + H^e_\mu(\mu^e_N).
\]

Remark. Ihara stated this result for the case \( E = L^2[0,T] \). His proof is based on a result of Huang and Johnson (1962) which holds only for a small class of signal and noise covariances (see Baker (1969)). However Baker (1978a, Lemma 6) extended Huang and Johnson's result to arbitrary Gaussian covariance operators for the signal and noise so that Ihara's result is valid for \( E = L^2[0,T] \). For locally convex spaces \( E \) use the same reasoning as in Theorem 3.2 for reduction to the case of Gaussian signals.

Let \( H \) denote a separable Hilbert space, \( W \) a covariance operator on \( H \) and let \( k \) be a real number. Let \( P \) be a zero mean Gaussian probability measure on \( H \) with covariance operator \( R \). Let \( c^{-1}_Q = k^2 + \text{Tr} WR \) and \( q(x) = c_Q(k^2 + |W^k x|^2) \). Then \( q(x) \geq 0 \), \( \int_H q(x)\text{d}P = 1 \) so that the relation \( dQ = q(x)\text{d}P \) defines a probability measure \( Q \) on \( H \). \( Q \) is said to be a QN-law with parameters \( k, W \) and \( R \), and we write \( Q \doteq \text{QN}(k, W, R) \). \( ||W|| \) and \( \text{Tr} WR \) are rough measures of the "degree of deviation of \( Q \) from normality." \( Q \) has mean zero and covariance operator \( \frac{1}{2}(I + 2c_Q R^w WR^w)R^w \) (see Gualtierotti (1979)). Let \( \{Q_n, n \geq 1\} \) be a sequence of QN-laws, \( Q_n \doteq \text{QN}(k_n, W_n, R) \), with \( R \) and \( P \) fixed. We pose the following question: If \( Q_n \) converges to \( P \) (in some sense) then does the information capacity for the channel with noise \( Q_n \) converge to that with noise \( P \)?
In order to give a satisfactory answer to this question we shall make some preliminary definitions. Let \( H_R \) denote the RKHS of \( R \), \( H_n \) the RKHS of the covariance operator of \( Q_n \). We consider the following constraints on input messages \( \mu_X \). Define, for \( S < \infty \) fixed,

\[
X = \{ \mu_X : \mu_X(H_R) = 1 \text{ and } \int_H ||x||^2_{H_R} \, d\mu_X(x) \leq S \}
\]

\[
X_n = \{ \mu_X : \mu_X(H_n) = 1 \text{ and } \int_H ||x||^2_{H_n} \, d\mu_X(x) \leq S \}.
\]

We note that \( X \) and \( X_n \) are constraints on the covariance operator of \( \mu_X \) and so fall under the framework for Ihara's Theorem in Section 6.4.

**Theorem 4.2** Suppose that \( Q_n \sim QN(k_n, W_n, R) \) for \( n \geq 1 \) and \( \text{Tr}(W_n R) \rightarrow 0 \), \( \liminf \limits_{n \rightarrow \infty} k_n > 0 \). Then \( C(Q_n; X_n) \rightarrow C(P; X) \) as \( n \rightarrow \infty \).

**Proof.** First suppose that \( Q \sim QN(k, W, R) \). Let \( Q^o \) denote the zero mean Gaussian measure on \( H \) with the same covariance operator as \( Q \), namely \( R^2(I + 2cQ R^2 WR^2)R^2 \). The operator \( T = 2cQ R^2 WR^2 R^2 \) is Hilbert-Schmidt, and since \( c_Q > 0 \) and \( W \) is non-negative, \( T \) does not have \(-1\) as an eigenvalue. Hence, by Theorem 5.1 of Rao and Varadarajan (1963), we have that \( P \) and \( Q^o \) are mutually absolutely continuous. Therefore \( Q \ll Q^o \) and

\[
H_{Q^o}(Q) = \int_H \left[ \log \frac{dQ}{dQ^o} \right] dQ = \int \left[ \log \frac{dQ}{dP} \right] \frac{dQ}{dP} \, dP + \int \left[ \log \frac{dp}{dQ^o} \right] dQ.
\]
The first term on the r.h.s. of (*) is

\[
\int \left[ \log c_Q (k^2 + ||W^{1/2}x||^2) c_Q (k^2 + ||W^{1/2}x||^2) \right] dP
\]

\[
= \int \left[ \log c_Q k^2 + \log(1 + \frac{1}{k^2} ||W^{1/2}x||^2) c_Q (k^2 + ||W^{1/2}x||^2) \right] dP
\]

\[
\leq c_Q \int ||W^{1/2}x||^2 dP + \frac{c_Q}{k^2} \int ||W^{1/2}x||^4 dP \quad (\text{since } c_Q k^2 \leq 1)
\]

\[
= c_Q \text{Tr}(WR) + \frac{c_Q}{k^2} \{2 \text{Tr}(WR)^2 + (\text{Tr WR})^2 \}
\]

\[
= \frac{k^2 \text{Tr}(WR) + 2 \text{Tr}(WR)^2 + (\text{Tr WR})^2}{k^2 (k^2 + \text{Tr WR})}
\]

which tends to 0 as \( \text{Tr}(WR) \to 0 \) and \( k \) is bounded away from 0. The second term on the r.h.s. of (*) equals

\[
\int \left\{ \frac{1}{2} \sum_{i=1}^{\infty} \left[ \log(1 + \tau_{1,i}) - (\eta_{1,i}(x))^2 \tau_{1,i}(1 + \tau_{1,i})^{-1} \right] \right\} dQ(x)
\]

where \( \eta_{1,i}(x) = \sum_{j=1}^{\infty} \lambda_j^{-1/2} \langle e_j, v_i \rangle \langle e_j, x \rangle \); \( \{ \lambda_j, e_j, j \geq 1 \} \) are the eigenvalues, orthonormal eigenvectors of \( R \), and \( \{ \tau_{1,i}, v_i, i \geq 1 \} \) are the eigenvalues, orthonormal eigenvectors of \( 2c_Q R^{1/2}WR^{1/2} \). (See Rao and Varadarajan (1963, p. 318).) Therefore the second term is less than \( \frac{1}{2} \sum_{i=1}^{\infty} \log(1 + \tau_{1,i}) \leq c_Q \text{Tr}(R^{1/2}WR^{1/2}) \), which also tends to 0 as \( \text{Tr}(WR) \to 0 \) and \( k \) is bounded away from 0. Thus \( H_{Q_n}(Q_n) \to 0 \) as \( n \to \infty \). Note that \( \dim(H_R) = \dim(H_n) \) since \( p \sim Q_n^0 \). Therefore, by Theorem 3.2, if \( \dim(H_R) < \infty \) (or by
Theorem 3.3 if \( \dim(H_R) = \infty \) we have \( C(Q^\circ_n; X_n) = C(P; X) \). However, by Ihara's inequality

\[
C(Q^\circ_n; X_n) \leq C(Q^\circ_n; X_n) \leq C(Q^\circ_n; X_n) + H_{Q^\circ_n}(Q_n),
\]

and we conclude that \( C(Q_n; X_n) \to C(P; X) \) as \( n \to \infty \). \( \square \)
5. Calculation of the Coding Capacity

A basic tool for the construction of codes is a lemma of Feinstein (see Ask (1965, p. 232)) and its generalization by Kadota (1970) to channels for which the noise process does not possess a probability density. A version of these results, to be used in this section, is stated below.

Suppose that we are given measurable spaces \((\Omega, \mathcal{B})\) of transmitted signals and \((\Lambda, \mathcal{F})\) of received signals. Let \(f: \Omega \times \Lambda \to \Lambda\) be a \(\mathcal{B} \times \mathcal{F}/\mathcal{F}\) measurable function and \(\mu_{N}\) a probability measure on \((\Lambda, \mathcal{F})\). Here, \(\mu_{N}\) represents a noise process. If a message \(x \in \Omega\) is transmitted then the received signal is a random element of \(\Lambda\) having distribution \(\mu_{N} \circ f_{x}^{-1}\), where \(f_{x}: \Lambda \to \Lambda\) is defined by \(f_{x}(y) = f(x, y)\). Note that \(f_{x}\) is \(\mathcal{F}/\mathcal{F}\) measurable (see Halmos (1950, p. 142)). A code \((k, F, \varepsilon)\) is a set of \(k\) code words \(x_{1}, \ldots, x_{k}\) belonging to \(F \subset \Omega\), and a measurable partition of \(\Lambda\) into \(k\) decoding sets \(C_{1}, \ldots, C_{k}\) such that \(\mu_{N} \circ f_{x_{i}}^{-1}(C_{j}) > 1 - \varepsilon\), for \(i = 1, \ldots, k\). Such a code is used as follows: If a received signal belongs to \(C_{i}\) then the receiver concludes that \(x_{i}\) has been transmitted. The probability of error is less than \(\varepsilon\), regardless of which code word is transmitted.

Now let \(\mu_{X}\) be an arbitrary probability measure on \((\Omega, \mathcal{B})\) and define \(\mu_{XY}\) on \(\Omega \times \Lambda\) by \(\mu_{XY}(D) = \mu_{X} \oplus \mu_{N}\{(x, y): (x, f(x, y)) \in D\}\). Let \(\mu_{Y}\) be the projection of \(\mu_{XY}\) onto \(\Lambda\).

**Lemma 5.1** (Baker (1979b)). Suppose \(\mu_{N} \circ f_{x}^{-1} \sim \mu_{N}\) a.e. \(d\mu_{X}(x)\). Then \(\mu_{XY} \sim \mu_{X} \oplus \mu_{Y}\) and \(\mu_{Y} \sim \mu_{N}\). Moreover, if \([d\mu_{N} \circ f_{x}^{-1} | d\mu_{Y}] (y)\) is \(\mathcal{B} \times \mathcal{F}\) measurable then

\[
[d\mu_{XY} | d\mu_{X} \oplus d\mu_{Y}](x, y) = [d\mu_{N} \circ f_{x}^{-1} | d\mu_{Y}](y) \text{ a.e. } d\mu_{X} \oplus \mu_{Y}(x, y) .
\]
Lemma 5.2 (Feinstein, Kadota). Suppose that \( \mu_N \circ f_X^{-1} \sim \mu_N \) a.e. \( d\mu_X(x) \) and \( [d\mu_N \circ f_X^{-1}] d\mu_Y(y) \) is \( \mathcal{B} \times \mathcal{F} \) measurable. For any real \( \alpha \) let \( \Lambda = \{(x,y) \in \Omega \times \Lambda : \log[d\mu_Y d\mu_X](x,y) > \alpha \} \). Then for each integer \( k \) and \( F \in \mathcal{B} \) there exists a code \((k,F,\epsilon)\) such that
\[
\epsilon \leq k e^{-\alpha} + \mu_{\Lambda Y}(\Lambda^c) + \mu_X(F^c).
\]

For the remainder of this section we take \( \Omega = \Lambda = E \), where \( E \) is a quasi-complete locally convex space. The function \( f: E \times E \to E \) is taken to be the addition operation \( (f(x,y) = x+y, \text{for } x,y \in E) \) which is \( \sigma(E') \times \sigma(E') | \sigma(E') \) measurable. Let \( \mu_N \) be a zero mean weak second-order probability measure on \( \sigma(E') \), with covariance operator \( R_N: E' \to E \) which has RKHS denoted \( H_N \). This set-up is known as the channel with additive noise. We do not assume that \( \mu_N \) is Gaussian unless explicitly mentioned.

A code \((k,n,\epsilon)\) for this channel will be as before but with the following constraints on the code words \( x_1,\ldots,x_k \):

a) \( x_i \in H_N \), for \( i = 1,\ldots,k \);

b) \( \|x_i\|_{H_N}^2 \leq n \rho_\epsilon \), for \( i = 1,\ldots,k \);

c) \( \dim \text{sp}\{x_i, i = 1,\ldots,k\} \leq n \).

These constraints form the natural generalization of the constraint for the time-discrete Gaussian channel with average power limitation \( P_\epsilon \), as in Shannon (1959). A real number \( R \geq 0 \) is said to be a permissible rate of transmission if there exist codes \((\epsilon_n R, n, \epsilon_n)\) with \( \epsilon_n \to 0 \) as \( n \to \infty \). The coding capacity, denoted \( C_\epsilon \), is defined as the supremum of all permissible transmission rates. Let \( \mu_N^\circ \) denote the Gaussian cylindrical measure on \( E \) having zero mean and the same covariance operator as \( \mu_N \). If \( \mu_N^\circ \) is countably additive then the entropy,
\( H_{N}^{\circ}(\mu_{N}) \), of \( \mu_{N} \) with respect to \( \mu_{N}^{\circ} \), may be defined as in section 4. Otherwise we define \( H_{N}^{\circ}(\mu_{N}) \) by regarding \( \mu_{N} \) and \( \mu_{N}^{\circ} \) as measures on \( E' \times a \)
where \( \mu_{N}^{\circ} \) is countably additive.

**Theorem 5.3** 1) If \( H_{N}^{\circ}(\mu_{N}) < \infty \) then \( C_{o} \leq \frac{1}{2} \log(1+P_{o}) \).

2) If \( H_{N}^{\circ}(\mu_{N}) < \infty \) and \( \dim(H_{N}) < \infty \) then \( C_{o} = 0 \).

3) If \( \mu_{N} \) is Gaussian and \( \dim(H_{N}) = \infty \) then \( C_{o} = \frac{1}{2} \log(1+P_{o}) \).

**Proof.** First let us suppose that \( \dim(H_{N}) = \infty \) and \( \mu_{N} \) is Gaussian. Let \( \{e_{n}, n \geq 1\} \) be an orthonormal sequence in \( H_{N} \). Choose \( Q < P_{o} \) and let \( \mu_{X} \) be the distribution of the random element \( X = \sqrt{Q} \sum_{i=1}^{n} Y_{i} e_{i} \), where \( \{Y_{i}, i = 1, \ldots, n\} \) are i.i.d. \( N(0,1) \) random variables. Then \( \mu_{X} \) has mean zero and covariance operator \( Q_{i=1}^{m} e_{i} \otimes e_{i} \).

By Theorem 2.1 \( \mu_{N}^{\circ} f_{X}^{-1} \sim \mu_{X} \) a.e. \( d\mu_{X}(x) \), since \( \mu_{X} \) is concentrated on \( H_{N} \). Using Theorem 2.1 it is possible to evaluate \( [d\mu_{N}^{\circ} f_{X}^{-1}/d\mu_{Y}](y) \) and check that it is \( \sigma(E') \times \sigma(E') \) measurable.

For full details see McKeage (1980). Thus Lemma 5.2 is applicable for \( \mu_{X} \). Let \( F_{n} = \{x \in \text{sp}(e_{1}, \ldots, e_{n}) : \|x\|_{H_{N}}^{2} \leq nP_{o}\} \) and note that \( \mu_{X}(F_{n}^{C}) = P\{n^{-1} \sum_{i=1}^{n} Y_{i}^{2} > P_{o}/Q\} \rightarrow 0 \) as \( n \rightarrow \infty \), by the law of large numbers. By a similar argument to Baker (1978a, proof of Proposition 2) it can be shown that

\[
\frac{d\mu_{XY}}{d\mu_{X} \otimes \mu_{Y}} = \frac{1}{2} \sum_{i=1}^{n} (a_{i}^{2} - b_{i}^{2}) + \frac{1}{2}n \log(1+Q) ,
\]

where, under \( \mu_{XY} \), \( \{a_{1}, \ldots, a_{n}, b_{1}, \ldots, b_{n}\} \) is a set of independent Gaussian random variables with mean zero and variance \( (Q/(1+Q))^{1/2} \). Let \( \delta > 0 \), \( \alpha_{n} = \frac{1}{2}n \log(1+Q) - n\delta \) and \( A_{n} = \{(x,y) : \log\frac{d\mu_{XY}}{d\mu_{X} \otimes \mu_{Y}}(x,y) > \alpha_{n}\} \), so that
\[ A_n = \{ (x, y) : \frac{1}{2n} \sum_{i=1}^{n} (a_i^2 - b_i^2) \leq -n\delta \} \text{ and } \mu_{XY}(A_n^c) \to 0 \text{ as } n \to \infty, \] by the law of large numbers. Let \( R < \frac{1}{2} \log(1 + P^*_o) \) and \( k_n = [e^{nR}] \). Note that
\[ k_n e^{-n\delta} \leq e^{-nR - \delta n \log(1 + Q) - n\delta} \to 0 \text{ as } n \to \infty, \] provided \( Q \) is chosen sufficiently close to \( P^*_o \) and \( \delta \) is sufficiently small. Thus by
Lemma 5.2 there exist codes \( ([e^{nR}], P_n, \epsilon_n) \) such that \( \epsilon_n \to 0 \) as \( n \to \infty \). Therefore \( C_o \geq \frac{1}{2} \log(1 + P^*_o) \).

For the converse inequality suppose that \( H_o(\mu_N) < \infty \) and do not assume that \( \mu_N \) is Gaussian. Let \( 0 < \epsilon < 1/4 \) and suppose that we have a code \( (k, n, \epsilon) \) with code words \( x_1, \ldots, x_k \) and corresponding decoding sets \( C_1, \ldots, C_k \). Write \( \mu_i \equiv \mu_N \cdot x_i^{-1} \) so that \( \mu_i(C_i) \geq 1 - \epsilon \), for \( i = 1, \ldots, k \).

Using a standard technique (see Ash (1965, p. 253)) it is possible to approximate the sets \( C_1, \ldots, C_k \) by disjoint cylinder sets \( D_1, \ldots, D_k \)
and produce a new code \( (k, n, 2\epsilon) \) with code words \( x_1, \ldots, x_k \) and decoding sets \( D_1, \ldots, D_k \). Let \( F \) be a finite dimensional subspace of \( \ell^1 \) with basis \( \{ f_1, \ldots, f_r \} \) such that all the cylinder sets \( D_i, i = 1, \ldots, k \)
are based on \( F \). Let \( \pi_r \) be the map \( \pi_r : \ell^1 \to \mathbb{R}^r \) defined by
\[ \pi_r(x) = (\langle f_1, x \rangle, \ldots, \langle f_r, x \rangle) \text{ so that } D_i = \pi_r^{-1}(A_i) \text{ for some } A_i \in \mathcal{B}(\mathbb{R}^r), \]
i = 1, \ldots, k.

Since the \( D_i \) are disjoint so are the \( A_i \). Define \( y_i = \pi_r(x_i) \),
\[ \mu_i^r = \mu_i \circ \pi_r^{-1} \text{ for } i = 1, \ldots, k. \] Then
\[ \mu_i^r(A_i) = \mu_i \circ \pi_r^{-1}(A_i) = \mu_i(D_i) \geq 1 - 2\epsilon \]
so that the \( (y_i, A_i) \) form a code for the channel on \( \mathbb{R}^r \) with noise \( \mu_N \circ \pi_r^{-1} \). Notice that the \( y_i, i = 1, \ldots, k \) are distinct because we have assumed that \( \epsilon < 1/4 \); if two coincide then we would have \( \mu_i^r(A_i) > 1/2, \mu_i^r(A_j) > 1/2 \) for some \( i \neq j \), which would be a contradiction.
Let \( \mu_X \) be the measure on \( E \) given by \( \mu_X = \frac{1}{k} \sum_{i=1}^{k} \delta_{x_i} \). Define \( p : E \times E \to \mathbb{R}^X \times \mathbb{R}^Y \) by \( \rho(x,y) = (\pi_x, \pi_y) \) and put \( \mu_{XY}^X = \mu_{XY} \circ \rho^{-1} \). Then \( I(\mu_{XY}^X) \leq I(\mu_{XY}) \) (see Baker (1978a, Lemma 4)). However, since \( x_1, \ldots, x_k \) satisfy the constraints (a), (b), (c) above, \( \mu_X \) satisfies the following: \( \mu_X \) is concentrated on an \( n \) dimensional subspace of \( H_N \) and

\[
\int |x|^2 \, d\mu_X(x) \leq n^p_0.
\]

Therefore, by Theorem 3.2 and Ihara's inequality, \( I(\mu_{XY}^X) \leq \frac{n}{2} \log(1 + p_0) + H^0_N(\mu_N) \). Since \( H^0_N(\mu_N) < \infty \), it follows (see Pinsker (1960)) that \( \mu_N \ll \mu_N^0 \). Thus, each \( \mu_{XY}^X \) has a density on \( \mathbb{R}^X \) and it is possible to define the conditional entropy \( H^X(X|Y) \) of \( X \) given \( Y \) on \( \mathbb{R}^X \), as in Ash (1965, p. 241). Since \( I(\mu_{XY}^X) = H^X(X) - H^X(X|Y) \), where \( H^X(X) \) is the entropy of \( \mu_X^X \), which equals \( \log k \), we have, using Fano's inequality (see Ash (1965, p. 244)),

\[
\log k = I(\mu_{XY}^X) + H^X(X|Y) \\
\leq \frac{n}{2} \log(1 + p_0) + H^0_N(\mu_N) + \log 2 + 2\varepsilon \log k.
\]

Therefore,

\[
\log k \leq \left[ \frac{n}{2} \log(1 + p_0) + K \right] / (1 - 2\varepsilon), \tag{\dag}
\]

where \( K \) is a constant which is independent of \( n \). Let \( R > \frac{1}{2} \log(1 + p_0) \) and suppose \( k \geq e^{nR} \). Then from (\dag) it follows that

\[
\varepsilon \geq \frac{n[R - \frac{1}{2} \log(1 + p_0)] - K}{2nR} = R - \frac{1}{2} \log(1 + p_0) > 0,
\]

as \( n \to \infty \). Thus, if \( R \geq \frac{1}{2} \log(1 + p_0) \), no sequence of codes \( ([e^{nR}], n, \varepsilon_n) \) can exist with \( \varepsilon_n \to 0 \) as \( n \to \infty \), so that \( C_e \leq \frac{1}{2} \log(1 + p_0) \). Finally, assuming \( \text{dim}(H_N) = M < \infty \), we have instead of (\dag),

\[
\log k \leq \left[ \frac{M}{2} \log\left(1 + \frac{np_0}{N}\right) + K \right] / (1 - 2\varepsilon).
\]
Let $R > 0$ and suppose $k \geq e^{nR}$. Then,

$$
\varepsilon \geq [nR - \frac{M}{2} \log(1 + \frac{np_0}{M}) - K]/2nR + \frac{1}{2},
$$

as $n \to \infty$. Thus, if $R > 0$, no sequence of codes $([e^{nR}], n, \varepsilon_n)$ can exist with $\varepsilon_n \to 0$ as $n \to \infty$, so that $C_0 = 0$. □
6. The Coding Capacity of Time-Continuous Channels

Throughout this section \((\Omega, F, P)\) will be a fixed probability space and \(N = \{N_t, -\infty < t < \infty\}\) a real-valued second-order stochastic process on \((\Omega, F, P)\), which represents the noise and is not assumed Gaussian unless explicitly stated. We shall assume that the RKHS of \(N\), denoted \(H_N\), is separable.

Code words will be signals \(S_t, t \in (-\infty, \infty)\) that vanish outside the time interval \([0, T]\). A decision is made as to the identity of the code word transmitted during the interval \([0, T]\) after observing the output during that interval. Let \(H^T_N\) denote the RKHS of \(\{N_t, 0 \leq t < T\}\). The following constraints are to be imposed on code words \(S = \{S_t, -\infty < t < \infty\}\):

a) \(S\) vanishes outside \([0, T]\);

b) \(S\) restricted to \([0, T]\) belongs to \(H^T_N\);

c) \(||S||^2_{H^T_N} \leq P_0 T\).

A code \((k, T, \varepsilon)\) for this channel is a set of code words \(S^{(1)}, \ldots, S^{(k)}\) satisfying the constraints (a), (b), (c) together with disjoint decoding sets \(C_1, \ldots, C_k\), belonging to \(B(\mathbb{R}[0, T])\) such that

\[
P\{ (S_t^{(i)} + N_t)_{0}^{T} \in C_i \} \geq 1 - \varepsilon, \text{ for } i = 1, \ldots, k.
\]

A real number \(R \geq 0\) is said to be a permissible rate of transmission if there exist codes \((e^{RT}, T, \varepsilon_T)\) with \(\varepsilon_T \to 0\) as \(T \to \infty\). The coding capacity, denoted \(C_o\), is defined as the supremum of all permissible transmission rates.

Let \(N^0\) denote a zero mean Gaussian stochastic process with the same covariance as \(N\) and let \(\mu_N, \mu_{N^0}\) be the measures on \(\mathbb{R}(-\infty, \infty)\) induced by \(N\) and \(N^0\) respectively. Denote by \(H_{N^0}(N)\) the entropy of \(\mu_N\) with respect to \(\mu_{N^0}\) as defined in section 4.
Theorem 6.1. For the time-continuous channel with additive noise N:

1) If $H_N \circ(N) < \infty$ then $C_o \leq P_o/2$.

2) If $H_N \circ(N) < \infty$ and $\dim(H_N) < \infty$ then $C_o = 0$.

3) If N is Gaussian and $\dim(H_N^{T_o}) = \infty$ for some $T > 0$ then $C_o = P_o/2$.

Proof. First let us suppose that N is Gaussian and $\dim(H_N^{T_o}) = \infty$, where $T_o > 0$. Let $m$ be a fixed positive integer and let $n = mT$ vary as $T \geq T_o$ varies over positive integers. Consider the general channel on $\mathbb{R}^{[0,T_o]}$ with noise $\{N_t, 0 \leq t \leq T_o\}$ and the following constraints on code words $x_1, \ldots, x_k \in \mathbb{R}^{[0,T_o]}$:

a) $x_i \in H_N^{T_o}$, for $i = 1, \ldots, k$;

b) $\|x_i\|_{H_N^{T_o}}^2 \leq n\left(\frac{P_o}{m}\right)$, for $i = 1, \ldots, k$;

c) $\dim \text{sp}\{x_i, i = 1, \ldots, k\} \leq n$.

Since N is Gaussian and $\dim(H_N^{T_o}) = \infty$ we have, by Theorem 5.3, that the coding capacity for this channel is $\frac{1}{2} \log(1 + \frac{P_o}{m})$. Let $\epsilon > 0$ and $R' < \frac{1}{2} \log(1 + \frac{P_o}{m})$. Then there exists a code ([e^{nR'}], n, $\epsilon$) with code words $S^{(i)}$, $i = 1, \ldots, [e^{nR'}]$ and corresponding decoding sets $C_i^* \in B(\mathbb{R}^{[0,T_o]})$. Let $C_i = C_i^* \times \mathbb{R}^{(T_o,T)} \in B(\mathbb{R}^{[0,T]})$, where T is determined by the choice of n. The $C_i$ will act as decoding sets since they are disjoint. Since there is a natural norm preserving injection of $H_N^{T_o}$ into $H_N^{T}$ we shall identify each $S^{(i)} \in H_N^{T_o}$ with its corresponding element in $H_N^{T}$. Define $S^{(i)}$ to be zero outside $[0,T]$. In this way we have a code ([e^{nR'}], T, $\epsilon$) for the time-continuous channel and it follows that if $R = mR' < (\frac{m}{2}) \log(1 + \frac{P_o}{m})$ there exists a code ([e^{RT}], T, $\epsilon$) for the time-continuous channel. Therefore $C_o \geq (\frac{m}{2}) \log(1 + \frac{P_o}{m})$ and letting $m \to \infty$, we conclude that $C_o \geq P_o/2$. 
\[ \mathbb{P}\{(S_n^{(i)} + N_t)^T \in D_i, t \leq 0\} = \mathbb{P}\{(S_n^{(i)} + N_t)^T \in A_i, t \in T\} \]
\[ = \int_{\mathbb{R}^T} \chi_{A_i}(x + S_n^{(i)}) \frac{d\mu_N^T (x)}{d\mu_N^o (x)} \frac{d\mu_T (x)}{d\mu_N^o (x)} \, dx. \]

Note that \( S_n^{(i)} \to S^{(i)} \) as \( n \to \infty \), for each \( t \in [0, T] \) and \( i = 1, \ldots, k \), since \( S_n^{(i)} \to S^{(i)} \) in \( H_N^T \) as \( n \to \infty \). If \( \mu_N^o \) is degenerate at 0 for some \( t \in [0, T] \) then \( S_n^{(i)} = S_t^{(i)} = 0 \), for all \( n \), \( i = 1, \ldots, k \). Therefore, since the \( A_i \) are finite unions of intervals in \( \mathbb{R}^T \), it can be seen that
\( \chi_{A_i}(x + S_n^{(i)}) \to \chi_{A_i}(x + S^{(i)}) \) as \( n \to \infty \), a.e. \( d\mu_N^o (x) \), for \( i = 1, \ldots, k \).

Thus, by the dominated convergence theorem
\[ \mathbb{P}\{(S_n^{(i)} + N_t)^T \in D_i, t \leq 0\} \Rightarrow \mathbb{P}\{(S_t^{(i)} + N_t)^T \in D_i, t \leq 0\} \text{ as } n \to \infty, \]
for \( i = 1, \ldots, k \). But we have already shown that \( \mathbb{P}\{(S_t^{(i)} + N_t)^T \in D_i, t \leq 0\} \geq 1 - 2\varepsilon \).

Therefore, for \( n \) sufficiently large
\[ \mathbb{P}\{(S_n^{(i)} + N_t)^T \in D_i, t \leq 0\} \geq 1 - 3\varepsilon, \text{ for } i = 1, \ldots, k. \]

Also
\[ \|S_n^{(i)}\|_{H_N^T}^2 \leq n\left(\frac{P_o T}{n}\right), \text{ for } i = 1, \ldots, k \]
and
\[ \dim \text{ sp}\{S_n^{(1)}, \ldots, S_n^{(k)}\} \leq n. \]

Therefore, by the converse for the general channel in Theorem 5.3.

we have
\[ (*) \log k \leq \left[ \frac{1}{2} \log(1 + \frac{P_o T}{n}) + H_{\{N_t^{(1)}\}^T} \{N_t^{(0)} \}^T + \log 2\right]/(1 - 3\varepsilon). \]

Note that \( H_{\{N_t^{(1)}\}^T} \{N_t^{(0)} \}^T \leq H_{N^o}(N) < \infty \) and \( \frac{n}{2} \log(1 + \frac{P_o T}{n}) \leq P_o T/2 \)
and it follows that \( C_o \leq P_o/2 \). Finally, assuming \( \dim(H_{N^o}) = M < \infty \), we
have instead of (*), \( \log k \leq \left[ \frac{M}{2} \log(1 + \frac{P_{\infty}T}{M}) + K \right] (1 - 3\epsilon) \), where \( K \) is independent of \( T \) and it follows that \( C_\epsilon = 0 \). \( \square \)

**Remark.** If, in Theorem 6.1 (1), we have instead of \( H_{N_\epsilon}(N) < \infty \),

\[
\lim_{T \to \infty} \frac{H_{\{N_\epsilon\}_T}^{\{N_\epsilon\}_T}}{T} = \overline{H}_{N_\epsilon}(N),
\]

where \( \overline{H}_{N_\epsilon}(N) \) is called the entropy rate of \( N \) with respect to \( N_\epsilon \) (see Pinsker (1960), p. 77), the most we can conclude is that \( C_\epsilon \leq P_{\infty}/2 + \overline{H}_{N_\epsilon}(N) \).

ACKNOWLEDGEMENT

This work is a part of the author's doctoral dissertation in Statistics at the University of North Carolina at Chapel Hill and he gratefully acknowledges the help and support of his advisor, Professor Charles R. Baker.
REFERENCES


Wiener et Fonctions Aléatoires Gaussiennes, Lecture Notes in

BAKER, C.R. (1969), Complete simultaneous reduction of covariance

BAKER, C.R. (1973), Zero-laws for Gaussian measures on Banach Spaces,

Inform. Contr. 37, 70-89.

BAKER, C.R. (1978b), Absolute continuity for a class of probability
measures, Vector Space Measures and Applications I, Proceedings,
Dublin 1977, Lecture Notes in Mathematics 644, Springer-Verlag,
Berlin, 44-50.

BAKER, C.R. (1979a), Mutual Information, Strong Equivalence, and Signal
Sample Path Properties for Gaussian Processes, Inform. Contr. 41,
156-164.

Anal. and Appl. 69, 115-123.

BETHOUX, P. (1962), Tests et estimations concernant certaines fonctions
aléatoires, en particulier Laplacienes, Ann. Inst. Henri Poincaré,
17, 255-337.

DOBUSHIN, R.L. (1959), General formulation of Shannon's basic theorems
of the theory of information, Uspehi Mat. Nauk. 14, 3-104; English

DOOB, J.L. (1937), Stochastic processes depending on a continuous parameter,
Trans. Amer. Math. Soc. 42, 107-140.

FELDMAN, J. (1958), Equivalence and perpendicularity of Gaussian processes,

FORTET, R. (1961), Hypothesis testing and estimation for Laplacian functions.
Fourth Berkeley Symp. on Math. Stat. and Probability, 1, 289-305,
University of California Press, Berkeley, California.

GUALTIEROTTI, A.F. (1979), Sur la détection des signaux, l'information de
Shannon, et la capacité de retransmission, quand les lois du
modèle ne sont pas exactement normales. Comptes Rendus Acad. Sci.


PERLMAN, M.D. (1972), Characterizing measurability, distribution and weak convergence of random variables in Banach space by total subsets of linear functionals, *J. Multivariate Anal.* 2, 174-188.


