Gaussian Feedback Capacity

THOMAS M. COVER, FELLOW, IEEE, AND SANDEEP POMBRA

Abstract —The capacity of time-varying additive Gaussian noise channels with feedback is characterized. Toward this end, an asymptotic equipartition theorem for nonstationary Gaussian processes is proved. Then, with the aid of certain matrix inequalities, it is proved that the feedback capacity C_{FB} in bits per transmission and the nonfeedback capacity C satisfy $C \leq C_{FB} \leq 2C$ (a result obtained by Pinsker and Ebert) and $C \leq C_{FB} \leq C + \frac{1}{2}$.

I. INTRODUCTION AND SUMMARY

WE WISH to characterize the capacity of time-varying additive Gaussian noise channels with feedback. At the same time, we wish to show that the feedback capacity C_{FB} and the nonfeedback capacity C satisfy the inequalities $C_{FB} \leq 2C$ and $C_{FB} \leq C + \frac{1}{2}$ in bits per transmission. The $C_{FB} \leq 2C$ result is due to Pinsker [1] and Ebert [2].

The channel $Y_i = X_i + Z_i$, $i = 1, 2, \cdots$, has additive Gaussian noise $Z_1, Z_2, Z_3, \cdots$ where $Z^n = (Z_1, \cdots, Z_n) \sim N_n(\mu_n, K_n)$. The output is given by $Y^n = X^n + Z^n$. For block length n we shall specify a $(2^{nR}, n)$ code with codewords $x^n(W, Y^{n-1}) = (x_1(W), x_2(W, Y^1), \cdots, x_n(W, Y^{n-1}))$, $W \in \{1, 2, \cdots, 2^{nR}\}$, and decoding function $g_n: R^n \rightarrow \{1, 2, \cdots, 2^{nR}\}$. The probability of error $P_e^{(n)}$ is defined by

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \Pr\{g_n(Y^n) \neq i | x^n = x^n(i, Y^{n-1})\}$$

$$= \Pr\{g_n(Y^n) \neq W\} \quad (1)$$

where W is uniformly distributed over $\{1, 2, \cdots, 2^{nR}\}$ and independent of Z^n. The object is to communicate the index W to the receiver at high rates R with low probability of error $P_e^{(n)}$. We begin by using Fano's inequality to show that, for any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$,

$$nR \leq I(W; Y^n) + n\epsilon_n$$

where $\epsilon_n \rightarrow 0$ and W is uniformly distributed over $\{1, 2, \cdots, 2^{nR}\}$.

It would now be a mistake to use the data processing inequality to replace $I(W; Y^n)$ by the upper bound $I(X^n; Y^n)$. Although these quantities are suitably close for channels without feedback, $I(X^n; Y^n)$ blows up (e.g., when

$X^n = (X_1, X_2, \cdots, X_n) = (0, Z_1, Z_2, \cdots, Z_{n-1}))$ when feedback is allowed. Instead we prove directly the known result

$$I(W; Y^n) = h(Y^n) - h(Z^n). \quad (2)$$

This we wish to maximize. We cannot affect the noise entropy

$$h(Z^n) = \frac{1}{2} \ln (2\pi e)^n |K_Z^{(n)}| \quad (3)$$

where $|K|$ denotes the determinant of K, so we are left with maximizing $h(Y^n)$, both with and without feedback.

From the entropy maximizing property of the Gaussian distribution we have

$$h(Y^n) \leq \frac{1}{2} \ln (2\pi e)^n |K_{X+Z}^{(n)}| \quad (4)$$

where

$$K_{X+Z}^{(n)} = E(X+Z)(X+Z)^t$$

$$= K_{XX} + K_{XZ} + K_{ZX} + K_{ZZ}. \quad (5)$$

(We shall often suppress the block length n and associated matrix size in the discussion.) We are thus led to believe that the capacity of the channel is

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(W; Y^n) = \lim_{n \rightarrow \infty} \max \frac{1}{2n} \log \frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|} \quad (6)$$

where the determinant $|K_{X+Z}^{(n)}|$ is maximized under the power constraint

$$E \frac{1}{n} \sum_{i=1}^n x_i^2(W, Z^{i-1}) = \frac{1}{n} \operatorname{tr}(K_X^{(n)}) \leq P. \quad (7)$$

There are a few problems with this formulation. First, $\max(1/2n)\log(|K_{X+Z}^{(n)}|/|K_Z^{(n)}|)$ may not have a limit as $n \rightarrow \infty$, because of the time-varying nature of the noise $\{Z_i\}$. (We have not assumed stationarity.) We handle this by generalizing the notion of capacity to

$$\max_{\frac{1}{n} \operatorname{tr}(K_X^{(n)}) \leq P} \frac{1}{2n} \log \frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|} \quad (8)$$

where this quantity can be thought of as the capacity in bits per transmission if the channel is to be used for the time block $\{1, 2, \cdots, n\}$. The relationship of X to Z in the maximization depends on whether or not we have feedback.

It is now time to distinguish the characterization of capacity in the feedback cases from that in the nonfeedback cases. Clearly, since the noise Z^n and the signal index W

are independent, it follows that $X^n(W)$ is independent of Z^n if no feedback is allowed, and thus

$$K_{X+Z}^{(n)} = K_X^{(n)} + K_Z^{(n)}. \tag{9}$$

Feedback adds cross terms K_{XZ} and K_{ZX}. We now have the following informal time-varying channel capacity statements for block length n.

1) The capacity $C_{n,\text{FB}}$ in bits per transmission of the time-varying Gaussian channel with feedback is

$$C_{n,\text{FB}} = \max_{\frac{1}{n}\text{tr}(K_X^{(n)}) \le P} \frac{1}{2n} \log \frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|} \tag{10}$$

where the maximization is taken over all X^n of the form

$$X_i = \sum_{j=1}^{i-1} b_{ij} Z_j + V_i, \qquad i = 1, 2, \cdots, n \tag{11}$$

and V^n is independent of Z^n. To verify that the maximization over (11) involves no loss of generality, note that the distribution on $X^n + Z^n$ achieving (4) is Gaussian. Since Z^n is also Gaussian, it can be verified that a jointly Gaussian distribution on $(X^n, Z^n, X^n + Z^n)$ achieves the maximization in (4) and consequently in (10). Since $Z^n = Y^n - X^n$, the most general jointly normal causal dependence of X^n on Y^n is however of the form (11), where V^n plays the role of the innovations process. Recasting (10), (11) by using $X = BZ + V$ and $Y = X + Z$, we can write

$$C_{n,\text{FB}} = \max \frac{1}{2n} \log \frac{|(B+I)K_Z(B+I)' + K_V|}{|K_Z|} \tag{12}$$

where the maximum is taken over all nonnegative definite K_V and strictly lower triangular B such that

$$\text{tr}\left(BK_Z B' + K_V\right) \le nP. \tag{13}$$

(Without feedback, B is necessarily 0.)

2) The capacity C_n of the time-varying Gaussian channel without feedback is given by

$$C_n = \max_{\frac{1}{n}\text{tr}(K_X^{(n)}) \le P} \frac{1}{2n} \log \frac{|K_X^{(n)} + K_Z^{(n)}|}{|K_Z^{(n)}|}. \tag{14}$$

This reduces to waterfilling on the eigenvalues $\{\lambda_i^{(n)}\}$ of $K_Z^{(n)}$. Thus

$$C_n = \frac{1}{2n} \sum_{i=1}^{n} \log\left(1 + \frac{\left(\lambda - \lambda_i^{(n)}\right)^+}{\lambda_i^{(n)}}\right) \tag{15}$$

where $(y)^+ = \max\{y, 0\}$ and λ is chosen so that

$$\sum_{i=1}^{n} (\lambda - \lambda_i)^+ = nP. \tag{16}$$

We have upper-bounded the achievable feedback rates by $C_{n,\text{FB}}$. We subsequently prove the achievability of $C_{n,\text{FB}}$ by proving the existence of $(2^{n(C_{n,\text{FB}} - \epsilon)}, n)$ codes with $P_e^{(n)} \to 0$, for any $\epsilon > 0$. To do this we use a random coding argument. This requires the use of the asymptotic equipartition property (AEP). Unfortunately, the AEP usually only holds for ergodic stochastic processes, and ergodicity

is too much to require of $\{Z_i\}_{i=1}^\infty$ and pointless to request of $\{Y_i\}_{i=1}^\infty$, because we would then be restricting the maximization with a resulting loss of generality. Surprisingly, the AEP holds for arbitrary (nonergodic) Gaussian processes as proved in Section V. (See also Pinsker [9].) Thus we can indeed prove that rates less than C_n bits per transmission can be achieved for n transmissions over this channel. We state the following theorem.

Theorem 1: Let $\{Z_i\}_{i=1}^\infty$ be an arbitrary Gaussian stochastic process such that $Z^n \sim N(\mu_n, K_Z^{(n)})$. Then there exists a sequence of $(2^{n(C_{n,\text{FB}} - \epsilon)}, n)$ feedback codes with $P_e^{(n)} \to 0$, as $n \to \infty$, for $\epsilon > 0$. Conversely, for $\epsilon > 0$, any sequence of $(2^{n(C_{n,\text{FB}} + \epsilon)}, n)$ codes has $P_e^{(n)}$ bounded away from zero for all n. The same statement holds in the special case without feedback upon substitution of C_n for $C_{n,\text{FB}}$.

We prove this theorem in Section VI.

In an earlier work, Butman [3] showed that feedback increases capacity for the first-order autoregressive channel. Tiernan and Schalkwijk [4] provided upper bounds to the capacity of band-limited first-order Gaussian autoregressive channels with feedback under an average energy constraint. Their development is based on "path energy increments" and does not require linear processing. In a subsequent paper [5], they analyzed the optimum linear system for an autoregressive forward channel with feedback. Finally, Butman [6] achieved tighter bounds on the capacity of general mth-order Gaussian autoregressive channels with linear feedback.

II. NECESSARY MATRIX INEQUALITIES

Let $|A|$ denote the determinant of A. To upper-bound $|K_{X+Z}|$ in the capacity formulas we need a number of matrix inequalities. It should come as a pleasant surprise that they all have information theoretic proofs. First we require the following.

Lemma 1: If A and B are nonnegative definite symmetric matrices, then

$$|A + B| \ge |A|. \tag{17}$$

Proof: Let $X \sim N_n(0, A), Y \sim N_n(0, B)$ be independent Gaussian n-vectors. Then

$$\frac{1}{2} \ln (2\pi e)^n |A + B| = h(X + Y) \ge h(X + Y|Y) = h(X)$$

$$= \frac{1}{2} \ln (2\pi e)^n |A|. \tag{18}$$

Lemma 2:

$$K_{X+Z} + K_{X-Z} = 2K_X + 2K_Z. \tag{19}$$

Proof: We have

$$K_{X+Z} = K_{XX} + K_{XZ} + K_{ZX} + K_{ZZ}$$

$$K_{X-Z} = K_{XX} - K_{XZ} - K_{ZX} + K_{ZZ}. \tag{20}$$

Summing the two equations yields the result.

The next inequality, which we get by combining Lemmas 1 and 2, is crucial in proving $C_{FB} \leq C + \frac{1}{2}$.

Lemma 3:

$$|K_{X+Z}| \leq |2K_X + 2K_Z| = 2^n|K_X + K_Z|. \qquad (21)$$

Finally, to prove $C_{FB} \leq 2C$, we need the (known) inequality given in Lemma 4 (see [7]). The proof is new. A number of determinant inequalities are developed in [12].

Lemma 4: For A, B nonnegative definite matrices and $0 \leq \lambda \leq 1$,

$$|\lambda A + (1 - \lambda)B| \geq |A|^\lambda |B|^{1-\lambda}. \qquad (22)$$

Proof: Let $X \sim N_n(0, A)$, and $Y \sim N_n(0, B)$. Let Z_θ be the mixture random vector

$$Z_\theta = \begin{cases} X, & \text{if } \theta = 1 \\ Y, & \text{if } \theta = 2, \end{cases} \qquad (23)$$

and let

$$\theta = \begin{cases} 1, & \text{with probability } \lambda \\ 2, & \text{with probability } 1 - \lambda. \end{cases}$$

Let X, Y, θ be independent. Then

$$K_Z = \lambda A + (1 - \lambda)B. \qquad (24)$$

We observe that

$$\frac{1}{2}\ln(2\pi e)^n|\lambda A + (1 - \lambda)B| \geq h(Z_\theta)$$
$$\geq h(Z_\theta|\theta)$$
$$= \lambda h(X) + (1 - \lambda)h(Y)$$
$$= \frac{1}{2}\ln(2\pi e)^n|A|^\lambda|B|^{1-\lambda}, \qquad (25)$$

which proves the result. The first inequality follows from the entropy maximizing property of the Gaussian distribution under a covariance constraint.

Definition: We say that the random vector X^n is *causally related* to Z^n if $f(x^n, z^n) = f(z^n)\prod_{i=1}^n f(x_i|x^{i-1}, z^{i-1})$. Note that feedback codes necessarily yield causally related (X^n, Z^n).

Lemma 5: If X^n and Z^n are causally related, then

$$h(X - Z) \geq h(Z) \qquad (26)$$

and

$$|K_{X-Z}| \geq |K_Z|. \qquad (27)$$

Proof: We have

$$h(X - Z) \overset{(a)}{=} \sum_{i=1}^n h\left(X_i - Z_i|X^{i-1} - Z^{i-1}\right)$$
$$\overset{(b)}{\geq} \sum h\left(X_i - Z_i|X^{i-1}, Z^{i-1}, X_i\right)$$
$$\overset{(c)}{=} \sum h\left(Z_i|X^{i-1}, Z^{i-1}, X_i\right)$$
$$\overset{(d)}{=} \sum h\left(Z_i|Z^{i-1}\right)$$
$$\overset{(e)}{=} h(Z). \qquad (28)$$

Here (a) is the chain rule, (b) is conditioning $h(A|B) \geq h(A|B, C)$, (c) follows from the conditional determinism of X_i and the invariance of differential entropy under translation, (d) follows from the causal relationship of X^n and Z^n, and (e) is the reverse chain rule.

Finally, suppose X^n and Z^n are causally related and the associated covariance matrices for X and $X - Z$ are K_Z and K_{X-Z}. There obviously exists a multivariate normal pair of (causally related) random vectors $\tilde{X}^n, \tilde{Z}^n$ with the same covariance structure. Thus from (28), we have

$$\frac{1}{2}\ln(2\pi e)^n|K_{X-Z}| = h(\tilde{X}^n - \tilde{Z}^n)$$
$$\geq h(\tilde{Z}^n)$$
$$= \frac{1}{2}\ln(2\pi e)^n|K_Z|,$$

thus proving (27).

III. Feedback Increases Capacity by at Most Half a Bit

We now show $C_{FB} \leq C + \frac{1}{2}$. Let

$$C_{n,FB} = \max \frac{1}{2n}\log \frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|} \qquad (29)$$

where the maximum is taken under the constraints in (12), (13). Let

$$C_n = \max_{\text{tr}(K_X^{(n)}) \leq nP} \frac{1}{2n}\log \frac{|K_X^{(n)} + K_Z^{(n)}|}{|K_Z^{(n)}|}. \qquad (30)$$

Although we shall be proving that both $C_{n,FB}$ and C_n are achievable communication rates, we do not need to show achievability at this stage.

Previous work by Gallager [8] guarantees that $C = \lim_{n \to \infty} C_n$ exists if $\{Z_i\}$ is a stationary Gaussian stochastic process and, furthermore, that C is the capacity for such stationary channels. Thus proving $C_{n,FB} \leq C_n + \frac{1}{2}$ for all n guarantees that

$$C_{FB} \leq \overline{\lim}_{n \to \infty} C_{n,FB} \leq \lim_{n \to \infty} C_n + \frac{1}{2} = C + \frac{1}{2}. \qquad (31)$$

We are now ready to prove that feedback adds to the capacity at most $\frac{1}{2}$ a bit per transmission.

Theorem 2: $C_{n,FB} \leq C_n + \frac{1}{2}$.

Proof: Let the $n \times n$ covariance matrix K_{X+Z} achieve feedback capacity $C_{n,FB}$ in (29). Then

$$nC_{n,FB} = \frac{1}{2}\log \frac{|K_{X+Z}|}{|K_Z|}$$
$$\leq \frac{1}{2}\log \frac{|2K_X + 2K_Z|}{|K_Z|}$$
$$= \frac{1}{2}\log \frac{2^n|K_X + K_Z|}{|K_Z|}$$
$$= \frac{1}{2}\log \frac{|K_X + K_Z|}{|K_Z|} + \frac{n}{2}$$
$$\leq n\left(C_n + \frac{1}{2}\right). \qquad (32)$$

Here, the first inequality, which relates feedback to non-feedback, follows from Lemma 3. Thus $C_{n,\mathrm{FB}} \le C_n + \frac{1}{2}$ for all n.

IV. FEEDBACK AT MOST DOUBLES CAPACITY

We now prove the celebrated result $C_{\mathrm{FB}} \le 2C$ of Pinsker [1] and Ebert [2].

Theorem 3: $C_{n,\mathrm{FB}} \le 2C_n$.

Proof: It is enough to show that

$$\frac{1}{2}\frac{1}{2n}\log \frac{|K_{X+Z}|}{|K_Z|} \le \frac{1}{2n}\log\frac{|K_X+K_Z|}{|K_Z|}, \qquad (33)$$

for it then follows, by maximizing each side in turn, that

$$\frac{1}{2}C_{n,\mathrm{FB}} \le C_n. \qquad (34)$$

We have

$$\frac{1}{2n}\log\frac{|K_X+K_Z|}{|K_Z|} \overset{(a)}{=} \frac{1}{2n}\log\frac{|\frac{1}{2}K_{X+Z}+\frac{1}{2}K_{X-Z}|}{|K_Z|}$$

$$\overset{(b)}{\ge} \frac{1}{2n}\log\frac{|K_{X+Z}|^{1/2}|K_{X-Z}|^{1/2}}{|K_Z|}$$

$$\overset{(c)}{\ge} \frac{1}{2n}\log\frac{|K_{X+Z}|^{1/2}|K_Z|^{1/2}}{|K_Z|}$$

$$= \frac{1}{2}\frac{1}{2n}\log\frac{|K_{X+Z}|}{|K_Z|} \qquad (35)$$

and the result is proved. Here (a) follows from Lemma 2, (b) from the inequality in Lemma 4 and (c) from Lemma 5 through the use of causality.

V. AEP FOR NONERGODIC GAUSSIAN PROCESSES

Gaussian stochastic processes apparently are special in the sense that they have the asymptotic equipartition property without the assumption of ergodicity or stationarity. We show, for $\{X_i\}$ jointly Gaussian,

$$-\frac{1}{n}\log f(X_1, X_2, \cdots, X_n) - \frac{h(X_1, \cdots, X_n)}{n} \to 0, \qquad (36)$$

with probability one. A similar result, without the distribution-free rate of convergence implicit in (45) and (47), is proved in Pinsker [9].

Let $X_1, X_2, \cdots$ be a time discrete Gaussian stochastic process. Let K_n denote the covariance matrix of $(X_1, X_2, \cdots, X_n)$. Let

$$h_n = \frac{h(X_1, X_2, \cdots, X_n)}{n} \qquad (37)$$

denote the (differential) entropy of $(X_1, X_2, \cdots, X_n)$ per unit time. Recall that if $X \sim f(X)$, the differential entropy is given (in nats) by

$$h(X_1, X_2, \cdots, X_n) = h(f) = -\int f(x)\ln f(x)\, dx. \qquad (38)$$

If $\lim_{n\to\infty} h_n = h$ exists, we say $\{X_i\}_{i=1}^{\infty}$ has entropy rate h. In particular, $\{X_i\}$ stationary implies the existence of an entropy rate. Since in this discussion, $X_1, X_2, \cdots, X_n$ are jointly normal, we know

$$X \sim N(\mu_n, K_n) \qquad (39)$$

and can calculate directly

$$h_n = \frac{1}{2n}\ln(2\pi e)^n |K_n|. \qquad (40)$$

Finally, recall the AEP as proved in full generality by Barron [11].

Theorem 4: If $\{X_i\}$ is stationary and ergodic with entropy rate h, then

$$-\frac{1}{n}\ln f(X_1, \cdots, X_n) \to h \qquad (41)$$

with probability one.

In a similar way to the AEP, we would like to show

$$-\frac{1}{n}\ln f(X_1, \cdots, X_n) \to h \qquad (42)$$

for arbitrary Gaussian processes. Without any further assumptions, h need not exist, so we wish to show the stronger result

$$-\frac{1}{n}\ln f(X_1, \cdots, X_n) - h_n \to 0. \qquad (43)$$

with probability one. Clearly, (43) implies (41) when $\lim h_n$ exists.

Note that Gaussian processes need not be ergodic. For example, consider $Z_0, Z_1, Z_2, \cdots$ i.i.d. $\sim N(0,1)$. Let $X_i = Z_0 + Z_i$, $i = 1, 2, \cdots$. Then $\{X_i\}$ is stationary but not ergodic. In particular $(1/n)\sum_{i=1}^{n} X_i \to Z_0$, a.e., which is a random variable. However, the AEP still holds. This is not the case for other apparently similar non-Gaussian constructions. For example, let $Z_i \sim \mathrm{Bern}(p_1)$ with probability λ and $Z_i \sim \mathrm{Bern}(p_2)$ with probability $1-\lambda$. Then $-(1/n)\log p(Z_1, \cdots, Z_n)$ does not converge to the entropy rate H.

Theorem 5: If $\{X_i\}$ is an arbitrary Gaussian stochastic process, then

$$-\frac{1}{n}\ln f(X_1, \cdots, X_n) - h_n \to 0 \qquad (44)$$

with probability one.

Proof: If $|K_n| = 0$, the result is trivially true since $h_n = -\infty$, and f is singular. Without loss of generality, let $\mu = 0$. We now assume $|K_n| > 0$, for all n. Then

$$-\frac{1}{n}\ln f(X_1, X_2, \cdots, X_n)$$

$$= -\frac{1}{n}\ln \frac{1}{(2\pi)^{n/2}|K_n|^{1/2}} e^{-X^t K_n^{-1} X/2}$$

$$= \frac{1}{2}\ln 2\pi + \frac{1}{2n}\ln|K_n| + \frac{1}{n}\frac{X^t K_n^{-1} X}{2}$$

$$= h_n - \frac{1}{2} + \frac{1}{n}\frac{X^t K_n^{-1} X}{2} \qquad (45)$$

where the last equality follows from (40).

However, it is well-known (Kendall and Stuart [10]) that, for $|K_n| > 0$, $X^t K_n^{-1} X$ has a chi-squared distribution with n degrees of freedom. That is, the distribution of $X^t K_n^{-1} X$ is the same as the distribution of $\sum_{i=1}^n Z_i^2$, Z_i i.i.d. $\sim N(0, 1)$.

For $\sum_{i=1}^n Z_i^2$ chi-squared with n degrees of freedom, the Chernoff bound yields

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^n Z_i^2 > (1 + \epsilon)\right\} \le e^{-(n/2)(\epsilon - \ln(1 + \epsilon))}. \quad (46)$$

Thus

$$\Pr\left\{\left|-\frac{1}{n}\ln f(X_1, X_2, \cdots, X_n) - h_n\right| > \epsilon\right\}$$
$$< e^{-n(\epsilon - (1/2)\ln(1 + 2\epsilon))} \quad (47)$$

a bound that does not depend on K_n.

Finally, by (47) and the Borel–Cantelli lemma,

$$-\frac{1}{n}\ln f(X_1, \cdots, X_n) - h_n \to 0 \quad (48)$$

with probability one.

VI. Converse for Theorem 1

We now show that $(2^{n(C_{n,\mathrm{FB}} + \epsilon)}, n)$ feedback codes have probability of error $P_e^{(n)}$ bounded away from zero. The same proof works in the special case without feedback upon substitution of C_n for $C_{n,\mathrm{FB}}$.

Consider a sequence of $(2^{nR}, n)$ feedback codes $(x_i(W, Y^{i-1}), g(\cdot))$, $W \in \{1, 2, \cdots, 2^{nR}\}$, $Y^i = (Y_1, Y_2, \cdots, Y_i)$, $g: R^n \to \{1, 2, \cdots, 2^{nR}\}$. Let $\lambda_i = P\{g(Y^n) \ne W | W = i\}$, and $P_e^{(n)} = (1/2^{nR})\sum_{i=1}^{2^{nR}}\lambda_i$. The joint distribution of (W, X, Y) is given by

$$W \sim \mathrm{unif}\{1, 2, \cdots, 2^{nR}\}$$
$$X = \left(x_1(W), x_2(W, Y^1), \cdots, x_n(W, Y^{n-1})\right)$$
$$Y = X + Z$$
$$Z \sim N(0, K_n)$$
$$W \text{ and } Z \text{ independent.} \quad (49)$$

We wish to show that a sequence of $(2^{nR_n}, n)$ codes with $P_e^{(n)} \to 0$ must have $nR_n < h(Y^n) - h(Z^n) + n\epsilon_n \le n(C_{n,\mathrm{FB}} + \epsilon_n)$ where $\epsilon_n \to 0$.

Proof: By Fano's inequality

$$nR_n = H(W)$$
$$= H(W | Y^n) + I(W; Y^n)$$
$$= I(W; Y^n) + n\epsilon_n \quad (50)$$

where $\epsilon_n \to 0$ if $P_e^{(n)} \to 0$. Now

$$I(W; Y^n) = h(Y^n) - h(Y^n | W) \quad (51)$$

and

$$h(Y^n | W) \stackrel{(a)}{=} \sum_{i=1}^n h(Y_i | W, Y^{i-1})$$
$$\stackrel{(b)}{=} \sum h(X_i + Z_i | W, Y^{i-1}, X_i(W, Y^{i-1}), Z^{i-1})$$
$$\stackrel{(c)}{=} \sum h(Z_i | W, Y^{i-1}, X_i, Z^{i-1})$$
$$\stackrel{(d)}{=} \sum h(Z_i | Z^{i-1})$$
$$\stackrel{(e)}{=} h(Z^n). \quad (52)$$

Here (a) follows from the chain rule, (b) merely adds functions of the conditions, (c) removes the conditionally deterministic constant X_i, (d) uses the conditional independence of (W, Y^{i-1}, X_i) and Z_i given Z^{i-1}, and (e) "unchains" the chain rule.

Thus, as shown by Tiernan and Schalkwijk [4],

$$I(W; Y^n) \le h(Y^n) - h(Z^n). \quad (53)$$

Finally, by the entropy maximizing property of the normal distribution, we have

$$h(Y^n) - h(Z^n) \le \frac{1}{2}\ln\frac{|K_{X+Z}^{(n)}|}{|K_Z^{(n)}|}$$
$$\le nC_{n,\mathrm{FB}}, \quad (54)$$

as given in (12), (13). The converse is thus proved.

VII. Achievability of $C_{n,\mathrm{FB}}$

Before we proceed to prove Theorem 1, we need the following definition of jointly ϵ-typical sequences.

Definition: Let (V^n, Y^n) be jointly distributed with density $f(v^n, y^n)$ and associated entropy rates as defined in (37),

$$h_n(V) = \frac{1}{n}h(V^n)$$
$$h_n(Y) = \frac{1}{n}h(Y^n)$$
$$h_n(V, Y) = \frac{1}{n}h(V^n, Y^n). \quad (55)$$

Then the set A_ϵ^n of jointly ϵ-typical (V^n, Y^n) is defined by

$$A_\epsilon^n = \Bigg\{(V^n, Y^n) \in R^n \times R^n:$$
$$\left|-\frac{1}{n}\ln f(V^n) - h_n(V)\right| < \epsilon$$
$$\left|-\frac{1}{n}\ln f(Y^n) - h_n(Y)\right| \le \epsilon$$
$$\left|-\frac{1}{n}\ln f(V^n, Y^n) - h_n(V, Y)\right| \le \epsilon\Bigg\}. \quad (56)$$

Let $V(A_\epsilon^n)$ denote the volume of A_ϵ^n. We have the following bound on the volume of the typical set.

Lemma 6:

$$V(A_\epsilon^n) \leq 2^{n(h_n(V, Y) + \epsilon)}. \tag{57}$$

Proof: By the definition of A_ϵ^n,

$$1 = \int f(v^n, y^n) \, dv^n \, dy^n$$

$$\geq \int_{A_\epsilon^n} f(v^n, y^n) \, dv^n \, dy^n$$

$$\geq \int_{A_\epsilon^n} 2^{-n(h_n(V, Y) + \epsilon)} \, dv^n \, dy^n$$

$$= V(A_\epsilon^n) 2^{-n(h_n(V, Y) + \epsilon)}, \tag{58}$$

which proves the lemma.

To show the achievability of $C_{n, \text{FB}}$, let $C_{n, \text{FB}}$ be as defined by (10) and (11), where (11) can be written as

$$X^n = BZ^n + V^n, \tag{59}$$

and where B is a strictly lower triangular matrix, V^n and Z^n are independent, and B and K_V achieve $C_{n, \text{FB}}$.

The proof uses random coding. Let $V(1), V(2), \cdots, V(2^{nR})$ be independent identically distributed n-vectors drawn according to $N_n(0, (1 - \delta)K_V + \delta P_V I)$, where $0 < \delta < 1$, I is the identity matrix, and $P_V = \text{tr}(K_V)$. Now by Lemma 1 in Section II, $(1 - \delta)K_V + \delta P_V I$ is nonsingular and the AEP will apply. Note that the expected power constraint on X is satisfied, i.e.,

$$E \frac{1}{n} \|X(W, Z)\|^2 = \frac{1}{n} \text{tr}(K_X) \leq P \tag{60}$$

since $(1/n) \text{tr} K_V = (1/n) \text{tr}((1 - \delta)K_V + \delta P_V I)$.

Transmission: To send W, the transmitter transmits $X(W, Z) = BZ + V(W)$.

Decoding: The receiver Y declares $\hat{W} \in 2^{nR}$ was sent if $(V(\hat{W}), Y)$ is the only ϵ-typical pair.

Error: An error is made if there is no typical $(V(\hat{W}), Y)$ pair, more than one such, or $\hat{W} \neq W$.

To analyze the probability of error $P_e^{(n)}$, assume without loss of generality that $W = 1$ was sent. Thus $Y = X + Z = BZ + V(1) + Z$. Define the events

$$E_i = (V(i), Y) \in A_\epsilon^n, \qquad i = 1, 2, \cdots, 2^{nR} \tag{61}$$

and E_i^c, the complement of E_i. Then

$$P_e^{(n)} \leq \Pr\{E_1^c | W = 1\} + 2^{nR} \Pr\{E_2 | W = 1\}. \tag{62}$$

By the AEP, (47), and the joint normality of $V(1)$ and Y,

$$\Pr\{E_1^c | W = 1\} \leq 3 e^{-n(\epsilon - (1/2)\ln(1 + 2\epsilon))}, \qquad \text{for all } n. \tag{63}$$

Moreover,

$$\Pr\{E_2 | W = 1\} = \int_{(v, y) \in A_\epsilon^n} f(v) f(y) \, dv \, dy$$

$$\leq \int 2^{-n(h_n(V) - \epsilon)} 2^{-n(h_n(Y) - \epsilon)} \, dv \, dy,$$

by the ϵ-typicality of $(V(1), Y)$,

$$\leq 2^{n(h_n(V, Y) - h_n(V) - h_n(Y) + 3\epsilon)},$$

by the volume bound on A_ϵ^n,

$$= 2^{-n(h_n(Y) - h_n(Y|V) - 3\epsilon)}$$

$$= 2^{-n(h_n(Y) - h_n(BZ + Z) - 3\epsilon)}$$

$$= 2^{-n(h_n(Y) - h_n(Z) - 3\epsilon)}, \text{ since } |B + I| = |I| = 1,$$

$$= 2^{-n((1/2n)\log\{|K_B + (1 - \delta)K_V + \delta P_V I| / |K_Z|\} - 3\epsilon)}, \tag{64}$$

where $K_B = (I + B)K_Z(I + B)^t$. By continuity of the determinant as a function of δ, we have

$$|K_B + (1 - \delta)K_V + \delta P_V I| \to |K_B + K_V| = |K_{X + Z}|, \text{ as } \delta \to 0. \tag{65}$$

Thus for δ sufficiently small,

$$\Pr\{E_2 | W = 1\} \leq 2^{-n(C_{n, \text{FB}} - 4\epsilon)}. \tag{66}$$

Combining, we have

$$P_e^n \leq \Pr\{E_1^c | W = 1\} + 2^{nR} \Pr\{E_2 | W = 1\}$$

$$\leq 3 e^{-n(\epsilon - (1/2)\ln(1 + 2\epsilon))} + 2^{n(R - C_{n, \text{FB}} + 4\epsilon)}. \tag{67}$$

Thus there exists a sequence of $(2^{n(C_{n, \text{FB}} - 5\epsilon)}, n)$ codes with $P_e^{(n)} \to 0$, as $n \to \infty$. Since $\epsilon > 0$ is arbitrary, the theorem is proved.

VIII. REMARKS ON POWER CONSTRAINT

Throughout this paper, we found the capacity under an expected power constraint

$$\frac{1}{n} \text{tr}(K_X) = E \frac{1}{n} \sum_{i=1}^{n} X_i^2(W, Y^{i-1}) \leq P. \tag{68}$$

From a stricter point of view, we should declare an error whenever the feedback causes X to use power greater than P. Thus we require the stronger condition

$$\Pr\left\{\frac{1}{n} \sum_{i=1}^{n} X_i^2(W, Y^{i-1}) > P\right\} \leq \epsilon. \tag{69}$$

A simple sufficiency condition for satisfying the power constraint (69) while achieving capacity would be

$$\frac{1}{n^2} \text{tr}\left(\left(K_X^{(n)}\right)^2\right) \to 0, \text{ as } n \to \infty \tag{70}$$

where $K_X^{(n)}$ achieves capacity in (10). This guarantees that

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2(W, Y^{i-1}) \to P. \tag{71}$$

Stationarity of $\{Z_i\}$ certainly suffices.

IX. Concluding Remarks

Distribution-free error bounds were found for achieving rates $C_{n, \text{FB}} - \epsilon$, for each n. Thus no asymptotic statement as $n \to \infty$ is required to impart significance to capacity. In general, feedback does not increase the capacity of a Gaussian channel by more than $\frac{1}{2}$ bit/transmission.

References

[1] M. Pinsker, talk delivered at the Soviet Information Theory Meeting, 1969 (no abstract published).
[2] P. Ebert, "The capacity of the Gaussian channel with feedback," *Bell Syst. Tech. J.*, pp. 1705–1712, Oct. 1970.
[3] S. A. Butman, "A general formulation of linear feedback communication systems with solutions," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 3, pp. 392–400, May 1969.
[4] J. C. Tiernan and J. P. M. Schalkwijk, "An upper bound to the capacity of the band limited Gaussian autoregression channel with noiseless feedback," *IEEE Trans. Inform. Theory*, IT-20, no. 3, pp. 311–316, May 1974.
[5] J. C. Tiernan, "Analysis of the optimum linear system for the autoregressive forward channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 3, pp. 359–363, May 1976.
[6] S. A. Butman, "Linear feedback rate bounds for regressive channels," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 3, pp. 363–366, May 1976.
[7] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic, 1979.
[8] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[9] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden Day, 1964.
[10] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vols. I and II. New York: MacMillan, 1977.
[11] A. R. Barron, "The strong ergodic theorem for densities: Generalized Shannon McMillan Breiman theorem," *Ann. Prob.*, vol. 13, no. 4, pp. 1292–1303, 1985.
[12] T. M. Cover and J. A. Thomas, "Determinant inequalities via information theory," *Matrix Anal. Appl. (SIAM J.)*, vol. 9, no. 3, pp. 384–392, July 1988.