# On the Capacity of the Noisy Runlength Channel

Chris D. Heegard, *Member, IEEE*, Alexandra Duel-Hallen, *Member, IEEE*, and Rajeev Krishnamoorthy

*Abstract* —**The theory and design of codes that limit the minimum and maximum runlengths in a modulation code are well developed. Until recently, most information theoretic models for these codes were noiseless. In this paper we consider a new, noisy model for runlength modulation codes. Through a series of lemmas and theorems, bounds on the capacity of this model are obtained. These bounds are evaluated and used to suggest that improvements in storage capacity are possible through the use of codes with designed noise tolerance.**

## I. Introduction

THE THEORY and design of codes that limit the minimum and maximum runlengths in a modulation code are well developed (e.g., [2]–[4], [12]–[16]). Until recently, most information theoretic models for these codes were noiseless. One notable exception is the example of a noisy channel model is the *binary symmetric channel* with a "$d,k$" constraint, as considered by Zehavi and Wolf [3], and Shamai and Kofman [17]. In this paper we consider a new, noisy model for runlength modulation codes. Through a series of lemmas and theorems, bounds on the capacity of this model are obtained. These bounds are evaluated and used to suggest that improvements in storage capacity are possible through the use of codes with designed noise tolerance.

In many communications and storage systems, binary waveforms are used to transmit or store digital information. The most common examples of such systems are magnetic and optical disk technology, magnetic tape and fiber optic channels. In all of these systems, information is encoded, via a runlength limiting modulation code, into a binary signal, $x(t) \in \{-1, +1\}$. In some real sense, the fact that the signal can take only two possible levels means that the information is essentially conveyed in the locations of the transitions (i.e., the times when the signal goes from plus to minus or vice versa) of the encoded waveform, $x(t)$. The fact that the signal is runlength-limited simply means that the intervals of time when the signal is constant (i.e., the "runlengths") are uniformly bounded from above and below. (The bounds are called the runlength constraints of the code.)

Many detectors are designed to estimate the locations of the transitions in the received waveform $y(t)$, a distorted and noisy version of the transmitted signal $x(t)$. For example, in magnetic recording, the channel produces a (signed) pulse at the output of the channel whenever a transition occurs in the input $x(t)$. A device, called a "peak detector," is used to determine that a transition has occurred and to find the location of the pulse. A typical peak detector has two components: 1) a threshold device that declares the presence of a transition (i.e., a "qualifier" circuit) and 2) a pulse locating device that estimates exactly where the transition has taken place. Analogous detection can be found in optical recording; in these systems, transitions in the input waveform translate into level changes at the output. In optical systems, the peak detector is replaced by an "edge detector," a device that declares the presence of an edge and estimates the location of the edge. In all of these recorders, once a transition is successfully qualified, an estimate of the transition location is computed.

In this paper, we consider a model for runlength coded systems that is based on peak or edge detection [1]. The model does not attempt to account for failures of the qualifying circuit; it is assumed that every transition is successfully declared (i.e., no missing or false qualifiers). In this model, it is the error in the estimate of the location of the transition that introduces uncertainty at the channel output. The model for this uncertainty is additive Gaussian noise; if a transition in $x(t)$ occurs at time $t_n$ and the detection estimate is $\tau_n$ then the error $t_n - \tau_n$ is assumed to be normally distributed with zero-mean and variance $\sigma^2$ and is independent of $t_n$.

The model is motivated by the problem of pulse location in white Gaussian noise. If a signal, $x(t) = p(t - t_0)$, with known shape, $p(t)$, and unknown location, $-T \le t_0 \le T$, is observed in additive white Gaussian noise, $y(t) = x(t) + w(t)$ ($Ew(t)w(s) = N_0\delta(t-s)$), then the maximum

Manuscript received June 30, 1990; revised February 15, 1991. This work was supported in part by IBM, AT&T, and NSF Grants ECS83-52220, MIP87-11793, and NCR89-03931. This work was presented in part at the 25th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 30–October 2, 1987 and the 1988 IEEE International Workshop on Information Theory, Bejing, China, July 1988.

C. D. Heegard and R. Krishnamoorthy are with the School of Electrical Engineering, 327 Engineering Theory Center, Cornell University, Ithaca, NY 14853.

A. Duel-Hallen is with AT&T Bell Laboratories, 600 Mountain Avenue, Room 2C-180, Murray Hill, NJ 07974.

IEEE Log Number 9144530.

0018-9448/91/0500-0712$01.00 

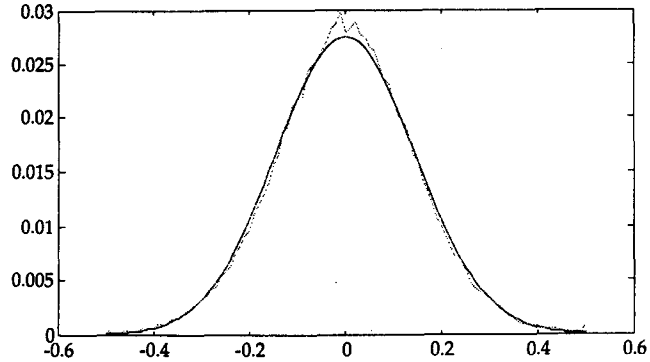


Fig. 1. Peak detection of Lorentz pulse in white Gaussian noise.

likelihood (ML) estimate of the location $\tau_0$ satisfies

$$\int_{-\infty}^{\infty} p(t-\tau_0)y(t)\,dt = \sup_{-T\le\tau\le T}\int_{-\infty}^{\infty} p(t-\tau)y(t)\,dt$$

(i.e., the ML estimate is one that maximizes the correlation). It is known, that for high signal to noise ratios (SNR) (i.e., $\|p\|^2/N_0$), the error in the estimate $t_0-\tau_0$ is approximately Gaussian with zero-mean and positive variance.

For example, the results of a computer experiment is displayed in Fig. 1. In this experiment, a Lorentz pulse, $p(t)=1/(1+t^2)$, is observed in white Gaussian noise with an SNR of 20 db. The distribution of the error of the correlation detector is estimated and compared with a Gaussian distribution. It is apparent from the figure that the Normal curve gives a good approximation to the distribution of the observed error.

The correlation can be computed by passing the observed signal through a filter with impulse response, $h(t) = p(-t)$,

$$s(\tau_0) = \sup_{-T\le\tau\le T} s(\tau), \qquad s(\tau)=\int_{-\infty}^{\infty} h(\tau-t)y(t)\,dt.$$

Such a filter, $h(t)=p(-t)$, is called a *matched filter*, since the response is "matched" to the known pulse shape. In practice, several prudent factors mitigate the use of the ideal matched filter. The first follows from the fact that the estimated transitions $t_n$ are not generated by isolated pulses; on the contrary, the pulses are very dense in the observed signal $y(t)$. Since the typical pulse width in recording is very long, the response of the matched filter would suffer from extreme *intersymbol interference* (ISI); the effects of "nearby" pulses on the estimate of $t_n$ would have a serious, detrimental effect on the estimate $\tau_n$. Thus, as a practical matter, a correlating filter $h(t)$, that makes a compromise between signal to noise ratio (i.e., "matched filter loss") and ISI is used (e.g., a "pulse slimming" filter). Furthermore, in practice, the pulse shape $p(t)$ is not exactly known (e.g., disk-to-disk variation) or can vary considerably (e.g., track-to-track variations due to constant angular velocity of the disk). Thus, again in reality, a compromise choice of response $h(t)$ is used to make the detector robust to variations in the pulse, $p(t)$. However, the fact that the error in the estimate, $t_n-\tau_n$, has a Gaussian distribution depends only on the correlation structure (not whether $h(t)$ is ideally matched to $p(t)$). The approximation will be a good one provided: 1) the signal to noise ratio at the output of the correlator is large and 2) the filter exhibits minimal ISI.

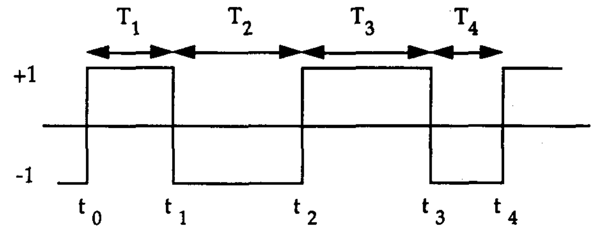


Fig. 2. Parameters of binary waveform.

Finally, the model considered in this paper assumes that a constraint is imposed on the runlengths of the signal, $T_{\min}\le T_n = t_n - t_{n-1}\le T_{\max}$. We allow the runlengths, $T_n$, to assume a continuum of possibilities in the interval from $T_{\min}$ to $T_{\max}$. Of course, in practice, this is not the case; a *minimum spacing* parameter $\Delta$ and *coding parameters* $d$ and $k$, are selected so that the runlengths take on a uniformly spaced set of values $T_n = i*\Delta$, $d+1 \le i \le k+1$ and $T_{\min}=(d+1)\Delta$, $T_{\max}=(k+1)\Delta$. It is easy to see that by shrinking the spacing parameter $\Delta$ (and increasing $d$ and $k$ appropriately), the runlength constraints, $T_{\min}$ and $T_{\max}$, can be maintained and the number of possible runlengths, $((T_{\max}-T_{\min})/\Delta+1)$ can grow. We can think of our problem as the limiting case, when spacing parameter goes to zero $\Delta\to 0$. Furthermore, the practical consideration that the runlengths are discrete imposes no real limitation, if one is allowed to vary $\Delta$. In fact, the choice of spacing parameter, $\Delta$, should depend on the capacity of the channel (i.e., the noise). For low SNR, a large $\Delta$ is acceptable, since the noise limits the capacity. On the other hand, for higher SNR, $\Delta$ must be made smaller to approximate the capacity of the channel or else the capacity will be limited by an insufficiently small signal set and not the noise! It is our feeling that, in practice, one would find that a suitable choice of $\Delta$ would not be unreasonably small.

## II. Noisy Runlength Model

In recording systems, a binary waveform, $x(t)\in\{-1,+1\}$, is used to represent stored information. Such a waveform can be described in terms of the instances of transition (i.e., when $x(t)$ goes from plus to minus or minus to plus) (Fig. 2). Define $t_0=0$ and the *nth transition time* by

$$t_n = \inf\{t > t_{n-1} | x(t)\ne x(t_{n-1})\}.$$

Then the *nth runlength* is equal to the difference

$$T_n = t_n - t_{n-1}.$$

When information is read in a recording system, estimates $\tau_n$ of the transition times $t_n$ are made and used to recover the information. Because of the bandlimited na-

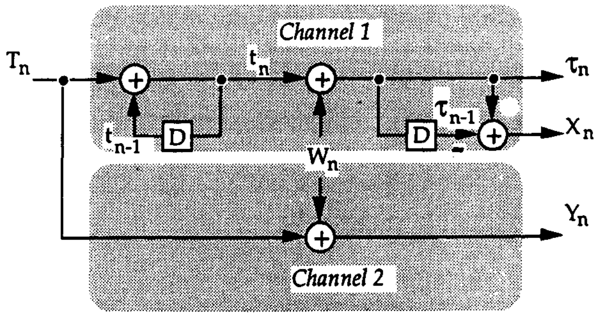


Fig. 3. Noisy runlength channel model.

ture of the recording channel, it is difficult to obtain reliable estimates for the transitions if the runlengths are small. For this reason, a constraint is placed on the minimum value, $T_{\min} > 0$, that a runlength can assume for an allowed waveform $x(t)$. A similar constraint is placed on the maximum runlength $T_{\max}$ for reasons of synchronization. Thus the channel imposes a constraint

$$T_{\min} \leq T_n \leq T_{\max}$$

on the runlengths. This suggests the runlength-limited channel model.

The runlength-limited channel accepts runlengths as inputs; the $n$th input is a bounded positive number, $T_{\min} \leq T_n \leq T_{\max}$. These inputs represent lengths of time intervals between channel events (i.e., transitions). The values of the runlength limits, $T_{\min}$ and $T_{\max}$, represent given constraints. The "true" or "absolute" time is obtained by summing the runlengths

$$t_n = \sum_{i=1}^{n} T_i = T_n + t_{n-1},$$

where $t_0 = 0$. The channel output is an estimate of the true time $\tau_n = t_n + W_n$, where $W_n$ is the estimation error. By assumption $W_n$ is an independent identically distributed (i.i.d.) additive Gaussian noise process with mean-zero and variance $\sigma^2$ (Fig. 3).

A closely related channel is obtained by differencing the noisy transition time $\tau_n$. This channel has output $X_n = \tau_n - \tau_{n-1} = T_n + Z_n$ where the noise $Z_n = W_n - W_{n-1}$ is a correlated (non-white) Gaussian sequence ($E(Z_n^2) = 2\sigma^2, E(Z_n Z_{n-1}) = -\sigma^2$). Both of these channels have the same capacity $C_1(T_{\min}, T_{\max}, \sigma^2)$. This capacity

$$C_1(T_{\min}, T_{\max}, \sigma^2) = \overline{\lim_{n \to \infty}} \max_{\mathscr{P}_n} \frac{I(T_1, \cdots, T_n; \tau_1, \cdots, \tau_n)}{E(t_n)}$$
$$= \overline{\lim_{n \to \infty}} \max_{\mathscr{P}_n} \frac{I(T_1, \cdots, T_n; X_1, \cdots, X_n)}{E(t_n)},$$

where $\mathscr{P}_n$ is the set of input distributions on the $n$-cube $[T_{\min}, T_{\max}]^n$.

Due to the integration on the channel input (in the $\tau$-channel case or equivalently the nonwhite noise in the $X$-channel case) the optimum input distribution on blocks of length $n$ is not an i.i.d. distribution. For this reason, a related channel, the channel with output $Y_n = T_n + W_n$ (i.e., without the integrator), is studied. The characterization of the second channel is easier since the optimum input distribution for this channel is i.i.d. In this case, the capacity

$$C_2(T_{\min}, T_{\max}, \sigma^2) = \max_{\mathscr{P}_1} \frac{I(T;Y)}{E(T)},$$

where $Y = T + W$, $T_{\min} \leq T \leq T_{\max}$, $W$ is zero mean Gaussian with variance $\sigma^2$ and $T$ and $W$ are independent.

In Section III, we obtain bounds on the capacity $C_1(T_{\min}, T_{\max}, \sigma^2)$ in the following order. First, three lemmas are introduced that are useful in dealing with this channel. Next, upper and lower bounds on the capacity of the noisy runlength channel, $C_1(T_{\min}, T_{\max}, \sigma^2)$, are derived in terms of the capacity of the second channel $C_2(T_{\min}, T_{\max}, \sigma^2)$ (Theorem 1). Theorems 2 and 3 provide two lower bounds to $C_2(T_{\min}, T_{\max}, \sigma^2)$. Theorem 4 describes an upper bound on $C_2(T_{\min}, T_{\max}, \sigma^2)$. These bounds on $C_2(T_{\min}, T_{\max}, \sigma^2)$, when combined with Theorem 1 give upper and lower bounds on $C_1(T_{\min}, T_{\max}, \sigma^2)$. Finally, an upper bound on $C_1(T_{\min}, T_{\max}, \sigma^2)$ itself is given in Theorem 5.

Section IV deals with computing the capacity $C_2(T_{\min}, T_{\max}, \sigma^2)$. The theorems and lemmas of Sections III and IV are used to find bounds on $C_1(T_{\min}, T_{\max}, \sigma^2)$ numerically as described in Section V.

Throughout the paper, logarithms with base 2 are used.

## III. Bounds on the Capacity

The following three lemmas are useful tools for determining bounds on the capacity of the runlength limited channel.

The first lemma follows from Lemma 2 of [5]. This lemma answers the question of which runlength random variable $T$ maximizes the ratio of the differential entropy $h(T)$ to the expected value $E(T)$.

*Lemma 1:* Let $T$ be a bounded random variable, $T_{\min} \leq T \leq T_{\max}$. For $\gamma > 0$, the bound

$$\frac{h(T) + \log(\gamma)}{E(T)} \leq \log(\lambda)$$

holds with equality if and only if the density of $T$, $p_T(t) = \gamma\lambda^{-t}$(a.e.), where

$$\int_{T_{\min}}^{T_{\max}} \gamma\lambda^{-t}\, dt = 1.$$

Note that this last equation uniquely determines $\lambda > 0$.

*Proof of Lemma 1:* More generally, let a random variable $T$ be restricted to a subset of real numbers $\mathscr{T}$, and $f(t)$ be a positive function. We show that

$$\frac{h(T) + \log(\gamma)}{E(f(T))} \leq \log(\lambda), \qquad \text{for } p_T(t) = \gamma\lambda^{-f(t)}\text{(a.e.)},$$

where

$$\int_{\mathscr{T}} \gamma\lambda^{-f(t)}\,dt = 1.$$

The proof consists of two simple steps.

First, let $T^*$ be the random variable with the density $q(t) = \gamma\lambda^{-f(t)}$. Then by definition

$$\begin{aligned} h(T^*) &\equiv -\int_{\mathscr{T}} q(t)\log(q(t))\,dt \\ &= \int_{\mathscr{T}} q(t)(f(t)\log(\lambda) - \log(\gamma))\,dt \\ &= Ef(T^*)\log(\lambda) - \log(\gamma). \end{aligned} \qquad (1)$$

Now, the *discrimination* (or *Kullback–Liebler number*) for probability densities is defined by [6]

$$D(p\|q) \equiv \int_{\mathscr{T}} p(t)\log\left(\frac{p(t)}{q(t)}\right)dt.$$

It is easy to show (and well known) that this quantity is nonnegative and equal to zero if and only if $p(t) = q(t)$ (a.e.). Thus,

$$\begin{aligned} 0 \le D(p\|q) &= -h(T) - \int_{\mathscr{T}} p(t)\log(\gamma\lambda^{-f(t)})\,dt \\ &= -h(T) + Ef(T)\log(\lambda) - \log(\gamma) \end{aligned}$$

or

$$h(T) + \log(\gamma) \le Ef(T)\log(\lambda).$$

Since $f(t)$ is a positive function, both $Ef(T)$ and $Ef(T^*)$ are positive and

$$\frac{h(T)+\log(\gamma)}{Ef(T)} \le \log(\lambda) = \frac{h(T^*)+\log(\gamma)}{Ef(T^*)}.$$

The last equality follows from (1). Note that equality holds only when $p(t) = q(t) = \gamma\lambda^{-f(t)}$(a.e.). □

*Lemma 2:* Let $T$ be a bounded random variable, $T_{\min} \le T \le T_{\max}$ and let $W$ be a Gaussian random variable with variance $\sigma^2$. Assume that $T$ and $W$ are independent. For $\gamma > 0$

$$\frac{h(T+W)+\log(\gamma)}{ET} \ge \frac{\log((2\pi e\sigma^2 + 2^{2h(T)})\gamma^2)}{2E(T)}.$$

*Proof of Lemma 2:* This result follows from the entropy power inequality for independent random variables [6]–[8]

$$2^{2h(T+W)} \ge 2^{2h(T)} + 2^{2h(W)}.$$

Taking the log of both sides and using the formula for the differential entropy of a Gaussian random variable, $h(W) = 1/2\log(2\pi e\sigma^2)$, shows

$$h(T+W) \ge \frac{1}{2}\log(2\pi e\sigma^2 + 2^{2h(T)}).$$

The lemma immediately follows from this. □

*Lemma 3:* Let $T$ be a bounded random variable, $T_{\min} \le T \le T_{\max}$, and let $W$ be a Gaussian random variable with variance $\sigma^2$. Assume that $T$ and $W$ are independent. For $\gamma > 0$ and $\delta > 0$,

$$\begin{aligned} \frac{h(T+W)+\log(\gamma)}{ET} &\le f(\sigma,\delta,\gamma) \\ &\equiv \log(\lambda_\delta)\left(1 + \frac{\sqrt{\frac{2}{\pi}}\,\sigma}{T_{\min}}\right) \\ &\quad + \frac{\overline{h_b(\xi)} - 1/2\,\overline{\xi\log\xi} + \xi/2\max(0,\log(2\pi e\sigma^2\gamma^2))}{T_{\min}}, \end{aligned}$$

where

$$\int_{T_{\min}-\delta}^{T_{\max}+\delta} \gamma\lambda_\delta^{-t}\,dt = 1,$$

and

$$\xi \equiv Q\left(\frac{\delta}{\sigma}\right) + Q\left(\frac{T_{\max} - T_{\min} + \delta}{\sigma}\right),$$

where the "$Q$" function

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}\,dy.$$

The function

$$\overline{h_b(\xi)} \equiv \begin{cases} h_b(\xi), & \text{if } \xi \le 1/2, \\ 1, & \text{otherwise,} \end{cases}$$

where the binary entropy function

$$h_b(\xi) \equiv -\xi\log(\xi) - (1-\xi)\log(1-\xi).$$

Finally,

$$\overline{\xi\log\xi} \equiv \begin{cases} \xi\log\xi, & \text{if } \xi \le 1/e, \\ -1/e\log e, & \text{otherwise.} \end{cases}$$

*Proof of Lemma 3:* Let $A$ be the random variable

$$A \equiv \begin{cases} 1, & T_{\min} - \delta \le T+W \le T_{\max} + \delta \\ 0, & \text{otherwise.} \end{cases}$$

Then $\Pr(A=0) = \epsilon$ and

$$\begin{aligned} h(T+W) &= I(A;T+W) + h(T+W|A) \\ &= h_b(\epsilon) + \epsilon h(T+W|A=0) \\ &\quad + (1-\epsilon)h(T+W|A=1). \end{aligned}$$

Let $Z$ be the random variable

$$Z \equiv \begin{cases} T+W-T_{\max}-\delta, & T_{\max}+\delta < T+W, \\ 0, & T_{\min}-\delta \le T+W \le T_{\max}+\delta, \\ T+W-T_{\min}+\delta, & T+W < T_{\min}-\delta. \end{cases}$$

For $z > 0$, the conditional density of $Z$ given $A = 0$

$$\begin{aligned} p_{Z|A}(z|0) &= \frac{1}{\epsilon}\int_{T_{\min}}^{T_{\max}} p_T(t)p_W(z+T_{\max}+\delta-t)\,dt \\ &= \frac{1}{\epsilon}p_W(z+T_{\max}+\delta-t_0) < \frac{1}{\epsilon}p_W(z), \end{aligned}$$

where $p_T$ and $p_W$ are the densities of $T$ and $W$, and $T_{\min} \le t_0 \le T_{\max}$. Since for negative values of $z$, the conditional density $p_{Z|A}(z|0)$ is also bounded above by

$1/\epsilon p_W(z)$, it follows that $E(Z^2|A=0) \le \sigma^2/\epsilon$. Thus,

$$h(T+W|A=0) = h(Z|A=0) \le \frac{1}{2}\log\left(\frac{2\pi e\sigma^2}{\epsilon}\right).$$

From Lemma 1,

$$\frac{h(T+W|A=1)+\log(\gamma)}{E(T+W|A=1)} \le \log(\lambda_\delta).$$

Also

$$E(T+W|A=1) = \frac{E(A(T+W))}{1-\epsilon} \le \frac{E(T)+E|W|}{1-\epsilon} = \frac{E(T)+\sigma\sqrt{\frac{2}{\pi}}}{1-\epsilon}.$$

Thus

$$h(T+W|A=1)+\log(\gamma) \le \log(\lambda_\delta)\frac{E(T)+\sigma\sqrt{\frac{2}{\pi}}}{1-\epsilon}.$$

Combining these inequalities yields

$$\frac{h(T+W)+\log(\gamma)}{ET} \le \log(\lambda_\delta) + \frac{h_b(\epsilon)+\frac{\epsilon}{2}\log\left(\frac{2\pi e\sigma^2}{\epsilon}\gamma^2\right)+\sigma\sqrt{\frac{2}{\pi}}\log(\lambda_\delta)}{E(T)}.$$

By using the fact that the function $Q(c-x)+Q(c+x)$ is increasing in $x$ on the interval $0 \le x \le c$, it can be shown that $\epsilon \le \xi$. Thus, the right-hand side of this inequality is bounded above by $f(\sigma,\delta,\gamma)$, where $f(\cdot)$ and $\xi$ are defined in the statement of this lemma. The last bound follows from the fact that $E(T) \ge T_{\min}$, and all terms of $f(\cdot)$ are nondecreasing in $\xi$. □

*Theorem 1:*

$$C_2(T_{\min},T_{\max},\sigma^2) \le C_1(T_{\min},T_{\max},\sigma^2) \le C_2(T_{\min},T_{\max},2\sigma^2)+\frac{1}{2T_{\min}}.$$

*Proof of Theorem 1:* The first inequality can be seen by comparing the first channel with output $\tau_j$ and the second channel with output $Y_j$ (see Fig. 3). Let the input distribution on the runlengths $T_1,\cdots,T_n$ be i.i.d. (this is the form of the optimum distribution for the second channel with output $Y$). Then

$$\begin{aligned} h(\tau_1,\cdots,\tau_n) &= \sum_{i=1}^n h(\tau_i|\tau_1,\cdots,\tau_{i-1}) \\ &\ge \sum_{i=1}^n h(\tau_i|\tau_1,\cdots,\tau_{i-1},t_{i-1}) \\ &= \sum_{i=1}^n h(Y_i) = h(Y_1,\cdots,Y_n), \end{aligned}$$

where the last equality follows from the i.i.d. assumption on the inputs and the white noise $W$.

The second inequality follows by comparing the first channel with output $X_j$ and the second channel with output $Y_j$ (see Fig. 3). First, for $Z_i = W_i - W_{i-1}$,

$$\begin{aligned} h(X_1,\cdots,X_n|T_1,\cdots,T_n) &= h(Z_1,\cdots,Z_n) \\ &= \sum_{i=1}^n h(Z_i|Z_1,\cdots,Z_{i-1}) \\ &\ge \sum_{i=1}^n h(Z_i|Z_1,\cdots,Z_{i-1},W_{i-1}) \\ &= \sum_{i=1}^n h(W_i) = h(W_1,\cdots,W_n) \end{aligned}$$

and

$$\begin{aligned} h(X_1,\cdots,X_n) &= \sum_{i=1}^n h(X_i|X_1,\cdots,X_{i-1}) \\ &\le \sum_{i=1}^n h(X_i) \\ &= \sum_{i=1}^n h\left(T_i+\sqrt{2}W_i\right). \end{aligned}$$

Thus

$$\begin{aligned} \frac{I(T_1,\cdots,T_n;X_1,\cdots,X_n)}{E(t_n)} &= \frac{h(X_1,\cdots,X_n)-h(X_1,\cdots,X_n|T_1,\cdots,T_n)}{E(t_n)} \\ &\le \frac{\sum_{i=1}^n \left(h\left(T_i+\sqrt{2}W_i\right)-h\left(\sqrt{2}W_i\right)+\frac{1}{2}\log(2)\right)}{E(t_n)} \\ &\le C_2\left(T_{\min},T_{\max},2\sigma^2\right)+\frac{1}{2T_{\min}}. \end{aligned}$$ □

To obtain lower bounds to $C_2(T_{\min} T_{\max},\sigma^2)$, write

$$\frac{I(T;Y)}{E(T)} = \frac{h(T+W)-\frac{1}{2}\log(2\pi e\sigma^2)}{E(T)}.$$

Then note that a lower bound to the capacity is obtained from Lemma 2 when $\gamma = 1/\sqrt{2\pi e\sigma^2}$

$$\frac{I(T;Y)}{E(T)} \ge \frac{\log\left(1+\frac{2^{2h(T)}}{2\pi e\sigma^2}\right)}{2E(T)}.$$

Then any choice of distribution for $T$ will provide a lower bound to $C_2(T_{\min},T_{\max},\sigma^2)$.

*Theorem 2:*

$$C_2(T_{\min},T_{\max},\sigma^2) \ge \frac{\log\left(1+\frac{(T_{\max}-T_{\min})^2}{2\pi e\sigma^2}\right)}{T_{\max}+T_{\min}}.$$

*Proof of Theorem 2:* Let $T$ have a uniform distribution on the interval $[T_{\min}, T_{\max}]$. Then

$$h(T) = \log(T_{\max} - T_{\min}), \qquad E(T) = \frac{T_{\max} + T_{\min}}{2}. \quad \square$$

*Theorem 3:*

$$C_2(T_{\min}, T_{\max}, \sigma^2) \geq \log(\lambda) \frac{\log(1 + 2^{2\theta})}{2\theta} > \log(\lambda),$$

where

$$\gamma = \frac{1}{\sqrt{2\pi e \sigma^2}}, \qquad \int_{T_{\min}}^{T_{\max}} \gamma \lambda^{-t}\, dt = 1,$$

$$\theta = \left(1 + \gamma\left(T_{\min} \lambda^{-T_{\min}} - T_{\max} \lambda^{-T_{\max}}\right)\right) \log e.$$

*Proof of Theorem 3:* Note that

$$\frac{\log\left(1 + \dfrac{2^{2h(T)}}{2\pi e \sigma^2}\right)}{2E(T)} > \frac{h(T) - \frac{1}{2}\log(2\pi e \sigma^2)}{E(T)}.$$

Let $T$ have the distribution, $\gamma\lambda^{-t}$; this maximizes this lower bound (c.f., Lemma 1). Then

$$h(T) = \theta - \log(\gamma), \qquad E(T) = \frac{\theta}{\log(\lambda)}. \quad \square$$

To state Theorems 4 and 5, we use definitions introduced in Lemma 3.

*Theorem 4:* For $\delta > 0$,

$$C_2(T_{\min}, T_{\max}, \sigma^2) \leq f\left(\sigma, \delta, \frac{1}{\sqrt{2\pi e \sigma^2}}\right).$$

*Proof of Theorem 4:* Apply Lemma 3 with $\gamma = 1/\sqrt{2\pi e \sigma^2}$. $\square$

*Theorem 5:* For $\delta > 0$

$$C_1(T_{\min}, T_{\max}, \sigma^2) \leq f\left(\sqrt{2}\sigma, \delta, \frac{1}{\sqrt{2\pi e \sigma^2}}\right).$$

*Proof of Theorem 5:* In the proof of the upper bound of Theorem 1, it is shown that

$$\frac{I(T_1, \cdots, T_n; X_1, \cdots, X_n)}{E(t_n)} \leq \frac{\sum_{i=1}^{n} h(T_i + \sqrt{2}W_i) - h(W_i)}{E(t_n)}$$

This inequality, when combined with Lemma 3 ($\gamma = 1/\sqrt{2\pi e \sigma^2}$) gives the result.

## IV. Computing the Capacity, $C_2(T_{\min}, T_{\max}, \sigma^2)$

This section deals with computing $C_2(T_{\min}, T_{\max}, \sigma^2)$, the capacity of the second channel. The capacity of an additive scalar Gaussian channel with amplitude (or amplitude and variance) constraint is achieved [10] by using a discrete input distribution. Following the results in [10], it was shown in [11] that the distribution that achieves capacity for the channel $C_2(T_{\min}, T_{\max}, \sigma^2)$ is also discrete. Furthermore, using the algorithm outlined in [11] it is possible to compute the capacity (and the capacity achieving distribution) for the channel $C_2(T_{\min}, T_{\max}, \sigma^2)$.

We consider a scalar additive Gaussian channel characterized by $Y = T + W$, where $T$, $Y$, and $W$ are the input, output, and noise random variables respectively. The input r.v. $T$ is constrained to take on values on the compact interval $K \triangleq [T_{\min}, T_{\max}]$. Let $\mathscr{F}_K$ be the class of distributions on the interval $[T_{\min}, T_{\max}]$. The random variable $W$ is taken to be Gaussian, $W \sim N(0, \sigma^2)$, and so

$$p_{Y|T}(y|t) = p_W(w = y - t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-t)^2/2\sigma^2}.$$

and for all $F \in \mathscr{F}_k$, we may write the marginal distribution,

$$p_Y(y) = \int_{T_{\min}}^{T_{\max}} p_W(y - t)\, dF(t).$$

Since for the fixed channel, the mutual information between the input and the output random variables depends only on the probability distribution of the input, we explicitly write $I(T;Y) = I(F)$. The capacity of the channel is

$$C_2(T_{\min}, T_{\max}, \sigma^2) \triangleq \sup_{F \in \mathscr{F}_K} \frac{I(F)}{E(F)},$$

where

$$E(F) \triangleq \int_{T_{\min}}^{T_{\max}} t\, dF(t)$$

is the mean of the input distribution $F$.

The marginal information density is defined as

$$i(t;F) \triangleq \int_{-\infty}^{\infty} p_{Y|T}(y|t) \log\left(\frac{p_{Y|T}(y|t)}{p_Y(y)}\right) dy, \qquad T_{\min} \leq t \leq T_{\max}$$

from which the mutual information $I(F)$ may be written as

$$I(F) = \int_{T_{\min}}^{T_{\max}} i(t;F)\, dF(t).$$

For any channel constraint $K = [T_{\min}, T_{\max}]$, noise power $\sigma^2$, and $\psi \geq 0$, denote

$$L(K, \sigma^2, \psi) \triangleq \sup_{F \in \mathscr{F}_k} (I(F) - \psi E(F)).$$

The set of distributions, $\mathscr{F}_k$ is convex and compact in the Lévy metric topology, the function $E(F)$ is linear, and the function $I(F)$ is strictly convex-cap, continuous, and weakly differentiable in $\mathscr{F}_k$. It can be shown [11] that

*Lemma 4:* The value $L(K, \sigma^2, \psi)$ is achieved by a unique input distribution function $F_\psi \in \mathscr{F}_k$; i.e.,

$$L(K, \sigma^2, \psi) = \max_{F \in \mathscr{F}_k} (I(F) - \psi E(F)) = I(F_\psi) - \psi E(F_\psi).$$

For any $F \in \mathscr{F}_k$ we define a *point of increase* $t$ to be such that $F(t) > F(t - \delta)$ for all $\delta > 0$. Let $G$ denote the set of the points of increase of $F$. Lemmas 5 and 6 state the necessary and sufficient conditions for $F$ to be optimal, and the discrete nature of the optimal distribution.

*Lemma 5:* Let $F$ be an arbitrary probability distribution in $\mathscr{F}_k$. Let $G$ denote the points of increase of $F$ on $[T_{\min}, T_{\max}]$. Then for a given parameter $\psi$, $F$ is "optimal," i.e., $F$ maximizes $L(K, \sigma^2, \psi)$, if and only if it satisfies the Kuhn–Tucker conditions [6],

$$i(t;F) - \psi t \le I(F) - \psi E(F), \qquad T_{\min} \le t \le T_{\max},$$
$$i(t;F) - \psi t = I(F) - \psi E(F), \qquad t \in G.$$

*Lemma 6:* The points of increase of an optimal $F$ in $[T_{\min}, T_{\max}]$, $G$ is a finite set.

Finally we show that the capacity $C_2(T_{\min}, T_{\max}, \sigma^2)$ can be achieved.

*Lemma 7:* The capacity, $C_2(T_{\min}, T_{\max}, \sigma^2)$, is achieved by a unique probability distribution function, $F^* \in \mathscr{F}_k$; i.e.,

$$C_2(T_{\min}, T_{\max}, \sigma^2) = \max_{F \in \mathscr{F}_k} \frac{I(F)}{E(F)} = \frac{I(F^*)}{E(F^*)},$$

for some $F^* \in \mathscr{F}_k$.

We use the function $L(K, \sigma^2, \psi)$ to find the constraint $\psi^*$ and the corresponding distribution $F_{\psi^*}$ that achieves $C_2(T_{\min}, T_{\max}, \sigma^2)$ in the following manner. Associated with every distribution $F \in \mathscr{F}_k$ are the following variables: the cardinality of the set of the points of increase of $F$, $|G| = n$; the locations of the points of increase of $F$ on $[T_{\min}, T_{\max}]$, $L$; and the respective probability mass functions associated with each member of the set $L, P$. If we fix $n$, then the following lemma outlines the iterative scheme to obtain $C_2(T_{\min}, T_{\max}, \sigma^2)$.

*Lemma 8:* Given a peak constrain, $K = [T_{\min}, T_{\max}]$, noise variance $\sigma^2$ and the number of inputs, $n$, let $\mathscr{F}_K^n$ denote the class of discrete distributions restricted to the interval $[T_{\min}, T_{\max}]$ having $n$ atoms. If, for any $\psi \ge 0$, $F \in \mathscr{F}_K^n$ is such that

$$\max_{F_\psi \in \mathscr{F}_K^n} [I(F) - \psi E(F)] = I(F_\psi) - \psi E(F_\psi),$$

then the set of iterations,

$$\psi_{j+1} = \frac{I(F_{\psi_j})}{E(F_{\psi_j})}$$

converges to $\psi^*$, where $\psi^*$ satisfies

$$\max_{\psi \ge 0} \frac{I(F_\psi)}{E(F_\psi)} = \frac{I(F_{\psi^*})}{E(F_{\psi^*})}.$$

The maximization algorithm (to find $F_{\psi_k}$) alternates between a variant of the Arimoto–Blahut algorithm to compute $P$, given $n$ and $L$, and a set of iterations to compute $L$ given $n$ and $P$. The optimality of the putative optimal distribution is checked by seeing if it satisfies the Kuhn–Tucker conditions, and increasing the number of points by one every time the optimality conditions fail. If the distribution satisfies the conditions of Lemma 5, then $F_{\psi^*}$ is the capacity achieving distribution $F^*$.

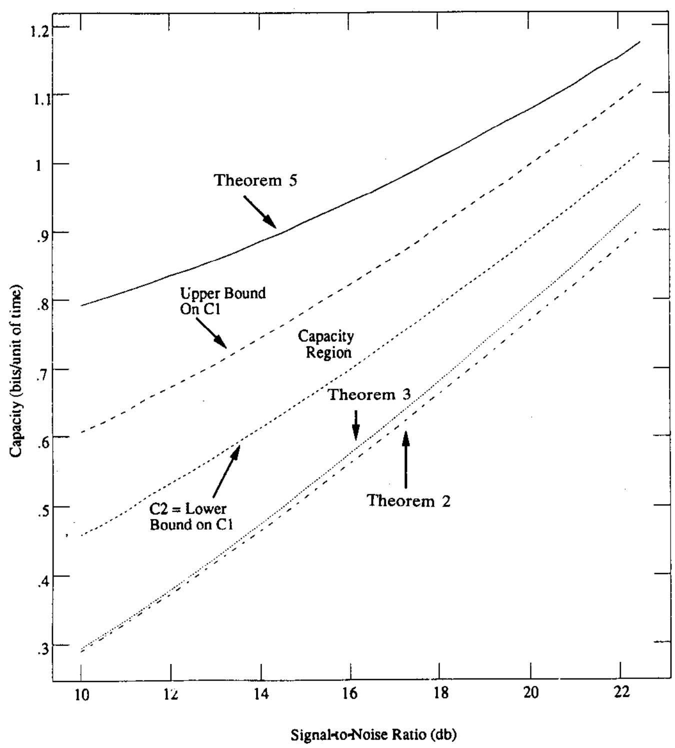


Fig. 4. Bounds on capacity of $T_{\min} = 2$, $T_{\max} = 4$, noisy runlength channel.

## V. Numerical Results

We computed bounds on $C_1(T_{\min}, T_{\max}, \sigma^2)$ numerically for several runlength parameters (Figs. 4–7). The bounds are plotted versus noise variance (SNR $= 1/\sigma^2$ db). By Theorem 1, $C_2(T_{\min}, T_{\max}, \sigma^2) \le C_1(T_{\min}, T_{\max}, \sigma^2)$. Thus, we used lower bounds on $C_2(T_{\min}, T_{\max}, \sigma^2)$ given by Theorems 2 and 3 to obtain lower bounds on $C_1(T_{\min}, T_{\max}, \sigma^2)$. We found that the bound of Theorem 2 was consistently worse than the bound of Theorem 3. However, we show both bounds since Theorem 2 utilizes uniform distribution on runlengths, and its bound is easier to compute. The difference between these lower bounds grows as the length of the interval $T_{\max} - T_{\min}$ increases. It is most pronounced for the ($T_{\min} = 3$, $T_{\max} = 100$) channel (Fig. 7).

Two different upper bounds on $C_1(T_{\min}, T_{\max}, \sigma^2)$ can be obtained from Theorems 1, 4 and 5. The first bound is found by bounding $C_2(T_{\min}, T_{\max}, \sigma^2)$ (Theorem 4) and using the second inequality of Theorem 1. The second bound is given by Theorem 5. Note that upper bounds of Theorems 4 and 5 depend on the parameter $\delta$. By varying $\delta$, we found the "best" upper bounds for each theorem and each set of parameters. For channels of consideration, the first upper bound (Theorems 1 and 4) was worse than the second upper bound (Theorem 5). Thus, we show only the second bound in the figures. We observe that lower and upper bounds depend much stronger on $T_{\min}$ than on $T_{\max}$. For example, for $T_{\min} = 3$, the bounds of Theorems 3 and 5 stay nearly the same when $T_{\max}$ is changed from 8 to 100 (Figs. 6, 7).

Following the approach of Section IV, we also computed the capacity of the second channel $C_2(T_{\min}, T_{\max}, \sigma^2)$

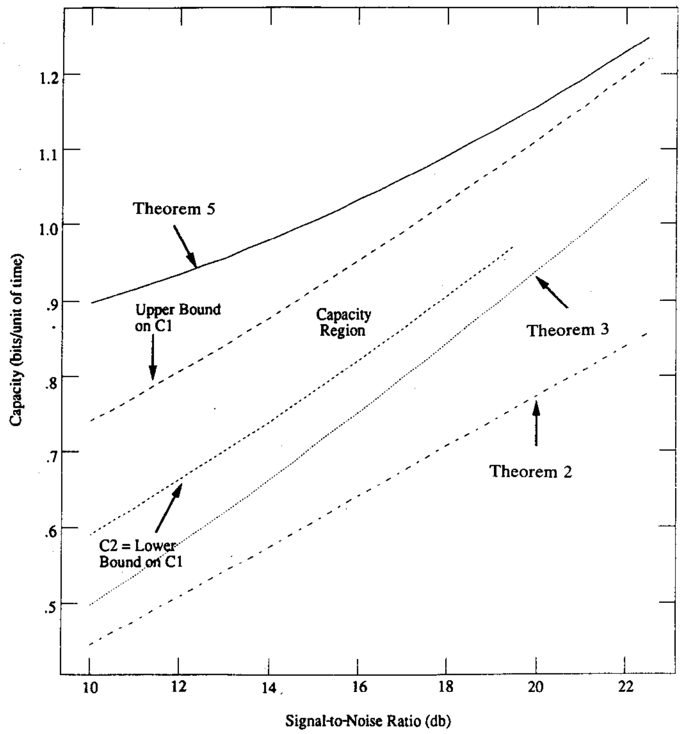


Fig. 5. Bounds on capacity of $T_{min} = 2$, $T_{max} = 8$, noisy runlength channel.

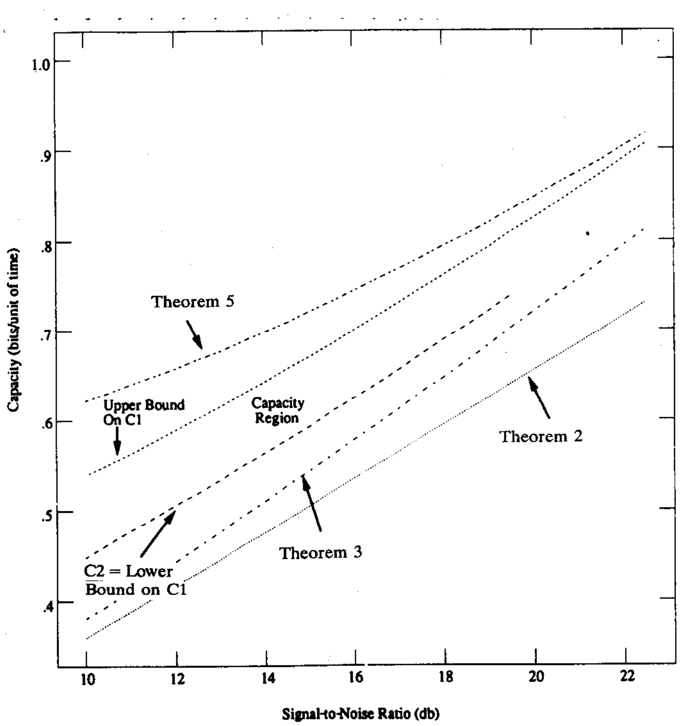


Fig. 6. Bounds on capacity of $T_{min} = 3$, $T_{max} = 8$, noisy runlength channel.

for ($T_{min} = 2$, $T_{max} = 4$). We used this capacity and Theorem 1 to find a lower and an upper bound on $C_1(T_{min}, T_{max}, \sigma^2)$. The resulting bounds ($C_2$ = lower bound on $C_1$ and upper bound on $C_1$) are better than the bounds obtained from theorems of Section III, but their computation is more complex.

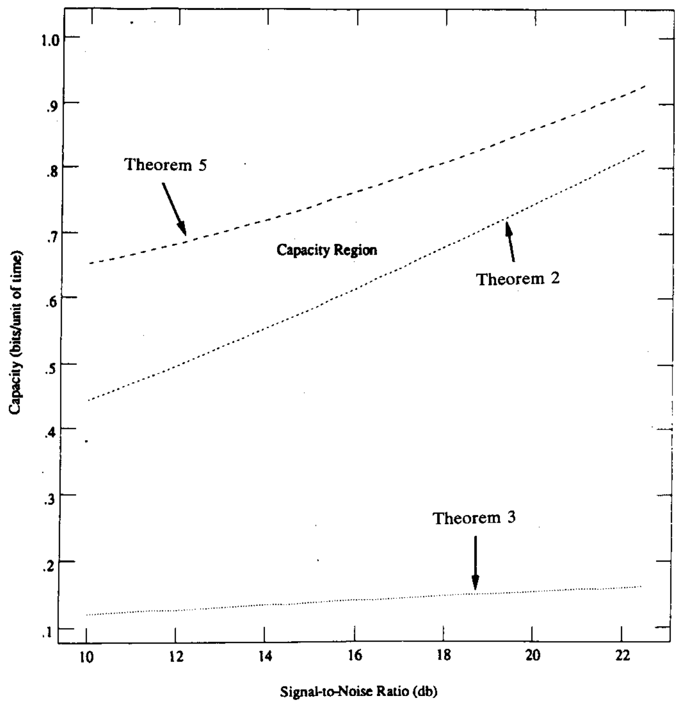


Fig. 7. Bounds on capacity of $T_{min} = 3$, $T_{max} = 100$, noisy runlength channel.

Note that the runlength parameters chosen in Figs. 4–6 correspond to $(d, k)$ constraints of (1,3), (1,7), and (2,7) recording codes, where $d$ and $k$ represent the minimum and the maximum number of consecutive zero (0) symbols between one (1) symbol. Thus, in the corresponding waveform, transitions occur at least $(d+1)\Delta$ and at most $(k+1)\Delta$ time units apart, where $\Delta$ is the symbol interval. (We use $\Delta = 1$.) For example, MFM is a ($d = 1$, $k = 3$) binary code of rate 1/2 [12]. We observe that our lower bound on the capacity $C_1$ of the corresponding ($T_{min} = 2$, $T_{max} = 4$) channel significantly exceeds 1/2 for high signal-to-noise ratios, i.e., in the region where the probability of "peak-shift" is small (Fig. 4). (Note that the SNR = 22 db corresponds to the "peak-shift" probability $2Q(\Delta/2\sigma) = 3.2 \times 10^{-10}$.) Popular rate 2/3 (1,7) codes are the Jacoby code [14] and the AHM (IBM) code [15], [16]. For high signal-to-noise ratios, their rate is much lower than our lower bound on $C_1(2, 8, \sigma^2)$ (Fig. 5). Similar comparisons can be made for rate 1/2 IBM [1], [2], [16] and Zerox [13] (2,7) codes (Fig. 6).

## Acknowledgment

C. D. Heegard would like to thank Paul Siegel and Razmik Karabed for useful discussions on the topic of this paper. The authors would like to thank Venkatachalam Anantharam for help with Lemma 3 and Joseph Abaya for computing some of the bounds.

## References

[1] P. H. Siegel, "Applications of a peak detection channel model," *IEEE Trans. Magn.*, vol. MAG-18, no. 6, pp. 1250–1252, 1982.

[2] R. L. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 1, pp. 5–22, Jan. 1983.
[3] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Trans. Inform. Theory*, vol. 34, no. 1, pp. 45-54, Jan. 1988.
[4] A. Gallopoulos, C. Heegard and P. Siegel, "The power spectrum of runlength limited codes," *IEEE Trans. Commun.*, vol. COM-37, no. 9, Sept. 1989.
[5] C. D. Heegard, "A pair of information theoretic lemmas with application to runlength coding," 25th Ann. Allerton Conf. Commun., Contr. and Comput., Sept. 30–Oct. 2, 1987.
[6] R. Blahut, *Principles and Practice of Information Theory*. New York: Addison-Wesley, 1987.
[7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948; pp. 623–656, Oct. 1948.
[8] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. Contr.*, vol. 2, pp. 101–112, 1959.
[9] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, April 1965.
[10] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inform. Contr.*, vol. 18, pp. 203–219, 1971.
[11] R. Krishnamoorthy, and C. D. Heegard, "On computing the capacity of certain peak- and power-constrained channels," in preparation.
[12] M. Hecht and A. Guida, "Delay modulation," *Proc. IEEE*, vol. 57, no. 7, pp. 1314–1316, July 1969.
[13] K. Norris, *Xerox Disclosure J.*, vol. 5, no. 6, pp. 647–648, Nov./Dec. 1980.
[14] G. Jacoby, M. Cohn, and A. Bates, III, U.S. Patent 4,337,458, 1982.
[15] R. Adler, M. Hassner, and J. Moussouris, U.S. Patent 4,413,25 1, 1982.
[16] T. Howell, "Statistical properties of selected recording codes," IBM Res. Rep., Aug. 1987.
[17] S. Shamai (Shitz) and Y. Kofman, "On the capacity of binary and Gaussian channels with runlength-limited inputs," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 584–684, May 1990.