Transactions Letters

Progressive Source Coding for a Power Constrained Gaussian Channel

Marc P. C. Fossorier, *Senior Member, IEEE*, Zixiang Xiong, and Kenneth Zeger, *Fellow, IEEE*

***Abstract*—We consider the progressive transmission of a lossy source across a power constrained Gaussian channel using binary phase-shift keying modulation. Under the theoretical assumptions of infinite bandwidth, arbitrarily complex channel coding, and lossless transmission, we derive the optimal channel code rate and the optimal energy allocation per transmitted bit. Under the practical assumptions of a low complexity class of algebraic channel codes and progressive image coding, we numerically optimize the choice of channel code rate and the energy per bit allocation. This model provides an additional degree of freedom with respect to previously proposed schemes, and can achieve a higher performance for sources such as images. It also allows one to control bandwidth expansion or reduction.**

***Index Terms*—Image compression, source and channel coding.**

I. Introduction

SHANNON'S "separation principle" ensures that transmission of a source across a discrete memoryless channel can asymptotically achieve the minimum possible distortion by independently choosing the source and channel coders. However, this theoretical result assumes unbounded delay and computational complexity. In practice, finite delay and complexity constraints motivate the search for source and channel codes which efficiently trade off the available transmission rate between source coding and channel coding.

Under high resolution assumptions, an optimum tradeoff between fixed-delay source coding and block channel coding was derived in [1] for the binary symmetric channel (BSC) and assumes the use of error correcting codes achieving exponentially small probability of error in terms of block length. Such codes are known to exist [10] but it is not known how to efficiently find and use them.

The "cost" of using a discrete-amplitude channel is generally described in terms of the transmission rate, measured in channel uses per source symbol. In contrast, the "cost" of using a power constrained channel, such as the additive white Gaussian noise (AWGN) channel, is often described by the average energy transmitted per source symbol. For a given modulation signal constellation, the number of signal constellation points transmitted per source symbol is a design parameter (determined off-line) that can be chosen to optimize the end-to-end quantization error of the system. The number of constellation points sent per source symbol is referred to as the *transmission rate* (equivalently we could describe the source and channel transmission rates per unit time).

For a power constrained channel, if one chooses a higher transmission rate, then on average there is less energy transmitted per constellation point, and a larger probability of error for each received constellation point results. The benefit, however, of a higher transmission rate is that more bits are available to code each source symbol and to provide error control coding. A given transmission rate implies a particular number of bits transmitted per source symbol. These bits are the output generated by sending source coding bits through a channel coder. An important problem is how to effectively allocate these bits between coding the source and providing protection against channel errors. This allocation is characterized by the choice of a channel code rate. The channel code rate determines what fraction of the transmission rate is for source coding and what fraction is for channel coding. There is thus a tradeoff between modulation, source coding, and channel coding. In this letter, we examine these components by jointly optimizing the transmission rate and the channel code rate for certain classes of source and channel coders. This model provides an additional degree of freedom with respect to previously proposed schemes, and can achieve a higher performance for sources such as images.

We consider two cases which are based upon progressively transmitting a lossy source across a power constrained Gaussian channel using binary phase-shift keying (BPSK) modulation. First, we mathematically analyze a theoretical case where the assumptions include infinite bandwidth, arbitrarily complex channel coding, and lossless transmission. Secondly, we experimentally analyze a practical case where the assumptions include a low complexity class of algebraic channel codes and progressive image coding. In the first case we rigorously derive the optimal channel code rate and the optimal energy allocation per transmitted bit. In the second case we numerically optimize the choice of channel code rate and the energy per bit allocation.

In Section II, we establish in Proposition 1 the theoretically achievable minimum distortion by calculating the largest reliable source rate that can be transmitted using Shannon-optimal channel codes and unlimited computational complexity, delay,

Paper approved by E. Ayanoglu, the Editor for Communication Theory and Coding Application of the IEEE Communications Society. Manuscript received February 3, 2000; revised December 13, 2000. This work was supported in part by the National Science Foundation. This paper was presented in part at the International Conference on Image Processing (ICIP), Chicago, IL, October 1998.

M. P. C. Fossorier is with the Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA (e-mail: marc@spectra.eng.hawaii.edu).

Z. Xiong is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zx@ee.tamu.edu).

K. Zeger is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: zeger@ucsd.edu).

Publisher Item Identifier S 0090-6778(01)06926-4.

and memory. The proof of Proposition 1 relies on Theorem 1, which proves that the capacity of the channel considered is a concave function of the energy per transmitted bit. Theorem 1 is an interesting result in its own right and to the best of our knowledge, has not previously been published.

In Section III, motivated by potential applications of image coding over noisy channels, we consider systems satisfying the following general conditions: 1) the source coder is progressive; 2) the channel decoder can detect decoding errors with high probability; 3) source decoding terminates at the first detected channel decoding error; and 4) the channel is power constrained. Condition 1 means that an image can be incrementally decoded as more bits correctly arrive at the receiver. Condition 2 is a property of many concatenated channel codes and is easy to achieve in practical systems. It allows the receiver to know when it cannot correctly decode received bits (and thus declare an erasure). Condition 3 is a byproduct of certain efficient image coders, such as Said and Pearlman's "Set Partitioning In Hierarchical Trees" (SPIHT) scheme [4]. Condition 4 is a classical communication theory assumption [9, p. 288]. No bandwidth assumption is made. This applies to unrestricted bandwidth channel models such as those used in deep space communications systems and also to constrained bandwidth channel models.

As an example, we examine a generalization of the image coding scheme of [2] from a discrete memoryless channel to a Gaussian channel. For a given power constrained Gaussian channel, we consider the optimum allocation of energy per bit for a BPSK transmitter when the performance is measured by end-to-end average quantizer distortion. We give a numerical optimization procedure that assumes a practical source coder and channel code family are used. Our experimental results demonstrate quantitatively how much one can improve system performance by carefully selecting the transmitted energy per bit and the channel code rate as a function of the given power constraint. Furthermore, it is showed how this optimization procedure can be used in the tradeoff between performance and bandwidth requirements.

II. Source and Channel Coding With a Power Constraint

Suppose that the quality of a sampled and encoded analog source is characterized by a distortion function $\mathcal{D}(r_s)$ in terms of a source coding rate r_s (measured in bits per source sample). The quantity $\mathcal{D}(r_s)$, for example, typically measures for a particular image coder the mean-squared quantization error of a decompressed image in terms of the number of bits per pixel (bpp) present in the compressed version of the image. Suppose a channel code with rate r_c acts on the output bits from the source encoder. The resulting bit stream is transmitted across an AWGN channel, whose noise has zero-mean and variance $N_0/2$, using a BPSK modulator with decoding performed by a hard-limiter on the received sampled values. Suppose the BPSK modulator emits a sequence of constellation points satisfying a fixed power constraint P, measured in units of energy per source sample. (It is a "power" constraint because the number of source samples per unit time is constant, and thus the amount of energy per unit time is constrained.) Then if R constellation signals per source sample are transmitted over the channel, the average energy E_s per transmitted signal satisfies

$$R \cdot E_s = P.$$

This channel model is referred to as a *power constrained BSC*. The number of bits per source sample available for source coding is

$$r_s = R \cdot r_c$$

and the probability of error for a transmitted channel bit is

$$p(E_s) = Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \tag{1}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/2} du.$$

The crossover probability of the resulting discrete BSC is $p(E_s)$ and the capacity is

$$C(E_s) = 1 - h(p(E_s)) \tag{2}$$

where h is the binary entropy function, defined by

$$h(x) = -x \log_2(x) - (1-x)\log_2(1-x).$$

Shannon's channel coding theorem [10] shows that for a fixed $E_s > 0$, if $r_c < C(E_s)$ then on average r_s information bits per source sample can be transmitted with arbitrarily small probability of error, and Shannon's separation principle shows that the distortion $\mathcal{D}(r_s)$ can be achieved. Shannon's theorem guarantees the achievability of the theoretically maximum reliable transmission rate, although it assumes unboundedly long block lengths.

In [2], the following related problem for image transmission is considered.

> *Given a fixed transmission rate R_0 and channel code blocklength n for transmission over a BSC with crossover probability ϵ, what is the lowest possible distortion achievable at the receiver?*

If the BSC model in [2] is induced from hard-limited BPSK on a power-constrained AWGN channel, then the ratio E_s/N_0 is determined by the BSC crossover probability ϵ. For a fixed noise level N_0 on a given Gaussian channel, the energy E_s per BPSK transmitted channel signal is then determined by ϵ. We denote the resulting value of E_s by E_b which can be computed from

$$E_b = \frac{N_0}{2}(Q^{-1}(\epsilon))^2 \tag{3}$$

and the power constraint is implicitly defined by

$$R_0 \cdot E_b = P. \tag{4}$$

In contrast, in the present letter, we fix the power constraint P and allow E_s to vary, so that the transmission rate becomes $R = P/E_s = R_0 \cdot (E_b/E_s)$. Consequently, for each channel code considered, the energy per bit E_s is optimized instead of being taken as a given. The model of [2] corresponds to the case

$E_{\rm s} = E_{\rm b}$. The case of $E_{\rm s} < E_{\rm b}$ can viewed as a degradation of the channel to allow the use of a more powerful channel code, and conversely for $E_{\rm s} > E_{\rm b}$. The image coding scheme in [2] terminates source coding when uncorrectable errors are detected. With high probability, therefore, the distortion for the system in [2] is $\mathcal{D}(r_{\rm s})$ where $r_{\rm s}$ is the amount of rate that successfully is transmitted prior to the first detected channel decoding error. This distortion is minimized when the rate $r_{\rm s}$ is maximized. The following proposition characterizes the largest possible rate $r_{\rm s}$, where the maximization is taken over all energy per bit allocations $E_{\rm s}$. A related and less rigorous result is given in [5].

Proposition 1: The maximum achievable source coding rate with arbitrarily small probability of channel decoding error on a power constrained BSC is

$$r_{\rm s} = P \cdot \left(\frac{2}{\pi N_0 \ln 2}\right). \tag{5}$$

Proof: We seek

$$\begin{aligned}\sup_{E_{\rm s}>0,\ r_{\rm c}\in[0,1]} \{r_{\rm s}\} &= \sup_{E_{\rm s}>0,\ r_{\rm c}\in[0,1]} \{R r_{\rm c}\}\\ &= P \cdot \sup_{E_{\rm s}>0}\ \sup_{r_{\rm c}\in[0,1]} \left\{\frac{r_{\rm c}}{E_{\rm s}}\right\}\\ &= P \cdot \sup_{E_{\rm s}>0} \left\{\frac{C(E_{\rm s})}{E_{\rm s}}\right\}\end{aligned} \tag{6}$$

where the last equality follows from Shannon's channel coding theorem [10]. In the Appendix, it is shown that $C(E_{\rm s})$ is a concave function of $E_{\rm s}$, so that for all $\alpha \in (0,1)$

$$C(\alpha E_{\rm s}) > \alpha C(E_{\rm s}) \tag{7}$$

or equivalently $C(E_{\rm s})/E_{\rm s} < C(\alpha E_{\rm s})/\alpha E_{\rm s}$. Thus, $C(E_{\rm s})/E_{\rm s}$ is monotonic decreasing and hence its supremum occurs in the limit as $E_{\rm s} \to 0$. Applying l'Hospital's rule, we obtain

$$\lim_{E_{\rm s}\to 0} \left\{\frac{C(E_{\rm s})}{E_{\rm s}}\right\} = \frac{2}{\pi N_0 \ln 2}. \tag{8}$$

■

Proposition 1 and its proof suggest that it is generally a good idea to transmit as little energy as possible per bit, by transmitting as many bits per second as possible, each with a high probability of error. This is because the energy per bit (proportional to the inverse of the number of bits per second transmitted) that achieves (5) is the one achieving the supremum in (6), which is shown in the Appendix to occur in the limit $E_{\rm s} \to 0$. Then, channel coding improves the overall effective probability of error on the channel. In other words, it is better to transmit many low quality bits per second with strong channel coding and use a only small fraction as information bits, than to send fewer but more reliable bits per second. This statement is valid only when a sufficiently large bandwidth and a sufficiently large decoding complexity are available in conjunction with a sufficiently long source symbol length. On the other hand, whenever one of these parameters is restricted, a different solution may arise, as demonstrated in Section III for progressive image coding with BPSK tranmission over an AWGN channel.

The optimal solution given by Proposition 1 applies to infinite bandwidth channels. When there exists a finite bandwidth constraint with no intersymbol interference the minimization of $\mathcal{D}(r_{\rm s})$ is achieved by the minimum value of $E_{\rm s}$ for which the bandwidth constraint is satisfied. Importantly, if the optimum tradeoff is achieved when $E_{\rm s} > E_{\rm b}$, then a bandwidth reduction can be obtained. With intersymbol interference the computation of $p(E_{\rm s})$ becomes more complicated and the general problem of maximizing $C(E_{\rm s})/E_{\rm s}$ is difficult to analyze.

III. Source-Channel Tradeoff for Progressive Image Coding

Specific implementable source coders and channel coders are chosen to illustrate the distortion minimization over various channel code rates and energy per bit allocations. The source coder chosen is the wavelet-based SPIHT image coder, which gives excellent data compression with a progressively transmitted bitstream. The channel codes considered are a family of BCH codes, whose performance is analytically tractable and for which small block lengths can be used.

Suppose an image is to be transmitted under a power constraint P (energy per pixel) over an AWGN channel with noise variance $N_0/2$ and modeled as a BSC using hard-limited BPSK modulation system. We address the following problem:

For an AWGN channel with noise power N_0 and power constraint P, determine from a family of channel codes the code of rate $r_{\rm c}$ and the average energy $E_{\rm s}$ per transmitted BPSK signal that jointly minimize the expected distortion $\mathcal{D}(r_{\rm s})$ of a transmitted image, where $r_{\rm s} = r_{\rm c} P/E_{\rm s}$.

In the following, we study this problem using the progressive image coder of [4]. Once a channel decoding error has occurred, source decoding stops. We restrict attention to the mean squared error (MSE) distortion function, and note that the MSE is a random variable with respect to the the statistics of the channel. For a given AWGN channel with average power constraint P, we attempt to trade off source coding and channel coding to minimize the expected MSE of a transmitted image using a block channel code of fixed length. We assume the Said-Pearlman wavelet-based image compressor is used in conjunction with an (n,k,d) BCH channel code of length n, dimension k, and minimum Hamming distance d [6]. We use BCH codes since they allow a tight error performance analysis by means of their known weight distributions. Also, efficient algebraic decoders for BCH codes have been devised.

For an image of size M pixels transmitted at a rate $R = P/E_{\rm s}$ bpp, the number of blocks of k source coding bits that are encoded by the (n,k,d) code is

$$b(E_{\rm s}) = \left\lceil \frac{M r_{\rm s}}{k} \right\rceil = \left\lceil \frac{MR}{n} \right\rceil. \tag{9}$$

Setting $t = \lfloor (d-1)/2 \rfloor$, the corresponding block error probability after decoding is bounded as

$$P_e \le \sum_{i=t+1}^{n} \binom{n}{i} p(E_{\rm s})^i (1 - p(E_{\rm s}))^{n-i} \tag{10}$$

with equality if only the errors of weight strictly less than $t+1$ are corrected (which is assumed in the following) [6]. The average MSE (averaged over both source and channel statistics), expressed as a function of r_c and E_s, is

$$\delta(r_c, E_s) = \mathcal{D}(r_s)(1-P_e)^{b(E_s)} + \sum_{i=0}^{b(E_s)-1} \mathcal{D}\left(\frac{r_s \cdot i}{b(E_s)}\right)(1-P_e)^i P_e \quad (11)$$

where $\mathcal{D}(r)$ represents the distortion of the image compressed at a rate of r bpp with no channel noise. In (11), the quantity $\mathcal{D}(r_s i/b(E_s))$ corresponds to the distortion achieved when the source decoder stops decoding after i blocks due to the first channel decoder error at the $(i+1)$-th block.

For a fixed channel codeword blocklength n, the smallest possible positive channel code rate is $1/n$. But as $E_s \to 0$ we have $p(E_s) \to 1/2$ and $C(E_s) \to 0$. So for sufficiently small $E_s > 0$ we have $C(E_s) < 1/n$. This implies that the block error probability $P_e \to 1$ as $E_s \to 0$. This in turn implies that

$$\lim_{E_s \to 0} \delta(r_c, E_s) = \mathcal{D}(0). \quad (12)$$

Also, for a given channel code of rate r_c, we have $P_e \to 0$ as $E_s \to \infty$ so that

$$\lim_{E_s \to \infty} \delta(r_c, E_s) = \lim_{E_s \to \infty} \mathcal{D}\left(\frac{R_0 r_c E_b}{E_s}\right) = \mathcal{D}(0). \quad (13)$$

Hence, the minimum value of $\delta(r_c, E_s)$ is achieved for some nonzero choice of E_s. We seek the value of E_s that achieves the minimum $\delta(r_c, E_s)$. Define

$$\delta^{(1)}_{\min}(r_c) = \inf_{E_s > 0} \delta(r_c, E_s).$$

For each (n, k, d) code in the family considered with $r_c = k/n$, the value E_s providing the minimum MSE value $\delta^{(1)}_{\min}(r_c)$ is determined. The corresponding source code rate is $r_s = Pr_c/E_s = R_0 r_c E_b/E_s$. Finally, the channel code of rate r_c that minimizes the MSE is chosen, and the minimum MSE is

$$\delta_{\min} = \min_{r_c}\left\{\inf_{E_s}\{\delta(r_c, E_s)\}\right\} = \min_{r_c}\left\{\delta^{(1)}_{\min}(r_c)\right\}. \quad (14)$$

In the approach of [2], the channel code chosen corresponds to the MSE

$$\delta^{(2)}_{\min}(E_b) = \min_{r_c}\{\delta(r_c, E_b)\} \geq \delta_{\min}. \quad (15)$$

Note that if a bandwidth constraint is imposed on the channel, then only values E_s which satisfy this bandwidth constraint have to be considered in (11).

The family of channel codes considered in Fig. 1 is the set of all binary primitive BCH codes with blocklength $n = 127$ (see [6, p. 584]). This family consists of 20 (n, k) codes with $k = 1$, 8, 15, 22, 29, 36, 43, 50, 57, 64, 71, 78, 85, 92, 99, 106, 113, 120, 126, 127. The *peak signal-to-noise ratio* (PSNR) is defined as

$$\text{PSNR} = 10 \log_{10}\left[\frac{255^2}{\delta(r_c, E_s)}\right]. \quad (16)$$

Fig. 1 plots the end-to-end PSNR of the transmitted image as a function of the ratio E_s/E_b. It has a fixed value for R_0, $p(E_b)$, and n, and shows a family of curves, one curve for each value of k (or equivalently for each channel code rate $r_c = k/n$). The quantity $p(E_b)$ allows one to deduce the value of E_b/N_0 from (3). For each E_s/E_b, the quantity $p(E_s)$ is then calculated using (1). When $E_s = E_b$, the transmission rate is by definition R_0 bpp, and for other values of E_s/E_b the transmission rate is calculated using $R = R_0 \cdot (E_b/E_s)$ from (1) and (4). Each such curve gives the end-to-end theoretically predicted image coder performance measured in PSNR, as a function of the quantity E_s/E_b. The PSNR is calculated for each E_s/E_b using (9) and (10) with equality, (11), and (16). With respect to the approach of [2] (represented by '+' in Fig. 1), it can be seen that an improved SNR (up to 1 dB) can be achieved at the expense of a bandwidth expansion, while maintaining the same power constraint. Although the optimum PSNR value 40.38 dB achieved for $E_s/E_b = 0.54$ results in 85% bandwidth expansion, bandwidth expansion and PSNR can be traded based on our optimization. In fact, this tradeoff can result in a bandwidth reduction: for $E_s/E_b = 1.3$, a PSNR value of 38.6 dB can be achieved in conjunction with about 25% bandwidth reduction for this channel. Similar results for other values of R_0 and $p(E_b)$ have been reported in [7]. In some cases, the optimization proposed in this letter was achieved in conjunction with a bandwidth reduction.

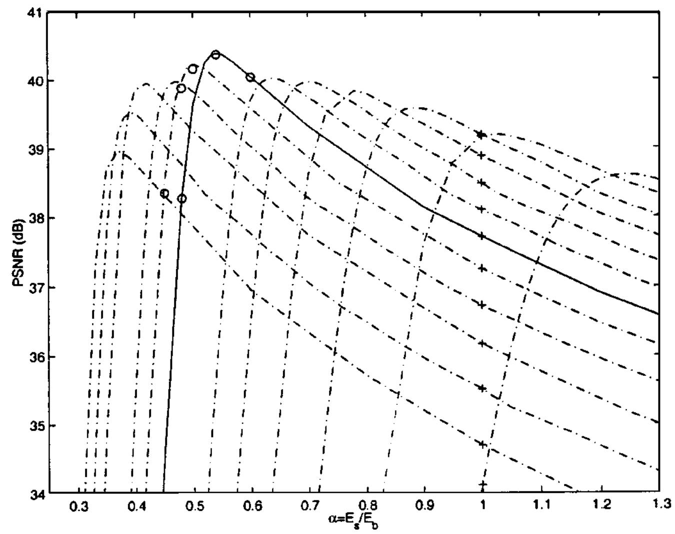

Fig. 1. PSNR values [computed from (16) using (10)] for coding 512×512 Lena image and transmitting with BPSK across a power constrained Gaussian channel quantized to two levels and using (n, k) BCH codes with $n = 127$ and $k = 36$, 43, 50, 57, 64, 71, 78, 85, 92, 99, 106, 113 (read from left to right near the bottom). The '+' symbols indicate the PSNR obtained using the method in [2] for this family of codes (i.e., $E_s = E_b$). The 'o' symbols indicate experimentally computed values of PSNR ($k = 64$ and $k = 71$). $p(E_b) = 10^{-3}$ and $R_0 = 1$ bpp.

The proposed method can be extended to other modulation techniques and to coded modulation if bandwidth is constrained. We also verified that this approach can be extended to soft decision decoding of the channel code after replacing (10) by the conventional union upper bound based on the weight distribution of the code considered.

Appendix

The proof of the concavity of $C(E_s)$ with respect to E_s is unexpectedly long. We first present three lemmas.

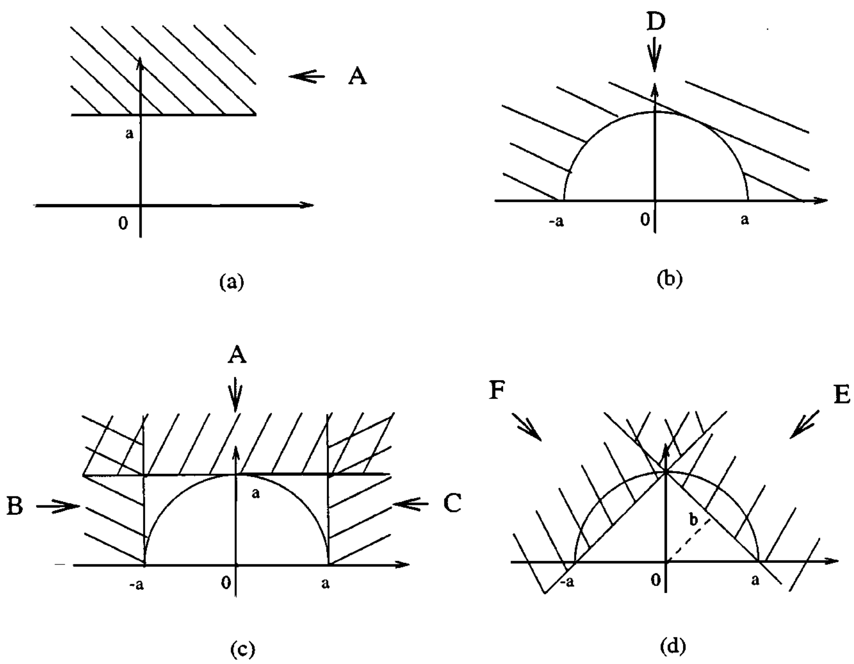

Fig. 2. The sets A, B, C, D, E, F used in the proof of Lemma 2.

Lemma 1: [8] For $x \geq 0$

$$Q(x) > \frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} e^{-x^2/2}. \tag{17}$$

Lemma 2: For $x \geq 0$

$$\frac{1}{4} e^{-x^2} \leq Q(x)(1-Q(x)) \leq \frac{1}{4} e^{-x^2/2}. \tag{18}$$

Proof: The proof is analogous to [9, p. 123, Problem 2–26]. Define the set function h by

$$h(S) = \frac{1}{2\pi} \int \int_s \exp\left[-\frac{(x^2+y^2)}{2}\right] dxdy$$

for each Lebesgue measurable set S in the plane. Also, let $a \geq 0$ and define the sets

$$\begin{aligned} A &= \{(x,y) : y \geq a\} \\ B &= \{(x,y) : x \leq -a \text{ and } y \geq 0\} \\ C &= \{(x,y) : x \geq a \text{ and } y \geq 0\} \\ D &= \{(x,y) : x^2 + y^2 \geq a \text{ and } y \geq 0\} \\ E &= \{(x,y) : y \geq a - x\} \\ F &= \{(x,y) : y \geq x + a\} \\ G &= \{(x,y) : y \geq 0\}. \end{aligned}$$

Since $A \subset D$ we have [see Fig. 2(a) and (b)]

$$\begin{aligned} Q(a) &= h(A) \leq h(D) \\ &= \frac{1}{2\pi} \int_0^{\pi} \int_a^{\infty} e^{-r^2/2} r dr d\theta = \frac{1}{2} e^{-a^2/2}. \end{aligned}$$

It is easy to see that $h(B) = h(C) = Q(a)/2$ and $h(A \cap B) = h(A \cap C) = Q^2(a)$. Since $A \cup B \cup C \subset D$, it follows that [see Fig. 2(b) and (c)]

$$\begin{aligned} \frac{1}{2} e^{-a^2/2} &= h(D) \\ &\geq h(A \cup B \cup C) \\ &= h(A) + h(B) + h(C) - h(A \cap B) - h(A \cap C) \\ &= 2(Q(a) - Q(a)^2) \end{aligned}$$

or equivalently by taking $x = a$

$$Q(x)(1-Q(x)) \leq \frac{1}{4} e^{-x^2/2}. \tag{19}$$

Since $D \subset (E \cup F) \cap G$ we have [see Fig. 2(b) and (d)]

$$\begin{aligned} h(D) &\leq h((E \cup F) \cap G) \\ &= h(E) + h(F) - h(E \cap F) \\ &\quad - h(E \cap G^c) - h(F \cap G^c) \\ &= Q\left(\frac{a}{\sqrt{2}}\right) + Q\left(\frac{a}{\sqrt{2}}\right) - Q^2\left(\frac{a}{\sqrt{2}}\right) \\ &\quad - \frac{1}{2} Q^2\left(\frac{a}{\sqrt{2}}\right) - \frac{1}{2} Q^2\left(\frac{a}{\sqrt{2}}\right) \\ &= 2\left(Q\left(\frac{a}{\sqrt{2}}\right) - Q^2\left(\frac{a}{\sqrt{2}}\right)\right) \end{aligned}$$

using the circular symmetry of the function h about the origin. Taking $x = a/\sqrt{2}$ completes the proof. ∎

In the following lemma, "sgn" represents the function

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Lemma 3: If f and g are nonnegative functions such that f is differentiable, $g(x) \geq f(0)$, and for all x

$$\operatorname{sgn}(f'(x)) = \operatorname{sgn}(-f(x) + g(x)) \tag{20}$$

then for all $x > 0$

$$f(x) \geq f(0). \tag{21}$$

Proof: Let $Z = \inf\{x \geq 0 : f(x) < f(0)\}$. If $Z = \infty$, then we are done, so assume Z is finite. Since f is continuous, we must have $f(Z) = f(0)$. Also, since f is continuous, there exists $d > 0$ such that $f(x) < f(0)$ for all $x \in (Z, Z+d)$. But $g(x) \geq f(0)$ implies that $f'(x) > 0$ in the interval $(Z, Z+d)$. Since $f(Z) = f(0)$, this implies $f(x) > f(0)$ in the interval $(Z, Z+d)$, which contradicts the original assumption about this interval. ■

Theorem 1: The capacity $C(E_s)$ is a concave function of the energy per bit E_s.

Proof: The capacity of the BSC can be written as

$$C_e(E_s) = \ln 2 - h_e\left(Q\left(\sqrt{\frac{2E_s}{N_0}}\right)\right) \tag{22}$$

where $h_e(\cdot)$ is the binary entropy function, defined by

$$h_e(x) = -x\ln(x) - (1-x)\ln(1-x).$$

After some algebra, we obtain for $q = Q(\sqrt{2E_s/N_0})$

$$\frac{\partial C_e(E_s)}{\partial E_s} = \frac{1}{2}(\pi N_0 E_s)^{-1/2}e^{-E_s/N_0}\ln\left(\frac{1-q}{q}\right) \tag{23}$$

$$\begin{aligned}\frac{\partial^2 C_e(E_s)}{\partial E_s^2} &= \frac{1}{2}(\pi N_0)^{-1/2}e^{-E_s/N_0}\\ &\quad\times\left(-\ln\left(\frac{1-q}{q}\right)E_s^{-1/2}\left(\frac{1}{2E_s}+\frac{1}{N_0}\right)\right.\\ &\quad+\frac{1}{2}(\pi N_0)^{-1/2}e^{-E_s/N_0}\\ &\quad\left.\times q^{-1}(1-q)^{-1}E_s^{-1}\right).\end{aligned} \tag{24}$$

Lemma 1 implies that

$$\begin{aligned}&\frac{1}{2}(\pi N_0)^{-1/2}e^{-E_s/N_0}q^{-1}(1-q)^{-1}E_s^{-1}\\ &\quad< E_s^{-1/2}\left(\frac{1}{2E_s}+\frac{1}{N_0}\right)(1-q)^{-1}\end{aligned} \tag{25}$$

so that $\partial^2 C_e(E_s)/\partial E_s^2 < 0$ for

$$\ln\left(\frac{1-q}{q}\right) > (1-q)^{-1}. \tag{26}$$

Define $f(E_s) = \ln((1-q)/q)E_s^{-1/2}$ with $f(0) = 4(\pi N_0)^{-1/2}$ (f is continuous by l'Hospital). We compute

$$\begin{aligned}\frac{\partial f(E_s)}{\partial E_s} &= \frac{1}{2}E_s^{-1}\left(-f(E_s) + (\pi N_0)^{-1/2}\right.\\ &\quad\left.\times e^{-E_s/N_0}q^{-1}(1-q)^{-1}\right).\end{aligned} \tag{27}$$

Lemma 2 gives

$$(\pi N_0)^{-1/2}e^{-E_s/N_0}q^{-1}(1-q)^{-1} \geq 4(\pi N_0)^{-1/2} \tag{28}$$

with $4(\pi N_0)^{-1/2} = f(0)$. As a result, after applying Lemma 3 to (27), we obtain

$$\ln\left(\frac{1-q}{q}\right)E_s^{-1/2} \geq 4(\pi N_0)^{-1/2}. \tag{29}$$

It follows that

$$\begin{aligned}&-\ln\left(\frac{1-q}{q}\right)E_s^{-1/2}\left(\frac{1}{2E_s}+\frac{1}{N_0}\right)\\ &\quad+\frac{1}{2}(\pi N_0)^{-1/2}e^{-E_s/N_0}q^{-1}(1-q)^{-1}E_s^{-1}\\ &\leq -4(\pi N_0)^{-1/2}\left(\frac{1}{2E_s}+\frac{1}{N_0}\right)\\ &\quad+\frac{1}{2}(\pi N_0)^{-1/2}e^{-E_s/N_0}q^{-1}(1-q)^{-1}E_s^{-1}\\ &\leq -4(\pi N_0)^{-1/2}\left(\frac{1}{2E_s}+\frac{1}{N_0}\right)+2(\pi N_0)^{-1/2}e^{E_s/N_0}E_s^{-1}\\ &= 2(\pi N_0)^{-1/2}E_s^{-1}\left(e^{E_s/N_0}-\left(1+\frac{2E_s}{N_0}\right)\right)\end{aligned} \tag{30}$$

where the second inequality follows from Lemma 2. Finally, the function $p(x) = e^x - (1+2x)$ is nonpositive for $0 \leq x < \ln 2$ as $p(0) = 0$ and $\partial p(x)/\partial x = e^x - 2 < 0$ for $x < \ln 2$. It follows from (24) and (30) that $\partial^2 C_e(E_s)/\partial E_s^2 \leq 0$ whenever $E_s/N_0 < \ln 2$. On the other hand, if we let $x = 1/(1-q)$, then $\ln((1-q)/q) > 1/(1-q)$ becomes $x < 1 + e^{-x}$, which holds for $x < 1.278\,46$ or equivalently $q < 0.2178$. Using $q = Q(\sqrt{2E_s/N_0})$ we see that $\partial^2 C_e(E_s)/\partial E_s^2 \leq 0$ also holds whenever $E_s/N_0 > 0.3042$. Since $0.3042 < \ln 2$, the proof is completed based on (26). ■

ACKNOWLEDGMENT

The authors would like to thank E. Ayanoglu for helpful suggestions which improved the letter.

REFERENCES

[1] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sept. 1997.

[2] P. G. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Lett.*, vol. 4, pp. 189–191, July 1997.

[3] ——, "Error protection for progressive image transmission over memoryless and fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1555–1559, Dec. 1998.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.

[5] R. E. Blahut, *Digital Transmission of Information*. Reading, MA: Addison-Wesley, 1990.

[6] S. Lin and D. J. Costello Jr, *Error Control Coding Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[7] M. Fossorier, Z. Xiong, and K. Zeger, "Joint source-channel image coding for a power constrained noisy channel," in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 2, Chicago, IL, Oct. 1998, pp. 122–126.

[8] P. O. Borjesson and C. W. Sundberg, "Simple approximations of the error function $Q(x)$ for communications applications," *IEEE Trans. Commun.*, vol. COM-27, pp. 639–643, Mar. 1979.

[9] J. M. Wozencraft and I. M. Jacobs, *Principles of Communications*. New York: McGraw-Hill, 1965.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.