

# A Gaussian Input Is Not Too Bad

Ram Zamir and Uri Erez \*

Dept. of Elect. Eng. - Systems, Tel Aviv University, ISRAEL  
*erez,zamir@eng.tau.ac.il*

*Submitted to the IEEE Tr. on Infor. Theory. June 2002; Revised Dec. 2003*

January 1, 2004

## Abstract

We consider the problem of choosing a robust input for communicating over an input constrained additive noise channel where the noise distribution is arbitrary. We show that the mutual information rate achievable using a white Gaussian input never incurs a loss of more than half a bit per sample with respect to the power constrained capacity. For comparison, for the family of colored Gaussian noise channels a white Gaussian input loses at most  $\log(e)/2e \approx 0.265$  bit per sample with respect to the optimum water-pouring solution. For general input constraints, we derive a formula for choosing the best input in the min-max capacity loss (bound) sense. The bound on the capacity loss is tight for Pulse Position Modulation in the presence of a bursty jammer.

*Index Terms:* Unknown channels, min-max rate loss, Gaussian codebook, white versus water-pouring spectrum.

## 1 Introduction

Consider additive noise channels of the form

$$Y = X + N \tag{1}$$

where the encoder is subject to a power constraint  $E\{X^2\} \leq A$ . It is well known that for high SNR capacity  $C(A)$  can be achieved with a Gaussian input, even when  $N$  is not Gaussian.

---

\*This work was presented in part in the annual Allerton Conference, Illinois, Oct. 2002.

A capacity achieving input, however, should maximize in general the entropy of the channel *output*. Since a Gaussian random variable has the greatest entropy under a power constraint (see, e.g, [1]), a natural question is the following: How much do we lose by using a Gaussian random variable  $X^* \sim \mathcal{N}(0, A)$  instead of the optimal input distribution? In particular, is the rate loss bounded, when considering all possible noise random variables  $N$ ? In the next section we state a general form of the following theorem.

**Theorem 1** *For any power constrained additive noise channel (1),*

$$C(A) \leq I(X^*; X^* + N) + \frac{1}{2} \text{ bit} \quad (2)$$

where  $X^* \sim \mathcal{N}(0, A)$ , and  $C(A)$  is the channel capacity. Thus a Gaussian input loses no more than half a bit with respect to capacity.

This theorem is dual to the following result regarding rate distortion proved in [11], which we recall here.

**Theorem 2** ([11]) *For any source  $Y$*

$$R(D) \geq I(Y; Y + X^*) - \frac{1}{2} \text{ bit} \quad (3)$$

where  $X^* \sim \mathcal{N}(0, D)$ , and  $R(D)$  is the rate-distortion function of  $Y$  at mean squared distortion  $D$ . Thus the redundancy of a Gaussian noise “test” channel is no more than half a bit above the quadratic rate-distortion function.

As noted earlier at high SNR a Gaussian input becomes optimal and the rate loss goes to zero. Theorem 1 gives an *additive* bound on the rate loss which is interesting mostly for “medium SNR”. At low SNR (more precisely, when the capacity is small), a *relative* bound however is more meaningful. See the discussion in the last section.

Section 2 shows how to choose an input which minimizes the worst capacity loss bound for a general input constraint. In Section 3 we specialize the results of Section 2 to power, to peak and to general  $r$ -th power constrained channels. Section 4 extends the half a bit bound to power constrained channels with memory. Section 5 compares this bound with the worst rate loss of a white Gaussian input within the family of colored Gaussian noise channels, which is about 0.265 bit per channel use. Section 6 constructs an example of a channel and input constraint in which the min-max capacity loss bound of Section 2 is tight: “pulse position modulation” in the presence of a “bursty jammer”. In the Discussion section we relate the results to communication over unknown channels.

## 2 Main result

We prove a variant of Theorem 1 for a general input cost function  $c(\cdot)$  and then specialize it to the case of a power constraint. Consider the additive noise channel (1) where  $X, N$  and  $Y$  are either real valued or discrete, and the  $+$  sign denotes either regular or modulo addition. Suppose the channel input satisfies the constraint

$$E\{c(X)\} \leq A. \quad (4)$$

Denote the set of random variables satisfying (4) by  $\mathcal{X}(A)$ . The following theorem is proved in the Appendix.

**Theorem 3** *For any input  $X$  satisfying (4), and any input  $X'$ ,*

$$I(X; X + N) \leq I(X'; X' + N) + C' \quad (5)$$

where  $C'$  is the capacity of an auxiliary channel with additive noise  $X'$  and input constraint (4), i.e.,

$$C' = \sup_{Z \in \mathcal{X}(A)} I(Z; Z + X'). \quad (6)$$

In particular

$$C(A) = \sup_{X \in \mathcal{X}(A)} I(X; X + N) \leq I(X'; X' + N) + C'. \quad (7)$$

Thus, the rate loss with respect to the capacity  $C(A)$  for using the input  $X'$  is bounded by  $C'$ . This bound holds for any  $X'$ , not necessarily in  $\mathcal{X}(A)$ . The minimum bound over all possible inputs  $X'$  in  $\mathcal{X}(A)$  is

$$C^* = \inf_{X' \in \mathcal{X}(A)} \sup_{Z \in \mathcal{X}(A)} I(Z; Z + X'). \quad (8)$$

Note that the figure  $C^*$  generally depends on the channel alphabet and on both  $c(\cdot)$  and  $A$ . If there is a noise  $X'$  achieving  $C^*$ , we denote it by  $X^*$  and call it the “minimax input” (or “minimax prior”). That is,  $X^*$  is the optimal constrained input in the sense of minimizing the bound on the maximum capacity loss within the class of additive noise channels. For any noise  $N$ , the loss with respect to the capacity  $C(A)$  for using  $X^*$  is bounded by  $C^*$ . The question of minimizing the maximum capacity loss *directly* (and not the *bound* of Theorem 3) is discussed in the last section.

### 3 Examples: Power, Peak and $r$ -th Power Constraints

Interestingly, the figure  $C^*$  appears as a universal bound on the rate loss also in other problems. In [9]  $C^*$  was defined and studied in the context of the Wyner-Ziv problem. As shown by Lastras and Berger [5], the quantity  $C^*$  also bounds the rate loss in successive refinement. Furthermore, in [10]  $C^*$  is shown to bound the loss in “writing on dirty paper”.

Note that the set  $\mathcal{X}(A)$  is convex and hence, due to the convexity of the mutual information, the order of the infimum and supremum in (8) may be reversed so  $C^*$  is also the *maximin* solution. Hence technically, the figure  $C^*$  amounts to the *compound capacity* relative to the class of auxiliary additive noise channels with input constrained to  $\mathcal{X}(A)$  and noise in  $\mathcal{X}(A)$ , while the minimax input  $X^*$  amounts to the worst noise in that class. See, e.g., [2] for the definition of compound capacity.

We calculate  $C^*$  and  $X^*$  for some common input constraints. While in general  $C^*$  may depend on  $A$ , fortunately for real alphabet and absolute  $r$ -th power constraint  $c(x) = |x|^r$  this is not the case. For the cases of  $r = 2$  and  $r = \infty$  we show that  $C^*$  is achieved by the *maximum entropy* noise distribution under the respective constraints.

- For  $r = 2$  a Gaussian noise  $X^* \sim \mathcal{N}(0, A)$  achieves  $C^*$ . This follows since a Gaussian noise is the worst in terms of capacity; see the “quadratic transmitter-jammer game” described in [1]. We get  $C^* = \frac{1}{2} \log(1 + 1) = \frac{1}{2}$  which yields Theorem 1.

Denoting by  $D(X' \| X^*)$  the Kullback-Leibler divergence of any  $X'$  with variance  $A$  from Gaussianity, we have (see, e.g., [12])  $h(X') = h(\mathcal{N}(0, A)) - D(X' \| X^*)$  and  $h(Z + X') \leq h(\mathcal{N}(0, 2A))$ , and therefore  $C' \leq C^* + D(X' \| X^*) = 1/2 + D(X' \| X^*)$ . Hence we obtain

$$C(A) \leq I(X'; X' + N) + D(X' \| X^*) + \frac{1}{2}, \quad (9)$$

i.e., the capacity loss for using  $X'$  is at most  $D(X' \| X^*) + 0.5$  bit. In particular, if  $X'$  is a uniform input then  $D(X' \| X^*)$  is exactly the “shaping gain”, and we get

$$C(A) \leq I(X'; X' + N) + \frac{1}{2} \log(2\pi e G_1) + \frac{1}{2} \quad (10)$$

where  $G_1 = 1/12$  and  $\frac{1}{2} \log 2\pi e G_1 \approx .2546$  bit. This is similar to Ziv’s result on universal dithered scalar quantization [13]. More generally, for an input  $\mathbf{X}'$  uniformly distributed over some  $k$ -dimensional “shaping” region  $\mathbb{R}$ , we have

$$C(A) \leq \frac{1}{k} I(\mathbf{X}'; \mathbf{X}' + \mathbf{N}) + \frac{1}{2} \log(2\pi e G(\mathbb{R})) + \frac{1}{2} \quad (11)$$

where  $G(\mathbb{R}) = \frac{\int_{\mathbb{R}} \|\mathbf{x}\|^2 d\mathbf{x}}{k \text{Vol}(\mathbb{R})^{1+2/k}}$  is the normalized second moment of  $\mathbb{R}$ . This result parallels a bound given in [13, 11].

- The limit  $r \rightarrow \infty$  corresponds to a peak constraint  $X \in [-A, A]$ . A uniform noise  $X^* \sim \text{Unif}(-A, A)$  achieves  $C^*$  in this case, and  $C^* = 1$  bit. To see this, note that the maximizing input ( $Z$  in (8)) places two equal point masses at  $+A, -A$ . With this particular  $Z$ , for any choice of  $X'$ , we have  $I(Z; Z + X') = 1$  and hence

$$1 = \inf_{X' \in [-A, A]} I(Z; Z + X') \leq C^* \leq \sup_{X \in [-A, A]} I(X; X + X^*) \leq \log 4A - \log 2A = 1. \quad (12)$$

**Corollary 1** *For the additive noise channel (1) with input subject to a peak constraint  $X \in [-A, A]$ ,*

$$C(A) \leq I(X^*; X^* + N) + 1 \quad (13)$$

*where  $X^* \sim \text{Unif}(-A, A)$ . Thus a uniform input loses no more than one bit with respect to capacity.*

- For general  $r \geq 1$  the maximum entropy noise  $X'$  guarantees  $C' \leq 1$  bit, and therefore

$$C(A) \leq I(X'; X' + N) + 1 \text{ bit}. \quad (14)$$

This of course implies that also  $C^* \leq 1$  bit, but the actual distribution of  $X^*$  might be different. The bound follows since if  $Z$  and  $X'$  are in  $\mathcal{X}(A)$  then  $Z + X'$  is in  $\mathcal{X}(2^r A)$  (see Appendix). Noting that the maximum entropy with absolute  $r$ -th power  $A$  has the form  $c_r + 1/r \log(A)$  [1], we write

$$I(Z; Z + X') = h(Z + X') - h(X') \leq c_r + \frac{1}{r} \log(2^r A) - h(X'),$$

and the 1 bit bound on  $C'$  follows by substituting the maximum noise entropy  $h(X') = c_r + \frac{1}{r} \log(A)$ .

## 4 Extension to Channels with Memory

We now discuss possible vector and process extensions of Theorem 1. These are dual to [12].

**Corollary 2** *Let  $C$  be the capacity of a power constrained (to power  $A$ ) discrete time additive noise channel with a stationary and ergodic noise  $\{N_i\}$ . That is, [3],*

$$C = \lim_{k \rightarrow \infty} \frac{1}{k} \sup_{\mathbf{X}: E\|\mathbf{X}\|^2/k \leq A} I(\mathbf{X}; \mathbf{X} + \mathbf{N}). \quad (15)$$

*Then*

$$C \leq \lim_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{X}^*; \mathbf{X}^* + \mathbf{N}) + 1/2 \quad (16)$$

*where  $\mathbf{X}^* \sim \mathcal{N}(0, A)$  is i.i.d. vector.*

Thus an *i.i.d. Gaussian* input loses no more than half a bit/sample with respect to capacity.

*Proof:* The proof is obtained by retracing the proof of Theorem 1, replacing the scalar random variables  $X$ ,  $X^*$ , and  $N$  with the corresponding random vectors. Note that the limits in  $k$  above exist by the stationarity of  $\{N_i\}$ .  $\square$

By standard arguments (e.g., [3, 12]), these results extend straightforwardly to a continuous time channel of bandwidth  $W$  Hz. Note that for this channel the term “Nyquist sample” means  $1/2W$  seconds.

**Corollary 3** *Let  $C$  be the capacity of a power constrained bandlimited continuous time channel of bandwidth  $W$  Hz, with stationary and ergodic additive noise. Denote by  $C_{WG}$  the mutual information rate corresponding to a Gaussian input which is white over the band  $(0, W)$ . Then*

$$C \leq C_{WG} + W \text{ [bit/second]}, \quad (17)$$

*i.e., the loss is at most  $W/2W = 1/2$  bit per Nyquist sample.*

## 5 Example: White versus Water-Pouring Spectrum

To assess the figures above, we compare them with an explicit calculation of the min-max loss relative to a specific family of channels. Consider power constrained *colored additive Gaussian* noise channels of bandwidth  $W$  Hz.

**Proposition 1** *The capacity loss for using a white Gaussian input for a colored additive Gaussian noise channel is bounded by*

$$C - C_{WG} \leq \frac{\log(e)}{2e} \approx 0.265 \text{ [bit per Nyquist sample]}$$

*(or  $W \log(e)/e \approx 0.53W$  bit per second). This bound is asymptotically tight (as  $\epsilon \rightarrow 0$ ) if the power spectral density of the noise has the form:*

$$N(f) = \begin{cases} \epsilon, & \text{a fraction } 1/e \text{ of the band} \\ \infty, & \text{rest of the frequencies.} \end{cases} \quad (18)$$

The proof of this proposition is given in the Appendix. It follows that the universal bound of  $\frac{1}{2}$  bit per Nyquist sample in (17) is *not* achieved within the family of Gaussian noise channels with memory.

We also mention that Schein and Trott [7] have analyzed a related question where one uses, as often done in practice, an input power spectrum that is uniform over a band of frequencies rather than the optimal “water pouring” input spectrum. They obtained tight bounds on the resulting loss when using the optimal input spectrum within this limited class.

## 6 Example: Pulse Position Modulation in the Presence of a Bursty Jammer

In the following example there exists a class of noises for which the available capacity is roughly equal to  $C^*$ , while for *every* valid channel input  $X$  (not necessarily the minimax input  $X^*$ ) there exists a noise in that class for which  $X$  achieves only a negligible portion of  $C^*$ . Thus, the worst capacity loss of  $X^*$  is close to  $C^*$ , and the bound  $C^*$  is tight and cannot be improved by any other input.

Consider the case where  $X, N, Y \in \{0, 1\}^m$  and the  $+$  sign denotes component-wise modulo-2 addition. That is, each letter is an  $m$ -tuple binary vector. Let  $\mathcal{A}$  denote the set of  $x \in \{0, 1\}^m$  with exactly one nonzero component, i.e.,  $\mathcal{A} = \{(10 \dots 0), (010 \dots 0), \dots, (0 \dots 01)\}$ . We have  $|\mathcal{A}| = m$ . Suppose the input cost function is

$$c(x) = \begin{cases} 0, & x \in \mathcal{A} \\ \infty, & x \notin \mathcal{A} \end{cases} \quad (19)$$

thus only  $x$ 's in  $\mathcal{A}$  are allowed at the channel input. We can view transmission using this set of channel input letters as “pulse position modulation” (PPM).

It is shown in the Appendix that for this case the min-max capacity loss bound (8) is

$$C^* = \log(m) - 1 + o(1) \text{ bit} \quad (20)$$

where  $o(1) \rightarrow 0$  as  $m \rightarrow \infty$ , and  $C^*$  is achieved by  $Z$  and  $X^*$  which are uniform on  $\mathcal{A}$ . Thus  $X^* \sim \text{Unif}(\mathcal{A})$  is the minimax input for pulse position modulation over block binary symmetric channels as above.

Consider now the channel  $Y = X + N$  with the input constraint above, and suppose the additive noise  $N$  belongs to the following class. Some  $m - m'$  components of  $N$  are i.i.d. binary symmetric, while the other  $m'$  components of  $N$  are fixed to zero. In other words,  $m - m'$  of the components are “very noisy” while the remaining  $m'$  components are noise-free. We can view this class of noises as “bursty jammers”. Clearly, if the location of the noisy components of  $N$  is known, then

$$C = \log(m' + 1) = \log(\alpha m + 1),$$

where  $\alpha = m'/m$  denotes the relative portion of noise-free components, and  $C$  is achieved by an input  $X$  which is uniform over  $m' + 1$  members of  $\mathcal{A}$  corresponding to the  $m'$  noise-free components and an arbitrary letter from the noisy components.

On the other hand, the mutual information associated with the minimax input  $X^* \sim \text{Unif}(\mathcal{A})$  is

$$I(X^*; X^* + N) = H_B(\alpha) + \alpha \log(\alpha m) \quad (21)$$

for any bursty jammer  $N$ . See the Appendix. Moreover, for any PPM input  $X'$  there exists a “bad” bursty jammer  $N$  in the class above for which

$$I(X'; X' + N) \leq I(X^*; X^* + N). \quad (22)$$

This  $N$  has its noise-free components located on the  $m'$  least probable positions of  $X'$ . Hence, by letting  $m \rightarrow \infty$  while keeping  $0 < \alpha < 1$  small, we can make  $C/C^*$  close to one while  $I(X'; X' + N)/C^*$  close to zero. The intuition behind this is that while an optimum input puts positive probability only on the (small subset of) noise-free positions, the mismatched input  $X'$  wastes its most probable positions on corrupted components.

In the example above  $C^* \rightarrow \infty$ , and the capacity loss  $C - I(X^*; X^* + N)$  approaches the bound  $C^*$  in a *relative sense*. To get a non-relative version of this example, imagine a continuous time implementation of this channel, where  $m$  is roughly the product of the channel bandwidth  $W$  and the symbol duration  $T$ . Assume further that while  $m = WT$  is taken to infinity, the ratio  $\log(WT)/T$  is kept constant. This way the capacity  $C^*$  in *bits per unit time* is approximately

$$C^* \approx \frac{\log WT}{T}$$

i.e., roughly constant. Similarly, all other information rates per unit time are bounded in the interval  $(0, \log(WT)/T)$ .

## 7 Discussion and Open Questions

One situation where Theorem 3 may be applied is in the case of communication over an unknown additive noise channel. In such a situation, the receiver may learn the channel noise distribution to any degree of required precision with asymptotically no rate loss. Therefore the receiver’s lack of prior knowledge about the channel noise does not result in any penalty in capacity. Indeed, universal decoders have been extensively studied, see [4] for relevant references.

However, provided there is no feedback link between receiver and transmitter, the transmitter remains totally ignorant of the noise distribution throughout the whole transmission period. Thus, it is this latter ignorance that inevitably must incur a rate loss. There are different approaches to defining both the goals and the methods for communication under channel uncertainty, see [4] for a comprehensive survey. One approach is that of considering the *compound channel capacity* of a class of channels. In this setting the goal is to maximize the capacity corresponding to the worst channel in the class. This results in a rather conservative design and in many cases is useless as the resulting maximin capacity is zero. This is indeed the case for the class of all power constrained additive noise channels considered in this paper. We may alternatively set as a goal to design a communication system that



operates “not too far” from capacity for any particular channel in the class. Our results are relevant for this design approach.

Similar questions on robust input distributions have been studied for non constrained discrete memoryless channels. For instance, Majani and Rumsey [6] and Shulman and Feder [8] have proved that for any binary input DMC the ratio of the uniform input capacity to the true capacity is always at least 0.94. Thus a uniform input is “never too bad” for a binary input channel. However their bound on the loss increases with the input alphabet.

Our analysis differs from the above in that it is restricted to the class of additive or symmetric noise channels, yet on the other hand it is virtually independent of the alphabet size. Our result indicates that for an arbitrary power constrained additive noise channel, if limited feedback is available it is sufficient to inform the transmitter of the *capacity* of the channel for reliable communication at rate half a bit smaller than capacity. We also note that even this limited feedback is unnecessary if variable rate decoding is allowed (see, e.g., [8]).

Having demonstrated that the penalty for using a mismatched codebook is bounded, we can define the best fixed codebook distribution, i.e., the input distribution that minimizes the rate loss (in a min max sense). Specifically, define

$$\mathcal{L} = \inf_{X' \in \mathcal{X}(A)} \sup_N \{C(A, N) - I(X'; X' + N)\} \quad (23)$$

where  $C(A, N)$  is the channel capacity defined in (7) as  $C(A)$ . Theorem 3 and (8) give the bound  $\mathcal{L} \leq C^*$ . As shown in the previous section, in some cases this bound is approximately tight, that is  $\mathcal{L}/C^* \approx 1$ . In these cases  $X^*$  approximates the best input  $X'$  in (23). However, for power constrained channels the half a bit bound is not tight, and therefore a *non*-Gaussian  $X'$  is possibly better.

A related open question, motivated by the *low SNR* regime, is whether there is a robust input distribution for which the *relative* rate loss is bounded for the class of power constrained additive noise channels. That is, we seek for an input  $X$  for which

$$I(X; X + N) > \alpha C(A, N) \quad (24)$$

for some  $\alpha > 0$  and any noise  $N$ . As we saw in Section 6, for PPM channels there are no such input and positive  $\alpha$  as the loss can be 100% of the available capacity for any input. Yet for power constrained channels this remains a question for future study.

## Appendix

### A. Proof of Theorem 3

We prove Theorem 3 using the following lemma.

**Lemma 1** *Let  $W, Y, Z$  be three real or discrete valued mutually independent random variables, and let the “+” sign denote real or modulo addition. Then*

$$I(W; W + Z) \leq I(W; W + Y) + I(Y; Y + Z). \quad (25)$$

*Proof:* By the independence of  $W, Y$  and  $Z$  and by the chain rule of mutual information we have

$$I(W; W + Z) = I(W; W + Z, W + Z + Y) \quad (26)$$

$$= I(W; W + Z + Y) + I(W; W + Z | W + Z + Y). \quad (27)$$

We bound the first term by

$$I(W; W + Z + Y) \leq I(W; W + Y). \quad (28)$$

The last term is bounded as follows:

$$I(W; W + Z | W + Z + Y) = I(W - [W + Z + Y]; W + Z - [W + Z + Y] | W + Z + Y) \quad (29)$$

$$= I(Y + Z; Y | W + Z + Y) \quad (30)$$

$$\leq I(Y; Y + Z) \quad (31)$$

where the first two equalities follow since an invertible transformation (addition, negation) given the condition does not affect the conditional mutual information; and the inequality follows from the convexity  $\cap$  of the mutual information with respect to the input distribution, noting that  $(W + Y + Z) \leftrightarrow (Y + Z) \leftrightarrow Y$  form a Markov chain. Combining the above yields the lemma.  $\square$

The theorem now follows by making the following identification

$$Y \leftrightarrow X', W \leftrightarrow X, Z \leftrightarrow N, \quad (32)$$

and noting that  $X \in \mathcal{X}(A)$  implies  $I(X; X + X') \leq C'$ .

## B. Proof of Proposition 1

Check first the achievability of the bound. The optimum “water-pouring” solution for the noise power spectral density  $N(f)$  of (18) is to put all the input power  $A$  on the low noise portion of the spectrum, resulting in capacity of  $C = W/e \log(1 + Ae/[W\epsilon])$ . The mutual information rate associated with white Gaussian input is given by (see, e.g., [3])  $C_{WG} = \int_0^W \log(1 + A/[WN(f)])df = W/e \log(1 + A/[W\epsilon])$ . Thus

$$C - C_{WG} = \frac{W}{e} \log \left( \frac{\epsilon W + Ae}{\epsilon W + A} \right)$$

which converges to  $W \log(e)/e$  as  $\epsilon \rightarrow 0$ .

To prove the inequality in the lemma, note that channel capacity has the form  $C = \int_{B_\theta} \log(\theta/N(f))df$ , where  $B_\theta$  is the frequency band in which  $\theta > N(f)$ , and  $\theta$  is the (unique) value satisfying  $\int_{B_\theta} [\theta - N(f)]df = A$ . This can be rewritten as  $C = \int_0^W \log(\max\{\theta, N(f)\}/N(f))df$ . See [1]. Let  $B'_\theta$  denote the frequency band in which  $\theta - N(f) > A/W$ . Clearly  $B'_\theta \subset B_\theta$ . Thus

$$C - C_{WG} = \int_0^W \log\left(\frac{\{\max\{\theta, N(f)\}}{N(f) + A/W}\right) df \quad (33)$$

$$\leq \int_{B'_\theta} \log\left(\frac{\theta}{N(f) + A/W}\right) df \quad (34)$$

$$\leq \int_{B'_\theta} \log\left(\frac{\theta - N(f)}{A/W}\right) df \quad (35)$$

$$\leq W'_\theta \log\left(\frac{1}{W'_\theta} \int_{B'_\theta} \frac{\theta - N(f)}{A/W} df\right) \quad (36)$$

$$\leq W'_\theta \log\left(\frac{W}{W'_\theta A} \int_{B_\theta} [\theta - N(f)]df\right) \quad (37)$$

$$= W'_\theta \log\left(\frac{W}{W'_\theta}\right) \quad (38)$$

$$\leq W \frac{\log(e)}{e} \quad (39)$$

where  $W_\theta$  and  $W'_\theta$  are the sizes of  $B_\theta$  and  $B'_\theta$ , respectively; here (34) follows because outside  $B'_\theta$  the integrand is nonpositive; (35) follows since  $b \geq c$  implies  $[a + b]/[a + c] \leq b/c$  for positive  $a, b, c$ ; (36) follows from Jensen's inequality; (37) follows since  $B'_\theta \subset B_\theta$  and  $\theta - N(f) \geq 0 \forall f \in B_\theta$ , so the integral can only increase by changing the integration domain from  $B'_\theta$  to  $B_\theta$ ; (38) follows from the definition of  $\theta$ ; and finally (39) follows since the function  $x \log(1/x)$  has a maximum at  $x = 1/e$ .

## C. Min-max Capacity Calculation for the Binary PPM Channel

To prove (20) we use the following lemma.

**Lemma 2** *Let  $U \sim \text{Unif}(\mathcal{A})$  where  $\mathcal{A}$  is the subset of  $\{0, 1\}^m$  defined in Section 5. For any random variable  $X'$  independent of  $U$  and supported on  $\mathcal{A}$*

$$H(U + X') \leq H_B(1/m) + \frac{m-1}{m} \log\binom{m}{2}$$

and

$$H(U|U + X') \leq \frac{m-1}{m} + \frac{\log(m)}{m}$$

with equality in both inequalities if  $X' \sim \text{Unif}(\mathcal{A})$ .

*Proof:* Let  $E$  be the indicator of the event  $U + X' = 0^m$ , or equivalently of the event  $U = X'$ . Then  $H(E|U + X') = 0$ , so that by the chain rule  $H(U + X') = H(U + X', E) = H(E) + H(U + X'|E) = H(E) + \Pr(E = 0)H(U + X'|E = 0)$ , because  $H(U + X'|E = 1) = 0$ . Since  $U$  is uniform we have  $\Pr(U + x' = 0^m) = 1/m$  for any value of  $x'$ , so  $\Pr(E = 1) = 1/m$  and  $H(E) = H_B(1/m)$ . Also, the  $m$ -tuple  $U + X'$  is either equal to  $0^m$  or it contains exactly two 1's and  $m - 2$  0's. Thus, conditional on the event  $U + X' \neq 0^m$  the  $m$ -tuple  $U + X'$  can take at most  $\binom{m}{2} = m(m-1)/2$  values, implying that  $H(U + X'|E = 0) \leq \log \binom{m}{2}$  with equality if  $X' \sim \text{Unif}(\mathcal{A})$ . Combining the above we get the first inequality.

To prove the second inequality, note that since the indicator  $E$  is a function of  $U + X'$  we have  $H(U|U + X') = H(U|U + X', E) = \Pr(E = 0)H(U|U + X', E = 0) + \Pr(E = 1)H(U|U + X', E = 1)$ . Conditional on  $E = 0$  the  $m$ -tuple  $U + X'$  contains exactly two 1's, so  $U$  can take at most two values and therefore  $H(U|U + X', E = 0) \leq 1$  bit. Since  $\Pr(E = 1) = 1/m$  and  $H(U|U + X', E = 1) \leq \log |\mathcal{A}| = \log(m)$  the second inequality of the lemma follows. The case of equality can easily be checked.  $\square$

We next prove that

$$C^* = \inf_{X' \in \mathcal{A}} \sup_{X \in \mathcal{A}} I(X; X + X') = \left(1 - \frac{1}{m}\right) \log \left(\frac{m}{2}\right) \quad (40)$$

and that  $C^*$  is achieved by  $X^* = U$ . This will prove (20). To get an upper bound for  $C^*$  we substitute  $X' = U$  resulting in  $C^* \leq \sup_{X \in \mathcal{A}} I(X; X + U)$ . Writing  $I(X; X + U) = H(X + U) - H(U) = H(X + U) - \log(m)$ , we have by the first inequality in Lemma 2

$$C^* \leq H_B(1/m) + \frac{m-1}{m} \log \binom{m}{2} - \log(m) \quad (41)$$

$$= (1 - 1/m) \log(m/2). \quad (42)$$

To get a lower bound for  $C^*$  we substitute  $X = U$  resulting in  $C^* \geq \inf_{X' \in \mathcal{A}} I(U; U + X')$ . Writing  $I(X; X + U) = H(U) - H(U|U + X') = \log(m) - H(U|U + X')$ ,

$$C^* \geq \log(m) - \frac{m-1}{m} - \frac{\log(m)}{m}$$

i.e.,  $C^* \geq (1 - 1/m) \log(m/2)$ , which together with (42) implies (40) as desired.

## D. Proof of (21)

Let  $B$  be the indicator of the event that  $X^*$  is equal to one of the  $m'$  members of  $\mathcal{A}$  which have a “1” in a noisy coordinate. Note that  $B$  is a function of  $X^*$ , and it is also a deterministic function of  $X^* + N$ . It follows that

$$I(X^*; X^* + N) = I(B, X^*; X^* + N) \quad (43)$$

$$= I(B; X^* + N) + I(X^*; X^* + N|B) \quad (44)$$

$$= H(B) + \Pr(B = 0)I(X^*; X^* + N|B = 0) + \Pr(B = 1)I(X^*; X^* + N|B = 1). \quad (45)$$

Now  $\Pr(B = 1) = 1 - \alpha$  since  $X^*$  is uniform on  $\mathcal{A}$ , and therefore  $H(B) = H_B(\alpha)$ . Also, since the noise  $N$  concentrates on  $m'$  components of the input  $m$ -tuple, we have  $I(X^*; X^* + N|B = 0) = \log(\alpha m)$  and  $I(X^*; X^* + N|B = 1) = 0$ , and (21) follows.

## E. Absolute $r$ -th Power Under Independent Summation

**Lemma 3** *If  $X$  and  $Z$  are statistically independent variables such that  $E|X|^r \leq A$  and  $E|Z|^r \leq A$ , then  $E|X + Z|^r \leq 2^r A$ .*

*Proof:*

$$E|X + Z|^r \leq E(|X| + |Z|)^r \quad (46)$$

$$= E(|X|^r)^{1/r} + |Z|^r \quad (47)$$

$$= E\{E\{|X|^r\}^{1/r} + |Z|^r \mid Z = z\} \quad (48)$$

$$\leq E\{(|E\{|X|^r \mid Z = z\}|)^{1/r} + |Z|^r\} \quad (49)$$

$$= E\{(|E\{|X|^r\}|)^{1/r} + |Z|^r\} \quad (50)$$

$$\leq E\{(A^{1/r} + |Z|^r)^r\} \quad (51)$$

$$\leq (A^{1/r} + A^{1/r})^r \quad (52)$$

$$= 2^r A \quad (53)$$

where the second inequality follows from Jensen's inequality since  $Z$  is fixed in the inner expectation and the function  $f(t) = (t^{1/r} + 1)^r$  is convex  $\cap$  for positive  $t$  and  $r \geq 1$ ; the third inequality follows since  $E|X|^r \leq A$ ; and the last inequality follows by repeating the same steps with respect to  $Z$ .  $\square$

## Acknowledgement

We thank the referees for helpful comments and suggestions.

## References

- [1] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, New York, 1991.
- [2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, N.Y., 1968.
- [4] A. Lapidoth and P. Narayan, *Reliable communication under channel uncertainty*, IEEE Trans. Information Theory **IT-44** (Oct. 1998), 2148–2177.
- [5] L. Lastras and T. Berger, *All sources are nearly successively refinable*, IEEE Trans. Information Theory **IT-47** (March 2001), 918–926.
- [6] E.E. Majani and H. Rumsey, *Two results on binary-input discrete memoryless channels*, Proc. of Int. Symp. Inf. Th. (ISIT91), Budapest, Hungary, June 1991, p. 104.
- [7] B. Schein and M. Trott, *Sub-optimal power spectra for colored Gaussian channels*, Proceedings, 1997 IEEE International Symposium on Information Theory, Ulm, Germany, June/July , 1997, p. 340.
- [8] N. Shulman and M. Feder, *The uniform distribution as a universal prior*, IEEE Trans. Information Theory (2002, submitted).
- [9] R. Zamir, *The rate loss in the Wyner-Ziv problem*, IEEE Trans. Information Theory (Nov. 1996), 2073–2084.
- [10] ———, “The half a bit loss of robust source/channel codebooks,” in *Proc. of the Info. Theory Workshop*, (Bangalore, India), Oct. 2002, p. 123.
- [11] R. Zamir and M. Feder, *On universal quantization by randomized uniform / lattice quantizer*, IEEE Trans. Information Theory (March 1992), 428–436.
- [12] ———, *Information rates of pre/post filtered dithered quantizers*, IEEE Trans. Information Theory (September 1996), 1340–1353.
- [13] J. Ziv, *On universal quantization*, IEEE Trans. Information Theory **IT-31** (May 1985), 344–347.

## **Ram Zamir: a Biography**

Ram Zamir was born in Ramat-Gan, Israel in 1961. He received the B.Sc., M.Sc. (summa cum laude) and D.Sc. (with distinction) degrees from Tel-Aviv University, Israel, in 1983, 1991, and 1994, respectively, all in electrical engineering. In the years 1994 - 1996 he spent a post-doctoral period at Cornell University, Ithaca, NY, and at the University of California, Santa Barbara. Since 1996 he has been with the department of Elect. Eng. - Systems at Tel Aviv University. Dr. Zamir has been consulting in the areas of radar and communications, mainly in developing algorithms and in the design of signals, and has been teaching information theory, communications systems and communications circuits at Tel Aviv University.

Dr. Zamir received the Israel Ministry of Communications Award in 1993, and the Wolfson Post-Doctoral Research Award in 1994, and visited the Technical University of Budapest under a joint program of the Israeli and Hungarian academies of science in summer 1995. His research interests include information theory, communication and remote sensing systems, and signal processing. He now serves as an Associate Editor for Source Coding in the IEEE tr. on Information Theory.

## **Uri Erez: a Biography**

Uri Erez was born in Tel-Aviv, Israel, on October 27, 1971. He received the B.Sc. degree in mathematics and physics and the M.Sc. and Ph.D. degrees in electrical engineering from Tel-Aviv University in 1996, 1999 and 2003, respectively. He is currently a Postdoctoral Associate at the Signals, Information and Algorithms Laboratory at MIT. His research interests include information theory and digital communication.