[10] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 581–586, Mar. 2004.
[11] M. Mandell and R. J. McEliece, "Some properties of memoryless multiterminal interference channels," in *Proc. 1991 IEEE Int. Symp. Information Theory*, Budapest, Hungary, June 1991, p. 212.
[12] H. Sato, "Two-user communication channels," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 295–304, May 1977.
[13] ——, "On degraded Gaussian two-user channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 637–640, Sept. 1978.
[14] ——, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 786–788, Nov. 1981.
[15] E. C. van der Meulen, *Some Reflections on the Interference Channel*. Boston, MA: Kluwer, 1994, pp. 409–421.
[16] F. M. J. Willems, "The maximal-error and average-error capacity regions of the broadcast channel are identical: A direct proof," *Probl. Control Inform. Theory*, vol. 19, no. 4, pp. 339–347, 1990.

The Uniform Distribution as a Universal Prior

Nadav Shulman and Meir Feder, *Fellow, IEEE*

***Abstract*—In this correspondence, we discuss the properties of the uniform prior as a universal prior, i.e., a prior that induces a mutual information that is simultaneously close to the capacity for all channels. We determine bounds on the amount of the mutual information loss in using the uniform prior instead of the capacity-achieving prior. Specifically, for the class of binary input channels with any output alphabet, we show that the Z-channel has the minimal mutual information with uniform prior, out of all channels with a given capacity. From this, we conclude that the degradation of the mutual information with respect to the capacity is at most 0.011 bit, and as was shown previously, at most 6%. A related result is that the capacity-achieving prior, for any channel, is not far from uniform. Some of these results are extended to channels with nonbinary input.**

***Index Terms*—Uniform distribution, universal prior, Z-channel.**

I. Introduction

A transmitter that wishes to communicate over an unknown channel faces several problems. First, the transmitter does not know the rate at which a reliable communication can be maintained since it does not know the channel capacity. Second, the encoder may not know how to design a good code, tuned to the channel, as the code may depend on the unknown optimal capacity-achieving prior. In this correspondence, we consider the second problem. We show that the uniform prior distribution has some universal properties, so that the degradation in using it universally, instead of the optimal prior is minimal in many cases. A related observation is that the capacity-achieving prior cannot be too far from uniform. This implies that codes based on a uniform prior assumption, such as linear codes, work well on a large class of channels, as discussed, e.g., in [5].

We show in this correspondence the following novel results.

- In the class of all channels with the same input alphabet and the same capacity, the uniform distribution maximizes the minimal mutual information.
- Among all binary input channels with a given capacity, the mutual information induced by the uniform prior is minimal at the Z-channel.
- As a result, the mutual information induced by the uniform distribution is never less than about 0.011 bit than the capacity. In relative terms, as was also shown before, it is never less than about 6% of the channel capacity.
- The capacity-achieving prior, for any channel, can allocate at most a probability mass $1 - e^{-1}$ to any input value.
- For nonbinary input channels, we conjecture that a generalized Z-channel has the maximal degradation in using the uniform prior. In any case, we show an upper bound on that degradation.

Much of this correspondence extends and proves conjectures discussed in [6], [1], [3], and [4]. Specifically, in [4] it was shown that the capacity-achieving prior for binary-input channels allocate at most $1 - e^{-1}$ to each symbol, but it only conjectures the extension for larger input alphabet. Also, in [4] it is shown that the degradation in using the uniform prior is at most 6%, but the extremal properties of the Z-channel, and thus the fact that maximal degradation is at most 0.011 bit, is not shown.

The correspondence is organized as follows. In the next section, we show the optimal max-min properties of the uniform prior. The results for binary-input channels are shown in Section III, and the extension to nonbinary-input alphabet is discussed in Section IV.

II. The Universal Prior

In this section, we explicitly investigate the *universal* prior, i.e., a single predetermined prior that can be used (and to design codes based on it) for all channels, so that the loss in using it instead of the optimal prior tuned to channel is monitored. While there may be several criteria to measure the goodness of that prior, we use a max-min approach—choose the prior distribution P, so that the achieved rate (measured by the mutual information it induces) as compared with the channel capacity, for the worst possible channel, is maximized. Specifically, one option is to look for P that attains

$$\alpha = \max_{P} \inf_{W} \frac{I(P;W)}{C(W)} \tag{1}$$

where $C(W) = \max_{P'} I(P';W)$ is the capacity of the channel W, and the infimum is taken over the class of the possible channels. An alternative criterion is to look for P that attains

$$\delta = \min_{P} \sup_{W} \left[C(W) - I(P;W)\right]. \tag{2}$$

Another case is where the universal prior is designed to work well for a class of channels that have the same capacity, $C > 0$. A modification of criterion (1) for this case is to find P (which may depend on C) that attains

$$\beta(C) = \max_{P} \inf_{\{W:C(W)=C\}} \frac{I(P;W)}{C}. \tag{3}$$

In the sequel, the class of channels we consider is the set of all discrete input memoryless channels with a given input alphabet size

Manuscript received August 16, 2002; revised November 24, 2003.

The authors are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel (e-mail: meir@eng.tau.ac.il).

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2004.828152

$A < \infty$ and any (even nondiscrete) output alphabet.[1] When the class of channels is all possible channels with input alphabet of size A, we denote the max-min ratios of (1), (2), and (3) by α_A, δ_A, and $\beta_A(C)$, respectively.

Theorem 1: The uniform prior, over the alphabet of size A, attains α_A, δ_A, and $\beta_A(C)$ for all C.

Proof: The theorem is proved by a simple symmetry argument. Assume that a distribution P attains α_A, i.e., solves the max-min problem of (1). Let P_π be a permutation of P, that is, P_π is the same probability distribution as P, defined over a permutation of the input symbols, for some permutation $\pi \in S_A$ where S_A is the set of all $A!$ permutations of the A symbols. Since all channels with the same input alphabet can be considered, P_π also attains (1) for any $\pi \in S_A$. Thus, for any channel W and any permutation π

$$\alpha_A \le \frac{I(P_\pi; W)}{C(W)}$$

and so, for any W, using the convexity of the mutual information with respect to the input distribution

$$\alpha_A C(W) \le \sum_{\pi \in S_A} \frac{I(P_\pi; W)}{|S_A|} \le I\left(\sum_{\pi \in S_A} \frac{P_\pi}{|S_A|}; W\right).$$

But $\sum_{\pi \in S_A} \frac{P_\pi}{|S_A|}$ is the uniform distribution, regardless of P. Thus, the uniform distribution attains (1).

The same proof holds for δ_A and $\beta_A(C)$. □

Corollary 1: $\alpha_A = \inf_C \beta_A(C)$ and $\delta_A = \max_C C(1 - \beta_A(C))$.

The following lemma shows that the relative degradation in using the uniform prior is smaller as the channel capacity grows.

Lemma 1: $\beta(C)$ is monotonically increasing with C.

Proof: Given a channel $W(y|x)$, define a new channel $W'(y|x)$ with an additional output symbol E as follows:

$$W'(y|x) = \begin{cases} (1-\epsilon)W(y|x), & \text{if } y \ne E \\ \epsilon, & \text{if } y = E. \end{cases}$$

W' is an *erasure* version of W, and for any input prior P we have $I(P; W') = (1-\epsilon)I(P; W)$ which implies $C(W') = (1-\epsilon)C(W)$. Now for a given C, assume W leads to $\beta(C)$, then, for any $C' < C$ we take ϵ such that $C(W') = C'$ and so we have

$$\beta(C) = \frac{I(U; W)}{C(W)} = \frac{I(U; W')}{C(W')} \ge \beta(C')$$

where U is the uniform prior □

Corollary 2 : $\alpha_A = \lim_{C \to 0} \beta(C)$.

Later in the correspondence, we investigate the values of α_A, δ_A, and $\beta_A(C)$. We begin with the binary-input case, and show that the degradation is small; α_2 is about 0.94 and δ_2 is about 0.011 bit. Unfortunately, for large A, the degradation can be significant. If one considers the A-ary input channel with binary output, where $W(1|1) = W(2|2) = 1, W(1|i) = W(2|i) = 1/2$ for $i = 3, \ldots, A$, its capacity is 1, but $I(U; W) = \frac{2}{A}$. Thus, $\alpha_A \le \frac{2}{A}$.

III. The Binary Input Case

We begin with some properties of the mutual information for binary-input channels. Let $q_1(y)$ and $q_0(y)$ be two probability distributions on $\mathcal{Y}$. Considering $\mathcal{Y}$ as the output alphabet, these two distributions define a binary-input channel Q where

$$Q(y|1) = q_1(y) \quad \text{and} \quad Q(y|0) = q_0(y). \tag{4}$$

We assume that the capacity of the channel Q is greater than zero, i.e., $q_1 \ne q_0$. For an input probability $P(1) = u$ and $P(0) = 1 - u$, the output distribution is

$$q_u(y) = uq_1(y) + (1-u)q_0(y). \tag{5}$$

Define

$$d_0(u) = D(q_0(y) \| q_u(y)) = \sum_{y \in \mathcal{Y}} q_0(y) \log \frac{q_0(y)}{q_u(y)} \tag{6}$$

$$d_1(u) = D(q_1(y) \| q_u(y)) = \sum_{y \in \mathcal{Y}} q_1(y) \log \frac{q_1(y)}{q_u(y)}. \tag{7}$$

These functions are strictly monotonic, $d_1(u)$ decreases and $d_0(u)$ increases, and $d_0(0) = d_1(1) = 0$. The mutual information over the channel Q, implied by the input distribution $P(1) = u$, is

$$I(u; Q) = ud_1(u) + (1-u)d_0(u). \tag{8}$$

It is known that the channel's capacity $C(Q)$ satisfies $C(Q) = d_1(u^*) = d_0(u^*)$, where u^* is the prior that maximizes the mutual information.

The derivatives of $d_1(u)$ and $d_0(u)$ with respect to u are given by

$$d_0'(u) = -\sum_{y \in \mathcal{Y}} q_0(y) \frac{q_1(y) - q_0(y)}{q_u(y)}$$

$$d_1'(u) = -\sum_{y \in \mathcal{Y}} q_1(y) \frac{q_1(y) - q_0(y)}{q_u(y)}$$

which leads to the interesting equalities

$$ud_1'(u) + (1-u)d_0'(u) = 0 \tag{9}$$

and

$$I'(u; Q) = d_1(u) - d_0(u) \tag{10}$$

for any $0 \le u \le 1$. Since $d_1(1) = 0$ we have

$$\begin{aligned} d_1(\tilde{u}) &= -\int_{\tilde{u}}^1 d_1'(u)du \\ &= \int_{\tilde{u}}^1 \frac{1-u}{u} d_0'(u)du \\ &= \int_{\tilde{u}}^1 \frac{1}{u^2} d_0(u)du - \frac{1-\tilde{u}}{\tilde{u}} d_0(\tilde{u}) \end{aligned} \tag{11}$$

where the second equality is due to (9) and the third is obtained by partial integration. Substituting (11) in (8) leads to

$$I(\tilde{u}; Q) = \int_{\tilde{u}}^1 \frac{\tilde{u}}{u^2} d_0(u)du. \tag{12}$$

Utilizing the properties above, we prove the following theorem stating that the capacity-achieving prior, for any binary input channel, cannot be too skewed.

Theorem 2: For any binary-input channel, the capacity-achieving prior assigns a probability greater than e^{-1} for both symbols.

As discussed earlier, this result was already shown in [4]. Our derivation is different and allows (as shown later in Section IV) the extension,

[1]Of course, the class of channels is such that all considered quantities are well defined.

too skewedconjectured in [4], to nonbinary-input alphabet. Our derivation starts with the following technical lemma which is proved in the Appendix .

Lemma 2: Let $f(x)$ be a positive, nonincreasing function such that $\int_0^1 f(x)dx = 1$ and t solves

$$\int_0^t \frac{1}{1-x} f(x)dx = 1.$$

Then $t \leq 1 - e^{-1}$, with equality if and only if $f(x) = 1$ for all $x \in (0,1)$.

Proof of Theorem 2: We assume that the capacity is greater than zero, since at zero capacity corresponding to $q_1(y) = q_0(y)$, any prior leads to the same, zero, mutual information.

Denote the capacity-achieving prior u^*. It satisfies $d_1(u^*) = d_0(u^*)$, and so

$$d_1(0) + \int_0^{u^*} d_1'(u)du = d_0(0) + \int_0^{u^*} d_0'(u)du.$$

Since $d_0(0) = 0$ we have from (9)

$$\int_0^{u^*} \frac{-1}{1-u} d_1'(u)du = d_1(0).$$

Suppose first that $d_1(0) < \infty$. Using Lemma 2 with $f(\cdot) = -d_1'(\cdot)/d_1(0)$, leads to $u^* \leq 1 - e^{-1}$. Equality holds iff $-d_1'(u)/d_1(0) = 1$ for $0 \leq u \leq 1$, but since $d_1'(1) = 0$, this implies that the channel capacity is zero. So for the binary channels with nonzero capacity, $u^* < 1 - e^{-1}$.

In case $d_1(0) = \infty$, pick some $0 < \epsilon < u^*$, and define

$$\tilde{d}_1(u) = \begin{cases} d_1(u), & \text{for } u \geq \epsilon \\ d_1(\epsilon) + (u-\epsilon)d_1'(\epsilon), & \text{for } u < \epsilon. \end{cases}$$

Define $\tilde{d}_0(u)$ by $\tilde{d}_0(0) = 0$ and its derivative $\tilde{d}_0'(u) = \frac{-u}{1-u}\tilde{d}_1'(u)$. Let u^+ solve

$$\int_0^{u^+} \tilde{d}_0'(u) - \tilde{d}_1'(u)du = \int_0^{u^+} \frac{-1}{1-u}\tilde{d}_1'(u)du = \tilde{d}_1(0).$$

Since for $u < \epsilon$ we have $\tilde{d}_0'(u) < d_0'(u)$, it follows that $u^+ \geq u^*$. Using this and Lemma 2 for u^+ leads to $u^* < u^+ < 1 - e^{-1}$.

We showed, then, that $u^* < 1 - e^{-1}$. Due to symmetry we also have $1 - u^* < 1 - e^{-1}$, i.e., $u^* > e^{-1}$. □

In view of our derivation, this result has an interesting geometrical interpretation. In solving for the capacity-achieving prior, we actually search for a distribution q satisfying $D(q_1\|q) = D(q_0\|q)$, where q_1 and q_0 are the output probability given each input symbol, and q is the output probability induced by the capacity-achieving prior. This q can be interpreted as the average of q_1 and q_0 in the sense of the divergence. Our result indicates that this informational average $q = uq_1 + (1-u)q_0$ is not so far from the arithmetic, Euclidean average $(q_1 + q_0)/2$.

Next we investigate the maximal degradation in the mutual information induced by the uniform prior. It turns out that the maximal degradation happens at the Z-channel due to its extremal behavior.

The Z-channel, Z_p, is a binary-input binary-output channel which is defined by a single parameter p, $P(0|0) = 1$, and $P(1|1) = p$, as in Fig. 1.

Denote the input prior $P(1) = u$. For the Z-channel we choose to denote the divergences $d_1(u)$ and $d_0(u)$, defined above, by

$$z_1(u) = D(\{1-p, p\}\|\{1-up, up\}) \tag{13}$$

$$z_0(u) = D(\{1,0\}\|\{1-up, up\}) = \log\frac{1}{1-up}. \tag{14}$$

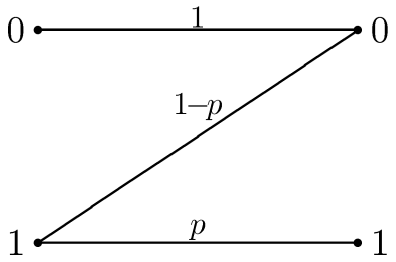

Fig. 1. The Z-channel.

The mutual information over the Z-channel satisfies

$$\begin{aligned} I(u;Z_p) = I_Z(u) &= uz_1(u) + (1-u)z_0(u) \\ &= h(up) - uh(p) \end{aligned}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy. The capacity is denoted $C(Z_p) = \max_u I(u;Z_p)$. The maximizing u is $u^* = \frac{q^{q/p}}{1+pq^{q/p}}$ where $q = 1-p$, see [3].

The following two lemmas demonstrate the extremal behavior of the divergence $z_0(u)$ and the mutual information $I(u;Z_p)$ as functions of the input prior u.

Lemma 3: For any binary-input channel, with divergence $d_0(\cdot)$, and any Z-channel, if for some $0 < \tilde{u} \leq 1$

$$z_0(\tilde{u}) = d_0(\tilde{u})$$

then

$$z_0(u) \geq d_0(u)$$

for any $0 \leq u < \tilde{u}$.

The proof of this lemma, and an interesting corollary to it are given in the Appendix .

Lemma 4: Let $I_Z(u) = I(u;Z_p)$ and $I(u) = I(u;Q)$ be the mutual informations associated with the Z-channel and a general binary input channel Q, respectively, both with input prior satisfying $P(1) = u$. If $I(\tilde{u}) = I_Z(\tilde{u})$ for some $0 < \tilde{u} \leq 1$ then $I(u) \leq I_Z(u)$ for $u \leq \tilde{u}$.

Again, the proof is given in the Appendix . We are now ready for the main result.

Theorem 3: Among all binary-input channels with the same capacity, the Z-channel has the smallest mutual information given uniform input distribution.

Proof: An equivalent statement of the theorem is that among all binary-input channels for which the mutual information induced by a uniform input distribution is the same, the Z-channel has the maximum capacity. We show this statement.

Given any binary-input channel Q, without loss of generality, assume that the capacity achieving prior for Q satisfies $P(1) \leq \frac{1}{2}$. Define a Z-channel Z_p such that it has the same mutual information for uniform input distribution. By Lemma 4, $I(u;Z_p) \geq I(u;Q)$ for $u \leq 1/2$. Hence,

$$C(Z_p) = \max_{0\leq u\leq 1/2} I(u;Z_p) \geq \max_{0\leq u\leq 1/2} I(u;Q) = C(Q)$$

which completes the proof. □

Theorem 3 implies that $\beta_2(C)$ is the ratio between the mutual information of the Z-channel with the uniform prior, and its capacity. This rate degradation is depicted in Fig. 2.

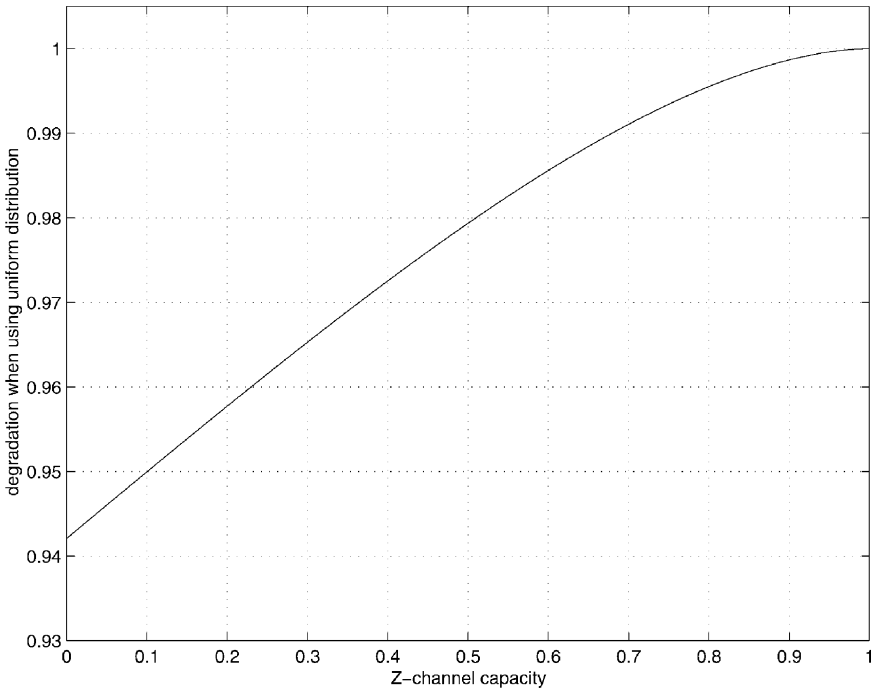

Fig. 2. $\beta_2(C)$—the degradation of the rate, when using the uniform prior in the Z-channel.

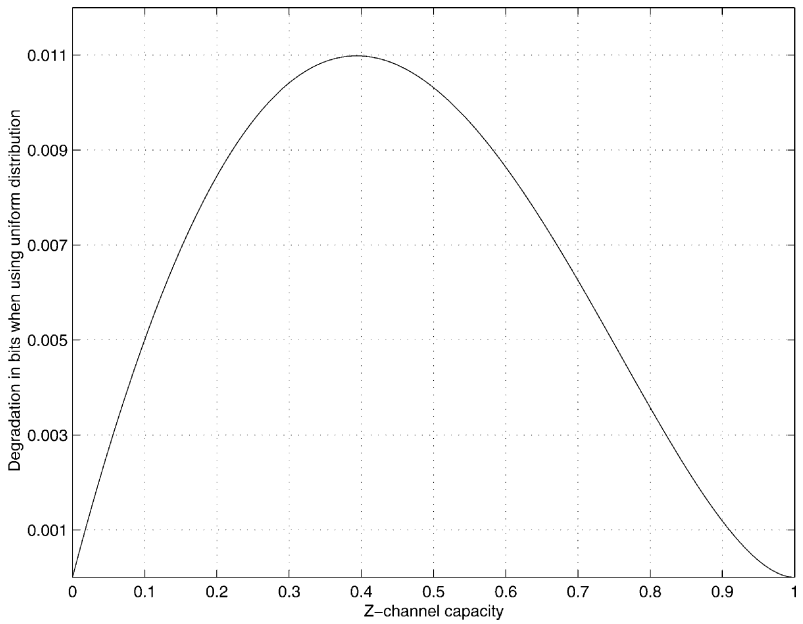

Fig. 3. $C(1-\beta_2(C))$—the degradation, in bits, when using the uniform prior in the Z-channel.

The maximal degradation occurs when C (and so p) goes to 0. Golomb showed [3] that the capacity-achieving prior of the Z-channel, for $p \to 0$, is $P(1) = u \to e^{-1}$. Thus, the value of this maximal degradation is

$$\alpha_2 = \lim_{p \to 0} \frac{I(1/2; Z_p)}{I(e^{-1}; Z_p)} = \frac{e}{2 \log_2 e} = 0.9420847\ldots$$

which is less than 6% loss.

As noted, $\beta_2(C)$ provides a refined bound on the degradation, as a function of the capacity. For example, for channels with $C = 1/2$, the degradation when using the uniform prior is less than $\sim 2\%$.

In terms of absolute values[2]

$$\delta_2 = \min_C C(1-\beta_2(C)) \approx 0.011.$$

Hence, the uniform distribution leads to a mutual information that is less than the capacity by at most ~ 0.011 bit. The absolute maximal

[2]To evaluate $\boldsymbol{\delta}_2$, one should solve the equation

$$\frac{1}{2}\log\frac{q}{1+q} = \frac{\log q}{(1-q)q^{-q/1-q} + (1-q)^2}.$$

We used Matlab.

degradation as function of the channel capacity $C(1-\beta(C))$ is plotted in Fig. 3.

Finally, we note that Lemma 4 leads to another result: among all binary-input channels with the same capacity, the Z-channel has a capacity-achieving prior which allocates the minimal mass, $1/e$ on the symbol "1." This proves Theorem 2, and the result in [4], in yet another way.

IV. Nonbinary Input Alphabet

This section extends some of the results, previously shown for binary input channels, to general input alphabets.

Theorem 4: The capacity-achieving prior of any memoryless, discrete-input channel is smaller than $1-e^{-1}$ for all symbols.

Proof: Given a channel $W(y|x)$ and capacity-achieving input distribution $P^*(x)$, we wish to show that $P^*(x') < 1-e^{-1}$, for any x'. Denote by $q(y)$ the output distribution. Then, $D(W(y|x)\|q(y)) = C(W)$ for $P^*(x) > 0$. Assume $P^*(x') > 0$. Create the following binary input channel:

$$Q(y|0) = W(y|x')$$
$$Q(y|1) = \frac{1}{1-P^*(x')}\sum_{x\neq x'} P^*(x)W(y|x).$$

Using $P(0) = P^*(x'), P(1) = 1-P^*(x')$ as the input distribution to the channel Q leads to the same $q(y)$ as the output distribution. Due to the convexity of the divergence we have

$$D(Q(y|0)\|q(y)) = C(W) \geq D(Q(y|1)\|q(y)).$$

Hence, the capacity-achieving prior for this binary-input channel gives greater or equal probability for the symbol 0 than $P^*(x')$. Thus, using Theorem 2 we have $P^*(x') < 1-e^{-1}$. □

Another bound on the capacity-achieving prior is given in the following lemma.

Lemma 5: Consider a memoryless channel with a discrete-input alphabet and a capacity C. Let P^* be the capacity-achieving prior. Then, for each input symbol x we have

$$P^*(x) \leq 2^{-C}.$$

The proof is given in the Appendix .

Generalized Z-channel: This channel is defined by three parameters: A is the input alphabet size, $1 < A_e \leq A$ is the output size (and the "effective" input size), and $0 \leq p \leq 1$ as follows:

$$Z_p^{A_e}(y|x) = \begin{cases} 1, & \text{if } x < A_e \text{ and } y = x \\ 0, & \text{if } x < A_e \text{ and } y \neq x \\ p, & \text{if } x = A_e \text{ and } y = x \\ (A_e-1)^{-1}(1-p), & \text{if } x = A_e \text{ and } y \neq x \\ A_e^{-1}\sum\limits_{x'=1}^{A_e} Z_p^{A_e}(y|x'), & \text{if } x > A_e. \end{cases} \quad (15)$$

Note that the input symbols greater than A_e are not used to achieve the capacity, and they have no contribution to the mutual information when uniform distribution on the input symbols is used. Fig. 4 shows a generalized Z-channel with $A = 4, A_e = 3$.

Due to the symmetry between the $A_e - 1$ symbols, the capacity-achieving input distribution is of the form $P(x = A_e) = u$, and $P(x) = \frac{1-u}{A_e-1}$ for $x < A_e$

$$I(u; Z_p^{A_e}) = h(up) - uh(p) + (1-u)\log(A_e - 1).$$

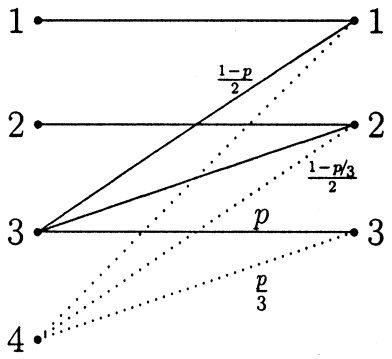

Fig. 4. Generalized Z-channel $A = 4, A_e = 3$.

Following [3], the capacity-achieving prior is

$$u^* = \frac{q^{q/p}}{(A_e-1)^{1/p} + pq^{q/p}}$$

where $q = 1-p$. For $p \to 0$, in general $u^* \to 0$, unlike the binary case ($A_e = 2$) for which $u \to e^{-1}$. This indicates that for nonbinary input channels shows there is no "minimum probability" for each of the (used) input symbols.

Substituting u^*, we get the capacity of the generalized Z-channel

$$C(Z_p^{A_e}) = \log\left(A_e - 1 + \left(\frac{pq}{A_e-1}\right)^{q/p}\right). \quad (16)$$

As seen easily from the definition of the channel, for each value of $C \leq \log A$, choosing A_e to be the minimal value such that $C < \log A_e$, there is a value for p that will lead to a generalized Z-channel with capacity C.

The mutual information over this channel, given a uniform distribution U over all A input symbols, is

$$I(U; Z_p^{A_e}) = A_e A^{-1}\left(h(A_e^{-1}p) - A_e^{-1}h(p) + (1-A_e^{-1})\log(A_e-1)\right).$$

Conjecture 1: For a memoryless channel with input alphabet size A, $\beta_A(C)$ is achieved by the generalized Z-channel.

If the conjecture is true, for a given capacity C, $\beta_A(C) \propto \frac{1}{A}$ for A's greater or equal to 2^C. And in particular

$$\alpha_A = \frac{2}{A}\alpha_2 = \frac{e}{A\log_2 e}.$$

The generalized Z-channel provides an upper bound on $\beta_A(C)$, conjectured to be the actual value. In the sequel, we shall derive a lower bound on $\beta_A(C)$.

The following lemma bounds the ratio between the mutual informations associated with two input probabilities in terms of of the ratio between these probabilities. The proof for this lemma is in the Appendix .

Lemma 6: Given a channel W and two prior distributions P and P', then

$$\frac{I(P;W)}{I(P';W)} \geq \min_x \frac{P(x)}{P'(x)}.$$

Combining Lemma 6 , Theorem 4 ,and Lemma 5 we obtain for a memoryless channel with discrete-input alphabet of size A

$$\beta_A(C) \geq \frac{1}{A\ \min(2^{-C}, 1-e^{-1})}. \quad (17)$$

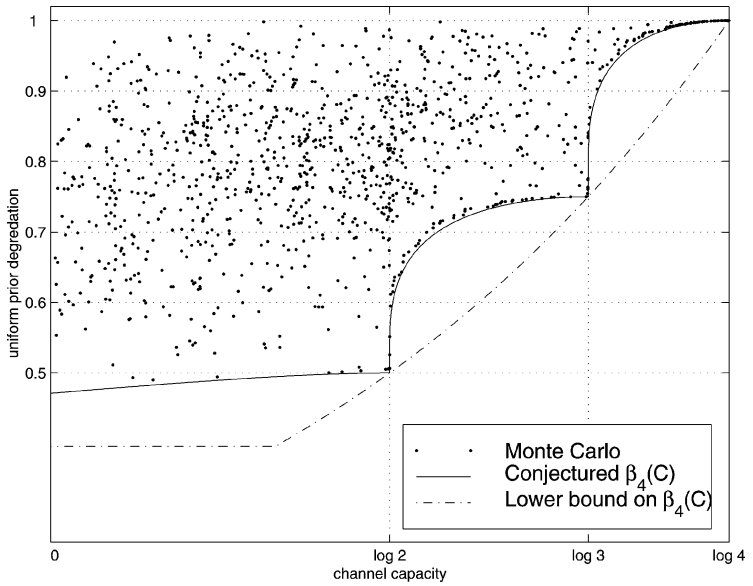

Fig. 5. The simulation results for $A = 4$.

Corollary 3: For a memoryless channel with input alphabet size of A, if K is an integer such that $2 \leq K \leq A$, then

$$\beta(\log K) = \frac{K}{A}. \tag{18}$$

We finally show simulation results that provide some insight on the value of $\beta_A(C)$. In this simulation, depicted in Fig. 5, channels with input and output alphabet of size 4 were selected randomly.[3] For each chosen channel, we calculated the capacity and the mutual information associated with the uniform prior. The simulation contains more than a million points. The fact that none of the randomly selected channels supplies a counterexample for our conjecture on $\beta_A(C)$, and that many examples lie on the border implied by the conjecture, supports indeed the conjecture validity. Also, note that for this case of $A = 4$, the conjecture leads to $\delta_4 = 0.5$, while the lower bound we proved on leads to $\delta_4 \leq 0.512$. So in terms of the difference in bits, the degradation in using the uniform prior is estimated quite well. Unfortunately, it is not as low as 0.011 bit, shown for binary-input channels.

V. Summary

This correspondence shows the optimal property of the uniform distribution that it attains, universally, over all memoryless channels, the maximal mutual information. It was further shown that the degradation in using it, instead of the true capacity-achieving prior, is worst for the Z-channel, and the amount of that degradation, for binary-input channels, is quite small.

The optimal properties of the uniform prior were largely observed, see, e.g., [5]. An additional interesting result [2] states that the expurgated error exponent is maximized by the uniform prior for all binary-input channels. All these results indicate that codes designed, implicitly or explicitly, with a uniform prior can work well for a large class of discrete-input channels.

For power-constrained channels, the role of the uniform prior is played by the Gaussian prior, which essentially can be regarded as a uniform distribution over a large-dimension ball, whose radius represents the power constraint. A recent result [7] analyzes the loss in using the Gaussian prior instead of the capacity-achieving prior and, indeed, shows that the loss is small. However, for this and other similar cases, the problem of finding an *optimal* min-max prior, analogous to the result for discrete memoryless channel (DMC) shown in this correspondence, is still open.

[3]We used a "smart" lottery scheme in order to have a good coverage of the different types of channels

Appendix

Proof of Lemma 2: Assume that out of all functions that satisfy the monotonicity and integrability conditions above $f(x)$ leads to the maximal possible value of t, denoted t^*. Without loss of generality, we can assume $f(x) = c$ for $x > t^*$ where c is a constant chosen to satisfy $\int_0^1 f(x)dx = 1$ and will be no greater than $f(t^*)$. Define a new function $g(x)$ that satisfies the conditions, as follows:

$$g(x) = \begin{cases} f(x) + f(x + \frac{v}{2}) - c, & \text{for } x \leq \frac{v}{2} \\ c, & \text{for } x > \frac{v}{2} \end{cases}$$

where $v = \inf\{x | f(x) = c\}$.

Since $v \leq t^*$ and $\int_0^{t^*} \frac{f(x)}{1-x} dx = 1$ we have

$$\int_0^{t^*} \frac{g(x)}{1-x} dx = 1 + \int_0^{\frac{v}{2}} \left(f\left(x + \frac{v}{2}\right) - c \right) \left(\frac{1}{1-x} - \frac{1}{1-(x+\frac{v}{2})} \right) dx \leq 1.$$

But, if it is strictly smaller than one it contradicts the maximality of t^*. Equality to one leads to $v = 0$, which means that $g(x) = f(x) = 1$ for $x \in (0, 1)$. Solving for t with $f(x) = 1$ leads to $t^* = 1 - e^{-1}$. □

Proof of Lemma 3: It is sufficient to show that if $z_0(\tilde{u}) = d_0(\tilde{u})$, then $z_0'(\tilde{u}) \leq d_0'(\tilde{u})$ with equality if and only if $z_0(\cdot) = d_0(\cdot)$. Now, $z_0(\tilde{u}) = d_0(\tilde{u})$ means

$$\log \frac{1}{1-\tilde{u}p} = \sum_{y \in \mathcal{Y}} q_0(y) \log \frac{q_0(y)}{\tilde{u} q_1(y) + (1-\tilde{u}) q_0(y)}. \tag{19}$$

Using Jensen inequality and taking the exponent of both sides

$$\frac{1}{1-\tilde{u}p} \leq \sum_{y \in \mathcal{Y}} q_0(y) \frac{q_0(y)}{\tilde{u} q_1(y) + (1-\tilde{u}) q_0(y)}. \tag{20}$$

Since $\log(\cdot)$ is strictly convex, equality in the last equation will hold only if q_1 and q_0 fit the Z-channel (actually it might fit an "effective Z-channel" where several output symbols have the same distribution

as the "0" symbol in the Z-channel, and a similar situation for the "1" symbol). Now

$$\begin{aligned}\tilde{u}z_0'(\tilde{u}) &= \frac{\tilde{u}p}{1-\tilde{u}p} = \frac{1}{1-\tilde{u}p} - 1 \\ &\le -1 + \sum_{y\in\mathcal{Y}} q_0(y)\frac{q_0(y)}{\tilde{u}q_1(y)+(1-\tilde{u})q_0(y)} \\ &= \sum_{y\in\mathcal{Y}} q_0(y)\frac{\tilde{u}q_0(y)-\tilde{u}q_1(y)}{\tilde{u}q_1(y)+(1-\tilde{u})q_0(y)} = \tilde{u}d_0'(\tilde{u})\end{aligned}$$

which completes the proof. □

Lemma 3 leads to the following interesting inequality for the divergence.

Corollary 4: For any distributions $p_1(y), p_0(y)$, let p be such that

$$D(p_0\|p_1) = \log\frac{1}{1-p}$$

then

$$D(p_0\|up_1+(1-u)p_0) \le \log\frac{1}{1-up}$$

for all $0 \le u \le 1$.

This corollary is stronger than the straightforward usage of the log-function monotonicity which is the above inequality with $p = 1$.

Proof of Lemma 4: Using (12)

$$0 = I_Z(\tilde{u}) - I(\tilde{u}) = \int_{\tilde{u}}^{1}\frac{\tilde{u}}{u^2}[z_0(u)-d_0(u)]du$$

where d_1, d_0, z_1 and z_0 were defined earlier. Hence, we must have that at least at one point u, $\tilde{u} \le u \le 1$ $z_0(u) = d_0(u)$. As a result, due to Lemma 3, $z_0(u) \ge d_0(u)$ for all $u < \tilde{u}$. Using again (12), completes the proof. □

Proof of Lemma 5: We need to prove that

$$C \le \log\frac{1}{max_x P^*(x)} \quad (= H_\infty(P^*)) \ .$$

Given that the channel law is $W(y|x)$, the output distribution given $P^*(x)$ is

$$Q^*(y) = \sum_x P^*(x)W(y|x)$$

then we have, for each x (such that $P^*(x) > 0$) $C = D(W(\cdot\|x)\|Q^*)$. But since for each y we have $Q^*(y) \ge P^*(x)W(y|x)$ we have

$$C = D(W(\cdot\|x)\|Q^*) \le -\log P^*(x)$$

which completes the proof. □

Proof to Lemma 6: Define $r = \max_x \frac{P'(x)}{P(x)}$, then the lemma claims that

$$I(P';W) \le rI(P;W).$$

Since for each x, $P'(x) \le rP(x)$, it is sufficient to show that the function

$$I(P;W) = \sum_{x,y} P(x)W(y|x)\log\frac{W(y|x)}{\sum_x W(y|x)\frac{P(x)}{\sum_{x'}P(x')}}$$

is *increasing* with respect to each $P(x)$ (without the constrain $\sum_x P(x) = 1$).

Define

$$Q(y) = \sum_x W(y|x)\frac{P(x)}{\sum_{x'}P(x')}$$

for each x_0, we have

$$\begin{aligned}\frac{dI(P;W)}{dP(x_0)} &= \sum_y W(y|x_0)\log\frac{W(y|x_0)}{Q(y)} \\ &\quad - \sum_{x,y} P(x)W(y|x)\frac{1}{Q(y)}\frac{dQ(y)}{dP(x_0)} \\ &= D(W(y|x_0)\|Q(y)) \\ &\quad - \sum_{x,y}\frac{P(x)W(y|x)}{Q(y)}\frac{W(y|x_0)-Q(y)}{\sum_{x'}P(x')} \\ &= D(W(y|x_0)\|Q(y)) \ge 0\end{aligned}$$

which completes the proof. □

REFERENCES

[1] E. M. Gabidulin, "Limits for the decoding error probability when linear codes are used in memoryless channel," *Probl. Inform. Transm.*, pp. 43–48, 1967. Translated from *Probl. Pered. Inform.*.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] S. W. Golomb, "The limit behavior of the Z-channel," *IEEE Trans. Inform. Theory*, vol. IT-26, p. 372, May 1980.

[4] E. E. Majani and H. Rumsey, "Two results on binary-input discrete memoryless channels," in *Proc. IEEE Int. Symp. Information Theory*, Budapest, Hungary, 1991, p. 104.

[5] R. J. McEliece, "Are turbo-like codes effective on nonstandard channels?," *IEEE Inform. Theory Soc. Newslett.*, vol. 51, pp. 1–8, Dec. 2001.

[6] R. A. Silverman, "On the binary channels and their cascades," *IRE Trans. Inform. Theory*, vol. PGIT–1, pp. 19–27, Dec., 1955.

[7] R. Zamir and U. Erez, "Gaussian input is not too bad," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1362–1367, June 2002.

A Gaussian Input Is Not Too Bad

Ram Zamir, *Senior Member, IEEE,* and Uri Erez, *Member, IEEE*

***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 (PPM) in the presence of a bursty jammer.**

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

I. INTRODUCTION

Consider additive-noise channels of the form

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

Manuscript received June 14, 2002; revised November 24, 2003. The material in this correspondence was presented in part at the 40th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, October 2002.

R. Zamir is with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel (e-mail: zamir@eng. tau.ac.il)

U. Erez is with the Signals Information and Algorithms Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02137 USA, on leave from the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel (e-mail: uri@allegro.mit.edu).

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2004.828153