Writing on Dirty Paper in the Presence of Difference Set Noise

Aaron S. Cohen*
Division of Applied Mathematics
Brown University
Providence, RI 02912
acohen@dam.brown.edu

Ram Zamir†
Dept. of Electrical Engineering
Tel Aviv University
Ramat Aviv 69978, Israel
zamir@eng.tau.ac.il

Abstract

Costa's celebrated "writing on dirty paper" (WDP) shows that the power-constrained channel $Y = X + S + Z$, with Gaussian Z, has the same capacity as the standard AWGN channel $Y = X + Z$, provided that the "interference" S (no matter how strong it is) is known at the transmitter. While this ability for perfect interference cancellation is very appealing, it relies heavily on the Gaussianity of the (unknown) noise Z. We construct an example of "bad" noise for writing on dirty paper, namely, "difference set noise" (DSN). If the interference S is strong, then DSN limits the WDP capacity to at most 2 bits. At the same time, the capacity of the zero-interference channel $Y = X + Z$, where Z is DSN, grows without bound as the input constraint grows. Thus almost 100% of the available capacity is lost in WDP in the presence of DSN. These results are based on the "entropy amplification property" of DSN, and they shed light on the potentials and limitations of writing on dirty paper.

1 Introduction

In Costa's *writing on dirty paper* (WDP) [1], a channel encounters two independent sources of additive white Gaussian noise. One source, S^n, is known to the transmitter non-causally and will be referred to as *interference*. The other source, Z^n, is not directly known to any part of the system and will be called simply *noise*. The input, x^n, can depend on the interference sequence S^n and nR independent information bits and must satisfy a power constraint, $\sum x_i^2 \leq nP$. Finally, the output is $Y^n = x^n + S^n + Z^n$. Costa showed that the capacity (highest achievable R) of WDP is the same as if there was no interference, i.e., $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$ bits/channel use,[1] where N is the variance of each unknown noise sample Z_i.

An important property of WDP is that the capacity does not depend on the variance of the interference. Thus, the capacity would be the same if there were no interference, or, equivalently, if the interference were also known at the receiver. This "interference

*Supported in part by the National Science Foundation under Grant DMS-0074276.
†Supported in part by the Israel Science Foundation grant #01/65.
[1]Logarithms throughout the paper are base 2.

cancellation property" is particularly interesting as unlike the receiver, the power constrained transmitter cannot simply subtract the interference. Extensions of Costa's result show that the interference cancellation property also holds more generally, in that the capacity of WDP does not depend on the distribution of the interference (i.e., it can be non-Gaussian or even arbitrarily varying) if [2, 3] and only if[2] [5] the noise is Gaussian. Furthermore, the difference between WDP capacity and the zero-interference capacity is at most 1/2 bit/channel use provided only that $E[Z_i^2] \leq P$ (i.e., the noise can be general but must be less powerful than the input) [6]. In the general framework, we refer to the difference between WDP capacity and zero-interference capacity as the *loss* resulting from the interference.

These results may give the impression that having side information (the interference) at the transmitter can be (almost) as efficient as having it at the receiver. In fact, the recent popularity of coding techniques for WDP, the related Gaussian Wyner-Ziv problem, and general algebraic binning for network communications settings heavily rely on small loss. However, in the general setting of channels with side information (SI), there may be a large gap between the transmitter-SI capacity [7, 8] and the receiver-SI capacity (e.g., [9]). Examples of a large gap can be generated using a discrete modulo-additive noise channel with state-dependent noise, i.e., $Y = X + Z_S$, where Z_S is conditionally independent of X given S [10]. Note that in such channels the gap can be positive (and large) even without an input constraint.

The objective of this work is to show that in the basic input-constrained additive state setting, $Y = X + S + Z$, the encoder is sometimes unable to use the SI efficiently if the receiver does not have access to it, i.e., the loss can be very large. In contrast to the conditions for zero or small loss, the noise must be non-Gaussian and more "powerful" than the input. Furthermore, the noise cannot be periodic, since periodicities can be overcome using the lattice-strategies that are successful in the Gaussian case. Thus, we consider strong interference (so that its knowledge at the receiver has the most benefit) and irregular noise (i.e., its distribution is spread in a highly non-uniform and non-periodic manner over the space). In a previous paper [11], we constructed such a noise distribution and demonstrated that the loss in causal WDP (also known as "writing on dirty tape") is at least one-half of the zero-interference capacity.

In this paper, we strengthen the results of [11] in two ways. First, we extend the analysis to non-causal WDP. Second, we find a family of noise distributions for which the WDP capacity is bounded by a small constant (2 bits/channel use), while the zero-interference capacity can be arbitrarily large.

Our main tool in the new results is the identification of structured irregular noise, namely, difference set noise. A difference set is an algebraic notion, describing a subset D of a group G for which $D \cap (D + g)$ has the same number of elements for all $g \neq 0$. We focus on planar difference sets, i.e., those with $|D \cap (D + g)| = 1$, and refer to a uniform distribution over such a set as *difference set noise* (DSN). In Section 3, we identify the "entropy amplification property" of additive DSN. In Section 4, we use this property to separate the effects of the different set noise and the input distribution, which allows us to derive our main results. We begin in Section 2 by precisely defining the problem and giving a general formula for WDP capacity with strong interference.

[2]For a large class of transmitter strategies that includes nested lattices [3] and quantization index modulation [4].

2 Capacity with Strong Interference

We consider a discrete version of WDP in which addition is done modulo some given integer L, with results in the alphabet $\mathcal{A}_L = \{1, \ldots, L\}$.[3] The interference S^n and the noise Z^n are both independent, identically distributed (i.i.d.) sequences with respective marginal distributions, P_S and P_Z. We focus on the case of *strong interference* in which the distribution P_S is uniform on $\mathcal{A}_L$. Instead of a power constraint, we require that each input symbol x_i be drawn from some constraint set $\mathcal{C} \subset \mathcal{A}_L$.

A WDP system consists of two components. The first is a rate-R blocklength-n encoder,

$$f_n : \{1, \ldots, 2^{nR}\} \times \mathcal{A}_L^n \mapsto \mathcal{C}^n, \tag{1}$$

which takes a message M consisting of nR bits and the interference sequence S^n and creates the constrained input sequence as $X^n = f_n(M, S^n)$. The other component is a decoder,

$$g_n : \mathcal{A}_L^n \mapsto \{1, \ldots, 2^{nR}\}, \tag{2}$$

which estimates the message from the output sequence, $\hat{M} = g_n(Y^n)$. The probability of error is measured by

$$P_e = \Pr\left\{g_n\left(f_n(M, S^n) + S^n + Z^n\right) \neq M\right\}, \tag{3}$$

where the message M is uniformly distributed on $\{1, \ldots, 2^{nR}\}$ and is independent of S^n and Z^n. A rate R is achievable if there exists a sequence of rate-R encoders and decoders such that the probability of error can be made as small as desired. The capacity is the supremum of all achievable rates and is written C_{WDP}, with implicit dependence on L, $\mathcal{C}$, P_Z and P_S.

A comparison zero-interference (ZI) system uses the same noise distribution P_Z and input constraint set $\mathcal{C}$, but the interference does not play a role. That is, given the (constrained) output sequence x^n, the output is given by $Y^n = x^n + Z^n$. Thus, for this system, the encoder is of the form $X^n = f_n(M)$ and the probability of error is measured as $P_e = \Pr\{g_n(f_n(M) + Z^n) \neq M\}$. The capacity is defined as above and is written C_{ZI}, with implicit dependence on L, $\mathcal{C}$, and P_Z.

The general formula for C_{WDP} is given in [8], but the assumption of strong interference allows us to derive a simpler expression. The proof of this lemma will be given in upcoming work [12].

Lemma 1. *For S uniformly distributed on $\mathcal{A}_L$, i.e., strong interference,*

$$C_{WDP} = \sup_{P_V, Q(\cdot)} H(V) - H(Q(V) + Z), \tag{4}$$

where the maximization is over distributions P_V on $\mathcal{A}_L$ and over quantizers with $Q(v) - v \in \mathcal{C}$ for all $v \in \mathcal{A}_L$.

If the side information can only be used causally, then the capacity is the same as (4), except that P_V is constrained to be uniform over $\mathcal{A}_L$ [3]. The zero-interference capacity is given by (see, e.g., [13]),

$$C_{\text{ZI}} = \sup_{P_X} H(X + Z) - H(Z), \tag{5}$$

where the maximization is over distributions P_X on $\mathcal{C}$.

[3] Addition throughout the rest of the paper will be done mod L with results in $\mathcal{A}_L$.

3 Difference Set Noise

In this section, we first define different sets and show that arbitrarily large ones exist. We next consider noise uniformly distributed on difference sets and show that adding such noise amplifies entropy as much as possible for a large class of input distributions.

A planar difference set D on the group $\mathcal{A}_L$ under addition modulo L satisfies

$$|D \cap (D+g)| = 1, \tag{6}$$

for all non-zero g in $\mathcal{A}_L$, where $D+g = \{d+g : d \in D\}$. In order to use difference sets in a non-trivial way, it is important that arbitrarily large ones exists. Indeed, the following lemma demonstrates that this is the case; see, e.g., [14] for a proof.

Lemma 2. *For any prime p and integer m, let $\alpha = p^m + 1$ and $L = \alpha(\alpha-1)+1$. A planar difference set D with $|D| = \alpha$ exists on the group $\mathcal{A}_L$ under addition modulo L.*

Not only do difference sets of size p^m+1 exist, but such a D can be easily constructed from any primitive cubic polynomial in the Galois field $\mathbb{F}(p^m)$.

We refer to the noise distribution P_Z that is uniform over a planar difference set as *difference set noise* (DSN). This noise is certainly non-periodic from the main property (6) of difference sets. Furthermore, we can make DSN "powerful" relative to the constraint set $\mathcal{C}$ since arbitrarily large difference sets exist. Thus, DSN is a good prospect for creating a large loss. In order to prove that the loss is large, we shall first show that additive DSN amplifies entropy as much as possible as long as the input support is smaller than the support of the DSN. The following lemma is proved in Appendix A.

Lemma 3 (Entropy Amplification Property). *Consider difference set noise Z of size α. For any random variable X with support $\beta \leq \alpha$,*

$$H(X) + H(Z) - \frac{\beta-1}{\alpha} \leq H(X+Z) \leq H(X) + H(Z). \tag{7}$$

4 Capacity with Difference Set Noise

In this section, we bound the WDP and ZI capacities in the presence of strong interference and DSN. We use the general capacity formulas introduced in Section 2 and the entropy amplification property of DSN given in Section 3.

Theorem 1. *Let P_S be uniformly distributed on $\mathcal{A}_L$ (i.e., S is strong interference) and let P_Z be uniformly distributed on a difference set D over $\mathcal{A}_L$ (i.e., Z is difference set noise). Then, for any constraint set $\mathcal{C}$ with $\beta = |\mathcal{C}| \leq \alpha = |D|$,*

$$C_{WDP} \leq \frac{\alpha-1}{\alpha} + \log\left(1+\frac{\beta}{\alpha}\right) \leq 2 \text{ bits}, \tag{8}$$

while

$$C_{ZI} \geq \log\beta - \frac{\beta-1}{\alpha} \geq \log\beta - 1 \text{ bits}. \tag{9}$$

By Lemma 2, we can choose α, and hence β, arbitrarily large. Thus, the loss can be arbitrarily large. Furthermore, since $\log\beta$ is an upper bound on any capacity with a constraint set $\mathcal{C}$ of size β, we see that the loss can approach 100% of the available capacity.

Proof. The bound on the ZI capacity follows directly from (5) and Lemma 3, by making the input distribution P_X uniform over the constraint set. In this case,

$$\begin{aligned} H(X+Z) - H(Z) &\geq H(X) + H(Z) - \frac{\beta - 1}{\alpha} - H(Z) && (10) \\ &= \log \beta - \frac{\beta - 1}{\alpha}. && (11) \end{aligned}$$

We now turn to the WDP capacity with strong interference (4) and provide a bound on $H(V) - H(Q(V) + Z)$ for any allowable distribution P_V and quantizer $Q(\cdot)$. We write the outputs of the quantizer as $\mathcal{Q} = \{q_1, \ldots, q_m\}$. In order to use the entropy amplification property, we consider a further quantization $K(\cdot)$ of the space $\mathcal{Q}$. This quantizer groups α points of $\mathcal{Q}$ together according to

$$K(q_i) = \left\lfloor \frac{\pi(i)}{\alpha} \right\rfloor, \tag{12}$$

where $\pi(\cdot)$ is a permutation of $\{1, \ldots, m\}$ that we specify below. We refer to the groupings formed by $K(\cdot)$ as K-bins, e.g., the set $K^{-1}(1)$ is the first K-bin. If we write Q for $Q(V)$ and K for $K(Q(V))$, we see that

$$\begin{aligned} H(Q+Z) &= H(Q+Z|K) + H(K) - H(K|Q+Z) \\ &\geq H(Q|K) + H(Z) - \frac{\alpha - 1}{\alpha} + H(K) - H(K|Q+Z) \\ &= H(Q) + H(Z) - H(K|Q+Z) - \frac{\alpha - 1}{\alpha}. && (13) \end{aligned}$$

Here, the inequality follows by Lemma 3 since given $K = k$ the support of Q is at most α. The final equality follows since K is a deterministic function of Q. Combining (13) with the fact that $H(K|Q+Z) \leq H(K)$, we observe that

$$H(V) - H(Q(V) + Z) \leq H(V|Q(V)) + H(K(Q(V))) - H(Z) + \frac{\alpha - 1}{\alpha}. \tag{14}$$

In order to upper bound (14), let us write the probability of each quantizer output as

$$p_i = \Pr(Q = q_i) \tag{15}$$

and let us also define

$$r_i = \frac{|Q^{-1}(q_i)|}{L}, \tag{16}$$

which is the probability of quantizer output i if P_V were uniform over $\mathcal{A}_L$. Consider a permutation $\pi(\cdot)$ so that $r_{\pi(i)}$ is a non-increasing sequence. In the sequel, we will use this permutation in the definition of K (12). Although this choice of $\pi(\cdot)$ does not minimize the RHS of (14),[4] it will allow us to prove the desired results. With this permutation, the first K-bin has the α largest quantization points according to the sequence $\boldsymbol{r}$, the second K-bin has the second α largest points according to $\boldsymbol{r}$, and so forth. Notice that $r_{\pi(1)}$ (and hence each r_i) is at most $\frac{\beta}{L}$. This follows from the constraint that $Q(v) - v \in \mathcal{C}$

[4]The minimizing choice of $\pi(\cdot)$ makes $p_{\pi(i)}$ a non-increasing sequence. The resulting distribution on K majorizes all other possible P_K, and hence $H(K)$ is minimized [15].

for all v, and thus each $v \in Q^{-1}(q)$ corresponds to exactly one of the β c's in $\mathcal{C}$. Let us next define

$$p_i^{K\text{-mean}} = \frac{1}{\alpha} \sum_{i':K(q_{i'})=K(q_i)} p_i \tag{17}$$

and

$$r_i^{K\text{-max}} = \max_{i':K(q_{i'})=K(q_i)} r_i. \tag{18}$$

That is, $p_i^{K\text{-mean}}$ is the average of the probabilities of the quantization points in the same K-bin as q_i. Similarly, $r_i^{K\text{-max}}$ is the maximum of the r-values in the same K-bin as q_i. We can compute the distribution of $K(Q(V))$ as

$$\begin{aligned}
\Pr(K(Q(V)) = k) &= \sum_{i:q_i \in K^{-1}(k)} p_i && (19)\\
&= \sum_{i:q_i \in K^{-1}(k)} p_i^{K\text{-mean}} && (20)\\
&= \alpha p_i^{K\text{-mean}},\ \forall i : q_i \in K^{-1}(k), && (21)
\end{aligned}$$

where the last equality follows since $p_i^{K\text{-mean}}$ is constant for all q_i in the same K-bin. It follows that

$$H(K(Q(V))) = \sum_i p_i^{K\text{-mean}} \log \frac{1}{\alpha p_i^{K\text{-mean}}}. \tag{22}$$

We can also upper bound $H(V|Q(V))$ using

$$\begin{aligned}
H(V|Q(V)) &\le \sum_i p_i \log L r_i && (23)\\
&\le \sum_i p_i \log L r_i^{K\text{-max}} && (24)\\
&= \sum_i p_i^{K\text{-mean}} \log L r_i^{K\text{-max}}. && (25)
\end{aligned}$$

The first bound follows since given $Q(V) = q_i$, the conditional distribution that maximizes the entropy is uniform over the Lr_i values in $Q^{-1}(q_i)$. The second bound follows since $r_i \le r_i^{K\text{-max}}$ for all i. The equality follows as in (20) since $r_i^{K\text{-max}}$ is constant over any K-bin. Combining the previous two results, we see that

$$\begin{aligned}
&H(V|Q(V)) + H(K(Q(V))) - \log \alpha\\
&\quad\le \sum_i p_i^{K\text{-mean}} \log \frac{r_i^{K\text{-max}}}{p_i^{K\text{-mean}}} + \log \frac{L}{\alpha^2} && (26)\\
&\quad= \log \sum_i r_i^{K\text{-max}} - D(\boldsymbol{p}^{K\text{-mean}} || \boldsymbol{r}^{K\text{-max}} / \sum_i r_i^{K\text{-max}}) + \log \frac{L}{\alpha^2} && (27)\\
&\quad\le \log \sum_i r_i^{K\text{-max}} + \log \frac{L}{\alpha^2} && (28)\\
&\quad\le \log \left(\frac{\alpha\beta}{L} + 1 \right) + \log \frac{L}{\alpha^2} && (29)\\
&\quad\le \log \left(1 + \frac{\alpha}{\beta} \right). && (30)
\end{aligned}$$

Here, (28) follows since the relative entropy term is non-negative and (29) follows since

$$r_{\pi(i)}^{K\text{-max}} \leq \begin{cases} \frac{\beta}{L} & \text{if } 1 \leq i \leq \alpha \\ r_{\pi(i-\alpha)} & i > \alpha \end{cases} \tag{31}$$

and since $\sum_i r_i = 1$. The bound on the WDP capacity follows by combining (4), (14) and (30). □

5 Discussion

In this paper, we introduced difference set noise (DSN), and we demonstrated that it contains the necessary irregularity to produce large loss for writing on dirty paper (WDP). There are several other interesting applications of difference sets and DSN. The first is to consider the role of DSN in the source coding with side information (Wyner-Ziv) problem, which is often considered the dual of WDP. In this problem, DSN can also be used to produce arbitrarily large loss. Another use of difference sets is in constructing expander graphs. For example, consider the graph $\mathcal{G} \subset \mathcal{A}_L \times \mathcal{A}_L$, in which $(x, y) \in \mathcal{G}$ if $y - x \in \mathcal{D}$. Then, every input is connected to α outputs and every set of $\beta \leq \alpha$ inputs is connected to at least $\beta\alpha/2$ outputs. Finally, the entropy amplification property of DSN given in Lemma 3 (or a generalized version to larger input support) could be used to guarantee output entropy in a variety of situations.

Our main result concerned a discrete version of WDP with a hard input constraint. We would like to extend these results to continuous alphabets and average input constraints. One analogous situation with continuous alphabets (but still with a hard input constraint) would be to let the interference S be a Unif($[0, 1]$) random variable and let the noise Z be DSN divided by L plus a Unif($[0, 1/L]$) random variable. Then, for a peak input constraint, $x \leq \beta/L$, the capacities should behave similarly to those in Theorem 1. Since $\beta \leq \alpha$ and $L \approx \alpha^2$, we see explicitly that the "signal to noise ratio" must be small in order for the loss to be large, which agrees with the results of [6].

A Proof of Entropy Amplification Property

In this section, we prove the entropy amplification property of difference set noise (DSN). Recall that DSN consists of a random variable Z that is uniformly distributed over a planar difference set D over $\mathcal{A}_L$ with addition modulo L. We also consider an arbitrary set $A_x = \{x_1, \ldots, x_\beta\} \subset \mathcal{A}_L$ with the only constraint that $\beta \leq \alpha = |D|$. We develop a lower bound on the entropy of the output of the additive DSN channel, $H(X + Z)$, for *any* input distribution P_X with support only on A_x. This lower bound will be within a constant of the upper bound

$$H(X + Z) \leq H(X) + H(Z), \tag{32}$$

which is achieved only if $x + z \neq x' + z'$ for all $x \neq x' \in A_x$ and $z \neq z' \in D$.

Let us consider an arbitrary distribution on X, which we write $\boldsymbol{p} = (p_1, \ldots, p_\beta)$ so that $\Pr(X = x_i) = p_i$. To describe the distribution of $Y = X + Z$, let us define the sets

$$B_i = x_i + D, \ \forall 1 \leq i \leq m. \tag{33}$$

y	$\alpha \Pr(Y = y)$	$\alpha \Pr(Y' = y)$
y_0	$p_1 + p_2 + \cdots + p_{n_0}$	$p_2 + \cdots + p_{n_0}$
y_j for $1 \leq j \leq n_0 - 2$	p_1	0
z_j for $2 \leq j \leq n_0$	p_j	$p_1 + p_j$

Table 1: Comparison of the distributions of Y and Y' in the proof of the entropy amplification property.

Given that $X = x_i$, the conditional distribution of Y is uniform over the set B_i. The unconditional distribution of Y is given by

$$\Pr(Y = y) = \sum_{i=1}^{\beta} \Pr(X = x_i) \Pr(Y = y | X = x_i) = \sum_{i=1}^{\beta} \frac{p_i}{\alpha} 1_{\{y \in B_i\}}. \tag{34}$$

Due to the properties of difference sets and the definition of B_i in (33), these sets must satisfy

$$|B_i| = \alpha, \tag{35}$$
$$|B_i \cap B_j| = 1, \ \forall i \neq j, \text{ and} \tag{36}$$
$$|\cup B_i| \leq L. \tag{37}$$

Let us now minimize $H(Y)$ where Y has the distribution (34) and the collection of sets $\{B_i\}$ must satisfy (35), (36) and (37). This will provide a lower bound on the entropy of Y when Z is difference set noise.

First, find a point y_0 such that y_0 is in at least three of the B_i's. If there is no such y_0, then the distribution of Y already has the form (40) and we can proceed from there. Otherwise, we can reduce the entropy of Y by changing the B_i's as follows. Without loss of generality, let us say that that y_0 is in $B_1, \ldots, B_{n_0}$ (but not in $B_{n_0-1}, \ldots, B_\beta$). There exists $y_1, \ldots, y_{n_0-2} \in B_1$ such that $y_i \notin B_i$ for $i \neq 1$. To see this, note that at most $\beta - n_0$ of the α points in B_1 intersect with any other set. There also exists $z_2 \in B_2, \ldots, z_{n_0} \in B_{n_0}$ such that $z_j \notin B_i$ for $i \neq j$. Again, this follows since at most $\beta - n_0$ of the α points in the relevant B_j's intersect with any other set. Create a set B_1' from B_1 by changing y_j to z_{j+2} for $0 \leq j \leq n_0 - 2$. Create sets B_j' for $j \geq 2$ by copying B_j. Let Y' have the distribution (34) with the sets $\{B_j'\}$ instead of $\{B_j\}$. Table 1 summarizes the resulting differences in the distributions of Y and Y'.

In order to compare $H(Y)$ and $H(Y')$ let us define $c_j = P_j/P_1$ for $2 \leq j \leq n_0$. With this definition,

$$H(Y) - H(Y') = \frac{P_1}{\alpha} \left[\log \frac{\prod_{j=2}^{n_0}(1 + c_j)}{1 + \sum_{j=2}^{n_0} c_j} + \sum_{j=2}^{n_0} c_j \log \left(\frac{1 + c_j}{c_j} \cdot \frac{\sum_{j'=2}^{n_0} c_{j'}}{1 + \sum_{j'=2}^{n_0} c_{j'}} \right) \right] \tag{38}$$
$$> 0. \tag{39}$$

The inequality follows since the terms in the logs are greater than 1 for positive c_j. The term in the first log of (38) is greater than 1 since the denominator contains just 2 of the n_0 positive terms in the expansion of the numerator. The term in the second log is greater than 1 since $(1 + x)/x > (1 + x + y)/(x + y)$ for positive x and y. Thus, every iteration of this procedure reduces the entropy of Y.

After a finite number of iterations of the above process there will be no y_0 which is in more than two B_i's. Notice that there must be at least one y in at least two B_i's and that the resulting distribution on Y is unique up to permutation. We have thus created sets $\{B_i^*\}$ under the constraints (35), (36), and (37) that minimize $H(Y)$. The main property of these sets is that each y appears in at most two B_i^*'s. The optimal distribution is thus of the form

$$\Pr(Y^* = y) = \begin{cases} \frac{p_j}{\alpha} & \text{for } \alpha - \beta + 1 \text{ values of } y \text{ for each } 1 \le j \le \beta, \\ \frac{p_j + p_{j'}}{\alpha} & \text{for one } y \text{ for each } 1 \le j < j' \le \beta. \end{cases} \tag{40}$$

In order to study the entropy of Y^*, let us consider a random variable Y_2^* that only contains the points y that are in two B_i^*'s. That is,

$$\Pr(Y_2^* = y) = \frac{p_j + p_{j'}}{\beta - 1} \quad \text{for one } y \text{ for each } 1 \le j < j' \le \beta. \tag{41}$$

The entropy of this random variable satisfies

$$H(Y_2^*) \ge H(X) + \log(\beta - 1) - 1. \tag{42}$$

To see this, we can take the convex combination of two reorderings of copies of the distribution of X. In particular, let

$$\boldsymbol{a} = \frac{1}{\beta - 1}[p_1\ p_1 \dots p_1\ p_2\ p_2 \dots p_2\ \dots\ p_\beta\ p_\beta \dots p_\beta], \text{ and} \tag{43}$$

$$\boldsymbol{b} = \frac{1}{\beta - 1}[p_2\ p_3 \dots p_\beta\ p_1\ p_3 \dots p_\beta\ \dots\ p_1\ p_2 \dots p_{\beta-1}]. \tag{44}$$

That is, $\boldsymbol{a}$ has $\beta - 1$ copies of p_1, then $\beta - 1$ copies of p_2 and so forth. On the other hand, $\boldsymbol{b}$ has every value p_j except p_1 in the first $\beta - 1$ places, then every value p_j except p_2 in the second $\beta - 1$ places, and so forth. We see that $H(\boldsymbol{a}) = H(\boldsymbol{b}) = H(X) + \log(\beta - 1)$. On the other hand, $H((\boldsymbol{a} + \boldsymbol{b})/2) = H(Y_2^*) + 1$. The inequality (42) can be seen by combining these equalities with the inequality $H((\boldsymbol{a} + \boldsymbol{b})/2) \ge (H(\boldsymbol{a}) + H(\boldsymbol{b}))/2$ by the concavity of entropy.

We conclude the proof of Lemma 3 by computing

$$\begin{aligned} H(X + Z) &\ge H(Y^*) && (45) \\ &= \log \alpha + \frac{\alpha - \beta + 1}{\alpha} H(X) + \frac{\beta - 1}{\alpha} \left(H(Y_2^*) - \log(\beta - 1) \right) && (46) \\ &\ge H(Z) + H(X) + \frac{\beta - 1}{\alpha} 1. && (47) \end{aligned}$$

Acknowledgment

The authors are indebted to Dave Forney for referring them to difference set theory.

References

[1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[2] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1639–1667, June 2002.

[3] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference." Submitted to *IEEE Trans. Info. Theory*, 2002.

[4] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May 2001.

[5] A. S. Cohen and A. Lapidoth, "Generalized writing on dirty paper," in *Proc. of ISIT*, (Lausanne, Switzerland), p. 227, 2002.

[6] R. Zamir, "The half a bit loss of robust source/channel codebooks," in *Proc. of the Info. Theory Workshop*, (Bangalore, India), 2002.

[7] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, Oct. 1958.

[8] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[9] J. Wolfowitz, *Coding Theorems in Information Theory*. Springer-Verlag, third ed., 1978.

[10] U. Erez and R. Zamir, "Noise prediction for channels with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1610–1617, July 2000.

[11] R. Zamir and A. S. Cohen, "Unbounded loss in writing on dirty paper in possible," in *DIMACS Workshop on Network Information Theory*, (Rutgers, NJ), Mar. 2003.

[12] A. S. Cohen and R. Zamir, "The rate loss in writing on dirty paper." In preparation, 2003.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[14] T. Beth, D. Jungnickel, and H. Lenz, *Design Theory*, vol. 1. Cambridge University Press, second ed., 1999.

[15] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.