Entropy Amplification Property and the Loss for Writing on Dirty Paper *

Aaron S. Cohen[†]
acohen@alum.mit.edu

Ram Zamir[‡]
zamir@eng.tau.ac.il

Final revision: December 24, 2007

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 cancelation 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". If the interference S is strong, then difference-set noise limits the WDP capacity to at most 2 bits. At the same time, like in the AWGN case, the zero-interference capacity grows without bound with the input constraint. Thus almost 100 % of the available capacity is lost in WDP in the presence of difference-set noise. This high capacity loss is due to the "entropy amplification property" (EAP) of noise with an aperiodic probability distribution. Using the EAP and the duality between WDP and Wyner-Ziv source coding, we also give an example of dramatic rate-loss in quantizing encrypted source.

Keywords: side information, writing on dirty paper, the Wyner-Ziv setup, difference sets.

1 Introduction

In Costa's *writing on dirty paper* (WDP) [6], a channel encounters two independent sources of additive white Gaussian noise. One source, S^n, is known to the transmitter

[*] This work was presented in part at the Allerton Conference, IL., October 2003, and at ISIT04, Chicago, IL., June 2004.

[†] Aaron Cohen was supported in part by the National Science Foundation under Grant DMS-0074276. He was with the Division of Applied Mathematics, Brown University, Providence, RI 02912.

[‡] This research was performed during the year 2002-2003, while Ram Zamir was on Sabbatical leave from Tel Aviv University and a visiting scientist at the Signals, Information and Algorithms group, the Research Laboratories of Electronics, MIT. He is now with the Dept. of Electrical Engineering - Systems, Tel Aviv University, Ramat Aviv 69978, Israel.

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 on 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 with vanishing decoding error probability) of WDP is given by

$$C_{WDP}^{G} = \frac{1}{2} \log \left(1 + \frac{P}{N} \right)$$

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 S. 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 cancellation property" is particularly interesting as unlike the receiver, a 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 S (i.e., it can be non-Gaussian or even arbitrarily varying) if [3, 10] and only if [4] the noise Z is Gaussian[2]. Furthermore, for general distribution of the noise Z 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) [20]. 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 [2], the related Gaussian Wyner-Ziv problem [18, 19], and general algebraic binning for network communications settings [21] 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 [16, 11] and the receiver-SI capacity (e.g., [17]). 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 [9]. 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 also 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. In contrast to the conditions for zero or small loss, the noise Z 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; at the same time it must be relatively sparse (so the zero-interference capacity can be high), yet dense enough to make a strong, negative impact on the WDP capacity. Thus, we consider strong interference S, so that its knowledge at the receiver has the most benefit, and irregular noise Z. In a previous paper [5], we constructed such a noise distribution and demonstrated that the loss in

[1] Logarithms throughout the paper are base 2.

[2] The "only if" part holds for a certain large class of transmitters.

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 [5] in two ways. First, we extend the analysis to *non-causal* WDP, for which capacity is potentially higher. Second, we find a family of irregular 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. Thus almost 100 % of the capacity is lost in WDP compared to the fully informed system case (or equivalently, compared to the zero-interference case). In addition, we construct a "dual" source coding problem with a large gap between the decoder-SI rate-distortion function (the Wyner-Ziv function) and the conditional (encoder + decoder-SI) rate-distortion function.

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 $\mathcal{G}$ for which $D \cap (D + g)$ has the same number of elements for all shifts $g \neq 0$. We focus on a variation on this notion, namely, sets with *unique differences*, where $|D \cap (D+g)| \leq 1$, and we refer to a uniform distribution over such a set as *difference set noise* or *aperiodic noise*. In Section 2, we identify the "entropy amplification property" (EAP) of additive difference set noise. In Section 3 we derive a simplified formula for the WDP capacity for strong interference and general noise. In Section 4 we apply this formula to the case of difference-set noise; we use the EAP to separate the effects of the noise and the input distribution, which allows us to derive our main result regarding the capacity loss in WDP. Finally, in Section 5 we derive a parallel result for the rate loss in the Wyner-Ziv problem.

2 Entropy Amplification Property

In this section we examine the behavior of the entropy of an independent sum of discrete random variables. It is well known that independent addition increases entropy in a sub-linear way [8]. That is, for any independent random variables X and Z in an additive group $\mathcal{G}$,

$$H(X) \leq H(X+Z) \leq H(X) + H(Z) \tag{1}$$

where $X + Z$ means addition of the two variables in $\mathcal{G}$. Both inequalities above follow from the convexity ($\cap$) of the entropy function. The upper bound is a special case of the fact that for any function f of the pair (x, z), and in particular for addition, the entropy of $f(X, Z)$ is less than or equal to the joint entropy of X and Z, which in the independent case is equal to their entropy sum [8]. The lower bound follows since conditioning can only reduce the entropy, and in the independent case the conditional entropy $H(X+Z|Z)$ is equal to $H(X)$ [8]. The latter inequality becomes equality if and only if $X + Z$ is statistically independent of Z, a condition that can be interpreted as if the channel from Z to $X + Z$ "masks" Z completely so $H(X+Z) - H(X) = I(Z; Z+X) = 0$.

Our main interest, however, is in the tightness of the upper bound; or at least, in achieving approximately

$$H(X+Z) \approx H(X) + H(Z). \tag{2}$$

Equality in (2) holds if and only if both X and Z can be obtained from their sum (i.e., the sum is "invertible") with probability one. Such a situation happens, e.g., if the distribution of X is *sparse* so that the spacing between any two points where X gets positive probability is larger than, say, Δ, while the distribution of Z is *concentrated*

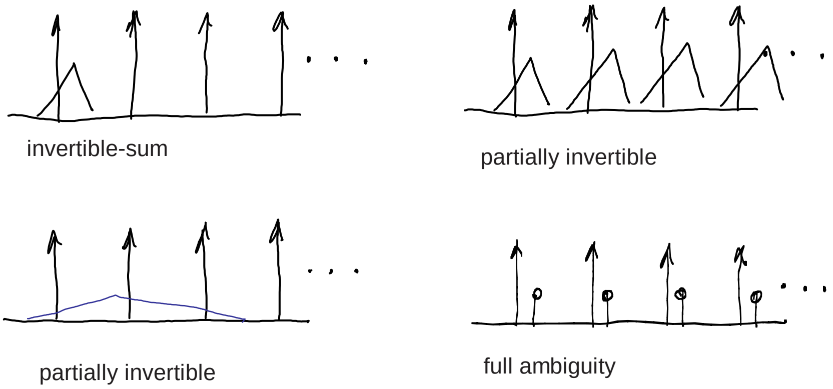

Figure 1: Cases of invertibility and ambiguity given the sum.

within an interval of size Δ. This invertibility condition is not met if both X and Z are concentrated in an interval, or if their support is *periodic* with the same period. See Figure 1.

Is there a "universal entropy amplifier"? That is, can we find a single Z such that (2) holds for *any* X? Clearly, we cannot hope to satisfy that without some limitation on the support of X; e.g., uniform distribution has the maximum entropy (the logarithm of the alphabet size) so its entropy cannot increase further by addition. Also, *exact* equality in (2) cannot hold uniformly, because for every Z we can find an X such that the ambiguity in X given $X + Z$ is not quite zero. Nevertheless, there exist sufficiently sparse and non-periodic distributions of Z, such that with a suitable limitation on the support of X, (2) holds up to a small constant gap. Such "aperiodic distributions" are related to the notion of a *difference set.*

2.1 Difference Set

We say that a subset $\mathcal{Z}$ of an Abelian group $\mathcal{G}$ has *unique differences* if for all non-zero $d \in \mathcal{G}$ the equation

$$z_1 - z_2 = d$$

has **at most one** solution for $z_1, z_2 \in \mathcal{Z}$. This is equivalent to saying that the intersection of $\mathcal{Z}$ with its shift either contains one point or is empty:

$$|\mathcal{Z} \cap (\mathcal{Z} + d)| \leq 1 \ \ \forall \ d \neq 0. \tag{3}$$

One example of a set with unique differences is $\mathcal{Z} = \{2^k : 1 \leq k \leq n\}$ for any n, where the group $\mathcal{G}$ is the integers. This follows since $2^{k_1} - 2^{k_2} = 2^{j_1} - 2^{j_2}$ if and only if $k_1 = j_1$ and $k_2 = j_2$. A special example of unique differences is when the equation $z_1 - z_2 = d$ has **exactly one** solution, i.e.,

$$|\mathcal{Z} \cap (\mathcal{Z} + d)| = 1 \ \ \forall \ d \neq 0. \tag{4}$$

For example, the set $\{0, 1, 4, 6\}$ in the group of integers modulo 13 has a unique (ordered) pair with difference d for every d in $\{1, \ldots, 12\}$. See Table 1. Such a set is known as a

i j	0	1	4	6
0	*	1	4	6
1	12	*	3	5
4	9	10	*	2
6	7	8	11	*

Table 1: A difference table for the planar difference-set of size 4 in the group of integers modulo 13. (Here $\alpha = 3^1 + 1 = 4$ and so $|\mathcal{G}| = 4(4-1)+1 = 13$). Every non-zero difference $d = i - j$ corresponds to a unique pair (i, j) in the group.

planar difference set, and is the largest possible set with unique differences for a fixed (finite) $\mathcal{G}$. In fact, we can bound the size of a set $\mathcal{Z}$ with unique differences by

$$|\mathcal{Z}|(|\mathcal{Z}|-1) \leq |\mathcal{G}| - 1, \tag{5}$$

and equality holds if and only if $\mathcal{Z}$ is a planar difference set. [3] Although it is not true that a set satisfying (4) exists in every finite group, it is interesting to note that for any prime p and integer m, a planar difference set exists with size $\alpha = p^m + 1$ in the group of integers modulo $\alpha(\alpha-1)+1$; see e.g. [1]. The modulo-13 group in Table 1 is the case $p = 3$ and $m = 1$. These examples demonstrate that arbitrarily large sets with unique differences exist, and that their size is roughly the square root of the group size (for planar difference sets) or smaller.

In the context of set sums, the bounds on the entropy of an independent sum in (1) are analogous to

$$|\mathcal{X}| \leq |\mathcal{X} + \mathcal{Z}| \leq |\mathcal{X}| \cdot |\mathcal{Z}| \tag{6}$$

where

$$\mathcal{X} + \mathcal{Z} = \{x + z : \ x \in \mathcal{X}, z \in \mathcal{Z}\}.$$

For finite sets, equality in the upper bound occurs if and only if for any y there is at most one pair $x \in \mathcal{X}, z \in \mathcal{Z}$ satisfying $x + z = y$. As we shall see below, if $\mathcal{Z}$ has unique differences, then the upper bound is "relatively tight", up to a factor of 2, for *any* $\mathcal{X}$ whose size is smaller than or equal to $|\mathcal{Z}| + 1$.

2.2 Entropy Amplification by Aperiodic Noise

When viewed as noise in an additive noise channel, uniform distribution over a set $\mathcal{Z}$ with unique differences amplifies the output entropy as much as possible for a large class of input distributions. Note that uniform distribution over $\mathcal{Z}$ is aperiodic in the sense that its autocorrelation function, $\sum_i P(i)P(i+d)$, is upper bounded by $1/|\mathcal{Z}|^2$ for all non-zero shifts d, which is the smallest possible bound. We thus call $Z \sim \text{Unif}(\mathcal{Z})$ *aperiodic noise.* The following lemma gives a precise statement of the Entropy Amplification Property (EAP) of aperiodic noise.

[3] By the unique differences property, every ordered pair (z_1, z_2) corresponds to a *different* difference $d = z_1 - z_2$; thus the number of possible (non-zero) differences in $\mathcal{G}$, which is $|\mathcal{G}| - 1$, is greater than or equal to the number of ordered pairs in $\mathcal{Z}$, which is $|\mathcal{Z}|(|\mathcal{Z}|-1)$. The reverse inequality holds if an ordered pair in $\mathcal{Z}$ exists for *every* (non-zero) difference d, as indeed implied by (4).

z x	0	1	4	6
0	0	1	4	6
3	3	4	7	9
6	6	7	10	12
9	9	10	0	2
12	12	0	3	5

Table 2: Sum table of the set $\mathcal{X} = \{0, 3, 6, 9, 12\}$ and the planar difference-set $\mathcal{Z} = \{0, 1, 4, 6\}$ in the modulo-13 group of Table 1. The sum 0 appears three times in the table; 3,4,6,7,9,10 and 12 appear twice; 1,2,5 appear once; and 8 and 11 do not appear at all.

Lemma 1. *(***EAP***) Let $\mathcal{Z} \subseteq \mathcal{G}$ have unique differences as defined in (3), and let Z be uniformly distributed on $\mathcal{Z}$ (i.e., Z is aperiodic noise). Then, for any random variable X in $\mathcal{G}$ that is statistically independent of Z and whose support size $|\mathcal{X}|$ is bounded from above by $|\mathcal{Z}| + 1$, the entropy of $X + Z$ is lower bounded by*

$$\begin{aligned} H(X+Z) &\geq H(X) + H(Z) - \frac{|\mathcal{X}| - 1}{|\mathcal{Z}|} \log(2) \\ &\geq H(X) + H(Z) - 1 bit. \end{aligned} \tag{7}$$

Lemma 1 implies that the ambiguity in the pair (X, Z) given the sum $X + Z$ is at most one bit on the average. For example, consider X and Z which are uniformly distributed over the set $\mathcal{X} = \{0, 3, 6, 9, 12\}$ and the planar difference-set $\mathcal{Z} = \{0, 1, 4, 6\}$, respectively, in the group of integers modulo 13. Table 2 shows that most sums $x + z$ result from two different pairs (x, z), so the ambiguity in the pair given the sum is indeed 1 bit; for three sums the ambiguity is zero, while in one case (given $x + z = 0$) there are three different pairs so the ambiguity is $\log(3) \approx 1.585$ bit. Explicit calculation gives

$$H(X+Z) = -7\frac{2}{20}\log\Big(\frac{2}{20}\Big) - 3\frac{1}{20}\log\Big(\frac{1}{20}\Big) - \frac{3}{20}\log\Big(\frac{3}{20}\Big) \approx 3.3842 \ bit \tag{8}$$

which is indeed larger than $H(X) + H(Z) - 1 \approx 2.32 + 2 - 1 = 3.32$ bit. Furthermore, using (8) we can calculate the (average) conditional entropy of the pair given the sum:

$$H(X, Z|X+Z) = H(X) + H(Z) - H(X+Z) \approx 0.94 \ bit$$

which is indeed smaller than 1 bit. We shall discuss a case of equality in the EAP in the next subsection.

Another interesting corollary of Lemma 1 concerns the tightness of the upper bound on the size of a set sum in (6). Specifically, by a maximum entropy argument we have on the one hand $H(X+Z) \leq \log|\mathcal{X} + \mathcal{Z}|$ for any distribution on X. On the other hand, we have $H(Z) = \log|\mathcal{Z}|$, and for $X \sim \text{Unif}(\mathcal{X})$ we have $H(X) = \log|\mathcal{X}|$. Substituting in (7), we obtain that the size of the sum of a set with unique differences and an arbitrary set of smaller size is always *greater* than or equal to half their size product:

$$|\mathcal{X} + \mathcal{Z}| \geq \frac{|\mathcal{Z}| \cdot |\mathcal{X}|}{2}. \tag{9}$$

		A	A			A	
B	B	B					
		C		C	C		
	D		D	D			
0	1	2	3	4	5	6	...

(A)

		A	A			A	
B	B	B					
C’				C		C’	
	D		D	D			
0	1	2	3	4	5	6	...

(B)

Figure 2: Unstructured sets satisfying the constraint in (11). (A) before modification. (B) after modification.

Thus, a set with unique differences is a “universal size amplifier” under set addition.

Proof. of Lemma 1: Let us define $\alpha = |\mathcal{Z}|$, $\beta = |\mathcal{X}|$, $\mathcal{X} = \{x_1, \ldots, x_\beta\}$, and $p_i = \Pr(X = x_i)$. Let us also define the random variable $Y = X + Z$, and define $\mathcal{Y}_i = x_i + \mathcal{Z}$, $i = 1, \ldots, \beta$, so $\mathcal{Y}_i$ is a shift of $\mathcal{Z}$ by x_i, and $|\mathcal{Y}_i| = \alpha$ for all i. Given $X = x_i$, Y is uniformly distributed on $\mathcal{Y}_i$. Hence we can write the distribution of Y as

$$\Pr(Y = y) = \sum_{i=1}^{\beta} \Pr(X = x_i) \Pr(Y = y | X = x_i) = \sum_{i=1}^{\beta} \frac{p_i}{\alpha} 1_{\{y \in \mathcal{Y}_i\}} \tag{10}$$

where $1_{\{\cdot\}}$ denotes indicator function. Note that $\{y \in \mathcal{Y}_i\}$ is equivalent to $\{y - x_i \in \mathcal{Z}\}$. Since $\mathcal{Z}$ has unique differences, we have from (3)

$$|\mathcal{Y}_i| = \alpha, \quad |\mathcal{Y}_i \cap \mathcal{Y}_j| \leq 1, \quad i, j = 1, \ldots, \beta, \quad i \neq j. \tag{11}$$

In order to provide a lower bound on $H(Y)$, let us consider the distribution (10) under the constraint that $\mathcal{Y}_1, \ldots, \mathcal{Y}_\beta$ must satisfy (11) (but these sets are not necessarily related to the underlying $\mathcal{Z}$ and $\mathcal{X}$). See Figure 2 (A) where all set pairs intersect at exactly one point (A and B at $y = 2$, A and C at $y = 2$, A and D at $y = 3$, B and C at $y = 2$, B and D at $y = 1$, and C and D at $y = 4$). We claim that

(Claim 1) with the constraint (11), a collection of sets $\{\mathcal{Y}_i\}$ that minimizes $H(Y)$ satisfies that each y is in **at most two** of the $\mathcal{Y}_i$’s.

If this property is not satisfied, then there exists a point which is common to three or more sets; assume (without loss of generality) that y_0 is in $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ where $\beta \geq n \geq 3$. In the example of Figure 2(A), the A, B and C sets intersect at $y_0 = 2$ so $n = 3$. We shall show that we can reduce the entropy of Y by decreasing the number of intersections at y_0 to two, without violating the constraint (11), and without generating new points with three or more intersections. Specifically, since $\beta \leq \alpha + 1$ and since the $\mathcal{Y}_i$’s must satisfy (11), there exists at least $n - 2$ points y in each $\mathcal{Y}_i$, for $1 \leq i \leq n$, which belong

y	$\alpha \Pr(Y = y)$	$\alpha \Pr(Y' = y)$
y_0	$p_1 + p_2 + \cdots + p_n$	$p_1 + p_2$
a removed point in $\mathcal{Y}_i$	p_i	0
a point in $\mathcal{Y}_i$ to which a point in $\mathcal{Y}_j$ is modified	p_i	$p_i + p_j$

Table 3: Comparison of the distributions of Y and Y' in the proof of the entropy amplification property.

to this $\mathcal{Y}_i$ alone, i.e., $y \notin \mathcal{Y}_k$ for $k = 1, \ldots, \beta$, $k \neq i$. [4] In the example of Figure 2(A), the lonely point in the A set is $y = 6$, in the B set it is $y = 0$ and in the C set it is $y = 5$. Thus, we can modify these $n(n-2)$ points and the point y_0 into $n(n-1)/2$ points that are in **exactly two** of the $\mathcal{Y}_i$'s in $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$. In the example of Figure 2, we modified the C set as shown in Figure 2(B), where the point $y = 2$ became $y = 0$ and the point $y = 5$ became $y = 6$. Note that these $n(n-1)/2$ two-set intersections consist of *all* pairs of sets in $\{\mathcal{Y}_1, \ldots, \mathcal{Y}_n\}$; hence the probabilities (10) of the modified points y consist of the values $\{\frac{1}{\alpha}(p_i + p_j) : 1 \leq i < j \leq n\}$. Let Y' have the distribution (10) with these modified sets; see Table 3. Then,

$$\begin{aligned} &\alpha(H(Y) - H(Y')) \\ &= \sum_{1 \leq i < j \leq n} (p_i + p_j) \log(p_i + p_j) - \left(\sum_{i=1}^{n} p_i\right) \log \left(\sum_{i=1}^{n} p_i\right) - (n-2) \sum_{i=1}^{n} p_i \log p_i \\ &= \sum_{i=1}^{n} p_i \log \frac{\prod_{j \neq i}(p_i + p_j)}{p_i^{n-2}\left(\sum_{j=1}^{n} p_j\right)} \\ &> 0, \end{aligned} \tag{12}$$

where the inequality follows since for each i the term in the log is greater than 1 (the numerator is equal to the denominator plus additional positive terms). By repeating this process a finite number of times (with each step reducing entropy as in (12)), we obtain a collection of sets $\{\mathcal{Y}_i\}$ with each y in at most two $\mathcal{Y}_i$'s. This proves Claim 1 above.

We also claim that a minimizing solution satisfies (11) with equality for all $i, j = 1, \ldots, \beta$, i.e.,

(Claim 2) a collection of sets $\{\mathcal{Y}_i\}$ that minimizes $H(Y)$ satisfies $|\mathcal{Y}_i \cap \mathcal{Y}_j| = 1$ for all $i \neq j$.

To see why, note that if after the modification above there are any remaining pairs of sets with $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, then we can further reduce the entropy while still satisfying (11) by changing a point $y_i \in \mathcal{Y}_i$ to a point $y_j \in \mathcal{Y}_j$ where both y_i and y_j are not in any other sets (such points exist since $\beta \leq \alpha + 1$), thus obtaining one point with the probability

[4] To see why, suppose we start with the set $\mathcal{Y}_1$ and add the sets $\mathcal{Y}_2, \ldots, \mathcal{Y}_\beta$ sequentially, while noting at each stage which point of $\mathcal{Y}_1$ (if any) intersects the new $\mathcal{Y}_i$. (We say that the point is "covered" by the new $\mathcal{Y}_i$.) If each new intersection occurs at a different point, then there will be at most $\beta - 1$ points of $\mathcal{Y}_1$ covered, and therefore at least $\alpha - \beta + 1$ points which are in $\mathcal{Y}_1$ alone. If there are m "re-coverings" in this process (i.e., m is the number of stages in which the new $\mathcal{Y}_i$ intersects a point of $\mathcal{Y}_1$ which is already covered), then there will be at least $\alpha - \beta + 1 + m$ points which are in $\mathcal{Y}_1$ alone. Since we assumed that $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ intersect at y_0, there must be at least $n - 2$ re-coverings; substituting $m \geq n - 2$ and $\beta \leq \alpha + 1$, we get that there are at least $\alpha - \beta + 1 + m \geq n - 2$ points which are in $\mathcal{Y}_1$ alone. Since the same argument applies to all the sets $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, each one of them must have at least $n - 2$ points which are not in any other $\mathcal{Y}_i$.

sum $p_i + p_j$. After this process is complete, the graph of transition distribution from X to Y has the following properties: each $x \in \mathcal{X}$ is connected to α points y with branch probabilities $1/\alpha$; each $y \in \mathcal{G}$ is connected to at most two x's (by Claim 1); and each pair x and x' in $\mathcal{X}$ are connected to each other through exactly one y (by Claim 2). We thus have a total of $\beta(\beta-1)/2$ points y connected to exactly two x's (all pairs from $\mathcal{X}$), leaving out $\alpha\beta - \beta(\beta-1)$ points y (that is, $\alpha-(\beta-1)$ in the span of each x) which are connected to a single x. It follows that the minimizing distribution under the constraint (11) has the following form:

$$\Pr(Y^* = y) = \begin{cases} \frac{p_j}{\alpha} & \text{for } \alpha-\beta+1 \text{ values of } y \text{ for each } 1 \le j \le \beta, \\ \frac{p_i+p_j}{\alpha} & \text{for one } y \text{ for each } 1 \le i < j \le \beta, \end{cases} \tag{13}$$

and the entropy of this distribution lower bounds our quantity of interest, i.e., $H(X+Z) = H(Y) \ge H(Y^*)$.

In order to study the entropy of the random variable Y^*, let us define E as a binary random variable that takes the vlaue $E=1$ if Y^* falls into the first case of (13), and takes the value $E=2$ if Y^* falls into the second case of (13). We have

$$\Pr(E=1) = \frac{\alpha-\beta+1}{\alpha},$$

and since E is a function of Y^*, we have $H(Y^*) = H(Y^*,E) = H(E) + H(Y^*|E)$. Thus

$$H(Y^*) = H_B\Big(\frac{\beta-1}{\alpha}\Big) + \frac{\alpha-\beta+1}{\alpha}H(Y^*|E=1) + \frac{\beta-1}{\alpha}H(Y^*|E=2) \tag{14}$$

where H_B denotes the binary entropy. It is easy to verify that

$$H(Y^*|E=1) = H(X) + \log(\alpha-\beta+1).$$

As for the second case, note that

$$\Pr(Y^* = y|E=2) = \frac{p_i+p_j}{\beta-1} \quad \text{for one } y \text{ for each } 1 \le i < j \le \beta. \tag{15}$$

The entropy of this distribution satisfies

$$H(Y^*|E=2) \ge H(X) + \log(\beta-1) - \log 2. \tag{16}$$

To see this, we can take the convex combination of two reorderings of copies of the distribution of X. In particular, define two $\beta(\beta-1)$-dimensional distributions

$$\boldsymbol{P}_a = \frac{1}{\beta-1}[p_1\, p_1 \ldots p_1\, p_2\, p_2 \ldots p_2\, \ldots\, p_\beta\, p_\beta \ldots p_\beta], \tag{17}$$

and

$$\boldsymbol{P}_b = \frac{1}{\beta-1}[p_2\, p_3 \ldots p_\beta\, p_1\, p_3 \ldots p_\beta\, \ldots\, p_1\, p_2 \ldots p_{\beta-1}]. \tag{18}$$

That is, $\boldsymbol{P}_a$ has $\beta-1$ copies of p_1, then $\beta-1$ copies of p_2 and so forth. On the other hand, $\boldsymbol{P}_b$ 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 $H(\boldsymbol{P}_a) = H(\boldsymbol{P}_b) = H(X) + \log(\beta-1)$. On the other hand, $H((\boldsymbol{P}_a + \boldsymbol{P}_b)/2) = H(Y^*|E=2) + \log 2$. The inequality (16) can be seen by combining these equalities with the inequality $H((\boldsymbol{P}_a+\boldsymbol{P}_b)/2) \geq (H(\boldsymbol{P}_a)+H(\boldsymbol{P}_b))/2$ by the concavity of entropy. Combining the minimum entropy property of Y^* and (14)-(16) above, we conclude the proof of Lemma 1 by computing

$$\begin{aligned} H(X+Z) &\geq H(Y^*) && (19)\\ &\geq \log\alpha + H(X) + \frac{\beta-1}{\alpha}\log 2. && (20) \end{aligned}$$

□

In the next subsection we will see that this lower bound can in fact hold with equality.

2.3 Tightness in the EAP

Combining the upper bound of (1) with the EAP (7), we see that for X and Z satisfying the conditions of the EAP lemma

$$H(X)+H(Z) \geq H(X+Z) \;\; \geq \;\; H(X)+H(Z) - \frac{|\mathcal{X}|-1}{|\mathcal{Z}|}\text{bit}. \tag{21}$$

We now show that the lower bound above (the EAP) can be tight while the upper bound above can be made as tight as desired by choosing a large enough set with unique differences. Let $\mathcal{Z}$ be a set with unique differences and let Z_1 and Z_2 be independent random variables each uniformly distributed on $\mathcal{Z}$. Furthermore, let $Y_+ = Z_1 + Z_2$ and $Y_- = Z_1 - Z_2$ (with addition and subtraction in the underlying group $\mathcal{G}$). The EAP applies to Y_+ and Y_- since Z_1 is aperiodic noise and the support of Z_2 (or $-Z_2$) is less than $|\mathcal{Z}|+1$. We first compute that

$$H(Y_+) = 2\log|\mathcal{Z}| - \left(1 - \frac{1}{|\mathcal{Z}|}\right), \tag{22}$$

which is exactly the lower bound in the EAP. To see (22), notice that if $z_{1,1}+z_{1,2} = z_{2,1}+z_{2,2}$ with $z_{i,j}\in\mathcal{Z}$, then unique differences implies that $\{z_{1,1},z_{1,2}\} = \{z_{2,1},z_{2,2}\}$ and thus $\Pr(Y_+ = y) \leq 2/|\mathcal{Z}|^2$ for all y. Indeed, $\Pr(Y_+ = y) = 2/|\mathcal{Z}|^2$ for $|\mathcal{Z}|(|\mathcal{Z}|-1)/2$ values of y (once for each distinct pair in $\mathcal{Z}$) and $\Pr(Y_+ = y) = 1/|\mathcal{Z}|^2$ for $|\mathcal{Z}|$ values of y (once for each element of $\mathcal{Z}$). On the other hand,

$$H(Y_-) = \left(2 - \frac{1}{|\mathcal{Z}|}\right)\log|\mathcal{Z}|, \tag{23}$$

which can be made arbitrarily close to the upper bound in the EAP by choosing sufficiently large set $\mathcal{Z}$. To see (23), we note that $\Pr(Y_- = 0) = \frac{1}{|\mathcal{Z}|}$ and that $\Pr(Y_- = y) = \frac{1}{|\mathcal{Z}|^2}$ for all non-zero y with non-zero probability.

We have not only shown tightness in the EAP, but we have shown that the difference between the sum-entropy and the difference-entropy of two i.i.d. random variables Z_1 and Z_2 can be arbitrarily close to 1 bit:

$$|H(Z_1+Z_2) - H(Z_1-Z_2)| \approx 1 \; \textit{bit}.$$

2.4 Extension to Un-restricted Set Size

One drawback of the EAP lemma is the restriction on the support of X. Since the size of a set with unique differences is upper bounded by $\approx \sqrt{|\mathcal{G}|}$ (achieved by the case of a planar difference set), this also limits the EAP to X's with support size $\approx \sqrt{|\mathcal{G}|}$. Nevertheless, Lemma 2 below gives a lower bound on the entropy of $X + Z$ which can be optimized for a *specific* distribution of X.

In order to circumvent the restriction on the size of $\mathcal{X}$ in the EAP, we consider a quantization $K(\cdot)$ of the group $\mathcal{G}$. This quantizer groups $|\mathcal{Z}|$ points of $\mathcal{G}$ together according to

$$K(g) = \left\lfloor \frac{\pi(g)}{|\mathcal{Z}|} \right\rfloor, \quad g \in \mathcal{G}, \tag{24}$$

where $\lfloor \cdot \rfloor$ is the round-down operation, and $\pi(\cdot)$ is some permutation of $\mathcal{G}$ to be optimized later. We refer to the groupings formed by $K(\cdot)$ as K-bins, e.g., the set $K^{-1}(1)$ is the first K-bin.

Lemma 2. *(***Extended EAP***) Let Z be aperiodic noise as in Lemma 1. For any independent random variable X and any quantizer $K(\cdot)$ as above,*

$$H(X + Z) \quad \geq \quad H(X) + H(Z) - H(K(X)) - 1bit. \tag{25}$$

The best (tightest) lower bound is achieved by a quantizer $K(\cdot)$ which minimizes the entropy of $K(X)$. By majorization theory [14], a permutation $\pi(\cdot)$ in (24) for which the probabilities $\Pr(X = \pi(i))$ for $i = 1, \ldots, |\mathcal{G}|$ are in a non-increasing (or non-decreasing) order results in a distribution of $K(X)$ which majorizes that of any other permutation; hence this $K(\cdot)$ is the minimum-entropy quantizer. Note that for such a permutation, if $|\mathcal{X}| \leq |\mathcal{Z}|$, then $H(K(X)) = 0$, and Lemma 2 coincides with Lemma 1.

Proof. Using the chain rule for the joint entropy $H(X + Z, K(X))$, we have

$$\begin{aligned} H(X+Z) &= H(X+Z|K(X)) + H(K(X)) - H(K(X)|X+Z) \\ &\geq H(X|K(X)) + H(Z) - \frac{|\mathcal{Z}|-1}{|\mathcal{Z}|} + H(K(X)) - H(K(X)|X+Z) \\ &\geq H(X) + H(Z) - H(K(X)) - 1\text{bit}. \end{aligned} \tag{26}$$

Here, the first inequality follows by Lemma 1 since, given $K(X) = k$, the size of the support of X is at most $|\mathcal{Z}|$, so the conditions of the lemma are satisfied; the second inequality follows since $H(K(X)|X + Z) \leq H(K(X))$, and since $K(X)$ is a deterministic function of X so $H(X|K(X)) + H(K(X)) = H(X)$. □

3 WDP Capacity with Strong Interference

We turn to the second part of this paper, which deals with cases of high rate loss in side information problems. The next two sections are devoted to the "dirty-paper" channel. We start with a definition of the problem, followed by a simple characterization of the WDP capacity in the case of general unknown noise and strong interference. In the next section we then use this characterization for the case of aperiodic unknown noise, and apply the EAP to derive an upper bound for the resulting WDP capacity.

We consider a discrete version of WDP in which addition is done modulo some given integer L, with results in the alphabet $\mathcal{A}_L = \{1, \ldots, L\}$:

$$Y^n = x^n + S^n + Z^n$$

where additions are modulo L. 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, and S^n is known at the encoder prior to encoding. 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 $\mathcal{C} \subset \mathcal{A}_L$:

$$x_i \in \mathcal{C}.$$

A WDP system consists of two components. The first is a rate-R blocklength-n encoder, which takes a message M consisting of nR bits and the interference sequence S^n and creates the constrained input sequence as $X^n = f_n(M, S^n)$, where each component X_i is in the constrained set $\mathcal{C}$. The other component is a decoder, which estimates the message from the output sequence, $\hat{M} = g_n(Y^n)$. A rate R is achievable if there exists a sequence of rate-R encoders and decoders such that the probability of decoding error, $P_e = \Pr\{g_n(Y^n) \neq M\}$, can be made as small as desired. The capacity is the supremum of all achievable rates and is written $C_{\mathrm{WDP}}(L, \mathcal{C}, P_Z, P_S)$.

The comparison zero-interference system uses the same noise distribution P_Z and input constraint set $\mathcal{C}$, but the interference does not play a role, i.e., $Y = x + Z$. Thus, for this system, the encoder is of the form $X^n = f_n(M)$. The zero-interference capacity is defined as above and is written $C_{\mathrm{ZI}}(L, \mathcal{C}, P_Z)$. It is straightforward to see that (see, e.g., [8]),

$$C_{\mathrm{ZI}}(L, \mathcal{C}, P_Z) = \sup_{P_X} H(X + Z) - H(Z), \tag{27}$$

where the maximization is over distributions P_X on $\mathcal{C}$, and where X is independent of Z.

By the general Gelfand-Pinsker capacity formula, [11], the WDP capacity is given by

$$C_{\mathrm{WDP}}(L, \mathcal{C}, P_Z, P_S) = \sup_{U,X} \Big\{ I(U;Y) - I(U;S) \Big\} \tag{28}$$

with $Y = X + S + Z$, where the maximization is over all random variables (X, U) such that (S, X, U) are independent of Z, and X is in $\mathcal{C}$. In fact, as shown in [11] it is enough to maximize over U and functions f such that $X = f(U, S)$ is in $\mathcal{C}$. We call such (X, U)'s *admissible pairs*. This general formula for C_{WDP} involves maximization over a conditional distribution ($P_{U|S}$) and a function ($f(\cdot)$). But the assumption of strong interference allows us to derive a simpler expression, involving an *un*conditional distribution and a function.

Lemma 3. (Strong Interference Capacity) *For P_S uniformly distributed on $\mathcal{A}_L$, i.e., strong interference,*

$$C_{WDP}(L, \mathcal{C}, P_Z, P_S) = \sup_{P_V,\ Q(\cdot) \in \mathcal{Q}_{\mathcal{C}}} \Big\{ H(V) - H(Q(V) + Z) \Big\}, \tag{29}$$

where the maximization is over distributions P_V on $\mathcal{A}_L$ such that V is independent of Z, and over functions $Q : \mathcal{A}_L \to \mathcal{A}_L$ belonging to the set

$$\mathcal{Q}_{\mathcal{C}} = \{Q : \ Q(v) - v \in \mathcal{C} \textit{ for all } v \in \mathcal{A}_L\}.$$

The proof is given below after we make a few comments. Equation (29) is similar to the formula for the capacity if the side information can only be used *causally* (the "dirty-tape" problem) [10, Thm. 1]. In the causal case the capacity is the same as (29), except that P_V is constrained to be uniform over $\mathcal{A}_L$:

$$C_{\mathrm{WDP}}^{causal} = H(U) - \inf_{Q \in \mathcal{Q}_{\mathcal{C}}} H(Q(U) + Z), \tag{30}$$

where $U \sim Unif(\mathcal{A}_L)$. It is interesting to note that for a continuous channel in the extreme case of high signal-to-noise ratio, the optimum function $Q(\cdot)$ in either of the capacity expressions above tends to take discrete values hence can be regarded as a quantizer. Thus a good choice for Q in (29) and (30) is roughly a minimum entropy quantizer [10].

The negative term in (29) is the entropy of the sum of independent random variables; if Z is a difference-set noise, then by the EAP this term is large and almost cancels out completely the first entropy term. This would be the key for showing small WDP capacity in the next section.

Proof. Every admissible pair (U, X) in (28) gives a lower bound for C_{WDP}. Specifically, let us substitute $U = (S - V)$ and $X = Q(V) - V$, where $V, Q(v)$ achieve the maximum in (29), and where V is statistically independent of (S, Z). By the conditions of the maximization (29), $X \in \mathcal{C}$, so (U, X) is admissible pair as desired. We also have for this pair and *any* P_S

$$\begin{aligned} I(U;Y) - I(U;S) &= H(U|S) - H(U|Y) & (31)\\ &= H(S-V|S) - H(S-V|Q(V) - V + S + Z) & (32)\\ &= H(V) - H(Q(V) + Z|Q(V) - V + S + Z) & (33)\\ &\geq H(V) - H(Q(V) + Z) & (34)\\ &= C^*(L, \mathcal{C}, P_Z), & (35) \end{aligned}$$

where $C^*(L, \mathcal{C}, P_Z)$ denotes the right hand side of (29), and where the inequality above follows since conditioning reduces entropy (with equality for a uniform S). This establishes the "direct" part, i.e., for *any* P_S

$$C_{\mathrm{WDP}}(L, \mathcal{C}, P_Z, P_S) \geq C^*(L, \mathcal{C}, P_Z). \tag{36}$$

We now turn to prove the reverse inequality for P_S uniform over $\mathcal{A}_L$. For arbitrary P_V and $Q(\cdot)$, let us define

$$F(P_V, Q) = H(V) - H(Q(V) + Z) \tag{37}$$

where V is independent of Z, so the RHS of (29) is given by

$$C^*(L, \mathcal{C}, P_Z) = \sup_{P_V, Q \in \mathcal{Q}_{\mathcal{C}}} F(P_V, Q). \tag{38}$$

For any admissible pair $(U, X = f(U, S))$ in (28), we can write

$$\begin{aligned} I(U;Y) - I(U;S) &= \Big\{H(S|U) - H(Y|U)\Big\} - \Big\{H(S) - H(Y)\Big\} & (39)\\ &\leq H(S|U) - H(Y|U) & (40)\\ &= E_U\Big\{H(S|U=u) - H(f(u,S) + S + Z|U=u)\Big\} & (41)\\ &= E_U\Big\{F(P_{S|U=u}, \tilde{Q}_u)\Big\} & (42)\\ &\leq \sup_{P_V, Q \in \mathcal{Q}_{\mathcal{C}}} F(P_V, Q), & (43) \end{aligned}$$

where the function $\tilde{Q}_u$ in (42) is defined as $\tilde{Q}_u(S) = f(u,S) + S$. The first inequality above follows since S is uniform, so its entropy ($\log L$) is the largest among all random variables in $\mathcal{A}_L$. To justify the second inequality, note that conditioned on any $U = u$, we have that $\tilde{Q}_u(S) - S$ is in $\mathcal{C}$; thus we can think of $(S, \tilde{Q}_u)$ as an admissible pair (V, Q) in the right hand side of (29), so $F(P_{S|U=u}, \tilde{Q}_u)$ is bounded for every u by its maximum $C^*(L, \mathcal{C}, P_Z)$ of (38). By the Gelfand-Pinsker formula (28), this establishes the "converse part" $C_{\mathrm{WDP}}(L, \mathcal{C}, P_Z, P_S) \leq C^*(L, \mathcal{C}, P_Z)$ for a uniform S. Combining with the direct part (36), this proves the lemma. □

4 Rate Loss in WDP in the Presence of Difference-Set Noise

Aperiodic (noise) distributions arise as extreme cases in the investigation of the rate loss in side-information problems. Two side-information settings which attracted much attention are channel coding with additive interference known at the encoder (or in its popular name "writing on dirty paper") [11, 6], and lossy source coding with side-information at the decoder (called "the Wyner-Ziv problem") [18]. In this section we focus on the former setting. We use the EAP above to show that WDP in the presence of planar difference-set noise suffers a large capacity loss with respect to the zero-interference case.

Our main result, regarding the WDP and the zero-interference capacities defined in Section 3 above, is summarized in Theorem 1 below.

Theorem 1. (Upper Bound for $C_{\mathbf{WDP}}$) *Let S be strong interference (i.e., P_S is uniformly distributed on $\mathcal{A}_L$). For any aperiodic unknown noise (i.e., P_Z uniform on a set $\mathcal{Z}$ with unique differences (3)), and for any constraint set $\mathcal{C}$,*

$$C_{WDP} \leq 1 + \log\left(\frac{L}{\alpha^2} + \frac{\beta}{\alpha}\right) bit, \tag{44}$$

where $\beta = |\mathcal{C}|$ and $\alpha = |\mathcal{Z}|$. In particular, if Z is a planar difference-set noise (i.e., $\mathcal{Z}$ is a planar difference-set) and $\beta \leq \alpha$, then

$$C_{WDP} \leq 2 bit. \tag{45}$$

On the other hand,

$$C_{ZI} \geq \log\beta - \frac{\beta - 1}{\alpha} \geq \log\beta - 1 bit \tag{46}$$

provided $\beta \leq \alpha + 1$.

Hence the capacity loss in the presence of planar difference-set noise is at least $\log\beta - 3$ bit (provided $\beta = |\mathcal{C}| \leq \alpha = |\mathcal{Z}|$). Since we can choose α (and hence β) arbitrarily large, this loss can be arbitrarily large, and approach 100 % of the available capacity.

Proof. The bound on the zero-interference capacity follows directly from (27) and Lemma 1 by making the input distribution P_X uniform over the constraint set. In this case,

$$\begin{aligned} C_{\mathrm{ZI}}(L, \mathcal{C}, P_Z) &\geq H(X) + H(Z) - \frac{\beta - 1}{\alpha} - H(Z) && (47) \\ &= \log\beta - \frac{\beta - 1}{\alpha}. && (48) \end{aligned}$$

We now turn to the WDP capacity with strong interference (29) and provide an upper bound on $H(V) - H(Q(V) + Z)$ for any allowable distribution P_V and quantizer $Q(\cdot) \in \mathcal{Q}_{\mathcal{C}}$. By the extended-EAP (Lemma 2), we see that

$$H(Q(V) + Z) \geq H(Q(V)) + H(Z) - H(K(Q(V))) - 1 \text{ bit}$$

for any grouping $K(\cdot)$ into bins of size α. We shall optimize the permutation $\pi(\cdot)$ used in the definition of $K(\cdot)$ in (24) later. Since $H(V) = H(Q(V)) + H(V|Q(V))$, this implies that

$$H(V) - H(Q(V) + Z) \leq H(V|Q(V)) + H(K(Q(V))) - H(Z) + 1\text{bit}. \tag{49}$$

In order to upper bound (49), let us write the probability of each value of Q as

$$p_i = \Pr(Q = q_i) \tag{50}$$

and let us also define

$$r_i = \frac{|Q^{-1}(q_i)|}{L}, \tag{51}$$

which is the probability of q_i if P_V were uniform over $\mathcal{A}_L$. Let us next define $p_i^{K\text{-mean}} = \Pr(K(Q(V)) = k)/\alpha$ where k is the bin index of q_i, i.e.,

$$p_i^{K\text{-mean}} = \frac{1}{\alpha} \sum_{i' \in K^{-1}(K(q_i))} p_{i'}. \tag{52}$$

That is, $p_i^{K\text{-mean}}$ is the average of the probabilities of the points of Q in the same K-bin as q_i. Note that $p_i^{K\text{-mean}}$ is identical for all i in the same K-bin. We can compute the distribution of $K(Q(V))$ as

$$\begin{aligned}
\Pr(K(Q(V)) = k) &= \sum_{i:q_i \in K^{-1}(k)} p_i && (53)\\
&= \sum_{i:q_i \in K^{-1}(k)} p_i^{K\text{-mean}} && (54)\\
&= \alpha p_i^{K\text{-mean}}, \ \forall i : q_i \in K^{-1}(k), && (55)
\end{aligned}$$

where the last equality follows since $p_i^{K\text{-mean}}$ is constant for all q_i in the same K-bin. It follows from (53) and (55) that

$$H(K(Q(V))) = \sum_i p_i \log \frac{1}{\alpha p_i^{K\text{-mean}}}. \tag{56}$$

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

$$\begin{aligned}
H(V|Q(V)) &\leq \sum_i p_i \log |Q^{-1}(q_i)| && (57)\\
&= \sum_i p_i \log L r_i. && (58)
\end{aligned}$$

The bound follows since given $Q(V) = q_i$, the conditional distribution that maximizes the entropy is uniform over the Lr_i values in $Q^{-1}(q_i)$. Combining (49) with the expressions above and substituting $H(Z) = \log \alpha$, we see that

$$H(V) - H(Q(V) + Z) \leq \sum_i p_i \log \left(\frac{r_i}{p_i^{K\text{-mean}}} \right) + \log \left(\frac{L}{\alpha^2} \right) + 1 \text{ bit}. \tag{59}$$

Using the log-sum inequality [8], we bound the first term above by

$$\sum_i p_i \log \frac{r_i}{p_i^{K\text{-mean}}} = \sum_i p_i \log \frac{\left(\frac{p_i \cdot r_i}{p_i^{K\text{-mean}}}\right)}{p_i} \tag{60}$$

$$\leq \log \left(\sum_i \frac{p_i \cdot r_i}{p_i^{K\text{-mean}}} \right). \tag{61}$$

To bound the sum inside the log, consider a permutation $\pi(\cdot)$ in the definition of $K(\cdot)$ in (24) so that $r_{\pi(i)}$ is a non-increasing sequence. Although this choice of $\pi(\cdot)$ does not necessarily minimize the entropy of $K(Q(V))$ in (49) (see the discussion after Lemma 2), it will allow us to get the desired bound. With this permutation, the first K-bin has the α largest cells of Q (= the α most probable q_i's according to the sequence $\boldsymbol{r}$), the second K-bin has the second α largest cells of Q, 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 $Q(v) - v \in \mathcal{C}$ for all v, and thus each quantization cell $Q^{-1}(q)$ contains at most β elements. We can write the sum inside the log as

$$\sum_{i=1}^{\infty} \frac{p_i \cdot r_i}{p_i^{K\text{-mean}}} = \alpha \sum_{n=0}^{\infty} \sum_{i=n\alpha+1}^{(n+1)\alpha} \frac{p_{\pi(i)}}{\alpha p_{\pi(i)}^{K\text{-mean}}} \cdot r_{\pi(i)} \tag{62}$$

$$\leq \alpha \Big(r_{\pi(1)} + \frac{1}{\alpha} \sum_{i=1}^{\infty} r_{\pi(i)} \Big) \tag{63}$$

$$\leq \alpha \frac{\beta}{L} + 1. \tag{64}$$

The first inequality follows because for each n the inner sum can be viewed as a weighted average over the r's in a window of size α after the permutation; since the sequence $r_{\pi(1)}, r_{\pi(2)}, \ldots$ is non-increasing, this weighted average can be upperbounded for all $n \geq 1$ by a uniform average of the r's in the previous window, and for $n = 0$ (the first window) by $r_{\pi(1)}$. The second inequality follows since as explained above $r_{\pi(1)} \leq \beta/L$, and the sum of r_i's is one. Combining (59), (61) and (64), we get that for any P_V and admissible function $Q(\cdot)$,

$$H(V) - H(Q(V) + Z) \leq \log \left(\alpha \frac{\beta}{L} + 1 \right) + \log \left(\frac{L}{\alpha^2} \right) + 1 \tag{65}$$

$$= \log \left(\frac{\beta}{\alpha} + \frac{L}{\alpha^2} \right) + 1 \text{ bit}. \tag{66}$$

This holds for any aperiodic noise, i.e., for any $\mathcal{Z}$ with unique differences, and for any constraint set $\mathcal{C}$. In particular, if $\mathcal{Z}$ is a planar difference-set, then $L \leq \alpha^2$ (see (5)), and in view of Lemma 3 the desired 2 bit bound on C_{WDP} follows. □

5 Rate Loss in the Wyner-Ziv Problem Revisited

The Gelfand-Pinsker (channel coding with side-information) setting, [11], and the Wyner-Ziv (source coding with side-information) setting, [18], are widely considered as "dual", in accordance with Shannon's concept of source-channel coding duality. See, e.g., [7, 21, 15] and the references therein. However, there are several indications to the contrary. One aspect in which the behavior of the two settings seems different is the rate loss relative to the complete information case. Specifically, in the Wyner-Ziv problem, for r-th power distortion measures the extra rate relative to the case where the side information is available to both the encoder and the decoder is upper bounded by a small constant, e.g., 0.5 bit for $r = 2$ (MSE) [19]. On the other hand, in the previous section we saw that in writing on dirty paper, the capacity loss for not knowing the interference at the decoder can be *unbounded* if the noise is irregular, e.g., if the noise is uniform over a difference set.

Does this seemingly very different behavior of the rate loss in the two settings contradicts source/channel duality? In fact, it turns out that this "duality gap" is a matter of interpretation. Specifically, the rate loss in the Wyner-Ziv problem may be unbouded for an *irregular distortion measure*. Thus, the distortion measure in the Wyner-Ziv problem plays a role dual to the *noise* in writing on dirty paper, perhaps against the more natural view of the distortion measure as the counterpart of the channel input cost function. This observation provides another angle for the general framework for source/channel duality developed for side information problems [7, 15].

We consider below a special case of the Wyner-Ziv problem called "quantizing encrypted source" [13]. For this special setting we construct an example for unbounded rate loss, using an irregular distortion measure based on a difference set.

5.1 Quantizing Encrypted Source

We consider the situation of quantizing an encrypted source $X+K$, where K (the "key") is statistically independent of the clean source X and is known only to the decoder. The decoder outputs the reconstruction $\hat{X}$, which is required to satisfy a given distortion constraint relative to X. We assume that X, K and $\hat{X}$ belong to an additive group $\mathcal{G}$, and define the quantization error as the difference $\hat{X} - X$, where all subtractions and additions are modulo the group $\mathcal{G}$. Since the distortion measure is only a function of the quantization error, i.e., $d(x, \hat{x}) = d(x - \hat{x})$, and since K is known at the decoder, reconstructing the clean source as $\hat{X}$ with some distortion is equivalent to reconstructing the encrypted source as $\hat{X} + K$ with the same distortion. Therefore, this setting is equivalent to the standard Wyner-Ziv problem if we view $X + K$ as the source and K as the side information.

We further assume that K is uniform over $\mathcal{G}$, and the decoder has access to another correlated source W, where K is statistically independent of (X, W). As shown in [19], the Wyner-Ziv rate-distortion function of $X + K$ when (K, W) is known at the decoder is equal to the *additive rate-distortion function* of X given W:

$$R^{WZ}_{X+K|K,W}(D) = R^{add}_{X|W}(D)$$

where

$$R^{add}_{X|W}(D) = \min_{N} I(X; X + N|W).$$

Here, $I(X; X+N|W)$ denotes the conditional mutual information between X and $X+N$ given W, and the minimization is over all *additive* noises N (i.e., N independent of (X, W)) satisfying $E\{d(f(W, X + N) - X)\} \leq D$ for some function $f(\cdot)$ (i.e., such that the best estimate of X from $X + N$ and W satisfies the distortion constraint D).

The reference setting is when both the encoder and the decoder have access to K and W. The optimum performance is characterized in this case by the conditional rate-distortion function, which is given by (see [19]),

$$R_{X+K|K,W}(D) = R_{X|W}(D) = \min_U I(X;U|W)$$

where the minimization is over all random variables U (jointly distributed with (X,W)) who satisfy the distortion constraint D.

Clearly $R^{WZ}_{X+K|K,W}(D)$ cannot be smaller than $R_{X+K|K,W}(D)$. It was shown in [19] that the loss in the Wyner-Ziv setting, defined as the excess rate of the Wyner-Ziv rate-distortion function above the conditional rate-distortion function, is bounded by a universal constant which is independent of the source/side-infromation distribution. This constant is given by the "min-max capacity"

$$C^* = C^*(d,D) = \min_N \max_U I(U;U+N) \tag{67}$$

where U and N are statistically independent, and the maximization and minimization are over all U and N who satisfy $Ed(U) \leq D$ and $Ed(N) \leq D$. As mentioned above, for real alphabets and r-th power distortion measures C^* varies between 0.5 and 1 bit. For a (finite) additive group $\mathcal{G}$, the min-max capacity is upper bounded for any difference distortion measure d by

$$C^* \leq 0.5 \log |\mathcal{G}|. \tag{68}$$

To see that, note that $H(U+N)$ is upper bounded by $H(U)+H(N)$ and also by $\log|\mathcal{G}|$; while under the constraint $E\{d(U)\} \leq D$, the entropy $H(U)$ is upper bounded by the maximum entropy [8]

$$H_{max}(d,D) = \max_{Z:\ E\{d(Z)\}\leq D} H(Z).$$

It follows that for any U, $\min_N I(U;U+N)$ is upper bounded by the minimum between $H_{max}(d,D)$ and $\log|\mathcal{G}| - H_{max}(d,D)$, which is clearly at most $0.5\log|\mathcal{G}|$. Equality in (68) holds if and only if there exist (maximum-entropy) random variables U and N satisfying the distortion constraint, for which $H(U+N) = H(U) + H(N) = \log|\mathcal{G}|$.

We next construct an example of a source and distortion measure for which both C^* and the rate loss approach $0.5\log|\mathcal{G}|$. In this example W is binary, and the source X is a mixture of two components: given $W=1$ it is uniform over a set $\mathcal{Z} \subset \mathcal{G}$, while given $W=2$ it is uniform over the whole alphabet $\mathcal{G}$. The source mode indicator W equals 2 with probability ϵ and 1 with probability $1-\epsilon$. (Note that due to the encryption, the encoder cannot guess the mode of the source by observing $X+K$.) The distortion constraint is such that the quantization error $\hat{X} - X$ must belong to the set $-\mathcal{Z}$ (the reflection of $\mathcal{Z}$), that is, the distortion measure is of the form

$$d(x,\hat{x}) = \begin{cases} 0, & \hat{x} - x \in -\mathcal{Z} \\ \infty, & \hat{x} - x \notin -\mathcal{Z}. \end{cases}$$

The conditional rate-distortion function of the source is zero for $W=1$ (by mapping every value of X to zero), while for $W=2$ it is $\log(|\mathcal{G}|) - \log(|\mathcal{Z}|)$ (realized by an additive noise test channel with noise uniform over $-\mathcal{Z}$). Thus

$$R_{X|W}(D) = \epsilon\Big[\log|\mathcal{G}| - \log|\mathcal{Z}|\Big].$$

As for the Wyner-Ziv function in this setup, since the mixture source X has a uniform component (assuming $\epsilon > 0$), the noise in the test channel above **must** belong to the set $-\mathcal{Z}$ (otherwise the distortion in the case $W = 2$ would be infinite). Thus, the best possible noise is uniform over $-\mathcal{Z}$, and we get

$$R^{add}_{X|W}(D) = (1-\epsilon)\Big[H(U_Z - U'_Z) - \log|\mathcal{Z}|\Big] + \epsilon\Big[\log|\mathcal{G}| - \log|\mathcal{Z}|\Big]$$

where U_Z and U'_Z are statistically independent and uniform over the set $\mathcal{Z}$. Comparing the two rate-distortion functions, we conclude that the rate loss for not having (K, W) at the encoder is given in this case by

$$R^{WZ}_{X+K|K,W}(D) - R_{X|W}(D) = (1-\epsilon)\Big[H(U_Z - U'_Z) - \log|\mathcal{Z}|\Big].$$

Now, if we choose $\mathcal{Z}$ to be a planar difference set, then for a large group we have $|\mathcal{Z}| \approx \sqrt{|\mathcal{G}|}$ (see (5)), and from the asymptotic tightness of the EAP (Section 2.3), $H(U_Z - U'_Z) \approx H(U_Z) + H(U'_Z) \approx \log|\mathcal{G}|$. We thus proved the following theorem.

Theorem 2. (Maximum rate loss in WZ coding) *We can choose ϵ, $\mathcal{G}$ and $\mathcal{Z}$ in the example above such that the rate loss approaches the min-max capacity bound (68). Specifically, the conditional rate distortion function $R_{X|W}(D)$ is arbitrarily small, while the Wyner-Ziv rate distortion function $R^{WZ}_{X+K|K,W}(D)$ is arbitrarily close to $0.5\log|\mathcal{G}|$.*

Hence, the rate loss in the Wyner-Ziv setting above is close to 100% of the rate, similarly to the situation described in Theorem 1 regarding the capacity loss in WDP.

6 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. Admittedly, this loss seems to rely quite heavily on the EAP, which in return relies on the special structure of DSN. Yet on the positive side, we can expect that for more "standard" noise models the loss for writing on dirty paper would be small. Furthermore, the machinery in the proof (in particular Lemma 3) can assist in designing WDP schemes for general noise distributions.

There are several other interesting applications of difference sets and DSN. One is their role in the "dual" problem of source coding with side information (the Wyner-Ziv problem), which we briefly investigated in the last section. We can also use DSN to construct examples of large capacity loss in the broadcast setting, e.g., for the loss of sum-rate due to lack of cooperation between the receivers (for background, see [12]). Another use of difference sets is in constructing expander graphs. For example, consider the bi-partite graph $G \in \mathcal{G} \times \mathcal{G}$, in which $(x, y) \in G$ if $y - x \in \mathcal{Z}$. Then, every input is connected to $\alpha = |\mathcal{Z}|$ outputs, and for a difference set $\mathcal{Z}$, every set of $\beta \leq \alpha + 1$ inputs is connected to at least $\alpha\beta/2$ outputs (see the corollary in (9) after the EAP lemma). Finally, the EAP 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 to be uniform over $[0, 1]$, and let the noise Z be DSN (in the alphabet $\mathcal{A}_L$) divided by L plus a random variable uniform over $[0, 1/L]$.

Then, for a peak input constraint, $0 \leq x \leq \beta/L$, 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 [20].

Acknowledgement

The authors are indebted to Dave Forney for referring them to difference set theory. They also thank Yuval Kochman for simplifying the proof of Theorem 1. Thanks are also due to Greg Wornell, Uri Erez and the students at the DSP and SIA labs at MIT for many fruitful discussions.

References

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

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

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

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

[5] A. Cohen and R. Zamir, “Unbounded loss in writing on dirty paper is possible,” in American Mathematical Society (2004), DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Advances in Network Information Theory, P. Gupta, G. Kramer, A. van Wijngaarden Editors, pp. 79-86, March 2003.

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

[7] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information”, *IEEE Trans. Info. Theory*, IT-48:1629–1638, June 2002.

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

[9] 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.

[10] U. Erez, S. Shamai, and R. Zamir, “Capacity and lattice-strategies for cancelling known interference,” *IEEE Trans. Info. Theory*, pp. 3820–3833 Nov. 2005.

[11] 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.

[12] E. Haim and R. Zamir, “Broadcast channels and input-cost side information”, in the 24th IEEE CONVENTION OF ELECTRICAL AND ELECTRONICS ENGINEERS IN ISRAEL November 15-17, 2006 Eilat, Israel.

[13] P. Ishwar, V. Prabhakaran, and K. Ramchandran, “Compressing encrypted sources using side-information coding”, in *ISIT 2004*, (Chcago, Illinois), June 2004.

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

[15] S. S. Pradhan and K. Ramchandran, “A comprehensive view of duality in multiuser source and channel coding”, in *DIMACS Workshop on Network Information Theory*, (Rutgers, NJ), Mar. 2003.

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

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

[18] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory*, IT-22:1–10, Jan., 1976.

[19] R. Zamir, “The rate loss in the Wyner-Ziv problem,” *IEEE Trans. Inform. Theory*, pp. 2073-2084, November 1996.

[20] R. Zamir, “The half a bit loss of robust source/channel codebooks,” in *Proc. of the Info. Theory Workshop*, pp. 123, Bangalore, India, Oct. 2002.

[21] R. Zamir, S. Shamai, and U. Erez. Nested Linear/Lattice Codes for Structured Multiterminal Binning. *IEEE Trans. Info. Theory*, IT-48:1250–1276, June 2002.

List of Tables

List of Figures

List of Footnotes

1. Logarithms throughout the paper are base 2.

2. The "only if" part holds for a certain large class of transmitters.

3. By the unique differences property, every ordered pair (z_1, z_2) corresponds to a *different* difference $d = z_1 - z_2$; thus the number of possible (non-zero) differences in $\mathcal{G}$, which is $|\mathcal{G}| - 1$, is greater than or equal to the number of ordered pairs in $\mathcal{Z}$, which is $|\mathcal{Z}|(|\mathcal{Z}| - 1)$. The reverse inequality holds if an ordered pair in $\mathcal{Z}$ exists for *every* (non-zero) difference d, as indeed implied by (4).

4. To see why, suppose we start with the set $\mathcal{Y}_1$ and add the sets $\mathcal{Y}_2, \ldots, \mathcal{Y}_\beta$ sequentially, while noting at each stage which point of $\mathcal{Y}_1$ (if any) intersects the new $\mathcal{Y}_i$. (We say that the point is "covered" by the new $\mathcal{Y}_i$.) If each new intersection occurs at a different point, then there will be at most $\beta - 1$ points of $\mathcal{Y}_1$ covered, and therefore at least $\alpha - \beta + 1$ points which are in $\mathcal{Y}_1$ alone. If there are m "re-coverings" in this process (i.e., m is the number of stages in which the new $\mathcal{Y}_i$ intersects a point of $\mathcal{Y}_1$ which is already covered), then there will be at least $\alpha - \beta + 1 + m$ points which are in $\mathcal{Y}_1$ alone. Since we assumed that $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ intersect at y_0, there must be at least $n - 2$ re-coverings; substituting $m \geq n - 2$ and $\beta \leq \alpha + 1$, we get that there are at least $\alpha - \beta + 1 + m \geq n - 2$ points which are in $\mathcal{Y}_1$ alone. Since the same argument applies to all the sets $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, each one of them must have at least $n - 2$ points which are not in any other $\mathcal{Y}_i$.

Aaron S. Cohen Bio

Aaron S. Cohen received the S.B. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 1997 and 2001, respectively. From 2001 to 2002, he was a Postdoctoral Associate in the Signals, Information, and Algorithms Laboratory at MIT. From 2002 to 2004, he was a Visiting Research Associate in the Pattern Theory Group at Brown University, Providence, RI. Since 2005, he has been with Knight Capital Group in Purchase, NY, where he is currently a Vice President in the Electronic Trading Group.

Rami Zamir Bio

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. In 2002 he spent a Sabbatical year at MIT. 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, data compression, random processes, 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. He served as an Associate Editor for Source Coding in the IEEE transactions on Information Theory (2001-2003), and headed the Information Theory Chapter of the Israeli IEEE society (2000-2005). His research interests include information theory, communication and remote sensing systems, and signal processing.