Capacity and Lattice Strategies for Canceling Known Interference

Uri Erez, Shlomo Shamai (Shitz) and Ram Zamir

Submitted August 2002, Revised October 2004 *

Abstract

We consider the generalized dirty-paper channel $Y = X+S+N,\ E\{X^2\} \leq P_X$, where N is not necessarily Gaussian, and the interference S is known causally or noncausally to the transmitter. We derive worst-case capacity formulas and strategies for "strong" or arbitrarily varying interference. In the causal side information case, we develop a capacity formula based on minimum noise entropy strategies. We then show that strategies associated with entropy-constrained quantizers provide lower and upper bounds on the capacity. At high SNR conditions, i.e., if N is weak relative to the power constraint P_X, these bounds coincide, the optimum strategies take the form of scalar lattice quantizers, and the capacity loss due to not having S at the receiver is shown to be exactly the "shaping gain" $\frac{1}{2}\log(\frac{2\pi e}{12}) \approx 0.254$ bit. We extend the schemes to obtain achievable rates at any SNR and to noncausal side information, by incorporating MMSE scaling, and by using k-dimensional lattices. For Gaussian N, the capacity loss of this scheme is upper bounded by $\frac{1}{2}\log 2\pi e G(\Lambda)$, where $G(\Lambda)$ is the normalized second moment of the lattice. With a proper choice of lattice, the loss goes to zero as the dimension k goes to infinity, in agreement with the results of Costa. These results provide an information theoretic framework for the study of common communication problems such as precoding for intersymbol interference channels and broadcast channels.

Keywords: Dirty-paper channel, interference, causal side information, noncausal side information, precoding, common randomness, dither, randomized code, MMSE estimation.

---

*Parts of this work were presented in ISITA 2000, Honolulu, Hawaii, November 2000 and at the Cornell Summer Workshop on Information Theory, August 2000, Ithaca, NY. This work was supported in part by the Israel Academy of Science, grant 65/01.

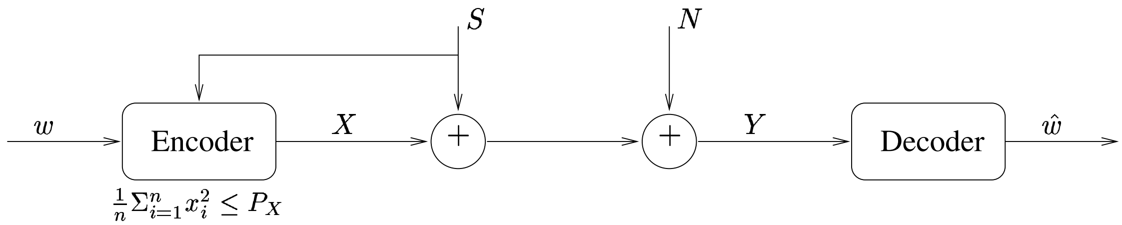

Figure 1: The generalized dirty-paper channel

1 Introduction

We consider power-constrained additive noise channels where part of the noise is known at the transmitter as side information (SI), as shown in Figure 1. That is, the channel is of the form

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

where S is known at the encoder and N is a statistically independent random variable (not necessarily Gaussian) with variance P_N, and where the encoder has power P_X. We refer to S, the known part of the noise, as interference. This choice of terminology will be made clear in the sequel. This channel model has recently received much attention as it has been demonstrated that it models well various important communication problems, among them precoding for intersymbol interference (ISI) channels [14, 18], digital watermarking (e.g., [12, 4]) and various broadcast schemes (e.g., [3],[37]). The channel model was proposed by Cover with Gaussian S and N, where he considered an encoder that has unlimited anticipation, i.e., has knowledge of the entire interference sequence $S_1, S_2 \ldots, S_n$ at the beginning of transmission. In [9], Costa showed that in this case the capacity is equal to $\frac{1}{2}\log(1+P_X/P_N)$. Therefore the interference S does not incur any loss in capacity. We follow [9] and refer to this channel model as the "dirty-paper" channel.

This result has been extended by several authors. In [14, 18] it was shown that it holds for *arbitrarily* varying interference, and also for non-Gaussian noise at high SNR. In [38] and [6], the result was extended to ergodic Gaussian noise [1]. In [7], the case of arbitrarily varying noise was studied.

A different transmission setting is that of a causal SI encoder. A formal definition is given in the next section. In this setting the encoder at each time instance prior to the transmission of x_i has knowledge only of the interference terms up to and including the current instance, i.e., of $S_1, S_2 \ldots, S_i$. We refer to this causal counterpart of the dirty-paper channel as the dirty-tape channel (where "tape" signifies the sequential (causal) availability of the side information). This setting, just as the former, corresponds to many applications.

[1]It was also shown in [6] that the interference need not be Gaussian. However, this result can in fact be deduced from the extension to arbitrary interference provided in [14, 18].

These may be communication problems where the nature of the interference is indeed causal, but may also correspond to dirty-paper coding where we restrict the encoder to be causal in the interest of lower complexity of encoder implementation.

The general formula for the capacity of channels with causal side information at the transmitter was found by Shannon [31], while the capacity with noncausal side information was found by Gelfand and Pinsker [21] (see Section 2). Both formulas are involved in the sense that they are given in terms of maximization over an auxiliary random variable (or function). For the Gaussian dirty-paper channel, however, the solution can be found explicitly [9]. This is of course due to the fact that there is no rate loss in this case with respect to the interference free AWGN channel. Since this does not hold for the dirty-tape channel, finding explicit solutions is a harder problem in this case. Willems was the first to consider the dirty-tape channel in [35]. He suggested a causal encoding scheme in which the encoder uses some of its power to convert the interference S into a discrete random variable whose support is an equally spaced lattice $(\ldots,-3A,-2A,-A,0,A,2A,3A,\ldots)$ which effectively leaves us (when A^2 is large compared to P_N) with a Gaussian noise channel. However, this scheme entails a power loss due to this "noise concentration" process, equal to $A^2/12$ (assuming A is much smaller than the amplitude of the interference signal). In [36], Willems refers to schemes which circumvent the power loss of "noise concentration".

In this paper we are concerned with both the causal and noncausal settings, as well as with the case of side-information with *finite anticipation*. We focus our attention on the *worst interference* case, which we show to be equivalent to "strong and smooth" interference, and to arbitrarily varying interference. We derive capacity formulas and bounds as well as coding strategies for these settings in a unified approach. This allows to bridge the causal and the noncausal settings. We also investigate how much is lost in capacity by imposing the causality constraint. Our coding scheme is based on *minimum noise entropy strategy*, a concept proposed earlier for unconstrained modulo-additive noise channels in [15]. We addressed these issues in a preliminary version of this paper [14, 18]. Schemes similar to those presented in [14, 18] were independently proposed by Chen and Wornell [4] as well as by Su, Eggers and Girod [12] in the context of information embedding. The present paper gives a detailed account of the results reported in [14] for the dirty-tape as well as dirty-paper channel, where N may or may not be Gaussian.

One of the insights developed in [14] is that the dirty-paper channel model offers a theoretical framework for precoding techniques, and in particular the link to Tomlinson-Harashima precoding [32, 23] was established. Since then, considerable work has been done (and published) by the authors as well as by others, building on this insight. We will thus not delve into applications in this paper. Instead we refer the reader to [41] and the references therein for a survey of some of the recent works. A noteworthy implication of this work is that the capacity of the Gaussian intersymbol interference (ISI) channel may be achieved using precoding at the transmitter and that there is no inherent precoding loss. Another important application [3, 34] is to precoding for broadcast over multiple-input multipleoutput (MIMO) antennas, allowing to achieve the capacity region [34]. Finally, the present work lead in turn to a transmision scheme [17, 19] that allows the capacity of the AWGN channel to be achieved using lattice encoding and decoding, a problem that was open for many years.

A distinctive feature of our approach, as proposed in [14], is the introduction of *common randomness* at the transmitter and receiver ends which enters in the form of a "dither" that is added to the interference. This serves a dual purpose. With the exception of Section 4.3, its primary role is as an analytic tool in the direct part of the capacity formulas: the dither greatly simplifies the treatment and allows for a rigorous treatment as well as enables us to relate coding for the dirty-tape channel to well established results in quantization theory. In this respect, the dither may be regarded as merely a method of proof while the capacity results *do not* depend on the availability of common randomness in practice. This is due to the fact that common randomness does not result in a greater capacity for fixed probabilistic channels with SI at the transmitter (unlike Arbitrarily Varying Channels (AVC) as will be noted); see [28, 1]. In Section 4.3, on the other hand, the dither will prove essential where we discuss the issue of cancellation of *arbitrary* interference. In this case, common randomness, (e.g. a randomized codebook known to the receiver) may be in fact advantageous [28]. That is, the capacity formula we give for this case will assume that common randomness is indeed available.

The paper is organized as follows. Section 2 summarizes known results for channels with causal/noncausal side information at the transmitter and the associated (non-explicit) capacity formulas. Section 3 treats the worst-interference, general noise, dirty-tape channel, for which a semi-explicit capacity formula is derived in terms of a minimum noise entropy strategy. Lattice encoding schemes are proposed and are shown to be optimal in the limit of high SNR. Furthermore, for general SNR, upper bounds for the rate loss of *inflated* lattice strategies are given. Section 4 proposes efficient schemes for side information known with finite anticipation, linking the dirty-tape and dirty-paper settings, and develops techniques for cancellation of arbitrary interference. Section 5 offers a summary of the results and discusses some extensions of the results.

2 Channels with side information at the transmitter

The channel model (1) is a special case of a channel with side information at the transmitter. Such channels were introduced by Shannon in [31]. He considered a discrete memoryless channel whose transition matrix is dependent on the channel "state", as shown in Figure 2. The transmitter has knowledge of this state prior to transmission [2]. More precisely, let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ denote the input, output and state alphabets of the channel, respectively, with transition probability $p(y|x,s)$ and with state probabilities given by $p(s)$. The transmitter

[2]The interference term S of the dirty-paper channel corresponds to the channel "state".

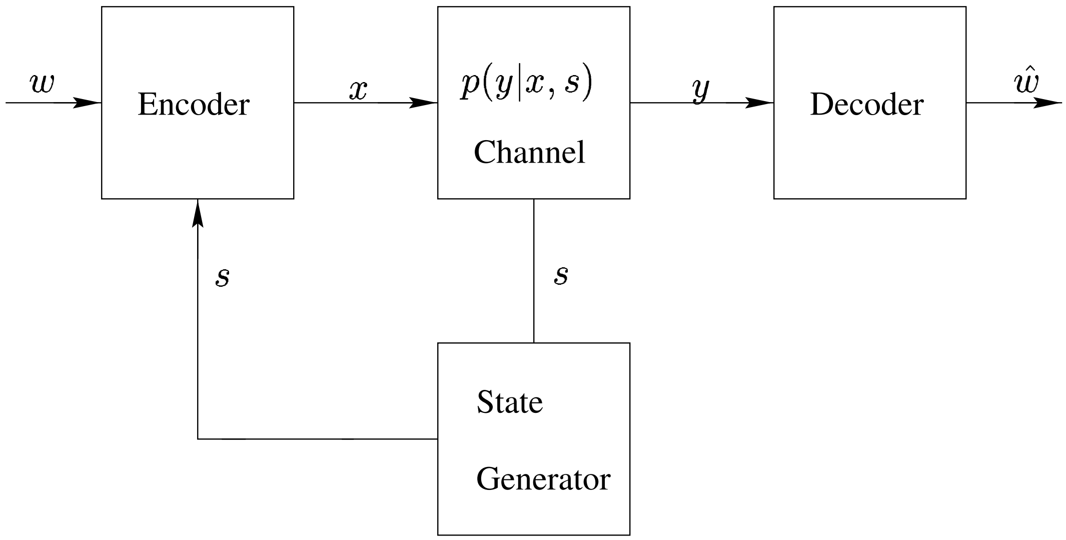

Figure 2: Channel with side information at transmitter

(but not the receiver) has access to the side information. This problem divides into two categories, according to whether the encoder observes the state process *causally*, or *anticipates future states* (corresponding to the dirty-tape/dirty-paper scenarios when the channel is given by (1)).

In the causal case, considered by Shannon [31], the encoder maps the message $w \in \{1, 2, ..., M = 2^{nR}\}$ into $\mathcal{X}^n$ using functions

$$x_i = f_i(w, s_1^i) \quad 1 \leq i \leq n \tag{2}$$

where $s_1^i = s_1, \ldots, s_i$ are the states up to time i. Shannon found the capacity of such channels as described below.

Shannon's work remained largely an isolated result for many years (with the notable exception of [25]). Renewed interest was sparked in the Russian literature by Kuznetsov and Tsybakov during the 1970's in the context of coding for memories with defective cells [27]. The study of this problem eventually led to the general formulation of Gelfand and Pinsker [21] for coding with *noncausal* side information at the transmitter. In this case, the encoder observes the entire state sequence before transmitting the code sequence, thus

$$x_i = f_i(w, s_1^n) \quad 1 \leq i \leq n. \tag{3}$$

In either case (causal or non-causal), the receiver decodes the message w from the whole received vector as $\hat{w} = g(y_1^n)$. For the causal scenario, the (average) probability of error is

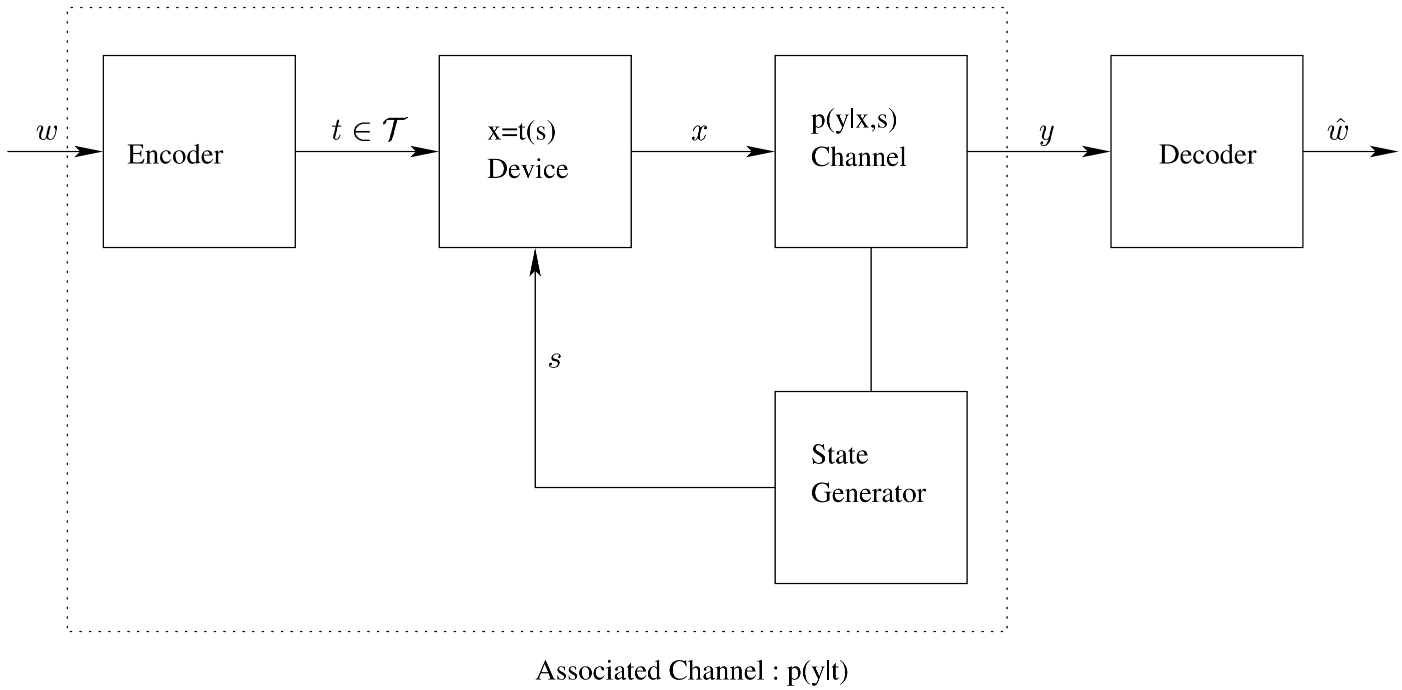

Figure 3: Shannon's associated channel.

given by

$$P_e = \frac{1}{M}\sum_{w=1}^{M}\sum_{\{y_1^n : g(y_1^n)\neq w\}}\sum_{s_1^n} p(s_1^n)\prod_{i=1}^{n} p\left(y_i | f_i(w, s_1^i), s_i\right). \tag{4}$$

The probability of error for the noncausal case is similarly defined.

We consider also randomized codes. That is, transmission schemes involving common randomness. In such cases, the transmitter and receiver operation may depend on the value of a random variable which is known at both transmission ends. Denote this random variable by U. For the causal (Shannon) scenario the encoder mapping is then given by functions of the form

$$x_i = f_i(w, s_1^i, u) \quad 1 \leq i \leq n. \tag{5}$$

Likewise the decoding function is given by $\hat{w} = g(y_1^n, u)$. The (average) probability of error is then defined by

$$P_e^{\mathrm{RC}} = \frac{1}{M}\sum_{w=1}^{M}\left\{E_U \sum_{\{y_1^n : g(y_1^n, u)\neq w\}}\sum_{s_1^n} p(s_1^n)\prod_{i=1}^{n} p\left(y_i | f_i(w, s_1^i, u), s_i\right)\right\}. \tag{6}$$

The probability of error for the noncausal case is similarly defined. Note that by interchanging the expectation with the outer summation in (6) it follows that there must be some specific value u, i.e., some deterministic code, with a probability of error no greater than that of (6). In this sense, randomized codes do not yield better performance than deterministic ones. However, this optimal u depends in general on the state distribution; thus randomization may be advantageous for arbitrary varying or unknown state sequences as discussed in the Section 4.3. We note that in the sequel, we will consider transmission (and hence codebooks) that are subject to a power contraint. In this case, *both* the probability of error as well as the codeword power depend on the value of u. Nonetheless, using a Langrangian formulation, it can be shown that a randomized code does not improve on a deterministic code in this setting as well.

2.1 Non-explicit expressions for capacity

For a general memoryless channel $p(y|x, s)$, with memoryless states, Shannon [31] showed that the capacity with causal SI at the transmitter is equal to the regular capacity of an *associated* DMC as shown in Figure 3. The input alphabet of the associated channel, denoted $\mathcal{T}$, is the set of all possible mappings

$$t : \mathcal{S} \longrightarrow \mathcal{X}$$

which we refer to as *strategies* or *strategy functions.* The output y of the associated channel is related to the input t according to the transition probability

$$p(y|t) = \sum_s p(s) p\Big(y \Big| x = t(s), s\Big) \tag{7}$$

and also

$$p(y_1^n|t_1^n) = \prod_{i=1}^{n} p(y_i|t_i). \tag{8}$$

The capacity with side information at the transmitter is given by [31],

$$C^{\text{causal}} = \max_{p(t)} I(T;Y), \tag{9}$$

where the maximization is taken over the distribution $p(t)$ of the random variable $T \in \mathcal{T}$. The main feature of Shannon's capacity formula is that it involves strategy functions that are functions only of the *current* state. This in turn means that to achieve capacity it is sufficient to have an encoder that takes into account only the current state of the channel. See also [15].

This result can readily be extended to the case where the alphabets $\mathcal{X}$,$\mathcal{Y}$,$\mathcal{S}$ are the real line and where the transmitter is subject to an average power constraint P_X to yield

$$C^{\text{causal}}(P_X) = \sup_{p(t): E\{T(S)^2\} \leq P_X} I(T;Y) \tag{10}$$

where the expectation is relative to the *product* measure $p(s,t) = p(s)p(t)$. The capacity with noncausal SI at the transmitter is given by [3] [21]

$$C^{ncausal} = \max_{p(t|s)} \{I(T;Y) - I(T;S)\} \tag{11}$$

where T is a random strategy, i.e., a random element of the set of functions $\{t : \mathcal{S} \to \mathcal{X}\}$, and the maximization is taken over all joint distributions satisfying

$$p(t,s,x,y) = p(s)p(t|s)\delta(x - t(s))p(y|x,s)$$

where $\delta(\cdot)$ denotes the Kronecker delta function. Note that unlike in (10) here $p(s,t)$ is a *general* joint distribution. This expression coincides with the causal capacity (9) if the maximization is restricted to distributions satisfying $p(t,s,x,y) = p(s)p(t)\delta(x-t(s))p(y|x,s)$, i.e. when T and S are independent. As in the causal case, the capacity formula may be extended to the power-constrained/continuous alphabet case (see [2]). The capacity formula is then given by

$$C^{ncausal}(P_X) = \sup_{p(t|s):E\{T(S)^2\} \leq P_X} \{I(T;Y) - I(T;S)\}. \tag{12}$$

3 Results for Causal Side Information

3.1 Capacity formula via minimum noise entropy

Let us turn our attention back to the generalized dirty-paper channel model (1). In this section, we treat the causal SI scenario (or dirty-tape channel). We use the general capacity formula of Shannon (10) to find the capacity of this channel for the worst-case interference, which will turn out to be the asymptotic case of strong and smooth interference. This greatly simplifies the treatment, while still incurring only a *finite penalty* relative to the case of $S \equiv 0$ which we shall quantify. We assume that the noise N has a finite differential entropy and finite first and second moments. We define the *worst interference capacity* of the dirty-tape channel as

$$C^{causal,worst}(P_X) = \inf_S C^{causal}(P_X, S) = \inf_S \sup_{T:E\{T(S)\}^2 \leq P_X} I(T;Y), \tag{13}$$

where $C^{causal}(P_X, S)$ is the capacity expression in (10) with the dependence on S made explicit. We now present an expression for the worst-case capacity of the dirty-tape channel which translates the maximization in (10) into noise entropy minimization. In this sense,

[3]This is a modified form [13, 5] of the Gelfand-Pinsker formula, which better shows the relation to Shannon's formula (see 9) for the causal case. We identify the random variable U in the Gelfand-Pinsker capacity expression with the random function T.

the resulting capacity formula is "semi-explicit". The result is derived by transforming the original channel into an effective modulo-additive noise channel, whose noise distribution depends on a chosen strategy. In the sequel, we propose explicit lattice-strategy encoding schemes and prove their optimality in the limit of high SNR.

For $L > 0$, let $U \sim \mathrm{Unif}([-L/2, L/2))$. Let $t(\cdot)$ be a strategy function from $[\frac{-L}{2}, \frac{L}{2})$ to $\mathbb{R}$. Define the *minimum effective noise entropy*

$$h_{\min}(L,P) = \inf_{t \in \mathcal{T}(L,P)} h(t(U) + U + N), \tag{14}$$

and the effective noise channel capacity

$$\tilde{C}_L(P) \;=\; \log L - h_{\min}(L,P), \tag{15}$$

where $h(\cdot)$ denotes differential entropy, and the class of admissible strategies is defined as

$$\mathcal{T}(L,P) = \{t : E[t(U)]^2 \leq P\}. \tag{16}$$

Define

$$C^*_L(P) = \text{upper convex envelope of } \tilde{C}_L(P), \tag{17}$$

and let

$$C^*(P) = \limsup_{L \to \infty} C^*_L(P). \tag{18}$$

Note that any point in the convex envelope may be obtained by time-sharing of at most two points [10].

Theorem 1 (causal worst-case capacity) *The worst-case (causal) SI capacity of the channel (1), defined in (13), is given by*

$$C^{causal,worst}(P_X) = C^*(P_X). \tag{19}$$

The theorem is proved in Section 3.3. We next describe a universal interference canceling scheme that will play a central role in the proof.

3.2 Universal interference cancelling scheme

We present a randomized transmission scheme, which is independent of the statistics of the interference S and achieves the worst interference capacity $C^*(P)$ for any S. We transform in effect the channel into a modulo additive noise channel over the alphabet $\mathcal{A}_L = [-L/2, L/2)$. The transmission scheme is outlined in Figure 4. The transmitter uses an input alphabet that consists of strategies belonging to

$$\mathcal{T}_{0,L} = \{t_v : t_v(s) = t_0(s - v \bmod \mathcal{A}_L),\ v \in \mathcal{A}_L\}, \tag{20}$$

where $t_0(S)$ is some strategy function. That is, all strategies are a shift modulo $\mathcal{A}_L$ of a single strategy. Let $U \sim \mathrm{Unif}(\mathcal{A}_L)$ be a dither available at both transmission ends.

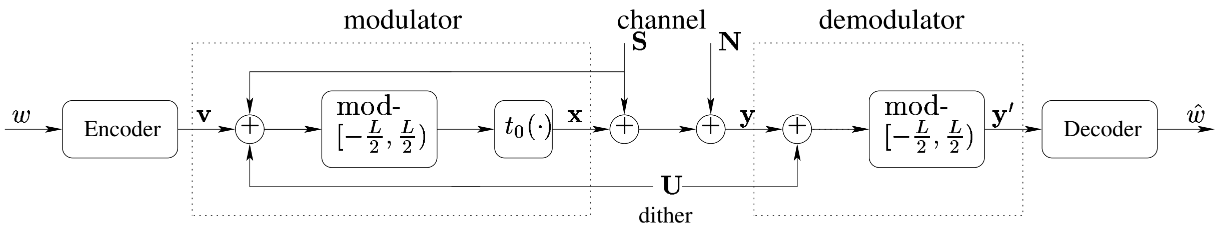

Figure 4: Universal interference cancelling scheme for the dirty-tape channel.

- *Modulator:* For any $v \in \mathcal{A}_L$, the encoder sends:

$$x = t_v(s+u) = t_0(s+u-v \bmod \mathcal{A}_L). \tag{21}$$

- *Demodulator:* Computes

$$y' = [y+u] \bmod \mathcal{A}_L. \tag{22}$$

We thus arrive at the following channel from v to Y'.

Lemma 1 (effective additive-noise channel) *The channel from v to Y' defined by (1), (21) and (22) is equivalent in distribution to the additive noise channel*

$$Y' = [v + N_{\text{eff}}] \bmod \mathcal{A}_L$$

where the effective noise N_{eff} is independent of v and is given by

$$N_{\text{eff}} = [t_0(U) + U + N] \bmod \mathcal{A}_L.$$

Note that the effective channel from v to Y' is independent of the interference S. This is not only useful for proving Theorem 1, but it also has an important consequence for arbitrarily varying interference, as discussed in Section 4.3.

Proof:

$$\begin{aligned} Y' &= [t_0(U+S-v \bmod \mathcal{A}_L) + U + S + N] \bmod \mathcal{A}_L & (23)\\ &= [t_0(U+S-v \bmod \mathcal{A}_L) + (U+S-v) + v + N] \bmod \mathcal{A}_L & (24)\\ &= [v + t_0(U'') + U'' + N] \bmod \mathcal{A}_L & (25)\\ &= [v + N''] \bmod \mathcal{A}_L & (26) \end{aligned}$$

where $U'' = [U+S-v] \bmod \mathcal{A}_L$ and

$$N'' = [t_0(U'') + U'' + N] \bmod \mathcal{A}_L. \tag{27}$$

Due to the dither U, for any $V = v$ and $S = s$ the random variable U'' is uniformly distributed over $\mathcal{A}_L$, i.e., U'' has the same distribution as U. Consequently, S does not have any effect on the associated channel and N'' is statistically independent of V and S. Thus, the resulting channel (26) is a *modulo additive noise channel* and N'' has the same distribution as N_{eff}. □

Applying a uniform distribution upon the class of strategies $\mathcal{T}_{0,L}$, i.e. $V \sim \text{Unif}(\mathcal{A}_L)$ yields for any S

$$I(V;Y') = h(Y') - h(Y'|V) = \log L - h(N_{\text{eff}}). \tag{28}$$

3.3 Proof of Theorem 1

As noted in Section 2, common randomness does not increase capacity for the channel models we study, i.e., channels with side information at the transmitter, see e.g., [28]. Nonetheless, it will prove useful in the proof to examine the case where common randomness is available. Let the random variable U be available at both transmission ends. As noted, allowing the strategy functions to depend on the dither U does not increase capacity. We may therefore rewrite the worst-case capacity of (13) as

$$C^{causal,worst}(P_X) = \inf_S C^{random}(P_X, S) \tag{29}$$

where

$$C^{random}(P_X, S) = \sup_{U,T:E\{T(S,U)^2 \leq P_X\}} I(T;Y). \tag{30}$$

Theorem 1 is proved using the following two lemmas.

Lemma 2 *(Direct) For any interference S*

$$C^{causal}(P_X, S) \geq C_L^*(P_X) \tag{31}$$

for every L.

Lemma 3 *(Converse) For $S \sim \text{Unif}(\mathcal{A}_L)$*

$$C^{causal}(P_X, S) \leq C_L^*(P_X) + o_L(1) \tag{32}$$

where $o_L(1) \to 0$ as $L \to \infty$.

Since the worst interference capacity is defined as an infimum over all interferences S (see (13)), every S gives an upper bound on $C^{causal,worst}(P_X)$, in particular $S \sim \text{Unif}(\mathcal{A}_L)$. Thus, the two lemmas above imply that

$$C_L^*(P_X) \leq C^{causal,worst}(P_X) \leq C_L^*(P_X) + o_L(1)$$

for every L, and the desired result follows by taking the limsup in L. [4] We are left to prove the two lemmas.

Proof of Lemma 2

We employ the universal interference canceling scheme described in Section 3.2. From (28), for any choice of basic strategy $t_0(\cdot)$ we can achieve the mutual information

$$I(V;Y') = \log L - h\left(t_0(U) + U + N \bmod \mathcal{A}_L\right). \tag{33}$$

But

$$h\left(t_0(U) + U + N \bmod \mathcal{A}_L\right) \leq h\left(t_0(U) + U + N\right) \tag{34}$$

since the modulo operation can only reduce the entropy. For any P, we may take a strategy $t_0(\cdot)$ that achieves a value arbitrarily close to the minimum effective noise entropy in (14). Combining with the definition of $\tilde{C}_L(P)$ in (15), we conclude that we can achieve mutual information

$$I(V;Y') \geq \tilde{C}_L(P) - \epsilon \tag{35}$$

for any $\epsilon > 0$. By the definition of $C_L^*(P)$ in (17), we may achieve mutual information of $C_L^*(P_X) - \epsilon$ by time-sharing (at most two basic strategies), and the lemma follows.

Proof of Lemma 3

The proof is similar to the proof of Theorem 1 of [15]. Let T be any strategy random variable. We have

$$\begin{aligned} I(T;Y) &= h(S + T(S) + N) - h(S + T(S) + N|T) & (36)\\ &= h(S + X + N) - E_T\left\{h(S + T(S) + N)|T = t(s)\right\} & (37)\end{aligned}$$

where $X = T(S)$ and $E_T\{\cdot\}$ denotes expectation over T. In Appendix 6, we prove the following lemma.

Lemma 4 *If $S \sim \mathrm{Unif}(\mathcal{A}_L)$ and $E\{X^2\} \leq P_X$ then*

$$h(S + X + N) \leq \log L + o_L(1) \tag{38}$$

where (X, S) are independent of N (but X may depend on S) and $o_L(1) \to 0$ as $L \to \infty$.

Therefore by (37) for $S \sim \mathrm{Unif}(-L/2, L/2)$ we have

$$I(T;Y) \leq \log L - E_T\left\{h(S + T(S) + N|T = t(s))\right\} + o_L(1). \tag{39}$$

Let $t(\cdot)$ be any function participating in the expectation of (39). Denote $P_t = E\{t(S)^2\} = E\left\{T(S)^2|T = t(s)\right\}$. Since by the definition of $\tilde{C}_L(P)$ in (15) we have

$$h(t(S) + S + N) \geq h_{min}(L, P_t) = \log L - \tilde{C}_L(P_t), \tag{40}$$

[4] Note that this implies also that $C^*(P) = \lim_L C_L^*(P)$.

it follows that equation (39) reduces to

$$I(T;Y) - o_L(1) \le E_T\left\{\tilde{C}_L(P_T)\right\} \le E_T\left\{C_L^*(P_T)\right\} \le C_L^*\left(E_T\{P_T\}\right) \le C_L^*(P_X) \tag{41}$$

where the inequalities follow from the definition of $\tilde{C}_L(\cdot)$ and $C_L^*(\cdot)$ and from the convexity and monotonicity of $C_L^*(\cdot)$, and since the power constraint implies $E_T\{P_T\} \le P_X$. Since this inequality holds for any T satisfying the power constraint, the lemma follows.

Remark: In fact, Lemma 4 holds with respect to any S of the form LS_0 where S_0 has a density. Thus the worst-case capacity occurs whenever the interference is "strong and smooth".

3.4 Bounds via entropy-constrained quantization

From Theorem 1 we see that the capacity formula involves finding an optimal $t(\cdot)$ that minimizes $h(t(U)+U+N)$ subject to the power constraint $t \in \mathcal{T}(L,P)$. The following theorem links this problem to that of finding the optimal entropy-constrained quantizer of $U \sim \mathrm{Unif}(-\frac{L}{2}, \frac{L}{2})$. Let

$$H_{\min}(U,D) = \inf H(Q(U)) \tag{42}$$

denote the minimum entropy in quantizing U with mean squared distortion D, where $H(\cdot)$ denotes regular entropy and the infimum is over all quantizers Q satisfying $E[Q(U)-U]^2 \le D$.

Lemma 5 *Suppose N has a finite differential entropy. Then*

$$\begin{aligned}\tilde{C}_L(P_X) &\ge h(U) - H_{\min}(U,P_X) - h(N) &(43)\\ &\ge \frac{1}{2}\log 12P_X - h(N) - \log\left(1+\frac{\sqrt{12P_X}}{L}\right). &(44)\end{aligned}$$

On the other hand, for any $a>0$

$$\begin{aligned}\tilde{C}_L(P_X) &\le h(U) - H_{\min}(U,[\sqrt{P_X}+a/2]^2) - h(N) + I(N;N+Z_a) &(45)\\ &\le \frac{1}{2}\log 12\left(\sqrt{P_X}+\frac{a}{2}\right)^2 - h(N) + I(N;N+Z_a) &(46)\end{aligned}$$

where Z_a is independent of N and is uniformly distributed over $(-a/2,+a/2)$.

The proof of the lemma is given in Appendix 6.

We now restrict our attention to the case of high SNR, i.e, to the limit $P_X \to \infty$. Define $\tilde{C}^*(P) = \lim_{L\to\infty}\tilde{C}(P)$. From (44), taking $L \to \infty$, we have

$$C^*(P_X) \ge \tilde{C}^*(P_X) \ \ \ge \ \ \frac{1}{2}\log 12P_X - h(N). \tag{47}$$

From (46), we have

$$\tilde{C}^*(P_X) \;\leq\; \min_a \left[\frac{1}{2}\log 12\left(\sqrt{P_X}+\frac{a}{2}\right)^2 + I(N;N+Z_a)\right] - h(N). \tag{48}$$

From (47) and (48), we have

$$0 \leq C^*(P_X) - \left[\frac{1}{2}\log(12P_X) - h(N)\right] \leq \min_a \left[\log\left(1+\frac{a/2}{\sqrt{P_X}}\right) + I(N;N+Z_a)\right]. \tag{49}$$

Clearly $I(N;N+Z_a) \to 0$ as $a \to \infty$ whenever $P_N < \infty$; see [29]. For any $\epsilon > 0$, let a_ϵ be large enough so that $I(N;N+Z_a) < \epsilon$. But $\log\left(1+\frac{a/2}{\sqrt{P_X}}\right) \to 0$ as $P_X \to \infty$. We thus have,

Corollary 1 *If N has a finite second moment and finite differential entropy, then*

$$C^*(P_X) = \frac{1}{2}\log(12P_X) - h(N) + o(1) \tag{50}$$

where $o(1) \to 0$ as $P_X \to \infty$.

3.5 Optimality of lattice strategies at high SNR

From (50), we see that the asymptotic (high SNR) rate loss with respect to the no-interference case $S=0$ (or equivalently, to having S also at the receiver), is equal to the "shaping gain", $\frac{1}{2}\log\frac{2\pi e}{12} \approx 0.254$ bit. The role of the shaping gain here will be made clear in Section 4, where we discuss the use of multidimensional lattice strategies for coding with finite anticipation side-information. Note that this result holds for general N, not necessarily Gaussian.

It also follows from (50), that entropy-constrained quantizers generate efficient strategies for the universal interference cancelling scheme at high SNR. From the well known result by Gish and Pierce, we know that at "high resolution" conditions the quantizer achieving the minimum entropy $H_{\min}(U,P_X)$ is uniform, see [22]. Thus, at high SNR the dirty-tape channel capacity may be approached using the error of a uniform quantizer as $t(\cdot)$ in (14). That is, we choose $t_0(s) = Q_\Delta(s) - s$ where $Q_\Delta(\cdot)$ is a "mid-thread" uniform scalar quantizer with step size

$$\Delta = \sqrt{12P_X}. \tag{51}$$

The function $t_0(s)$ is depicted in Figure 5. We now apply a uniform distribution upon the class of strategies which are shifts of t_0

$$\begin{aligned} t_v(s) &= Q_\Delta(s-v) + v - s \\ &= [v-s] \bmod \Delta \end{aligned} \tag{52}$$

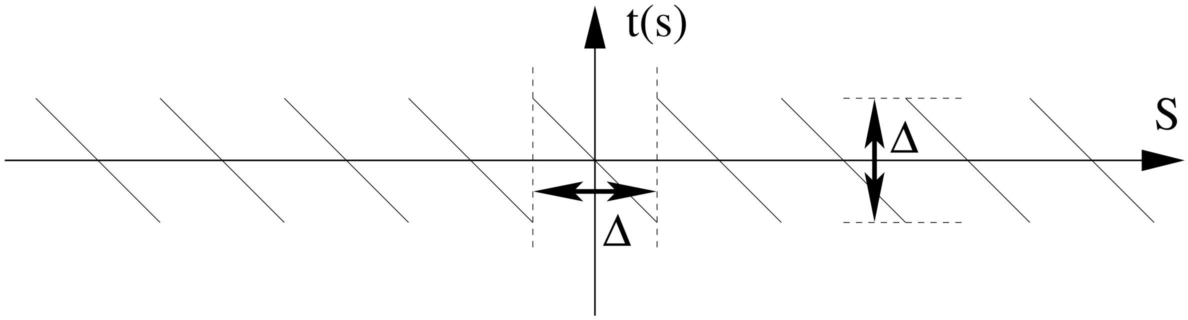

Figure 5: Uniform lattice strategy: $t_0(s) = -s \bmod \Delta$ with $\Delta = \sqrt{12P_X}$.

where the modulo operation is to the interval

$$\mathcal{A}_\Delta = (-\frac{\Delta}{2}, +\frac{\Delta}{2}]. \tag{53}$$

Due to the periodic nature of t_0, the shift v may be limited to the interval $\mathcal{A}_\Delta$, and it is sufficient to take the dither to be $U \sim \text{Unif}(\mathcal{A}_\Delta)$. Also, due to the dither U being added at the receiver side, reducing the output (after the dither is added) modulo Δ produces a sufficient statistic at the receiver. We therefore use the following transmission scheme:

- *Transmitter:* For any $v \in \mathcal{A}_\Delta$, the encoder sends:

  $$x = t_0(s+u-v) = Q_\Delta(s+u-v) + v - s - u = [v - s - u] \bmod \Delta. \tag{54}$$

  Note that since U is uniform over $\mathcal{A}_\Delta$, so is the transmitted signal X. It follows from (51) that the transmitted power is $EX^2 = \frac{\Delta^2}{12} = P_X$.

- *Receiver:* The receiver computes

  $$\begin{aligned} y' &= [y+u] \bmod \Delta && (55)\\ &= [v+N] \bmod \Delta. && (56) \end{aligned}$$

  where (56) follows by specializing Lemma 1 to this case, noting that $t_0(u)+u \bmod \Delta = Q_\Delta(u) \bmod \Delta = 0$ for all u.

Taking $V \sim \text{Unif}(\mathcal{A}_\Delta)$ gives rise to the rate

$$\begin{aligned} I(V;Y') &= h(V) - h(N \bmod \Delta) && (57)\\ &\approx h(V) - h(N) && (58)\\ &= \log \Delta - h(N) && (59)\\ &= \frac{1}{2}\log 12P_X - h(N) && (60) \end{aligned}$$

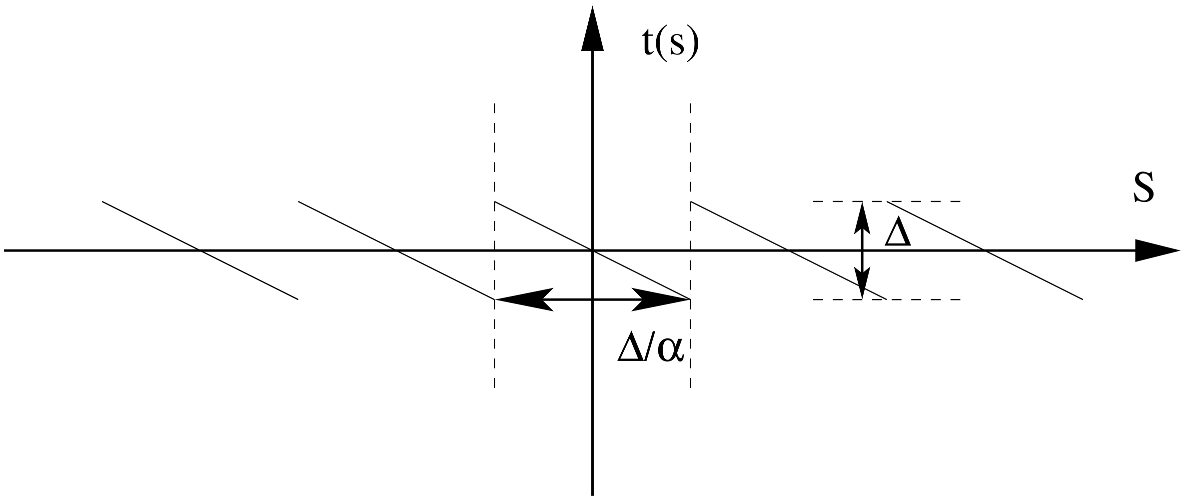

Figure 6: Inflated lattice strategy: $t_0(s) = -\alpha s \bmod \Delta$ with $\Delta = \sqrt{12 P_X}$.

where the approximation in (58) becomes tight as $P_X \to \infty$ and therefore $\Delta \to \infty$. Hence, in light of (50) this scheme is asymptotically optimal. The scheme (52) is similar to the technique for information embedding of [4] and is closely linked to Tomlinson-Harashima prcoding [32, 23].

3.6 Inflated lattice strategies for general SNR

In principle, the optimal noise entropy minimizing strategy function $t(\cdot)$ as defined in (15), gives us a capacity achieving encoding scheme as depicted in Figure 4. Unfortunately, we have only been able to determine this optimal function in the case of asymptotically high SNR. For general SNR, we resort to a judicious choice of a suboptimal strategy function. The scheme we propose, based on an "inflated" lattice strategy, is motivated by the encoding scheme of Costa [9].

The development up to this point of the paper did not necessitate that the constraint be a power constraint, and could be extended to more general constraints. The MMSE scaling that we next introduce *does* fit specifically the case of a power constraint.

The scheme uses a scaling coefficient, $0 < \alpha \leq 1$, effectively producing at the receiver end a lattice with cells of length $\sqrt{12 P_X/\alpha^2}$, at the expense of adding an additional noise component with variance $\left(\frac{1-\alpha}{\alpha}\right)^2 P_X$. The basic strategy takes the form

$$t_0(s) = -\alpha s \bmod \Delta \tag{61}$$

where as before $\Delta = \sqrt{12 P_X}$, and the modulo operation is to the interval $\mathcal{A}_\Delta$ defined in (53). Since $t_0(s)$ is periodic, it is sufficient now to restrict the shift v to the expanded interval $\mathcal{A}_{\Delta/\alpha} = [-\frac{\Delta}{2\alpha}, \frac{\Delta}{2\alpha})$, and the dither to be $U \sim \text{Unif}(\mathcal{A}_{\Delta/\alpha})$. For any $v \in A_{\Delta/\alpha}$, the encoder sends:

$$x = [\alpha(v - s - u)] \bmod \Delta \tag{62}$$

and the receiver computes:

$$y' = [y + u] \bmod \Delta/\alpha \tag{63}$$

where, as before, reducing the output modulo the period Δ/α produces sufficient statistics.

Note that the input and output alphabet (after applying the modulo operation) is scaled or "inflated" by a factor of $1/\alpha$ relative to the basic lattice transmission scheme of Section 3.5. Hence, we refer to these strategy functions as "inflated lattice strategies". See Figure 6. Alternatively, we may restrict the input alphabet to $\mathcal{A}_\Delta$ as in Section 3.5 (defined in (53) and (51)) and take $U \sim \text{Unif}(\mathcal{A}_\Delta)$, if we scale instead the interference S prior to subtracting it off at the transmitter, and scale the receiver input prior to adding the dither. The transmission scheme then takes the form,

- *Transmitter:* For any $v \in \mathcal{A}_\Delta$, the encoder sends:

$$x = [v - \alpha s - u] \bmod \Delta. \tag{64}$$

- *Receiver:* The receiver computes

$$y' = [\alpha y + u] \bmod \Delta. \tag{65}$$

This gives rise to an equivalent modulo lattice channel described by the following lemma.

Lemma 6 (Inflated lattice lemma: scalar case) *The channel defined by (1),(64) and (65) is equivalent in distribution to the channel*

$$Y' = v + N' \bmod \Delta \tag{66}$$

where N' is independent of v and is given by

$$N' = \Big[(1-\alpha)U + \alpha N\Big] \bmod \Delta \tag{67}$$

and where $U \sim \text{Unif}(-\Delta/2, \Delta/2)$ and is statistically independent of N.

Proof: For any $v \in \mathcal{A}_\Delta$ we get

$$\begin{aligned}
Y' &= [\alpha Y + U] \bmod \Delta & (68)\\
&= [v - v + \alpha X + \alpha S + \alpha N + U] \bmod \Delta & (69)\\
&= [v + \alpha X + (\alpha S + U - v) + \alpha N] \bmod \Delta & (70)\\
&= [v + \alpha X + (\alpha S + U - v) \bmod \Delta + \alpha N] \bmod \Delta & (71)\\
&= [v + \alpha X + (-X) + \alpha N] \bmod \Delta & (72)\\
&= [v - (1-\alpha)X + \alpha N] \bmod \Delta. & (73)
\end{aligned}$$

Notice that due to the dither U, the channel input X is uniform over $\mathcal{A}_\Delta$ independently of v. Since U and $-U$ also have the same distribution the lemma follows. $\square$

The scaling factor α should be chosen so as to maximize the corresponding mutual information (minimize the entropy of N'). Alternatively, we may use an MMSE-scaling factor (as done by Costa), i.e., take

$$\alpha = \frac{P_X}{P_X + P_N} = \frac{\mathrm{SNR}}{1+\mathrm{SNR}}. \tag{74}$$

This minimizes the variance of the effective noise prior to the modulo operation, i.e., the variance of $(1-\alpha)U + \alpha N$ [5]. Thus

$$\begin{aligned}
\mathrm{Var}(N') &\leq \mathrm{Var}((1-\alpha)U + \alpha N) && (75)\\
&= (1-\alpha)^2 \mathrm{Var}(U) + \alpha^2 \mathrm{Var}(N) && (76)\\
&= \frac{P_N P_X}{P_N + P_X} && (77)
\end{aligned}$$

where (75) follows since the modulo operation may only reduce the variance of a random variable. The corresponding rate satisfies

$$\begin{aligned}
I(T;Y) &= \log\Delta - h(N') && (78)\\
&\geq \frac{1}{2}\log 12 P_X - \frac{1}{2}\log 2\pi e \frac{P_X P_N}{P_X + P_N} && (79)
\end{aligned}$$

where the inequality follows since a Gaussian random variable has the greatest entropy for a given variance [10]. We thus have,

Theorem 2 *For any noise N and arbitrary interference S, the capacity of the channel (1) with S known causally to the tranmitter satisfies*

$$C^{causal}(P_X) \geq \frac{1}{2}\log\Big(1 + \frac{P_X}{P_N}\Big) - \frac{1}{2}\log\frac{2\pi e}{12}. \tag{80}$$

Notice that this bound may be tighter than the lower bound in (47). To recognize this consider the case of Gaussian N where (47) would give us the weaker bound

$$C(P_X) \geq \tilde{C}(P_X) \geq \frac{1}{2}\log\left(\frac{P_X}{P_N}\right) - \frac{1}{2}\log\frac{2\pi e}{12}. \tag{81}$$

It is interesting to find a lower bound for the achievable rate at the limit of very low SNR, i.e. as $\mathrm{SNR} = \frac{P_X}{P_N} \to 0$. We do this for the case of Gaussian noise N by numerically computing

$$\beta(\mathrm{SNR}) = I(T;Y)/\mathrm{SNR}, \tag{82}$$

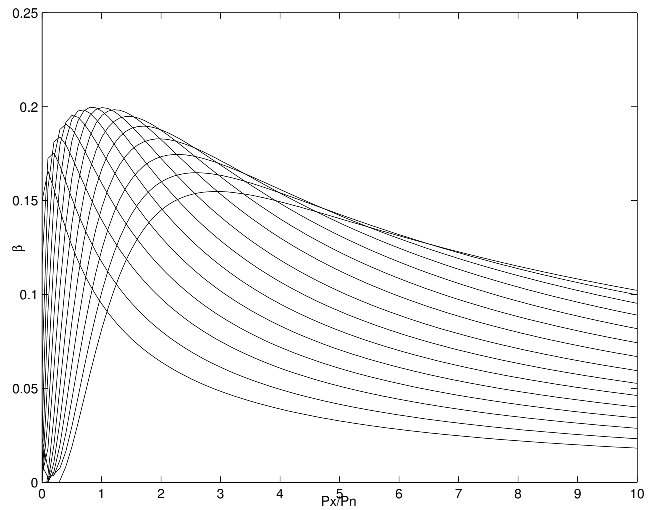

Figure 7: $\beta(\text{SNR}) = I(T;Y)/SNR$ as a function of $SNR = P_X/P_N$ for inflated lattice strategies. The different lines correspond to different values of α.

where $I(T;Y)$ is given in (78), as a function of the SNR. This is shown in Figure 7. Due to the convex hull in the expression for capacity (17), it follows that

$$\beta^* = \max_{\text{SNR}} \left\{ \frac{I(T;Y)}{\text{SNR}} \right\} \tag{83}$$

is the slope of the capacity as a function of the SNR at SNR $= 0$. In effect, this value is the maximum information per unit power that can be conveyed using an inflated lattice scheme.

From Figure 7 we see that $\beta^* \approx 0.2$. This yields a rate of approximately $0.2P_X/P_N$ nats at low SNR, whereas the capacity with noncausal SI is $\frac{1}{2}P_X/P_N$ nats. This indicates that the rate loss due to causality is bounded by 4dB. This performance can be obtained by time-sharing the zero power strategy and the optimal operating point, i.e., the SNR that maximizes $\beta(\text{SNR})$, which is approximately at 0dB. We note that the above derivation is equivalent to applying the result of Verdú on the capacity per unit cost [33] for the class of inflated lattice strategies. The technique of [33] also relies on "time-sharing" between the zero strategy (symbol) and an optimal strategy (symbol). The divergenece to SNR ratio in [33] reduces to the ratio $\beta(\text{SNR})$. For lower bounds for the achievable transmision rates at low SNR, when the interference is Gaussian of *finite* variance, we refer the reader to [24].

Having seen that inflated lattice strategies are preferable to ordinary lattice strategies (corresponding to $\alpha = 1$), we may attempt a further generalization by using some *nonlinear* characteristic function instead of αs. Let $g(\cdot)$ be an antisymmetric function, i.e, $g(-x) = -g(x)$, as well as satisfying $0 \leq g(x) \leq x$. Let $\delta > 0$ be such that $Eg^2(U) = P_X$ where

[5] It turns out that the improvement in mutual information possible using an optimal choice of α instead of α_{MMSE} is negligible when time-sharing is taken into account (the convex envelope in (17)).

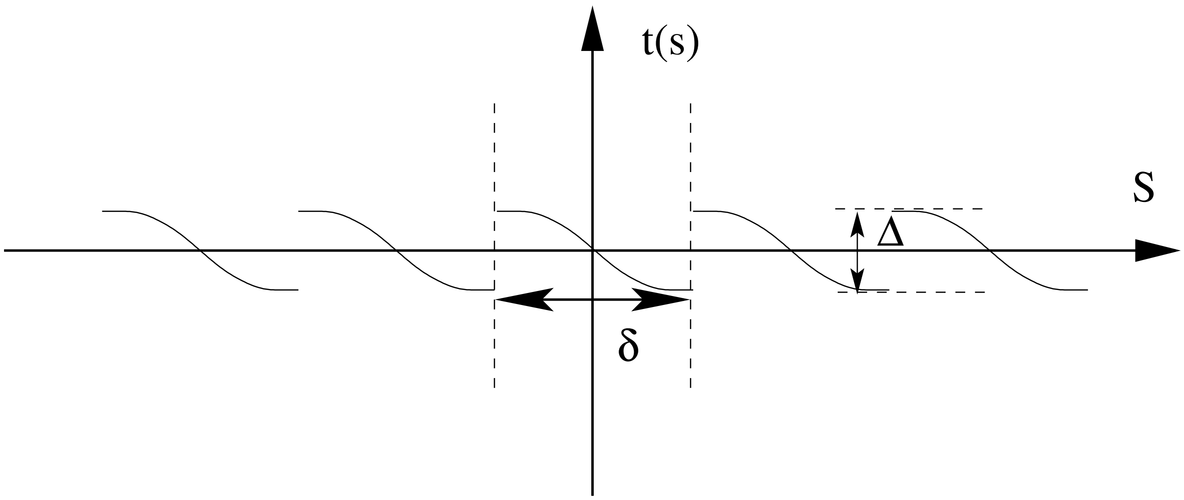

Figure 8: Generalized lattice strategy: $t_0(s) = -g(s \bmod \delta)$ with $\frac{1}{\delta}\int_{-\delta/2}^{\delta/2} g^2(s)ds = P_X$ and $\Delta = 2g(\delta/2)$.

$U \sim \mathrm{Unif}(-\delta/2, \delta/2)$. The transmision scheme would then take the form:

- *Transmitter:* For any $v \in [-\delta/2, \delta/2)$, the encoder sends:
$$x = -g(s + u + v \bmod \delta). \tag{84}$$
- *Receiver:* The receiver computes
$$y' = y + u \bmod \Delta, \tag{85}$$
where $\Delta = 2g(\delta/2)$.

Figure 8 depicts such generalized strategies. In effect, we consider general strategies as in Section 3.2, but restrict attention to *periodic* functions. The function $g(\cdot)$ therefore allows us some freedom to "shape" the self noise. Thus far, however, the attempts of the authors as well as of others [26] have not been successful to improve with such generalized lattice strategies upon the results obtained using regular inflated lattice strategies.

4 Lattice Strategies for Finite Anticipation Side Information

4.1 Rates and capacity

We can link our results for the causal setting to Costa's noncausal dirty-paper channel by allowing the encoder to anticipate k states ahead. Thus $k = 1$ corresponds to the dirty-tape channel while $k \to \infty$ corresponds to the dirty-paper channel. We obtain achievable rates for transmission with anticipation of order k. For Gaussian noise N, when k goes to infinity the corresponding rate will equal the no interference capacity, in agreement with the result of Costa [9]. The results in this section (in their preliminary version [14]) were the basis for the *nested lattice binning schemes* which were developed for the dirty-paper channel in [41].

It is important to note that for $1 < k < \infty$, we derive achievable rates but without a converse. The reason for this is two-fold: (i) we restrict attention to lattice strategies, which as we already saw in the $k = 1$ case are not necessarily optimal for general SNR and/or general noise; (ii) optimum coding with finite anticipation may also take advantage of *past* interference samples, while we consider schemes that operate only on *blocks* of length k. As a consequence, we also do not make use of the Gelfand-Pinsker capacity formula (11), but rather use k-dimensional Shannon strategies, i.e., functions of the form $\mathbf{x} = \mathbf{t}(\mathbf{s})$.

We generalize the inflated lattice encoding scheme of Section 3 by employing a lattice vector quantizer $Q_\Lambda(\cdot)$ instead of a scalar one and also having a vector dither $\mathbf{U} \sim \mathrm{Unif}(\mathcal{V}_0)$ where $\mathcal{V}_0$ is the basic Voronoi region of the lattice Λ having a second moment P_X. The transmission scheme is given by:

- *Transmitter:* For any $\mathbf{v} \in \mathcal{V}_0$, the encoder sends:

$$\mathbf{x} = [\mathbf{v} - \alpha\mathbf{s} - \mathbf{u}] \bmod \Lambda \tag{86}$$

  where $\mathbf{x} \bmod \Lambda$ is defined as $\mathbf{x} - Q_\Lambda(\mathbf{x})$.

- *Receiver:* The receiver computes

$$\mathbf{y}' = [\alpha\mathbf{y} + \mathbf{u}] \bmod \Lambda. \tag{87}$$

The resulting channel is a modulo-Λ additive noise channel described by the following lemma:

Lemma 7 (Inflated lattice lemma: vector case) *The channel defined by (1),(86) and (87) satisfies*

$$\mathbf{Y}' = \mathbf{V} + \mathbf{N}' \bmod \Lambda. \tag{88}$$

with

$$\mathbf{N}' = \Big[(1-\alpha)\mathbf{U} + \alpha\mathbf{N}\Big] \bmod \Lambda. \tag{89}$$

where $\mathbf{U}$ *is a random variable distributed uniformly over the Voronoi region of* Λ *and* $\mathbf{x} \bmod \Lambda$ *is defined as* $\mathbf{x} - Q_\Lambda(\mathbf{x})$.

The proof is the same as that of Lemma 6, replacing all scalars with their vector counterparts. We refer to this derived channel as a Modulo Lattice Additive Noise (MLAN) channel. The capacity of the MLAN channel is achieved by $V \sim \text{Unif}(\mathcal{V})$, and is given by

$$
\begin{aligned}
C_{\Lambda_k} &= \frac{1}{k} I(\mathbf{V};\mathbf{Y}) && (90)\\
&= \frac{1}{k} h(\mathbf{Y}') - \frac{1}{k} h(\mathbf{N}') && (91)\\
&= \frac{1}{2}\log(P_X/G(\Lambda)) - \frac{1}{k} h(\mathbf{N}'). && (92)
\end{aligned}
$$

Since U and N are uncorrelated and $E\{U\}=0$, we have

$$
\frac{1}{k} E\left[\|(1-\alpha)\mathbf{U} + \alpha\mathbf{N}\|^2\right] = (1-\alpha)^2 P_X + \alpha^2 P_N. \qquad (93)
$$

The minimizing α (the MMSE or "Wiener" coefficient) is $\alpha = \frac{P_X}{P_X+P_N}$ and we obtain

$$
\begin{aligned}
\frac{1}{k} E\left[\|(1-\alpha)\mathbf{U} + \alpha\mathbf{N}\|^2\right] &= \frac{P_N P_X}{P_N + P_X} && (94)\\
&= \alpha P_N. && (95)
\end{aligned}
$$

Since for a given second moment a Gaussian random vector has the greatest entropy [10] it follows that

$$
\frac{1}{k} h(\mathbf{N}') \le \frac{1}{k} h((1-\alpha)\mathbf{U} + \alpha\mathbf{N}) \le \log\left(2\pi e \frac{P_N P_X}{P_N + P_X}\right). \qquad (96)
$$

We thus have,

Theorem 3 *For any noise N and arbitrary interference S, the capacity of the MLAN channel (1) satisfies*

$$
C_{\Lambda_k} \ge \frac{1}{2}\log\left(1 + \frac{P_X}{P_N}\right) - \frac{1}{2}\log 2\pi e G(\Lambda). \qquad (97)
$$

By taking a sequence of lattices such that $G(\Lambda_k) \to \frac{1}{2\pi e}$ (see [39]), we may approach the interference free capacity arbitrarily closely for Gaussian N. Therefore for Gaussian noise as $k \to \infty$ there is no rate loss at all. This agrees with the results of [9]. Note that this result holds at any SNR. The encoding scheme is shown in Figure 9.

It is interesting to note that while in general, for any dimension k, the input maximizing the mutual information of the MLAN channel is uniquely the uniform input $\mathbf{V} \sim \text{Unif}(\mathcal{V})$, this is not the case as the dimension $k \to \infty$. In fact, for any $\frac{\text{SNR}}{1+\text{SNR}} \le \gamma \le 1$, an input $\mathbf{V} \sim \text{Unif}(\gamma\mathcal{V})$, will also be asymptotically capacity achieving. This follows from the fact that for any such input the output $\mathbf{Y}'$ in (87) will be nearly uniform. A similar result holds for Gaussian inputs $V \sim \mathcal{N}(\gamma P_X)$. We refer the reader to [19] Section 2.5 for a discussion of the implications of this fact.

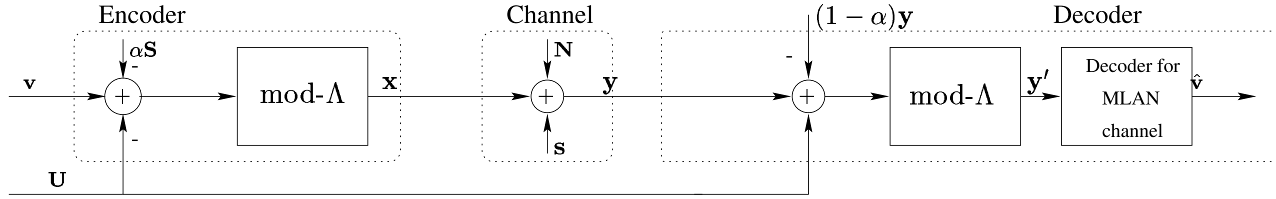

Figure 9: Encoding and decoding scheme for the dirty-paper channel.

We also note that similarly to the treatment of Section 3, in the limit of high SNR the capacity for *any* noise N, not necessarily Gaussian, is

$$C^{ncausal} = \frac{1}{2}\log(2\pi e P_X) - h(N) - o(1) \tag{98}$$

where $o(1) \to 0$ as $P_X \to \infty$. Thus, in the high SNR regime for *general* noise N the interference S does not cause rate loss, irrespective of its severeness.

4.2 Implications for the no interference case

The MLAN channel transformation is oblivious, due to the dither, to the characteristics of the interference S. Thus, we may apply the transformation even in the interference free case, i.e., in the case of an AWGN channels. It turns out that this has some nontrivial implications.

Forney et al. [20] introduced a mod-Λ channel transformation for the AWGN and showed that at high SNR, the error exponent of the resulting channel is lower bounded by the Poltyrev exponent. They also proposed strutured coset coding schemes, allowing to benefit from the group symmetry of the mod-Λ channel.

The MLAN transformation as proposed in this work generalizes the approach of [20] by incorporating MMSE scaling and by introducing dithering. This allows us to transform the power-constrained AWGN channel into an unconstrained MLAN channel, having asymptotically (in dimension) the same capacity as the original channel at *any* SNR. This insight led to the work in [17] where lattice codes are used for coding for the AWGN channel. Conversely, since the starting point for the derivation of the results of [17] is the MLAN channel they equally apply to the dirty-paper channel. In particular, it follows from [17] that the error exponent of the dirty-paper channel is lower bounded by the Poltyrev exponent [6] at any SNR.

[6] In fact, a recent result by Liu et al. [30] shows that the random coding error exponent of the MLAN channel (but with $\alpha \neq \alpha_{\mathrm{MMSE}}$) in fact is equal to that of the original AWGN channel. This implies that at rates sufficiently close to capacity, the error exponent of the dirty-paper channel equals that of an AWGN channel (at the same SNR).

4.3 Arbitrarily varying interference

We note that while we assumed that S is i.i.d. in Theorem 1, this assumption is not necessary for the universal interference cancelling scheme of Figure 4 which is virtually independent of the statistics of S. However, unlike the results of the previous sections, the dither is essential now to guarantee the achievability of these rates and cannot be regarded as an analytic tool. We modify Theorem 1 for the case of arbitrary interference as follows.

Theorem 4 (causal case) *The* randomized code *capacity of the causal SI channel (1) with arbitrarily varying interference sequence* $\{s_i\}$ *is equal to the worst-case capacity* $C^*(P_X)$ *of (18).*

Likewise, for the noncausal case we may use the lattice transmission scheme of Section 4. Thus, equation (97) holds for any interference sequence, even an arbitrarily varying one. In particular, for Gaussian N, the effect of any interference known at the transmitter noncausally can be canceled completely, with *no power loss.*

Theorem 5 (noncausal case) *The* randomized code *capacity of the noncausal SI channel (1) with arbitrarily interference sequence* $\{s_i\}$ *and Gaussian i.i.d. noise* N *is equal to the zero-interference capacity* $\frac{1}{2}\log\left(1+\frac{P_X}{P_N}\right)$.

We note that the fact that the result of Costa does not depend on the interference being Gaussian was also recognized by Cohen and Lapidoth [7, 6]. They showed that in the noncausal case with ergodic Gaussian noise N, no loss in capacity is incurred by any ergodic interference S known to the transmitter. Although the arbitrarily varying interference case treated here is more general, it necessitates common randomness which is not necessary in the ergodic interference case.

5 Summary and Extensions

We have presented a structured transmission scheme for the generalized dirty-paper channel model. Our treatment encompasses both the causal Shannon setting and the noncausal Gelfand-Pinsker setting. For the Shannon setting, an explicit capacity formula is given for the first time, albeit only for the asymptotic case of strong interference. When the intereferece is not as severe, performance may be improved and this calls for futher research. For the Gelfand-Pinsker setting, we generalized the results of Costa to arbitrary interference. The main features of the proposed schemes are lattice strategies, MMSE estimation and dithering.

The results presented may be extended in many directions. We briefly outline two generalizations. We first present a capacity theorem analogous to Theorem 1 for the *noncausal* case. This is an extension of a result presented in [8] to the case of a continuous alphabet. Similarly to (13), define the worst interference capacity of the dirty-*paper* channel as

$$C^{ncausal,worst}(P_X) = \inf_S C^{ncausal}(P_X, S). \tag{99}$$

Let

$$C^{**}(P) = \text{upper convex envelope}\{\ \tilde{\tilde{C}}(P)\ \} \tag{100}$$

where

$$\tilde{\tilde{C}}(P) = \sup_{V,t(v)} \Big\{ h(V) - h(t(V) + V + N) \Big\},$$

and where the supremum is over all continuous random variables V which are independent of N, and all functions $t(v)$ such that $E\{t(V)^2\} \leq P$.

Proposition 1 (noncausal worst-case capacity)

$$C^{ncausal,worst}(P_X) = C^{**}(P_X).$$

Since the derivation in [8] is for a finite alphabet, for completeness we include the proof in the Appendix. Note that $C^{**}(P)$ reduces to $C_L^*(P)$ of the causal case in (17) if we substitute a uniformly distributed V. Achievability of $1/2\log(1 + P_X/P_N)$ for Gaussian N can be seen by substituting $V \sim N(0, P_X/\alpha^2)$ and $t(v) = -\alpha v$, with $\alpha = \frac{P_X}{P_X+P_N}$.

We next extend the results of Section 3.5 to more general additive noise channels with side information at the transmitter than the channel model (1). Consider an additive noise channel

$$Y = X + S' + Z_S, \tag{101}$$

where S' is independent of the pair (S, Z_S). Here S' is an interference term, and the noise Z_S is dependent on S. We assume that the double side-information (S', S) is available causally to the transmitter, so X depends on (S', S) but is conditionally independent of Z_S given S. In [15], the case of a *modulo* additive noise channel (with no constraints) and with $S' = 0$ was considered.

Let $\hat{z}(S)$ be the optimal estimator of Z given S in an entropy sense. That is,

$$\hat{z}(\cdot) = \arg\min_{t:\mathcal{S}\to\mathcal{X}} h([Z - t(S)]). \tag{102}$$

We assume worst case interference S' as above. We furthermore assume high SNR in the sense that for any s we have $P_X \gg E(Z^2|S = s)$. We have the following result.

Proposition 2 (additive interference and state-dependent noise) *The (causal) capacity of the channel (101) under high SNR and strong interference conditions satisfies*

$$C^{\text{causal}}(P_X) = \frac{1}{2}\log 12P_X - h(Z - \hat{z}(S)) + o(1) \tag{103}$$

where $o(1) \to 0$ *as* $P_X \to \infty$.

Finally, an analysis of the error exponent is possible using the results of [16].

Acknowledgement

The authors wish to thank Dave Forney and Amos Lapidoth for helpful comments.

6 Appendix

A. Proof of Lemma 4

We show that

$$\limsup_{L\to\infty} h(S+X+N) - \log L \leq 0, \tag{104}$$

where $S \sim \text{Unif}(-L/2, L/2)$, (S, X) are independent of N, and where $EX^2 \leq P_X$. Let $S_1 \sim \text{Unif}(-1/2, 1/2)$. It follows that we may rewrite (104) as

$$\limsup_{L\to\infty} h\left(S_1 + (X+N)/L\right) \leq 0. \tag{105}$$

Denote $S_L = S_1 + \epsilon_L$ with $\epsilon_L = (X+N)/L$. Now let S_1^* be a Gaussian random variable having the same variance as that of S_1, i.e., $S_1^* \sim \mathcal{N}(0, 1/12)$ and let S_L^* be a Gaussian random variable having the same variance as that of S_L, i.e., $S_L^* \sim \mathcal{N}(0, \text{Var}\,(S_1 + \frac{1}{L}X + \frac{1}{L}N))$. We have

$$h(S_1) = h(S_1^*) - D(S_1 \| S_1^*) \tag{106}$$

and

$$h(S_L) = h(S_L^*) - D(S_L \| S_L^*) \tag{107}$$

where $D(\cdot\|\cdot)$ denotes Kullback-Leibler divergence (see derivation of maximum entropy property in [10]). Combining (106) and (107) we obtain

$$h(S_L) - h(S_1) = h(S_L^*) - h(S_1^*) - D(S_L \| S_L^*) + D(S_1 \| S_1^*). \tag{108}$$

Now since EX^2 and EN^2 are bounded,we have $\lim_{L\to\infty} E\epsilon_L^2 = 0$. It follows that $S_L \to S_1$ and $S_L^* \to S_1^*$ as $L \to \infty$ in the M.S. sense and in distribution. Hence, by the lower semicontinuity of the Kullback-Leibler divergence [11, 29] we have

$$\liminf_{L\to\infty} D(S_L \| S_L^*) \geq D(S_1 \| S_1^*). \tag{109}$$

Clearly, since $\text{Var}\,(S_L^*) \to \text{Var}\,(S_1^*)$, we have $\lim_{L\to\infty} [h(S_L^*) - h(S_1^*)] = 0$. Along with (109) this implies that

$$\limsup_{L\to\infty} [h(S_L) - h(S_1)] \leq 0 \tag{110}$$

which since $h(S_1) = 0$ implies (105) and thus the lemma is proved.

B. Proof of Lemma 5

In this section, we prove that

$$\tilde{C}_L(P_X) \geq h(U) - H_{\min}(U, P_X) - h(N) \tag{111}$$

$$\geq \frac{1}{2}\log(12P_X) - h(N) - \log\left(1 + \frac{\sqrt{12P_X}}{L}\right). \tag{112}$$

and also that for any $a > 0$

$$\tilde{C}_L(P_X) \leq h(U) - H_{\min}(U, [\sqrt{P_X} + a/2]^2) - h(N) + I(N; N + Z_a) \tag{113}$$

$$\leq \frac{1}{2}\log 12\left(\sqrt{P_X} + \frac{a}{2}\right)^2 - h(N) + I(N; N + Z_a) \tag{114}$$

where Z_a is independent of N and is uniformly distributed over $(-a/2, +a/2)$. To that end we first note that (see [22])

$$\log\left(\frac{L}{\sqrt{12D}}\right) \leq H_{\min}(U, D) \leq \log\left(1 + \frac{L}{\sqrt{12D}}\right) \tag{115}$$

which justifies the second step in the bounds, i.e., equations (112) and (114).

We now turn to prove (111). Let the quantizer $Q(\cdot)$ achieve (42) up to ϵ, i.e., $E(Q(U) - U)^2 \leq D$ and $H(Q(U)) = H_{\min}(U, D) + \epsilon$. We have

$$H(Q(U)) \geq I(Q(U); Q(U) + N) \tag{116}$$

$$\geq \min_{f: f(u) - u \in \mathcal{T}} I(f(U); f(U) + N) \tag{117}$$

$$= \min_{f: f(u) - u \in \mathcal{T}} h(f(U) + N) - h(N) \tag{118}$$

$$= h_{\min}(L, P_X) - h(N). \tag{119}$$

Combining (119) and the definition of $\tilde{C}_L(\cdot)$ in (15), we obtain (111).

We next prove (113). Let $Q_a(\cdot)$ be a uniform quantizer with step a, and let $Z \sim \text{Unif}(-\frac{a}{2}, \frac{a}{2})$. Since

$$Q_a(f(U) + Z) - Z \leftrightarrow f(U) \leftrightarrow f(U) + N \tag{120}$$

forms a Markov chain for any value $Z = z$, by the data processing lemma for mutual information [10] we have

$$I(f(U); f(U) + N) \geq I(Q_a(f(U) + Z); f(U) + N|Z) \tag{121}$$

$$= H(Q_a(f(U) + Z)|Z) - H(Q_a(f(U) + Z)|Z, f(U) + N). \tag{122}$$

For any value of Z, the error $Q_a(f(U) + Z) - Z - S$ of the "dithered" quantizer with respect to U is at most $|f(U) - U| + a/2$, thus the distortion is at most $(\sqrt{P_X} + a/2)^2$, so the first term above is lower bounded by $H_{\min}(U, (\sqrt{P_X} + a/2)^2)$. As for the second term , by the properties of entropy coded dithered quantization (ECDQ) [40] it can be written as

$$
\begin{aligned}
H(Q_a(f(U) + Z)|Z, f(U) + N) &= I(f(U); f(U) - Z|f(U) + N) && (123)\\
&= I(N; N + Z|f(U) + N) && (124)\\
&< I(N; N + Z). && (125)
\end{aligned}
$$

Using the l.h.s. of (115) the proof is complete.

6.1 C. Proof of Proposition 1

Note first that the upper convex envelop operation in (100) can be replaced by conditioning on a "time sharing" variable, while letting the function $t(\cdot)$ depend on this variable, i.e.,

$$
C^{**}(P) = \sup_{V,W,t(v,w)} \Big\{ h(V|W) - h(t(V,W) + V + N|W) \Big\}, \tag{126}
$$

where the supremum is over all continuous random variables V and abstract random variables W such that (V, W) are independent of N, and over all functions $t(v, w)$ such that $E\{t(V, W)^2\} \le P$. We next show that for any random interference S

$$
C^{ncausal}(P_X, S) \ge C^{**}(P_X). \tag{127}
$$

By the Gelfand-Pinsker formula (11), the capacity of the channel $Y = X + S + N$ with noncausal side information S at the transmitter is lower bounded by $I(T; Y) - I(T; S)$, for any pair of random variables X, T such that S, X, T are independent of N, and $E\{X^2\} \le P$. Let us make the following specific choice[7]: $T = (S - V, W)$ and $X = t(V, W) = t'(S, T)$, where $V, W, t(v, w)$ achieve the maximum in (126), and where (V, W) are statistically independent of (S, N). By the definition of $C^{**}(P)$ above, we have $Et'(S, T)^2 = Et(V, W)^2 \le P$. We also have

$$
\begin{aligned}
I(T; Y) - I(T; S) &= I(S - V, W; t(V, W) + S + N) - I(S - V, W; S) && (128)\\
&\ge I(S - V; S + t(V, W) + N|W) - I(S - V; S|W) && (129)\\
&= -h(S - V|S + t(V, W) + N, W) + h(S - V|S, W) && (130)\\
&= h(V|W) - h(-V - t(V, W) - N|S + t(V, W) + N, W) && (131)\\
&\ge h(V|W) - h(V + t(V, W) + N|W) && (132)\\
&= C^{**}(P_X). && (133)
\end{aligned}
$$

where the first inequality follows from the nonnegativity of the mutual information, after using the chain rule and substituting $I(W; S) = 0$; the second inequality follows since taking

[7] Here we view T as an abstract random variable.

out conditions increases the conditional differential entropy; and the last equality follows from our specific choice of V, W and t. This establishes (127). We turn to prove the converse part. We shall show that for S uniform over $(-L/2, L/2)$, we have

$$C^{ncausal}(P_X, S) \leq C^{**}(P_X) + o_L(1) \tag{134}$$

where $o_L(1)$ goes to zero as L goes to infinity. We restrict attention to the case where N has finite differential entropy, otherwise both capacities in Proposition 1 go to infinity. For any admissible T, if $I(T;Y) - I(T;S) \geq 0$ then we have

$$I(T;Y) - I(T;S) \;=\; \Big\{h(S|T) - h(Y|T)\Big\} + \Big\{h(Y) - h(S)\Big\}. \tag{135}$$

This expansion is possible since $h(S|T)$ must be finite. To see why, note that $I(T;Y)$ is finite because $h(Y|T) \geq h(N)$ is finite and $h(Y)$ is finite; thus, if $h(S|T)$ did not exist (or was minus infinity), then $I(T;S)$ would be infinite, and $I(T;Y) - I(T;S)$ would be negative. Now, from the alternative definition for $C^{**}(P)$ in (126), we see that the expression in the first brackets above is upper bounded by $C^{**}(P_X)$ (view S and T as possible choices for V and W, respectively), while by Lemma 4 the second expression in brackets above goes to zero as $L \to \infty$. This establishes (134), and together with (127) completes the proof of the proposition.

References

[1] R. Ahlswede. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Information Theory*, IT-32:621–629, Sept. 1986.

[2] R. J. Barron and B. Chen and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Information Theory*, IT–49:1159–1180, May 2003.

[3] G. Caire and S. Shamai (Shitz). On the achievable throughput of a multi-antenna gaussian broadcast channel. *IEEE Trans. Information Theory*, IT–49:1649–1706, July 2003.

[4] 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.

[5] A. S. Cohen. Communication with side information. Graduate Seminar 6.962, MIT, Cambridge, MA, Spring 2001 (http://web.mit.edu/6.962/www/www_spring_2001/schedule.html).

[6] A. S. Cohen and A. Lapidoth. Generalized writing on dirty paper. In *Proc. Int. Symp. Inform. Theory (ISIT), Lausanne, Swtizerland*, page 227, July 2002.

[7] A. S. Cohen and A. Lapidoth. The Gaussian watermarking game. *IEEE Trans. Information Theory*, June, 2002, 1639–1667. See also, *On the Gaussian Watermarking Game*, Proc. Int. Symp. Inform. Theory (ISIT), page 48, Sorrento, Italy, June 2000.

[8] A. S. Cohen and R. Zamir. The rate loss in writing on dirty paper. In preparation. See also *Allerton Conference*, Oct. 2003, pages 819–828.

[9] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Information Theory*, IT-29:439–441, May 1983.

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

[11] I. Csiszar. On an extremum problem of information theory. *Studia Scientiarum Mathematicarum Hungarica*, pages 57–70, 1974.

[12] J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. In *IEE Colloquium: Secure images and image authentication, London, UK*, April 2000.

[13] U. Erez. Noise prediction for channel coding with side information at the transmitter. Master's thesis, Tel-Aviv University, 1998.

[14] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice-strategies for cancelling known interference. In *Proceedings of ISITA 2000, Honolulu, Hawaii*, pages 681–684, Nov. 2000. See also, invited talk presented at the 2000 Cornell Summer Workshop on Information Theory, Cornell University, Ithaca, NY, USA, pages 18-19, August, 2000.

[15] U. Erez and R. Zamir. Noise prediction for channel coding with side-information at the transmitter. *IEEE Trans. Information Theory*, 46:1610–1617, July 2000.

[16] U. Erez and R. Zamir. Error exponents of modulo additive noise channels with side information at the transmitter. *IEEE Trans. Information Theory*, 47:210–218, Jan. 2001.

[17] U. Erez and R. Zamir. Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. *IEEE Trans. Information Theory*, Oct. 2004.

[18] U. Erez, R. Zamir, and S. Shamai (Shitz). Additive noise channels with side information at the transmitter. In *21st IEEE Conv. of the Electrical and Electronic Engineers in Israel, Tel Aviv, Israel*, pages 373–376, Apr. 2000.

[19] G. D. Forney, Jr. On the role of MMSE estimation in approaching the information- theoretic limits of linear Gaussian channels: Shannon meets Wiener. In *41st Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, pages 430-439, Oct. 2003.

[20] G. D. Forney, Jr., M. D. Trott, and S.-Y. Chung Sphere-bound-achieving coset codes and multilevel coset codes. *IEEE Trans. Information Theory*, IT–46:820–850, May 2000.

[21] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.

[22] A. Gyorgy and T. Linder. Optimal entropy-constrained scalar quantization of a uniform source. *IEEE Trans. Information Theory*, IT-46:2704–2711, Nov. 2000.

[23] H. Harashima and H. Miyakawa. Matched-Transmission Technique for Channels With Intersymbol Interference. *IEEE Trans. Communications*, COM–20:774–780, Aug. 1972.

[24] D. Hoesli. On the capacity per unit cost of the dirty tape channel. In *Winter School on Coding and Information Theory Monte Verita, Switzerland*, Feb. 24-27, 2003.

[25] F. Jelinek. Indecomposible channels with side information at the transmitter. *Inform. Control*, 8:36–55, 1965.

[26] R. Koetter. Private communication.

[27] A. V. Kuznetsov and B. S. Tsybakov. Coding in a memory with defective cells. *translated from Prob. Peredach. Inform.*, 10:52–60, April-June, 1974.

[28] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Trans. Information Theory*, IT–44:2148–2177, Oct. 1998.

[29] T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Information Theory*, IT-40:2026–2031, Nov. 1994.

[30] T. Liu, P. Moulin, and R. Koetter. On error exponents of nested lattice codes for the AWGN channel. *IEEE Transactions on Information Theory*, submitted (2004).

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

[32] M. Tomlinson. New Automatic Equalizer Employing Modulo Arithmetic. *Electronic Lett.*, 7:138–139, Mar. 1971.

[33] S. Verdú. On channel capacity per unit cost. *IEEE Trans. Information Theory*, IT–36:1019–1030, Sept. 1990.

[34] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the gaussian MIMO broadcast channel. In *Int. Symp. Inform. Theory (ISIT2004), Chicago, Illinois*, page 174, June, 2004.

[35] F. M. J. Willems. On Gaussian channels with side information at the transmitter. In *Proc. of the Ninth Symposium on Information Theory in the Benelux*, Enschede, The Netherlands, 1988.

[36] F. M. J. Willems. Signalling for the Gaussian channel with side information at the transmitter. In *Proc. Int. Symp. Inform. Theory (ISIT), Sorrento, Italy*, page 348, June 2000.

[37] W. Yu and J.M. Cioffi. Trellis precoding for the broadcast channel, preprint, Globecom 2001. See also: The sum capacity of a Gaussian vector broadcast channel, submitted to IEEE Trans. Inform. Theory.

[38] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang. Writing on colored paper. In *Proc. Int. Symp. Inform. Theory (ISIT2001)*, page 302, Washington, DC., USA, June 24–29, 2001.

[39] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, IT-42:1152–1159, July 1996.

[40] R. Zamir and M. Feder. On universal quantization by randomized uniform / lattice quantizer. *IEEE Trans. Information Theory*, pages 428–436, March 1992.

[41] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Information Theory*, IT–48:1250 –1276, June 2002.