On the Robustness of Dirty Paper Coding

Anatoly Khina and Uri Erez, *Member, IEEE*

***Abstract*—A dirty-paper channel is considered, where the transmitter knows the interference sequence up to a constant multiplicative factor, known only to the receiver. Lower bounds on the achievable rate of communication are derived by proposing a coding scheme that partially compensates for the imprecise channel knowledge. We focus on a communication scenario where the signal-to-noise ratio is high. Our approach is based on analyzing the performance achievable using lattice-based coding schemes. When the power of the interference is finite, we show that the achievable rate of this lattice-based coding scheme may be improved by a judicious choice of the scaling parameter at the receiver. We further show that the communication rate may be improved, for finite as well as infinite interference power, by allowing randomized scaling at the transmitter.**

***Index Terms*—Dirty paper coding, Tomlinson-Harashima precoding, channel estimation error, compound channel, channel state information.**

I. Introduction

THE dirty-paper (DP) channel, first introduced by Costa [1], provides an information theoretic framework for the study of interference cancellation techniques for interference known to the transmitter. The DP channel model has since been further studied and applied to different communication scenarios such as ISI channels (see, e.g., [2]), the MIMO Gaussian broadcast channel [3], [4], [5] and information embedding [6]. The DP channel is given by[1]

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

where X is the channel input and is subject to an average power constraint P_X, N is AWGN with variance P_N and S is interference which is known causally ("causal DP") or non-causally ("non-causal DP") to the transmitter but not to the receiver. We note that the DP channel expressed in (1) models communication scenarios where the channel (i.e., all channel coefficients) is known *perfectly* to both transmission ends.

Costa [1] showed that, for an i.i.d. Gaussian interference with arbitrary power, the capacity in the non-causal scenario is equal to that of the interference-free AWGN channel, $\frac{1}{2}\log(1+\mathrm{SNR})$, where $\mathrm{SNR} \triangleq P_X/P_N$. This result was extended in [7], [8] to the case of general interference.

In this work we focus our attention on scalar precoding, both since it results in simpler coding schemes but also since the benefit of using a vector approach (at least using the methods we study) diminishes in the presence of imprecise channel knowledge, as will be shown in Section VI. Note that scalar precoding is applicable when the interference is known causally, whereas vector approaches require non-causal knowledge, see, e.g., [8]. We consider the *real* channel case; for treatment of the case of imperfect phase knowledge, in the *complex* channel case, see [9], [10].

The capacity of the DP channel with causal knowledge of the interference is not known but upper and lower bounds were found in [8], which coincide in the high-SNR regime, thus establishing the capacity for this case to be $\frac{1}{2}\log(1+\mathrm{SNR}) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - o(1)$, where $o(1) \to 0$ as $\mathrm{SNR} \to \infty$. Thus, causality incurs a rate loss of $\frac{1}{2}\log(\frac{2\pi e}{12})$ (commonly known as the shaping loss), relative to the capacity of the interference-free AWGN channel. This result implies that in the limit of strong interference and high SNR, the well-known Tomlinson-Harashima precoding (THP) technique [11], [12] is optimal. For general SNRs, the lattice-based coding techniques of [13], [14], [8] are an extension of Tomlinson-Harashima precoding where a scaling parameter is introduced at the transmitter and receiver. In this work the term Tomlinson-Harashima precoding is used in this wider sense. A review of THP and its extensions is presented in [8].

In many cases of interest, the transmitter has imprecise channel knowledge. For instance in a multi-user broadcast scenario, the interference sequence S corresponds to the signal intended to another user *multiplied* by a channel gain. While the transmitter knows the transmitted interfering signal, only an estimate of the channel gain is known (for instance by quantized feedback; see, e.g., [15]). This leads to the question, studied in this paper, of how sensitive dirty-paper coding (DPC) is to imprecise channel knowledge. We address this question by adapting the extended Tomlinson-Harashima precoding as presented in [8] to the case of imprecise channel knowledge.

The paper is organized as follows: in Section II we discuss the compound causal dirty-paper channel. We then turn, In Section III, to the more general problem of the compound state-dependent discrete memoryless channel (DMC) and determine its capacity where the state is known causally. In Section IV we consider the case where the interference S is i.i.d. (of some distribution) with power P_S, and show how using a modified front-end can outperform the regular DP channel receiver, which ignores the inaccuracy in the channel knowledge. We then concentrate on the high-SNR regime and show that using *random* scaling improves performance further, in Section V. Finally in Section VI, we discuss the extension of the scheme to the non-causal case, as well as some implications to multiple-input multiple-output (MIMO) broadcast channels with imperfect channel knowledge at the transmitter.

The material in this paper was presented in part at the 2008 IEEE Information Theory Workshop, Porto, Portugal. This work was supported in part by the Israeli ministry of trade and commerce as part of Nehusha/iSMART project. The work of A. Khina was supported in part by a fellowship of the Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel Aviv University. The work of U. Erez was supported in part by the ISF under grant # 1234/08 and by the Braun-Roger-Siegl Foundation.

[1]We denote random variables by uppercase letters. Vectors are denoted by bold and random vectors by bold uppercase letters.

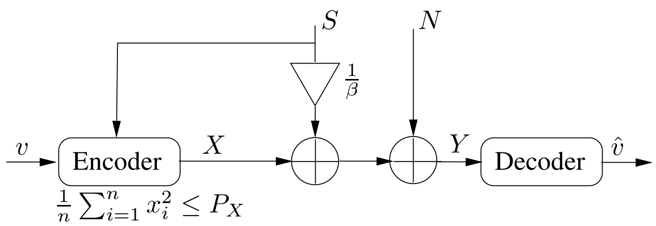

Fig. 1. The compound dirty-paper channel.

II. Channel Model and Motivation

We consider the channel model, depicted in Figure 1:

$$Y = X + \frac{S}{\beta} + N, \tag{2}$$

where $\beta \in \mathcal{I}_\Delta = [1-\Delta, 1+\Delta]$ is a constant that is unknown to the transmitter. Thus, Δ is a measure of the degree of channel uncertainty.

Consider first the limit of high SNR. At first glance, one might suspect that a reasonable approach could be to use standard THP since, as mentioned above, THP is optimal at high SNR in the perfect channel knowledge case. This would correspond to pre-subtracting the interference S at the transmitter, applying a modulo operation and treating the residual interference as noise. However, the residual interference, namely $(1-\frac{1}{\beta})S$, left at the receiver, may be large if the power of the interference is large. In fact, in the limit $P_S \to \infty$, the achievable rate, for reliable communication using this approach, would vanish. Thus naïve implementation of THP is not robust to channel uncertainty.

We observe, in Section IV, that by using a carefully chosen scaling parameter at the receiver, reliable communication, at strictly positive rate, is possible, regardless of the interference power. The optimal scaling parameter does, however, depend on the power of the interference and should strike a balance between the residual interference, the "self-noise" component, and the Gaussian noise.

We then show, in Section V, that performance may further be improved by using randomized (time-varying) scaling at the transmitter. We begin by examining the more general problem of compound channel with side information, considered previously by Mitran, Devroye and Tarokh [16].

III. Compound Channels with Causal State Information at the Transmitter

The causal DP channel model (1) is a special case of the more general problem of a channel with side information (SI) at the transmitter. This problem was first introduced by Shannon [17], who considered a DMC whose transition matrix depends on the channel state s, where the latter is independent of the message W that is sent, i.i.d. and known causally to the transmitter but not to the receiver. This channel is described by

$$p(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{x}) = \prod_i p(y_i|s_i,x_i)$$
$$p(\boldsymbol{s}) = \prod_i p(s_i),$$

where $x \in \mathcal{X}$ is the channel input, $y \in \mathcal{Y}$ is the channel output, $s \in \mathcal{S}$; and $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$ denote the channel input alphabet, channel output alphabet and state alphabet, respectively, all of which are finite sets. Shannon showed that the capacity of the above channel is equal to that of an equivalent DMC whose inputs are mappings $t \in \mathcal{T}$, which will be referred to hereafter as strategies, from $\mathcal{S}$ to $\mathcal{X}$ ($\mathcal{T}$ denotes the set of all mappings from $\mathcal{S}$ to $\mathcal{X}$), and the corresponding transition probabilities of this channel are

$$p(y|t) = \sum_s p(s)p(y|x=t(s),s).$$

Note that this result uses mappings of the *current* state only, even though the transmitter has access to all past states.

A compound (discrete memoryless) channel is a channel whose transition matrix depends on a parameter β, which is constant and not known to the transmitter but is known to the receiver[2] and takes values from $\mathcal{B}$, where the alphabet $\mathcal{B}$ is a finite set.[3]

The ("worst-case") capacity of this channel was found, by several different authors [19], [20], [21] (see also [18]), to be

$$C = \sup_{p(x)\in\mathcal{P}(\mathcal{X})} \inf_{\beta\in\mathcal{B}} I_\beta(X;Y),$$

where $I_\beta(X;Y)$ denotes the mutual information of X and Y with respect to the transition matrix $p_\beta(y|x)$ and $\mathcal{P}(\mathcal{X})$ is the set of all probability vectors over $\mathcal{X}$. This result may be easily extended to the case of a compound channel with SI available *causally* to the transmitter, as implied by the following theorem, which is proved in Appendix A.

Theorem 1: The worst-case capacity of a compound DMC with causal SI at the transmitter is given by

$$C = \sup_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} I_\beta(T;Y),$$

where $\mathcal{T}$ denotes the set of all random strategy functions of the form $t : \mathcal{S} \longrightarrow \mathcal{X}$, and $\mathcal{P}(\mathcal{T})$ is the set of all probability vectors over $\mathcal{T}$.

Remark 1: The case of non-causal SI is more difficult. The converse of Gel'fand-Pinsker [22] is not easily extended to the compound scenario. In [16] Mitran, Devroye and Tarokh derived upper and lower *single-letter* bounds for the capacity with non-causal SI. Using Theorem 1, a *non single-letter* expression for the worst-case capacity in the non-causal SI case, using k-dimensional vector strategies and taking k to infinity, follows:

$$C^{\text{non-casual}} = \limsup_{k\to\infty} \sup_{p(\boldsymbol{t})} \inf_{\beta} \frac{1}{k} I_\beta(\boldsymbol{T};\boldsymbol{Y}).$$

IV. Compensation for Channel Uncertainty at Receiver

The compound DP channel was defined in (2). In this section, we consider the case of i.i.d. interference of finite

[2] Sometimes a channel is said to be compound if β is not known at *both* ends. The capacity however is the same in both scenarios (see, e.g., Wolfowitz [18, chap. 4]), as the receiver may estimate β to within any desired accuracy (with probability going to one), using a negligible portion of the block length.

[3] $\mathcal{B}$ plays the role of the interval $\mathcal{I}_\Delta$ of Section II.

power P_S. The results of Section III may readily be extended to continuous alphabet and to incorporate an input constraint (see [16], Sec. IV). Thus, Theorem 1 holds for this setting as well.

Since the capacity of the dirty-paper channel with causal SI is unknown even in the standard (non compound) setting, we do not attempt to explicitly find the capacity in the compound setting. Rather, we shall examine the performance of THP-like precoding schemes and suggest methods by which the lack of perfect channel knowledge at the transmitter may be taken into account, and partially compensated for.

A. THP With Imprecise Channel Knowledge

Denote the one-dimensional lattice, whose basic interval is $\mathcal{V}_0 \triangleq \left[-\frac{L}{2}\frac{L}{2}\right)$, by $\Lambda = L\mathbb{Z}$ where L is chosen such that $P_X = \frac{L^2}{12}$, and by SIR $\triangleq \beta^2 \frac{P_X}{P_S}$ the signal-to-interference ratio. Let $U \sim \mathrm{Unif}(\mathcal{V}_0)$ be a random variable (dither) known to both transmitter and receiver. We consider an extended THP scheme given by:

- Transmitter: for any $v \in \mathcal{V}_0$, the transmitted signal is
$$X = [v - \alpha_T S - U] \bmod \Lambda.$$
- Receiver: the receiver computes,
$$Y' = [\alpha_R Y + U] \bmod \Lambda.$$

The channel from v to Y' can be rewritten as:

$$\begin{aligned} Y' &= [\alpha_R Y + U] \bmod \Lambda \\ &= \Big[v - (v - \alpha_T S - U) + \alpha_R X \\ &\qquad +(\alpha_R - \alpha_T\beta)\frac{S}{\beta} + \alpha_R N\Big] \bmod \Lambda \\ &= \Big[v - (1-\alpha_R)X + \\ &\qquad +(\alpha_R - \alpha_T\beta)\frac{S}{\beta} + \alpha_R N\Big] \bmod \Lambda. \end{aligned}$$

Due to the dither U, X is independent of S and of the information signal V, and is uniform over Λ (see, e.g., [8], [23]). Therefore, this channel is equivalent to the modulo-additive channel:

$$Y' = [v + N_{\text{eff}}^{\beta}] \bmod \Lambda \tag{3}$$

$$N_{\text{eff}}^{\beta} \triangleq (1-\alpha_R)U + (\alpha_R - \alpha_T\beta)\frac{S}{\beta} + \alpha_R N, \tag{4}$$

where N_{eff}^{β} is the "effective noise", composed of a "self noise" component $(1-\alpha_R)U$, a residual interference component $(\alpha_R - \alpha_T\beta)\frac{S}{\beta}$ and a Gaussian noise component $\alpha_R N$. The average power of the effective noise is

$$P_{N_{\text{eff}}^{\beta}} = (1-\alpha_R)^2 P_X + (\alpha_R - \alpha_T\beta)^2 \frac{P_S}{\beta^2} + {\alpha_R}^2 P_N.$$

and the corresponding signal-to-effective noise power is

$$\begin{aligned} \mathrm{SNR}_{\text{eff}} &\triangleq \frac{P_X}{P_{N_{\text{eff}}^{\beta}}} \\ &= \left[(1-\alpha_R)^2 + \frac{(\alpha_R - \alpha_T\beta)^2}{\mathrm{SIR}} + \frac{{\alpha_R}^2}{\mathrm{SNR}}\right]^{-1}. \end{aligned}$$

We denote the maximal achievable rate under these settings by R^d_{THP}, where "d" stands for "deterministic" (choice of) α_T (in contrast to the random strategies treated in Section V), and the achievable rate for a specific triplet $(\alpha_T, \alpha_R, \beta)$ by $R^d_{\text{THP}}(\alpha_T, \alpha_R, \beta)$.

Lemma 1: The maximal achievable rate using the scheme described above is lower-bounded by:

$$\begin{aligned} R^d_{\text{THP}} \geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} & \frac{1}{2}\log(\mathrm{SNR}_{\text{eff}}) \\ &+ \varepsilon(\beta, \alpha_T, \alpha_R) - \frac{1}{2}\log\left(\frac{2\pi \mathrm{e}}{12}\right), \end{aligned}$$

where $\varepsilon(\beta, \alpha_T, \alpha_R) \triangleq h\left(N^{\beta}_{\text{eff,G}}\right) - h\left(N^{\beta}_{\text{eff}}\right)$, $h(\cdot)$ denotes the differential entropy and $N^{\beta}_{\text{eff,G}}$ is Gaussian with the same variance as N^{β}_{eff}.
Thus $\varepsilon(\beta, \alpha_T, \alpha_R)$ is a measure of non-Gaussianity. Note that $\varepsilon(\beta, \alpha_T, \alpha_R) \geq 0$.

Proof: First note that for any triplet $(\alpha_T, \alpha_R, \beta)$, the mutual information is maximized by taking $V \sim \mathrm{Unif}(\mathcal{V}_0)$. Hence:

$$\begin{aligned} R^d_{\text{THP}}(\alpha_T, \alpha_R, \beta) &= h(Y') - h(Y'|V) \\ &= \log(L) - h([N^{\beta}_{\text{eff}}] \bmod \Lambda). \end{aligned}$$

The maximal achievable rate R^d_{THP} is therefore lower-bounded by

$$\begin{aligned} R^d_{\text{THP}} &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} R^d_{\text{THP}}(\alpha_T, \alpha_R, \beta) \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\log(L) - h([N^{\beta}_{\text{eff}}] \bmod \Lambda)\right] \\ &\geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \left[\frac{1}{2}\log(L^2) - h(N^{\beta}_{\text{eff}})\right] \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \Big[\frac{1}{2}\log(L^2) - h\left(N^{\beta}_{\text{eff,G}}\right) \\ &\qquad + \varepsilon(\beta, \alpha_T, \alpha_R)\Big] \\ &= \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \Big[\frac{1}{2}\log(\mathrm{SNR}_{\text{eff}}) \\ &\qquad + \varepsilon(\beta, \alpha_T, \alpha_R) - \frac{1}{2}\log\left(\frac{2\pi \mathrm{e}}{12}\right)\Big]. \end{aligned}$$

where $\varepsilon(\beta, \alpha_T, \alpha_R) \triangleq h\left(N^{\beta}_{\text{eff,G}}\right) - h\left(N^{\beta}_{\text{eff}}\right)$ and $N^{\beta}_{\text{eff,G}}$ is Gaussian with the same power as N^{β}_{eff}. ■

We are left with the task of choosing $\alpha_T, \alpha_R, \beta$.

B. Naïve Approach

One could ignore the presence of the inaccuracy factor β and apply standard THP, using the parameters $\alpha_R = \alpha_T = \alpha_{\text{MMSE}} \triangleq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$, which is the best selection of α_R and α_T in this case (see [8]). This gives rise to the following signal-to-effective noise ratio at the receiver:

$$\begin{aligned} \mathrm{SNR}_{\text{eff}} &= \lambda_{\text{Naïve}}(\beta)(1+\mathrm{SNR}) \\ \lambda_{\text{Naïve}}(\beta) &\triangleq \frac{1}{1 + \frac{1}{\mathrm{SIR}} + \frac{\mathrm{SNR}}{\mathrm{SIR}}(1-\beta)^2}. \end{aligned}$$

Note that since $(1+\text{SNR})$ is the output SNR in the perfect SI case, the loss due to the imprecision $(1-\beta)$ is manifested in the multiplicative factor $0 < \lambda_{\text{Naïve}}(\beta) \leq 1$.

Moreover, when the interference is very strong, i.e., SIR $\rightarrow 0$, even if the SNR is high, the effective SNR goes to zero along with the rate (as further explained in Section V-A1). Nonetheless, a strictly positive rate can be achieved in this scheme, using a smarter Rx-Tx pair, as is shown in the following sections.

C. Smart Receiver - Ignorant Transmitter

Using the fact that $\varepsilon(\beta, \alpha_T, \alpha_R) \geq 0$, we can further loosen the lower-bound of Lemma 1 to

$$R^d_{\text{THP}} \geq \max_{\alpha_T} \min_{\beta \in \mathcal{I}_\Delta} \max_{\alpha_R} \frac{1}{2}\log(\text{SNR}_{\text{eff}}) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right). \tag{5}$$

Note that optimizing the r.h.s. of (5) is equivalent to maximizing SNR_{eff} with respect to $\{\alpha_T, \alpha_R\}$. In this section we shall optimize with respect to α_R ("smart receiver") and use $\alpha_T = \alpha_{\text{MMSE}} \triangleq \frac{\text{SNR}}{1+\text{SNR}}$ ("ignorant transmitter") as was done in Section IV-B, and leave the treatment of a smarter selection of α_T ("smart transmitter") to Section V.

By solving the problem of maximizing the signal-to-effective noise ratio, the following α_R value and corresponding SNR_{eff} are obtained:

$$\begin{aligned} \alpha_T^{\text{MMSE}} &= \alpha_{\text{MMSE}} \triangleq \frac{\text{SNR}}{1+\text{SNR}} \\ \alpha_R^{\text{MMSE}} &= \frac{1 + \frac{\alpha_T^{\text{MMSE}}\beta}{\text{SIR}}}{1 + \frac{1}{\text{SIR}} + \frac{1}{\text{SNR}}} \\ \text{SNR}_{\text{eff}} &= \lambda_{\text{MMSE}}(\beta)(1+\text{SNR}) \\ \lambda_{\text{MMSE}}(\beta) &\triangleq \frac{1 + \frac{1}{\text{SIR}} + \frac{1}{\text{SNR}}}{1 + \frac{1}{\text{SIR}} + \frac{1}{\text{SNR}} + \frac{\text{SNR}}{\text{SIR}}(1-\beta)^2}, \end{aligned} \tag{6}$$

where again, the loss due to β is captured by $0 < \lambda_{\text{MMSE}} \leq 1$. Note that the loss in SNR_{eff} is smaller than that of the naïve approach since $\lambda_{\text{Naïve}}(\beta) < \lambda_{\text{MMSE}}(\beta)$, for every β.

Using α_R^{MMSE} rather than the standard $\alpha_R = \frac{\text{SNR}}{\text{SNR}+1}$ improves SNR_{eff}, for all values of β. A lower-bound on the achievable rate is therefore given by,

$$R^d_{\text{THP}} \geq \frac{1}{2}\log(1+\text{SNR}) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(\frac{1}{\lambda_{\text{MMSE}}(\beta = 1+\Delta)}\right).$$

The gains of the this approach over the naïve one of Section IV-B, for different SNR values and $\Delta = 1/3$, are depicted in Figure 2.

Remark 2:

1. In the weak interference regime, SIR $\rightarrow \infty$, we have $\lambda_{\text{MMSE}}(\beta) \rightarrow 1$ (for all β) and hence $\alpha_R = \frac{\text{SNR}}{\text{SNR}+1}$ and $\text{SNR}_{\text{eff}} = 1 + \text{SNR}$. This is of course a non-interesting case as THP is unattractive in this regime.
2. In the strong interference regime, SIR $\rightarrow 0$, the residual interference component of N^β_{eff} has to be completely cancelled. This is done by selecting $\alpha_R = \alpha_T \beta$ and results in an effective noise with finite power (dictated by the magnitude of Δ). Thus reliable communication is possible at strictly positive rates, even when the interference is arbitrarily strong.

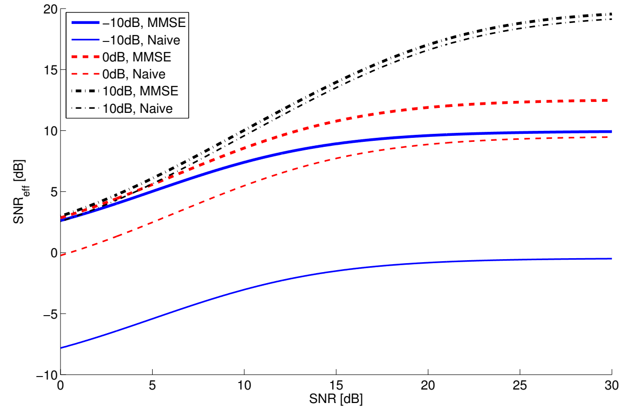

Fig. 2. SNR_{eff} as a function of SNR for different SIR values and $\Delta = 1/3$. Continuous line - (-10)dB; dashed line - $0dB$; dot-dashed - $10dB$. Within each pair: Thick line - "Smart Rx" approach; thin line - "Naïve" approach.

D. High SNR Regime

In the high SNR regime, i.e., SNR $\gg 1$, the choice $\alpha_T = 1$ becomes optimal. Using this choice of α_T in (6), we have the following effective SNR:

$$\text{SNR}_{\text{eff}} \geq \frac{1+\text{SIR}}{(1-\beta)^2}\Big(1 - o(1)\Big),$$

where $o(1) \rightarrow 0$ as SNR $\rightarrow \infty$. By substituting this effective SNR in the lower-bound of Lemma 1, we obtain the following achievable rate:

$$\begin{aligned} R^d_{\text{THP}} \geq\ & \frac{1}{2}\log(1+\text{SIR}) + \log\left(\frac{1}{\Delta}\right) \\ & - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) + \min_{\beta \in \mathcal{I}_\Delta} \varepsilon(\beta, \alpha_T = 1, \alpha_R) - o(1), \end{aligned} \tag{7}$$

where again, $o(1) \rightarrow 0$ as SNR $\rightarrow \infty$.

Remark 3:

1. In the case of strong interference and high SNR (SIR $\rightarrow 0$, SNR $\rightarrow \infty$), with the choice of $\alpha_T = 1$ and the corresponding optimal choice of $\alpha_R^{\text{MMSE}} = \frac{1}{\beta}$, the effective noise N^β_{eff} has virtually only a self-noise component, i.e., $N^\beta_{\text{eff}} \approx (1-\alpha_R)U$. Hence, $\varepsilon(\beta, \alpha_T = 1, \alpha_R^{\text{MMSE}}) \rightarrow \frac{1}{2}\log\left(\frac{2\pi e}{12}\right)$ as SNR $\rightarrow \infty$ (for $\forall \beta \in \mathcal{I}_\Delta$). Thus, there is no shaping loss compared to high-dimensional lattices in this case, as further explained in Section VI-A, and the corresponding achievable rate is $R^d_{\text{THP}} = \log\left(\frac{1}{\Delta}\right) - o(1)$.
2. The lower bound of (7) can be evaluated for any specific distribution of S, by calculating $\min_\beta \varepsilon(\beta, \alpha_T = 1, \alpha_R^{\text{MMSE}})$. For instance, if S is uniform, that is the limit of an M-PAM constellation ($M \rightarrow \infty$), then R^d_{THP} can be lower-bounded by

$$R^d_{\text{THP}} \geq \frac{1}{2}\log(1+\text{SIR}) + \log(\frac{1}{\Delta}) - \frac{1}{2}\log\left(\frac{e}{2}\right) - o(1),$$

where $o(1) \to 0$ as SNR $\to \infty$. This can be done for a general SNR as well, viz., not only in the limit of high-SNR.

3. Even in the limit of strong interference, i.e., SIR $\to 0$, in the "smart-receiver" approach, $\text{SNR}_{\text{eff}} \geq 1$, due to the extra 1 in the nominator. Hence a strictly positive rate is achieved in this regime, contrary to the effective SNR of the naïve approach, $\frac{\text{SIR}}{(1-\beta)^2}$, which goes to zero along with the achievable rate.
4. In the case of equal interference and signal powers, $\text{SIR} = 1$, there is a gain of 3dB over the naïve approach, as is seen in Figure 2.
5. When the signal and interference have the same power, $\text{SIR} = 1$, α_R^{MMSE} strikes a balance between the two effective noise components, the powers of which become both equal to $\frac{1}{4}(1-\beta)^2 P_X$ for $\alpha_R = \alpha_R^{\text{MMSE}}$. Thus, α_R^{MMSE} gives a total noise power of $P_{N_{\text{eff}}^\beta} = \frac{1}{2}(1-\beta)^2 P_X$, which is half the noise power obtained by cancelling out the interference component completely ($\alpha_R = \beta$), or alternately, half the noise power obtained by cancelling out completely the self-noise component ($\alpha_R = 1$).
6. Due to the modulo operation at the receiver side and since the effective noise is not Gaussian, the choice $\alpha_R = \alpha_R^{\text{MMSE}}$ does not strictly maximize the mutual information $I(V;Y')$, but rather is a reasonable approximate solution. Moreover, in the compound case, in contrast to the perfect SI case, minimizing the mean-square error (MSE) is not equivalent to maximizing the effective SNR or the rate, as may be seen in Example 1 below.

V. Randomized Scaling at Transmitter

For simplicity, we now restrict our attention to the case of strong interference and high SNR, i.e., $\text{SIR} \to 0, \text{SNR} \to \infty$. More specifically, we consider a noise-free channel model:

$$Y = X + \frac{S}{\beta}\,.$$

In this case, the receiver must completely cancel out the interference by choosing $\alpha_R = \beta \cdot \alpha_T$. Note that if β were known at the transmitter, the capacity would be infinite.

We now investigate whether performance may be improved by introducing a random scaling factor α at the transmitter ($\alpha_T = \frac{1}{\alpha}$), which is chosen in an i.i.d. manner at each time instance and is assumed known to both transmitter and receiver. Thus, we consider the following transmission scheme:

- Transmitter: for any $v \in \mathcal{V}_0$, sends
$$X = [v - \frac{1}{\alpha} S - U] \bmod \Lambda.$$
- Receiver: the receiver applies the front end operation,
$$Y' = [\alpha_R Y + U] \bmod \Lambda,$$
where $\alpha_R = \beta/\alpha$.

The above channel can be shown (by retracing the steps of (3), (4)) to be equivalent to the modulo-additive channel

$$Y' = \left[v + N_{\text{eff}}^\beta\right] \bmod \Lambda,$$

where $N_{\text{eff}}^\beta \triangleq \frac{\alpha-\beta}{\alpha} U$. Note that the average power of N_{eff}^β now varies from symbol to symbol according to the value of α.

The rationale for considering such scaling at the transmitter is that had the transmitter known β, it would choose $\alpha = \beta$ to match the actual interference as experienced at the receiver. By using randomization, this will occur some of the time. Since β is unknown however (to the transmitter), one might suspect that using a deterministic selection of $\alpha = 1$ may be optimal, as was done in Section IV-A. However, due to convexity, it turns out that a better approach is to let α vary[4] from symbol to symbol (or block to block) within the interval of uncertainty $\mathcal{I}_\Delta$.

Example 1: To further motivate this we shall look at the simple case of a binary alphabet $\beta \in \mathcal{B} = \{1 \pm \Delta\}$. In this case the best deterministic selection of α is $\alpha = 1$, which gives rise to a finite rate for every $\beta \in \mathcal{B}$. However, consider choosing α at random, in an i.i.d. manner for each symbol, according to

$$P(\alpha = 1-\Delta) = P(\alpha = 1+\Delta) = \frac{1}{2}\,.$$

When the transmitter uses this selection policy of α, approximately for half of the transmitted symbols the chosen α will equal β, even though β is unknown to the transmitter; while for the other half of the symbols, the mismatch between β and the chosen α will be greater than that obtained by taking $\alpha = 1$. Since, whenever the chosen α is (exactly) equal to β, the mutual information between the conveyed message v and the channel output Y is *infinite*, since the channel is noiseless, the total rate is *infinite* as well.

Remark 4: In the absence of noise, if β takes only a finite number of values, i.e. $|\mathcal{B}| < \infty$, then the achievable rate is infinite. The achievability is shown by generalizing the idea of the binary case: by varying α in an .i.i.d manner from symbol to symbol according to the uniform distribution $\alpha \sim \text{Unif}(\mathcal{B})$. However a straightforward extension to the case of an infinite countable cardinality (all the more to a continuous alphabet), is not possible.

We denote the maximal achievable rate of this scheme by R_{THP}^r, where "r" stands for "random". It is given by:

$$R_{\text{THP}}^r = \max_{f(\alpha)} R_{\text{THP}}^r(f) = \max_{f(\alpha)} \min_{\beta \in \mathcal{I}_\Delta} I_\beta(V;Y'|\alpha), \qquad (8)$$

where $f(\alpha)$ is the PDF according to which α is drawn and $R_{\text{THP}}^r(f)$ denotes the mutual information corresponding to the specific choice of $f(\alpha)$. Note that in this case the distribution of α that minimizes the MSE is not necessarily the one that maximizes SNR_{eff}. Hence a direct optimization of (8) needs to be done. Note that in this case the effective noise will vary with time along with α.

Lemma 2: The maximal achievable rate, when $\Delta \leq \frac{1}{3}$, for the noiseless DP channel, using the "extended THP scheme", is

$$R_{\text{THP}}^r = \max_{\substack{f(\alpha):\\ \text{Supp}\{f(\alpha)\} \subseteq \mathcal{I}_\Delta}} \min_{\beta \in \mathcal{I}_\Delta} -E_\alpha \left[\log\left|\frac{\alpha-\beta}{\alpha}\right|\right]. \qquad (9)$$

[4]Note that by doing so, we in effect extend the class of strategies used in the transmission scheme.

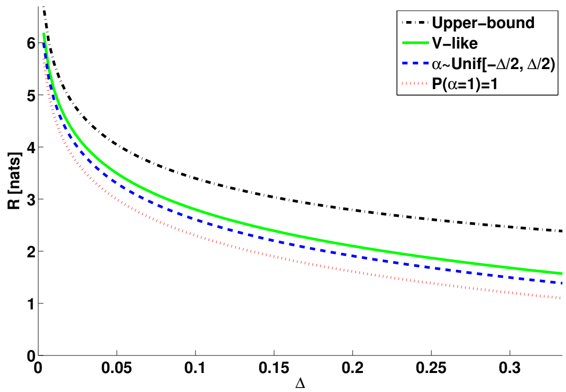

Fig. 3. Achievable rates and upper-bound on the deterministic THP scheme.

The proof of this lemma is given in Appendix B along with the treatment of the case of $\Delta > \frac{1}{3}$.

Finding the optimal distribution of $f(\cdot)$ in (8) is cumbersome. Instead, we suggest several choices for the distribution $f(\cdot)$ which achieve better performance than that of any deterministic selection of α as well as derive an upper bound on R^r_{THP}.

A. Quantifying the Achievable Rates

As indicated by Lemma 2, we restrict attention to the case of $\Delta \leq \frac{1}{3}$. We consider three different distributions for α: deterministic selection, a uniform distribution and a V-like distribution.

1) Deterministic Selection: One easily verifies that the value of α, which achieves the maximal rate, is $\alpha = 1$ and the corresponding rate is

$$R^r_{\text{THP}}(f_{\text{Deter}}) = -\log\Delta = \log\frac{1}{\Delta}.$$

Note that this result coincides with the result for R^d_{THP} of Section IV-A ($\varepsilon(\beta, \alpha_R) - \frac{1}{2}\log(\frac{2\pi e}{12})$ is equal to zero in this case, as mentioned in Remark 3).

2) Uniform Distribution: Taking $\alpha \sim \text{Unif}(\mathcal{I}_\Delta)$ yields the following achievable rate:

$$\begin{aligned} R^r_{\text{THP}}(f_{\text{Unif}}) =& \frac{1}{2\Delta}\big[(1+\Delta)\log(1+\Delta) \\ &- (1-\Delta)\log(1-\Delta) - 2\Delta\log(2\Delta)\big]. \end{aligned}$$

Hence, even this simple randomization improves on the deterministic selection, as may be seen in Figure 3.

3) V-like Distribution: A further improvement is obtained by taking a V-like distribution,

$$f_{\text{V-like}}(\alpha) = \frac{|\alpha - 1|}{\Delta^2}, \qquad |\alpha - 1| \leq \Delta.$$

The resulting rate is

$$\begin{aligned} R^r_{\text{THP}}(f_{\text{V-like}}) = &-\frac{1}{2\Delta^2}\big[(1-\Delta^2)\log(1-\Delta^2) \\ &+ \Delta^2\log(\Delta^2)\big]. \end{aligned}$$

We have not pursued numerical optimization of $f(\cdot)$. We note that none of the three distributions above are optimal since $I_\beta(V;Y')$ varies with β. Moreover, the optimal PDF will not be totally symmetric around 1 due to the first term $\log(\alpha)$ in (16). This term becomes, however, less and less significant (and hence the optimal PDF more and more symmetrical) for small Δ. We next derive an upper bound on the achievable rate which holds for any choice of $f(\cdot)$.

B. Upper-Bound on Achievable Rates

Lemma 3: The rate achievable, using THP with randomized scaling, is upper bounded by

$$R^r_{\text{THP}} \leq \log(1+\Delta) - \log(\Delta) + 1$$

for any distribution $f(\alpha)$, when $\Delta \leq \frac{1}{3}$.

Proof: Using (16), for every distribution $f(\alpha)$, we have

$$\begin{aligned} I_\beta(V;Y') &= \min_\beta \left\{E_\alpha[\log\alpha] - E_\alpha[\log|\tilde{\alpha} - \beta|]\right\} \\ &\overset{(a)}{\leq} \min_\varepsilon \left\{\log(1+\Delta) - E_\alpha[\log(|\alpha - \beta| \bmod \Lambda)]\right\} \\ &\overset{(b)}{\leq} \log(1+\Delta) - \frac{1}{2\Delta}\int_{-\Delta}^{\Delta}\log|x|dx \\ &= \log(1+\Delta) - \log(\Delta) + 1, \end{aligned}$$

where (a) holds since $\text{Supp}\{f(\alpha)\} \subseteq \mathcal{I}_\Delta$ and (b) is true due to the monotonicity of the log function where equality is achieved for $\alpha \sim \text{Unif}(\mathcal{I}_\Delta)$. ■

C. Noisy case

The randomized approach taken may be extended to the noisy case:

$$Y' = \left[v + N^\beta_{\text{eff}}\right] \bmod \Lambda$$
$$N^\beta_{\text{eff}} = (1-\alpha_R)U + \left(\alpha_R - \frac{\beta}{\alpha}\right)\frac{S}{\beta} + \alpha_R N.$$

This result is easily proved simply by choosing $\alpha_T = \frac{1}{\alpha}$ in (4).

Consider the case of SIR $\to 0$ (and finite SNR). In this case, α_R has to be chosen to be $\frac{\beta}{\alpha}$, in order to eliminate the residual interference component in the effective noise. The effective noise in this case is hence:

$$N^\beta_{\text{eff}} = \frac{\alpha - \beta}{\alpha}U + \frac{\beta}{\alpha}N.$$

Unlike in the noiseless case, in which the effective noise had only a finite support ("self-noise") component $\frac{\alpha-\beta}{\alpha}U$, here the noise has an additional Gaussian component $\frac{\beta}{\alpha}N$.

We only examine the deterministic and uniform distributions from Section V and minor variations on them, taking $\alpha_T = \frac{\alpha_{\text{MMSE}}}{\alpha} \triangleq \frac{1}{\tilde{\alpha}}$, where α is selected according to the distributions of Section V and $\alpha^{\text{MMSE}} \triangleq \frac{\text{SNR}}{1+\text{SNR}}$. The performances of the different choices for $f(\cdot)$ are shown in Figure 4.

Note that in the high-SNR regime, the non-deterministic distributions prove to be more effective than the best deterministic scheme, whereas in the low-SNR regime the deterministic selection becomes superior. This threshold phenomenon can be explained by considering the two components of N^β_{eff}: in the high-SNR regime, the dominant noise component is the "self-noise" component $\frac{\tilde{\alpha}-\beta}{\tilde{\alpha}}U$, which is minimized by a "smart" selection of $f(\cdot)$; in the low-SNR regime, on the other hand, the

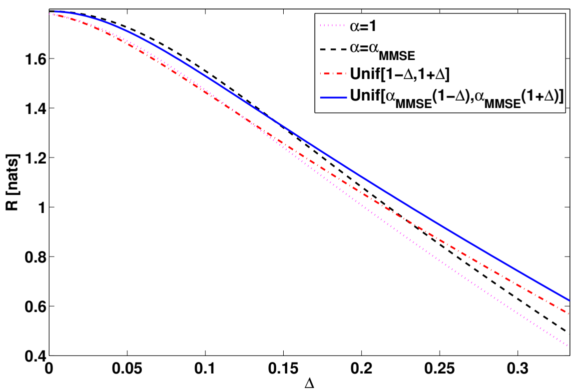

Fig. 4. Achievable rates in the random THP scheme for SNR $= 17$dB.

dominant noise component is the Gaussian part $\frac{\beta}{\alpha}N$, whose multiplicative factor $\frac{\beta}{\alpha}$ should be deterministic to minimize its average power. In general, there is a tradeoff between the best deterministic selection of α_T which minimizes the power of the Gaussian component and the self-noise component, which is to be minimized by a random α_T selection.

VI. Extensions and Implications

A. Multi-Dimensional Lattices

It is well known, that a multi-dimensional extension of THP (i.e., lattice-based precoding), allows to approach the full capacity of the DP channel (with non-causal knowledge of the interference) with *perfect channel knowledge*. Somewhat surprisingly, we observe that the channel knowledge is *imperfect* multi-dimensional lattice precoding yields identical results to those obtained by scalar (one-dimensional lattice) precoding, in the limit of high-SNR. This is seen by simply repeating the proof of Lemma 2 for a multi-dimensional lattice Λ. It can be explained by the fact that, in the "noiseless case" no shaping gain can be obtained using higher dimensional lattices, as the self-noise "gains shaping" just as the signal. Hence, using high-dimensional lattices does not increase the achievable rates for lattice-based precoding schemes in the absence of channel noise. In the noisy case, however, multi-dimensional strategies allow gaining some of the shaping gain, due to the presence of a Gaussian noise component as was discussed in Section V-C.

Note that multi-dimensional lattice correspond to non-causal SI case, .i.e, S is known non-causally. For this case an inner-bound on the capacity, due to Mitran, Devroye and Tarokh [16], is given by

$$C \geq \sup_{p(u|x,s,w),p(x|s,w),p(w)} \inf_{\beta \in \mathcal{B}} \left[I_\beta(U;Y|W) - I(U;S|W) \right], \tag{10}$$

where U is auxiliary, W is a time-sharing random and X, S and Y are the channel input, side-information at the transmitter and channel output, respectively.[5] The authors of [16], suggest using an auxiliary random variable, similar to that of Costa [1], viz.,

$$U = X + \alpha S. \tag{11}$$

By selecting the parameter α of (11), in the same manner as α_T of the THP schemes of Section IV and Section V, we arrive to the same performances when using multi-dimensional lattices of dimensions going to infinity, in all scenarios (finite/infinite SIR, finite/infinite SNR).

B. Implications to MIMO Broadcast Channels

Consider a Gaussian MIMO channel. For simplicity we consider a real-valued channel model. We further consider the case of two transmitting antennas and two receivers, with one receiving antenna each:

$$Y_i = \boldsymbol{h}_i^T \boldsymbol{X} + N_i, \tag{12}$$

where Y_i is the channel output received by user $i = 1, 2$, $\boldsymbol{X}$ is the 2×1 channel input, and N_i is an AWGN. Without loss of generality, we take the power of N_i to be 1. We consider the case where only private messages are sent to the two users. Hence, for linear zero-forcing or linear MMSE as well as for dirty paper coding, the transmitted signal can be decomposed into the form

$$\boldsymbol{X} = W_1 \boldsymbol{t}_1 + W_2 \boldsymbol{t}_2,$$

where W_i is the signal intended for user i of average power P_i, and $\boldsymbol{t}_i$ is a unit vector in the direction of the transmitted direction of this message. Without loss of generality, we shall assume that $P_2 \geq P_1$. In a similar way we shall rewrite the channel vectors $\boldsymbol{h}_i$ in the form

$$\boldsymbol{h}_i = h_i \boldsymbol{e}_i,$$

where h_i is the signed-amplitude and $\boldsymbol{e}_i$ is a unit vector in the direction of $\boldsymbol{h}_i$. Let us denote the acute angle between $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ by θ (see Figure 5):

$$\theta \triangleq \min \Big\{ \arccos\big(\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle\big), \arccos\big(-\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle \big) \Big\}.$$

In practice, the channel vectors $\boldsymbol{h}_i$ are known up to some finite accuracy, due to estimation errors or limited feedback, at the transmitter end. We assume that the transmitter knows the channel vectors $\boldsymbol{h}_i$ up to some *small* angular errors $\varepsilon_i \in [-\Delta, \Delta]$ ($\Delta \ll 1$),[6] that is:

$$\begin{aligned} h_i &= \tilde{h}_i \\ \langle \boldsymbol{e}_i, \tilde{\boldsymbol{e}}_i \rangle &= \cos(\varepsilon_i), \end{aligned} \tag{13}$$

where $\boldsymbol{h}_i$ ($i = 1, 2$) are the estimations of the channel vectors available at the transmitter and $\tilde{\boldsymbol{h}}_i = \tilde{h}_i \tilde{\boldsymbol{e}}_i$ are the true channel realizations.

Linear zero forcing

According to this strategy, the transmitter avoids interferences by transmitting $\boldsymbol{x}_1$ in an orthogonal direction to $\boldsymbol{h}_2$, and $\boldsymbol{x}_2$ - orthogonally to $\boldsymbol{h}_1$, as depicted in Figure 5 (see, e.g., [24]). However, in the case of imperfect channel knowledge at the transmitter, described by (13), the presence of an additional residual noise component is inevitable. The simplest approach

[5]The upper bound, given in [16], tends to infinity, and thus, not interesting in our problem.

[6] One may assume a presence of a *small* magnitude error as well. However, such an error would have no effect when performing first-order approximations.

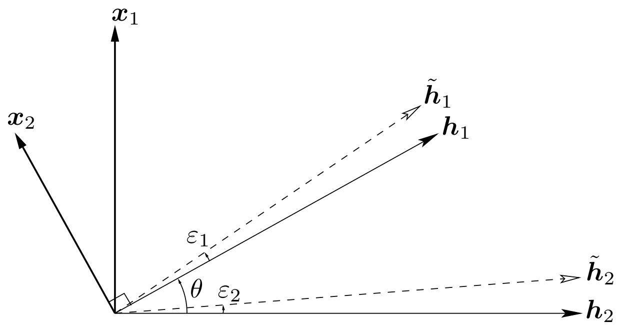

Fig. 5. Pictorial representation of the zero-forcing technique in the MIMO broadcast channel

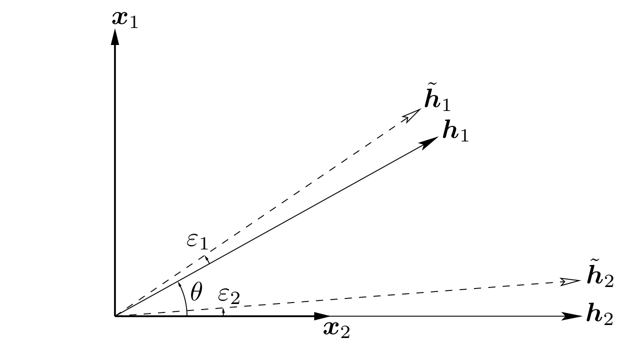

Fig. 6. Pictorial representation of the DPC technique in the MIMO broadcast channel

and only approach we consider here, is ignoring the estimation inaccuracy, that is, transmitting as if ε_i were 0. Hence by using codebooks that achieve capacity for the (interference-free) AWGN channel, any rate pair (R_1, R_2) satisfying:

$$R_1 \geq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_1 \left\|\boldsymbol{h}_1\right\|^2 \sin^2(\theta)}{\mathrm{SNR}_2 \left\|\boldsymbol{h}_1\right\|^2 \Delta^2 + 1}\right) - o(1)$$
$$R_2 \geq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_2 \left\|\boldsymbol{h}_2\right\|^2 \sin^2(\theta)}{\mathrm{SNR}_1 \left\|\boldsymbol{h}_2\right\|^2 \Delta^2 + 1}\right) - o(1) \qquad (14)$$

is achievable, under first-order approximations ($\Delta \ll 1$), where $o(1) \to 0$ as $\Delta \to 0$ (see details in [25]).

Dirty paper coding

Instead of using linear precoding approaches, one may transmit the message to user 1 in an orthogonal direction to the channel vector of user 2, and apply dirty paper coding to eliminate the interference of the user 2 on its own channel vector, as depicted in Figure 6. The expressions we provide below are for the non-causal case, i.e., correspond to using multi-dimensional THP where the dimension goes to infinity.[7] For a more thorough discussion of this approach see [3], [24].

Using the "smart receiver" approach and taking $\alpha_T = 1$ and assume Δ small enough such that $\Delta \ll \mathrm{ctg}(\theta)$, any rate pair (R_1, R_2) satisfying:

$$R_1 \geq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_1 \left\|\boldsymbol{h}_1\right\|^2 \sin^2(\theta)}{\mathrm{SNR}_1 \left\|\boldsymbol{h}_1\right\|^2 \sec^2(\theta)\Delta^2 + 1}\right) - o(1)$$
$$R_2 \geq \frac{1}{2}\log\left(1 + \frac{\mathrm{SNR}_2 \left\|\boldsymbol{h}_2\right\|^2}{\mathrm{SNR}_1 \left\|\boldsymbol{h}_2\right\|^2 \Delta^2 + 1}\right) - o(1). \qquad (15)$$

is achievable, under first-order approximations, where $o(1) \to 0$ as $\Delta \to 0$ (see details in [25]).

It has been speculated in some works, e.g., [26], [27], [28], that DPC has a significant drawback in the presence of channel estimation errors, compared to linear approaches such as linear-ZF. However, by comparing the achievable rates using DPC (15), to those achievable by linear-ZF (14), one sees that, at least for small Δ values, DPC is "*more robust*" than linear-ZF, when

$$\mathrm{SNR}_2 \sec^2(\theta) > \mathrm{SNR}_1.$$

We note that the performance of both schemes, i.e., the dirty paper coding scheme and the linear-ZF scheme, can further be improved by employing MMSE optimization at the receiver as was done in Section IV-A in conjunction with a randomized "guessing" of ε_1 and ε_2 (chosen in an i.i.d. manner according to a judicious distribution function), as was explained in Section V.

VII. Summary

In this work, the compound dirty paper channel was considered. We studied the performance that may be achieved by an extended Tomlinson-Harashima precoding scheme and derived lower bounds on the capacity of the channel. We derived the MMSE scaling that can be applied at the receiver to compensate for imprecise channel knowledge at the transmitter. We further showed that randomized α scaling at the transmitter may further improve the achievable rate. It was also shown that the potential shaping gain of higher lattice dimensions diminishes with the increase of the channel estimation inaccuracy.

This work focused exclusively on the performance achievable using THP-like schemes. It would be interesting to obtain an upper bound on the capacity (without any restriction on the coding technique) of noiseless DP channel under channel uncertainty.

Appendix A
Proof of Theorem 1

Direct: Denote by $\mathcal{T}$ the family of all mappings from $\mathcal{S}$ to $\mathcal{X}$. Use a transmitter that sends $x = t(s)$, where t is chosen in an i.i.d. manner, according to some predefined probability distribution $p(t)$. In this case, the problem reduces to that of a compound channel with *no side-information*, with an input alphabet $\mathcal{T}$ (see [17]), the same output alphabet $\mathcal{Y}$ and the corresponding transition probabilities

$$p(y|t) = \sum_{s \in \mathcal{S}} p(y|x = t(s), s)\,.$$

[7]The results for the causal case are identical up to a subtraction of the shaping loss $\frac{1}{2}\log\left(\frac{2\pi e}{12}\right)$ in (14).

Hence, by maximizing over all possible input probabilities, $p(t)$, of the equivalent channel, we have an inner bound on the (worst-case) capacity (see, e.g., [19]):

$$C \geq \sup_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} I_\beta(T;Y)\,.$$

Converse: For each n, let the information message W be drawn according to a uniform distribution over $\{1,...,2^{nR}\}$. Denote the error probability corresponding to $\beta\in\mathcal{B}$ by $P_{e,\beta}^{(n)}$ and the error probability of the scheme as the supremum of these probabilities, $P_e^{(n)} \triangleq \sum_{\beta\in\mathcal{B}} P_{e,\beta}^{(n)}$. Then we have:

$$\begin{aligned}nR = H(W) &\leq 1 + P_{e,\beta}^{(n)} nR + I_\beta(W;Y_1^n)\\ &\leq 1 + P_e^{(n)} nR + \sum_{i=1}^{n} I_\beta(W;Y_i|Y_1^{i-1})\,,\end{aligned}$$

where the first inequality is due to Fano's inequality (see, e.g., [29]) and the second inequality follows from the chain-rule for mutual information. By retracing the steps of Shannon in [17], for every $\beta\in\mathcal{B}$, we have $I_\beta(W;Y_i|Y_1^{i-1}) \leq I_\beta(W,S_1^{i-1};Y_i)$. Since $\left\{W,S_1^{i-1}\right\}$ does not depend on the value of β, the following inequality holds true as explained in detail in [17]:

$$nR \leq P_e^{(n)} nR + \sum_{i=1}^{n} I_\beta(T_i;Y_i), \qquad \forall\beta\in\mathcal{B}$$

The inequality above needs to be held for all $\beta\in\mathcal{B}$ simultaneously, and hence can be rewritten as

$$nR \leq \sup_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} P_e^{(n)} nR + nI_\beta(T;Y)\,.$$

Finally, dividing by n, taking $P_e \to 0$ and letting $n\to\infty$, we obtain

$$R \leq \sup_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} I_\beta(T;Y_i)\,.$$

■

Appendix B
Proof of Lemma 2 and treatment for $\Delta > 1/3$

Lemma 2: The term

$$I_\beta(V;Y'|\alpha) = h_\beta(Y'|\alpha) - h_\beta(Y'|V,\alpha)$$

is maximized by taking $V\sim\mathrm{Unif}(\Lambda)$. Moreover, it is easily seen that the support of $f(\alpha)$ should be restricted to $\mathcal{I}_\Delta$. It follows that,

$$\begin{aligned}&I_\beta(V;Y'|\alpha) = h_\beta(Y'|\alpha) - h_\beta(Y'|V,\alpha)\\ &\quad= \log(L) - h_\beta(Y'|V,\alpha)\\ &\quad= \log(L) - h([N_{\text{eff}}^\beta] \bmod \Lambda)\\ &\quad= \log(L) - E_\alpha\left[h\left(\left[\frac{\alpha-\beta}{\alpha}U\right] \bmod \Lambda\right)\right].\end{aligned}$$

The term $\frac{\alpha-\beta}{\alpha}$ is maximized when $\alpha = 1-\Delta$ and $\beta = 1+\Delta$, and is equal to $\frac{2\Delta}{1-\Delta}$. Hence, for $\Delta\leq\frac{1}{3}$, we have $\frac{\alpha-\beta}{\alpha}\leq 1$. Therefore,

$$\begin{aligned}I_\beta(V;Y'|\alpha) &= \log(L) - E_\alpha\left[h\left(\frac{\alpha-\beta}{\alpha}U\right)\right]\\ &= \log(L) + E_\alpha\left[-\log(\Delta) - \log\left|\frac{\alpha-\beta}{\alpha}\right|\right]\\ &= -E_\alpha \log\left|\frac{\alpha-\beta}{\alpha}\right| = E_\alpha\left[\log(\alpha) - \log|\alpha-\beta|\right].\end{aligned} \tag{16}$$

■

The case of $\Delta > 1/3$ can be treated in a similar manner by employing the following lemma.

Lemma 4: Suppose $U\sim\mathrm{Unif}(\mathcal{V}_0)$. Then for every $a>1$, the entropy of $\big([aU] \bmod \Lambda\big)$ is bounded by

$$\log(L) - \log\left(\frac{\lceil a\rceil}{a}\right) \leq h([aU] \bmod \Lambda) \leq \log(L)\,.$$

Proof: The upper-bound follows easily from the fact that differential entropy is maximized by a uniform distribution, when subject to an amplitude constraint, see, e.g. [29]. To prove the lower-bound, note that there is a unique index $k\in\mathbb{Z}$ which satisfies

$$aU = [aU] \bmod \Lambda + kL,$$

and the cardinality of k equal to $|\mathcal{K}| = \lceil a\rceil$. One may easily verify [30], that the following relation holds

$$h(aU) = h([aU] \bmod \Lambda) + H(k|[aU])$$

which leads to the desired bound:

$$\begin{aligned}h([aU] \bmod \Lambda) &= \log(aL) - H(k|[aU])\\ &\geq \log(aL) - H(k)\\ &= \log(aL) - \log(a+1)\\ &= \log(L) - \log\left(\frac{\lceil a\rceil}{a}\right).\end{aligned}$$

■

Acknowledgment

The authors wish to thank Emin Martinian for suggesting to incorporate random scaling into THP schemes at the transmitter.

References

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

[2] E. A. Lee and D. G. Messerschmitt. *Digital Communication*. Kluwer, 3rd ed. Newell, MA, 2004.

[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] N. Jindal S. Vishwanath and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels. *IEEE Trans. Information Theory*, 49:2658–2668, Oct. 2003.

[5] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Information Theory*, 52:3936–3964, Sept. 2006.

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

[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] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice strategies for cancelling known interference. *IEEE Trans. Information Theory*, pages 3820–3833, Nov. 2005.

[9] P. Grover and A. Sahai. What is needed to exploit knowledge of primary transmissions? *available at arXiv:cs/0702071*.

[10] A. Bennatan, V. Aggarwal, Y. Wu, A. R. Calderbank, J. Hoydis, and A. Chindapol. Bounds and lattice-based transmission strategies for the phase-faded dirty-paper channel. *IEEE Trans. Wireless Communications*.

[11] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Lett.*, vol. 7:138–139, Mar. 1971.

[12] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Trans. Communications*, COM-20:774–780, Aug. 1972.

[13] 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.

[14] J. J. Eggers, R. Bäumel, R. Tzschoppe, and B. Girod. Scalar Costa scheme for information embedding. *IEEE Trans. Signal Processing*, pages 1003–1019, Apr., 2003.

[15] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Trans. Information Theory*, Vol. 52, No. 11:5045–5060, Nov. 2003.

[16] P. Mitran, N. Devroye, and V. Tarokh. On compound channels with side information at the transmitter. *IEEE Trans. Information Theory*, 52:1745–1755, April 2006.

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

[18] J. Wolfowitz. *Coding Theorems of Information Theory*. Springer-Verlag, Berlin Heidelberg, New York, third edition, 1978.

[19] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30:1229–1241, Dec. 1959.

[20] R. L. Dobrushin. Optimal information transmission over a channel with unknown parameters (in Russian). *Radiotekh. i Elektron.*, Vol. 4, No. 12:1961–1956, Dec. 1959.

[21] J. Wolfowitz. Simultaneous channels. *Arch. Rational Mech. Anal.*, 4:371–386, 1960.

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

[23] 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.

[24] J. Lee and N. Jindal. High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding. *IEEE Trans. Information Theory*, Vol. 53, No. 12:4787–4792, Dec. 2007.

[25] A. Khina. On the robustness of dirty paper coding. Master's thesis, Tel Aviv University, 2009.

[26] S. Yang and J. Belfiore. The impact of channel estimation error on the DPC region of the two-user Gaussian broadcast channel. In *43th Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, Oct. 2002.

[27] R. Chen, R.W. Heath Jr., and J.G. Andrews. Transmit selection diversity for unitary precoded multiuser spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Processing*, vol. 55, no. 3.

[28] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero-forcing based achievable rates. In *Proc. Int. Symp. Inform. Theory (ISIT), Nice, France*, pages 2046–2050, June 2007.

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

[30] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden Day, San Francis. CA., 1964.