On Robust Dirty Paper Coding

Anatoly Khina
Dept. EE - Systems, Tel Aviv University
Tel Aviv, Israel
Email: anatolyk@eng.tau.ac.il

Uri Erez
Dept. EE - Systems, Tel Aviv University
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

***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. We derive lower bounds on the achievable rate of communication by proposing a coding scheme that partially compensates for the imprecise channel knowledge. We focus on a communication scenario where the Gaussian noise is small while the interference is strong. Our approach is based on analyzing the performance achievable using extended Tomlinson-Harashima like coding schemes.**

When the power of the interference is finite, we show that this may be achieved 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.

I. Introduction

The dirty-paper (DP) channel, first introduced by Costa [3], provides an information theoretical 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, the MIMO Gaussian broadcast channel and information embedding. The DP channel is given by

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

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 ("dirty-tape") or noncausally ("dirty-paper") to the transmitter but not to the receiver. In this work we consider both the case where S is i.i.d. (of some distribution) with power P_S (Section IV-A) as well as arbitrary (Section IV-B).

Costa [3] showed that, for an i.i.d. Gaussian interference with an arbitrary power, the capacity in the noncausal scenario is equal to that of the interference-free AWGN channel, i.e., $\frac{1}{2}\log(1 + \text{SNR})$, where $\text{SNR} \triangleq P_X/P_N$. This result was extended in [5], [2] to the case of arbitrary interference.

In this work we focus our attention to the causal (or scalar) scenario, both since it results in simpler coding schemes but also since the benefit from using a vector approach (at least using the methods we propose) diminishes in the presence of imprecise channel knowledge.

The capacity for the causal case is not known but upper and lower bounds were found in [5], which coincide in the high-SNR regime, thus establishing the capacity for this case to be $\frac{1}{2}\log(1 + \text{SNR}) - \frac{1}{2}\log(\frac{2\pi e}{12})$. That is, a rate loss of $\frac{1}{2}\log(\frac{2\pi e}{12})$, relative to the capacity of the AWGN channel. This result implies that in the limit of strong interference and high SNR, the well-known Tomlinson-Harashima precoding (THP) technique [11], [7] is optimal. For general SNRs, the lattice-based coding techniques of [1], [4], [5] are an extension of Tomlinson-Harashima precoding where a scaling parameter is introduced in the transmitter and receiver. In this work the term Tomlinson-Harashima precoding is used in this wider sense.

The work of one of the authors (AK) was supported by a fellowship of the Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel Aviv University.

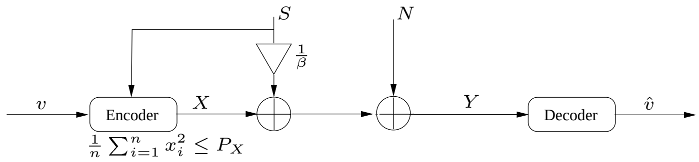

Fig. 1. The compound dirty-paper channel.

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., [8]). This leads to the question of how sensitive is DP coding to imprecise channel knowledge.

II. Channel Model and Motivation

We consider the channel model:

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

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

Consider the limit of high SNR. At first glance, one might assume that a reasonable approach could be to use standard THP as the SNR is high. 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.

We observe in Section IV-A that by using a carefully chosen scaling parameter at the receiver in Section IV-A, 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 residual interference and a "self noise" component.

We then show in Section IV-B 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.

III. Compound Channels with Causal SI at the Transmitter

The causal DP channel model (1) is a special case of the more general problem of a channel with side information at the transmitter. This problem was first introduced by Shannon [10], who considered a DMC whose transition matrix depends on the channel state $s \in \mathcal{S}$, 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

$$p(\boldsymbol{y}, |\boldsymbol{s}, \boldsymbol{x}) = \prod_i p(y_i|s_i, x_i) \tag{3}$$

$$p(\boldsymbol{s}) = \prod_i p(s_i), \tag{4}$$

where $x \in \mathcal{X}$ is the channel input and $y \in \mathcal{Y}$ is the channel output. 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 $\mathcal{X}$, and the corresponding transition probabilities of this channel are

$$p(y|t) = \sum_s p(s)p(y|x = t(s), s). \tag{5}$$

Note that this result uses mappings from the *current* state only although the transmitter has access to all the past states.

A compound (discrete memoryless) channel is a channel whose transition matrix depends on a parameter $\beta \in \mathcal{B}$ which is constant and not known to the transmitter but is known to the receiver. [1] The capacity of this channel is (see, e.g., [12]),

$$C = \sup_{p(x)\in\mathcal{P}(\mathcal{X})} \inf_{\beta\in\mathcal{B}} I_\beta(X;Y), \tag{6}$$

where $I_\beta(X;Y)$ denotes the mutual information of X and Y with respect to the transition matrix $p_\beta(y|x)$. This result may be easily extended to the case of a compound channel with SI available *causally* to the transmitter.

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

$$C = \sup_{p(t)\in\mathcal{P}(\mathcal{T})} \inf_{\beta\in\mathcal{B}} I_\beta(T;Y), \tag{7}$$

where $\mathcal{T}$ denotes the set of all random strategy functions of the form $t : \mathcal{S} \longrightarrow \mathcal{X}$.

Proof: The achievability follows by the same methods used in the achievability proof of the standard compound channel over the extended alphabet $\mathcal{T}$, whereas the converse is proved by following the steps of Shannon in [10]. ■

Remark 1: The case of noncausal SI is more difficult. The converse of Gelfand-Pinsker [6] is not easily extended to the compound scenario. In [9] Mitran et al. 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 noncausal SI case, using k-dimensional vector strategies and taking k to infinity, follows:

$$C^{\mathrm{non-casual}} = \limsup_{k\to\infty} \max_{p(\boldsymbol{t})} \min_\beta \frac{1}{k} I_\beta(\boldsymbol{T};\boldsymbol{Y}). \tag{8}$$

IV. Compound noiseless Dirty Paper Channel

The compound DP channel was defined in (2). For simplicity, we consider the noiseless case $P_N = 0$, i.e.,

$$Y = X + \frac{1}{\beta}S. \tag{9}$$

The results of Section III may readily be extended to the continuous case and incorporating an input constraint (see e.g., [9], 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.

A. THP With Imprecise Channel Knowledge

Denote the interval $\left[-\frac{\Delta}{2}\frac{\Delta}{2}\right)$ by $\mathcal{A}_\Delta$ where Δ is chosen such that $P_X = \frac{\Delta^2}{12}$. Suppose that S has finite power P_S and denote by $\mathrm{SIR} = \beta^2\frac{P_X}{P_S}$ the signal-to-interference ratio. Let $U \sim \mathrm{Unif}(\mathcal{A}_\Delta)$ be a random variable (dither) [2] known to both transmitter and receiver. We consider an extended THP scheme given by:

- Transmitter: For any $v \in \mathcal{A}_\Delta$, the transmitted signal is
$$X = [v - S - U] \bmod \mathcal{A}_\Delta. \tag{10}$$
- Receiver: The receiver computes,
$$Y' = [\gamma Y + U] \bmod \mathcal{A}_\Delta. \tag{11}$$

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

$$\begin{aligned} Y' &= [\gamma Y + U] \bmod \mathcal{A}_\Delta & (12)\\ &= [v - (v - S - U) + \gamma X - (\beta-\gamma)\frac{S}{\beta}] \bmod \mathcal{A}_\Delta & (13)\\ &= [v - (1-\gamma)X - (\beta-\gamma)\frac{S}{\beta}] \bmod \mathcal{A}_\Delta. & (14)\end{aligned}$$

[1] 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 [12, chap. 4]), as the receiver may estimate to within any desired accuracy (with probability going to one), using a negligible portion of the block length.

[2] It can be shown that common randomness is not needed in the case of an i.i.d. interference sequence.

Due to the dither U, X is independent of S and of the information signal V, and is uniform over $\mathcal{A}_\Delta$. Therefore, this channel is equivalent to the modulo-additive channel:

$$Y' = [v + N_{\text{eff}}^{\beta}] \bmod \mathcal{A}_\Delta \tag{15}$$

$$N_{\text{eff}}^{\beta} \triangleq (1-\gamma)U - (\beta - \gamma)\frac{S}{\beta}, \tag{16}$$

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

$$P_{N_{\text{eff}}^{\beta}} = (1-\gamma)^2 P_X + (\beta-\gamma)^2 \frac{P_S}{\beta^2}. \tag{17}$$

We denote the maximal achievable rate under this setting by R_{THP}^{d}, where "d" stands for "deterministic" (in contrast to the random strategies treated in Section IV-B), and the achievable rate for a specific pair (γ, β) by $R_{\text{THP}}^{d}(\gamma,\beta)$.

For any pair (γ,β), the mutual information is maximized by taking $V \sim \text{Unif}(\mathcal{A}_\Delta)$. Hence:

$$\begin{aligned} R_{\text{THP}}^{d}(\gamma,\beta) &= h(Y') - h(Y'_\varepsilon|V) &(18)\\ &= \log(\Delta) - h([N_{\text{eff}}^{\beta}] \bmod \mathcal{A}_\Delta). &(19)\end{aligned}$$

A lower bound on R_{THP}^{d} is obtained by minimizing the effective noise power, $P_{N_{\text{eff}}^{\beta}}$, with respect to γ. This results in taking

$$\gamma_{\text{MMSE}} = \frac{1+\frac{\beta}{\text{SIR}}}{1+\frac{1}{\text{SIR}}} = \frac{\text{SIR}+\beta}{\text{SIR}+1} \tag{20}$$

for which the corresponding power is

$$P_{N_{\text{eff}}^{\beta}}^{\text{MMSE}} = \frac{(1-\beta)^2}{1+\text{SIR}} P_X. \tag{21}$$

The maximal achievable rate R_{THP}^{d} is lower-bounded by

$$\begin{aligned} R_{\text{THP}}^{d} &= \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} R_{\text{THP}}^{d}(\gamma,\beta) &(22)\\ &= \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} \log(\Delta) - h([N_{\text{eff}}^{\beta}] \bmod \mathcal{A}_\Delta) &(23)\\ &\geq \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} \frac{1}{2}\log(\Delta) - h(N_{\text{eff}}^{\beta}) &(24)\\ &= \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} \frac{1}{2}\log(\Delta) - h\left(N_{\text{eff,G}}^{\beta}\right) + \varepsilon(\beta,\gamma) &(25)\\ &\geq \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} \frac{1}{2}\log(12P_X) - \frac{1}{2}\log\left(2\pi e P_{N_{\text{eff}}^{\beta}}\right) + \varepsilon(\beta,\gamma) &(26)\\ &\geq \min_{\beta\in\mathcal{I}_\delta}\max_{\gamma\in\mathcal{I}_\delta} \frac{1}{2}\log\left(\frac{P_X}{(1-\gamma)^2P_X + (\gamma-\beta)^2 P_S}\right) &(27)\\ &\quad + \varepsilon(\beta,\gamma) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right), &(28)\end{aligned}$$

where $\varepsilon(\beta,\gamma) \triangleq h\left(N_{\text{eff,G}}^{\beta}\right) - h\left(N_{\text{eff}}^{\beta}\right)$ and $N_{\text{eff,G}}^{\beta}$ is Gaussian with the same variance as N_{eff}^{β}. Note that $\varepsilon(\beta,\gamma) > 0$. Substituting $\gamma = \gamma_{\text{MMSE}}$ and using the fact that $\varepsilon(\beta,\gamma) > 0$, yields the lower bound

$$R_{\text{THP}}^{d} \geq \frac{1}{2}\log\left(\frac{1+(1-\delta)^2\text{SIR}}{\delta^2}\right) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right). \tag{29}$$

Remark 2:

1. The above lower bound can be further tightened, for any specific distribution of S, by calculating $\varepsilon(\beta,\gamma)$. For instance, if S is uniform then R_{THP}^{d} may be lower-bounded by
$$R_{\text{THP}}^{d} \geq \frac{1}{2}\log\left(\frac{1+(1-\delta)^2\text{SIR}}{\delta^2}\right) - \frac{1}{2}\log\left(\frac{e}{2}\right).$$
2. In the weak interference region, SIR $\to \infty$, we have $\gamma \to 1$ and the achievable rate goes to infinity. This is of course an uninteresting case as THP is unattractive in this regime.
3. In the strong interference region, SIR $\to 0$, the residual interference component of N_{eff}^{β} has to be completely cancelled. This is done by selecting $\gamma = \beta$. This results in an effective noise with finite power (dictated by the magnitude of δ) and thus reliable communication is possible at strictly positive rates. In this case, $\varepsilon(\beta,\gamma)$ goes to zero.
4. When the signal and interference have the same power, SIR $= 1$, γ_{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 $\gamma = \gamma_{\text{MMSE}}$. Thus, γ_{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 canceling out the interference component completely ($\gamma = \beta$), or alternately, half of the noise power obtained by canceling out completely the self-noise component ($\gamma = 1$).
5. Due to the modulo operation at the receiver side and since the effective noise is not Gaussian, the choice $\gamma = \gamma_{\text{MMSE}}$ does not strictly maximize the mutual information $I(V;Y')$, but rather is a reasonable approximate solution.

B. Randomized Scaling at Transmitter

For simplicity, we now restrict our attention to the case of strong interference, i.e., SIR $\to 0$. In this case, the receiver must totally cancel out the interference by choosing $\gamma = \beta$.

We now investigate whether performance may be improved by introducing a random scaling factor α at the transmitter, which is chosen at random at each time instance and is assumed known to both transmitter and receiver. Thus, we consider the following transmission scheme:

- Transmitter: For any $v \in \mathcal{A}_\Delta$, sends
$$X = [v - \frac{1}{\alpha}S - U] \bmod \mathcal{A}_\Delta. \tag{30}$$
- Reciever: The receiver applies the front end operation,
$$Y' = [\gamma Y + U] \bmod \mathcal{A}_\Delta, \tag{31}$$
where $\gamma = \beta/\alpha$.

The above channel can be shown (by retracing the steps of (15), (16)) to be equivalent to the modulo-additive channel

$$Y' = \left[v + N_{\text{eff}}^{\beta}\right] \bmod \mathcal{A}_\Delta, \tag{32}$$

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[3] from symbol to symbol (or block to block) within the interval of uncertainty $\mathcal{I}_\delta$. We denote the maximal achievable rate of this scheme by R^r_{THP}, where "r" stands for "random". It is given by:

$$R^r_{\text{THP}} = \max_{f(\alpha)} R^r_{\text{THP}}(f) = \max_{f(\alpha)} \min_{\beta \in \mathcal{I}_\delta} I_\beta(V; Y'|\alpha), \tag{33}$$

where $f(\alpha)$ is the PDF according to which α is selected and $R^r_{\text{THP}}(f)$ denotes the mutual information corresponding to the specific choice of $f(\alpha)$.

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

$$R^r_{\text{THP}} = \max_{f(\alpha):\,\text{Supp}\{f(\alpha)\} \subseteq \mathcal{I}_\delta} \min_{\beta \in \mathcal{I}_\delta} -E_\alpha \left[\log \left| \frac{\alpha - \beta}{\alpha} \right| \right]. \tag{34}$$

Proof: The term

$$I_\beta(V; Y'|\alpha) = h_\beta(Y'|\alpha) - h_\beta(Y'|V, \alpha) \tag{35}$$

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

$$\begin{aligned} I_\beta(V; Y'|\alpha) &= h_\beta(Y'|\alpha) - h_\beta(Y'|V, \alpha) && (36)\\ &= \log(\Delta) - h_\beta(Y'|V, \alpha) && (37)\\ &= \log(\Delta) - h([N^\beta_{\text{eff}}] \bmod \mathcal{A}_\Delta) && (38)\\ &= \log(\Delta) - E_\alpha \left[h\left(\left[\frac{\alpha - \beta}{\alpha} U \right] \bmod \mathcal{A}_\Delta \right) \right]. && (39) \end{aligned}$$

For $\delta \geq \frac{1}{3}$, we have

$$\frac{\alpha - \beta}{\alpha} < 1, \tag{40}$$

since the term $\frac{\alpha - \beta}{\alpha}$ is maximized when $\alpha = 1 - \delta$ and $\beta = 1 + \delta$, and its value in this case is $\frac{2\delta}{1-\delta}$. Therefore,

$$\begin{aligned} I_\beta(V; Y'|\alpha) &= \log(\Delta) - E_\alpha \left[h\left(\frac{\alpha - \beta}{\alpha} U \right) \right] && (41)\\ &= \log(\Delta) + E_\alpha \left[-\log(\delta) - \log \left| \frac{\alpha - \beta}{\alpha} \right| \right] && (42)\\ &= -E_\alpha \log \left| \frac{\alpha - \beta}{\alpha} \right| = E_\alpha \left[\log(\alpha) - \log|\alpha - \beta| \right]. && (43) \end{aligned}$$

■

Finding the optimal distribution f in (33) is cumbersome. Instead, we suggest several choices for the distribution f which achieve better performance than that of any deterministic selection of α as well as give an upper bound on R^r_{THP}.

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

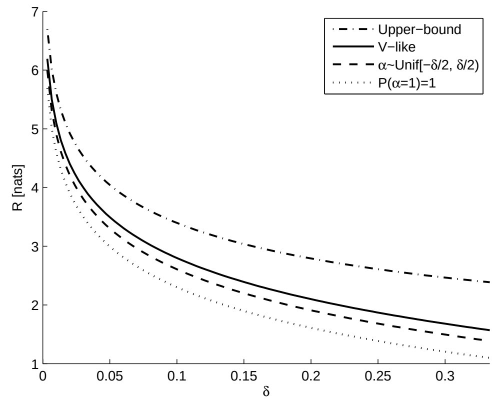

Fig. 2. Achievable rates and upper-bound on the deterministic THP scheme.

C. Quantifying the Achievable Rates

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

$$R^r_{\text{THP}}(f_{\text{Deter}}) = -\log \delta = \log \frac{1}{\delta}. \tag{44}$$

Note that this result coincides with the result for R^d_{THP} of Section IV-A (where $\varepsilon(\beta, \gamma)$ is equal to zero in this case as mentioned in Remark 2).

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

$$\begin{aligned} R^r_{\text{THP}}(f_{\text{Unif}}) = &\frac{1}{2\delta} \big[(1+\delta)\log(1+\delta) && (45)\\ &- (1-\delta)\log(1-\delta) - 2\delta \log(2\delta) \big]. && (46) \end{aligned}$$

Hence, even this simple randomization improves on the deterministic selection, as seen in Fig. 2.

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

$$f_{\text{V-like}}(\alpha) = \begin{cases} -\frac{\alpha - 1}{\delta^2} & 1 - \delta \leq \alpha < 1 \\ \frac{\alpha - 1}{\delta^2} & 1 \leq \alpha \leq 1 - \delta \end{cases}. \tag{47}$$

The resulting rate is

$$R^r_{\text{THP}}(f_{\text{V-like}}) = -\frac{1}{2\delta^2} \big[(1-\delta^2)\log(1-\delta^2) + \delta^2 \log(\delta^2) \big]. \tag{48}$$

We have not pursued numerically optimizing $f(\cdot)$. We note that the optimal PDF will not be totally symmetric around 1 due to the first term $\log(\alpha)$ in (43). 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 $f(\cdot)$.

Remark 3: None of the three distributions above are optimal since $I_\beta(V; Y')$ varies with β.

D. Upper-Bound on the Scheme

Lemma 2: The rate achievable, using THP with randomized scaling, is upper bounded by

$$R^r_{\text{THP}} \leq \log(1+\delta) - \log(\delta) + 1 \tag{49}$$

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

Proof: Using (43), for every distribution $f(\alpha)$, we have

$$I_\beta(V;Y') = \min_\beta \{E_\alpha [\log \alpha] - E_\alpha [\log |\tilde{\alpha} - \beta|]\} \tag{50}$$

$$\stackrel{(a)}{\leq} \min_\varepsilon \{\log(1+\delta) - E_\alpha [\log(|\alpha - \beta| \bmod \Delta)]\} \tag{51}$$

$$\stackrel{(b)}{\leq} \log(1+\delta) - \frac{1}{2\delta}\int_{-\delta}^{\delta} \log |x| dx \tag{52}$$

$$= \log(1+\delta) - \log(\delta) + 1, \tag{53}$$

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

Remark 4: Somewhat surprisingly, multi-dimensional lattices give identical results to those obtained by one-dimensional lattices. This 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 like the signal. Hence, knowing the SI non-causally does not increase the achievable rates for lattice-based precoding schemes in the absence of noise. In the noisy case, however, multi-dimensional strategies allow gaining some of the shaping gain.

E. Noisy case

The approach taken may be extended to the noisy case. Using the "robust THP scheme" for the *noisy* compound DP channel (2), gives rise to the following equivalent modulo-additive channel:

$$Y' = \left[v + N^\beta_{\text{eff}}\right] \bmod \Lambda \tag{54}$$

$$N^\beta_{\text{eff}} = \frac{\alpha - \beta}{\alpha} U + \frac{\beta}{\alpha} N. \tag{55}$$

Unlike in the noiseless case, where the effective noise has a finite support, here the noise has a Gaussian component.

We only examine the deterministic and uniform distributions from Section IV-B and minor variations on them, taking $\tilde{\alpha} = \alpha_{\text{MMSE}} \cdot \alpha$, where α is selected according to the distributions of Section IV-B and $\alpha_{\text{MMSE}} \triangleq \frac{\text{SNR}}{1+\text{SNR}}$. The performances of the different choices for $f(\cdot)$ are shown in Fig. 3.

In the high-SNR region, the non-deterministic distributions prove to be more effective than the best deterministic scheme, whereas in the low-SNR region the deterministic selection becomes superior. This threshold phenomenon can be explained by the two components of N^β_{eff}: In the high-SNR region, the dominant noise component is the "self-noise" component $\frac{\alpha-\beta}{\alpha}U$, which is minimized by a "smart" selection of $f(\cdot)$; In the low-SNR region, on the other hand, the dominant noise component is the Gaussian part $\frac{\beta}{\alpha}N$, whose multiplicative factor $\frac{\beta}{\alpha}$ should be deterministic to minimize this component, due to concavity.

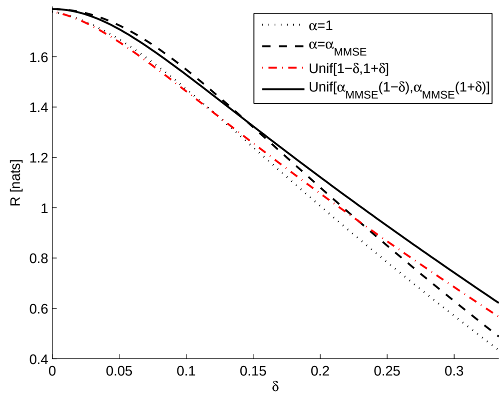

Fig. 3. Achievable rates in the random THP scheme for SNR $= 17dB$.

V. Summary

In this work, the compound dirty tape channel was investigated. We studied the performance obtained by an extended Tomlinson-Harashima precoding scheme and derived lower bounds to 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 maximal achievable rate.

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 the noiseless DP channel under channel uncertainty.

References

[1] 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.

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

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

[4] 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.

[5] 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.

[6] 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.

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

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

[9] 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.

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

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

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