DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Writing on Dirty Tape with LDPC Codes

Giuseppe Caire and Shlomo Shamai (Shitz)

ABSTRACT. We consider channels with interference known to the transmitter but unknown to the receiver. A precoding technique based on lattice quantization is capable of achieving Costa's "Dirty-Paper" limit for arbitrary interference signals, when interference is known non-causally. The same technique can be applied when interference is known with given anticipation k, fixed and independent of the code blocklength n. We review the main information-theoretic results in the non-causal $k = n$ and causal ($k = 1$) settings, and provide specific examples for a modulo-2 additive-noise binary channel with binary interference and for an AWGN channel with additive interference.

Then, we turn to coding design for the AWGN case with causally-known interference (nicknamed "Dirty-Tape" channel, in analogy with Costa's "Dirty-Paper"). We provide explicit code constructions based on Low-Density Parity-Check codes and on M-PAM modulation, able to approach the rates achievable by lattice precoding for the AWGN Dirty-Tape channel. Finally, we point out some research problems such as the wideband limit of the causally-known interference channel (and the related optimal signaling) and the extension of our coding technique to the case of non-causally known interference (AWGN Dirty-Paper channel).

1. Introduction

Memoryless channels with input X, output Y and state-dependent transition probability $P_{Y|X,S}$ where the channel state S is i.i.d., known to the transmitter and unknown to the receiver, date back to Shannon [**1**], who considered the case of state sequence known causally, and to Kusnetsov and Tsybakov [**2**], who considered the case of state sequence known non-causally. Gel'fand and Pinsker [**3**] proved the capacity formula

$$C = \sup_{P_{T|S}} \{I(T;Y) - I(T;S)\} \tag{1.1}$$

for the non-causal case, where T is an auxiliary random variable with conditional distribution $P_{T|S}$ and X is a deterministic function of S and T. This yields Shannon's capacity formula [**1**]

$$C = \sup_{P_T} I(T;Y) \tag{1.2}$$

for the causal case with i.i.d. state sequence, by restricting the supremization in (1.1) to T independent of S [**6**].

In the case where the channel is $Y = X + S + Z$, with $Z \sim \mathcal{N}(0, \sigma^2)$, $E[|X|^2] \leq \mathcal{E}$ and the interference S is also Gaussian and known non-causally at the transmitter, Costa [**4**]

1991 *Mathematics Subject Classification.* Primary 94B99, 94A15.

Key words and phrases. Low-Density Parity-Check codes, Channels with Side Information at the Transmitter.

The work of G. Caire has been partially aupported by the French *Réseau national de Recherche en Télécommunications*.

The Work of S. Shamai has been supported by the Israeli Academy of Science and Humanities (Israel Science Foundation).

proved that the capacity in (1.1) is equal to the standard AWGN capacity $\frac{1}{2}\log(1+\mathcal{E}/\sigma^2)$, as if interference was not present. From the title of Costa's paper, coding strategies for the non-causal known interference case are nicknamed "Dirty-Paper" coding and, by analogy, coding strategies for the causal case are nicknamed "Dirty-Tape" coding [**5, 6**].

While in early works such problems were motivated by data storage on defective media [**2**], more recently "Dirty-Paper" coding gained renewed attention because it arises as the main tool in several important settings such as broadcast vector Gaussian channels [**7, 8, 9, 10, 11, 12**], precoding for ISI channels [**13, 6**] and data hiding [**14, 15, 16**]. A widescope recent survey of applications of Dirty-Paper coding is provided in [**6**].

From the information theoretic point of view, Gel'fand, Pinsker and Costa's results were generalized in many ways (see [**6**] and references therein). However, efficient coding strategies able to approach the theoretical results are still not a common practice (preliminary results can be found in [**14, 15**]) although, based on random coding arguments, it can be shown that sequences of good lattice codes, mimicking Gel'fand and Pinsker random binning scheme, do exist and can approach Costa's result in the additive interference AWGN case [**17**].

The reminder of this paper is organized as follows. In Sections 2 and 3 we illustrate the capacity formulas (1.1) and (1.2) through two well-known examples: an input-constrained Binary-Symmetric Channel (BSC) with binary equiprobable interference, and Costa's AWGN channel with arbitrary interference. Then, Section 4 is devoted to the construction of coded modulation schemes based on Low-Density Parity-Check codes (LDPCs) for the AWGN Dirty-Tape problem (the example in Section 3 with causal interference knowledge). Conclusions and discussion are pointed out in Section 5.

2. BSC with known binary interference

Suppose that we wish to hide data into the least significant bits (LSBs) of the pixels of a gray-scale image (host signal) with a Hamming distortion constraint W, and assume that the host signal goes through some transformation that, acting on the LSBs, can be modeled as a BSC with transition probability p.

This problem can be modeled by the binary modulo-2 adder channel $Y = X + S + Z$ where S is the host signal LSB sequence, Z is the BSC noise and the input X is subject to the constraint

$$\frac{1}{n}\sum_{i=1}^{n} d_H(x_i, 0) \leq W \tag{2.1}$$

where $d_H(a,b)$ denotes Hamming distance.

In the following we shall assume that S is i.i.d. with $P(S=0) = P(S=1) = 1/2$. If the encoder has available the whole interference sequence realization $\mathbf{s}$ non-causally (i.e., before encoding), it can be shown (see [**18**] and references therein) that the maximum achievable rate is given by

$$C = \left\{ \begin{array}{ll} K(W) & \text{for } W_c \leq W \leq 1/2 \\ \alpha W & \text{for } 0 \leq W \leq W_c \end{array} \right. \tag{2.2}$$

where the function $K(w)$, defined for $w \in [0, 1/2]$, is given by

$$K(w) = \left\{ \begin{array}{ll} h(w) - h(p) & \text{for } p \leq w \leq 1/2 \\ 0 & \text{for } 0 \leq w \leq p \end{array} \right. \tag{2.3}$$

where $h(p)$ denotes the binary entropy function and where we let $W_c = 1 - \exp(-h(p))$ and $\alpha = \log((1-W_c)/W_c)$.[1]

The rate $K(W)$ (for $W > p$) is achieved by Dirty-Paper coding. Namely, we construct a binary random codebook $\mathcal{C}$ of size $|\mathcal{C}| = \exp(n(1 - h(p) - \epsilon/2))$ and make a random partition of $\mathcal{C}$ into subsets $\{\mathcal{C}_m : m = 1, \ldots, M\}$ of size $|\mathcal{C}_m| = \exp(n(1 - h(W) + \epsilon/2))$.

[1] When information rates are measured in bits, log and exp are base-2, when it is measured in nats, log and exp are base-e.

For large n, the number of subsets is given by $M = \exp(n(K(W) - \epsilon))$. Both $\mathcal{C}$ and the partition $\{\mathcal{C}_m\}$ are revealed to encoder and decoder.

Let $\mathbf{s}$ be the interference sequence and m be the information message to be sent. The encoder finds a codeword $\mathbf{c} \in \mathcal{C}_m$ such that

$$\frac{1}{n}\sum_{i=1}^{n} d_H(s_i, c_i) \leq W$$

Since each subset is "large enough", such sequence $\mathbf{c}$ can be found with probability $1 - \epsilon/2$ for sufficiently large n. Then, the encoder sends $\mathbf{x} = \mathbf{c} + \mathbf{s}$. The decoder observes the standard BSC output $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z} = \mathbf{c} + \mathbf{z}$, where $\mathbf{z}$ is the noise realization, finds the minimum distance codeword

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{c}, \mathbf{y})$$

and outputs the message $\widehat{m}$ as the index of the subset containing $\widehat{\mathbf{c}}$. Since the codebook $\mathcal{C}$ is "small enough", $\Pr(\widehat{m} \neq m) \leq \epsilon/2$ for sufficiently large n. Eventually, the rate $K(W) - \epsilon$ can be transmitter with error probability not larger than ϵ for sufficiently large n.

Notice the different roles of the $\mathcal{C}$ and its subsets $\mathcal{C}_m$: the codebook $\mathcal{C}$ must be a good channel code for the BSC with transition probability p while its bins must be good Hamming quantizers for the interference sequence $\mathbf{s}$ (a Bernoulli i.i.d. unbiased source) with Hamming distortion W.

In the range $0 \leq W \leq W_c$, capacity is achieved by time-sharing with duty-cycle $\theta = W/W_c$ the above scheme for Hamming distortion equal to W_c and "silence" (zero rate and zero Hamming distortion).

The above intuitive argument can be made formal and proves achievability of (2.2). A converse is provided, for example, in [**18**].

Let's consider now the Dirty-Tape case, where the transmitter at time i knows only $(s_1, \ldots, s_i)$. In this case, Shannon's formula (1.2) is given explicitly by

$$C = 2W(\log 2 - h(p)) \tag{2.4}$$

The proof of the above capacity formula serves as an exemple of the rather abstract idea of "coding over strategies" underlying the general capacity formula (1.2). The auxiliary variable T in (1.2) takes on values in the set $\mathcal{T}$ of memoryless functions (or "strategies") mapping the state S into the input X. Therefore, a code for the Dirty-Tape channel (with i.i.d. states) is given by a set $\mathcal{C} \subseteq \mathcal{T}^n$ of sequences $\mathbf{c}_m = (c_{m,1}, \ldots, c_{m,n})$ of mappings $c_{m,i} : S \to X$, such that the input sequence $\mathbf{x}$ corresponding to the information message m and the state sequence $\mathbf{s}$ is obtained as $x_i = c_{m,i}(s_i)$.

When both S and X are binary, then $\mathcal{T} = \{\mathbf{0}, \mathbf{1}, \mathbf{id}, \mathbf{not}\}$, i.e., the identically zero, identically 1 functions, identity and negation, respectively. The transition probability of the associated channel with input T and output Y, given by

$$P_{Y|T}(y|t) = \sum_{s} P_{Y|X,S}(y|t(s), s)P_S(s),$$

is given by the table

Y/T	**0**	**1**	**id**	**not**
0	$1/2$	$1/2$	$1-p$	p
1	$1/2$	$1/2$	p	$1-p$

We start with an upperbound to capacity. Due to the concavity of capacity as a function of the input average "cost" W and to the fact that the input symbol **0** has zero asscoiated cost, we have [**21**]

$$C \leq W \sup_{t\in\mathcal{T}} \frac{D(P_{Y|T=t} \| P_{Y|T=\mathbf{0}})}{E[d_H(t(S), 0)]} = 2W(\log 2 - h(p)) \tag{2.5}$$

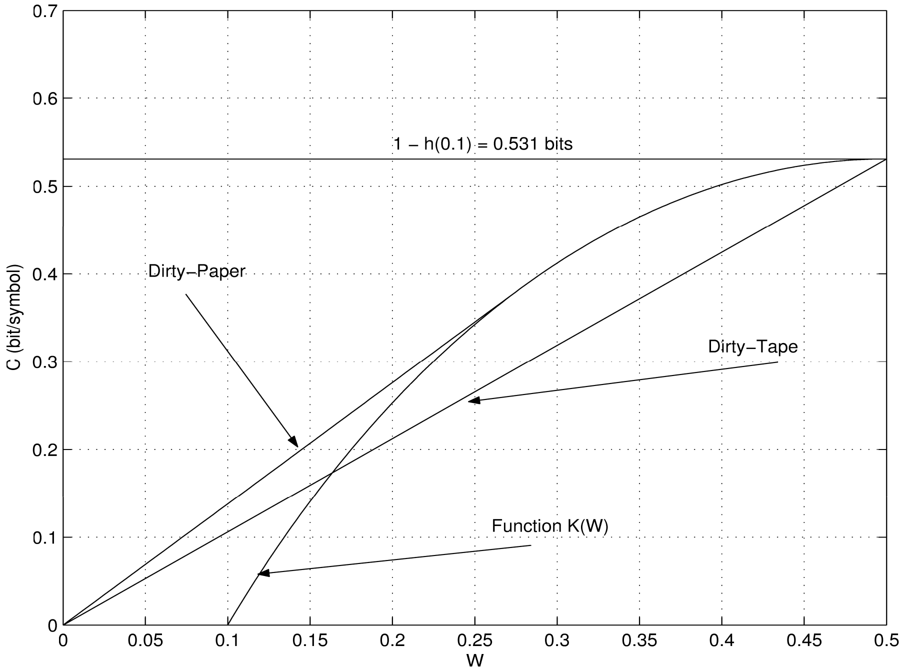

FIGURE 1. Capacity of the input-constrained BSC with causally and non-causally known interference, for $p = 0.1$.

where the supremization is achieved by either $t = \mathbf{id}$ or by $t = \mathbf{not}$. In order to show that (2.5) is indeed the capacity, we find an input probability assignment P_T such that $I(T;Y)$ equals (2.5). The input constraint is given by

$$E[d_H(T(S),0)] = P_T(\mathbf{0}) \cdot 0 + P_T(\mathbf{1}) \cdot 1 + \frac{1}{2}\left(P_T(\mathbf{id}) + P_T(\mathbf{not})\right) \leq W \tag{2.6}$$

By direct inspection, it is clear that the optimal input probability assignment must put zero probability on the input $\mathbf{1}$, since $P_T(\mathbf{1}) > 0$ would increase the input average Hamming weight, without increasing mutual information, as $I(T = \mathbf{1}; Y) = 0$. Moreover, by symmetry, it must be $P_T(\mathbf{id}) = P_T(\mathbf{not})$ (notice that this choice makes the output distribution uniform). Hence, we choose $P_T(\mathbf{id}) = P_T(\mathbf{not}) = \theta/2$, and $P_T(\mathbf{0}) = 1 - \theta$ for some $\theta \in [0, 1]$. The resulting mutual information is given by

$$I(T;Y) = \theta\left(\log 2 - h(p)\right) \tag{2.7}$$

that, by letting $\theta = 2W$, coincides with the upperbound (2.5).

Fig. 1 shows the Dirty-Paper and the Dirty-Tape capacities vs. the input constraint W for $p = 0.1$. Causal knowledge of the interference sequence incurs a noticeable capacity loss with respect to non-causal knowledge. The capacity per unit cost [**21**], defined as the maximum number of information bits per input "one" (the input cost function is $d_H(x, 0)$) and given by $U = \frac{d}{dW} C(W)\big|_{W=0}$, yields

$$\begin{aligned} U^{\text{non-causal}} &= \log \frac{1}{1 - \exp(-h(p))} - h(p) \\ U^{\text{causal}} &= 2(\log 2 - h(p)) \end{aligned} \tag{2.8}$$

3. AWGN channel with known arbitrary interference

Consider the real additive noise channel $Y = X + S + Z$, where $Z \sim \mathcal{N}(0, \sigma^2)$ is white Gaussian noise, the input is constrained by $E[X^2] \leq \mathcal{E}$ and S is an interference signal arbitrarily distributed according to some P_S. This model differ from Costa's channel since P_S is not necessarily known, therefore, the random coding argument for achievability (outlined in Section 2 for the BSC) does not hold any longer, as it makes use of typicality and, implicitly, exploits the knowledge of P_S. Nevertheless, in [**5, 6**] an explicit coding strategy referred to as *inflated lattice precoding* is shown to achieve capacity $\frac{1}{2}\log(1+\mathcal{E}/\sigma^2)$ when the interference signal *realization* is known non-causally.

Inflated lattice precoding is based on lattice dithered quantization and on the scaling of the received signal (inflation) followed by lattice decoding. Dithering requires that encoder and decoder share common randomness [2], and allows the extension of Costa's result to any interference statistics and even to arbitrary interference sequences [**6**], making it an efficient coding approach for the arbitrary varying channel with states known to the transmitter [**22**].

Moreover, the inflated lattice scheme can be applied to the case where interference is known non-causally with anticipation k, i.e., the encoder at time i knows the realization of the interference signal $(s_1,\dots,s_{\min\{i+k-1,n\}})$. In particular, for $k=1$ is provides an effective coding strategy for the Dirty-Tape setting.

The inflated lattice precoding scheme is illustrated in Fig. 2. Let k be an integer dividing the block length n, and consider a k-dimensional lattice Λ with fundamental Voronoi cell $\mathcal{V}$, with normalized second-order moment $\mathcal{E}$ (i.e., such that $\frac{1}{V(\Lambda)}\int_{\mathcal{V}}|\mathbf{x}|^2 d\mathbf{x}=k\mathcal{E}$).

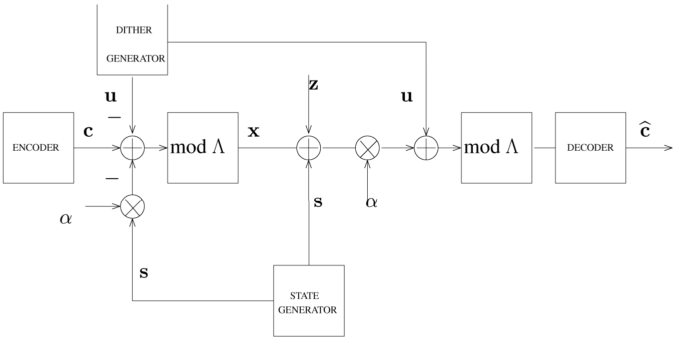

FIGURE 2. The inflated lattice precoding scheme.

A code $\mathcal{C}$ of rate R and block length n is constructed according to a uniform distribution over $\mathcal{V}$, so that

$$\mathcal{C}\subseteq\mathcal{V}^{n/k}\triangleq\underbrace{\mathcal{V}\times\cdots\times\mathcal{V}}_{n/k \text{ times}}$$

Furthermore, the transmitter and the receiver generate the same *dither* signal $\mathbf{u}$, uniformly distributed over $\mathcal{V}^{n/k}$.

For an n-dimensional vector $\mathbf{v}$, the vector $\mathbf{v} \bmod \Lambda^{n/k}$ is defined as

$$\mathbf{v} \bmod \Lambda^{n/k}=\mathbf{v}-\boldsymbol{\lambda}(\mathbf{v}) \tag{3.1}$$

where

$$\boldsymbol{\lambda}(\mathbf{v})\triangleq\arg\min_{\boldsymbol{\lambda}\in\Lambda^{n/k}}|\mathbf{v}-\boldsymbol{\lambda}|^2 \tag{3.2}$$

In other words, reducing $\mathbf{v}$ modulo $\Lambda^{n/k}$ consists of quantizing $\mathbf{v}$ by using $\Lambda^{n/k}$ as a lattice quantizer, and computing the total quantization error vector.

Let $\mathbf{c}\in\mathcal{C}$ be the codeword to be transmitted. After observing the interference signal $\mathbf{s}$, the transmitter produces the channel input sequence

$$\mathbf{x}=[\mathbf{c}-\alpha\mathbf{s}-\mathbf{u}] \bmod \Lambda^{n/k} \tag{3.3}$$

[2] Notice that sharing randomness is common practice in wireless communications. For example, in standard randomly spread CDMA transmitter and receiver share the (pseudo-)random spreading code generator.

where $\alpha \in [0,1]$ is a scaling coefficient (to be optimized), and sends $\mathbf{x}$. Thanks to the dither signal $\mathbf{u}$, $\mathbf{x}$ is uniformly distributed on $\mathcal{V}^{n/k}$ and its average energy per symbol is $\mathcal{E}$, so that the power input constraint is satisfied.

After receiving $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z}$, the receiver computes

$$\mathbf{y}' = [\alpha \mathbf{y} + \mathbf{u}] \bmod \Lambda^{n/k} \tag{3.4}$$

It can be shown [**5, 6**] that the channel from the encoder output to the decoder input (see Fig. 2) is equivalent to the additive modulo$-\Lambda^{n/k}$ noise channel

$$\mathbf{y}' = [\mathbf{c} + \mathbf{z}'] \bmod \Lambda^{n/k} \tag{3.5}$$

where $\mathbf{z}'$ is distributed as $[(1-\alpha)\mathbf{u} + \alpha\mathbf{z}] \bmod \Lambda^{n/k}$. Finally, the decoder computes (or approximates) the ML decision

$$\widehat{\mathbf{c}} = \arg \max_{\mathbf{c} \in \mathcal{C}} \, p_{Z'}(\mathbf{y}' - \mathbf{c}) \tag{3.6}$$

where $p_{Z'}$ denotes the pdf of $\mathbf{z}'$ defined above.

As anticipated above, inflated lattice precoding for $k = n$ applies to the non-causal Dirty-Paper case and, by choosing a sequence of good n-dimensional lattices, it achieves the AWGN capacity $\frac{1}{2}\log_2(1 + \mathcal{E}/\sigma^2)$, where the optimized inflation factor is given by [**4**]

$$\alpha = \frac{\mathcal{E}}{\sigma^2 + \mathcal{E}} \tag{3.7}$$

For the Dirty-Tape case ($k = 1$), we have $\Lambda = \Delta\mathbb{Z}$, with $\Delta = \sqrt{12\mathcal{E}}$, and the Voronoi region $\mathcal{V}$ is the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$. The scheme is a special case of Shannon's coding over strategies, where the code is *randomized* (via the dither signal) and the strategy alphabet is given by

$$\mathcal{T} = \{t_{v,u}(s) = [v - \alpha s - u] \bmod [-\Delta/2, \Delta/2] \, : \, v \in [-\Delta/2, \Delta/2]\} \tag{3.8}$$

Random coding over $\mathcal{T}$ where the code ensemble is generated according to a uniform distribution achieves the rate

$$R^{\text{prec.}} = \max_{\alpha \in [0,1]} \{\log \Delta - h(Z')\} \tag{3.9}$$

where the optimization of the inflation coefficient α is obtained numerically. Fig. 3 shows the optimal α as a function of the SNR $\stackrel{\Delta}{=} \mathcal{E}/\sigma^2$. The optimal Dirty-Paper value (3.7) is shown for comparison.

By choosing α as in (3.7), we obtain the lower bound

$$R^{\text{prec.}} \geq \frac{1}{2}\log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right) - \frac{1}{2}\log\frac{2\pi e}{12} \tag{3.10}$$

where $\frac{1}{2}\log_2 \frac{2\pi e}{12} = 0.254$ bits is the "shaping gain". Moreover, in [**5**] it is proved that the above rate is asymptotically optimal, i.e., it coincides with the capacity of the AWGN Dirty-Tape channel with arbitrary interference, in the limit of high SNR. It is interesting to notice that, for high SNR, the penalty incurred by causal versus non-causal interference knowledge is only the shaping gain.

4. LDPC coded-modulation for writing on Dirty-Tape

In this section we consider the AWGN Dirty-Tape channel and we construct inflated lattice coding schemes approaching the performance of random coding with uniform probability over the alphabet $\mathcal{T}$ defined in (3.8). The obtained codes can be seen either as "Dirty-Tape" codes, either as "Dirty-Paper" codes with a suboptimal choice of the lattice Λ in the general inflated lattice precoding scheme outlined in Section 3.

It is natural to construct the code $\mathcal{C}$ by concatenating a linear binary code with M-PAM signal set

$$\mathcal{A} = \left\{ \frac{\Delta}{2M}(2m - M + 1) \, : \, m = 0, \dots, M-1 \right\}$$

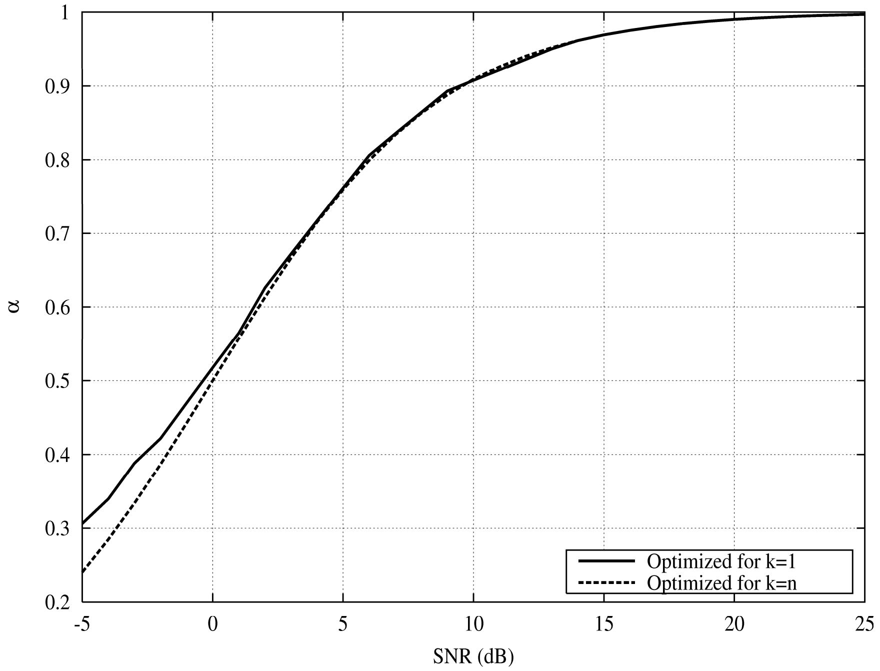

FIGURE 3. Optimal inflation factor for $k = 1$ and $k = n \to \infty$.

thus inducing an input marginal distribution close to the uniform distribution over $[-\Delta/2, \Delta/2]$, for large PAM alphabet size M.

The first-order transition pdf of the corresponding modulo-noise channel (defined in general by (3.5)) is given by

$$p_{Z'}(z) = \sum_{k\in\mathbb{Z}} \frac{P_Z\left(\frac{z+k\Delta+(1-\alpha)\Delta/2}{\alpha}\right) - P_Z\left(\frac{z+k\Delta-(1-\alpha)\Delta/2}{\alpha}\right)}{(1-\alpha)\Delta} \tag{4.1}$$

where $P_Z(z)$ denotes the cdf of the noise of the original channel. In this work we consider AWGN, therefore $P_Z(z)$ is the Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

Next, we consider two code constructions for $\mathcal{C}$: LDPC multilevel coding and direct design of LDPC-coded modulation.

4.1. Multilevel coding. Multilevel coding (see [**23**] and references therein) is a general method for constructing coded modulation schemes. Fig. 4 shows a block diagram of the multilevel encoder, where m binary codes produce codewords $\mathbf{c}_1, \ldots, \mathbf{c}_m$ of length n. These are arranged as rows of a $m \times n$ matrix, and a binary labeling function $\phi : \mathbb{F}_2^m \to \mathbf{A}$ is applied columnwise, in order to form the corresponding codeword of $\mathcal{C}$.

As binary component codes we choose LDPCs from the database of LDPC ensembles optimized for the binary-input AWGN channel provided in [**24**]. Following [**23**], the choice of the component code rates is dictated by the mutual information chain rule. Namely, let A be uniformly distributed on $\mathcal{A}$ and let $(b_1, \ldots, b_m)$ be binary uniform random variables, then

$$I(A; Y') = I(b_1, \ldots, b_m; Y') = \sum_{i=1}^{m} R_i$$

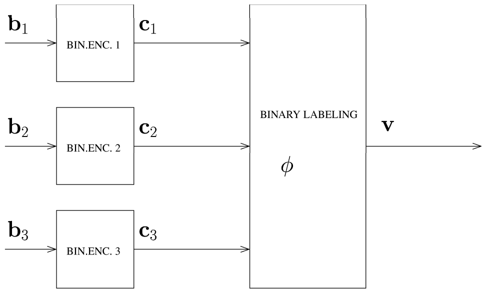

FIGURE 4. Multilevel coded modulator with three levels.

where we define the rate at level i as

$$\begin{aligned} R_i &= I(Y'; b_i | b_1, \ldots, b_{i-1}) \\ &= E\left[\log_2 \frac{\sum_{a \in \mathcal{A}(b_1,\ldots,b_i)} p_{Z'}(Y'-a)}{\sum_{a' \in \mathcal{A}(b_1,\ldots,b_{i-1})} p_{Z'}(Y'-a')}\right] \end{aligned} \tag{4.2}$$

where $\mathcal{A}(b_1, \ldots, b_i)$ denotes the subset of points of $\mathcal{A}$ whose label first i positions are given by $(b_1, \ldots, b_i)$. The rates R_i are achievable by a multistage decoder that considers the levels in sequence, by decoding each level i assuming that the decoding outcomes at previous levels $1, \ldots, i-1$ are correct, and by treating the levels $i+1, \ldots, m$ as a random nuisance. The multistage decoder is analogous to the well-known successive interference cancellation decoder used in Gaussian multiaccess and broadcast channels [**20**].

The binary labeling ϕ affects the level rates R_i but, as long as ϕ is a one-to-one mapping, the total mutual information is independent of ϕ. Fig. 5 shows an example of the set-partitioning labeling [**25**] considered in this work for 8-PAM codes and Fig. 6 shows the corresponding achievable rate of the 8-PAM modulo-noise channel, with the level rates R_i.

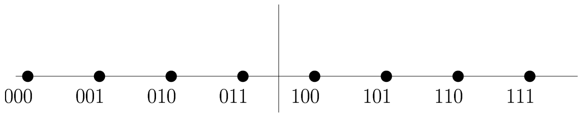

FIGURE 5. Set-partitioning labeling of 8-PAM.

Levels are decoded from right to left of the labels. We notice that for high SNR only the first level need coding at rate $R_1 < 1$, while the other two can be transmitted uncoded (i.e., $R_2 = R_3 = 1$). This provides a very simple scheme requiring a single decoding stage followed by symbol-by-symbol detection of the remaining stages. Hence, multilevel coding with set-partitioning labeling is particularly attractive in the high-SNR region. Incidentally, this is also the region where the inflated lattice precoding scheme with $k = 1$ pays the smallest relative penalty with respect to the full non-causal case (as already noticed, the shaping gain 0.254 bit/symbol).

4.2. Direct optimization. Our second approach consists of directly optimizing the ensemble of LDPC-coded modulation schemes constructed over the M-PAM alphabet $\mathcal{A}$, by exploiting the *density-evolution* method developed to analyze LDPC codes under message-passing decoding in the limit of infinite block length [**26, 27**]. Fig. 7 shows the Tanner graph of the code, where the bitnodes are partitioned into subsets of m nodes, each of which is associated to the m label positions of a M-PAM symbol. The super-nodes corresponding to modulation symbols will be referred to as "$\mathcal{A}$-nodes".

We say that a $\mathcal{A}$-node is of *type* $(d_1, \ldots, d_m)$ if its i-th label bitnode has degree d_i. We enumerate the $\mathcal{A}$-node types in lexicographic order, and let $d_{t,i}$ be the degree of the i-th

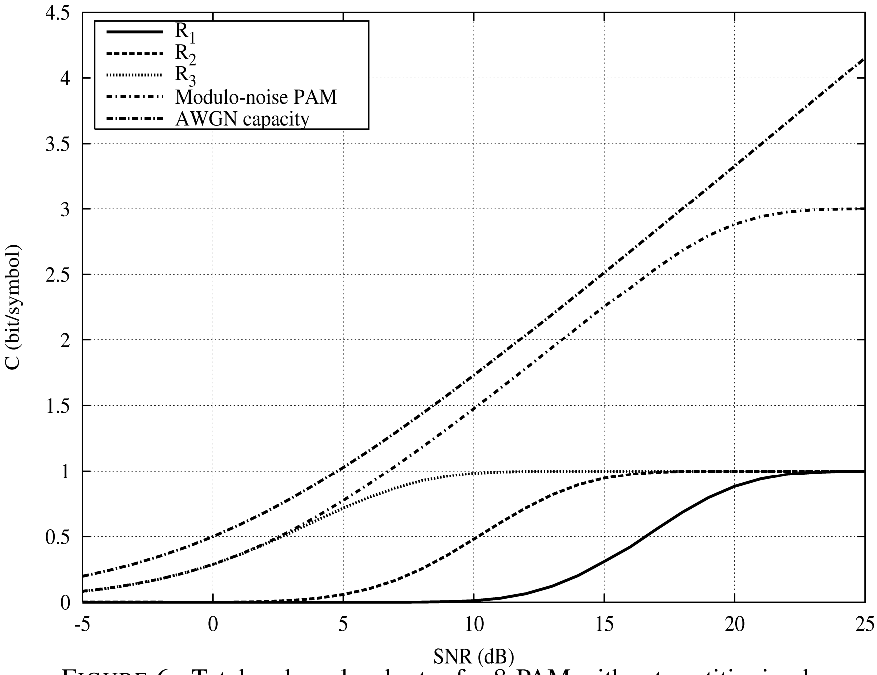

FIGURE 6. Total and per-level rates for 8-PAM with set-partitioning labeling in the modulo-noise channel.

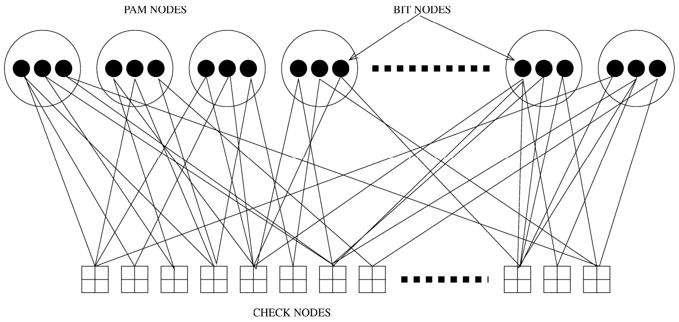

FIGURE 7. Tanner graph of an LDPC-coded modulation scheme.

bitnode in the $\mathcal{A}$-nodes of type t. We let $\lambda_{t,i}$ denote the fraction of edges connected with $\mathcal{A}$-nodes of type t in position i. For a graph with e edges, the number of $\mathcal{A}$-nodes of type t is given by $n_t = e\lambda_{t,i}/d_{t,i}$, therefore $\lambda_{t,i}/d_{t,i}$ must not depend on i.

As in standard LDPC notation, we let ρ_j be the fraction of edges connected to checknodes of degree j. The ensemble optimization consists of finding, for any given SNR, the set of $\mathcal{A}$-node types, the left degree distribution $\{\lambda_{t,i}\}$ and the right degree distribution $\{\rho_j\}$ such that the coding rate is maximized subject to the constraint that the density evolution (DE) converges to zero bit-error probability as the number of iterations goes to infinity (see [**26, 28**] for the details). Since optimizing the degree distributions based on the exact DE is computationally very intensive, we propose a design method based on a one-dimensional approximation of DE, obtained by approximating the message densities by Gaussian pdfs (see [**27**] and especially [**29**] for similar approaches in different contexts).

From $\sum_{t,i} \lambda_{t,i} = 1$ we obtain the constraint $\sum_t \lambda_{t,1} \sum_{i=1}^m \frac{d_{t,i}}{d_{t,1}} = 1$. The design coding rate is given by

$$R = \log_2 M - \frac{\sum_j \rho_j / j}{\sum_t \lambda_{t,1}/d_{t,1}} \quad \text{bit/symbol} \tag{4.3}$$

For given right degree sequence $\{\rho_j\}$, we wish to obtain an optimization problem in the variables $\{\lambda_{t,1}\}$.

Consider an $\mathcal{A}$-node of type $(d_1, \dots, d_m)$ and consider an edge o connected in position i. The message-passing transformation of the iterative belief-propagation decoder associated to the message output by the node onto edge o is given by

$$\mathcal{L}_{i,o}^{\text{out}} = \log \frac{\sum_{a \in \mathcal{A}_0^i} p_{Z'}(y' - a) \exp\left(-\sum_{j=1}^m b_j \sum_{u=1}^{d_j} \mathcal{L}_{j,u}^{\text{in}}\right)}{\sum_{a \in \mathcal{A}_1^i} p_{Z'}(y' - a) \exp\left(-\sum_{j=1}^m b_j \sum_{u=1}^{d_j} \mathcal{L}_{j,u}^{\text{in}}\right)} - \mathcal{L}_{i,o}^{\text{in}} \tag{4.4}$$

where $\mathcal{L}_{j,u}^{\text{in}}$ denotes the input message from edge u connected to the bitnode in position j, and $\mathcal{A}_0^i$ (resp. $\mathcal{A}_1^i$) denotes the signal subset of all points of $\mathcal{A}$ having symbol 0 (resp. 1) in label position i and where y' denotes the channel output corresponding to the given $\mathcal{A}$-node.

It has been shown in [**28**] that the message pdfs generated by the belief propagation algorithm at each iteration satisfy a *symmetry condition*. If the pdfs are Gaussian, the symmetry condition imposes that the variance must be twice the mean. Assuming that all input messages are Gaussian i.i.d. $\sim \mathcal{N}(\mu, 2\mu)$, we can obtain by Monte Carlo simulation of (4.4) the distribution of $\mathcal{L}_{i,o}^{\text{out}}$ for every i and node type t, parameterized in the input mean value μ.

Following [**29**], we shall replace the pdf $f_{\mathcal{L}}(z) \stackrel{\Delta}{=} \frac{d}{dz} \Pr(\mathcal{L} \leq z | b = 0)$ of a message $\mathcal{L}$ relative to a bitnode b by the value of the mutual information functional $I(b; \mathcal{L})$ which, for symmetric pdfs, is given by

$$I(b; \mathcal{L}) = 1 - \int_{-\infty}^{\infty} \log_2\left(1 + e^{-z}\right) f_{\mathcal{L}}(z) dz \tag{4.5}$$

Then, from the Gaussian assumption of the input messages and the explicit mapping (4.4), for any $\mathcal{A}$-node type and bitnode position i we can find (numerically) a mutual information transfer function

$$\mathsf{y} = \Gamma_{t,i}(\mathsf{x}) \tag{4.6}$$

where $\mathsf{x} \stackrel{\Delta}{=} I(b; \mathcal{L}^{\text{in}})$ and $\mathsf{y} \stackrel{\Delta}{=} I(b; \mathcal{L}^{\text{out}})$.

Again as a consequence of the message pdf symmetry, the mutual information between a message $\mathcal{L}^{\text{out}}$ and its associated bitnode value b on a randomly selected graph edge, chosen with probability $\lambda_{t,i}$, is given by

$$\mathsf{y} = \sum_{t,i} \lambda_{t,i} \Gamma_{t,i}(\mathsf{x}) \tag{4.7}$$

Fig. 8 shows the functions $\Gamma_{t,i}$ for an 8-PAM-node of type $(2, 5, 10)$.

For the message-passing mapping at the checknodes we use the approximated duality property [**30**], stating that the mutual information transfer function of a checknode is closely approximated by the mutual information transfer function of a bitnode with the same degree, by applying the mapping $\mathsf{x} \mapsto 1 - \mathsf{x}$ to the input and $\mathsf{y} \mapsto 1 - \mathsf{y}$ to the output. Assuming the input messages i.i.d. Gaussian $\sim \mathcal{N}(\mu, 2\mu)$ and by defining the binary-input Gaussian mutual information functional

$$J(\mu) \stackrel{\Delta}{=} 1 - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-z^2} \log_2\left(1 + e^{-2\sqrt{\mu}z - \mu}\right) dz, \tag{4.8}$$

the mutual information transfer function of a checknode of degree j is given by

$$\mathsf{y} = 1 - J\left((j-1) J^{-1}(1 - \mathsf{x})\right)$$

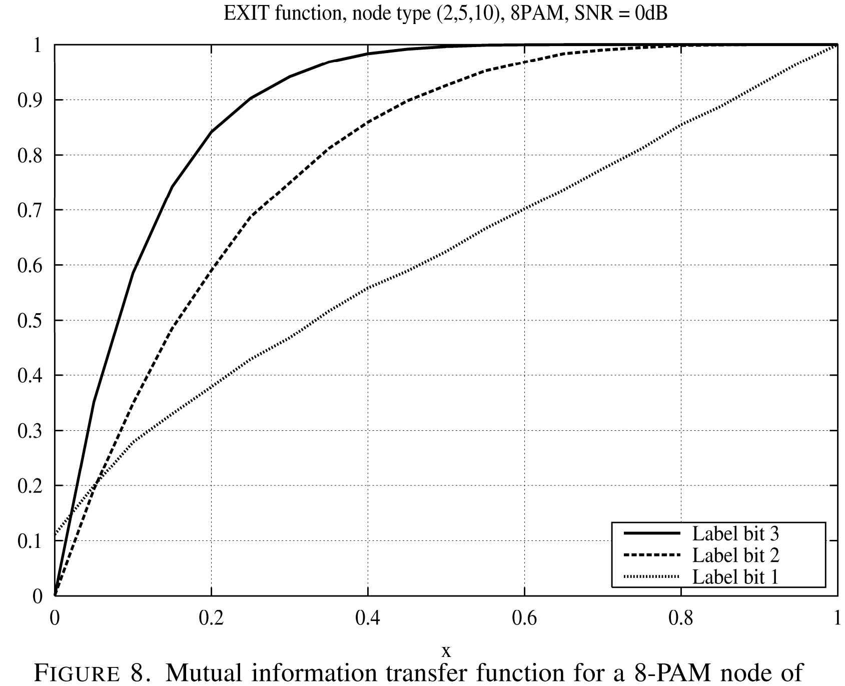

FIGURE 8. Mutual information transfer function for a 8-PAM node of type $(2, 5, 10)$ for SNR= 0 dB.

Eventually, the one-dimensional DE approximation is given by the recursion

$$\mathsf{x}^{(\ell)} = \sum_{t,i} \lambda_{t,i}\Gamma_{t,i}\left(1 - \sum_j \rho_j J\left((j-1)J^{-1}\left(1 - \mathsf{x}^{(\ell-1)}\right)\right)\right) \tag{4.9}$$

for $\ell = 1, 2, \ldots$, with initial condition $\mathsf{x}^{(0)} = 0$.

The recursion (4.9) has a unique stable fixed point in $\mathsf{x} = 1$ (corresponding to vanishing bit-error probability), if and only if

$$\mathsf{x} > \sum_{t,i} \lambda_{t,i}\Gamma_{t,i}\left(1 - \sum_j \rho_j J\left((j-1)J^{-1}(1 - \mathsf{x})\right)\right), \quad \forall\, \mathsf{x} \in [0, 1) \tag{4.10}$$

Hence, we can sample the above equation for x taking on values in a fine grid of points in the interval $[0, 1)$ and for each point we obtain a linear constraint in the variables $\lambda_{t,1}$ (recall that $\lambda_{t,i} = \frac{d_{t,i}}{d_{t,1}}\lambda_{t,1}$, therefore the only independent variable of the optimization problem are $\{\lambda_{t,1}\}$). Since both the constraints and the objective function (4.3) are linear in the $\{\lambda_{t,1}\}$, the solution is readily obtained by linear programming.

4.3. Results. Fig. 9 shows the rate versus E_b/N_0 of some multilevel LDPC codes and LDPC-coded M-PAM codes obtained according to the methods described above, for the AWGN Dirty-Tape channel. In all cases, the block length (in PAM symbols) is 20000 and the marks correspond to bit-error probability $\leq 10^{-4}$ with a maximum of 100 decoder iterations.

Marks labeled by "2PAM" correspond to standard binary LDPC, marks labeled by "MCM MPAM" correspond to scheme obtained by multilevel coded modulation and marks labeled by "LDPC-CM MPAM" correspond to direct optimization of LDPC-coded modulation.

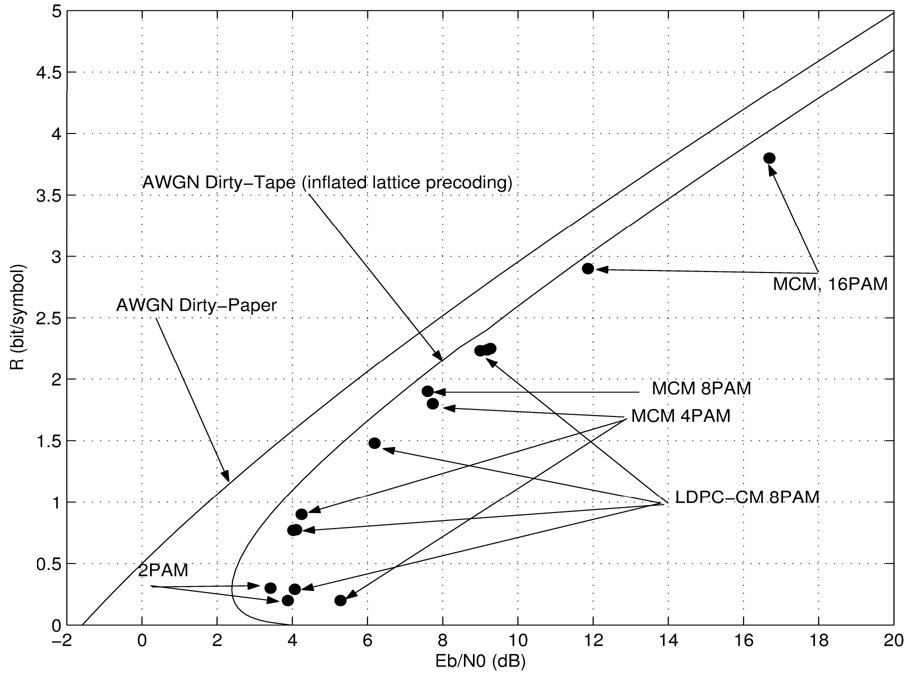

FIGURE 9. Performance of LDPC-coded modulation schemes for the dirty-tape channel with AWGN.

We observe that the theoretical achievable rate $R^{\rm prec.}$ given by (3.9) can be closely approached by our LDPC-coded PAM modulation, especially for high SNR. For low SNR, the best results are provided by standard binary LDPC codes ($M = 2$), while for high SNR the multilevel construction with one or two coding levels and the remaining levels left uncoded proves to be both efficient and simple in terms of complexity. There exists a region of intermediate SNR where it is indeed worthwhile to construct explicitly optimized LDPC-coded modulation, although better understanding and more refined optimization are called for, since the simple schemes that we experimented do not show a dramatic improvement with respect to multilevel codes. For direct optimization we restricted to codes over the 8-PAM alphabet, but the coding design ideas apply immediately to other cases.

It is interesting to notice that the minimum E_b/N_0 (corresponding to the inverse of the capacity per unit-cost) achieved by inflated lattice precoding with $k = 1$ is not obtained for vanishing SNR. This is due to the fact that the rate $R^{\rm prec.}(\mathcal{E}/\sigma^2)$ (seen as a function of $\mathcal{E}/\sigma^2$) is not concave, therefore, in a region of low SNR a better rate is obtained by time-sharing between SNR$=0$ and some positive SNR*. Fig. 10 shows $(E_b/N_0)^{\rm prec.}$, defined implicitly by the equation

$$\left(\frac{E_b}{N_0}\right)^{\rm prec.} R^{\rm prec.}(\mathcal{E}/\sigma^2) = \frac{\mathcal{E}}{2\sigma^2} \tag{4.11}$$

versus SNR. From this figure we find SNR$^\star \approx 0$ dB.

5. Discussion

From the results of previous section we observe that inflated lattice precoding incurs a significant loss in achievable rates for low SNR. It is natural to ask whether this loss is due to the suboptimality of the inflated lattice strategy alphabet, or it is due to causal versus non-causal interference knowledge. Assuming i.i.d. interference with known distribution P_S, the capacity per unit-cost of the AWGN Dirty-Tape channel can be investigated by using the general formula of [**21**]. In particular, let $\mathcal{T}$ denote the set of all functions $t : \mathbb{R} \to \mathbb{R}$

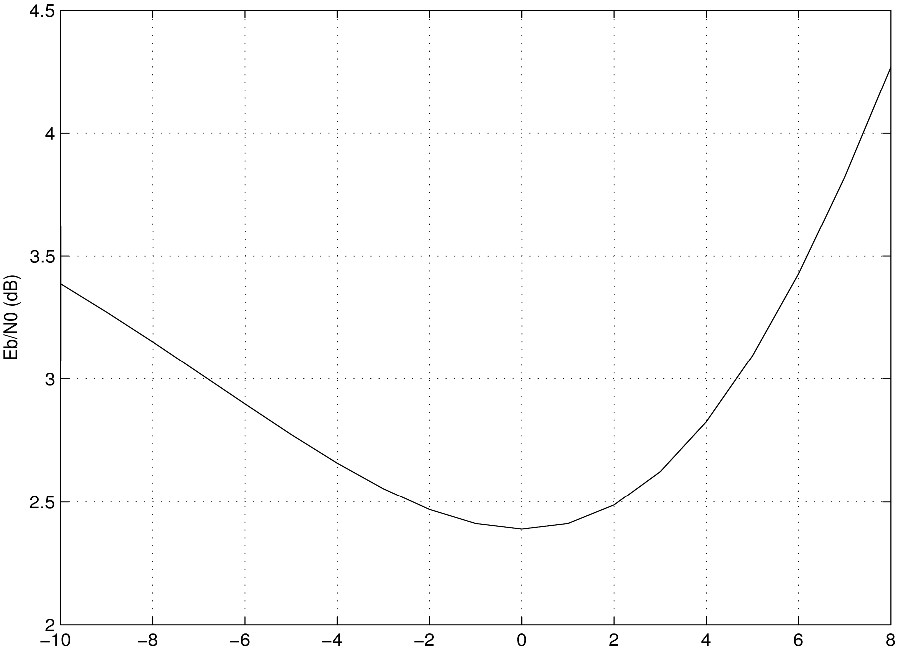

FIGURE 10. E_b/N_0 corresponding to the Dirty-Tape inflated lattice achievable rate vs. SNR.

mapping S into X, and let

$$P_{Y|T}(y|t) = \int P_{Y|X,S}(y|t(s), s)dP_S(s)$$

where $P_{Y|X,S}(y|x,s) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-x-s)^2}{2\sigma^2}}$. Then,

$$U^{\text{non-causal}} = \sup_{t \in \mathcal{T}} \frac{D\left(P_{Y|T=t} \| P_{Y|T=\mathbf{0}}\right)}{E[t(S)^2]} \tag{5.1}$$

where, as before, **0** denotes the indentically zero function. The supremization in (5.1) does not yield, in general, a simple solution. However, a lower-bound to $U^{\text{non-causal}}$ for some specific P_S can be found by restricting t to take on a particular form. For example, assume $S \sim \mathcal{N}(0, \sigma_S^2)$, define $\gamma = \sigma^2/\sigma_S^2$, and consider the set of affine functions

$$t(s) = as + b$$

We find [**31**]

$$U^{\text{non-causal}} \geq \frac{1}{2\sigma_S^2} \sup_{a \geq -1} \frac{-\frac{1-(1+a)^2}{1+\gamma} - \log\left(1 - \frac{1-(1+a)^2}{1+\gamma}\right)}{a^2} \tag{5.2}$$

where the RHS is the rate per unit-cost achieved by on-off affine-strategy signaling. The "on" input is $t^\star(s) = a^\star s$, with $a^\star$ given by the supremization in (5.2), and the "off" input is **0**, where the duty-cycle is $\theta = \mathcal{E}/((a^\star)^2\sigma_S^2)$.

Fig. 11 shows the minimum E_b/N_0 corresponding to the above on-off signaling versus γ. This is compared with the minimum $(E_b/N_0)^{\text{prec.}}$ (which is independent of γ). As expected, for weak interference power (i.e., for large γ) on-off signaling approaches -1.59 dB, the minimum E_b/N_0 of the AWGN interference-free channel. However, on-off affine signaling is not robust to the interference power, as its minimum E_b/N_0 increases as γ decreases. On the contrary, the inflated lattice strategy yields minimum E_b/N_0 independent of the interference power. This example shows that for some interference distributions P_S

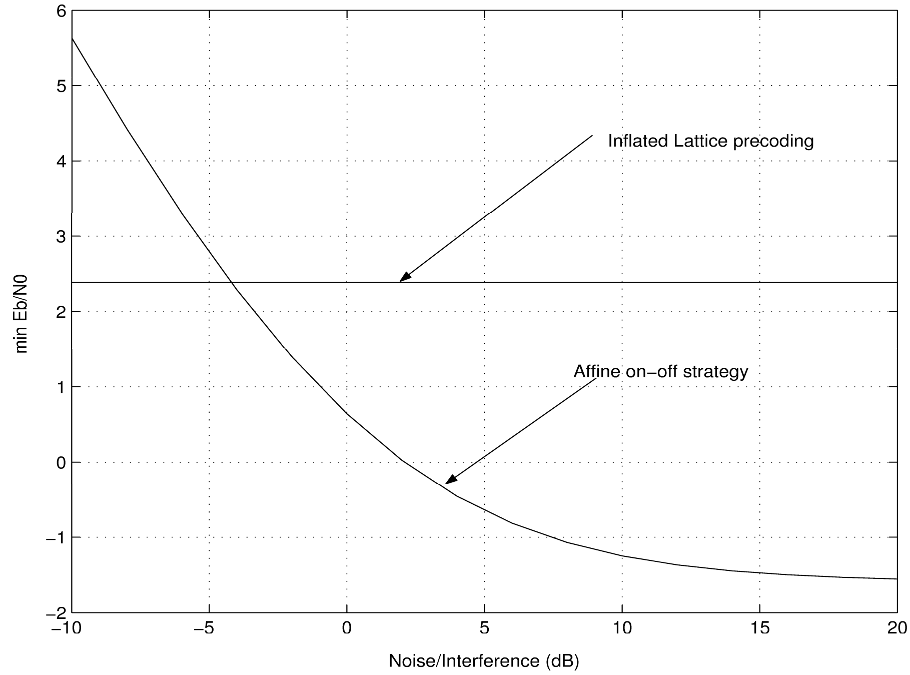

FIGURE 11. Minimum E_b/N_0 corresponding to the on-off affine strategy signaling and to inflated lattice precoding for the AWGN Dirty-Tape channel.

we can find coding strategies that outperform inflated lattice precoding (in some range of non-asymptotically large SNR). Whether the convex upper envelope of the achievable rates $R^{\text{prec.}}$ is the worst-case Dirty-Tape capacity over all possible interference distributions P_S is still an open conjecture.

An important generalization of the coding schemes presented here considers the case of finite anticipation $k > 1$. The same approach of multilevel coding or direct optimization can be applied to a k-dimensional lattice Λ. Let $\mathbf{G} \in \mathbb{R}^{k \times k}$ be the generator matrix of Λ. Then, the code $\mathcal{C}$ will be constructed on the constellation $\mathcal{A} = \mathbf{G}(M\text{−PAM})^k$, of M^k points in the fundamental Voronoi cell $\mathcal{V}$ of Λ (after appropriate scaling and translation). Depending on Λ, the modulo-Λ operations at the transmitter and receiver can be implemented either by some ML lattice decoder (e.g., based on a trellis representation of Λ) or by the general-purpose sphere decoder [**32, 33**]. As noticed in [**6**], the shaping gain plays a very important role also in the low-SNR region, and there is no hope of approaching Costa's Dirty-Paper limit for low SNR by using low-dimensional lattices (small k).

Finally, an interesting information-theoretic problem consists of studing the achievable rates with some restrictions of the code alphabet, e.g., the constraint that $\mathcal{C}$ must be a binary code.

References

[1] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. & Dev.*, pp. 289–293, 1958.

[2] A. Kusnetsov and B. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, no. 2, pp. 52–60, 1974.

[3] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, January 1980.

[4] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[5] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," submitted to IEEE Trans. on Inform. Theory. See also ISITA 2000, Honolulu, Hawaii (USA) November 2000.

[6] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[7] G. Caire and S. Shamai, "On achievable rates in a multi-antenna broadcast downlink," in *38th Annual Allerton Conf. on Commun., Cont. and Comp.*, Monticello, IL, October 2000.

[8] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," submitted to IEEE Trans. on Inform. Theory, July 2001.

[9] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," submitted to IEEE Trans. on Inform. Theory, August 2002.

[10] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum capacity of Gaussian MIMO broadcast channels," submitted to IEEE Trans. on Inform. Theory, August 2002.

[11] M. Schubert and H. Boche, "Joint "Dirty-Paper" precoding and downlink beamforming," in *Proc. of Int. Symp. on Inform. Theory and Appl.*, Prague, September 2002.

[12] G. Kramer, S. Vishwanath, S. Shamai, and A. Goldsmith, "Information-theoretic issues concerning broadcasting," in *Proc. Workshop on Sig. proc. for Wireless Comm. (DIMACS 2002)*, Rutgers University, New Jersey, October 2002.

[13] G. Ginis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *34-th Asilomar Conference on Signals, Systems and Computers*, Pacific-Groove, CA, USA, November 2000.

[14] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[15] J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," to appear on IEEE Trans. on Sig. Proc., 2002.

[16] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

[17] U. Erez and R. Zamir, "Lattice decoded nested codes achieve the Poltyrev exponent," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT 2002)*, Lausanne, Switzerland, June 2002.

[18] S. Pradhan, J. Chou and K. Ramchandran, "Duality between source coding and channel coding with side information," UCB/ERL Technical Memorandum No. M01/34, UC Berkeley, Dec. 2001.

[19] R. Gallager, *Information theory and reliable communication*, Wiley, New York, 1968.

[20] T. Cover and J. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[21] S. Verdú, "On Channel Capacity per Unit Cost," *IEEE Trans. on Inform. Theory,* Vol. 36, No. 5, pp. 1019–1030, September 1990.

[22] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. on Inform. Theory*, vol. 32, no. 5, pp. 621–629, September 1986.

[23] U. Wachsmann, R. Fisher, and J. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

[24] R. Urbanke et al., "`http://lthcwww.epfl.ch/research/ldpcopt/`," 2002.

[25] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Inform. Theory*, vol. 28, no. 1, pp. 55–67, January 1982.

[26] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[27] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 657–670, February 2001.

[28] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[29] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Commun.*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[30] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: A model and two properties," in *Proc. 36th Annual Conf. on Inform. Sc. and Syst. (CISS 2002)*, Princeton, New Jersey, March 2002.

[31] S. Shamai and A. Lapidoth, Private communication.

[32] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, No. 7, pp. 1639–1642, July 1999.

[33] E. Agrell, T. Eriksson, A. Vardy and K. Zeger, "Closest point search in lattices," *IEEE Trans. on Inform. Theory*, vol. 48, No. 8, pp. 2201–2214, Aug. 2002.

INSTITUT EURECOM, DEPARTMENT OF MOBILE COMMUNICATIONS, 2229 ROUTE DES CRETES, 06094 SOPHIA-ANTIPOLIS, FRANCE

E-mail address: `giuseppe.caire@eurecom.fr`

TECHNION, ISRAEL INSTITUT OF TECHNOLOGY, DEPARTMENT OF ELECTRICAL ENGINEERING, 32000 HAIFA, ISRAEL

E-mail address: `sshlomo@ee.technion.ac.il`