

## Superposition Coding for Costa Channels

Amir Bennatan, David Burshtein\*

Giuseppe Caire†

Shlomo Shamai†

Tel Aviv University  
School of Electrical Engineering  
abn@eng.tau.ac.il, burstyn@eng.tau.ac.il

Institut Eurecom  
caire@eurecom.fr

Technion  
Faculty of Electrical Engineering  
sshlomo@ee.technion.ac.il

### Abstract

We present practical codes designed for the Gaussian dirty paper (Costa) channel. We show that the dirty paper decoding problem can be transformed into an equivalent multiple-access problem, for which we apply superposition coding. Our approach is a generalization of the nested lattices approach of Zamir, Shamai and Erez. We present simulation results which confirm the effectiveness of our methods.

### I. Introduction

The Gaussian dirty paper channel is given by  $Y = X + S + Z$ .  $S$  is called the *interference* and constitutes the channel state, which is known only to the encoder.  $X$  is the channel input, subject to a power constraint  $P_X$ .  $Z$  is distributed as a zero-mean Gaussian variable with variance  $P_Z$ , and we make no assumptions on the distribution of  $S$ . Costa [6] obtained the remarkable result that the interference, known only to the encoder, incurs no loss of capacity in comparison with the standard interference-free channel<sup>1</sup>.

The achievability of the capacity of dirty paper channels is proven by means of a random construction of codes and a random partitioning of their codewords into “bins”. This method typically produces unstructured codes, which are infeasible for practical implementation. Zamir, Shamai and Erez [16] suggested

a framework for introducing structure into the above “random binning” method. Their technique involves nested codes (and nested lattices). That is, they use a *fine* code  $\mathcal{C}$  and a *coarse* code  $\mathcal{C}_0$  such that  $\mathcal{C}_0 \subset \mathcal{C}$ . Their construction requires that the fine code  $\mathcal{C}$  be designed as a good channel-code, while the coarse code  $\mathcal{C}_0$  be designed as a good source-code.

LDPC codes are likely candidates for codes  $\mathcal{C}$  and  $\mathcal{C}_0$ . However, although LDPC codes are well suited for channel coding, the problem of finding a good source-coding algorithm for them remains open. Unless such an algorithm is found, the codes in their current form are unsuitable for selection as  $\mathcal{C}_0$ . We would like to select  $\mathcal{C}$  as an LDPC code, but the nested structure of  $\mathcal{C}$  and  $\mathcal{C}_0$  means that the codes are entangled in a way that restricts the independent selection of  $\mathcal{C}$ . One approach for challenging this problem, called *syndrome dilution*, was considered by Philosof *et al.* [14].

In [3] we presented an alternative to the nested lattices method of [16] using superposition of codes, that enables independent selection of a *quantization* code  $\mathcal{C}_0$  and an *auxiliary* code  $\mathcal{C}_1$ . This construction was designed for the binary dirty-paper channel. In this work we extend this construction to the Gaussian dirty-paper channel.

### II. Insight

A precise definition of superposition coding is provided in section III. We begin, however, with an informal, intuitive derivation that provides insight into how and why the approach works.

In this section we assume the interference  $S$  to be spherically uniform (asymptotically) with arbitrary large power  $P_S$  (this can be approximated by adding a large Gaussian distributed dither signal). The reason for this will be clarified shortly.

Our encoder begins by encoding data into code-

---

<sup>1</sup>The first two authors were supported by the Israel Science Foundation, grant no. 22/01-1, by an equipment grant from the Israel Science Foundation to the school of Computer Science in Tel Aviv University and by a fellowship from The Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel Aviv University.

The last two authors were supported by the EU 6th framework programme, via the NEWCOM network of excellence.

<sup>1</sup>Costa’s result was obtained for the case of Gaussian distributed interference. This was later extended in [9, 5] to arbitrarily distributed interference.

words of an *auxiliary* code  $\mathcal{C}_1$ . Our objective is to achieve the no-interference channel capacity, and hence we set the rate of the code at approximately  $R_1 = \frac{1}{2} \log(1 + P_X/P_Z)$ . We select  $\mathcal{C}_1$  to be a good Gaussian channel code with components individually selected according to a Gaussian distribution with variance  $Q = \alpha P_X$ .  $\alpha$  is set at  $P_X/(P_X + P_Z)$ .

We select  $\mathcal{C}_0$  as a good *channel and quantization*. We require the code to be able to quantize the signal  $-\mathbf{c}_1 + \alpha \mathbf{s}$  with a mean-square distortion of  $P_X$ . We achieve this by random i.i.d selection of the codewords' components according to a Gaussian distribution with power  $E(c_0^2) = \alpha P_X + \alpha^2 P_S - P_X$  (for large enough  $P_S$  we obtain  $E(c_0^2) \geq 0$ ). The code rate  $R_0$  is accordingly set at  $R_0 = \frac{1}{2} \log((\alpha P_X + \alpha^2 P_S)/P_X)$ .

**Encoder:** As mentioned above, the encoder first selects a codeword  $\mathbf{c}_1$  according to the message to be transmitted. It then transmits  $\mathbf{x} = \left[ -\mathbf{c}_1 + \alpha \mathbf{s} \right]_{\mathcal{C}_0}$ , defined as the difference between  $\mathbf{c}_0$ , the nearest codeword of  $\mathcal{C}_0$ , and  $-\mathbf{c}_1 + \alpha \mathbf{s}$ . Thus, by our above selection of  $\mathcal{C}_0$ , we are guaranteed with high probability to satisfy the transmitter's power constraint. Equivalently,

$$\mathbf{x} = \mathbf{c}_0 - (-\mathbf{c}_1 + \alpha \mathbf{s}) = \mathbf{c}_0 + \mathbf{c}_1 - \alpha \mathbf{s},$$

The received signal is

$$\mathbf{y} = \mathbf{c}_0 + \mathbf{c}_1 + (1 - \alpha)\mathbf{s} + \mathbf{z}$$

**Decoder:** The decoder constructs

$$\begin{aligned} \hat{\mathbf{y}} &= \alpha \mathbf{y} \\ &= \mathbf{c}_0 + \mathbf{c}_1 - (1 - \alpha)\mathbf{x} + \alpha \mathbf{z} \\ &= \mathbf{c}_0 + \mathbf{c}_1 + \hat{\mathbf{z}} \end{aligned} \quad (1)$$

where  $\hat{\mathbf{z}} \triangleq -(1 - \alpha)\mathbf{x} + \alpha \mathbf{z}$  is the *effective noise*. The decoder seeks the pair  $(\hat{\mathbf{c}}_0, \hat{\mathbf{c}}_1)$  such that  $\hat{\mathbf{c}}_0 + \hat{\mathbf{c}}_1$  is closest to  $\hat{\mathbf{y}}$ , and  $\hat{\mathbf{c}}_1$  is announced as the decoded codeword.

The equivalence to the MAC decoder model is evident from (1). We now analyze the probability of decoding error assuming successive decoding.

The successive decoding process is composed of two steps. The decoder first decodes  $\hat{\mathbf{c}}_0$ , treating  $\mathbf{c}_1 + \hat{\mathbf{z}}$  as an unknown random noise vector  $\mathbf{z}_0$ . Assuming  $P_S \gg 1$ , it can be claimed that  $\mathbf{z}_0$ , which contains elements of correlated self-noise, acts as an independent factor<sup>2</sup>. Using this approximation, the power  $P_{Z_0}$  of  $\mathbf{z}_0$  is shown [1] to be exactly  $P_X$ . The code  $\mathcal{C}_0$ , whose parameters were

<sup>2</sup>In Section III we will use a dithered signal to ensure decorrelation.

designed for quantization with a mean-square distortion of  $P_X$ , is shown [1] to be also capable of decoding with a noise variance of  $P_X$ .

The decoder next removes the interference caused by  $\mathbf{c}_0$  in order to compute  $\hat{\mathbf{c}}_1$ . It now decodes  $\mathbf{c}_1$  from the signal  $\hat{\mathbf{y}} - \mathbf{c}_0$ , which is equal to  $\mathbf{c}_1 + \hat{\mathbf{z}}$ .

The power of  $\mathcal{C}_1$  is  $Q = \alpha P_X = P_X^2/(P_X + P_Z)$ , while the power of the effective noise  $P_{\hat{\mathbf{z}}} = P_X P_Z/(P_X + P_Z)$ . We therefore obtain [1] that the achievable rate is exactly  $\frac{1}{2} \log(1 + P_X/P_Z)$  as desired. Hence  $\mathcal{C}_1$  is decodable with high probability, yielding the no-interference capacity rate.

Important insight provided by this formulation involves the practical selection of codes  $\mathcal{C}_0$  and  $\mathcal{C}_1$ .  $\mathcal{C}_0$  operates always in extremely good SNR conditions (as  $P_S$  is taken to be large). This is why in practice it is important to choose  $\mathcal{C}_0$  as an efficient trellis based coding/quantization code a la Forney's coset trellis codes [11] and relevant references therein. The code  $\mathcal{C}_1$  frequently operates at low SNR and hence could be selected as a binary LDPC code.

### III. Formal Definition

We again consider two codes, a quantization code  $\mathcal{C}_0$  and an auxiliary code  $\mathcal{C}_1$ . The superposition code is defined as  $\mathcal{C} = \mathcal{C}_0 + \mathcal{C}_1 \bmod A$ , where the operation  $\bmod A$  is applied componentwise as follows: Given a scalar  $x$ ,  $x \bmod A \triangleq x - Q_A(x)$  such that  $Q_A(x)$  is the nearest multiple of  $A$  to  $x$ . The dynamic range of  $x$  is thus reduced to  $[-A/2, A/2]$ .

The modulo- $A$  operation is borrowed from the construction A approach to generating lattices from linear codes. Its effect can be equivalently modelled as the tessellation of the entire space  $\mathbb{R}^n$  with replicas of the  $n$ -dimensional cube  $[-A/2, A/2]^n$ . Note that it must not be confused with the modulo-lattice operation of the nested lattices scheme of [16], which serves a different purpose.

**Encoder:** The encoder selects a codeword  $\mathbf{c}_1 \in \mathcal{C}_1$ , and sends the sequence  $\mathbf{x} = [\alpha \mathbf{s} + \mathbf{d} - \mathbf{c}_1 \bmod A]_{\mathcal{C}_0} \bmod A$ .

$A$  and  $\alpha$  are arbitrary constants.  $\mathbf{d}$  is a randomly selected *dither* signal, borrowed from the nested lattices approach of [16]. However, unlike [16], the elements of the dither are defined to be uniformly i.i.d in the range  $[-A/2, A/2]$ .

$[\boldsymbol{\xi}]_{\mathcal{C}_0} \triangleq Q_{\mathcal{C}_0}(\boldsymbol{\xi}) - \boldsymbol{\xi} \bmod A$ ,  $Q_{\mathcal{C}_0}(\boldsymbol{\xi})$  being the codeword of  $\mathcal{C}_0$  that is closest to  $\boldsymbol{\xi}$  assuming a *modulo A distance metric*. The *mod A* distance between two vectors  $\mathbf{x}$  and  $\mathbf{y}$  is given by  $\|\mathbf{y} - \mathbf{x}\|_A^2 \triangleq \sum_{i=1}^n (y_i - x_i \bmod A)^2$ .  $A$  and  $\alpha$  are parameters that may be optimized to obtain best

performance. We obtain,

$$\begin{aligned}\mathbf{x} &= [Q_{\mathcal{C}_0}(\alpha\mathbf{s} + \mathbf{d} - \mathbf{c}_1 \bmod A) - (\alpha\mathbf{s} + \mathbf{d} - \mathbf{c}_1)] \bmod A \\ &= \mathbf{c}_0 + \mathbf{c}_1 - \alpha\mathbf{s} - \mathbf{d} \bmod A\end{aligned}$$

where  $\mathbf{c}_0 \triangleq Q_{\mathcal{C}_0}(\alpha\mathbf{s} + \mathbf{d} - \mathbf{c}_1 \bmod A)$ .

We select  $\mathcal{C}_0$  to be capable of quantization with a mean square distortion  $P_X$ , assuming  $\bmod A$  distance. Thus  $\mathbf{x}$  is guaranteed to satisfy the power constraint,  $1/n \cdot \sum_{i=1}^n x_i^2 \leq P_X$ . The received signal is  $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z}$ .

**Decoder:** The decoder computes

$$\begin{aligned}\hat{\mathbf{y}} &\triangleq \alpha\mathbf{y} + \mathbf{d} \bmod A \\ &= \mathbf{c}_0 + \mathbf{c}_1 - (1 - \alpha)\mathbf{x} + \alpha\mathbf{z} \bmod A \\ &= \mathbf{c}_0 + \mathbf{c}_1 + \hat{\mathbf{z}} \bmod A\end{aligned}\quad (2)$$

where  $\hat{\mathbf{z}} \triangleq -(1 - \alpha)\mathbf{x} + \alpha\mathbf{z}$  is the effective noise. The decoder evaluates the pair  $(\hat{\mathbf{c}}_0, \hat{\mathbf{c}}_1)$  such that  $\hat{\mathbf{c}}_0 + \hat{\mathbf{c}}_1$  is closest to  $\hat{\mathbf{y}}$ , assuming  $\bmod A$  distance.  $\hat{\mathbf{c}}_1$  is announced as the decoded codeword.

From the above construction, it is clear that  $\mathbf{c}_0$  and  $\mathbf{c}_1$  are independent, implying the analogy to the MAC channel. However, the effective noise  $\hat{\mathbf{z}}$  contains a “self-noise” element  $\mathbf{x}$  that, for particular choices of  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , is not independent of  $\mathbf{c}_0$  and  $\mathbf{c}_1$ , undermining an assumption of the Gaussian MAC model. Nonetheless, in Section IV we show that under a random-coding assumption, the MAC model is valid and a decoder designed for the MAC channel is capable of achieving the Costa capacity.

#### IV. Random Coding Analysis

We now provide a rigorous proof of the validity of the MAC model, assuming the formulation provided in Section III. For this purpose, we consider the following channel model:

$$Y = X + S + Z \bmod A/\alpha \quad (3)$$

$Y$  in this model corresponds to the channel output as in (2) after the modulo operation was performed, but without multiplication by  $\alpha$ . Hence the argument to the modulo operation is  $A/\alpha$  instead of  $A$ . For simplicity of our model, we encapsulate the random known dither into the interference  $S$ , and assume that the interference is uniformly distributed in the range  $[-A/(2\alpha), A/(2\alpha)]$ .

We define the following set of auxiliary variables:  $X$  is Gaussian with variance  $P_X$ .  $U_1$  is distributed as a Gaussian variable with variance  $Q$  (which will be determined later). The variables  $S$ ,  $X$  and  $U_1$  are independent. We also define  $U_0$  to satisfy the equation:

$$U_0 = \alpha S + X - U_1 \bmod A \quad (4)$$

Hence,  $U_0$  is uniformly distributed in the range  $[-A/2, A/2]$  and is dependent on  $S$ ,  $X$  and  $U_1$ . Note that  $X$  is identical to a similar definition by Costa [6].  $U_0$  and  $U_1$  replace Costa’s auxiliary  $U$ .

We construct codes  $\mathcal{C}_0$  and  $\mathcal{C}_1$  by random i.i.d selection according to the distribution  $U_0$  and  $U_1$ , respectively, at rates  $R_0$  and  $R_1$ . The code  $\mathcal{C} = \mathcal{C}_0 + \mathcal{C}_1 \bmod A$  is the dirty paper code. For each  $\mathbf{c}_1$ , the set  $\mathbf{c}_1 + \mathcal{C}_0$  is equivalent to a bin of the theoretical analysis of Costa.

To simplify our analysis, we consider an encoder/decoder pair that employ joint-typicality rather than minimum-distance metrics.

**Encoder:** The encoder selects a codeword  $\mathbf{c}_1 \in \mathcal{C}_1$ , and seeks a word  $\mathbf{c}_0 \in \mathcal{C}_0$  such that the pair  $\mathbf{c}_0$  and  $(\alpha\mathbf{s} - \mathbf{c}_1 \bmod A)$  are jointly strongly  $\epsilon$ -typical with respect to the distribution of the random variables  $U_0$  and  $(\alpha S - U_1 \bmod A)$  ( $\epsilon$  will be determined later). If no such  $\mathbf{c}_0$  is found, the encoder declares an error. Otherwise, it transmits the sequence  $\mathbf{x} = \mathbf{c}_0 + \mathbf{c}_1 - \alpha\mathbf{s} \bmod A$ .

Note that the encoder requires *strong* typicality. The justification for this is similar to the one in the theoretical analysis of Gel’fand and Pinsker [12] and is clarified later, in the proof of Lemma 2, which is provided in [1]. Lemma 1 examines the probability of an encoder error.

**Lemma 1** *Let  $\delta > 0$  be an arbitrary number and assume  $R_0$  satisfies*

$$R_0 > \log A - \frac{1}{2} \log(2\pi e P_X) + \delta \quad (5)$$

*Then for  $A$  large enough, there exists a constant  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$ , ( $\epsilon$  having been defined above) then the average probability of an encoder error approaches zero with the block length  $n$ .*

The proof of the lemma is provided in [1]. We now define  $\hat{Y} = \alpha Y$  and combine (3) and (4) to obtain,

$$\begin{aligned}\hat{Y} &= U_0 + U_1 - (1 - \alpha)X + \alpha Z \bmod A \\ &= U_0 + U_1 + \hat{Z} \bmod A\end{aligned}\quad (6)$$

where  $\hat{Z} \triangleq -(1 - \alpha)X + \alpha Z$ .  $X$  and  $Z$  are independent of  $U_0$  and  $U_1$ , and hence  $\hat{Z}$  is also independent of  $U_0$  and  $U_1$ .  $\hat{Z}$  is distributed as a Gaussian variable with variance  $P_{\hat{Z}} \triangleq (1 - \alpha)^2 P_X + \alpha^2 P_Z$ .

The effective noise element  $\hat{Z}$  in (6) is independent of  $U_0$  and  $U_1$ , thus overcoming the obstacle in (2). Since  $\mathcal{C}_0$  and  $\mathcal{C}_1$  were constructed according to  $U_0$  and  $U_1$ , we would expect the probability of error to approach zero if  $(R_0, R_1)$  lie within the capacity region of the MAC channel as defined in Lemma 14.3.1 of [7].

The proof, however, is slightly more involved than the proof of [7]. This is because the channel output  $\hat{\mathbf{y}}$  was not generated according to the true MAC channel model (6). Specifically, the self-noise element  $\mathbf{x}$  of  $\hat{\mathbf{z}}$  was not generated by random selection according to  $X$ . This obstacle is overcome in the proof of Lemma 2 [1]. We begin by replacing the decoder of [7] with a decoder that requires strong (rather than weak) typicality.

**Decoder:** The decoder seeks  $\hat{\mathbf{c}}_0 \in \mathcal{C}_0$  and  $\hat{\mathbf{c}}_1 \in \mathcal{C}_1$  such that the triplet  $(\hat{\mathbf{c}}_0, \hat{\mathbf{c}}_1, \mathbf{y})$  are jointly strongly  $\epsilon$ -typical with respect to the distribution of  $(U_0, U_1, Y)$ .

We now examine the probability of a decoder error.

**Lemma 2** *Let  $\delta > 0$  be an arbitrary number and assume  $R_0$  and  $R_1$  satisfy*

$$R_0 + R_1 < \log A - \frac{1}{2} \log(2\pi e P_{\hat{Z}}) - \delta \quad (7)$$

$$R_1 < \frac{1}{2} \log \left( 1 + \frac{Q}{P_{\hat{Z}}} \right) - \delta \quad (8)$$

Then for  $A$  large enough, there exists a constant  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$ , then the average probability of a decoder error approaches zero with the block length  $n$ .

The proof of this lemma is provided in [1]. Finally, we combine Lemmas 1 and 2 to obtain:

**Theorem 1** *Let  $\delta > 0$  be an arbitrary number and assume the above defined superposition coding scheme. Let  $\epsilon$  be the argument to the strong typicality tests of the above described encoder and decoder. Assume  $R_0$  and  $R_1$  satisfy (5), (7) and (8). Then if  $A$  is large enough, there exists  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$ , the average probability of error approaches zero with the block length  $n$ .*

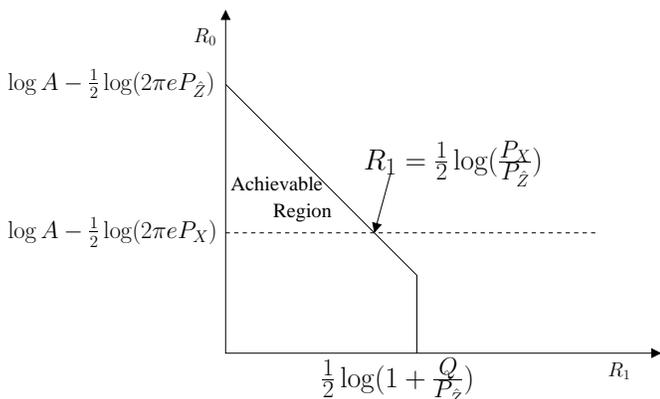


Figure 1: Capacity region of the equivalent AWGN MAC channel.

Fig. 1 presents the capacity region prescribed by Theorem 1. A dashed line marks the constraint imposed by equation (5) of Lemma 1. Transmission is possible at any point that is within the MAC capacity region and is above the dashed line.

The MAC capacity region is a function not only of the power constraint  $P_X$  and the noise variance  $P_Z$  but also of parameters  $\alpha$  and  $Q$ . A selection of  $\alpha = P_X/(P_X + P_Z)$  and  $Q \geq P_X^2/(P_X + P_Z)$  produces a capacity region where the Costa capacity is achieved at the point  $R_1 = 1/2 \log(1 + P_X/P_Z)$  and  $R_0 = \log A - 1/2 \log(2\pi e P_X)$  (note that  $R_1$  is the effective rate of the coding scheme).

This result is valid for any selection of  $Q$  that satisfies  $Q \geq P_X^2/(P_X + P_Z)$ . A selection of  $Q = P_X^2/(P_X + P_Z)$  produces a MAC capacity region where the Costa capacity is achieved at a vertex point. This point is interesting from a practical implementation viewpoint. At this point, successive decoding of  $\mathcal{C}_0$  and  $\mathcal{C}_1$  is possible. The vertex point is also preferred by joint iterative belief-propagation decoding.

## V. Comparison with Nested Lattices

We now compare superposition coding as defined in Section III with the nested lattices approach of [16]. We assume lattices produced by construction-A, as used by Philosof *et al.* [13, 14] and by Erez and ten Brink [10].

Assume a pair of nested lattices  $\Lambda$  and  $\Lambda_0$  ( $\Lambda_0$  a coarse lattice nested in a fine lattice  $\Lambda$ ) constructed from a pair of nested codes  $\mathcal{C}$  and  $\mathcal{C}_0$ . We can construct an equivalent pair  $\mathcal{C}_0$  and  $\mathcal{C}_1$  for superposition coding by leaving  $\mathcal{C}_0$  unchanged and setting  $\mathcal{C}_1$  to contain the codewords of  $\mathcal{C}$  that fall within the basic Voronoi cell of  $\mathcal{C}_0$ . We select  $A$  to be the argument to the construction-A of lattices  $\Lambda_0$  and  $\Lambda$  from  $\mathcal{C}_0$  and  $\mathcal{C}$  (i.e.  $\Lambda = \mathcal{C} + A \cdot Z^n$ ).  $\alpha$  remains unaltered from the nested lattices construction. The obtained superposition encoding and decoding operations become identical to the equivalent nested lattices operations. Thus, superposition coding is a generalization of nested lattices. Nevertheless, superposition coding focuses on settings where  $\mathcal{C}_0$  and  $\mathcal{C}_1$  are designed differently.

One difference between the two approaches lies in the selection  $\mathcal{C}_0$  and  $\mathcal{C}$ . For construction-A to produce a lattice, both codes need to be linear under modulo- $q$  arithmetic ( $q$  being some positive integer). With superposition coding, the equivalent codes  $\mathcal{C}_0$  and  $\mathcal{C}$  are allowed to be nonlinear.

The use of a random dither is common to both methods. The dither of [16] is uniformly distributed over the basic Voronoi cell of the coarse lattice  $\Lambda_0$ , rather than as defined in Section III. However,

Philosof [13] has shown that assuming construction-A, the dither may equivalently be defined in the range  $[-A/2, A/2]$ , thus simplifying its random generation. This is identical to the dither used in superposition coding.

The theoretical results of Section IV are valid for  $A$  approaching infinity. This requirement appears to be common also to nested lattices, where  $A$  is the argument for construction-A as defined above. For nested lattices coding to achieve capacity, the lattices need to be “good” for both source and channel coding [16]. Good nested lattices were presented in [8]. The lattices were generated using construction-A, and exhibit values of  $A$  that approach infinity<sup>3</sup>. However, further research is necessary in order to establish if this requirement is indeed inherent to any method that produces good lattices.

An important difference between the two approaches is in their preferred variance  $Q$  for the code  $\mathcal{C}_1$ . Nested lattices coding is equivalent to a selection of  $\mathcal{C}_1$  that is uniformly distributed in the Voronoi cell of  $\mathcal{C}_0$ . This roughly corresponds to a selection of  $Q = P_X$ . Superposition coding, as noted in Section IV, prefers a smaller value of  $Q$  which corresponds to a vertex of the MAC capacity region.

Lastly, a small advantage of superposition coding over syndrome dilution is a simpler encoder design. With syndrome dilution, encoding involves selection of a coset leader to represent the transmitted codeword. This step is not required by superposition coding.

## VI. Simulation Results

We experimented at a rate of 0.25 bits per real dimension (which is equivalent to the 0.5 bits per complex dimension as used by Erez and ten Brink [10]), at an SNR of -2.5dB. The dirty paper Costa (and Shannon) limit at this rate is -3.82 dB.

For the quantization code  $\mathcal{C}_0$ , we selected a Trellis code borrowed from Ungerboeck [15] of memory 9. The feedback polynomials are given by the octal digits (1072, 0342), the output alphabet consisted of the 4-PAM signals  $[-0.75 -0.25 0.25 0.75]$ . Simulation results indicate that the code is capable of quantization with a mean square distortion of  $P_X = 0.061$ . The random-coding achievable distortion for rate 1 bit per channel use is approximately 0.0585, and hence  $\mathcal{C}_0$  op-

<sup>3</sup>The construction of [8] examines lattices with a constant  $A = 1$ . However, it is assumed that the lattices are to be scaled to fit any practical problem setting. The effective radius  $r_A^{\text{effec}}$  of the lattices is held (approximately) constant while the block length  $n$  is taken to infinity. This implies that the effective power approaches zero, requiring a scaling factor that approaches infinity.

erates close to the limit. The noise variance was thus set to  $P_Z = 0.108$ , in accordance with the above selected SNR of -2.5dB.

We selected a binary LDPC code for  $\mathcal{C}_1$ . We mapped the code bits to the BPSK signals  $\pm 0.1482$ , approximately corresponding to an energy of  $Q = P_X^2 / (P_X + P_Z) = 0.0219$ . The code’s edge distribution is given by  $\lambda(2, 3, 4, 5, 7, 35) = (0.531357, 0.147539, 0.249499, 0.0525727, 0.00601187, 0.0130199)$ , and  $\rho(3, 4) = (0.501, 0.499)$ . The rate of  $\mathcal{C}_1$  is 0.25.

We used a joint iterative belief-propagation decoder to decode  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , using a BCJR decoder for  $\mathcal{C}_0$  and a belief-propagation decoder for  $\mathcal{C}_1$ . The decoders exchanged soft data in accordance with the concepts of Boutros and Caire [2]. The decoders alternated at a rate of at least 10 LDPC iterations per BCJR iteration (the ratio was changed throughout the decoding process).

We obtained the edge distribution for  $\mathcal{C}_1$  using a method that is based on Chung *et al.* [4]. The method requires “singleton” error probabilities, which in [4] are produced by density evolution. In our work we have instead used the output of simulations.

An additional improvement was obtained by applying a non-random ordering on the bits in the code  $\mathcal{C}_1$ . i.e. bits corresponding to same-degree variable-nodes were grouped together in consecutive indices, rather than spread randomly. Since the edges of the LDPC graph were selected by random permutation, this had no effect on the performance of the LDPC code. However, this produced a partitioning of the LDPC codeword into segments of equal-degree variable-nodes and inferred a similar partitioning on the bits of  $\mathcal{C}_0$ . Segments of the  $\mathcal{C}_0$  code bits that were connected to different segments of  $\mathcal{C}_1$  enjoyed an unequal degree of reliability of information transferred from the LDPC decoder. This resulted in irregularity in the BCJR decoder (a desirable quality in iterative soft-decoders), thus significantly improving the performance of the joint decoder.

Simulation results for the above dirty paper scheme of rate 0.25 indicate a bit error rate of approximately  $4.8 \cdot 10^{-5}$  at a block length of  $2 \cdot 10^5$  (50 simulations). As noted above, these results were obtained at an SNR approximately 1.3 dB away from the Shannon limit. These results are similar to results reported using nested lattices. In [10], reliable transmission was reported within 1.8dB of the Shannon limit, and improvement to approximately 1.3dB was reported later.

## VII. Comparison with Dirty Tape

It is worth while to compare these results to the

*dirty tape* case of  $S$  known causally. Dirty tape schemes can be used as low-complexity solutions for dirty paper problems, by simply ignoring the noncausal portion of the known interference. Ignoring part of the known data is clearly suboptimal, but nonetheless provides an important benchmark for the performance of the more complex dirty paper schemes. For a dirty paper scheme to be interesting, it must surpass the performance of schemes for the equivalent dirty tape channel.

The Gaussian dirty tape channel was examined by Erez, Shamai and Zamir [9]. They derived expressions for the capacity at the asymptotic case of strong interference. Furthermore, they suggested an efficient coding scheme capable of approaching the computed capacity. Erez and ten Brink [10] applied this scheme to achieve reliable transmission within 0.4dB of the dirty tape Shannon limit, at a rate of 0.667 bits per complex dimension.

The gap between the Shannon limits for the Gaussian dirty tape and dirty paper problems is greatest at low SNR, and approaches approximately 4dB. This was the motivation for the preference of low SNR for the dirty paper simulations of [13] and in this paper, because this is the point where there is the most potential for gain by exploiting the non-causally known data. The dirty tape Shannon limit at rate 0.25 bits per real dimension is -0.6 dB (SNR) (see [10]). This limit is thus surpassed by the above described dirty paper scheme at SNR -2.5dB.

## VIII. Conclusion

Superposition coding can be viewed as a generalization of the nested lattices approach of [16]. An advantage of our approach over nested lattices is that our codes are not required to be components of a lattice, and hence are allowed to be nonlinear, adding an extra degree of freedom to their design.

In [1], we suggested a framework for extending superposition coding to the general Gel'fand-Pinsker problem.

## References

- [1] A. Bennatan, D. Burshtein, G. Caire and S. Shamai, "Superposition Coding for Side-Information Channels," submitted for publication, *IEEE Trans. on Inform. Theory*.
- [2] J. Boutros and G. Caire "Iterative Multiuser Joint Decoding: United Framework and Asymptotic Analysis," *IEEE Trans. on Inform. Theory*, Vol. 48, No. 7, July 2002.
- [3] G. Caire, A. Bennatan, D. Burshtein and S. Shamai, "Coding Schemes for the Binary Symmetric Channel with Known Interference," *The 41st Annual Allerton Conference on Commun., Control and Computing*, Monticello, IL, Oct. 1-3, 2003.
- [4] S.-Y. Chung, J. G. D. Forney, T. Richardson and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit", *IEEE Commun. Lett.*, vol. 5, pp. 58–60, February 2001.
- [5] A. S. Cohen and A. Lapidoth, "The Gaussian Watermarking Game" *IEEE Trans. on Inform. Theory* vol. 48, no. 6, June 2002.
- [6] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [7] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
- [8] U. Erez, S. Litsyn and R. Zamir, "Lattices which are good for (almost) everything", submitted for publication *IEEE Trans. on Inform. Theory*, June 2003.
- [9] U. Erez, S. Shamai and R. Zamir, "Capacity and Lattice-Strategies for Cancelling Known Interference", *IEEE Trans. on Inform. Theory*, submitted August 2002.
- [10] U. Erez and S. ten Brink, "Approaching the dirty paper Limit for Canceling Known Interference," *41st Annual Allerton Conf. on Commun., Cont. and Comp.*, Monticello, IL, October 2003.
- [11] G.D. Forney, "Approaching the Capacity of the AWGN Channel with Coset Codes and Multilevel Coset Codes," *ISIT'97*, p. 164, Ulm, Germany, June 29–July 4, 1997.
- [12] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, January 1980.
- [13] T. Philosof, "Precoding for Interference Cancellation at Low SNR", MSc. dissertation, Jan 2003.
- [14] T. Philosof, U. Erez and R. Zamir "Combined Shaping and Precoding for Interference Cancellation at Low SNR" *Proc. 2003 IEEE International Symposium on Information Theory*, July 2003.
- [15] G. Ungerboeck, "Channel coding with multi-level/phase signals," *IEEE Trans. on Inform. Theory*, vol. 28, no. 1, pp. 55–67, January 1982.
- [16] 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.