On Universal Coding for Parallel Gaussian Channels

Maryam Modir Shanechi
Dept. EECS, MIT
Cambridge, MA
Email: shanechi@mit.edu

Uri Erez
Tel Aviv University
Ramat Aviv, Israel
Email: uri@eng.tau.ac.il

Gregory Wornell
Dept. EECS, MIT
Cambridge, MA
Email: gww@mit.edu

***Abstract*— Two classes of approximately universal codes are developed for parallel Gaussian channels whose state information is not available at the encoder. Both architectures convert the channel into a set of scalar additive white Gaussian noise (AWGN) channels to which good AWGN base codes can be applied, and both are layered schemes used in conjunction with successive interference cancellation to keep decoding complexity low. The first construction uses a concatenated code structure, and with perfect base codes achieves an arbitrarily high fraction of capacity in the limit of high SNR. The second construction uses a layer-dither-repeat structure, and with perfect base codes can achieve better than 90% of capacity at typical target spectral efficiencies, corresponding to a roughly 1 dB gap to capacity.**

I. Introduction

The design of practical universal codes for parallel Gaussian channels, when the capacity is known at the transmitter but the channel parameters themselves are not, is of significant interest in a variety of emerging wireless applications and standards based on, for example, orthogonal frequency-division multiplexing (OFDM) and multi-input multi-output (MIMO) technologies.

In principle, one can simply code across the subchannels in such applications. Indeed, the usual Gaussian random codes are capacity-achieving when used in this way, even though the signal-to-noise ratio (SNR) varies across the subchannels. However, practical codes such as low density parity check (LDPC) codes, and their associated low-complexity iterative decoding algorithms, are typically designed for use in situations where the SNR is uniform across the block. As such, it is often difficult to predict the performance of such codes when used in this way.

In this paper we develop two code architectures with low-complexity decoding that avoid this problem, and that allow standard additive white Gaussian (AWGN) channel codes to be used efficiently. Our first construction is a concatenated code that is strongly inspired by (and therefore adopts the key features of) both the permutation codes developed in [1] and the rateless codes introduced in [2], and can be viewed as modest refinement of both. The code is asymptotically perfect at high SNR, i.e., it achieves an arbitrarily high fraction of capacity in the limit of high SNR. However, as we show, at typical target spectral efficiencies, the gap to capacity for such codes is large.

This work was supported in part by NSF under Grant No. CCF-0515122, Draper Laboratory, by ONR under MURI Grant No. N00014-07-1-0738, and by an NSERC scholarship.

As a more practical architecture, we then develop a second class of codes based on a layer-dither-repeat code structure in the spirit of the rateless codes of [3]. We show that numerically optimized versions of these codes substantially reduce the gap to capacity even with small numbers of layers.

II. Channel Model and Problem Formulation

The channel model of interest takes the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1}$$

where

$$\mathbf{A} = \operatorname{diag}(\boldsymbol{\alpha}), \qquad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K), \tag{2}$$

where channel input $\mathbf{x}$ and output $\mathbf{y}$ at any particular time are K-dimensional vectors, and where the associated noise $\mathbf{w}$ is $\mathcal{CN}(0, \mathbf{I})$ and independent over time. The channel input is constrained to a total average power of KP.

In our formulation, the channel realization $\boldsymbol{\alpha}$ is known to the receiver but not to the transmitter. However, the transmitter does know the white-input capacity of the channel

$$C(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \log\left(1 + P|\alpha_k|^2\right). \tag{3}$$

The encoder generates from the message a collection of K codewords of length N, for transmission over the respective subchannels. There are 2^{NR} possible messages, corresponding to a rate R code. The receiver, in turn, collects and jointly uses the K received blocks to decode the message. The block length of N symbols is assumed to be large, but otherwise plays no role in the analysis.

A code of rate R is a perfect universal code if the message is decodable (with high probability) for any channel with parameters $\boldsymbol{\alpha} \in \mathcal{A}(C)$ such that

$$\mathcal{A}(C) = \{\boldsymbol{\alpha} : C(\boldsymbol{\alpha}) = C\} \tag{4}$$

when C is chosen arbitrarily close to R.

III. An Asymptotically Perfect Construction

In the low SNR limit, simple repetition coding across subchannels is a universal code in a meaningful asymptotic sense. By contast, in the high SNR limit, more effort is required, as we now discuss.

First, there are a variety of notions of asymptotic optimality that may be of interest. As perhaps the coarsest notion, we say a family of codes of fixed block length N and increasing rate R is asymptotically universal in the *coarse* sense if for some $f(\cdot)$ such that $f(R)/R \to 1$ as $R \to \infty$,

$$\lim_{R\to\infty} \sup_{\boldsymbol{\alpha}\in\mathcal{A}(f(R))} \overline{P}_{\mathrm{e}}(\mathcal{C}_N(R), \boldsymbol{\alpha}) = 0, \tag{5}$$

where $\mathcal{C}_N(R)$ denotes the sequence of codes, $\overline{P}_{\mathrm{e}}$ denotes the (maximal) probability of a decoding error, and $\mathcal{A}(R)$ is as defined in (4). Constructions satisfying this notion of universality are developed in [1].

Note that codes optimum in the coarse sense need not have vanishing error probability for any fixed rate and channel realization. As a somewhat more refined notion, then, we say a family of codes is asymptotically universal in the *fine* sense if $R/C_{\min}(R) \to 1$ as $R \to \infty$, where $C_{\min}(R)$ is the minimum value of C such that

$$\sup_{\boldsymbol{\alpha}\in\mathcal{A}(C)} \lim_{N\to\infty} \overline{P}_{\mathrm{e}}(\mathcal{C}_N(R), \boldsymbol{\alpha}) = 0. \tag{6}$$

A code that is asymptotically universal in the fine sense, can be constructed as the concatenation of an outer binary erasure code and an inner AWGN code. Specifically, the message is first encoded using the erasure code, the result of which is then divided among the K subchannels, and then further broken down into layers in each subchannel. For each subchannel, each layer is independently encoded using an AWGN code at a fixed rate and according to a geometric power progression over the layers. The encoded layers are then superimposed to form the block to be transmitted over the subchannel.

At the receiver, as many layers as possible on each subchannel are (independently) decoded via successive interference cancellation. The number of layers decoded depends on the realized SNR in the subchannel. The result is fed as input to the decoder for the outer erasure code, which treats the undecoded layers as erasures and reconstructs the message.

We now describe how the parameters of the code are chosen, and analyze the resulting performance. We begin with the inner AWGN codes.

For each subchannel k, we decompose its (length-N) input as $\mathbf{x}_k = \mathbf{x}_k^{(0)} + \mathbf{x}_k^{(1)} + \dots$ corresponding to the different layers (the number of which we discuss shortly), with which we use a geometric power allocation of total power P of the form

$$P_l = P(1-a)a^l, \qquad l = 0, 1, \dots \tag{7}$$

for some parameter $0 < a < 1$. As a result, the signal-to-interference ratio (SIR) is identical for all layers:

$$\mathrm{SIR} = \frac{1-a}{a}, \tag{8}$$

while the signal-to-interference+noise ratio (SINR) experienced in the decoding of the lth layer in the kth subchannel is

$$\mathrm{SINR}_k^l = \frac{|\alpha_k|^2 P_l}{|\alpha_k|^2 P_l \frac{a}{1-a} + 1}. \tag{9}$$

We allocate the same rate to all the layers of all the subchannels, viz.,

$$R_l = \log\left(1 + (1-\epsilon)\mathrm{SIR}\right), \qquad l = 0, 1, \dots \tag{10}$$

for some choice of parameter $0 < \epsilon < 1/2$. Thus, to determine how many layers are decodable, we determine for which values of l

$$\mathrm{SINR}_k^l \geq (1-\epsilon)\mathrm{SIR}. \tag{11}$$

To this end, let l_k be the (not necessarily integer) value of l in the kth subchannel such that (11) is satisfied with equality. Hence, substituting (9) into (11), we get l_k as the solution to

$$\frac{|\alpha_k|^2 P_{l_k}}{|\alpha_k|^2 P_{l_k} \frac{a}{1-a} + 1} = (1-\epsilon)\frac{1-a}{a},$$

which, using (7), yields

$$|\alpha_k|^2 P a^{l_k+1} = (1-\epsilon)/\epsilon. \tag{12}$$

Hence, the number of decodable layers in the K subchannels is

$$L = \sum_{\{k\,:\,l_k \geq 0\}} \lfloor l_k + 1 \rfloor, \tag{13}$$

which, solving (12) for l_k and substituting into (13), and noting that $\mathcal{K} = \{k : l_k \geq 0\} = \{k : |\alpha_k|^2 \geq (1-\epsilon)/(Pa\epsilon)\}$, yields

$$\begin{aligned}
L &= \sum_{k\in\mathcal{K}} \left\lfloor \frac{1}{\log(1/a)} \left(\log(|\alpha_k|^2 P) - \log\frac{1-\epsilon}{\epsilon} \right) \right\rfloor \\
&\geq \sum_{k\in\mathcal{K}} \frac{1}{\log(1/a)} \left(\log(|\alpha_k|^2 P) - \log\frac{1-\epsilon}{\epsilon} \right) - K \\
&\geq \frac{-1}{\log(a)} \left(C(\boldsymbol{\alpha}) - \sum_{k\notin\mathcal{K}} P|\alpha_k|^2 - \sum_{k\in\mathcal{K}} \frac{|\alpha_k|^{-2}}{P} - K\log\frac{1-\epsilon}{a\epsilon} \right) \qquad (14) \\
&\geq \frac{1}{\log(1/a)} \left(C(\boldsymbol{\alpha}) - K\frac{1-\epsilon}{a\epsilon} - \frac{Ka\epsilon}{1-\epsilon} - K\log\frac{1-\epsilon}{a\epsilon} \right), \qquad (15)
\end{aligned}$$

where (14) follows from using (3), $\log(x) = \log(1+x) - \log(1+1/x)$, and $\log(1+y) \leq y$ (for $x > 0$ and $y > 0$).

Next we consider the outer erasure code. We begin by observing from (14) that since $C(\boldsymbol{\alpha}) = C$ is fixed, L is also fixed (at least in the high SNR regime), though how the decodable layers are distributed across subchannels depends on the α_k's. In the most extreme case, all L decodable layers are in a single subchannel. Thus, we need a total of L layers per subchannel in our scheme, which in turn implies that the rate of the outer erasure code should be $1/K$. In practice, the outer erasure code can be implemented via a maximal distance separable (MDS) or near-MDS code designed to recover L source packets from any L of KL encoded packets.

Finally, the rate R achievable via this concatenated code

TABLE I
ACHIEVABLE EFFICIENCIES η, LAYER REQUIREMENTS L, AND OPTIMIZING PARAMETER ϵ, FOR CONCATENATED CODE AS A FUNCTION OF THE TOTAL CHANNEL CAPACITY C FOR THE CASE OF $K = 2$ SUBCHANNELS.

C, b/s/Hz	a	ϵ	L	η
20	1	0.18	∞	64%
20	0.5	0.25	14.8	60%
20	0.1	0.47	3.9	50%
50	1	0.06	∞	79%
50	0.5	0.09	41.4	77%
50	0.1	0.18	11.7	72%

satisfies

$$R = LR_l \tag{16}$$

$$\geq \frac{1}{\log(1/a)} \left(C(\boldsymbol{\alpha}) - \frac{Ka\epsilon}{1-\epsilon} - K \log \frac{1-\epsilon}{a\epsilon} \right) \cdot \log \left(1 + (1-\epsilon) \frac{1-a}{a} \right) \tag{17}$$

$$\geq (1-\epsilon) \left(C(\boldsymbol{\alpha}) - K \frac{1-\epsilon}{a\epsilon} - \frac{Ka\epsilon}{1-\epsilon} - K \log \frac{1-\epsilon}{a\epsilon} \right) \tag{18}$$

where to obtain (17) we have substituted (14) and (10) with (8) into (16), and where to obtain (18) we have exploited the inequality $\log(1+px) \geq p \log(1+x)$, which holds for any $0 \leq p \leq 1$ and $x \geq 0$, and that follows from concavity of the log function. Finally, if we let, e.g., $\epsilon = 1/\sqrt{C(\boldsymbol{\alpha})}$ and $a = 1 - e^{-C(\boldsymbol{\alpha})}$ then from (18) we obtain $R/C(\boldsymbol{\alpha}) \to 1$ as $C(\boldsymbol{\alpha}) \to \infty$, and thus the code is universal in the sense of (6).

While fine-sense asymptotically universal, this scheme is rather far from capacity even at large but finite spectral efficiencies, as the numerical results of Table I reveal. In these simulations, there are $K = 2$ subchannels, and we use the lower bound (17), optimized over the choice of ϵ for a given choice of a. The resulting lower bound and the associated number of decodable layers both increase monotonically with a. Note that with the proper choice of ϵ, the achievable efficiency is not especially sensitive to the number of layers in the scheme.

The results of Table I suggest that a concatenated code of this type, with the associated decoding procedure, while asymptotically perfect, is too far from capacity at typical target spectral efficiencies to be practical. Accordingly, in the next section we consider an alternative construction with better efficiency characteristics in the regime of interest.

IV. A NEAR-PERFECT CONSTRUCTION

Our layered construction in this section replaces the outer code with dithered repetition, and minimum mean-square error (MMSE) combining during successive interference cancellation (SIC), following the approach in [3] for rateless coding.

Specifically, the contruction for a rate R code is as follows. First, we choose the number of layers L and the associated codebooks $\mathcal{C}_1, \ldots, \mathcal{C}_L$. We further constrain the codebooks to have equal rate R/L, which has the advantage of allowing them to be derived from a single base code. We assume capacity achieving independent identically distributed Gaussian codebooks are used.

Given codewords $\mathbf{c}_l \in \mathcal{C}_l$, $l = 1, \ldots, L$, the blocks $\mathbf{x}_1, \ldots, \mathbf{x}_K$ to be distributed across the K subchannels take the form

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_L \end{bmatrix} \tag{19}$$

where $\mathbf{G}$ is an $K \times L$ matrix of complex gains and where $\mathbf{x}_k$ for each k and $\mathbf{c}_l$ for each l are row vectors of length N. The power constraint enters by limiting the rows of $\mathbf{G}$ to have squared norm P and by normalizing the codebooks to have unit power.

In addition to the layered code structure, there is additional decoding structure, namely that the layered code be successively decodable. Specifically, to recover the message, we decode one layer, treating the remaining layers as (colored) noise, then subtract its effect from the received codeword. Then we decode another layer from the residual, treating further remaining layers as noise, and so on.

There are several degrees of freedom in this code design, which can be used to maximize the efficiency of the code for a given C. At the encoder, we are free to choose $\mathbf{G}$, which has KL phase and $K(L-1)$ magnitude degrees of freedom available. At the decoder, the decoding order can be chosen as a function of the realized channel parameters $\boldsymbol{\alpha}$. Since there are $L!$ possible decoding orders, this degree of freedom can be described by a discrete variable m.

The optimization of the rate of the code over the available degrees of freedom can be expressed as

$$R = L \cdot \max_{\mathbf{G} \in \mathcal{G}} \min_{\boldsymbol{\alpha} \in \mathcal{A}(C)} \max_{1 \leq m \leq L!} \min_{1 \leq l \leq L} I_l(\boldsymbol{\alpha}, \mathbf{G}, m) \tag{20}$$

where $\mathcal{G}$ is the set of $\mathbf{G}$ satisfying the power constraint, i.e.,

$$\mathcal{G} = \{ \mathbf{G} : [\mathbf{G}\mathbf{G}^\dagger]_{k,k} = P, \; k = 1, \ldots, K \}$$

where $\mathcal{A}(C)$ is as given in (4), and where $I_l(\boldsymbol{\alpha}, \mathbf{G}, m)$ is the mutual information in the lth layer with respect to the decoding order specified by m, i.e., for $l = 1, \ldots, L$,

$$\sum_{l'=l}^{L} I_{l'}(\boldsymbol{\alpha}, \mathbf{G}, m) = \log\det \left(\mathbf{I} + \mathbf{A} [\mathbf{G}\boldsymbol{\Pi}(m)]_{l:L} [\mathbf{G}\boldsymbol{\Pi}(m)]_{l:L}^\dagger \mathbf{A}^\dagger \right), \tag{21}$$

with $\mathbf{A} = \mathrm{diag}(\boldsymbol{\alpha})$, with $\boldsymbol{\Pi}(m)$ denoting the matrix that permutes the columns of $\mathbf{G}$ according to m, and with $[\cdot]_{i:j}$ denoting the submatrix consisting of columns i through j of its argument. Finally, using (20), we obtain the resulting efficiency of the optimized code as $\eta_L = R/C$ as a function of the number of layers L in the code.

In the sequel, we restrict our attention to the case of $K = 2$ subchannels to simplify the exposition. Moreover, we let $P = 1$ without loss of generality. In this case, we note that the set $\mathcal{A}(C)$ can be expressed in the equivalent form $\mathcal{A}(C) = \{\boldsymbol{\alpha}(t), |t| \leq 1\}$ where

$$|\alpha_1(t)|^2 = 2^{(1-t)C/2} - 1, \qquad |\alpha_2(t)|^2 = 2^{(1+t)C/2} - 1. \quad (22)$$

When $L = 1$, our construction specializes to a simple repetition code across subchannels, and serves as a useful baseline. Such a code can achieve rate $R_1 = \log\left(1 + |\alpha_1|^2 + |\alpha_2|^2\right)$, and hence yields an efficiency of

$$\eta_1 = (1/C) \log\left(1 + 2(2^{C/2} - 1)\right). \quad (23)$$

As we will see, this efficiency is significantly lower than that achievable by our construction.

In addition, a simpler version of our construction in which the decoding order is fixed (independent of the realized channel gains) incurs a significant price in performance. Indeed, such a scheme has an efficiency η'_L that is strictly less than unity (for $C > 0$) even as $L \to \infty$. In particular, we have the following claim, whose proof we omit due to space constraints.

Claim 1: For the case of $K = 2$ subchannels, the efficiency of the fixed-decoding-order variant of our code is bounded according to

$$\eta'_L(C) \leq \eta'_\infty = R_+(C)/C, \quad (24a)$$

where

$$R_+ = 2\log\left(1 + \tilde{P}|\tilde{\alpha}|^2\right) + \log\left(\frac{1 + |\tilde{\alpha}'|^2}{1 + \tilde{P}|\tilde{\alpha}'|^2}\right) \quad (24b)$$

with

$$|\tilde{\alpha}|^2 = 2^{C/2} - 1, \ |\tilde{\alpha}'|^2 = 2^C - 1, \ \tilde{P} = \frac{1}{|\tilde{\alpha}|^2} - \frac{2}{|\tilde{\alpha}'|^2}. \quad (24c)$$

It is straightforward to verify that η_1 and η'_∞ both approach 1 as $C \to 0$ and $1/2$ as $C \to \infty$. It can also be verified numerically that the η'_∞ is at most a gain of 13.8% in efficiency over η_1, which occurs at $C \approx 3.76$ b/s/Hz. Thus, without the use of a variable decoding order, our code construction cannot achieve large gains over a simple repetition code.

We evaluate the efficiency of our code *with* variable decoding order via a numerical evaluation of (20). In our simulations, both L and C are varied, but we continue to restrict our attention to $K = 2$ subchannels. The results are summarized in Table II. As the table reflects, efficiencies in the range of 90% are possible over typical target spectral efficiencies. The table also shows the loss in efficiency that is incurred when a fixed decoding order is used. As described above, the efficiency with fixed decoding order also implies a system with substantially greater complexity, since it requires the use of many layers.

In Table II, we also show the SNR gaps Δ (in dB) to capacity corresponding to the calculated efficiencies. Note that the SNR gap for each value of C depends on the realized channel gain pair (α_1, α_2). In the table, we indicate the worst-case SNR gap, which corresponds to, e.g., the case $\alpha_2 = 0$.

In summary, the efficiency improvements of our scheme relative to a simple repetition code vary from 23% (3.3 dB) at $C = 4.33$ b/s/Hz, to 29% (10.5 dB) at $C = 12$ b/s/Hz. Moreover, at the typical target spectral efficiency of 4.33 b/s/Hz, our code achieves within approximately 1 dB of capacity, neglecting losses due to the base code.

TABLE II

ACHIEVABLE EFFICIENCIES η_L AND η'_L, AND CORRESPONDING SNR GAPS Δ_L AND Δ'_L TO CAPACITY, FOR THE LAYERED-DITHER-REPEAT CODE WITH VARIABLE AND FIXED DECODING ORDERS RESPECTIVELY, AS A FUNCTION OF THE NUMBER OF LAYERS L AND THE TOTAL CHANNEL CAPACITY C (B/S/HZ) FOR THE CASE OF $K = 2$ SUBCHANNELS.

	$C = 4.33$	$C = 8$	$C = 12$
η_3 (Δ_3)	92% (1.1 dB)	90% (2.4 dB)	87% (4.7 dB)
η_2 (Δ_2)	88% (1.7 dB)	82% (4.4 dB)	77% (8.3 dB)
η_1 (Δ_1)	69% (4.4 dB)	62% (9.3 dB)	58% (15.2 dB)
η'_∞ (Δ'_∞)	83% (2.4 dB)	73% (6.6 dB)	66% (12.3 dB)

V. MIMO AND RATELESS EXTENSIONS

Both the asymptotically-perfect and near-perfect universal code constructions of this paper for the parallel Gaussian channel can be readily extended both to the Gaussian MIMO channel and to a rateless scenario, where even the capacity C is not known a priori. Codes for the MIMO channel can be constructed, for example, by combining the codes of this paper with the diagonal Bell Labs space-time codes (D-BLAST) via concatenation. When, in addition, we are interested in the rateless scenario, we need not only spatial redundancy blocks across channel inputs as in this paper, but temporal redundancy blocks as well. In particular, if we desire that, given some C, the code be decodable whenever the MIMO channel matrix $\mathbf{H}$ is such that $C_{\text{MIMO}}(\mathbf{H}) = C/m$ for some $1 \leq m \leq M$, where $C_{\text{MIMO}}(\mathbf{H})$ is the capacity of the associated (K-input) MIMO channel, then a total of KM redundancy blocks are required. Since the spatial and temporal dimensions are otherwise indistinguishable, it suffices to replace K with $K' = KM$ in the constructions of this paper. In the rateless extension of our concatenated code, this implies using a good rateless erasure code as the outer code, such as a raptor code [4]. Such an architecture therefore leverages the existance of good rateless codes for the erasure channel in the construction of good rateless codes for Gaussian channels. We omit the details due to space constraints.

REFERENCES

[1] S. Tavildar and P. Viswanath, "Approximately universal codes over slow fading channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3233–3258, July 2006.

[2] L. Zhao and S.-Y. Chung, "Perfect rate-compatible codes for the Gaussian channel," in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, Illinois, 2007.

[3] U. Erez, M. Trott, and G. Wornell, "Rateless coding for Gaussian channels," *IEEE Trans. Inform. Theory*, 2007, submitted for publication. Also available on arXiv; see http://arxiv.org/abs/0708.2575v1.

[4] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.