Rateless Codes for MIMO 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 W. Wornell
Dept. EECS, MIT
Cambridge, MA
Email: gww@mit.edu

***Abstract*— Two rateless code constructions are developed for efficient communication over multi-input multi-output (MIMO) Gaussian channels. The key ingredients in both architectures are layering, dithering, and repetition. Both employ successive cancellation decoding along with minimum mean-square error (MMSE) combining and convert the MIMO channel into a scalar channel to which classical Gaussian base codes can be applied. The first construction based on simple layering induces an additive white Gaussian noise (AWGN) scalar channel and with perfect base codes achieves a substantial efficiency gain over a baseline repetition scheme. At typical spectral efficiencies, this scheme can achieve better than 85% of capacity for a rateless construction with two effective rates. The second construction which uses a diagonal layering (DL) structure is capacity achieving at any SNR and induces a particular time-varying scalar channel. This holds even if the MIMO channel is block constant but time-varying.**

I. Introduction

The design of practical and efficient codes for communicating over multi-input multi-output (MIMO) Gaussian channels, when channel state information is not available a priori at the transmitter or in broadcast scenarios with multiple receivers, is of significant interest in a variety of emerging wireless applications and standards. These include multi-antenna systems as well as orthogonal frequency-division multiplexing (OFDM) systems.

For such problems, a rateless approach is rather natural, whereby an encoder maps a message into an infinite-length codeword for transmission. The decoder attempts to decode from successive prefixes of its received sequence. When it succeeds, it sends a single-bit acknowledgment (ACK) to the encoder to terminate the transmission. Rateless codes are hence instances of hybrid ARQ (HARQ) schemes.

Of particular interest are rateless codes that achieve rates close to capacity and require low decoding complexity. Examples of such codes for erasure channels are the well-known Raptor codes [1] developed from LT codes [2]. Such codes have also been applied to other discrete-input channels. By contrast, example rateless constructions for the scalar Gaussian channel are developed in [3], [4], and involve the transmission of sequences of redundancy blocks. For MIMO systems, however, the design of HARQ protocols has received comparatively little attention to date in the literature.

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 a NSERC scholarship.

An important sub-problem of the rateless design problem for MIMO channels is that of universal coding when capacity is known at the transmitter but channel parameters themselves are not. Example universal codes for the parallel Gaussian channels, which are of significant interest for example in OFDM technologies, have been developed in [5].

In this paper, we develop two rateless code architectures for the MIMO Gaussian channel that map the MIMO channel into an equivalent scalar Gaussian channel to which a standard base code can be applied. The ingredients used in these constructions are similar to those of the rateless codes of [4]. In both architectures, layered coding with successive cancellation decoding is used to develop efficient codes and redundancy is introduced using dithered repetition coding with minimum mean-square error (MMSE) combining.

The first construction, which is based on simple layering, was introduced for the case of universal coding for parallel channels in [5]. We extend this design formulation to the case of rateless coding for parallel channels and then combining it with the D-BLAST space-time codes [6] to MIMO channels. This code structure allows standard AWGN channel codes to be used efficiently. We show that even with small numbers of layers, numerically optimized versions of these rateless codes substantially reduce the gap to capacity.

The second construction, in contrast to the first construction and the approach in [4], satisfies the additional constraint of time-invariant encoding, which means that message recovery can start with any temporal redundancy block. This construction is based on a diagonal layering (DL) structure and was briefly introduced in [7]. In this paper we fully analyze this scheme and show that when used with suitable base codes, it is capacity-approaching with design parameters that are independent of the spectral efficiency of the channel (in contrast to the first construction) and hence easy to obtain. We additionally show that these codes are capacity approaching even if the MIMO channel is block-constant but time-varying.

In both architectures, we emphasize the finite signal-to-noise ratio (SNR) regime. As such, our approach differs from, e.g., that of [8] and [9], which focus on high SNR analysis.

II. Channel Model and Problem Formulation

The channel model of interest takes the form

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

where channel input $\mathbf{x}$ and output $\mathbf{y}$ at a particular time are N_t- and N_r-dimensional vectors, respectively, 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 $\bar{P}$. The transmission is arranged in redundancy blocks of N symbols, which is assumed to be large but otherwise plays no role in the analysis. The channel matrix realization, $\mathbf{H}$, is known to the receiver but not to the transmitter. Furthermore, our target capacity is the white input capacity

$$C_{\mathrm{MIMO}}(\mathbf{H}) = \log\det\left(\mathbf{I} + P\mathbf{H}\mathbf{H}^\dagger\right), \quad (2)$$

with $P = \bar{P}/N_t$.

The encoder generates from the message a series of incremental redundancy blocks and transmits them over space (the transmit antennas) and time. The receiver in turn collects as many redundancy blocks as needed to accumulate enough mutual information to decode the message. The number of blocks that need to be collected directly depends on $C_{\mathrm{MIMO}}(\mathbf{H})$. We use C^* to denote the *ceiling rate* of the code, i.e., a parameter specifying the highest rate at which the code can operate.

We require our rateless codes to be capacity-approaching in the following sense. For any channel matrix $\mathbf{H}$ that satisfies

$$C_{\mathrm{MIMO}}(\mathbf{H}) = C^*/m \qquad 1 \le m \le M, \quad (3)$$

for some integer m, we require the message to be decoded after sending a sequence of m sets of N_t redundancy blocks.We denote this set of channel matrices by $\mathcal{H}(C^*/m)$, i.e.,

$$\mathcal{H}(C^*/m) = \{\mathbf{H} : C_{\mathrm{MIMO}}(\mathbf{H}) = C^*/m\}. \quad (4)$$

Here M is the *range* of the code, i.e., the maximum number of temporal redundancy blocks the code is designed for.

Note that a parallel channel is a special case of the MIMO channel with $\mathbf{H} = \mathrm{diag}(\alpha_1, \ldots, \alpha_K)$ where K is the number of subchannels and where (2) becomes

$$C_{\|}(\mathbf{H}) = \sum_{k=1}^{K} \log\left(1 + P|\alpha_k|^2\right). \quad (5)$$

We denote the set of channel vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ with $C_{\|}(\mathrm{diag}(\boldsymbol{\alpha})) = C^*/m$ by $\mathcal{A}(C^*/m)$, i.e.,

$$\mathcal{A}(C^*/m) = \{\boldsymbol{\alpha} : C_{\|}(\mathrm{diag}(\boldsymbol{\alpha})) = C^*/m\}. \quad (6)$$

III. A Near-Perfect Rateless Construction

Here we develop the first class of rateless codes based on a layer-dither-repeat structure first for the case of parallel channels and then extend them to the MIMO channels.

A. Layer-Dither-Repeat Codes for Parallel Channels

To construct a rate R code, we first choose the number of layers L and the associated codebooks $\mathcal{C}_1, \ldots, \mathcal{C}_L$ that we assume are capacity achieving independent identically distributed Gaussian codebooks. We divide the message into L submessages that are independently coded using the corresponding codebooks. We further constrain the codebooks to have equal rate R/L since this has the advantage of allowing them to be derived from a single base code.

The redundancy blocks, $\mathbf{X}_1, \ldots, \mathbf{X}_M$, sent over K subchannels and in M redundancy intervals are constructed from the codewords $\mathbf{c}_l \in \mathcal{C}_l$, $l = 1, \ldots, L$ as

$$\begin{bmatrix}\mathbf{X}_1 & \cdots & \mathbf{X}_M\end{bmatrix}^T = \mathbf{G}\begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_L\end{bmatrix}^T \quad (7)$$

where $\mathbf{G}$ is a $KM \times L$ matrix of complex gains and where $\mathbf{X}_m$ for each m is a $K \times N$ matrix and $\mathbf{c}_l$ for each l is a row vector of length N with N the block length of the code. 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, we require the code to be successively decodable to keep the complexity low. 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. We use MMSE combining of the redundancy blocks during successive interference cancellation (SIC) which is information-lossless.

Using this construction results in several design degrees of freedom. At the encoder, we can choose $\mathbf{G}$, which has KML phase and $KM(L-1)$ magnitude degrees of freedom available. At the decoder, the decoding order can be chosen as a function of the realized channel parameters $\boldsymbol{\alpha}$. We denote this degree of freedom by a discrete variable n which could take $L!$ possible values corresponding to the $L!$ possible decoding orders. We can use these degrees of freedom to maximize the efficiency of the code for a given C^* which is equivalent to maximizing the rate of the code over these degrees of freedom as follows

$$R = L \cdot \max_{\mathbf{G}\in\mathcal{G}} \min_{1\le m\le M} \min_{\boldsymbol{\alpha}\in\mathcal{A}(C^*/m)} \max_{1\le n\le L!} \min_{1\le l\le L} I_l(\boldsymbol{\alpha}, \mathbf{G}, m, n) \quad (8)$$

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

$$\mathcal{G} = \{\mathbf{G} : [\mathbf{G}\mathbf{G}^\dagger]_{i,i} = P, \ i = 1, \ldots, KM\},$$

$\mathcal{A}(C^*/m)$ is as given in (6), and $I_l(\boldsymbol{\alpha}, \mathbf{G}, m, n)$ is the achievable rate in the lth layer with respect to the decoding order specified by n after collecting $\mathbf{X}_1, \ldots, \mathbf{X}_m$. The equivalent matrix channel seen after receiving these m sets of blocks is given by the block diagonal matrix

$$\mathbf{A}_m = \mathrm{diag}(\underbrace{\boldsymbol{\alpha}, \ldots, \boldsymbol{\alpha}}_{m}) \quad (9)$$

with $\boldsymbol{\alpha}$ repeated m times on the diagonal. For $l = 1, \ldots, L$

$$\sum_{l'=l}^{L} I_{l'}(\boldsymbol{\alpha}, \mathbf{G}, m, n) = \log\det\left(\mathbf{I} + \mathbf{A}_m[\mathbf{G}_m\boldsymbol{\Pi}(n)]_{l:L}[\mathbf{G}_m\boldsymbol{\Pi}(n)]_{l:L}^\dagger\mathbf{A}_m^\dagger\right), \quad (10)$$

where $\mathbf{G}_m$ is the submatrix consisting of the first Km rows of $\mathbf{G}$ assigned to the first m sets of K redundancy blocks, $\boldsymbol{\Pi}(n)$ denotes the matrix that permutes the columns of $\mathbf{G}_m$ according to n, and $[\cdot]_{i:j}$ denotes the submatrix consisting of columns

i through j of its argument. Finally, using (8), 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 remainder of this section, we consider the case of $K = 2$ subchannels for simplicity. Moreover, we let $P = 1$ without loss of generality. In this case, we can express the set $\mathcal{A}(C^*/m)$ in the equivalent form $\mathcal{A}(C^*/m) = \{\boldsymbol{\alpha}^{(m)}(t), |t| \leq 1\}$ where

$$|\alpha_1^{(m)}(t)|^2 = 2^{(1-t)C^*/(2m)} - 1, \tag{11a}$$

$$|\alpha_2^{(m)}(t)|^2 = 2^{(1+t)C^*/(2m)} - 1. \tag{11b}$$

When $L = 1$, our construction specializes to a simple repetition code across subchannels and in time, and serves as a useful baseline. Such a code can achieve a rate $R_1 = \log\left(1 + m|\alpha_1^{(m)}(t)|^2 + m|\alpha_2^{(m)}(t)|^2\right)$, after collecting m temporal sets of 2 blocks that has its minimum at $m = M$ and $t = 0$ and hence yields an efficiency of

$$\eta_1 = (1/C^*) \log\left(1 + 2M(2^{C^*/(2M)} - 1)\right). \tag{12}$$

We will show that our construction achieves an efficiency that is significantly higher than this repetition efficiency.

Moreover, fixing the decoding order independent of the realized channel significantly degrades the efficiency performance even in the case of universal coding, i.e., $M = 1$, as was demonstrated in [5]. Even for $M = 1$, 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.

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

$$\eta'_L(C^*) \leq \eta'_\infty = R_+(C^*)/C^*, \tag{13a}$$

where

$$R_+ = 2\log\left(1 + \tilde{P}|\tilde{\alpha}|^2\right) + \log\left(\frac{1 + |\tilde{\alpha}'|^2}{1 + \tilde{P}|\tilde{\alpha}'|^2}\right) \tag{13b}$$

with

$$|\tilde{\alpha}|^2 = 2^{C^*/2} - 1,\ |\tilde{\alpha}'|^2 = 2^{C^*} - 1,\ \tilde{P} = \frac{1}{|\tilde{\alpha}|^2} - \frac{2}{|\tilde{\alpha}'|^2}. \tag{13c}$$

Note that in the case of $M = 1$, η'_L could be made arbitrarily close to η'_∞ for large enough L. However, for $M > 1$, η'_∞ is only an upperbound on the performance of fixed-decoding-order variant of the scheme since the rateless performance of the code could at best be equal to its universal performance.

It can be verified numerically that for a code range of $M = 1$, the maximum efficiency gain of η'_∞ over the repetition efficiency η_1 is 13.8% that occurs at $C^* \approx 3.76$ b/s/Hz. Hence, without the use of a variable decoding order, our code does not offer large improvements over a simple repetition code.

We evaluate the efficiency of our code *with* variable decoding order by numerically evaluating (8). In our simulations, both L and C^* are varied, but we continue to restrict our attention to $K = 2$ subchannels. We perform the analysis for the universal case, i.e., $M = 1$, and also for the rateless case with $M = 2$. The results are shown in Table I.

As the table shows, for the universal design using only 3 layers, efficiencies in the range of 90% are possible over typical target spectral efficiencies. For the rateless design and using only 3 layers, efficiencies are in the range of 82% to 90% in typical spectral efficiencies and in all cases have a substantial gain of around 40% over the simple repetition scheme. Also the efficiency loss incurred going from a universal design to a rateless design with $M = 2$ is small and less than 5%.

In the table we also show the SNR gaps Δ_L (in dB) to capacity corresponding to the calculated efficiencies for different numbers of layers. 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 the rateless case for $M = 2$ we show the worst-case SNR gap for channels in both $\mathcal{A}(C^*)$ and $\mathcal{A}(C^*/2)$ denoted by $(\Delta_L^{(1)}, \Delta_L^{(2)})$.

The table also shows η'_∞ that corresponds to a lower bound on efficiency loss due to a fixed decoding order. Note that this fixed decoding order variant of our scheme requires many layers and hence imposes substantially greater complexity.

In summary, the efficiency improvements of our scheme relative to a simple repetition code vary from 33% (4.8 dB,3.2 dB) at $C^* = 4.33$ b/s/Hz, to 42% (15.3 dB,8.3 dB) at $C^* = 12$ b/s/Hz in the rateless case with $M = 2$, and from 23% (3.3 dB) at $C^* = 4.33$ b/s/Hz, to 29% (10.5 dB) at $C^* = 12$ b/s/Hz in the universal case. Moreover, at the typical target spectral efficiency of 4.33 b/s/Hz, our code achieves within approximately 1.4 dB of capacity in the rateless case and 1 dB of capacity in the universal case, neglecting losses due to the base code.

B. MIMO Extension

Codes for the MIMO channel can be constructed by concatenating the optimized layered codes for the parallel channels (outer code) with the D-BLAST architecture (inner code) and associated MMSE-SIC receiver [6] that maps a MIMO channel into a parallel channel as follows. The message is divided into U submessages that are independently coded using the layer-dither-repeat codes. These messages are then transmitted over the N_t transmit antennas according to the following encoding of length $N(N_\mathrm{t} + U - 1)$

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{x}_1^{N_\mathrm{t}} \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ \mathbf{x}_1^{N_\mathrm{t}-1} \\ \mathbf{x}_2^{N_\mathrm{t}} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{x}_1^1 \\ \vdots \\ \mathbf{x}_{N_\mathrm{t}-1}^{N_\mathrm{t}-1} \\ \mathbf{x}_{N_\mathrm{t}}^{N_\mathrm{t}} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{x}_U^1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $\mathbf{x}_u^n$ denotes the layer-dither-repeat encoding of the uth submessage sent over the nth channel input.

Since the D-BLAST architecture combined with MMSE-SIC decoding is capacity achieving, the overall concatenated code has the same efficiency characteristics as that for the parallel channel assuming $U \gg N_\mathrm{t}$ such that the efficiency loss due to zero-padding in the D-BLAST is negligible.

TABLE I

ACHIEVABLE EFFICIENCIES η_L AND THE CORRESPONDING SNR GAP Δ_L TO CAPACITY, FOR THE LAYERED-DITHER-REPEAT CODE WITH VARIABLE DECODING ORDER, AS A FUNCTION OF THE RANGE OF THE CODE M, NUMBER OF LAYERS L, AND CEILING RATE C^* (B/S/HZ) FOR THE CASE OF $K = 2$ SUBCHANNELS. TABLE ALSO SHOWS THE EFFICIENCY η'_∞ FOR THE FIXED-DECODING-ORDER VARIANT OF THE CODE IN THE LIMIT OF LARGE L AND THE GAP TO CAPACITY Δ'_∞ FOR $M = 1$.

$M = 1$	$C^* = 4.33$	$C^* = 8$	$C^* = 12$
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)
$M = 2$			
η_3	90%	86%	82%
$\left(\Delta_3^{(1)}, \Delta_3^{(2)}\right)$ dB	(1.4, 0.9)	(3.4, 1.8)	(6.5, 3.3)
η_2	78%	70%	65%
$\left(\Delta_2^{(1)}, \Delta_2^{(2)}\right)$ dB	(3.08, 1.9)	(7.3, 4.0)	(12.6, 6.5)
η_1	57%	46%	40%
$\left(\Delta_1^{(1)}, \Delta_1^{(2)}\right)$ dB	(6.2, 4.1)	(13.3, 7.6)	(21.8, 11.6)
η'_∞ (Δ'_∞)	83% (2.4 dB)	73% (6.6 dB)	66% (12.3 dB)

IV. DIAGONALLY LAYERED RATELESS CODING

The layered codes developed in Section III achieve substantial gain over the baseline repetition scheme with few layers. In this section we develop and fully analyze a class of rateless codes constructed via a diagonal layering approach briefly introduced in [7] that have the following differences with the first construction: First they are time-invariant. Second, they can be directly applied to MIMO channels without the need to be concatenated with the D-BLAST. Third, they are capacity achieving when used with suitable base codes (as opposed to near-perfect) at any SNR, with design parameters that are independent of the spectral efficiency, at the price of the induced scalar channel being a particular time-varying one. Moreover this even holds for block constant but time-varying MIMO channels.

We now develop this DL construction.

A. DL Coding for Scalar Channels

With DL coding, the message is divided into LJ submessages, each of which is independently coded into a corresponding subcodeword. These subcodewords are then superimposed in a staggered (diagonal) manner, where the offset with respect to the previous subcodeword is fixed and determined by the choice of L. The resulting equivalent layered structure of the encoding is as depicted in Fig. 1, where we see that there are J subcodewords per layer, L layers, and L segments per subcodeword that see interference from different sets of subcodewords.

Such an encoding can be decoded using successive cancellation. In particular, as shown in Fig. 1, we decode starting from the lower left in a cyclic manner, subtracting the effect

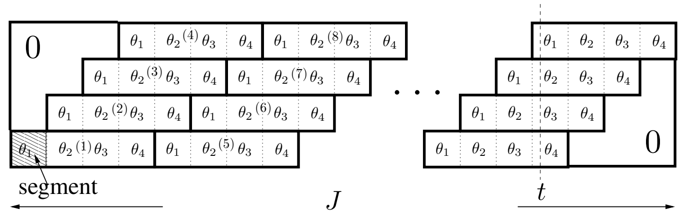

Fig. 1. Encoding structure for diagonally layered communication. In this example, $L = 4$. Each "brick" represents a subcodeword, and its associated parenthesized index indicates the order in which decoding proceeds. The phases θ_l applied within the subcodewords are parameters exploited to develop a rateless code from DL encoding.

of subcodewords as they are decoded, and treating undecoded layers as noise. With this approach, the effective channel for all subcodewords are identical. Thus, using perfect base codes, the resulting DL encoding is capacity achieving when the same rate and power is used for each subcodeword (provided J is chosen large enough that the efficiency due to zero-padding in the construction is negligible). This symmetry makes the DL architecture well-suited for the development of our rateless codes.

B. DL-based Rateless Coding for MIMO Channels

A rateless code for the MIMO channel is constructed by creating a collection of redundancy blocks in space and time that are distinct linear transformed versions of the basic DL-coded block. In particular, as Fig. 1 reflects, prior to superposition in the DL encoding, the lth segment in each subcodeword is multiplied by $e^{j\theta_l}$, where θ_l is a phase parameter. These phases must be chosen for each redundancy block in the rateless code.

The redundancy blocks at an arbitrary time t (see Fig. 1) can be expressed in the form

$$\begin{bmatrix} \mathsf{x}_1 \\ \vdots \\ \mathsf{x}_M \end{bmatrix} = \sqrt{\frac{P}{L}}\mathbf{Q} \begin{bmatrix} c_1 \\ \vdots \\ c_L \end{bmatrix} \tag{14}$$

where c_l is an arbitrary element of a (normalized) subcodeword in layer l, x_m is the N_t-dimensional channel input during incremental redundancy interval m, and $\mathbf{Q}$ is a $N_\mathrm{t}M \times L$ matrix whose (i, l)th component $e^{j\theta i,l}$ describes the phase for the segment at time t in the lth layer of the redundancy block indexed by i. Note that we can write the redundancy blocks at any other time keeping $\mathbf{Q}$ the same but using the appropriate cyclic shift of $c_1, \dots, c_L$ as shown in Fig. 1.

The receiver collects as many spatial and/or temporal redundancy blocks as it needs to accumulate enough mutual information to recover the message. In our rateless setup we require the message to be decoded for a channel in $\mathcal{H}(C^*/m)$ after collecting m sets of N_t redundancy blocks. The receiver combines and decodes these redundancy blocks using the information-lossless MMSE-SIC front end, creating an equivalent scalar channel. Similar to (9), for a channel realization $\mathbf{H}$, the equivalent channel matrix after receiving m

temporal sets of blocks is given by the block diagonal matrix,

$$\mathbf{H}_m = \text{diag}(\underbrace{\mathbf{H}, \cdots, \mathbf{H}}_{m}). \quad (15)$$

The choice of $\mathbf{Q}$ determines the performance of the rateless code. We prove the following proposition.

Proposition 2: Provided perfect base codes of rate C^*/L are used, the DL-based rateless coding architecture is perfect, i.e., capacity achieving for any channel realization $\mathbf{H}$ such that $C_{\text{MIMO}}(\mathbf{H}) = C^*/m$ for every $m \in \{1, \ldots, M\}$ if and only if $L \geq N_{\text{t}}M$ and $\mathbf{Q}/\sqrt{L}$ has orthonormal rows.

Proof: To see that the two conditions are necessary, note that to achieve the white input capacity $C_{\text{MIMO}}(\mathbf{H})$ we need the transmitted symbols on the different transmit antennas and temporal blocks to be i.i.d. Gaussian, which means that the N_tM rows of the $\mathbf{G}$ matrix must be orthogonal. Hence $\mathbf{Q}$ must at least have N_tM columns.

To prove the sufficiency of these conditions note that the achievable rate of the code after collecting m redundancy sets is given by the mutual information as

$$\begin{aligned} R &= I(c_1, \ldots, c_L; \mathbf{y}_1, \ldots, \mathbf{y}_m) \\ &= \log\det\left(\mathbf{I} + \frac{P}{L}\mathbf{H}_m\mathbf{Q}_m\mathbf{Q}_m^\dagger\mathbf{H}_m^\dagger\right), \end{aligned} \quad (16)$$

with $\mathbf{y}_m$ the channel output corresponding to $\mathbf{x}_m$ and $\mathbf{Q}_m$ the submatrix containing the first N_tm rows of $\mathbf{Q}$. Having $L \geq N_tM$ and that the rows of $\mathbf{Q}/\sqrt{L}$ are orthonormal, the argument in (16) reduces to

$$\frac{1}{L}\mathbf{H}_m\mathbf{Q}_m\mathbf{Q}_m^\dagger\mathbf{H}_m^\dagger = \mathbf{H}_m\mathbf{H}_m^\dagger = \text{diag}\,(\underbrace{\mathbf{H}\mathbf{H}^\dagger, \ldots, \mathbf{H}\mathbf{H}^\dagger}_{m})$$

Substituting back in (16), the achievable rate of the code is given by

$$R = \log\det\left(\mathbf{I} + P\mathbf{H}_m\mathbf{H}_m^\dagger\right) = m\log\det\left(\mathbf{I} + P\mathbf{H}\mathbf{H}^\dagger\right) = C^*$$

where the last equality follows since $\mathbf{H} \in \mathcal{H}(C^*/m)$. ■

The constraints of Proposition 2 are straightforward to satisfy. For example, it is sufficient to take $\mathbf{G}$ to be the $N_{\text{t}}M \times N_{\text{t}}M$ discrete Fourier transform (DFT) matrix, or, when it exists, the Hadamard matrix. Having this, assuming perfect base codes, we can pick the base code rate to be $R = C^*/L$.

C. Block Constant Time-Varying Channels

The DL-based rateless code is capacity achieving even for block constant but time-varying MIMO channels. This means that with this construction, the message could be decoded after collecting m redundancy blocks as long as the sequence of realized channels satisfy

$$\sum_{i=1}^{m} \log\det\left(\mathbf{I} + P\mathbf{H}(i)\mathbf{H}(i)^\dagger\right) = C^* \quad (17)$$

for any $m = 1, \ldots, M$ where $\mathbf{H}(i)$ is the channel realization in the ith temporal redundancy block and M is the code range.

The proof is exactly the same as the proof of Proposition 2 only replacing $\mathbf{H}_m$ in (15) with $\mathbf{H}_m = \text{diag}\,(\mathbf{H}(1), \ldots, \mathbf{H}(m))$.

D. Base Code Design

The symmetric structure of the DL-based code makes it suitable for rateless coding. However, it has the disadvantage of inducing an effective scalar channel in each layer which is time-varying. In particular, this effective channel has a fixed SINR in each of the L segments within a subcodeword that changes from segment to segment. In general, the base code and its decoding structure should be designed having this time-variation in mind. However, simulation results in [7], [10], [11] suggest that the performance of standard LDPC base codes in these low rate induced channels is not degraded by much.

E. Comparison to a Baseline Alamouti Scheme

As a baseline for comparison we consider a rateless code using standard Alamouti code [12] with redundancy blocks generated by repetition. We can show that the efficiency of this scheme on the 2×2 MIMO channel is given by

$$\eta_{\text{ala-rep}} = (1/C^*)\log\left(1 + 2M(2^{C^*/(2M)} - 1)\right)$$

For example, on the 2×2 channel with spectral efficiency of 2.4 b/s/Hz and assuming perfect base codes, a DL-based code consisting of 2 sets of 2 redundancy blocks and $L = 4$ layers has an efficiency gain of 33% over the baseline that has an efficiency of 67%. Moreover, the base code rate in the DL code is lower than that in the baseline that requires a good base code of rate $2.4 \times 0.67/2 = 0.8$ b/1D, the (non-binary) design of which is more challenging. Thus as a practical MIMO rateless code, the DL-based construction shows considerable promise.

References

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

[2] M. Luby, "LT-codes," in *Proc. IEEE Symp. Found. Comp. Sci. (FOCS)*, Vancouver, BC, Canada, 2002.

[3] U. Erez, G. W. Wornell, and M. D. Trott, "Faster-than-Nyquist coding," in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, Illinois, 2004.

[4] 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.

[5] M. M. Shanechi, U. Erez, and G. W. Wornell, "On universal coding for parallel Gaussian channels," in *Proc. Int. Zurich Seminar Commun.*, Zurich, Switzerland, Mar 2008.

[6] G. J. Foschini, "Layered space-time architectures for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[7] M. M. Shanechi, U. Erez, and G. W. Wornell, "Time-invariant rateless codes for MIMO channels," in *IEEE Intern. Symp. Inform. Theory*, Toronto, Canada, July 2008, pp. 2247–2251.

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

[9] 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.

[10] C. R. Jones, J. V. T. Tian, and R. D. Wesel, "The universal operation of LDPC codes over scalar fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 122–132, Jan 2007.

[11] C. Jones, T. Tian, A. Matache, R. D. Wesel, and J. Villasenor, "Robustness of LDPC codes on periodic fading channels," in *Proc. GLOBECOM*, Taipei, Taiwan, Nov 2002, pp. 1282–1284.

[12] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct 1998.