Rateless Coding for MIMO Fading Channels: Performance Limits and Code Construction

Yijia (Richard) Fan, Lifeng Lai, Elza Erkip, and H. Vincent Poor

***Abstract*—In this letter the performance limits and design principles of rateless codes over fading channels are studied. The diversity-multiplexing tradeoff (DMT) is used to analyze the system performance for all possible transmission rates. It is revealed from the analysis that the design of such rateless codes follows the design principle of approximately universal codes for multiple-input multiple-output (MIMO) channels. It is also shown that for a single-input single-output (SISO) channel, simple permutation codes of unit length for parallel channels can be transformed directly into rateless codes that achieve the DMT performance limit of the channel.**

***Index Terms*—MIMO, block fading channels, rateless codes.**

I. Introduction

Rateless codes present a class of codes that can be truncated to a finite number of lengths, each of which has a certain likelihood of being decoded to recover the entire message. Compared with conventional coding schemes having a single rate R, such codes can achieve multiple rate levels $(R, RL/(L-1), RL/(L-2), \ldots, LR)$, where L is the number of blocks over which the rateless codeword is sent. The actual rate achieved depends on the channel conditions. A rateless code is said to be *perfect* if each part of its codeword is capacity achieving. Compared with conventional codes, rateless codes offer a potentially *higher rate*. Several results have been obtained on the design of perfect rateless codes over erasure channels and additive white Gaussian noise (AWGN) channels (see [6] and the references therein).

Unlike in the fixed channel scenario, non-zero error probability always exists in fading channels, when the instantaneous channel state information (CSI) is not available at the transmitter and a codeword spans only one or a small number of fading blocks. In this scenario, it is well known that there is a fundamental tradeoff between the information rate and error probability over fading channels, which can be characterized as the diversity-multiplexing tradeoff (DMT) [1].

Definition 1 (DMT): Consider a multiple-input multiple-output (MIMO) system and a family of codes C_η operating at average SNR η per receive antenna and having rates R. The multiplexing gain and diversity order are defined respectively as

$$r \triangleq \lim_{\eta\to\infty} \frac{R}{\log_2 \eta} \quad \text{and} \quad d \triangleq -\lim_{\eta\to\infty} \frac{\log_2 P_e(R)}{\log_2 \eta}, \qquad (1)$$

where $P_e(R)$ is the average error probability at the transmission rate R.

In this letter, we analyze the DMT performance of rateless codes. The results show that, compared with conventional coding schemes having multiplexing gain r, rateless codes having multiple multiplexing gains $(r, rL/(L-1), rL/(L-2), \ldots, Lr)$ offer an *effective* multiplexing gain $r_e = Lr$, for the same diversity gain level, when r is *small*. As r increases, the performance of rateless codes degrades and ultimately becomes the same as that of a conventional scheme. Also while increasing L lifts the overall system DMT curve, it does not necessarily improve the system multiplexing gain for every fixed value of r. It is then revealed that the design of such rateless codes follows the principle of codes that are *approximately universal* [3] over fading channels. Thus approximately universal space time codes for MIMO channels can be directly transformed into rateless codes that achieve the optimal DMT of the channel. It is also shown that for a single-input single-output (SISO) channel, the simple unit length permutation codes for parallel channels [3] can be transformed directly into rateless codes of length L to achieve the DMT performance limit of the channel.

The performance of rateless coding over fading channels has also been considered in [4] and [5], in which the throughput and error probability are discussed. However, the tradeoff between these two was not analyzed explicitly. For example, the results in [4] show that increasing the value of L will decrease the system error probability in certain scenarios and is therefore desirable. In this letter we show that while this discovery is true, the system throughput, i.e., the multiplexing gain, might decrease when L becomes larger for every fixed value of r. Overall, our results reveal that the optimal design of rateless codes requires the consideration of both r and L.

Rateless coding may be considered a type of Hybrid-Automatic Repeat reQuest (ARQ) scheme [2]. The DMT for ARQ has been derived in [2]. However, it will be shown in the paper that this DMT curve represents the performance of rateless codes only when $r < \min(M, N)/L$ in which M and N are the number of transmit and receive antennas. The *complete* DMT curve for rateless coding, including those parts for higher values of r, has never been revealed before, and will be given in this paper. In addition to this, the results in this letter also offer a relationship between the design parameters (i.e., r and L) and the effective multiplexing gain r of the system, thus providing further insights into system design and operational meaning compared to conventional coding schemes. Furthermore, we construct new rateless codes specifically for fading channels.

Y. Fan and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544, USA (e-mail: {yijiafan, poor}@princeton.edu).

L. Lai is with the Department of Systems Engineering, University of Arkansas at Little Rock, Little Rock, AR, 72204, USA (e-mail: lxlai@ualr.edu).

E. Erkip is with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY, 11201, USA (e-mail: elza@poly.edu).

This research was supported in part by the U.S. National Science Foundation under Grants CNS-09-05398.

This work was presented in part at the the 2008 IEEE International Symposium on Information Theory, Toronto, Canada, July 2008.

Digital Object Identifier 10.1109/TWC.2010.04.090780

The rest of this paper is organized as follows. The system model is described in Section II. In Section III, the DMT performance of rateless codes is studied. In Section IV, design of specific rateless codes over fading channels is discussed. Finally, concluding remarks are given in Section V.

II. SYSTEM MODEL

We consider a frequency-flat fading channel with M transmit antennas and N receive antennas. We assume that the transmitter does not know the instantaneous CSI on its corresponding forward channel, while CSI is available at the receiver. Each message is encoded into a codeword of L blocks. Each block takes T channel uses. We assume that the channel remains static for the entire codeword length (i.e., up to L blocks)[1]. The system input-output relationship can be expressed as

$$\mathbf{Y} = \sqrt{\frac{P}{M}}\mathbf{H}\mathbf{X} + \mathbf{N} \tag{2}$$

where $\mathbf{X} \in \mathbb{C}^{M\times TL}$ is the input signal matrix; $\mathbf{H} \in \mathbb{C}^{N\times M}$ is the channel transfer matrix whose elements are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero means and unit variances; $\mathbf{N} \in \mathbb{C}^{N\times TL}$ is a noise matrix with i.i.d. zero mean and unit variance complex Gaussian elements; and $\mathbf{Y} \in \mathbb{C}^{N\times TL}$ is the output signal matrix. P is the total transmit power, which also corresponds to the average SNR η (per receive antenna) at the receiver.

The input signal matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_L \end{array}\right] \tag{3}$$

where $\mathbf{X}_l \in \mathbb{C}^{M\times T}$ is the codeword matrix being sent during the lth block, and its corresponding receiver noise matrix is denoted by $\mathbf{N}_l \in \mathbb{C}^{N\times T}$. We impose a power constraint on each $\mathbf{X}_l$ so that[2]

$$E\left[\frac{1}{T}\left\|\mathbf{X}_l\right\|_F^2\right] \leqslant M, \tag{4}$$

for $l = 1, ..., L$.

A. Conventional Schemes

Assume that the transmitter sends the codeword at a rate R bits per channel use. L messages, each of which is of size RT, are encoded into codewords $\mathbf{X}_l$ ($l = 1, \ldots, L$) and transmitted in T channel uses. An alternative method is to encode a message of size RLT into $\mathbf{X}$. Both encoding methods will offer the same performance provided that T is sufficiently large.

B. Rateless Coding

When rateless coding is applied, we wish to decode a message of size RLT with the codeword structure as shown in (3). During the transmission, the receiver measures the total mutual information I between the transmitter and the receiver and compares it with RLT after it receives each codeword block $\mathbf{X}_l$. Note that because of fading I is in fact a random variable at each decoding block. If $I < RLT$ after the lth block, the receiver remains silent and waits for the next block. If $I \geq RLT$ after the lth block, it decodes the received codeword $\left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_l \end{array}\right]$ and sends one bit of positive feedback to the transmitter[3]. Upon receiving the feedback, the transmitter stops transmitting the remaining part of the current codeword and starts transmitting the next message immediately.

Unlike conventional schemes, this process will bring multiple rate levels $(R, RL/(L-1), RL/(L-2), \ldots, LR)$. For example, if $I \geq RLT$ after the first block is received (i.e., $l = 1$) , the receiver will be able to decode the entire message and the rate becomes LR. Similar observations can be made for $l = 2, \ldots, L$. Therefore, compared with conventional schemes, the corresponding transmission rate achieved by using rateless codes is always *equal or higher*. Specifically, we define the multiplexing gains for the rate levels to be $(r, rL/(L-1), rL/(L-2), \ldots, Lr)$ where

$$r \triangleq \lim_{\eta\to\infty} \frac{R}{\log_2 \eta}.$$

Later we will show through the DMT analysis that rateless coding can retain the same diversity gain as conventional schemes, but with a much higher multiplexing gain especially when the corresponding r is small.

III. PERFORMANCE ANALYSIS

Denote by ε_l the decoding error when decoding is performed at the end of the lth block ($0 \leq l \leq L$) and by $\Pr(\varepsilon_l, l)$ the joint probability that decoding is performed at the end of lth block and a decoding error occurs. The overall error probability can be expressed as

$$P_e = \sum_{l=1}^{L} \Pr(\varepsilon_l, l).$$

Define $p(l)$ ($0 \leq l \leq L$) to be the probability with which $I < RLT$ after the lth block, and note that $p(0) = 1$. We define rate $\bar{R}$ (in bits per channel use) as the long term average transmission rate for each message, i.e., the expected value of realized rate for each transmitted rateless codeword. Following the steps in Section II.B in [2], $\bar{R}$ is given by

$$\bar{R} = \frac{RL}{\sum\limits_{l=0}^{L-1} p(l)}. \tag{5}$$

Note that this $\bar{R}$ describes the average rate with which the message is removed from the *transmitter*; i.e., it quantifies how quickly the message is decoded at the receiver. We define the effective multiplexing gain of the system as

$$r_e = \lim_{\eta\to+\infty} \frac{\bar{R}}{\log_2 \eta}.$$

Define $f(k)$ to be the piecewise linear function connecting the points $(k, (M-k)(N-k))$ for integral $k = 0, ..., \min(M, N)$. Recall that a conventional scheme operating

[1]Note, however, that the analysis in this paper can be extended straightforwardly to a faster fading scenario in which the channel varies from block to block during each codeword transmission.

[2]Note that this is a stricter constraint than letting $E\left[\frac{1}{TL}\left\|\mathbf{X}\right\|_F^2\right] \leqslant M$, which offers at least the same performance.

[3]Note that the feedback is sent to the transmitter even when the codeword is decoded in error, though the probability of this event can be made arbitrarily small.

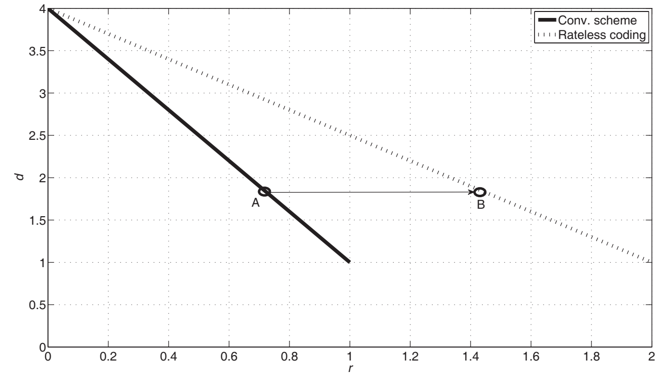

Fig. 1. The DMTs for conventional schemes and rateless coding for $0 \leq r \leq 1$. $M = N = 2$, $L = 2$.

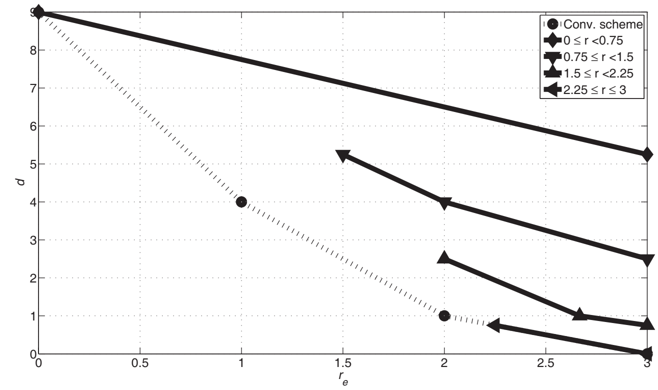

Fig. 2. The DMTs for different schemes for $0 \leq r \leq 3$. $M = N = 3$, $L = 4$.

at multiplexing gain r $(0 \leq r \leq \min(M,N))$ would have the diversity gain $f(r)$. The following theorem gives the performance of rateless coding for $0 \leq r < +\infty$.

Theorem 1: For rateless codes having multiple multiplexing gain levels $(r, rL/(L-1), rL/(L-2), \ldots, Lr)$, and for sufficiently large T, the corresponding DMT can be expressed as (r_e, d) where

$$r_e = r \cdot \frac{L}{l} \quad \text{and} \quad d = f\left(\frac{lr_e}{L}\right)$$

for

$$\frac{l-1}{L} \min(M,N) \leqslant r < \frac{l}{L} \min(M,N)$$

and $l = 1, 2, ..., L$. Finally, $d = 0$ for $r \geq \min(M,N)$.

Proof: See Appendix A. ■

Note that for rateless coding to achieve the performance in *Theorem 1*, we do not necessarily require $T \to \infty$. As long as T is large enough such that the error probability $\Pr(\varepsilon_l, l) \dot{\leqslant} \eta^{f(r)}$ for each l, the DMT in *Theorem 1* can be achieved. For example, it will be shown later that for SISO channels, $T = 1$ is sufficient to achieve the optimal DMT in *Theorem 1*.

Comparing rateless coding with conventional schemes, it can be shown that for $0 \leq r < \min(M,N)/L$, $r_e = Lr$ for $d = f(r)$. In this scenario rateless coding can improve the multiplexing gain up to L times that of conventional schemes, given the same diversity gain. Fig. 1 gives an example when $M = N = 2$ and $L = 2$, and $0 \leq r \leq 1$. The operating point A on the curve for a conventional scheme for $0 \leq r \leq 1$ corresponds to point B on the curve for rateless coding.

An important observation from *Theorem 1* is that the system DMT will not be improved after r (*almost*) reaches $\min(M,N)/L$. This is mainly due to the fact that the first block can no longer support the message size when the message rate reaches $\min(M,N)/L$. Thus the system multiplexing gain *decreases* for the same diversity gain, and finally offers the same DMT as conventional schemes when the first $L-1$ blocks all fail to decode the message. Fig. 2 shows an example when $M = N = 3$ and $L = 4$. This observation also implies that for any fixed value of r, simply increasing the value of L does *not necessarily* improve the system DMT. Although the optimal system DMT (for small values of r) will increase when L is larger, the multiplexing gain might decrease for certain larger values of r. A convenient choice for L would be in the range $L < \min(M,N)/r$. However, note that the maximal multiplexing gain $\min(M,N)$ can be achieved only with zero diversity gain, and this happens when $r = \min(M,N)$ *regardless of* the value of L.

IV. Design of Rateless Codes

Note that codewords $\mathbf{X}_l$ $(1 \leq l \leq L)$ in (3) are transmitted through different channels that are *orthogonal* in time. This is analogous to transmitting $\mathbf{X}_l$ through different channels that are parallel in *space*. In the (space) parallel channel model, elements in $\{\mathbf{X}_l\}$ can be jointly (simultaneously) decoded. However, for the channel model considered in this paper, which we now call the *rateless channel*, the decoding process needs to follow a certain direction in time, i.e., we start decoding from $\mathbf{X}_1$, then $[\mathbf{X}_1 \;\; \mathbf{X}_2]$ if $\mathbf{X}_1$ is not decoded, etc. This comparison implies that while good parallel channel codes can be used as the basis for rateless coding, they might need modifications in order to offer good performance over the rateless channel.

Specifically, for the rateless channel expressed in the form of (2), we consider the corresponding parallel MIMO channel, in which each sub-channel is a MIMO channel, having the following input-output relationship:

$$\mathbf{Y} = \sqrt{\frac{P}{M}} \begin{pmatrix} \mathbf{H} & & 0 \\ & \ddots & \\ 0 & & \mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_L \end{pmatrix} + \begin{pmatrix} \mathbf{N}_1 \\ \vdots \\ \mathbf{N}_L \end{pmatrix} \tag{6}$$

where $\mathbf{H}$, $\mathbf{X}_l$ and $\mathbf{N}_l$ are the same as in (2). It is easy to see that the DMT for this system is $d = f\left(\frac{r}{L}\right)$ for $0 \leq r \leq L\min(M,N)$. Assuming a code that achieves this DMT, when we implement its transformation $\left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_L \end{array}\right]$ into the rateless channel having multiple rates $(r, 2r, \ldots, Lr)$, it is not difficult to show that

$$\Pr(\varepsilon_L, L) \dot{\leqslant} \eta^{-f(r)}. \tag{7}$$

In order to make the overall $P_e \dot{\leqslant} \eta^{-f(r)}$, we need to ensure that $\Pr(\varepsilon_l, l) \dot{\leqslant} \eta^{-f(r)}$ for $1 \leq l \leq L-1$. However, those conditions are not *essential* in order to achieve the optimal DMT for the parallel channel shown in (6), which requires only the condition (7). Thus stricter code design criteria are required for the rateless channel. One example of such a criterion is the *approximately universal* criterion [3].

Codes being *approximately universal* for any channel ensures that the highest error probability when decoding *any* subset of $\{\mathbf{X}_l\}$ in the set of all non-outage events decays *exponentially* in SNR (i.e., in the form of $e^{-\eta^{\delta}}$ for some $\delta > 0$) under *any* fading distribution, and thus can be ignored compared with the outage probability under the same fading distribution when the SNR goes to infinity. It has been shown that being approximately universal is sufficient for a scheme or code to achieve the DMT of the channel [3]. Now, we consider the following MIMO channel with LM transmit antennas and LN receive antennas:

$$\mathcal{Y} = \sqrt{\frac{P}{M}}\mathcal{H}\mathcal{X} + \mathcal{N} \tag{8}$$

where $\mathcal{X} = \left[\begin{array}{ccc} \mathbf{X}_1^T & \cdots & \mathbf{X}_L^T \end{array}\right]^T$, $\mathcal{N} = \left[\begin{array}{ccc} \mathbf{N}_1^T & \cdots & \mathbf{N}_L^T \end{array}\right]^T$ and $\mathcal{H}$ is the $LN \times LM$ channel matrix in which each element follows an arbitrary distribution. We have the following result regarding the construction of rateless codes.

Theorem 2: Suppose a code $\mathcal{X} = \left[\begin{array}{ccc} \mathbf{X}_1^T & \cdots & \mathbf{X}_L^T \end{array}\right]^T$ is *approximately universal* for the channel with LM transmit antennas and LN receive antennas shown in (8). Then, its transformation $\bar{\mathcal{X}} = \left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_L \end{array}\right]$, when applied to the rateless channel with M transmit antennas and N receive antennas shown in (2) aiming at multiple multiplexing gains $(r, rL/(L-1), \ldots, Lr)$, can achieve the DMT given in *Theorem 1*.

Proof: See Appendix B. ■

Because of *Theorem 2*, existing approximately universal space time codes (see Section VII.A of [3] for an overview) can be applied directly to rateless channels. Examples of such codes include cyclic division algebra codes [9] with block lengths $T = ML$.

A. Simple codes for SISO channel

The decoding algorithms for universal codes for MIMO channels are usually very complex. So are the decoding algorithms for the corresponding rateless codes. In the following, we propose a family of codes for SISO channels that have a particularly simple coding algorithm. They are transformed from permutation codes designed for parallel channels [3].

Permutation codes are a class of codes generated from quadrature amplitude modulation (QAM) constellations. In the encoding process, a message is mapped into different QAM constellation points across all subchannels. The constellation points over one subchannel is a permutation of the points in the constellation over any other subchannel. The permutation is optimized such that the minimal codeword difference is large enough to satisfy the approximate universality criterion. Explicit permutation codes can be constructed using *universally decodable matrices*. We refer the readers to [3] and the references therein for details. It has been shown that permutation codes are approximately universal for parallel channels and have a particularly simple structure. For example, the codewords are of *unit* length.

Assume the transmission rates over a rateless channel are $(R, RL/(L-1), \ldots, LR)$ bits per channel use. To implement permutation codes, we choose a codebook of size 2^{LR} (messages) for the parallel channel in (6). Each message is mapped into a code $\left[\begin{array}{ccc} \mathbf{X}_1^T & \cdots & \mathbf{X}_L^T \end{array}\right]^T$, in which each $\mathbf{X}_l$ is drawn from a 2^{LR}-point QAM constellation. The message

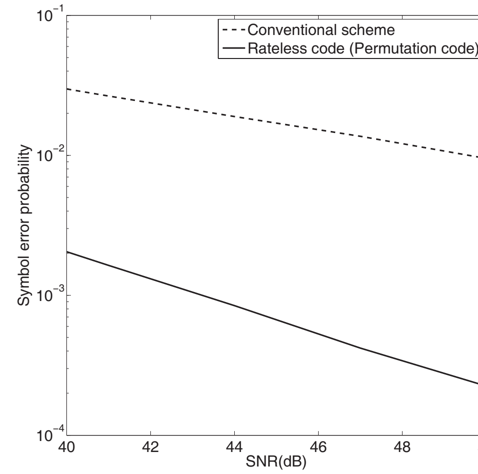

Fig. 3. The error probability performance over a SISO channel for (a) the conventional scheme, in which each message is mapped into a 16-QAM point, and (b) a rateless permutation code generated from 16-QAM constellations with $L = 2$.

can be fully recovered as long as any subset of $\{\mathbf{X}_l\}$ can be correctly decoded. Now, we transform this code into the form $\left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_L \end{array}\right]$ for the rateless channel. Since $\Pr(\varepsilon_l, l)$ decays exponentially in SNR due to the approximate universality of such codes, the overall error probability is always dominated by the decoding error probability upon receiving all $\mathbf{X}_l$, when the SNR tends to *infinity*. We summarize the above observations as the following corollary.

Corollary 1: Rateless codes that are transformed from permutation codes for parallel channels can offer exactly the same performance as that shown in *Theorem 1* over the SISO rateless channel.

Proof: The proof is a direct extension of the proof of *Theorem 2* and is omitted. ■

Fig. 3 shows the error probability performance over a SISO channel for (a) a conventional scheme, in which each message is mapped into a 16-QAM point (4 bits per channel use), and (b) rateless permutation code generated from 16-QAM constellations with $L = 2$. The SNR range (40-50dB) approximately corresponds to a multiplexing gain range of 0.8-1. Clearly the rateless coding offers a higher diversity gain than the conventional scheme, while they have (nearly) the same rate at high SNR [4].

V. CONCLUSIONS

The performance limits of rateless codes have been studied for MIMO fading channels in terms of the DMT. The analysis shows that design principles for rateless codes can follow those of the approximately universal codes for MIMO channels. Simple rateless codes that are DMT optimal for a SISO channel have also been constructed.

APPENDIX

A. Proof of Theorem 1

Define $r_L = Lr$. Following the steps in [1], it is easy to show that $p(l) \doteq \eta^{-f\left(\frac{r_L}{l}\right)}$ for $l \neq 0$. We write the error

[4]Note that the diversity gain for a multiplexing gain in the error probablity curve cannot be exactly matched to that in the DMT curve due to the finite SNR values[10].

probability as

$$P_e = \sum_{l=1}^{L-1} (1 - p(l)) \Pr(\varepsilon_l) + \Pr(\varepsilon_L, L). \tag{9}$$

In (9), $\Pr(\varepsilon_l)$ is the error probability when $lI_b \geq LTR$, where I_b is the mutual information of the channel in each block. Using Fano's inequality we can obtain the error probability lower bound [1]:

$$P_e \geq \Pr(\varepsilon_L, L) \dot{\geqslant} \eta^{-f\left(\frac{r_L}{L}\right)}.$$

Since $r_e \leq r_L$, we have $\eta^{-f\left(\frac{r_L}{L}\right)} \geq \eta^{-f\left(\frac{r_e}{L}\right)}$, and thus the desired performance upper bound is obtained.

Now we prove the achievability part. Consider $\Pr(\varepsilon_l)$. Following the same argument as in the proof of Theorem 10.1.1 in [8], we obtain

$$\Pr(\varepsilon_l) \leqslant 3\epsilon \tag{10}$$

for sufficiently large T. Note that a very similar argument has been made in *Lemma 1* in [7], although it is claimed there that both T and L are required to be sufficiently large in order to satisfy (10). Now (9) can be further rewritten as

$$\begin{aligned} P_e &\leqslant 3(L-1)\epsilon + \eta^{-f\left(\frac{r_L}{L}\right)} + (1 - p(L)) \Pr(\varepsilon_L) \\ &\doteq \eta^{-f\left(\frac{r_L}{L}\right)}. \end{aligned} \tag{11}$$

Note that

$$\bar{R} \doteq \frac{LR}{1 + \sum_{i=1}^{L-1} \eta^{-f\left(\frac{r_L}{l}\right)}} \doteq LR$$

for $0 \leq r_L < \min(M, N)$. Thus $r_e = r_L$ and diversity gain $f\left(\frac{r_e}{L}\right)$ is achievable in the range $0 \leq r_e < \min(M, N)$. Note that $r_L = Lr$, and thus we have $d = f(r)$ for

$$r_e = rL, 0 \leq r < \frac{\min(M, N)}{L}.$$

So far we have considered only the scenario in which $r < \frac{\min(M,N)}{L}$. Now we consider what happens if we increase the value of r to $\frac{\min(M,N)}{L}$ and beyond. In this scenario, $f\left(\frac{r_L}{1}\right) = 0$, and thus $\bar{R} \doteq \frac{LR}{2}$. The message rate r_e is decreased to $r_L/2$ due to the fact that after the first block the receiver has no chance of decoding the message correctly and it always needs the second block. However, the system error probability P_e is not changed. Therefore the message rate becomes

$$r_e = r \cdot \frac{L}{2}, \frac{\min(M, N)}{L} \leq r < \frac{2\min(M, N)}{L}, \tag{12}$$

and the system DMT becomes

$$d = f\left(\frac{2r_e}{L}\right), \frac{\min(M, N)}{2} \leq r_e < \min(M, N). \tag{13}$$

Similarly, when r_e reaches $\min(M, N)$ again, i.e., r reaches $\frac{2\min(M,N)}{L}$, $f\left(\frac{r_L}{2}\right) = f\left(\frac{2r_e}{2}\right) = 0$. Thus $\bar{R} \doteq \frac{LR}{3}$ and

$$r_e = r \cdot \frac{L}{3}, \frac{2\min(M, N)}{L} \leq r < \frac{3\min(M, N)}{L}; \tag{14}$$

the system DMT becomes

$$d = f\left(\frac{3r_e}{L}\right), \frac{2\min(M, N)}{3} \leq r_e < \min(M, N). \tag{15}$$

Continuing the above until $\bar{R} \doteq R$, we obtain the desired result and the proof is complete.

B. Proof of Theorem 2

Codes being approximately universal means that such codes can achieve the optimal DMT of the channel for any distribution of $\mathcal{H}$. This includes the scenario in which

$$\mathcal{H} = \begin{pmatrix} \mathbf{H}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}_L \end{pmatrix} \tag{16}$$

where each channel matrix in $\{\mathbf{H}_l\}$ $(1 \leq l \leq L)$ follows an *arbitrary* distribution. Assume that the system in (6) transmits at a rate $LR = r_L \log_2 \eta$. The probability of any decoding error can be upper bounded by [1]

$$P \leqslant P_O + P_{e|O^c}$$

where P_O is the outage probability and $P_{e|O^c}$ is the average error probability given that the channel is not in outage. Approximate universality means that for such codes $P_{e|O^c} = e^{-\eta^\delta}$ under *any* fading distribution. For the system in (16), these include the fading distributions in which $\mathbf{H}_1 = \cdots = \mathbf{H}_l$ follow the same distribution as the $\mathbf{H}$ in (2) and $\mathbf{H}_{l+1} = \cdots = \mathbf{H}_L \equiv 0$ for all $1 \leq l \leq L-1$. When such codes are transformed into the rateless channels shown in (2), it is a simple matter to show that $\Pr(\varepsilon_l) = P_{e|O^c} = e^{-\eta^\delta}$ for any $1 \leq l \leq L$, where $\Pr(\varepsilon_l)$ is given in (9). Thus the system error probability for the rateless channel in (2) is always upper bounded by

$$P_e \leqslant Le^{-\eta^\delta} + \eta^{-f\left(\frac{r_L}{L}\right)} \doteq \eta^{-f\left(\frac{r_L}{L}\right)}.$$

The rest of the proof follows that of *Theorem 1* and is omitted.

References

[1] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[2] H. El Gamal, G. Caire and M. O. Damen, "The MIMO ARQ channel: diversity-multiplexing-delay tradeoff," *IEEE. Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601-3619, Aug. 2006.

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

[4] J. Castura, Y. Mao, and S. Draper, "On rateless coding over fading channels with delay constraints," in *Proc. 2006 Int'l Sym. Inf. Theory* (ISIT 2006), Seattle, WA, July, 2006.

[5] J. Castura and Y. Mao, "Rateless coding over fading channels," *IEEE Commun. Lett.* vol. 10, no. 1, pp. 46-48, Jan. 2006.

[6] U. Erez, M. Trott, and G. Wornell, "Rateless coding for Gaussian channels," submitted to *IEEE Trans. Inf. Theory*. [Online]. Available: arxiv.org/PS_cache/arxiv/pdf/0708/0708.2575v1.pdf

[7] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12 pp. 4152-4172, Dec. 2005.

[8] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] P. Elia, K. R. Kumar, S. A. Pawar, and P. V. Kumar, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869-3884, Sept. 2006.

[10] K. Azarian and H. El Gamal, "The throughput-reliability tradeoff in block fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 488-501, Feb. 2007.