The MIMO ARQ Channel: Diversity–Multiplexing–Delay Tradeoff

Hesham El Gamal, *Senior Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, and Mohamed Oussama Damen, *Senior Member, IEEE*

***Abstract*—In this paper, the fundamental performance tradeoff of the delay-limited multiple-input multiple-output (MIMO) automatic retransmission request (ARQ) channel is explored. In particular, we extend the diversity–multiplexing tradeoff investigated by Zheng and Tse in standard delay-limited MIMO channels with coherent detection to the ARQ scenario. We establish the three-dimensional tradeoff between reliability (i.e., diversity), throughput (i.e., multiplexing gain), and delay (i.e., maximum number of retransmissions). This tradeoff quantifies the *ARQ diversity gain* obtained by leveraging the retransmission delay to enhance the reliability for a given multiplexing gain. Interestingly, ARQ diversity appears even in long-term static channels where all the retransmissions take place in the same channel state. Furthermore, by relaxing the input power constraint allowing variable power levels in different retransmissions, we show that power control can be used to dramatically increase the diversity advantage. Our analysis reveals some important insights on the benefits of ARQ in slow-fading MIMO channels. In particular, we show that 1) allowing for a sufficiently large retransmission delay results in an *almost* flat diversity–multiplexing tradeoff, and hence, renders operating at high multiplexing gain more advantageous; 2) MIMO ARQ channels quickly approach the ergodic limit when power control is employed. Finally, we complement our information-theoretic analysis with an incremental redundancy lattice space–time (IR-LAST) coding scheme which is shown, through a random coding argument, to achieve the optimal tradeoff(s). An integral component of the optimal IR-LAST coding scheme is a list decoder, based on the minimum mean-square error (MMSE) lattice decoding principle, for joint error detection and correction. Throughout the paper, our theoretical claims are validated by numerical results.**

***Index Terms*—Automatic retransmission request (ARQ) protection, diversity–multiplexing tradeoff, incremental redundancy coding, lattice coding, multiple-input multiple-output (MIMO) channels, space–time coding.**

Manuscript received October 31, 2004; revised January 26, 2006. The work of H. El Gamal was supported in part by the National Science Foundation under Grants CCR 0118859, ITR 0219892, and CAREER 0346887.

H. El Gamal is with the Electrical and Computer Engineering Department, The Ohio State University, Columbus, IH 43210 USA (e-mail: helgamal@ece.osu.edu).

G. Caire was with The Mobile Communication group at Eurecom Institute, Sophia-Antipolis, France. He is now with the Electrical Engineering Department, the University of Southern California, Los Angeles, CA 90089 USA (e-mail: caire@usc.edu).

M. O. Damen is with the Electrical and Computer Engineering Department, University of Waterloo, Waterloo ON N2L 3G1, Canada (e-mail: modamen@ece.uwaterloo.ca).

Communicated by B. Hassibi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2006.878173

I. Introduction

THE seminal work of Telatar [1], Foschini and Gans [2], Tarokh *et al.* [3], and Guey *et al.* [4] has spurred interest in multiple-antenna wireless systems. Loosely speaking, two-dimensional signaling schemes that exploit the spatial domain to improve both the reliability and throughput of wireless channels are nicknamed *space–time codes* after [3]. The literature on space–time coding is huge (see, for example, [5] and references therein). Several settings have been considered and, for each setting, information-theoretic results and associated coding schemes have been developed.

Arguably, the coherent delay-limited (or quasi-static) multiple-input multiple-output (MIMO) setting is the most studied model. In this scenario, the channel is random but fixed during the whole codeword duration and the channel state information (CSI) is assumed to be perfectly known at the receiver and not known at the transmitter. The transmitter, though, knows the channel *statistics*. The best achievable error probability on this channel is essentially given by the so-called information outage probability, i.e., the probability that the mutual information as a function of the channel realization is below the transmitted coding rate [1].

Several classes of coherent space–time codes, targeting different optimization criteria, have been proposed. Zheng and Tse developed a powerful tool that serves as a benchmark for comparing existing space–time coding schemes and guiding the design of new approaches [6]. This tool, referred to as the diversity–multiplexing tradeoff, is inspired by rigorous information -theoretic definitions of the diversity and multiplexing gains and establishes the necessary tradeoff between reliability and throughput in *outage-limited* fading channels. In [7], the authors have established the optimality of space–time lattice coding and decoding in delay-limited MIMO channels with respect to the delay–multiplexing tradeoff [7]. More recently, different variants of the algebraic space–time constellations presented in [8], [9] were shown to achieve the optimal tradeoff under the more complex maximum-likelihood decoding rule [10]–[13].

Zheng–Tse formulation applies to channels where the transmitter does not have CSI and a codeword error results in the loss of the corresponding information message. In this work, we extend this formulation to automatic retransmission request (ARQ) MIMO channels. In this case, the receiver feeds back to the transmitter a one-bit success/failure indicator. In the success case, the transmitter moves on to the next information message in the transmission queue whereas in the failure case the transmitter retransmits a (possibly different) encoded version of

the same message. We refer to the successive transmissions of coded versions of the *same* information message as "ARQ protocol rounds." The ARQ protocol is allowed to use a given maximum number of rounds, denoted by L. If after L rounds no successful decoding has occurred, an error is declared. In this case, we assume that the message will be dropped from the transmission queue (i.e., delay-sensitive application). Therefore, we define the probability of error as the probability of no successful decoding within L protocol rounds.

We investigate and completely characterize the three-dimensional diversity–multiplexing–delay tradeoff in MIMO ARQ channels.[1] This tradeoff establishes, rigorously, the fact that the ARQ retransmission delay can be exploited as a potential source for diversity. We investigate two extreme cases of channel dynamics: long-term and short-term static channels. In the long-term static case, the MIMO channel matrix is assumed to be constant over all the ARQ rounds. This scenario applies to very fast ARQ protocols and/or very slow fading environments, such as wireless LANs [14]. In the short-term static case, the MIMO channel matrix is constant over each transmission round of the ARQ protocol but changes independently from round to round. This scenario applies to slow ARQ protocols where the time between the consecutive rounds is larger than the channel coherence time, or to frequency-selective fading, where each ARQ transmission takes place at a different frequency according to some frequency hopping scheme.

It is worthwhile noticing that the performance improvement of ARQ holds even under the more restrictive case of long-term static channel, where no time diversity can be exploited. However, as shown in the sequel, the long-term static assumption limits the ARQ diversity at low multiplexing gains. In fact, allowing for larger values of the maximum ARQ delay translates into *flatter* diversity–multiplexing tradeoff curves in this scenario (i.e., in the limit $L \to \infty$ one can achieve simultaneously the maximum multiplexing gain and maximum diversity advantage).

We then show that the limited ARQ diversity advantage at low multiplexing gains, in long-term static channels, can be significantly increased by combining ARQ retransmissions with a properly constructed power control algorithm. This algorithm does not require any additional feedback beyond the standard one-bit ARQ feedback signal, and is inspired by the power control diversity gain reported in [15]. Contrary to most earlier works on MIMO channels with feedback, the proposed power control ARQ scheme avoids the unrealistic assumption of noncausal channel state information knowledge at the transmitter. This feature is expected to translate into enhanced robustness in practical implementations. We also observe that the proposed ARQ algorithms with and without power control are, loosely speaking, analogous to the Schalkwijk–Kailath and Schalkwijk–Barron coding scheme for communication over additive white Gaussian noise (AWGN) channels with feedback [16], [17], respectively. Similar to our case, allowing for a large peak-to-average power ratio in [16] dramatically improves the achievable reliability with respect to the case of a strict peak power constraint [17]. Furthermore, the common feature of all such schemes is that the whole information message can be decoded from the first block alone, if the channel is well-behaved, and retransmission are used as a "refinement," in order to tame atypical (outage) channel events. This boosts reliability (or diversity, in our case), without sacrificing the coding rate.

The achievability of our information-theoretic results relied on using random Gaussian codebooks coupled with incomplete decoders. This motivates our next step where we construct an incremental redundancy lattice space–time (IR-LAST) coding scheme that achieves the optimal tradeoff. An important ingredient in this construction is a list lattice decoding algorithm optimized for joint error correction and detection. Finally, we validate our theoretical claims with numerical examples based on explicit code constructions, demonstrating significant performance gains in certain representative scenarios.

Recently, there has been a growing interest in MIMO ARQ schemes (e.g., [18]–[22]). Those works have been largely motivated by heuristic arguments. The theoretical foundation developed here should serve as a benchmark for evaluating previously proposed schemes and inspiring more innovative approaches.

Throughout the paper, we use the following notation. The superscript ^{c} denotes complex quantities, $^{\mathsf{T}}$ denotes transpose, and $^{\mathsf{H}}$ denotes Hermitian transpose. The notation $\boldsymbol{v} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that $\boldsymbol{v}$ is a circular symmetric complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For real Gaussian random vector we use the notation $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The acronym i.i.d. means "independent and identically distributed." We use $\doteq$ to denote exponential equality, i.e., $f(z) \doteq z^b$ means that $\lim_{z\to\infty} \frac{\log f(z)}{\log z} = b$, $\dot{\geq}$ and $\dot{\leq}$ are used similarly. For a bounded Jordan-measurable region $\mathcal{R} \subset \mathbb{R}^m$, $V(\mathcal{R})$ denotes the volume of $\mathcal{R}$. $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix, and $\otimes$ denotes the Kronecker product. The complement of a set $\mathcal{A}$ is denoted by $\overline{\mathcal{A}}$. The positive part of a real variable x is denoted by $[x]_+$.

The rest of the paper is organized as follows. In Section II, we define the MIMO ARQ channel model and its performance measures in terms of diversity gain, multiplexing gain, and delay. Section III establishes the fundamental diversity–multiplexing–delay tradeoff of MIMO ARQ channels. In Section IV, we present the IR-LAST coding scheme, that achieves the optimal tradeoff, along with representative numerical results that demonstrate the gains offered by it. Finally, we offer some concluding remarks in Section V. In order to enhance the flow of the paper, we collect all the proofs in the Appendix.

II. Background

A. Channel and ARQ Protocol Models

We consider a frequency-flat fading M-transmit N-receive MIMO channel with no CSI at the transmitter and perfect CSI at the receiver. The following ARQ protocol is considered. The transmitter has an infinite buffer of information messages to send.[2] The information message to be transmitted is encoded by a space–time encoder, and mapped into a sequence of L matrices, or blocks, $\{\boldsymbol{X}_\ell^c \in \mathbb{C}^{M\times T} : \ell = 1, \ldots, L\}$. The trans-

[1] Here, delay refers to the maximum number of transmission rounds L of the ARQ protocol.

[2] In this infinite backlog case, the *stability* of the protocol is irrelevant [23].

mission of each block takes T channel uses, by transmitting the matrix columns in parallel over the M transmit antennas, as in standard space–time coding. At the ℓth round of the current information message, $\boldsymbol{X}_\ell^c$ is transmitted. The decoder is allowed to process the received signal over all the ℓ received blocks, in order to decode the message. If successful decoding is detected, a positive acknowledgment signal (ACK) is sent back to the transmitter whereas a negative acknowledgment (NACK) signal is sent in case of detection of a decoding failure. The ACK/NACK one-bit message is the only feedback allowed in our model and the ARQ feedback channel is assumed to be error-free and zero-delay. Upon reception of the ACK, the transmitter sends the first block of the next message in the buffer whereas the reception of the NACK triggers the transmission of the next block of the current message, $\boldsymbol{X}_{\ell+1}^c$. The only exception to the above rule is when the maximum number of protocol rounds, L, is reached. In this case, a NACK bit will be interpreted as an error, the current message is removed from the transmission buffer, and the transmission of the next message is started anyway. Error in the system occur either when the decoder makes a decoding error at round $\ell < L$ and it fails to detect it (undetected error event) or when the decoder makes a decoding error at round L.

We notice that the encoding rule that maps the information message into the blocks is generally different for each block. Hence, the protocol implements a form of incremental redundancy [23]: the space–time codes defined by $1, 2, \ldots, L$ blocks can be seen as progressively punctured version of the same space–time code with block length LT.

Let us focus on the transmission of the current information message. The complex baseband model of our channel is defined by

$$\boldsymbol{y}_{\ell,t}^c = \sqrt{\frac{\rho}{M}}\boldsymbol{H}_\ell^c \boldsymbol{x}_{\ell,t}^c + \boldsymbol{w}_{\ell,t}^c \tag{1}$$

where the index $\ell = 1, 2, \ldots$, counts the protocol rounds and $t = 1, \ldots, T$ counts the channel uses in each block

$$\{\boldsymbol{x}_{\ell,t}^c \in \mathbb{C}^M : t = 1, \ldots, T\}$$

are the columns of the ℓth block $\boldsymbol{X}_\ell^c$

$$\{\boldsymbol{w}_{\ell,t}^c \in \mathbb{C}^N : t = 1, \ldots, T\} \text{ and } \{\boldsymbol{y}_{\ell,t}^c \in \mathbb{C}^N : t = 1, \ldots, T\}$$

denote the channel noise and the corresponding received signal block, respectively. The channel noise is assumed to be temporally and spatially white with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. The channel in the ℓth round is characterized by the matrix $\boldsymbol{H}_\ell^c \in \mathbb{C}^{N\times M}$ with the (i,j)th element $h_{ij,\ell}^c$ representing the fading coefficient between the jth transmit and the ith receive antenna. The fading coefficients are assumed to be i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ and remain fixed over each block, for $t = 1, \ldots, T$.

As anticipated in the Introduction, we consider two distinct scenarios of channel dynamics: 1) long-term static channels, where the channel coefficients remain constant during all ARQ rounds and change to new independent values with each new packet (i.e., $\boldsymbol{H}_\ell^c = \boldsymbol{H}^c$ for all $\ell = 1, \ldots, L$); 2) short-term static channels, where the channel remains constant during each round and changes independently at each round.

The long-term static model aims at decoupling the ARQ gain from the temporal (or frequency) interleaving gain. It represents the worst case scenario in terms of the achievable diversity with a maximum of L ARQ rounds. Assuming the channel coherence time to be a random variable (coinciding with the renewal event) in this model allows for some elegance and simplicity in the analysis. We also note that the long-term static channel model is justified in practice by considering a time-division multiple access (TDMA) network environment where the channel is allocated to a given (reference) transmitter–receiver pair only sporadically, according to some policy at the MAC layer. When the channel is allocated (i.e., the reference transmitter–receiver pair becomes active) the transmission takes place using the MIMO ARQ scheme at the physical layer. If the channel coherence time is much larger than L, but it is smaller than the idle time between two consecutive active times, our long-term static model precisely describes the channel from the perspective of the reference transmitter–receiver pair.

Also, we consider two different input power constraints: 1) short-term (or per-block) average power constraint; 2) long-term average power constraint. In the first case, we enforce

$$\mathbb{E}\left[\frac{1}{T}\|\boldsymbol{X}_\ell^c\|_F^2\right] \le M \tag{2}$$

for all $\ell = 1, \ldots, L$, where expectation is with respect to the uniform probability measure over the codebook. This means that the average transmitted power in each round of the ARQ protocol is the same, irrespective of the round index ℓ.

In the second case, we enforce

$$\limsup_{\tau\to\infty} \mathbb{E}\left[\frac{1}{T\tau}\sum_{s=1}^{\tau}\|\boldsymbol{X}^c[s]\|_F^2\right] \le M \tag{3}$$

where we have introduced the *absolute* index s of the transmitted block,[3] and now $\boldsymbol{X}^c[s]$ denotes the sth transmitted block since the beginning of transmission. Again, expectation is with respect to the uniform probability measure over the codebook. Clearly, in both cases, the parameter ρ in (1) takes on the meaning of *average* signal-to-noise ratio (SNR) per receiver antenna.

In order to simplify the presentation in the sequel, we will sometimes appeal to the following real channel model, equivalent to (1). After ℓ transmission rounds, the total received signal is given by

$$\boldsymbol{y}_\ell = \boldsymbol{H}_\ell \boldsymbol{x} + \boldsymbol{w}_\ell \tag{4}$$

where we define

$$\boldsymbol{x} = (\boldsymbol{x}_{1,1}^\mathsf{T}, \ldots, \boldsymbol{x}_{1,T}^\mathsf{T}, \ldots, \boldsymbol{x}_{L,1}^\mathsf{T}, \ldots, \boldsymbol{x}_{L,T}^\mathsf{T})^\mathsf{T}$$

with $\boldsymbol{x}_{\ell,t}^\mathsf{T} = \left[\mathrm{Re}\{\boldsymbol{x}_{\ell,t}^c\}^\mathsf{T}, \mathrm{Im}\{\boldsymbol{x}_{\ell,t}^c\}^\mathsf{T}\right]^\mathsf{T}$, and

$$\boldsymbol{w}_\ell = (\boldsymbol{w}_{1,1}^\mathsf{T}, \ldots, \boldsymbol{w}_{1,T}^\mathsf{T}, \ldots, \boldsymbol{w}_{\ell,1}^\mathsf{T}, \ldots, \boldsymbol{w}_{\ell,T}^\mathsf{T})^\mathsf{T}$$

[3]Notice that ℓ is a relative index, denoting the ℓth block in the transmission of the current message.

with $\boldsymbol{w}_{\ell,t}^{\mathsf{T}} = \left[\mathrm{Re}\{\boldsymbol{w}_{\ell,t}^{c}\}^{\mathsf{T}}, \mathrm{Im}\{\boldsymbol{w}_{\ell,t}^{c}\}^{\mathsf{T}}\right]^{\mathsf{T}}$. The vector $\boldsymbol{y}_{\ell} \in \mathbb{R}^{2NT\ell}$ represents the signal received over all transmitted blocks from 1 to ℓ.

The channel matrix $\boldsymbol{H}_{\ell}$ has dimensions $2NT\ell \times 2MTL$, and is formed by taking the first $2NT\ell$ rows of the matrix (5) at the bottom of the page, which is composed by L diagonal blocks. Each block has also a block-diagonal form, with T diagonal blocks equal to the $2N \times 2M$ real expansion of the complex channel matrix $\boldsymbol{H}_{\ell}^{c}$. In the case of long-term static channel, all these blocks are equal since $\boldsymbol{H}_{\ell}^{c}$ is constant with ℓ. Notice that for $\ell < L$, the matrix $\boldsymbol{H}_{\ell}$ can be partitioned into two blocks. The leftmost $2NT\ell \times 2MT\ell$ block is block-diagonal while the rightmost $2NT\ell \times 2MT(L-\ell)$ block is zero. This corresponds to the fact that at round ℓ the blocks $\boldsymbol{X}_{\ell+1}^{c}, \ldots, \boldsymbol{X}_{L}^{c}$ have not been transmitted yet, and in our real model they appear as multiplied by a zero channel matrix.

The design of a space–time code for the ARQ channel, therefore, reduces to the construction of a codebook $\mathcal{C} \subseteq \mathbb{R}^{2MTL}$ enjoying certain desirable properties.

B. Throughput, Transmitted Power, and Probability of Error

In this section, we use *renewal theory* (see [23] and references therein) in order to characterize the average throughput, the average transmitted power, and the probability of error of the ARQ scheme.

Consider the event that the transmission of the current information message is stopped, either because the receiver feeds back an ACK, or because the maximum number of rounds L is reached. In the long-term static channel case, we assume that the fading changes independently at each occurrence of such event (this assumption is automatically satisfied by the short-term static channel case). Under the above assumption, it is readily seen that stopping the current message transmission is a renewal event [23]: at each occurrence of such event the system resets and restarts anew.

Let $\mathcal{T}$ be a random variable indicating the inter-renewal time, i.e., the time (in slots) between two consecutive occurrences of the renewal event, and let $\mathcal{A}_{\ell}$ denote the event that an ACK is fed back at round ℓ. For all $\ell = 1, \ldots, L-1$, we have

$$\Pr(\mathcal{T} = \ell) = \Pr\left(\overline{\mathcal{A}}_1, \ldots, \overline{\mathcal{A}}_{\ell-1}, \mathcal{A}_{\ell}\right) \triangleq q(\ell). \tag{6}$$

At round L, since even in the case of NACK the transmitters moves on to the next message, we have

$$\Pr(\mathcal{T} = L) = 1 - \sum_{\ell=1}^{L-1} q(\ell). \tag{7}$$

It turns out that it is more convenient to work with the probabilities

$$p(\ell) \triangleq \Pr\left(\overline{\mathcal{A}}_1, \ldots, \overline{\mathcal{A}}_{\ell}\right) \tag{8}$$

where, by definition, we let $p(0) = 1$. It is a simple matter to verify the relation

$$q(\ell) = p(\ell-1) - p(\ell) \tag{9}$$

which yields

$$\sum_{\ell=1}^{L} \left[\sum_{k=1}^{\ell} a_k\right] \Pr(\mathcal{T} = \ell) = \sum_{\ell=1}^{L} a_{\ell} p(\ell-1) \tag{10}$$

for any $(a_1, \ldots, a_L) \in \mathbb{R}^{L}$.

Let b denote the size of the information messages in bits and let $B[s]$ denote the number of bits removed from the transmission buffer at slot s (absolute index). We have that $B[s] = b$ if the renewal event occurs at time s, and $B[s] = 0$ otherwise. The long-term average throughput of the ARQ protocol, expressed in *transmitted* bits per channel use (PCU), is given by [23]

$$\begin{aligned} \eta &= \liminf_{\tau \to \infty} \frac{1}{T\tau} \sum_{s=1}^{\tau} B[s] \\ &= \frac{b/T}{\mathbb{E}[\mathcal{T}]} \\ &= \frac{b/T}{\sum_{\ell=0}^{L-1} p(\ell)} \end{aligned} \tag{11}$$

where the last line follows by noticing that $\mathbb{E}[\mathcal{T}]$ is given by (10) for $a_1 = \cdots = a_L = 1$. In the following, we let $R_1 \triangleq b/T$ denotes the rate of the first block in bits PCU.

The long-term power constraint in (3) applies to any feasible power control rule including nonstationary and randomized algorithms. In the sequel, however, we shall restrict ourselves to the class of stationary power control policies, for which the power spent at round ℓ is just a deterministic function of ℓ. Let Γ_{ℓ} denote the average energy allocated to the ℓth round of transmission. Consequently, the limit in (3) takes on the form

$$\begin{aligned} \limsup_{\tau \to \infty} \mathbb{E}\left[\frac{1}{T\tau} \sum_{s=1}^{\tau} ||\boldsymbol{X}^{c}[s]||_F^2\right] &= \frac{1}{T} \frac{\mathbb{E}\left[\sum_{\ell=1}^{\mathcal{T}} \Gamma_{\ell}\right]}{\mathbb{E}[\mathcal{T}]} \\ &= \frac{1}{T} \frac{\sum_{\ell=1}^{L} \Gamma_{\ell}\ p(\ell-1)}{\sum_{\ell=0}^{L-1} p(\ell)} \end{aligned} \tag{12}$$

where the numerator in the last line of (12) follows again from (10) by letting $a_k = \Gamma_k$ for $k = 1, \ldots, L$.

The ARQ system incurs an error if decoding fails but it is not detected, so that an ACK is fed back, or if decoding fails

$$\boldsymbol{H}_L \triangleq \sqrt{\frac{\rho}{M}} \mathrm{diag}\left(\boldsymbol{I}_T \otimes \begin{bmatrix} \mathrm{Re}\{\boldsymbol{H}_1^{c}\} & -\mathrm{Im}\{\boldsymbol{H}_1^{c}\} \\ \mathrm{Im}\{\boldsymbol{H}_1^{c}\} & \mathrm{Re}\{\boldsymbol{H}_1^{c}\} \end{bmatrix}, \ldots, \boldsymbol{I}_T \otimes \begin{bmatrix} \mathrm{Re}\{\boldsymbol{H}_L^{c}\} & -\mathrm{Im}\{\boldsymbol{H}_L^{c}\} \\ \mathrm{Im}\{\boldsymbol{H}_L^{c}\} & \mathrm{Re}\{\boldsymbol{H}_L^{c}\} \end{bmatrix}\right) \tag{5}$$

at round L. Let $\mathcal{E}_\ell$ denote the event that the decoding outcome is not correct with ℓ received blocks. For a given code, power control, channel statistics, and decoding/error detection scheme, the probability of error can be written as

$$\begin{aligned} P_e &= \sum_{\ell=1}^{L} \Pr(\mathcal{E}_\ell, \mathcal{T} = \ell) \\ &= \sum_{\ell=1}^{L-1} \Pr(\mathcal{E}_\ell, \overline{\mathcal{A}}_1, \ldots, \overline{\mathcal{A}}_{\ell-1}, \mathcal{A}_\ell) + \Pr(\mathcal{E}_L, \overline{\mathcal{A}}_1, \ldots, \overline{\mathcal{A}}_{L-1}) \\ &\leq \sum_{\ell=1}^{L-1} \Pr(\mathcal{E}_\ell, \mathcal{A}_\ell) + \Pr(\mathcal{E}_L) \end{aligned} \tag{13}$$

where the terms in the last line have the following meaning: $\Pr(\mathcal{E}_\ell, \mathcal{A}_\ell)$ is the probability of *undetected* decoding error with $\ell \leq L-1$ received blocks, and $\Pr(\mathcal{E}_L)$ is the probability of decoding error with L received blocks.

C. Diversity–Multiplexing Tradeoff

In this work, we extend Zheng–Tse formulation of the diversity–multiplexing tradeoff [6] to the MIMO ARQ channel defined above. Zheng and Tse considered a family of space–time codes $\{\mathcal{C}_\rho\}$ indexed by their operating SNR ρ, such that the code $\mathcal{C}_\rho$ has rate $R(\rho)$ bits PCU and error probability $P_e(\rho)$. For this family, the multiplexing gain r and the diversity gain d are defined by

$$r \triangleq \lim_{\rho\to\infty} \frac{R(\rho)}{\log\rho} \quad \text{and} \quad d \triangleq -\lim_{\rho\to\infty} \frac{\log P_e(\rho)}{\log\rho}. \tag{14}$$

The optimal diversity–multiplexing tradeoff yields the maximum possible *SNR exponent* for every value of r. In the following, this optimal exponent is denoted by $d^*(r,1)$ in order to highlight the fact that transmission takes place over a single block. The main result of [6] is summarized by the following theorem.

Theorem 1: The optimal diversity gain of the coherent block-fading MIMO channel with M transmit, N receive antennas, and multiplexing gain r, is given by $d^*(r,1) = f(r)$, where $f(\cdot)$ is the piecewise linear function joining the points $(k,(M-k)(N-k))$ for $k = 0,\ldots,\min\{M,N\}$. In particular, $d^*(r,1)$ is achieved by the random Gaussian i.i.d. code ensemble for all block lengths $T \geq M+N-1$. □

In [7], the authors have shown that carefully constructed ensembles of LAST codes achieves $d^*(r,1)$ for $T \geq M+N-1$ under minimum mean-square error (MMSE) lattice decoding. In the sequel, we will show that this class of codes can be used as a building block for constructing optimal incremental redundancy codes for the MIMO ARQ channel. More recently, the existence of space–time constellations that achieve $d^*(r,1)$ for $T = M$ was established in [13].

In order to extend Zheng–Tse formulation of the diversity–multiplexing tradeoff to the ARQ case, we consider a family of ARQ protocols where the size of the information messages $b(\rho)$ depends on the operating SNR ρ. These protocols are based on a family of space–time codes $\{\mathcal{C}_\rho\}$ with first-block rate $R_1(\rho) = b(\rho)/T$ and overall block length TL. Then, we define the *effective* ARQ multiplexing gain as

$$r_e \triangleq \lim_{\rho\to\infty} \frac{\eta(\rho)}{\log\rho} \tag{15}$$

where $\eta(\rho)$ is given by (11), noticing that both b and the probabilities $p(\ell)$ depend on ρ. The *effective* ARQ diversity gain is defined as

$$d = -\lim_{\rho\to\infty} \frac{\log P_e(\rho)}{\log\rho} \tag{16}$$

where $P_e(\rho)$ is given by (13).

The optimal diversity–multiplexing tradeoff of MIMO ARQ channels yields the maximum possible SNR exponent, denoted by $d^*(r_e,L)$, for every value of r_e. As a consistency check, it is immediate to verify that these definitions reduce to the standard Zheng–Tse formulation when $L = 1$ (i.e., no ARQ).

III. The Fundamental Tradeoff

In this section, we find an explicit characterization for the exponent $d^*(r_e,L)$ of MIMO ARQ channels. In our study, we differentiate between two scenarios. In the first, a short-term power constraint is enforced and hence the same power level is used in all transmissions. In this case, $d^*(r_e,L)$ quantifies the **ARQ diversity gain** as a function of the maximum transmission delay L and illustrates the suboptimality of previously proposed schemes. In the second scenario, a long-term power constraint is enforced, and hence, we allow for varying the power level in every retransmission while keeping the overall average power fixed. We construct an asymptotically optimal power control policy which yields very significant diversity gains in long-term static channels, especially at low multiplexing gains. It is worth noting that the proposed power control algorithm does not require any additional feedback. The only information needed is the ACK/NACK feedback bit and *off-line* estimates of the probabilities $p(\ell)$ for $1 \leq \ell \leq L$.

A. ARQ Diversity

We are now ready to state our result on the diversity–multiplexing–delay tradeoff of MIMO ARQ channels with short-term average power constraint.

Theorem 2: The optimal diversity gain of the coherent block-fading MIMO ARQ channel with M transmit, N receive antennas, maximum number of ARQ rounds L, under the short-term power constraint, is given as follows.

In the case of long-term static channels

$$d^*_{ls}(r_e,L) = \begin{cases} f\left(\frac{r_e}{L}\right), & 0 \leq r_e < \min\{M,N\} \\ 0, & r_e \geq \min\{M,N\}. \end{cases} \tag{17}$$

In the case of short-term static channels,

$$d^*_{ss}(r_e,L) = \begin{cases} Lf\left(\frac{r_e}{L}\right), & 0 \leq r_e < \min\{M,N\} \\ 0, & r_e \geq \min\{M,N\}. \end{cases} \tag{18}$$

Furthermore, the optimal tradeoff is achieved by codes with finite block length T subject to the conditions

$$T \geq \left\lceil \frac{M+N-1}{L} \right\rceil, \text{ for long-term static channels} \quad (19)$$

$$T \geq M+N-1, \quad \text{for short-term static channels.} \quad (20)$$

Proof: (Sketch) The converse is shown through the judicious application of Fano inequality. We present two approaches for achieving the optimal tradeoff. We first establish the achievability of the exponents $d^*_{ls}(r_e, L)$ and $d^*_{ss}(r_e, L)$ in the limit of asymptotically large block length ($T \to \infty$) by employing the typical set decoder which has a built-in error detection capability. The achievability of the optimal tradeoff for finite T is then shown by using an incomplete bounded-distance decoder that mimics the behavior of the typical set decoder. In particular, we consider a decoder that accepts the message $\widehat{w}$ at round ℓ if 1) the channel is not in outage; 2) the corresponding codeword $\widehat{\boldsymbol{x}}$ is the unique codeword such that

$$|\boldsymbol{y}_\ell - \boldsymbol{H}_\ell \widehat{\boldsymbol{x}}|^2 \leq NT\ell(1+\delta)$$

for some $\delta > 0$ (which will be determined in the sequel). On the contrary, if either there is no such codeword or there are more than one, then a NACK is fed back. Since the noise $\boldsymbol{w}_\ell$ has dimension $2NT\ell$ and it is Gaussian i.i.d. with components $\sim \mathcal{N}(0, 1/2)$, the above condition is equivalent to saying that the noise is typical *and* the channel is not in outage. The term δ will be required to grow with the SNR in order to ensure that, despite the finite block length, the probability that the noise is outside the sphere of squared radius $NT\ell(1+\delta)$ vanishes with an SNR exponent at least equal to $d^*(L, r_e)$. The technical details of the proof are reported in Appendix A ∎

It is worth noting that in our proof technique, error detection is accomplished via an incomplete decoder [24], and hence, does not require additional redundancy. In most practical ARQ schemes, errors are detected by using an outer coding layer devoted to error detection (typically, a cyclic redundancy check (CRC)). Unfortunately, the following *intuitive* argument suggests that the "classical CRC" approach requires T growing to infinity in order to operate at the optimal tradeoff. Consider a MIMO ARQ scheme based on the following error detection rule: the transmitter and the receiver pre-agree on a check function

$$\mu : \{1, \ldots, \rho^{r_1 T}\} \to \{1, \ldots, 2^k\}$$

that maps information messages w into auxiliary check messages u. The composite message $w' = (w, \mu(w))$ is transmitted using the MIMO ARQ scheme. At each round $\ell \leq L-1$, the receiver decodes $(\widehat{w}, \widehat{u})$. If $\mu(\widehat{w}) = \widehat{u}$, the message is accepted and the transmission of the current message is stopped (ACK is fed back). If $\mu(\widehat{w}) \neq \widehat{u}$, an error is declared and the next round is requested (NACK is fed back). It is not difficult to see that the probability of undetected error at any round $\ell < L$ must vanish with SNR at least with exponent $d^*(r_e, L)$. Otherwise, the undetected error probability dominates the system performance. We assume that, if $\widehat{w} \neq w$, then $\mu(\widehat{w})$ is uniformly distributed over all possible messages u.[4] Hence, errors are not revealed with probability $\approx 2^{-k}$. The probability of undetected error at round ℓ is given by

$$\Pr(\mathcal{E}_\ell, \mathcal{A}_\ell) = \Pr(\mathcal{A}_\ell | \mathcal{E}_\ell) \Pr(\mathcal{E}_\ell) \doteq 2^{-k} \rho^{-d_\ell} \quad (21)$$

where d_ℓ denotes the SNR exponent of the probability of making an error with ℓ received blocks. Assuming, without loss of generality, that $d_1 \leq d_2 \leq \cdots \leq d_L = d^*(r_e, L)$, from the bound on error probability (13) we obtain that

$$2^{-k} \rho^{-d_1} \doteq \rho^{-d^*(r_e, L)}. \quad (22)$$

This implies that k must grow with SNR as

$$k(\rho) = (d^*(r_e, L) - d_1) \log \rho / \log 2.$$

The first-block rate of the CRC-based scheme, denoted by r'_1, is given by

$$r'_1 = \frac{r_1 T \log \rho - k(\rho)}{T \log \rho} = r_1 - \frac{d^*(r_e, L) - d_1}{T \log 2}.$$

If T is a constant independent of SNR, then r'_1 is strictly less than r_1. This prevents the CRC scheme from achieving the optimal tradeoff. However, if T grows without bounds at any speed as $\rho \to \infty$, then asymptotically optimal performance can be achieved by the CRC scheme.

Theorem 2 establishes the interesting fact that retransmission delay can be exploited to significantly improve diversity, especially at high multiplexing gain. The basic idea is that the multiplexing gain is determined by the rate assuming only one round whereas the diversity gain is determined by the rate of the *composite* code received at the end of the maximum number of rounds. This can be explained by the fact that *most* packets are decoded successfully in the first round and ARQ retransmissions are used to correct the *rare* error events, and hence, pushing the probability of error down with an asymptotically vanishing price in the transmission rate. This consideration is valid under the condition that errors in rounds $\ell < L$ are detected with high probability. As shown in Appendix A, this condition is always verified for sufficiently large T, for every given operating SNR ρ.

Interestingly, the ARQ diversity gain appears even in long-term static channels. In fact, as shown in Fig. 1, the tradeoff curve become flatter as L increases. This implies that one can approach the full diversity point, i.e., $d = MN$, for any multiplexing gain $0 < r_e < \min(N, M)$ by using a sufficiently large L. It is important to notice here that, in long-term static channels, larger values of L do not imply any increase of the *temporal diversity* (i.e., each codeword is still transmitted over a single realization of the channel matrix). It is also evident that, in long-term static channels, the diversity improvement due to ARQ disappears as the multiplexing gain tends to zero. In fact, we have $d^*_{ls}(0, L) = d^*(0, 1) = NM$, irrespectively of L. On the other hand, in short-term static channels ARQ provides

[4]This assumption is typically used in the design of error detection CRC schemes. It can be justified theoretically by letting the function μ be constructed by a random binning assignment of the codewords onto 2^k bins, and averaging over the ensemble of random binning assignments. However, this argument does not apply if the function μ and the code are designed jointly, as done in [25] in a different context.

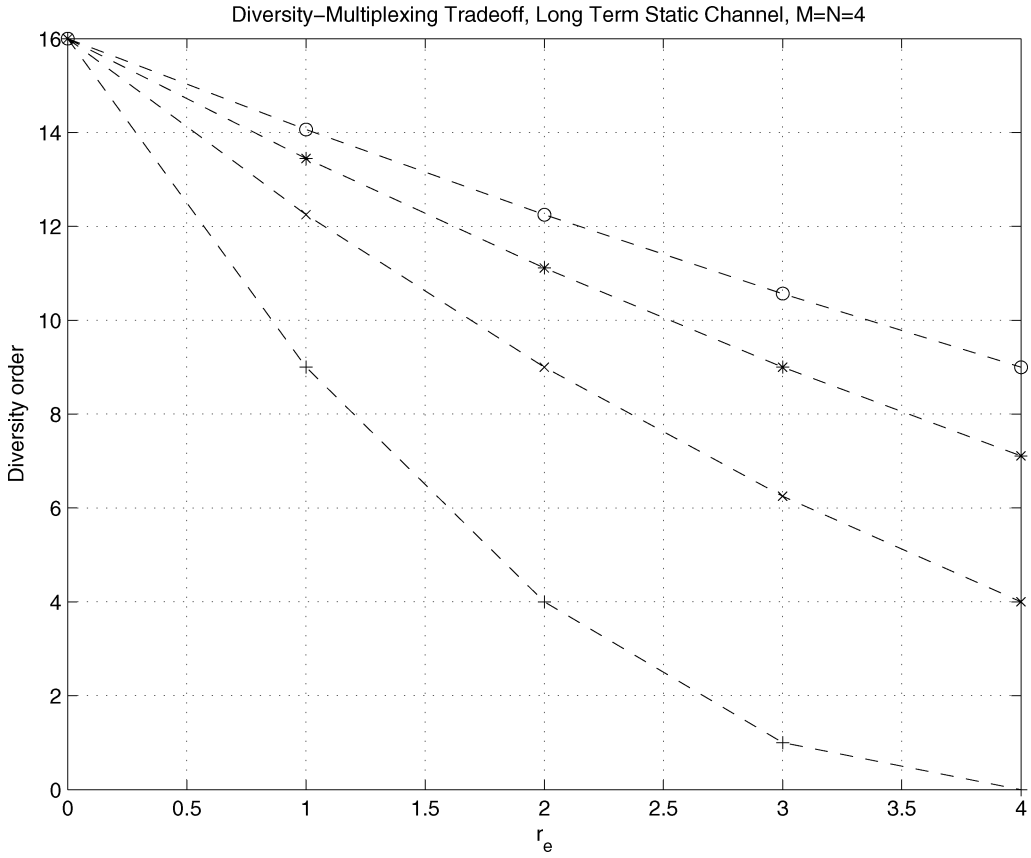

Fig. 1. The diversity–multiplexing tradeoff with different values of the maximum number of ARQ rounds "L."

also temporal diversity, as seen in the fact that $d^*_{ss}(r_e, L) = Ld^*_{ls}(r_e, L)$. This temporal diversity gain appears at both low and high multiplexing gains.

Next, we use Theorem 2 to quantify the loss incurred by some low-complexity suboptimal schemes. The first scheme we consider is the packet combining (PC) approach. In this approach, the same encoding rule is used in every retransmission and the received packets are combined (through maximum ratio combing) before decoding. The tradeoff achieved by this scheme is characterized in the following corollary.

Corollary 3: The packet combining (PC) diversity gain for long-term static and short-term static channels with M transmit, N receive antennas, maximum number of ARQ rounds L, and effective multiplexing gain $0 \leq r_e < \min(N, M)$, are given by

$$d^{(pc)}_{ls}(r_e, L) = f(r_e) \tag{23}$$

$$d^{(pc)}_{ss}(r_e, L) = Lf(r_e). \tag{24}$$

Proof: The proof is straightforward, and hence, is omitted for brevity. ∎

The suboptimality of the PC approach is manifested in the fact that it fails to exploit the ARQ diversity gain in long-term static channels. In these channels, the PC approach offers only a $10\log(L)$-decibel SNR increase, and hence, is limited by the same tradeoff of the channel without ARQ. In short-term channels, the PC approach only exploits the temporal diversity.

Another suboptimal scheme, targeting long-term static channels, was proposed in [18]. This scheme sends carefully chosen space–time constellations such that after M transmissions they form a square orthogonal constellation. The achievable diversity with this scheme, in long term static channels,[5] is upper-bounded in the following corollary.

Corollary 4: The diversity gain of the orthogonal ARQ scheme for long-term static channels with M transmit, N receive antennas, maximum number of ARQ rounds $L = M$, and effective multiplexing gain $0 \leq r_e < \min(N, M)$, is given by

$$d^{(o)}_{ls} \leq MN \max\left\{1 - \frac{r_e}{r_o M},\ 0\right\} \tag{25}$$

where r_o is the rate of the orthogonal constellation.[6]

Proof: (Sketch) Let $R_1 = r_1 \log(\rho)$ be the rate used in the first transmission. Without ARQ, orthogonal transmission transforms the $M \times N$ MIMO channel into a $1 \times NM$ single-input multiple-output (SIMO) channel with a maximum value of multiplexing gain equal to r_o. The tradeoff for this scheme is [6]

$$d^{(o)} = MN \max\left\{1 - \frac{r_1}{r_o}, 0\right\}. \tag{26}$$

One can then follow in the footsteps of Appendix A to see that the tradeoff for the long-term static channel with M rounds of

[5]We restrict the analysis to long-term static channels since the main property of the constellation, orthogonality, is destroyed in short-term static channels

[6]The rate r_o is expressed in modulation symbols per channel use. For example, $r_o = 1$ for the Alamouti constellation.

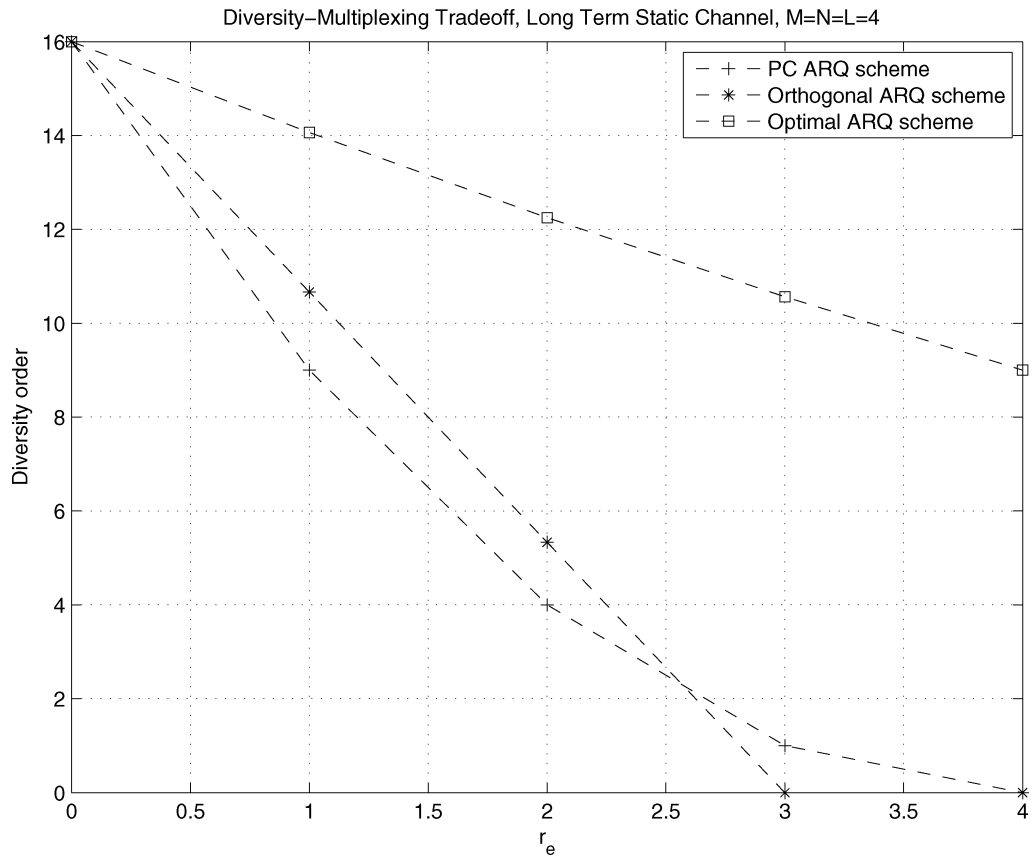

Fig. 2. The diversity–multiplexing tradeoff of several ARQ schemes.

ARQ transmissions is obtained by replacing r_1 in (26) with r_1/M, i.e.,

$$d_{ls}^{(o)} = MN \max\left\{1 - \frac{r_1}{r_o M}, 0\right\}. \tag{27}$$

The result follows by noting that $r_e \leq r_1$. ■

Fig. 2 compares the diversity gain of the PC and orthogonal ARQ schemes with that of the optimal tradeoff where it is apparent that the suboptimality of these approaches is more significant at high multiplexing gains.

B. Power Control Diversity

As shown in the previous subsection, in long-term static channels under the short-term power constraint the ARQ diversity advantage over conventional coherent space–time coding vanishes at low multiplexing gain. Here, we consider the long-term power constraint and construct an asymptotically optimal power control algorithm that yields very significant diversity advantage in long-term static channels especially at low multiplexing gains. A distinguishing feature of the proposed algorithm is that it avoids the noncausal feedback assumptions adopted in many earlier works. The proposed power control algorithm is enabled by the observation that the probability of transmitting the ℓ round, $p(\ell-1)$, decays polynomially with SNR. Therefore, the energy allocated to the ℓth block, Γ_ℓ, can be made proportional to $1/p(\ell-1)$, allowing for a significant increase in transmitted power without violating the long-term power constraint. The larger power level in round ℓ will result in a smaller $p(\ell)$, and hence, even larger $\Gamma_{\ell+1}$. Through this recursive procedure, the probability of error is minimized. Clearly, this power allocation policy only requires the knowledge of the probabilities $p(\ell)$'s, which can be estimated off-line.

The following theorem establishes the diversity–multiplexing–delay tradeoff of MIMO ARQ channels with long-term average power constraint (since this subsection treats only the long-term static channels, we drop the subscript "ls" in the following for brevity).

Theorem 5: The optimal diversity gain of the coherent block-fading MIMO ARQ channel with M transmit, N receive antennas, maximum number of ARQ rounds L, and effective multiplexing gain $0 \leq r_e < \min(N, M)$, under the long-term power constraint, is given by $d^*(r_e, L) = \xi_L$, where ξ_L is obtained recursively as follows. Let $\xi_0 = 0$. For $\ell = 1, \ldots, L$, let

$$\xi_\ell = \inf_{\boldsymbol{v} \in \mathcal{O}_\ell \cap \mathbb{R}_+^{\min\{M,N\}}} \left\{ \sum_{j=1}^{\min\{M,N\}} (2j - 1 + |M - N|)\, v_j \right\} \tag{28}$$

where $\mathcal{O}_\ell$ is the set defined by

$$\mathcal{O}_\ell = \left\{ \boldsymbol{v} \in \mathbb{R}^{\min\{M,N\}},\ v_1 \geq \cdots \geq v_{\min\{M,N\}} : \sum_{j=1}^{\min\{M,N\}} \left[\max_{k=1,\ldots,\ell} \left\{ \sum_{i=1}^{k} \xi_{\ell-i} + k(1 - v_j) \right\} \right]_+ \leq r_e \right\}. \tag{29}$$

Moreover, the exponent $d^*(r_e, L)$ is achievable by finite block-length codes if $T \geq M + N - 1$.

Proof: See Appendix B ∎

The more stringent requirement on T in the long-term static case of Theorem 5, as compared with Theorem 2, can be explained as follows. In Theorem 2, we only require the probability of error after one round of transmission to decay with the SNR (at any rate) to ensure that $r_e = r_1$. In Theorem 5, on the other hand, we need to maximize the rate of decay of the first round probability of error in order to maximize the power level in the second round. In Appendix B, we show that $T = M + N - 1$ is sufficient to achieve this goal.

For any $0 = \xi_0 \leq \xi_1 \leq \cdots \leq \xi_L$, and any $\ell = 1, \ldots, L$, the function

$$g_\ell(z) = \left[\max_{k=1,\ldots,\ell} \left\{ \sum_{i=1}^{k} \xi_{\ell-i} + k(1-z) \right\} \right]_+ \tag{30}$$

is convex, decreasing, piecewise linear with bounded support $[0, \xi_{\ell-1} + 1]$. Its maximum is attained at $z = 0$ and is given by

$$g_\ell(0) = \sum_{i=1}^{\ell-1} \xi_i + \ell.$$

It follows that the set defined by $\{\boldsymbol{v} \in \mathbb{R}_+^m : \sum_{j=1}^m g_\ell(v_j) \leq r_e\}$ is convex and bounded. Since the objective function in (28) is linear and hence convex, each of the minimizations in (28) has a well-defined unique solution that can be easily found by standard numerical optimization methods.

Unfortunately, at the moment we do not have a closed-form characterization of the optimal tradeoff curve in Theorem 5. To shed more light on the power control diversity gain, we derive easily computable lower and upper bounds on the optimal diversity gain $d^*(r_e, L)$ in the following lemma.

Lemma 6: Let $d^*(r_e, L)$ denote the optimal diversity gain under long-term power constraint given by Theorem 5. Then

$$\max\left(d_L^{(lb1)}, d_L^{(lb2)}\right) \leq d^*(r_e, L) \leq d_L^{(ub)} \tag{31}$$

where $d_L^{(lb1)}$, $d_L^{(lb2)}$, and $d_L^{(ub)}$ are obtained recursively as follows. Let

$$d_1^{(lb1)} = d_1^{(lb2)} = d_1^{(ub)} = f(r_e). \tag{32}$$

Then, for $\ell = 2, \ldots, L$ let

$$d_\ell^{(lb1)} = \left(1 + d_{\ell-1}^{(lb1)}\right) f\left(\frac{r_e}{1 + d_{\ell-1}^{(lb1)}}\right) \tag{33}$$

$$d_\ell^{(lb2)} = \frac{\ell + \sum_{k=1}^{\ell-1} d_k^{(lb2)}}{\ell} f\left(\frac{r_e}{\ell + \sum_{k=1}^{\ell-1} d_k^{(lb2)}}\right) \tag{34}$$

and

$$d_\ell^{(ub)} = \left(1 + d_{\ell-1}^{(ub)}\right) f\left(\frac{r_e}{\ell + \sum_{i=1}^{\ell-1} d_i^{(ub)}}\right). \tag{35}$$

Proof: See Appendix C. ∎

The lower bounds established in Lemma 6 have nice intuitive interpretations. The first lower bound, i.e., $d_\ell^{(lb_1)}$, corresponds to the outage probability achieved by only the round with the maximum power level. As a side result, this lower bound also corresponds to the diversity–multiplexing tradeoff of the power control algorithm proposed in [15] where the authors assume one round of transmission and the availability of the feedback information, needed for the power control algorithm, *a priori* (in this setting, L takes the meaning of the number of levels in the power control algorithm). The second lower bound, i.e., $d_\ell^{(lb_2)}$, corresponds to averaging the power levels[7] used in the ℓ ARQ rounds and then deriving the tradeoff under the assumption that this level is used in all the ℓ rounds. Fig. 3 depicts the upper and lower bounds on the optimal diversity advantage with power control. One can see in the figure the significant gain offered through power control, compared to ARQ with constant power, especially at low multiplexing gains. In fact, the remarkably large diversity gains observed for **all multiplexing gains** even with relatively small values of L indicates that very slow fading channels quickly approach the ergodic limit when ARQ and power control are used **jointly**. This phenomenon does not appear when only power control is used without ARQ retransmissions, as in [15] for example, since in this case the diversity advantage still approaches zero as the multiplexing gain approaches its maximum value of $\min(N, M)$. Moreover, at least in this scenario, it appears that the lower and upper bounds are very tight for a wide range of multiplexing gains.

IV. IR-LAST Coding

Thus far, our information-theoretic analysis has relied on using random Gaussian codes. In practice, the complexity resulting from using such *unstructured* codebooks may be prohibitive. Here, we replace the Gaussian codes with IR-LAST codes, the bounded distance decoder with a fixed radius list lattice decoder, and the maximum-likelihood (ML) decoder used in the final round with a closest point lattice decoder. We will show that this approach achieves the optimal tradeoff (with and without power control) for $T \geq M + N - 1$.[8] Furthermore, the simulation results, presented at the end of the section, will demonstrate the significant performance gains offered by this approach in certain representative scenarios.

In [7], we introduced the class of nested LAST codes and showed that it achieves the optimal diversity–multiplexing tradeoff in coherent MIMO channels. Here, we extend this paradigm to the MIMO ARQ scenario. For the sake of completeness, we review the basic definitions needed to describe the IR-LAST coding scheme. For more background information, the interested reader is referred to [7] and references therein. To simplify presentation, we focus on the constant-power scenario. When power control is allowed, the power allocation algorithm is combined with the code construction straightforwardly.

An m-dimensional lattice code $\mathcal{C}(\Lambda, \boldsymbol{u}_0, \mathcal{R})$ is the finite subset of the lattice translate $\Lambda + \boldsymbol{u}_0$ inside the *shaping region* $\mathcal{R}$, i.e., $\mathcal{C} = \{\Lambda + \boldsymbol{u}_0\} \cap \mathcal{R}$, where $\mathcal{R}$ is a bounded measurable

[7] Here, we average the power computed on a logarithmic scale.

[8] In a similar way, one can establish the fact that the minimum length needed for the long-term static channels with constant power is $T \geq \lceil \frac{M+N-1}{L} \rceil$ but we omit this part here to avoid redundancy.

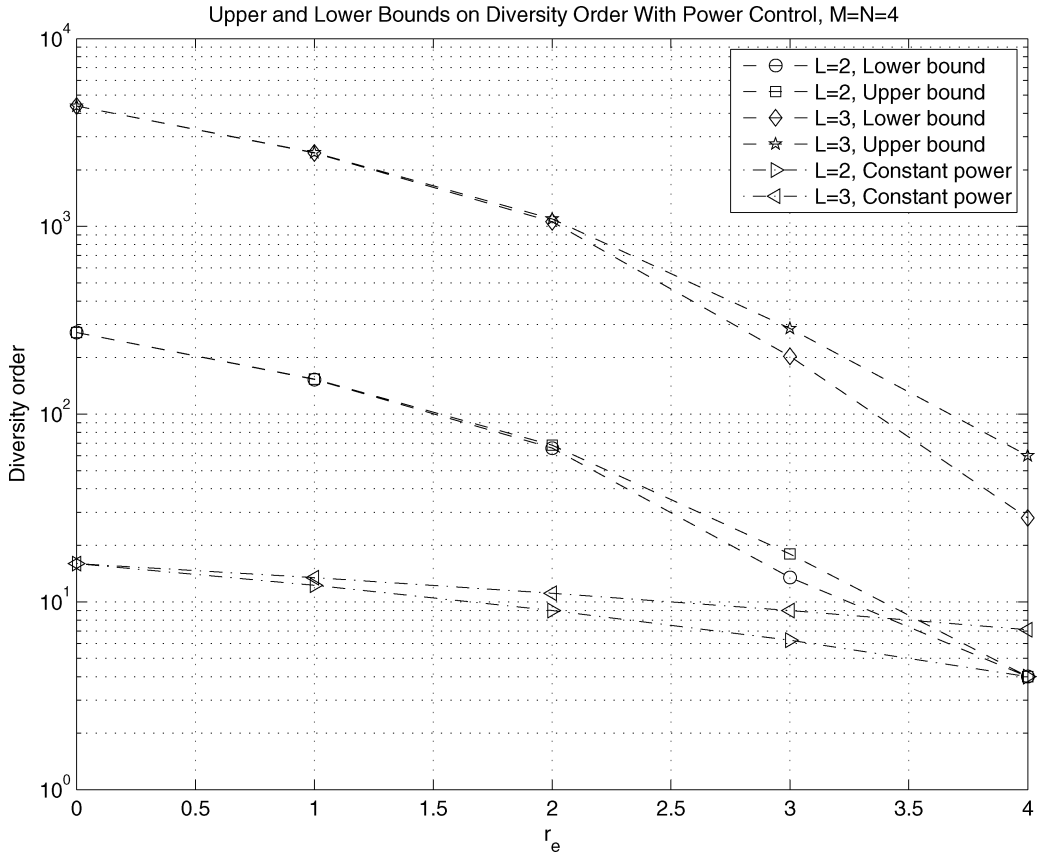

Fig. 3. The diversity–multiplexing tradeoff with power control on a log-scale (the upper and lower bound in Lemma 6).

region of $\mathbb{R}^m$. We say that a space–time coding scheme is a LAST code if its codebook is a lattice code. Next, we define nested lattice codes (or Voronoi codes).

Definition 7: Let Λ_c be a lattice in $\mathbb{R}^m$ and Λ_s be a sublattice of Λ_c. The nested lattice code defined by the partition Λ_c/Λ_s is given by

$$\mathcal{C} = \Lambda_c \cap \mathcal{V}_s$$

where $\mathcal{V}_s$ is the fundamental Voronoi cell of Λ_s. In other words, $\mathcal{C}$ is formed by the coset leaders of the cosets of Λ_s in Λ_c. We also define the lattice quantization function

$$Q_\Lambda(\boldsymbol{y}) \triangleq \arg \min_{\boldsymbol{\lambda}\in\Lambda} |\boldsymbol{y}-\boldsymbol{\lambda}|$$

and the modulo-lattice function

$$[\boldsymbol{y}] \mod \Lambda \triangleq \boldsymbol{y} - Q_\Lambda(\boldsymbol{y}). \qquad \diamondsuit$$

We say that a LAST code is nested if the underlying lattice code is nested. With nested codes, the information message is effectively encoded into the cosets of Λ_s in Λ_c.

The proposed incremental redundancy scheme works as follows. Consider the nested LAST code $\mathcal{C}$ defined by Λ_c (the *coding lattice*) and by its sublattice Λ_s (the *shaping lattice*) in $\mathbb{R}^{2MTL}$. Assume that Λ_s has a second-order moment $\sigma^2(\Lambda_s) = 1/2$ (so that $\boldsymbol{u}$ uniformly distributed over $\mathcal{V}_s$ satisfies $\mathbb{E}[|\boldsymbol{u}|^2] = MTL$). Assuming an effective multiplexing gain r_e, the rate of the code is $R = r_e \log(\rho)/L$. The transmitter selects a codeword $\boldsymbol{c} \in \mathcal{C}$, generates a dither signal $\boldsymbol{u}$ with uniform distribution over $\mathcal{V}_s$, and computes

$$\boldsymbol{x} = [\boldsymbol{c} - \boldsymbol{u}] \mod \Lambda_s. \tag{36}$$

The signal $\boldsymbol{x}$ is then partitioned into L vectors of size $2MT$ each. Those vectors are transmitted, sequentially, in the different ARQ rounds based on the ACK/NACK feedback. Upon completion of the $\ell < L$ transmission, the receiver attempts to decode the message using an incomplete list lattice decoder. In particular, the received signal, i.e., $\boldsymbol{y}_\ell$, is multiplied by the forward filter matrix $\boldsymbol{F}_\ell$ of the minimum mean-square-error decision feedback equalizer (MMSE-DFE) corresponding to the truncated matrix $\boldsymbol{H}_\ell$ [26]. Moreover, we add the dither signal filtered by the upper triangular feedback filter matrix $\boldsymbol{B}_\ell$ of the MMSE-DFE (the definitions and some useful properties of the MMSE-DFE matrices $(\boldsymbol{F}_\ell, \boldsymbol{B}_\ell)$ are given in [7]).

By construction, we have $\boldsymbol{x} = \boldsymbol{c} - \boldsymbol{u} + \boldsymbol{\lambda}$ with $\boldsymbol{\lambda} = -Q_{\Lambda_s}(\boldsymbol{c}-\boldsymbol{u})$. Then, we can write

$$\begin{aligned}
\boldsymbol{y}'_\ell &= \boldsymbol{F}_\ell \boldsymbol{y}_\ell + \boldsymbol{B}_\ell \boldsymbol{u} \\
&= \boldsymbol{F}_\ell \left(\boldsymbol{H}_\ell(\boldsymbol{c} - \boldsymbol{u} + \boldsymbol{\lambda}) + \boldsymbol{w}_\ell\right) + \boldsymbol{B}_\ell \boldsymbol{u} \\
&= \boldsymbol{B}_\ell(\boldsymbol{c} + \boldsymbol{\lambda}) - [\boldsymbol{B}_\ell - \boldsymbol{F}_\ell \boldsymbol{H}_\ell](\boldsymbol{c} - \boldsymbol{u} + \boldsymbol{\lambda}) + \boldsymbol{F}_\ell \boldsymbol{w}_\ell \\
&= \boldsymbol{B}_\ell(\boldsymbol{c} + \boldsymbol{\lambda}) - [\boldsymbol{B}_\ell - \boldsymbol{F}_\ell \boldsymbol{H}_\ell]\boldsymbol{x} + \boldsymbol{F}_\ell \boldsymbol{w}_\ell \\
&= \boldsymbol{B}_\ell(\boldsymbol{c} + \boldsymbol{\lambda}) + \boldsymbol{e}'.
\end{aligned} \tag{37}$$

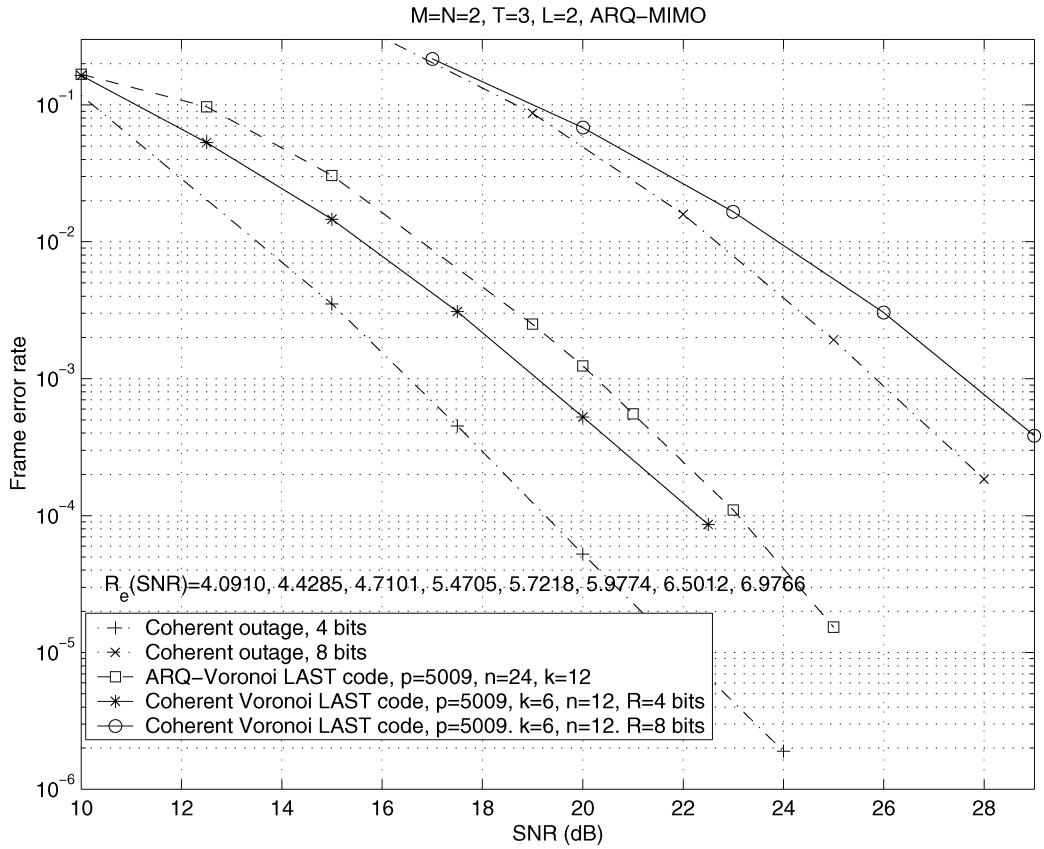

Fig. 4. The probability of error of incremental redundancy LAST codes.

By construction, $\boldsymbol{x}$ is uniformly distributed over $\mathcal{V}_s$ and is independent of $\boldsymbol{c}$. One can also rewrite (37) as

$$\boldsymbol{y}'_\ell = \boldsymbol{B}_\ell \boldsymbol{c}' + \boldsymbol{e}' \tag{38}$$

where $\boldsymbol{c}' \in \Lambda_s + \boldsymbol{c}$ and

$$\mathbb{E}\left[\boldsymbol{e}'\boldsymbol{e}'^{\mathsf{T}}\right] = \frac{1}{2}\boldsymbol{I}. \tag{39}$$

The desired signal $\boldsymbol{c}$ is now translated by an unknown lattice point $\boldsymbol{\lambda} \in \Lambda_s$. However, since $\boldsymbol{c}$ and $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{\lambda}$ belong to the same coset of Λ_s in Λ_c, this translation does not involve any loss of information (recall that information is encoded in the coset $\Lambda_s + \boldsymbol{c}$, rather than in the codeword $\boldsymbol{c}$ itself). It follows that in order to recover the information message, the decoder has to identify the coset $\Lambda_s + \boldsymbol{c}$ that contains $\boldsymbol{c}'$.

The basic idea in this approach is to use a list lattice decoder for joint error correction and detection. In this decoder, we first check if the channel is in outage. In this case, an error is declared and a NACK bit is sent back. If not, then we use a list lattice decoder to find all the lattice points that satisfy

$$\left\{\boldsymbol{z} \in \mathbb{Z}^{2MTL} : \left|\boldsymbol{y}' - \boldsymbol{B}_\ell \boldsymbol{G}\boldsymbol{z}\right|^2 \leq MTL(1 + \beta \log(\rho))\right\} \tag{40}$$

where $\boldsymbol{G}$ is the generator matrix of the channel coding lattice Λ_c, and β is chosen according to the proof of Theorem 8. Now, if no points are found or more than one point is found, an error is declared, and hence, a NACK bit is sent back. If only one point is found to satisfy (40), then we proceed to the next step to find the codeword as

$$\hat{\boldsymbol{c}} = [\boldsymbol{G}\hat{\boldsymbol{z}}] \mod \Lambda_s. \tag{41}$$

Here, we observe that the matrix $\boldsymbol{B}_\ell$ is always full rank even for the under-determined scenario $\ell < L$. This property is very critical for minimizing the complexity of the closest point search algorithm [27]. The only exception to this rule is after the L ARQ round where we replace this joint error correction and detection algorithm with the closest point lattice decoder described by

$$\hat{\boldsymbol{z}} = \operatorname{argmin} \min_{\boldsymbol{z} \in \mathbb{Z}^{2MTL}} \left|\boldsymbol{y}' - \boldsymbol{B}_L \boldsymbol{G}\boldsymbol{z}\right|^2. \tag{42}$$

The following result establishes the optimality of this approach for $T \geq M + N - 1$.

Theorem 8: Consider a long-term static MIMO ARQ channel with M transmit, N receive antennas, a maximum number of ARQ rounds L, an effective multiplexing gain $0 \leq r_e < \min(N, M)$, and $T \geq M + N - 1$. Then, the proposed IR-LAST coding scheme achieves the optimal diversity advantage $d^*(r_e, L)$ in Theorem 2 under the short-term average power constraint. Under the long-term power constraint, the IR-LAST coding scheme achieves the optimal diversity advantage $d^*(r_e, L)$ in Theorem 5 when coupled with the power control policy

$$\Gamma_\ell = \frac{MT \sum_{\ell=0}^{L-1} p(\ell)}{Lp(\ell - 1)}. \tag{43}$$

Fig. 4 compares the performance of the proposed IR-LAST coding scheme with the outage probability and the performance of the LAST coding scheme for the standard coherent channel in a long-term static channel. We have $M = N = L = 2, T = 3$,

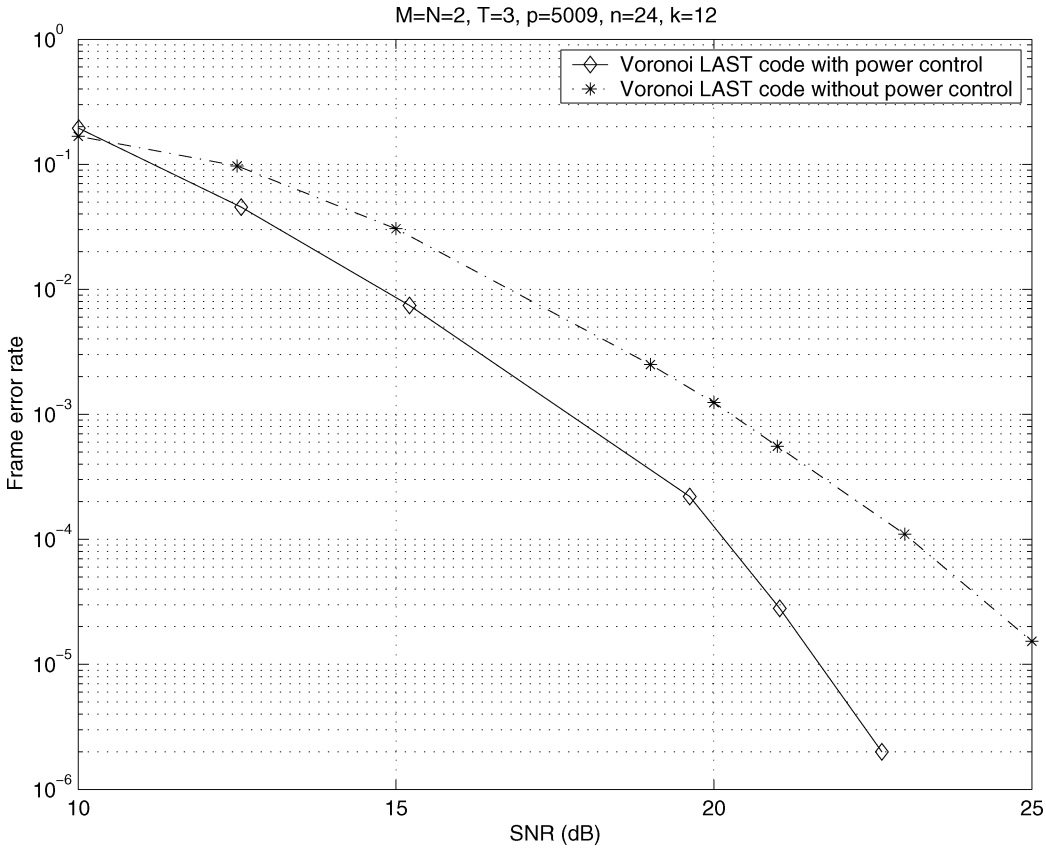

Fig. 5. The probability of error of incremental redundancy LAST codes with the asymptotically optimal power control algorithm.

and $R_1 = 8$ bits per channel use. The LAST code is obtained as an (n, k, p) Loeliger construction (please refer to [7] for a detailed description). In the coherent case, we report the performance with $R = 4$ and $R = 8$ bits per channel use. We also report the effective rate of the ARQ scheme, denoted by R_e, at the eight sampling points indicated on the curve. The IR-LAST coding scheme is shown to achieve probability of error very close to the coherent LAST code with $R = 4$. On the other hand, the effective rate of the IR-LAST coding scheme is shown to approach $R_1 = 8$ as the SNR grows. Overall, this results in a performance gain, compared to coherent systems with the same average rate, that increases with the SNR as predicted by the theory. Fig. 5 demonstrates the gain offered by the proposed power control policy. In this figure, we augment the IR-LAST coding scheme used in Fig. 4 with the power control strategy of Theorem 8. The power control diversity gain manifests itself in the steeper slope of the probability of error curve. Here, we remark that the proposed power control policy is only guaranteed to attain the optimal asymptotic slope of the probability of error curve. Therefore, there is still room for further optimization of the power control strategy to minimize the probability of error at small-to-moderate SNR.

V. Conclusion

In this paper, we investigated the fundamental tradeoff of MIMO ARQ channels. We have shown that the ARQ retransmission delay can be leveraged for significant gains in the diversity advantage. By characterizing the three-dimensional diversity–multiplexing–delay tradeoff, we have quantified this ARQ diversity gain. Our results show that, with the short-term power constraint, the ARQ diversity gain is significant only at high multiplexing gains. This limitation is overcome by combining the retransmission strategy with a carefully constructed power control policy, that allocates the power in the ℓth round to be inversely proportional to the probability of having to transmit ℓ rounds. In this way, very high power levels can be used to "correct" the very rare error events which determine the high-SNR behavior of error probability. We showed that the diversity gain achieved by ARQ with power control is dramatically large at all multiplexing gains, so that the performance approaches rapidly the ergodic (no-outage) behavior, according to which the multiplexing gain $\min\{M, N\}$ can be achieved with arbitrarily large reliability. Finally, we presented an IR-LAST explicit coding scheme which achieves the optimal tradeoff curve (with and without power control). In this scheme, the list lattice decoder emerged as a powerful tool for joint error correction and detection.

Overall, our work established a theoretical foundation for evaluating previously proposed MIMO ARQ schemes and, hopefully, inspiring more innovative approaches. For example, our approach for achieving the optimal diversity–multiplexing–delay tradeoff highlights the importance of incremental redundancy schemes coupled with list decoders for joint error correction and detection. The optimality of the proposed list decoder, however, is only limited to the high-SNR regime. An interesting venue for future work is, therefore, to design

more sophisticated decoders inspired by the elegant framework of [24].

APPENDIX I
PROOF OF THEOREM 2

We start by considering the long-term static channel. Let $\frac{1}{T}I_{\boldsymbol{H}^c}(\boldsymbol{x};\boldsymbol{y}_\ell)$ denote the mutual information per channel use over ℓ consecutive slots for a given channel matrix realization $\boldsymbol{H}^c$, where $\boldsymbol{x}$ is the vectorized input codeword and $\boldsymbol{y}_\ell$ is the corresponding ℓ-slot channel output as defined in (4).

In order to derive the upper bound on $d^*_{ls}(r_e, L)$ we consider a system that accumulates mutual information over consecutive slots and compares it with a threshold $R_1 = r_1 \log\rho$. If mutual information is larger than the threshold or if a maximum number L of slots is reached, the system resets and both the slot index and the mutual information count are restarted anew. Under the assumption that $\boldsymbol{H}^c$ changes in an i.i.d. fashion each time the system resets, the event of resetting is a renewal event and the results developed in Section II-B apply directly, by redefining the event $\mathcal{A}_\ell$ as the mutual information level-crossing event

$$\mathcal{A}_\ell = \{\boldsymbol{H}^c \in \mathbb{C}^{N\times M} : \frac{1}{T} I_{\boldsymbol{H}^c}(\boldsymbol{x};\boldsymbol{y}_\ell) > R_1\}.$$

We define the *information outage* event with ℓ received blocks as $\mathcal{O}(\rho,\ell) = \overline{\mathcal{A}}_\ell$, with the associated *outage probability* $P_{\rm out}(\rho,\ell) = \Pr(\mathcal{O}(\rho,\ell))$ where, by definition, $P_{\rm out}(\rho,0) = 1$. Outage probability is minimized, for every given SNR ρ, by choosing $\boldsymbol{x}$ i.i.d. in time and such that $\boldsymbol{x}^c_{\ell,t} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{Q}_\ell)$ for some covariance matrix $\boldsymbol{Q}_\ell$ such that $\mathrm{tr}(\boldsymbol{Q}_\ell) \le M$. It is straightforward to show that the outage probability minimized with respect to the input covariance matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_\ell$ satisfies the bounds [6]

$$\begin{aligned}&\Pr\left(\ell \log\,\det\left(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}^c\boldsymbol{H}^{c\mathsf{H}}\right) \le R_1\right)\\ &\ge \min_{\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_\ell} P_{\rm out}(\rho,\ell) \ge \Pr\left(\ell \log\,\det\left(\boldsymbol{I} + \rho\boldsymbol{H}^c\boldsymbol{H}^{c\mathsf{H}}\right) \le R_1\right)\end{aligned} \tag{44}$$

obtained by choosing $\boldsymbol{Q}_\ell = \boldsymbol{I}$ (lower bound) and $\boldsymbol{Q}_\ell = M\boldsymbol{I}$ (upper bound) for all ℓ. It follows that the optimal outage probability and the outage probability achieved by i.i.d. Gaussian inputs $\boldsymbol{x}^c_{\ell,t} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{I})$ have the same exponential order with respect to ρ and thus, for the sake of establishing the high-SNR behavior, it suffices to consider $P_{\rm out}(\rho,\ell)$ defined for such white Gaussian input distribution.

We denote by $d_{\rm out}(\ell)$ the SNR exponent of the ℓth round outage probability, namely,

$$d_{\rm out}(\ell) = \lim_{\rho\to\infty} \frac{-\log\left(P_{\rm out}(\rho,\ell)\right)}{\log(\rho)}. \tag{45}$$

It follows immediately from the results in [6] that

$$d_{\rm out}(\ell) = f\left(\frac{r_1}{\ell}\right) \tag{46}$$

where $f(\cdot)$ is the piecewise linear function defined in Theorem 1.

Now, consider any given MIMO ARQ system operating at SNR ρ, with given block-length T, codebook $\mathcal{C}_\rho$, first-block rate $r_1\log\rho$, and some decoding rule $\phi = (\phi_1, \ldots, \phi_L)$ such that, for all $\ell = 1, \ldots, L$, $\phi_\ell : \mathbb{R}^{2NT\ell} \to \{0, 1, \ldots, \rho^{r_1 T}\}$, and the decoded message at round ℓ is given by $\widehat{w} = \phi_\ell(\boldsymbol{y}_\ell)$. Message 0 corresponds to "error detection": if $\phi_\ell(\boldsymbol{y}_\ell) = 0$ a NACK is sent back to the transmitter. Notice that ϕ_ℓ takes as arguments both $\boldsymbol{y}_\ell$ and $\boldsymbol{H}^c$, since we assume that the channel matrix is known to the receiver. However, we omit the second argument for notation simplicity and since it is clear from the context.

We wish to show that $d_{\rm out}(L)$ defined in (46) is an upper bound to the SNR exponent of any such sequence of MIMO ARQ systems. Letting w denote the transmitted information message, uniformly distributed over $\{1, \ldots, \rho^{r_1 T}\}$, and recalling the general expression of the error probability (13), we can write the conditional error probability of the scheme $(\mathcal{C}_\rho, \phi)$ for given $\boldsymbol{H}^c$ as (47) at the bottom of the page. By definition, the error probability in (47) is lower-bounded by the probability of error of the optimal ML decoder $\phi_{\rm ml}$ that operates on the whole received signal vector $\boldsymbol{y} = \boldsymbol{y}_L$ knowing the channel matrix $\boldsymbol{H}^c$. Hence, Fano inequality yields [6]

$$\begin{aligned}P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho, \phi) &\ge P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho, \phi_{\rm ml})\\ &\ge 1 - \frac{1}{r_1 T\log\rho} I_{\boldsymbol{H}^c}(\boldsymbol{x};\boldsymbol{y}) - \frac{1}{r_1 T\log\rho}.\end{aligned} \tag{48}$$

Following the same steps as those in the proof of Theorem 2 in [6], it is a simple matter to show that, for any MIMO ARQ system

$$P_e(\rho|\mathcal{C}_\rho, \phi) = \mathbb{E}[P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho, \phi)] \dot{\ge} \rho^{-d_{\rm out}(L)}. \tag{49}$$

$$\begin{aligned}P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho, \phi) = &\sum_{\ell=1}^{L-1} \Pr\left(\{\phi_1(\boldsymbol{y}_1) = 0\}, \ldots, \{\phi_{\ell-1}(\boldsymbol{y}_{\ell-1}) = 0\}, \bigcup_{\substack{\widehat{w}\ne w\\ \widehat{w}>0}} \{\phi_\ell(\boldsymbol{y}_\ell) = \widehat{w}\} \,\middle|\, \boldsymbol{H}^c\right)\\ &+ \Pr\left(\{\phi_1(\boldsymbol{y}_1) = 0\}, \ldots, \{\phi_{L-1}(\boldsymbol{y}_{L-1}) = 0\}, \bigcup_{\widehat{w}\ne w} \{\phi_L(\boldsymbol{y}_L) = \widehat{w}\} \,\middle|\, \boldsymbol{H}^c\right).\end{aligned} \tag{47}$$

Noticing that, for any MIMO ARQ system, $r_e \leq r_1$ and $f(\cdot)$ is nonincreasing and by using (46), we eventually obtain the upper bound $d_{ls}^*(r_e, L) \leq f(r_e/L)$ as desired.

The achievability of the exponent upper bound for asymptotically large T is shown as follows. For each value of ρ, consider a sequence of MIMO ARQ systems with first-block rate $R_1(\rho) = r_1 \log \rho$, codebook $\mathcal{C}_\rho$ with randomly generated codewords $\boldsymbol{x} \in \mathbb{R}^{2MTL}$ with i.i.d. components $\sim \mathcal{N}(0, 1/2)$, and increasing block length T. Let ϕ be the typical-set decoder defined by the following decision rule:

1. $\phi_\ell(\boldsymbol{y}_\ell) = \widehat{w} > 0$ if $\boldsymbol{H}^c \notin \mathcal{O}(\rho, \ell)$ and the codeword corresponding to $\widehat{w}$ is the unique codeword in $\mathcal{C}_\rho$ jointly typical with the output $\boldsymbol{y}_\ell$ over slots 1 to ℓ.
2. $\phi_\ell(\boldsymbol{y}_\ell) = 0$ in any other case.

We use the upper bound to the MIMO ARQ error probability given by (13) where, for the typical-set decoder defined above, the decoding error event $\mathcal{E}_\ell$ is expressed in terms of ϕ as

$$\mathcal{E}_\ell = \{\phi_\ell(\boldsymbol{y}_\ell) \neq w\} \tag{50}$$

and the event of sending an ACK at round ℓ is given by

$$\mathcal{A}_\ell = \{\phi_\ell(\boldsymbol{y}_\ell) \neq 0\}. \tag{51}$$

Then, we have

$$P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho, \phi) \leq \sum_{\ell=1}^{L-1} \Pr\left(\bigcup_{\substack{\widehat{w} \neq w \\ \widehat{w} > 0}} \{\phi_\ell(\boldsymbol{y}_\ell) = \widehat{w}\} \middle| \boldsymbol{H}^c, \mathcal{C}_\rho \right) + \Pr\left(\{\phi_L(\boldsymbol{y}_L) \neq w\} | \boldsymbol{H}^c, \mathcal{C}_\rho \right). \tag{52}$$

Following in the footsteps of [23, Appendix A], it is immediate to show that for each $\rho, \epsilon > 0$, and sufficiently large T, there exists a code $\mathcal{C}_\rho^*$ such that

$$\Pr\left(\bigcup_{\substack{\widehat{w} \neq w \\ \widehat{w} > 0}} \{\phi_\ell(\boldsymbol{y}_\ell) = \widehat{w}\} \middle| \boldsymbol{H}^c, \mathcal{C}_\rho^* \right) < \epsilon \tag{53}$$

and

$$\Pr\left(\{\phi_L(\boldsymbol{y}_L) \neq w\} | \boldsymbol{H}^c, \mathcal{C}_\rho^* \right) < \epsilon + 1\{\boldsymbol{H}^c \in \mathcal{O}(\rho, L)\}. \tag{54}$$

Without repeating here the details of [23, Appendix A], we shall just illustrate qualitatively the above result: inequality (53) follows from the fact that the event of undetected decoding error is contained in the event that the input and the output of the channel are not jointly typical, whose probability is vanishing for large T; (54) follows from the existence of codes with arbitrarily small error probability for all fading matrices in the nonoutage set.

It follows that, for sufficiently large T

$$P_e(\rho|\boldsymbol{H}^c, \mathcal{C}_\rho^*, \phi) \leq L\epsilon + 1\{\boldsymbol{H}^c \in \mathcal{O}(\rho, L)\}$$

and by taking expectation of both sides with respect to $\boldsymbol{H}^c$ we obtain

$$P_e(\rho|\mathcal{C}_\rho^*, \phi) \doteq P_{\text{out}}(\rho, L) \doteq \rho^{-f(r_1/L)}. \tag{55}$$

On the other hand, for such a family of MIMO ARQ schemes $(\mathcal{C}_\rho^*, \phi)$ we have that, for all $1 \leq \ell \leq L-1$, for all $\epsilon > 0$, and for sufficiently large T

$$\begin{aligned} p(\ell) &\leq \Pr(\overline{\mathcal{A}}_\ell) \\ &= \Pr\left(\{\phi_\ell(\boldsymbol{y}_\ell) = 0\}\right) \\ &\overset{(a)}{\leq} \Pr\left(\boldsymbol{H}^c \in \mathcal{O}(\rho, \ell)\right) + \epsilon \\ &= P_{\text{out}}(\rho, \ell) + \epsilon \end{aligned} \tag{56}$$

where (a) follows form the fact that, for T large enough, the probability that there are more than one codeword jointly typical with the output can be made as small as desired for all $\boldsymbol{H}^c \notin \mathcal{O}(\rho, \ell)$, therefore, the event $\{\phi_\ell(\boldsymbol{y}_\ell) = 0\}$ is essentially given by the information outage event. Therefore, $p(\ell) \dot{\leq} \rho^{-d_{\text{out}}(\ell)}$. Using (46), we obtain

$$\eta = \frac{R_1}{1 + \sum_{\ell=1}^{L-1} \rho^{-f(r_1/\ell)}} \doteq R_1$$

which results in $r_e = r_1$. This, together with (55), proves that $f(r_e/L)$ is achievable for sufficiently large block length T.

The proof for the short-term static channels follows the same lines with the exception that the mutual information with i.i.d. Gaussian inputs takes on the expression

$$\frac{1}{T} I_{\boldsymbol{H}_1^c, \ldots, \boldsymbol{H}_\ell^c}(\boldsymbol{x}; \boldsymbol{y}_\ell) = \sum_{j=1}^{\ell} \log \det \left(\boldsymbol{I} + \frac{\rho}{M} \boldsymbol{H}_j^c \boldsymbol{H}_j^{c\mathsf{H}} \right)$$

so that (45) is replaced by

$$P_{\text{out}}(\rho, \ell) \doteq \rho^{-\ell f(\frac{r_1}{\ell})}. \tag{57}$$

This concludes the proof for achievability with $T \to \infty$. For finite T, we first assume long-term static channels. As before, the result for short-term static channels follows easily and the difference between (19) and (20) will be explained toward the end of the proof. The proof is composed of two steps. First, we consider an ensemble of Gaussian i.i.d. random codes with block length T and analyze their error probability and their throughput in the ensemble average sense. Second, we have to show via a simple expurgation argument that there are codes in the ensemble that perform at least as well as the ensemble average and thus achieve the same error probability and throughput.

Let $\mathcal{C}_\rho$ denote a random code generated with i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$ components, block length LT, and rate $r_1 \log \rho$. We define the following bounded distance decoder ϕ: at each round $\ell \leq L-1$

1. $\phi_\ell(\boldsymbol{y}_\ell) = \widehat{w} > 0$ if $\boldsymbol{H}^c \notin \mathcal{O}(\rho, \ell)$ and the codeword $\widehat{\boldsymbol{x}}$ (corresponding to $\widehat{w}$) is the unique codeword in $\mathcal{C}_\rho$ such that $|\boldsymbol{y}_\ell - \boldsymbol{H}_\ell \widehat{\boldsymbol{x}}|^2 \leq NT\ell(1+\delta)$, where $\delta > 0$ will be specified later;
2. $\phi_\ell(\boldsymbol{y}_\ell) = 0$ in any other case;
3. at round L, the decoder outputs the index of the minimum distance codeword, i.e., $\phi_L(\boldsymbol{y}_L) = \phi_{\text{ml}}(\boldsymbol{y}_L)$.

For the above ensemble, we wish to analyze the probability of error and the throughput. For the probability of error we bound each term in (13). Since the decoder at round L is the standard ML decoder, using the results of [6] we obtain immediately

$$\Pr(\mathcal{E}_L) \doteq \rho^{-f(r_1/L)} \tag{58}$$

under the condition that $LT \geq M+N-1$. In order to bound the undetected error probability $\Pr(\mathcal{E}_\ell, \mathcal{A}_\ell)$, let $\mathcal{D}_{\widehat{w}} \subseteq \mathbb{R}^{2NT\ell}$ be the region of received signal vectors $\boldsymbol{y}_\ell$ such that $\widehat{\boldsymbol{x}}$ is the unique codeword in $\mathcal{C}_\rho$ for which $|\boldsymbol{y}_\ell - \boldsymbol{H}_\ell \widehat{\boldsymbol{x}}|^2 \leq NT\ell(1+\delta)$. Then, we have

$$\begin{aligned}\Pr(\mathcal{E}_\ell, \mathcal{A}_\ell) &= \Pr\left(\bigcup_{\widehat{w}\neq w} \{\boldsymbol{y}_\ell \in \mathcal{D}_{\widehat{w}}\}\right) \\ &\leq \Pr\left(|\boldsymbol{w}_\ell|^2 \geq NT\ell(1+\delta)\right)\end{aligned} \tag{59}$$

where the inequality follows by noticing that the union of all $\mathcal{D}_{\widehat{w}}$ is included in the complement of the sphere centered in $\boldsymbol{x}$ (corresponding to the transmitted message w) of squared radius $NT\ell(1+\delta)$. We notice that $|\boldsymbol{w}_\ell|^2$ is central Chi-squared with $2NT\ell$ degrees of freedom. We can use the Chernoff bound to upper-bound the tail of the Chi-squared distribution, and find

$$\begin{aligned}&\Pr\left(|\boldsymbol{w}_\ell|^2 \geq NT\ell(1+\delta)\right) \\ &\quad\leq \min_{\lambda\geq 0} \exp\left(-NT\ell(\lambda(1+\delta)+\log(1-\lambda))\right) \\ &\quad= (1+\delta)^{NT\ell}\exp(-NT\ell\delta).\end{aligned} \tag{60}$$

For some $\beta > 0$, we let $\delta = \beta \log\rho$ and obtain

$$\Pr(\mathcal{E}_\ell, \mathcal{A}_\ell) \doteq \rho^{-NT\ell\beta}. \tag{61}$$

Eventually, the ensemble average error probability is given by

$$P_e(\rho) \doteq \rho^{-NT\beta} + \rho^{-f(r_1/L)} \doteq \rho^{-\min\{NT\beta, f(r_1/L)\}}. \tag{62}$$

In order to have the desired exponent $f(r_1/L)$, we need to ensure that $NT\beta \geq f(r_1/L)$. By choosing a large enough β, we can easily see that this is achieved under the condition on T given by (19).

In order to achieve $d^*(r_e, L) = f(r_e/L)$, we still need to show that $r_e = r_1$, i.e., that the probabilities $p(\ell)$ are $o(1)$ for large ρ. Fix $1 \leq \ell \leq L-1$. We partition the channel output space (formed by all possible received vectors $\boldsymbol{y}_\ell$ and channel matrices $\boldsymbol{H}_\ell$) into the following regions: $\mathcal{O}(\rho,\ell)$ is the usual outage event, $\mathcal{R}_0$ is the region of channel outputs not included in any of the spheres of squared radius $NT\ell(1+\delta)$ and centered around the codewords, and $\mathcal{R}_1$ is the region of channel outputs included in more than one of such spheres. Moreover, we partition $\mathcal{R}_1$ into $\mathcal{R}_{1,w}$ and $\mathcal{R}_{1,\overline{w}}$, where the former is the region of the sphere centered in $\boldsymbol{x}$ (the transmitted codeword, corresponding to w) included in other spheres, and the latter is its complement in $\mathcal{R}_1$. Then, we have

$$\begin{aligned}p(\ell) &\leq \Pr(\overline{\mathcal{A}}_\ell) \\ &= \Pr\left(\mathcal{O}(\rho,\ell) \cup \mathcal{R}_0 \cup \mathcal{R}_1\right) \\ &= \Pr\left(\mathcal{O}(\rho,\ell)\right) + \Pr\left(\overline{\mathcal{O}}(\rho,\ell) \cap \{\mathcal{R}_0 \cup \mathcal{R}_1\}\right) \\ &\leq \Pr\left(\mathcal{O}(\rho,\ell)\right) + \Pr\left(|\boldsymbol{w}_\ell|^2 \geq NT\ell(1+\delta)\right) \\ &\quad + \Pr\left(\overline{\mathcal{O}}(\rho,\ell) \cap \mathcal{R}_{1,w}\right).\end{aligned} \tag{63}$$

The high-SNR behavior of the first two terms in the last line is given by $\rho^{-f(r_1/\ell)}$ and by $\rho^{-NT\ell\beta}$, respectively. We shall focus on the third term. More explicitly, this can be written as (64) at the bottom of the page. We shall upper-bound the above probability as follows. We condition with respect to a given channel matrix $\boldsymbol{H}^c$ with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_m$, for $m = \min\{M,N\}$. We use the union bound over all possible codewords pairs $\boldsymbol{x}, \widehat{\boldsymbol{x}}$, and we average over the code ensemble the pairwise error probability. Eventually, we average the result with respect to $\boldsymbol{H}^c \in \overline{\mathcal{O}}(\rho, L)$. Fix the channel eigenvalues and define

$$v_j \triangleq \frac{-\log(\lambda_j)}{\log(\rho)}, \qquad j = 1, \ldots, m. \tag{65}$$

Using the fact that $\boldsymbol{y}_\ell = \boldsymbol{H}_\ell \boldsymbol{x} + \boldsymbol{w}_\ell$, we can write

$$\begin{aligned}&\Pr\left(|\boldsymbol{H}_\ell(\boldsymbol{x}-\widehat{\boldsymbol{x}}) + \boldsymbol{w}_\ell|^2 \leq NT\ell(1+\delta), |\boldsymbol{w}_\ell|^2 \leq NT\ell(1+\delta)\right) \\ &\quad\overset{(a)}{\leq} \Pr\left(|\boldsymbol{H}_\ell(\boldsymbol{x}-\widehat{\boldsymbol{x}})|^2 \leq 4NT\ell(1+\delta)\right) \\ &\quad\overset{(b)}{=} \Pr\left(\sum_{j=1}^m \lambda_j\chi_j \leq 2NMT\ell(1+\delta)\rho^{-1}\right) \\ &\quad\dot{\leq} \Pr\left(\bigcap_{j=1}^m \left\{\chi_j \leq \frac{2NMT\ell(1+\delta)}{\rho\lambda_j}\right\}\right) \\ &\quad\doteq \rho^{-T\ell\sum_{j=1}^m [1-v_j]_+}\end{aligned} \tag{66}$$

where (a) follows by letting $\boldsymbol{a} = \boldsymbol{H}_\ell(\boldsymbol{x}-\widehat{\boldsymbol{x}})$, $\boldsymbol{b} = \boldsymbol{w}_\ell$, and $\Delta = NT\ell(1+\ell)$ and by noticing that, for any random vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\Delta > 0$, it holds

$$\begin{aligned}\left\{|\boldsymbol{a}+\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{b}|^2 \leq \Delta\right\} &= \left\{|\boldsymbol{a}+\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{a}|^2 \leq 4\Delta\right\} \\ &\quad\cup \left\{|\boldsymbol{a}+\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{a}|^2 > 4\Delta\right\} \\ &= \left\{|\boldsymbol{a}+\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{a}|^2 \leq 4\Delta\right\} \\ &\subseteq \left\{|\boldsymbol{a}|^2 \leq 4\Delta\right\}\end{aligned}$$

since the event $\left\{|\boldsymbol{a}+\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{b}|^2 \leq \Delta, |\boldsymbol{a}|^2 > 4\Delta\right\}$ is empty. Then, (b) follows from the fact that, for the random coding

$$\Pr\left(\overline{\mathcal{O}}(\rho,\ell) \cap \mathcal{R}_{1,w}\right) = \Pr\left(\bigcup_{\substack{\widehat{w}\neq w \\ \widehat{w}>0}} \left\{|\boldsymbol{y}_\ell - \boldsymbol{H}_\ell\widehat{\boldsymbol{x}}|^2 \leq NT\ell(1+\delta), |\boldsymbol{w}_\ell|^2 \leq NT\ell(1+\delta)\right\}, \overline{\mathcal{O}}(\rho,\ell)\right). \tag{64}$$

ensemble, $\frac{1}{\sqrt{2}}(\boldsymbol{x}-\widehat{\boldsymbol{x}})$ is an i.i.d. Gaussian real vector with components $\sim \mathcal{N}(0,1/2)$. Therefore, using the singular value decomposition $\boldsymbol{H}_\ell = \sqrt{\frac{\rho}{M}}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^{\mathsf{H}}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_m)\otimes \boldsymbol{I}_{2T\ell}$ and $\boldsymbol{V}$ with orthonormal columns, we have that

$$|\boldsymbol{H}(\boldsymbol{x}-\widehat{\boldsymbol{x}})|^2 = \frac{2\rho}{M}\left|\boldsymbol{\Lambda}^{1/2}\frac{1}{\sqrt{2}}\boldsymbol{V}^{\mathsf{H}}(\boldsymbol{x}-\widehat{\boldsymbol{x}})\right|^2 = \frac{2\rho}{M}\sum_{j=1}^{m}\lambda_j\chi_j$$

where the χ_j's are i.i.d. central Chi-squared random variables $2T\ell$ degrees of freedom. Finally, the last line follows from the fact that, for such Chi-squared random variables, $\Pr(\chi_j \le a) = O(a^{T\ell})$ for small a and $\Pr(\chi_j \le a) = O(1)$ for large a (see [6, p. 1082] for a very similar development).

The last line of (66) yields an upper bound on the pairwise error probability averaged over the coding ensemble and conditioned with respect to the channel matrix. Summing over all distinct message pairs and averaging over the channels in the nonoutage set we obtain (see [6] for a very similar expression)

$$\Pr\big(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\big) \doteq \int_{\overline{\mathcal{O}}(\rho,\ell)} p_{\boldsymbol{v}}(\boldsymbol{v})\rho^{-T\ell\left[\sum_{j=1}^{m}[1-v_j]_+ - \frac{r_1}{\ell}\right]}d\boldsymbol{v}. \tag{67}$$

We make use the following result from [6]:

Lemma 9: Let $\boldsymbol{v} = (v_1,\ldots,v_m)$ be defined by (65), let $p_{\boldsymbol{v}}(\boldsymbol{v})$ denote the joint density of $\boldsymbol{v}$, which can be computed from the Wishart density of the ordered eigenvalues $\{\lambda_1,\ldots,\lambda_m\}$. For any set $\mathcal{S}\subseteq\mathbb{R}^m$, and any function $g:\mathbb{R}^m\to\mathbb{R}$ for which the integrals below exists

$$\lim_{\rho\to\infty}\frac{-\log\int_{\mathcal{S}} p_{\boldsymbol{v}}(\boldsymbol{v})\rho^{-g(\boldsymbol{v})}d\boldsymbol{v}}{\log\rho} = \inf_{\boldsymbol{v}\in\mathcal{S}\cap\mathbb{R}_+^m}\left\{\sum_{j=1}^{m}(2j-1+|M-N|)v_j + g(\boldsymbol{v})\right\} \tag{68}$$

□

Applying the lemma, we obtain that

$$\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\right) \doteq \rho^{-d_\ell}$$

where we get the equation at the bottom of the page, and where the set $\overline{\mathcal{O}}_\ell$ is the limit of $\rho\to\infty$ of $\overline{\mathcal{O}}(\rho,\ell)$, given by

$$\overline{\mathcal{O}}_\ell = \left\{\boldsymbol{v}\in\mathbb{R}^m,\ v_1\ge\cdots\ge v_m : \sum_{j=1}^{m}[1-v_j]_+ \ge r_1/\ell\right\}. \tag{69}$$

Following [6], for any $T\ge 1$ we find that $d_\ell > 0$ for all ℓ and $r_1 < \min\{M,N\}$. Moreover, if $T\ell \ge M+N-1$ then $d_\ell = f(r_1/\ell)$, which is the maximum possible SNR exponent for codes with multiplexing gain r_1/ℓ and block length $T\ell$.

By collecting all results and recalling (63), we have that

$$p(\ell)\dot{\le}\rho^{-\min\{f(r_1/\ell),NT\ell\beta,d_\ell\}}.$$

This implies that $r_1 = r_e$ and, therefore, that $d^*(r_e,L)$ is achievable by finite-length codes subject to the condition (19), provided that we can show that there exists a single codebook that achieves at the same time the above exponents for $p(\ell)$, for all $\ell = 1,\ldots,L-1$, as well as the exponent of error probability $\Pr(\mathcal{E}_L)$. In other words, we have to show that not only all these exponents can be achieved by averaging over the code ensemble, but that there exist codes that achieve them *simultaneously*.

The expurgation argument is stated by the following lemma in slightly more general terms. The application to our case is then immediate and the proof of Theorem 2 is concluded.

Lemma 10: Consider a sequence of random coding ensembles $\{\mathcal{C}_\rho\}$, indexed by SNR. For each value of ρ, let $\{\mathcal{U}_1,\ldots,\mathcal{U}_K\}$ be a finite set of events in the joint probability space of the code ensemble and of the channel parameters (noise, channel matrix). Let

$$p_k(\rho) = \mathbb{E}_{\mathcal{C}_\rho}[\mathbb{E}_{\mathrm{ch}}[\Pr(\mathcal{U}_k|\mathcal{C}_\rho,\mathrm{channel})]]$$

denote the average probability of the event $\mathcal{U}_k$, where expectation is with respect to both the channel parameters and to the code ensemble. Assume that, for all $k=1,\ldots,K$ there exist positive constants $d_1,\ldots,d_K$ such that

$$p_k(\rho) \doteq \rho^{-d_k}. \tag{70}$$

Then, the probability of the subset of codes $\mathcal{C}_\rho$ such that

$$\mathbb{E}_{\mathrm{ch}}[\Pr(\mathcal{U}_k|\mathcal{C}_\rho,\mathrm{channel})] \le p_k(\rho), \qquad \text{for all } k=1,\ldots,K$$

goes to 1 as $\rho\to\infty$, thus showing that there exist codes that perform at least as good as the ensemble average for all criteria $k=1,\ldots,K$.

Proof: For $\mathcal{C}_\rho$ random in the ensemble of codes, the probabilities

$$p_k(\rho,\mathcal{C}_\rho) \triangleq \mathbb{E}_{\mathrm{ch}}[\Pr(\mathcal{U}_k|\mathcal{C}_\rho,\mathrm{channel})]$$

are random variables whose mean value is equal to $p_k(\rho)$. For any $\epsilon>0$, by using Markov inequality we can write

$$\Pr\left(p_k(\rho,\mathcal{C}_\rho) \ge \rho^{-d_k+\epsilon}\right) \le \rho^{-\epsilon}. \tag{71}$$

$$d_\ell = \inf_{\boldsymbol{v}\in\overline{\mathcal{O}}_\ell\cap\mathbb{R}_+^m}\left\{\sum_{j=1}^{m}(2j-1+|M-N|)v_j + T\ell\left[\sum_{j=1}^{m}[1-v_j]_+ - \frac{r_1}{\ell}\right]\right\}$$

We have

$$\Pr\left(\bigcup_{k=1}^{K}\left\{p_k(\rho,\mathcal{C}_\rho)\geq \rho^{-d_k+\epsilon}\right\}\right)\leq \sum_{k=1}^{K}\rho^{-\epsilon}=K\rho^{-\epsilon}.$$

Hence, the probability of the complement event is lower-bounded by

$$\Pr\left(\bigcap_{k=1}^{K}\left\{p_k(\rho,\mathcal{C}_\rho)< \rho^{-d_k+\epsilon}\right\}\right)\geq 1-K\rho^{-\epsilon}=1-\epsilon'$$

for ρ sufficiently large. This shows that the probability of the set of codes $\mathcal{C}_\rho$ that achieve simultaneously probabilities $p_k(\rho,\mathcal{C}_\rho) < \rho^{-d_k+\epsilon}$ is as large as desired. Since $\epsilon > 0$ is arbitrary, for this set of codes, the SNR exponent of $p_k(\rho,\mathcal{C}_\rho)$ is not smaller than d_k of the ensemble average. To see this, just choose $\epsilon=\frac{\log\log\rho}{\log\rho}$. ∎

Using Lemma 10, the proof for the long-term static channels is concluded. For short-term static channels, the only difference is that, as shown in [6], we need $T \geq M+N-1$ in order to ensure that

$$\Pr(\mathcal{E}_L)\doteq \rho^{-Lf(r_1/L)}. \tag{72}$$

Appendix II
Proof of Theorem 5

We restrict ourselves to stationary power control policies, i.e., such that the total energy Γ_ℓ allocated to the ℓth transmitted block is a time-invariant deterministic function of the relative slot index ℓ. We start by noticing that for any fixed L-tuple $\boldsymbol{\Gamma}=(\Gamma_1,\ldots,\Gamma_L)$, the upper bound on the achievable diversity gain based on Fano inequality and the achievability part for large T in the proof of Theorem 2 hold. Moreover, the achievability result with finite length in Theorem 2 also extend to this scenario after the small modification of requiring $T \geq M+N+1$. This straightforward extension will be outlined at the end of the proof. Therefore, we can focus on studying the SNR exponent of the mutual information level-crossing system as described at the beginning of Appendix I, suitably modified in order to take into account the power control policy.

For each power control policy $\boldsymbol{\Gamma}$, let

$$\frac{1}{T}I_{\boldsymbol{H}^c,\boldsymbol{\Gamma}}(\boldsymbol{x};\boldsymbol{y}_\ell)=\sum_{j=1}^{\ell}\log\,\det\left(\boldsymbol{I}+\frac{\Gamma_j\,\rho}{TM}\;\boldsymbol{H}^c\boldsymbol{H}^{c\mathsf{H}}\right) \tag{73}$$

be the mutual information corresponding to i.i.d. white Gaussian inputs and define $\mathcal{A}_\ell$ as the mutual information level-crossing event

$$\mathcal{A}_\ell=\{\boldsymbol{H}^c\in\mathbb{C}^{N\times M}:\frac{1}{T}I_{\boldsymbol{H}^c,\boldsymbol{\Gamma}}(\boldsymbol{x};\boldsymbol{y}_\ell)>R_1\}.$$

As before, we define the *information outage* event with ℓ received blocks as $\mathcal{O}(\rho,\ell)=\overline{\mathcal{A}}_\ell$, with the associated *outage probability* $P_{\text{out}}(\rho,\ell)=\Pr(\mathcal{O}(\rho,\ell))$ where, by definition, $P_{\text{out}}(\rho,0)=1$. We define also the set of *feasible* power control policies as

$$\mathcal{F}=\left\{\boldsymbol{\Gamma}\in\mathbb{R}_+^L:\frac{1}{T}\frac{\sum_{\ell=1}^{L}\Gamma_\ell P_{\text{out}}(\rho,\ell-1)}{\sum_{\ell=0}^{L-1}P_{\text{out}}(\rho,\ell)}\leq M\right\} \tag{74}$$

where we have used the fact that, for the event $\mathcal{A}_\ell$ defined above

$$p(\ell)=\Pr(\overline{\mathcal{A}}_1,\ldots,\overline{\mathcal{A}}_\ell)=\Pr(\overline{\mathcal{A}}_\ell)=P_{\text{out}}(\rho,\ell)$$

and we used the long-term average transmit power formula (12).

Again, we denote by $d_{\text{out}}(\ell)$ the SNR exponent of the ℓth round outage probability

$$d_{\text{out}}(\ell)=\lim_{\rho\to\infty}\frac{-\log\left(P_{\text{out}}(\rho,\ell)\right)}{\log(\rho)}. \tag{75}$$

Then, $\boldsymbol{\Gamma}\in\mathcal{F}$ implies that

$$\frac{1}{T}\Gamma_\ell\leq\frac{M\sum_{j=0}^{L-1}P_{\text{out}}(\rho,j)}{P_{\text{out}}(\rho,\ell-1)}\leq\frac{ML}{P_{\text{out}}(\rho,\ell-1)}$$

where we used the fact that the average inter-renewal time, given by $\sum_{j=0}^{L-1}P_{\text{out}}(\rho,j)$, cannot be larger than the maximum inter-renewal time L. Letting $\frac{1}{T}\Gamma_\ell\doteq\rho^{\gamma_\ell}$, this yields

$$\gamma_\ell\leq d_{\text{out}}(\ell-1). \tag{76}$$

The condition (76) is clearly also *sufficient* for feasibility, in the sense that if (76) holds then weights $W_\ell>0$ independent of ρ exist such that $\{\Gamma_\ell=TW_\ell\rho^{\gamma_\ell}:\ell=1,\ldots,L\}$ is a feasible policy.

An asymptotically optimal feasible policy must achieve (76) with equality for all ℓ and maximize in sequence the outage exponents $d_{\text{out}}(\ell)$, for $\ell=1,\ldots,L$. This fact can be shown by contradiction: suppose that $\boldsymbol{\Gamma}\in\mathcal{F}$ is optimal and there exists $\boldsymbol{\Gamma}'\in\mathcal{F}$ such that for some $1\leq\ell\leq L$ we have

$$\begin{aligned}&d_{\text{out}}(j)=d'_{\text{out}}(j),\qquad 1\leq j<\ell\\&d_{\text{out}}(\ell)<d'_{\text{out}}(\ell).\end{aligned}$$

Then, a feasible policy $\boldsymbol{\Gamma}''$ with $\gamma''_j=\gamma'_j$ for all $1\leq j\leq\ell$ and

$$\gamma''_{\ell+1}=d'_{\text{out}}(\ell)>d_{\text{out}}(\ell)\geq\gamma_{\ell+1}$$

can be found. Outage probability is a strictly decreasing function of the transmitted powers. Hence, $d''_{\text{out}}(\ell+1)>d_{\text{out}}(\ell+1)$. Going on with this argument, we can show that $d''_{\text{out}}(L)>d_{\text{out}}(L)$, thus contradicting the assumption that $\boldsymbol{\Gamma}$ is asymptotically optimal.

Sequential maximization of the exponents $d_{\text{out}}(\ell)$ yields the following recursive algorithm. We let $m=\min(N,M)$ and denote the m not identically zero eigenvalues of $\boldsymbol{H}^c\boldsymbol{H}^{c\mathsf{H}}$ by

$0 \le \lambda_1 \le \cdots \le \lambda_m$. We let $v_1, \ldots, v_m$ be defined by (65) and we notice that, for all ℓ

$$\sum_{k=1}^{\ell} \log\det\left(\boldsymbol{I} + \rho^{\gamma_k+1}\boldsymbol{H}^c\boldsymbol{H}^{c\mathsf{H}}\right) = \log \prod_{j=1}^{m}\prod_{k=1}^{\ell}\left(1+\rho^{\gamma_k+1-v_j}\right). \tag{77}$$

For $\ell = 1$, we have $P_{\text{out}}(\rho, 0) = 1$ and therefore $\gamma_1 = d_{\text{out}}(0) = 0$. By using (77) for $\ell = 1$, we can write the outage event as

$$\mathcal{O}(\rho,1) = \left\{\boldsymbol{v} \in \mathbb{R}^m : v_1 \ge \cdots \ge v_m, \prod_{j=1}^{m}(1+\rho^{1-v_j}) \le \rho^{r_1}\right\} \tag{78}$$

which, for asymptotically large ρ, yields

$$\mathcal{O}_1 = \left\{\boldsymbol{v} \in \mathbb{R}^m : v_1 \ge \cdots \ge v_m, \sum_{j=1}^{m}[1-v_j]_+ \le r_1\right\}. \tag{79}$$

Writing $P_{\text{out}}(\rho,1) = \int_{\mathcal{O}(\rho,1)} p_{\boldsymbol{v}}(\boldsymbol{v})d\boldsymbol{v}$ and using Lemma 9 we obtain $d_{\text{out}}(1) = f(r_1)$.

Then, let $\gamma_2 = d_{\text{out}}(1)$. By using (77) for $\ell = 2$, we can write the outage event as

$$\mathcal{O}(\rho,2) = \left\{\boldsymbol{v} \in \mathbb{R}^m : v_1 \ge \cdots \ge v_m, \prod_{j=1}^{m}(1+\rho^{1-v_j})(1+\rho^{\gamma_2+1-v_j}) \le \rho^{r_1}\right\} \tag{80}$$

which, for asymptotically large ρ, yields

$$\mathcal{O}_2 = \left\{\boldsymbol{v} \in \mathbb{R}^m : v_1 \ge \cdots \ge v_m, \sum_{j=1}^{m}[\max\{\gamma_2+1-v_j, \gamma_2+2(1-v_j)\}]_+ \le r_1\right\}. \tag{81}$$

From Lemma 9 we obtain

$$d_{\text{out}}(2) = \inf_{\boldsymbol{v}\in\mathcal{O}_2\cap\mathbb{R}_+^m}\left\{\sum_{j=1}^{m}(2j-1+|M-N|)v_j\right\}. \tag{82}$$

Next, we let $\gamma_3 = d_{\text{out}}(2)$ and proceed similarly for $\ell = 3, \ldots, L$. The resulting sequence of optimal exponents $d_{\text{out}}(\ell)$ is upper-bounded by the sequence $\{\xi_\ell\}$ defined in Theorem 5. The upper bound comes from the fact that the sequence $\{\xi_\ell\}$ is given by the same recursion that generates the sequence $\{d_{\text{out}}(\ell)\}$ by replacing r_1 with $r_e \le r_1$. It follows that the optimal exponent of $P_{\text{out}}(\rho, L)$ is upper-bounded by ξ_L.

As anticipated at the beginning of this section, since the converse argument based on Fano inequality holds for any power control policy, it follows that $d^*(r_e, L) \le \xi_L$. Moreover, since the achievability argument for $T \to \infty$ holds for any power control policy, it follows that $d^*(r_e, L) = \xi_L$ (achieved by Gaussian codes in the limit of large T).

The final step is to prove the achievability of $d^*(r_e, L)$ for $T \ge M + N - 1$. The result hinges on the Gaussian i.i.d. code ensemble and on the use of the bounded distance decoder defined in the proof of Theorem 2. Notice that here the probabilities $p(\ell)$ must vanish with exponent $d_{\text{out}}(\ell)$ such that we can allocate power $\Gamma_\ell/T \doteq \rho^{d_{\text{out}}(\ell-1)}$ to the ℓth block while still satisfying the long-term average power constraint. Omitting steps analogous to the proof of Theorem 2 for the sake of conciseness, we find that the proof only requires showing the existence of codes such that

$$p(\ell) \dot{\le} \rho^{-d_{\text{out}}(\ell)} \tag{83}$$

when $\gamma_i = d_{\text{out}}(i-1)$ for $1 \le i \le \ell$. This holds provided that we show the existence of codes that achieve

$$\Pr\left(\overline{\mathcal{O}}(\rho,\ell) \cap \mathcal{R}_{1,w}\right) \dot{\le} \rho^{-d_{\text{out}}(\ell)} \tag{84}$$

for all ℓ, where $\mathcal{R}_{1,w}$ is as defined in Theorem 2 proof and $\overline{\mathcal{O}}(\rho,\ell)$ is the outage event after ℓ ARQ rounds. Replicating the arguments that lead to (66) and (67), we obtain that, with power control

$$\begin{aligned}\Pr\left(|\boldsymbol{H}_\ell(\boldsymbol{x}-\widehat{\boldsymbol{x}})|^2 \le 4NT\ell(1+\delta)\right) &= \Pr\left(\sum_{i=1}^{\ell}\sum_{j=1}^{m}\lambda_j\chi_{i,j}\frac{\Gamma_i}{T} \le \frac{2NT\ell(1+\delta)}{\rho}\right)\\ &\le \Pr\left(\bigcap_{\substack{i=1,\ldots,\ell\\ j=1,\ldots,m}}\left\{\chi_{i,j} \le \frac{2NT\ell(1+\delta)}{\rho\lambda_j\Gamma_i/T}\right\}\right)\\ &\doteq \rho^{-T\sum_{i=1}^{\ell}\sum_{j=1}^{m}[1+\gamma_i-v_j]_+}\end{aligned}$$

where $\chi_{i,j}$ are i.i.d. central Chi-squared random variables with $2T$ degrees of freedom.

By using the above upper bound on the pairwise error probability in the union bound, and averaging over the channel realizations in the no-outage set, we find that (84) holds if the condition at the bottom of the page holds, which is quaranteed for $T \ge M + N - 1$ [6]. Finally, our expurgation lemma (i.e., Lemma 10) yields the existence of codes that achieve the power-control exponent $d^*(r_e, L)$ for finite $T \ge M + N - 1$. This concludes the proof.

$$\inf_{\boldsymbol{v}\in\overline{\mathcal{O}}_\ell\cap\mathbb{R}_+^m}\left\{\sum_{j=1}^{m}(2j-1+|M-N|)v_j + T\sum_{i=1}^{\ell}\left[\sum_{j=1}^{m}[1+\gamma_i-v_j]_+ - \frac{r_1}{\ell}\right]\right\} \ge d_{\text{out}}(\ell).$$

APPENDIX III
PROOF OF LEMMA 6

In general, in order to obtain a lower bound on d^* it is sufficient to enlarge the feasible set of one or more of the optimization problems given in Theorem 5. Notice that the constraint functions $g_\ell(z)$ defined in (30) are piecewise linear, decreasing, and convex. By taking any one of the straight lines whose upper convex envelope forms $g_\ell(z)$, we obtain a *linear* constraint which is strictly looser than the original convex constraint, thus leading to a lower bound.

In particular, the two lower bounds are obtained by taking, for each ℓ, the linear constraints

$$\sum_{j=1}^{m}[\xi_{\ell-1}+1-v_j]_+ \le r_e \tag{85}$$

and

$$\sum_{j=1}^{m}\left[\sum_{i=1}^{\ell}\xi_{\ell-i}+\ell(1-v_j)\right]_+ \le r_e \tag{86}$$

respectively. These correspond to the straight lines for $k=1$ and for $k=\ell$ in the expression of $g_\ell(z)$, respectively.

By considering the sequence of linear optimization problems given by the constraints (85) we obtain explicitly

$$d_0^{(lb1)} = 0 \tag{87}$$
$$d_1^{(lb1)} = f(r_e) \tag{88}$$
$$d_\ell^{(lb1)} = \inf \sum_{j=1}^{m}(2j-1+|M-N|)\,v_j \tag{89}$$

subject to

$$(1+d_{\ell-1}^{(lb1)})\sum_{j=1}^{m}\left[1-\frac{v_j}{1+d_{\ell-1}^{(lb1)}}\right]_+ \le r_e. \tag{90}$$

Through the change of variables $\nu_j = v_j/(1+d_{\ell-1}^{(lb1)})$, and by noticing that $f(r_e)$ is the solution to the linear program

$$\inf \sum_{j=1}^{m}(2j-1+|M-N|)\,\nu_j \tag{91}$$

subject to the constraint

$$\sum_{j=1}^{m}[1-\nu_j]_+ \le r_e \tag{92}$$

we obtain

$$d_\ell^{(lb1)} = (1+d_{\ell-1}^{(lb1)})f\left(\frac{r_e}{1+d_{\ell-1}^{(lb1)}}\right)$$

as stated in the lemma. The second lower bound is established in a similar manner by considering the constraint (86).

To prove the upper bound, we observe that $g_\ell(z)$ attains its maximum value $g_\ell(0)=\sum_{i=1}^{\ell-1}\xi_i+\ell$ at $z=0$, and it is zero for $z \ge \xi_{\ell-1}+1$. Hence, the piecewise linear function

$$g_\ell(0)\left[1-\frac{z}{\xi_{\ell-1}+1}\right]_+ \tag{93}$$

is strictly above $g_\ell(z)$ for all z. By replacing $g_\ell(z)$ by (93), we obtain the sequence of linear programs

$$d_0^{(ub)} = 0 \tag{94}$$
$$d_1^{(ub)} = f(r_e) \tag{95}$$
$$d_\ell^{(ub)} = \inf \sum_{j=1}^{m}(2j-1+|M-N|)\,v_j \tag{96}$$

subject to

$$(\ell+\sum_{i=1}^{\ell-1}d_i^{(ub)})\sum_{j=1}^{m}\left[1-\frac{v_j}{1+d_{\ell-1}^{(ub)}}\right]_+ \le r_e \tag{97}$$

which yields

$$d_\ell^{(ub)} = (1+d_{\ell-1}^{(ub)})f\left(\frac{r_e}{\ell+\sum_{i=1}^{\ell-1}d_i^{(ub)}}\right)$$

as stated in the lemma.

APPENDIX IV
PROOF OF THEOREM 8

The proof is essentially the same with either the short-term or long-term average power constraint. Therefore, we only discuss the long-term static channel with the short-term power constraint (i.e., Theorem 2) for conciseness. We start with the Loeliger ensemble of $\mathrm{mod}\text{-}p$ lattices defined in [28] (see also [29], [30]). For the sake of completeness, we recall here its definition. Let p be a prime. The ensemble is generated via Construction A, as the set of all lattices given by

$$\Lambda_p = \kappa\left(\boldsymbol{g}\mathbb{Z}_p + p\mathbb{Z}^{2MTL}\right) \tag{98}$$

where $p\to\infty$, $\kappa\to 0$ is a scaling coefficient adjusted such that the fundamental volume

$$V_f = \kappa^{2MTL}p^{2MTL-1} = 1,$$

$\mathbb{Z}_p$ denotes the field of $\mathrm{mod}\text{-}p$ integers, and $\boldsymbol{g}\in\mathbb{Z}_p^{2MTL}$ is a vector with i.i.d. components. We use a pair of self-similar lattices for nesting. In particular, we take the shaping lattice to be $\Lambda_s=\zeta\Lambda_p$, where ζ is chosen such that $r_{\mathrm{cov}}^2=1/2$ in order to satisfy the input power constraint. The coding lattice is obtained as $\Lambda_c = 1/\tau\Lambda_s$, where $\tau=\lfloor\rho^{r_1/2ML}\rfloor$ in order to satisfy the transmission rate of the first round is $R_1(\rho)\doteq r_1\log\rho$. This yields the fundamental volumes

$$V_f(\Lambda_s)\triangleq V_s = \zeta^{2MTL} \tag{99}$$
$$V_f(\Lambda_c)\triangleq V_c = \left(\frac{\zeta}{\tau}\right)^{2MTL}. \tag{100}$$

In order to exclude bad shaping lattices, we expurgate the ensemble by removing all lattices whose covering efficiency is larger than $\log(\rho)$. The new ensemble, i.e., Λ_{exp}, will be used

throughout the proof. Now, we proceed in the same lines as the proof of Theorem 2. The only differences resulting from using an ensemble of lattice codes instead of the Gaussian ensemble and the list MMSE-lattice decoder instead of the bounded distance decoder is that we now need to upper-bound $\Pr\left(|\boldsymbol{e}'|^2 \geq MTL(1+\beta\log(\rho))\right)$ and the ensemble average $\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\right)\right]$. The fundamental challenge in the first task is the non-Gaussianity of $\boldsymbol{e}'$. In [7], however, we showed that this non-Gaussianity does not change the exponential order of the Chernoff upper bound assuming $\boldsymbol{e}'$ Gaussian (taking a form similar to (60)). Therefore, we have

$$\Pr\left(|\boldsymbol{e}'|^2 \geq MTL(1+\beta\log(\rho))\right) \dot{\leq} \rho^{-MTL\beta}$$

which settles our first task. Toward the second goal, we first observe that

$$\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\right)\right] \leq \mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1}\right)\right]$$

and $\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1}\right)\right]$ is the ensemble average of the probability of error achieved by the ambiguity decoder proposed by Loeliger [28]. In [7], we have shown that

$$\begin{aligned}&\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1}\right)\right]\\ &\dot{\leq}(1+\epsilon)\int_{\overline{\mathcal{O}}} p(\boldsymbol{H})\,\rho^{r_1 T}(1+\beta\log(\rho))^{NT\ell}\det\left(\boldsymbol{B}_\ell\boldsymbol{B}_\ell^{\mathsf{T}}\right)^{-1/2} d\boldsymbol{H}\end{aligned}$$

where $\epsilon > 0$ can be made arbitrarily small by increasing p. From elementary properties of MMSE-DFE equalization [7], we know that

$$\det\left(\boldsymbol{B}_\ell\boldsymbol{B}_\ell^{\mathsf{T}}\right) = \left(\det\left(\boldsymbol{I}+\frac{\rho}{M}\boldsymbol{H}^{c}\boldsymbol{H}^{c\mathsf{H}}\right)\right)^{2T\ell}. \tag{101}$$

Using this result and following in the footsteps of [6] we obtain

$$\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\right)\right] \dot{\leq} P_{\text{out}}(\rho,\ell) \tag{102}$$

which implies

$$\lim_{\rho\to\infty}\frac{-\log\left(\mathbb{E}_{\Lambda_{\exp}}\left[\Pr\left(\overline{\mathcal{O}}(\rho,\ell)\cap\mathcal{R}_{1,w}\right)\right]\right)}{\log(\rho)} \geq f\left(\frac{r_1}{\ell}\right) \tag{103}$$

for $T \geq M+N-1$. Then, we use the same arguments as in the proof of Theorem 2 to see that $r_e = r_1$ and the ensemble average of the probability of error achieves the optimal diversity advantage $d^*(r_e, L) = f\left(\frac{r_e}{L}\right)$. The final step follows from Lemma 10 which establishes the existence of a lattice in the ensemble $\Lambda_{\exp}$ such that the corresponding nested LAST code achieves *simultaneously* the condition in (103) for all ℓ. The proofs for the short-term channel and the power-controlled ARQ scheme follow exactly the same arguments and are omitted for the sake of conciseness.

REFERENCES

[1] E. Teletar, "Capacity of multi-antenna Gaussian channels," AT&T-Bell Labs, Tech. Rep., 1995.

[2] G. J. Foschini and M. Gans, "On the limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[4] J.-C. Guey, M. R. Bell, M. P. Fitz, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf.*, Atlanta, GA, 1996, pp. 136–140.

[5] B. M. Hochwald, G. Caire, B. Hassibi, and T. Marzetta, "Special Issue on Space–Time Transmission, Reception, Coding and Signal processing," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, Oct. 2003.

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

[7] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity–multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, Jun. 2004.

[8] H. El Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[9] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full diversity, high rate, space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[10] H. Yao and G. W. Wornell, "Achieving the full MIMO diversity-vs-multiplexing frontier with rotation-based space-time codes," in *41th Annu.Allerton Conf. Commication Control, and Computing*, Monticello, IL, Oct. 2003.

[11] P. Dayal and M. K. Varanasi, "An optimal two transmit antenna space-time code and its stacked extensions," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Monterey, CA, Nov. 2003, vol. 1, pp. 987–991.

[12] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden code: A 2×2 full-rate space-time code with non-vanishing determinants," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 308.

[13] P. Elia, R. Kumar, S. Pawar, P. V. Kumar, and H.-F. Liu, "Explicit, minimum-delay space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, submitted for publication.

[14] S. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE (Special Issue on Gigabit Wireless)*, vol. 92, no. 2, pp. 219–270, Feb. 2004.

[15] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay bounded packet scheduling of bursty sources over wireless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.

[16] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-PartI:No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[17] J. P. Schalkwijk and M. E. Barron, "Sequential signaling under a peak power constraint," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 3, pp. 278–282, May 1971.

[18] A. Hottinen and O. Tirkkonen, "Matrix modulation and adaptive retransmission," in *Proc. Int. Symp. Signal Processing and Its Applications*, Paris, France, Jul. 2003, vol. 1, pp. 221–224.

[19] E. N. Onngosanusi, A. Dabak, Y. Hui, and G. Jeong, "Hybrid ARQ transmission and combining for MIMO systems," in *Proc Int. Conf. Communications (ICC'03)*, Anchorage, AK, May 2003, vol. 5, pp. 3205–3209.

[20] H. Zheng, A. Lozano, and M. Haleem, "Multiple ARQ processes for MIMO systems," in *Proc. 13th IEEE Int. Symp. Personal Indoor and Mobile Radio Communications (PIMRC'02)*, Lisbon, Portugal, Sep. 2002, vol. 3, pp. 1023–1026.

[21] Z. Ding and M. Rice, "Type-I hybrid ARQ using MTCM spatio-temporal vector coding for MIMO systems," in *Proc. Int. Conf. Communications (ICC'03)*, Anchorage, AK, May 2003, vol. 4, pp. 2758–2762.

[22] T. Koike, H. Murata, and S. Yoshida, "Hybrid ARQ scheme suitable for coded MIMO transmission," in *Proc. Int. Conf. Communications (ICC'04)*, Paris, France, Jun. 2004, vol. 5, pp. 2919–2923.

[23] G. Caire and D. Tuninetti, "ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1971–1988, Jul. 2001.

[24] G. D. Forney Jr., "Exponential error bounds for list, erasure, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 206–220, Mar. 1968.

[25] S. C. Draper, K. Ramchandran, B. Rimoldi, A. Sahai, and D. N. C. Tse, "Attaining maximal reliability with minimal feedback via joint channel-code and hash-function design," in *Proc. 43rd Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2005.

[26] J. M. Cioffi and G. D. Forney, Jr, "Generalized decision feedback equalization for packet transmission with isi and Gaussian noise," in *Communications, Computation, Control, and Signal Processing*, A. Paulraj, Ed. *et al.* Stanford, CA: Stanford Univ., 1997, p. 79.

[27] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood decoding and the search of the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[28] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

[29] U. Erez and R. Zamir, "Lattice decoding can achieve $\frac{1}{2}\log(1+snr)$ on the AWGN channel using nested codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[30] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct.. 2005.