Lattice Coding and Decoding Achieve the Optimal Diversity-vs-Multiplexing Tradeoff of MIMO Channels [*]

Hesham El Gamal
Ohio-State University

Giuseppe Caire
Eurecom Institute

Mohamed Oussama Damen
University of Alberta

November 6, 2003

Abstract

This paper considers communication over coherent multiple-input multiple-output (MIMO) flat fading channels where the channel is only known at the receiver. For this setting, we introduce the class of LAttice Space-Time (LAST) codes. We show that these codes achieve the optimal diversity-vs-multiplexing tradeoff defined by Zheng and Tse under generalized minimum Euclidean distance lattice decoding. Our scheme is based on a generalization of Erez and Zamir mod-Λ scheme to the MIMO case. In our construction the scalar "scaling" of Erez-Zamir and Costa Gaussian "Dirty-Paper" schemes is replaced by the minimum mean square error generalized decision-feedback equalizer (MMSE-GDFE). This result settles the open problem posed by Zheng and Tse on the construction of *explicit* coding and decoding schemes that achieve the optimal diversity-vs-multiplexing tradeoff. Moreover, our results shed more light on the structure of optimal coding/decoding techniques in delay limited MIMO channels, and hence, opens the door for novel approaches for space-time code constructions. In particular; 1) we show that MMSE-GDFE plays a fundamental role in approaching the limits of delay limited MIMO channels in the high SNR regime, unlike the AWGN channel case and 2) our random coding arguments represent a major departure from traditional space-time code designs based on the rank and/or mutual information design criteria.

Keywords: Lattice coding and decoding, minimum mean square error (MMSE) equalization, multiple-input multiple-output (MIMO) channels, diversity-vs-multiplexing tradeoff.

[*] The work of Hesham El Gamal was funded in part by the National Science Foundation.

1 Introduction

Since the seminal work of Teletar [1], Foschini and Gans [2], Tarokh *et al.* [3], and Guey *et al.* [4], multiple antenna transmission/reception has emerged as a key tool to achieve high spectral and power efficiency in wireless communications. Loosely speaking, schemes that exploit both the classical Shannon degrees of freedom (time-frequency) and the additional spatial degrees of freedom (antennas) in order to achieve reliable transmission of information 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 developed on the basis of different physical motivations and, for each setting, information theoretic results and associated coding schemes have been developed.

Perhaps the most basic setting (originally treated in [1, 2] from the information theoretic viewpoint and in [3, 4] from the code construction viewpoint) consists of the quasi-static frequency-flat fading M-transmit N-receive multiple-input multiple-output (MIMO) channel with no channel knowledge at the transmitter and perfect channel knowledge at the receiver. The complex baseband model is defined by[1]

$$\mathbf{y}_t^c = \sqrt{\frac{\rho}{M}}\mathbf{H}^c\mathbf{x}_t^c + \mathbf{w}_t^c, \quad t = 1, \dots, T \tag{1}$$

where $\{\mathbf{x}_t^c \in \mathbb{C}^M : t = 1, \dots, T\}$ is the transmit signal, $\{\mathbf{y}_t^c \in \mathbb{C}^N : t = 1, \dots, T\}$ is the received signal, $\{\mathbf{w}_t^c \in \mathbb{C}^N : t = 1, \dots, T\}$ denotes the channel Gaussian noise, assumed temporally and spatially white with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and $\mathbf{H}^c$ is the $N \times M$ channel matrix with the (i, j)-th element h_{ij}^c representing the fading coefficient between the j-th transmit and the i-th receive antenna. The fading coefficients are further assumed to be i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and remain fixed for $t = 1, \dots, T$, where T is the duration of a space-time codeword (block length). By enforcing the input constraint

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}|\mathbf{x}_t^c|^2\right] \le M, \tag{2}$$

the parameter ρ takes on the meaning of *average* signal-to-noise ratio (SNR) per receiver antenna. The channel matrix $\mathbf{H}^c$ is assumed to be perfectly known at the receiver and completely unknown at the transmitter.

[1] **Notation:** the superscript ^{c} denotes complex quantities, $^\mathsf{T}$ denotes transpose and $^\mathsf{H}$ denotes Hermitian transpose. The notation $\mathbf{v} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that $\mathbf{v}$ is a proper complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For real Gaussian random vector we use the notation $\mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 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 $\mathcal{R}$.

The quasi-static MIMO fading channel defined above is not information-stable [6] and has a zero Shannon capacity. This means that the channel reliability function is also equal to zero and that the error probability of any coding scheme with positive rate is bounded away from zero. On the other hand, it can be easily shown that the error probability of a given coding scheme with a constant rate behaves like $O(\rho^{-d})$ for large ρ, where $d \leq MN$ depends on the coding scheme and is called *the diversity gain* [3, 4]. It is also well known that the capacity of the ergodic MIMO channel treated in [1][2] behaves like $O(\min\{M, N\} \log \rho)$ for large ρ.

The linear growth of the "pre-log" factor in the ergodic MIMO channel capacity motivated several researchers to consider coding schemes that send $r \leq \min(N, M)$ (independent) information symbols per channel use (PCU)[3]. The integer r was referred to as the *multiplexing gain of the scheme*, as it (roughly) corresponds to creating r parallel communication channels between the transmitter and receiver. In [3], the coding problem for the above channel was stated as follows: for each desired diversity gain d, maximize the multiplexing gain r. Early works considered *constrained* coding ensembles (e.g., trellis or block codes over discrete QAM signal sets of a given size). For these ensembles, there exists indeed a tradeoff between d and r as dictated by the Singleton bound [3]. Later on, it was recognized that this tradeoff is not a fundamental feature of the channel defined above, but it is due to the additional constraints put on the coding ensemble (see [7]). If no additional constraint beyond the standard average input power (2) is imposed, *structured* space-time coding schemes achieving "full-rate" ($r = \min(M, N)$) and "full-diversity" ($d = MN$) can be explicitly constructed [8, 7]. Furthermore, when these codes are linear over the field of complex numbers (e.g., [7]), they lend themselves to efficient decoding algorithms using number theoretic tools [9].

The problem of characterizing the optimal diversity-vs-multiplexing tradeoff was well-posed and solved by Zheng and Tse in [10]. For given M, N and T, the authors 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 as follows

$$r \triangleq \lim_{\rho \to \infty} \frac{R(\rho)}{\log \rho} \quad \text{and} \quad d \triangleq -\lim_{\rho \to \infty} \frac{\log(P_e(\rho))}{\log \rho}. \tag{3}$$

In [10], the optimal tradeoff curve $d^\star(r)$, yielding for each r the maximum possible d, was found for unrestricted coding and ML decoding. In particular, for any block length $T \geq N + M - 1$ the optimal tradeoff is given by the piecewise linear function joining the points $(k, (M-k)(N-k))$ for $k = 0, \ldots, \min\{M, N\}$.

[2] The ergodic MIMO channel is obtained by replacing the random constant matrix $\mathbf{H}^c$ by the ergodic matrix process $\{\mathbf{H}^c_t\}$ such that each $\mathbf{H}^c_t$ is identically distributed as $\mathbf{H}^c$ in the model (1).

[3] A channel use corresponds to the transmission of the input vector $\mathbf{x}^c_t$ in parallel from the M transmit antennas.

Zheng and Tse further showed that the optimal tradeoff curve is achieved in the $M = 2, N = 1$ case by the Alamouti scheme [11] and, in the limit of $T \to \infty$, by the so-called D-BLAST scheme [12] with **Gaussian code ensembles** and MMSE-DFE processing. The conclusion of their work provides the motivation for this paper: *"It should be noted that other than for the 2×1 channel (for which the Alamouti scheme is optimal), there is no explicitly constructed coding scheme that achieves the optimal tradeoff curve for any $r > 0$. This remains an open problem."* [10].

The quest for explicit coding schemes that exploit the multiplexing and diversity gains available in MIMO channels has generated very intensive work. Earlier works have been largely inspired by suboptimal schemes like the orthogonal designs [13] or the BLAST architecture [12] (e.g., [14]). More recent works have been inspired by Zheng and Tse characterization of the fundamental diversity-vs-multiplexing tradeoff. For example, a structured coding scheme achieving the optimal tradeoff in the case $M = N = 2$ for block length $T = 2$ under ML decoding was recently presented in [15]. In this paper we provide a general answer to the problem posed by Zheng and Tse by exhibiting explicit coding schemes that achieve $d^\star(r)$ for any M and N, and block length $T \geq M + N - 1$.

In order to facilitate the goal of achieving the optimal tradeoff, we first introduce a novel class of space-time codes obtained from lattices. The main idea of LAttice Space-Time (LAST) codes is to carve the space-time code directly from a properly constructed lattice. LAST coding is a non-trivial generalization of linear dispersion (LD) coding [16] as shown in the sequel. Here, we observe that some code constructions that have been proposed under the name *lattice space-time codes* in recent literature do not benefit from the generalization proposed in this paper, and hence, are more appropriately categorized as LD codes (e.g., [17], [18]). One important feature of lattice codes is that they can be decoded by a class of efficient decoders known as *lattice decoders*. Lattice decoding algorithms disregard the boundaries of the lattice code and find the point of the underlying (infinite) lattice closest (in some sense) to the received point. If a point outside the lattice code boundaries is found, an error is declared. Lattice decoding allows for significant reductions in complexity, compared to maximum likelihood (ML) decoding, since 1) it avoids the need for complicated *boundary control* [9] and 2) It allows for using efficient preprocessing algorithms (e.g., the LLL algorithm [19]) which are known to offer significant complexity reduction.

It is well known that lattice codes achieve the capacity $\log(1 + \rho)$ of standard single-input single-output (SISO) unfaded additive white Gaussian noise (AWGN) channels under ML decoding [20, 21]. For a long time lattice codes were believed to achieve a rate equal to only $\log(\rho)$ under lattice decoding. Recently, Erez and Zamir have shown that lattice coding and decoding indeed achieve the full AWGN capacity $\log(1+\rho)$ provided that transmitter and receiver share a common randomness in the form of a dither signal [22]. Their construction is based on nested lattices (i.e., Voronoi codes) and lattice decoding is applied to the received signal after an appropriate scaling. The scaling coefficient that does the "magic" in Erez and Zamir scheme corresponds to the linear MMSE estimator of the channel input from the channel output. The fundamental role played by minimum mean square error (MMSE) estimation in this problem has been further illuminated by Forney in [23].

The main contribution of this paper is Theorem 6, stating that LAST codes achieve the optimal diversity-vs-multiplexing tradeoff under generalized minimum Euclidean distance lattice decoding. The key ingredient in our proof is a non-trivial extension of Erez and Zamir scheme to the case of MIMO channels. It turns out that MMSE estimation plays a fundamental role in this scenario as well, but the MMSE estimator takes on the form of the MMSE generalized decision feedback equalizer (MMSE-GDFE) introduced in [24].

In addition to the main result, the analysis and technical arguments developed here allow for many interesting insights on the structure of optimal space-time coding and decoding techniques. In particular:

1. We show that MMSE estimation plays a fundamental role in allowing lattice decoding to achieve the optimal diversity-vs-multiplexing tradeoff. In fact, through theoretical analysis and representative numerical examples, we show that the *naive* implementation of minimum Euclidean distance lattice decoding without MMSE estimation entails significant losses in the achievable diversity-vs-multiplexing tradeoff.

2. In our random coding arguments we use ensembles of lattice codes which are *good* for an AWGN channel. In other words, we do not impose any *space-time structure* on the ensemble of codes and yet we establish that these ensembles achieve the optimal diversity-vs-multiplexing tradeoff. This represents a marked departure from traditional space-time code design techniques aimed at optimizing the coding gain [3] and/or mutual information [16]. More surprisingly, our simulation experiments show that the performance of randomly selected LAST codes under lattice decoding (with MMSE estimation) rivals that of the state of the art codes available in the literature under ML decoding.

3. In addition to the optimality of the mod-Λ construction with respect to the diversity-vs-multiplexing tradeoff in the high SNR regime, we establish the asymptotic optimality of this construction in terms of closing the gap to the outage probability for an arbitrary SNR. Specifically, we show that as $T \to \infty$, the probability of error achievable with this scheme approaches the outage probability (assuming i.i.d. Gaussian inputs).

4. The optimality of lattice decoding proves that maximizing the well known "rank and

determinant design criteria" [3, 4] **is not** a necessary condition for approaching the fundamental limits of delay limited MIMO channels. These design criteria are inspired by a pairwise probability of error analysis which fails in predicting the performance of lattice decoding due to the infinite number of *virtual* codewords as seen by the decoder. It remains to be seen if these criteria will play a role in further optimizing the performance of certain LAST codes.

5. In the generalized mod-Λ construction, we follow in the footsteps of Erez and Zamir and use a random lattice translate (i.e., dither). It will become clear in the sequel, however, that the optimality of the proposed scheme will still hold if the dither is replaced by the optimal lattice translate[4]. We further elaborate on the role of the random dither in Section 3.3.

6. As a side result, we also establish the optimality of spherical lattice codes (the shaping region is a sphere) under lattice and ML decoding. For these codes, encoding requires the storage of the whole codebook. Voronoi codes (i.e., nested lattices), therefore, enjoy an important advantage over spherical lattice codes due to their low encoding complexity [25].

Before we proceed further, a brief remark about the notion of "explicit" coding schemes is in order. Zheng and Tse proved the achievability of $d^\star(r)$ by using a Gaussian i.i.d. random coding ensemble (i.e., unrestricted coding) and maximum likelihood (ML) decoding. These codes have no structure, and hence, encoding requires storage of the whole codebook and decoding requires exhaustive search over all the codewords. The codes proposed in this paper are explicit in the sense that we use ensembles of lattice codes and lattice decoding. This allows for developing efficient decoding techniques inspired by algorithms that search for the closest lattice point (e.g., [26, 9]). The complexity of the decoding algorithm is, therefore, **exactly** the same as lattice decoding for single-input single-output (SISO) AWGN channels[22][5] and no additional complexity is entailed by the use of multiple antennas. This is precisely the same sense in which Alamouti code is an *explicit* construction [10].

The rest of this paper is organized as follows. In Section 2, we review the required results from lattice theory and introduce the class of LAttice Space-Time (LAST) codes. We establish the optimality of LAST codes with generalized minimum Euclidean distance lattice decoding in Section 3. We develop our main result in two steps. First, in Section 3.1, we show that the *naive* implementation of lattice decoding can entail significant performance losses. Then,

[4] The optimal lattice translate will be defined rigorously in the sequel.

[5] Assuming the dimensionality and rate are the same.

we introduce the generalized mod-Λ construction and establish its optimality in Section 3.2. Section 4 presents numerical results that validate our theoretical claims. Finally, we offer some concluding remarks and an outlook on future work in Section 5. In order to enhance the flow of the paper, all the proofs are deferred to the Appendices.

2 Lattices and LAST codes

We will recall here some notation and results from lattice theory (e.g. [27, 22, 28]) that will be used throughout the paper. An m-dimensional real lattice Λ is a discrete additive subgroup of $\mathbb{R}^m$ defined as $\Lambda = \{\mathbf{G}\mathbf{u} : \mathbf{u} \in \mathbb{Z}^m\}$, where $\mathbf{G}$ is the $m \times m$ (full rank) generator matrix of Λ. The fundamental Voronoi cell $\mathcal{V}$ of Λ is the set of points $\mathbf{x} \in \mathbb{R}^m$ closest to $\mathbf{0}$ than to any other point $\boldsymbol{\lambda} \in \Lambda$. The fundamental volume of Λ is

$$V_f(\Lambda) \triangleq V(\mathcal{V}) = \int_{\mathcal{V}} d\mathbf{x} = \sqrt{\det(\mathbf{G}^\mathsf{T}\mathbf{G})}.$$

The second-order moment of Λ is defined as $\sigma^2(\Lambda) \triangleq \frac{1}{mV_f(\Lambda)} \int_{\mathcal{V}} |\mathbf{x}|^2 d\mathbf{x}$ and the normalized second-order moment is defined as

$$G(\Lambda) \triangleq \frac{\sigma^2(\Lambda)}{V_f(\Lambda)^{2/m}}.$$

The covering radius $r_{\text{cov}}(\Lambda)$ is the radius of the smallest sphere centered in the origin that contains $\mathcal{V}$. The effective radius $r_{\text{eff}}(\Lambda)$ is the radius of the sphere with volume equal to $V_f(\Lambda)$.

A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is *good for covering* [28] if their covering efficiency satisfies

$$\eta_{\text{cov}}(\Lambda_m) \triangleq \frac{r_{\text{cov}}(\Lambda_m)}{r_{\text{eff}}(\Lambda_m)} \to 1 \tag{4}$$

and it is *good for MSE quantization* if

$$G(\Lambda_m) \to \frac{1}{2\pi e} \tag{5}$$

It can be shown (see [28] and references therein) that goodness for covering implies goodness for MSE quantization. Such lattice sequences exist, as shown in [28] (see also [29]). It is also known that if $\{\Lambda_m\}$ is a sequence of lattices good for MSE quantization, with fixed second-order moment σ^2, then a random vector uniformly distributed over $\mathcal{V}(\Lambda_m)$ converges in distribution (in the sense of divergence) to a Gaussian i.i.d. random vector with per-component variance equal to σ^2 [30].

In the rest of this paper, ensembles of lattices satisfying the Minkowski-Hlawka theorem play a very important role. For the sake of completeness, we recall the Minkowski-Hlawka theorem in the form given in [27]:

Theorem 1 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a Riemann integrable function of bounded support (i.e., $f(\mathbf{z}) = 0$ if $|\mathbf{z}|$ exceeds some bound). For any $\epsilon > 0$ there exist ensembles $\{\Lambda\}$ of lattices with fundamental volume V_f and dimension m such that*

$$\left| \mathbb{E}_\Lambda \left[\sum_{\mathbf{z} \in \Lambda, \mathbf{z} \neq \mathbf{0}} f(\mathbf{z}) \right] - \frac{1}{V_f} \int_{\mathbb{R}^m} f(\mathbf{z}) d\mathbf{z} \right| \leq \epsilon \tag{6}$$

□

As a corollary we have that, for any bounded Jordan-measurable set $\mathcal{R} \subset \mathbb{R}^m$, there exist lattice ensembles $\{\Lambda\}$ such that

$$\mathbb{E}_\Lambda \left[|\Lambda^* \cap \mathcal{R}| \right] \approx \frac{V(\mathcal{R})}{V_f} \tag{7}$$

where $\Lambda^* = \Lambda - \{\mathbf{0}\}$ and the approximation in (7) can be made as tight as desired.

Finally, we will need the following result, proved in [22].

Theorem 2 *For any $R > 0$, there exist sequences of nested lattices $\Lambda_m \subseteq \Lambda'_m$ of increasing dimension m such that:*

1. *The cardinality of the partition Λ'_m / Λ_m satisfies*

$$\frac{1}{m} \log |\Lambda'_m / \Lambda_m| \to R.$$

2. *For each m, Λ'_m is randomly selected in an ensemble that asymptotically satisfies Theorem 1, in the limit of $m \to \infty$.*

3. *$\{\Lambda_m\}$ is a sequence of lattices that are good for covering (i.e., they satisfy (4)) and consequently is also a sequence of lattices that are good for MSE quantization (i.e., they satisfy (5)).*

□

An m-dimensional lattice code $\mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$ is the finite subset of the lattice translate $\Lambda + \mathbf{u}_0$ inside the *shaping region* $\mathcal{R}$, i.e., $\mathcal{C} = \{\Lambda + \mathbf{u}_0\} \cap \mathcal{R}$, where $\mathcal{R}$ is a bounded measurable region of $\mathbb{R}^m$. For any Λ and $\mathcal{R}$, there exists $\mathbf{u}_0^\star$ such that

$$|\mathcal{C}(\Lambda, \mathbf{u}_0^\star, \mathcal{R})| \geq \frac{V(\mathcal{R})}{V_f(\Lambda)}. \tag{8}$$

Now, we go back to our space-time coding problem and introduce the class of LAST codes. In order to simplify the presentation, it is useful to introduce the real channel model, equivalent to (1),

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{9}$$

where we define $\mathbf{x} = (\mathbf{x}_1^\mathsf{T}, \ldots, \mathbf{x}_T^\mathsf{T})^\mathsf{T}$ with $\mathbf{x}_t^\mathsf{T} = \left[\mathrm{Re}\{\mathbf{x}_t^c\}^\mathsf{T}, \mathrm{Im}\{\mathbf{x}_t^c\}^\mathsf{T}\right]^\mathsf{T}$, $\mathbf{w} = (\mathbf{w}_1^\mathsf{T}, \ldots, \mathbf{w}_T^\mathsf{T})^\mathsf{T}$ with $\mathbf{w}_t^\mathsf{T} = \left[\mathrm{Re}\{\mathbf{w}_t^c\}^\mathsf{T}, \mathrm{Im}\{\mathbf{w}_t^c\}^\mathsf{T}\right]^\mathsf{T}$, and

$$\mathbf{H} \triangleq \sqrt{\frac{\rho}{M}}\mathbf{I} \otimes \left(\begin{bmatrix} \mathrm{Re}\{\mathbf{H}^c\} & -\mathrm{Im}\{\mathbf{H}^c\} \\ \mathrm{Im}\{\mathbf{H}^c\} & \mathrm{Re}\{\mathbf{H}^c\} \end{bmatrix} \right) \tag{10}$$

is the $2NT \times 2MT$ block-diagonal real channel matrix consisting of the same $2N \times 2M$ diagonal block repeated T times ($\mathbf{I}$ is the identity matrix of dimension T here and $\otimes$ denotes the Kronecker product). The design of space-time signals, therefore, reduces to the construction of a codebook $\mathcal{C} \subseteq \mathbb{R}^{2MT}$ satisfying the input constraint

$$\frac{1}{|\mathcal{C}|} \sum_{\mathbf{x} \in \mathcal{C}} |\mathbf{x}|^2 \leq MT$$

(equivalent to (2)) and enjoying certain desirable properties. The space-time coding rate is given by $R = \frac{1}{T} \log_2 |\mathcal{C}|$ bits PCU.

We say that a space-time coding scheme is a *full-dimensional* LAST code if its codebook is a lattice code, i.e., if $\mathcal{C} = \mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$, for some $2MT$-dimensional lattice Λ, translation vector $\mathbf{u}_0$ and shaping region $\mathcal{R}$.

We used the term *full-dimensional* in the above definition to highlight the fact that the dimensionality of the underlying lattice is equal to the number of (real) degrees of freedom offered by the channel. As demonstrated in Section 4, one can obtain LD codes as special cases of the LAST coding framework for a particular choice of the shaping region. The generalization from LD to LAST coding is instrumental in approaching the fundamental limits of MIMO channels, as we will show next.

3 Achieving the optimal tradeoff with LAST codes

In this section we consider LAST codes under *lattice decoding*. By lattice decoding we refer to the class of decoding algorithms which *do not* take into account the shaping region $\mathcal{R}$. In other words, a lattice decoder finds the point of the underlying (infinite) lattice translate $\Lambda + \mathbf{u}_0$ that is closest (according to a suitable decoding metric) to the received point, irrespective of whether this point is in $\mathcal{R}$ or not. As observed in Section 1, this allows for exploiting the algebraic structure of the underlying lattice to reduce the complexity of the search algorithm. In order to further limit the complexity, we restrict ourselves to the class of generalized minimum Euclidean distance lattice decoders defined by

$$\hat{\mathbf{z}} = \text{arg} \min_{\mathbf{z} \in \mathbb{Z}^{2MT}} |\mathbf{\Gamma y} + \mathbf{a} - \mathbf{\Xi z}|^2 , \tag{11}$$

where $\mathbf{\Gamma}$ and $\mathbf{\Xi}$ are matrices that will be defined in the sequel and $\mathbf{a}$ is a translation vector. It is well-known that this class of algorithms lends itself to an efficient implementation using the so called *sphere decoder* (e.g., [9] and references therein).

3.1 The suboptimality of "naive" lattice decoding

Before introducing the optimal scheme, we investigate the achievable performance of LAST codes under straightforward application of lattice decoding. In this *naive* implementation, we set $\mathbf{\Gamma}$ to be the identity matrix and $\mathbf{\Xi} = \mathbf{HG}$, where $\mathbf{G}$ is the generator matrix of Λ. We shall observe that the suboptimality of this naive lattice decoding, as compared to ML, entails a significant loss in the achievable diversity-vs-multiplexing tradeoff in most cases.

For a fixed, non-random, channel matrix $\mathbf{H}^c$, we have the following result.

Theorem 3 *Suppose that $\mathbf{H}^c$ has rank M, then the rate*

$$R_{\text{ld}}(\mathbf{H}^c, \rho) \triangleq M \log \rho + \log \ \det \left(\frac{1}{M} (\mathbf{H}^c)^{\mathsf{H}} \mathbf{H}^c \right) \tag{12}$$

is achievable by LAST coding and minimum Euclidean distance lattice decoding.

Proof. The proof relies on using LAST codes obtained from lattice ensembles satisfying Theorem 1 with a suitable translation vector and a spherical shaping region. The proof is based on the technical machinery introduced by Loeliger in [27]. In particular, the enabling tool in the analysis is Loeliger's ambiguity decoder. The details are presented in Appendix B. □

We observe that the suboptimality of R_{ld} in Theorem 3 is analogous to the loss of *one* in the SNR suffered by lattice decoding in SISO AWGN channels [27]. Next, we consider a random channel matrix $\mathbf{H}^c$ as defined in (1) and obtain an achievable diversity-vs-multiplexing tradeoff curve for LAST codes under naive lattice decoding. Following [10], we consider a family of LAST codes $\mathcal{C}_\rho$ for fixed M and T, obtained from lattices of a given dimension $2MT$ and indexed by their operating SNR ρ. The code $\mathcal{C}_\rho$ has rate $R(\rho)$ and error probability $P_e(\rho)$ (this is the average block error probability for a fixed code, where averaging is with respect to the random channel matrix $\mathbf{H}^c$). We have the following result.

Theorem 4 *For $M \leq N$ and any block length $T \geq 1$, there exists a sequence of full dimensional LAST codes that achieves diversity gain*

$$d(r) = \min\{T, 1 + N - M\}(M - r) \tag{13}$$

for $r \in [0, M]$ under naive minimum Euclidean distance lattice decoding. This coincides with the optimal $d^\star(r)$ for $T = 1$ (space-only coding) and squared channel matrix ($M = N$), or for any $N \geq M$, $T \geq 1 + N - M$, in the high-rate segment $r \in [M - 1, M]$.

Proof. The proof is deferred to Appendix C.

□

Theorem 4 shows that full dimensional LAST coding with the *naive* application of lattice decoding is optimal (in terms of the diversity-vs-multiplexing tradeoff) only in a few cases. On the contrary, it fails to take full advantage of larger block lengths ($T > 1$) and/or larger receiver diversity ($N > M$) for multiplexing gains $r < M - 1$. In fact, the difference between this achievable tradeoff and the optimal tradeoff widens as r decreases. While we realize that this is only a lower bound on the achievable diversity gain, yet this bound highlights the loss in performance entailed by lattice coding under lattice decoding. Furthermore, the numerical examples in Section 4 will validate this claim. The reason of this suboptimality can be traced back to the loss in the achievable rate of Theorem 3 with respect to the optimal (under unrestricted coding and decoding) achievable rate. The following two remarks are now in order.

1. In SISO AWGN channels, one can easily see that the loss in performance entailed by the naive implementation of lattice decoding vanishes as the SNR increases. On the other hand, Theorem 4 argues that the corresponding loss in quasi-static MIMO fading channels **persists** even as $\rho \to \infty$. This can be explained by noting that even with high **average** SNR, some of the channel eigenvalues can assume very small instantaneous values. With the straightforward application of lattice decoding, those *faded* eigenvalues *absorb* all the energy of the transmitted signal, and hence, result in significant performance degradations. In the next section, we will show that by using a MMSE-GDFE front end one can *neutralize* the effect of those faded eigenvalues and achieve the optimal tradeoff.

2. One can achieve other points on the optimal diversity-vs-multiplexing tradeoff by reducing the dimensionality of the lattice code and using a *clever* multiplexing scheme. For example, one can show that diagonal LAST codes (i.e., only one antenna is active at any point in time), based on lattices of dimension $2M$, achieve the point $d = MN, r = 0$. The proof follows from the same technical machinery used to prove Theorem 4. The suboptimality of this approach manifests itself in the fact that **the same scheme** fails to

achieve all the points on the tradeoff curve **simultaneously**. This suboptimality yields performance degradation for the schemes with low multiplexing gain at high transmission rates and performance degradation for the schemes with low diversity gains at high SNR. The scheme proposed in the next section avoids this drawback, and hence, allows for approaching the fundamental limits of MIMO channels with arbitrary parameters.

3.2 The generalized mod-Λ scheme and its optimality

In [22], Erez and Zamir showed that nested lattice codes achieve the AWGN channel capacity under lattice decoding, provided that the lattice decoder is modified by including a linear MMSE estimation stage and a random dither signal is used (implying a common randomness at transmitter and receiver). Random dithering renders the MMSE estimation error signal independent of the transmitted codeword (see also [23]). For a reason that will appear clearly later, we shall nickname Erez-Zamir scheme the "mod-Λ scheme". In this section we present a non-trivial generalization of the mod-Λ scheme to MIMO channels. We show that for fixed $\mathbf{H}^c$ and $T \to \infty$ LAST codes with the mod-Λ scheme achieve rates up to the *optimal* information rate[6] $\log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)$. We show also that LAST codes with the mod-Λ scheme achieve all points on the optimal diversity-vs-multiplexing tradeoff curve $d^\star(r)$, for block length $T \geq M + N - 1$.

We start by defining nested lattice codes (or Voronoi codes).

Definition 1 *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

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

and the modulo-lattice function

$$[\mathbf{y}] \mod \Lambda \triangleq \mathbf{y} - Q_\Lambda(\mathbf{y}).$$

◇

[6] This is the largest achievable information rate under the input constraint $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{T}}] = \frac{1}{2}\mathbf{I}$.

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 mod-Λ 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 $\mathbb{R}^{2MT}$. Assume that Λ_s has a second-order moment $\sigma^2(\Lambda_s) = 1/2$ (so that $\mathbf{u}$ uniformly distributed over $\mathcal{V}_s$ satisfies $\mathbb{E}[|\mathbf{u}|^2] = MT$). The transmitter selects a codeword $\mathbf{c} \in \mathcal{C}$, generates a dither signal $\mathbf{u}$ with uniform distribution over $\mathcal{V}_s$ and computes

$$\mathbf{x} = [\mathbf{c} - \mathbf{u}] \mod \Lambda_s \tag{14}$$

The signal $\mathbf{x}$ is then transmitted on the MIMO channel. Let $\mathbf{y}$ denote the corresponding channel output (we use the real channel model (9)). We replace the *scalar* scaling of [22] by a matrix multiplication by the forward filter matrix $\mathbf{F}$ of the MMSE-GDFE [24]. Moreover, instead of just adding the dither signal $\mathbf{u}$ at the receiver (as in [22]), we add the dither signal filtered by the upper triangular feedback filter matrix $\mathbf{B}$ of the MMSE-GDFE. The definitions and some useful properties of the MMSE-GDFE matrices $(\mathbf{F}, \mathbf{B})$ are given in Appendix A.

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

$$\begin{aligned}
\mathbf{y}' &= \mathbf{F}\mathbf{y} + \mathbf{B}\mathbf{u} \\
&= \mathbf{F}\left(\mathbf{H}(\mathbf{c} - \mathbf{u} + \boldsymbol{\lambda}) + \mathbf{w}\right) + \mathbf{B}\mathbf{u} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) - [\mathbf{B} - \mathbf{F}\mathbf{H}]\,(\mathbf{c} - \mathbf{u} + \boldsymbol{\lambda}) + \mathbf{F}\mathbf{w} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) - [\mathbf{B} - \mathbf{F}\mathbf{H}]\,\mathbf{x} + \mathbf{F}\mathbf{w} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) + \mathbf{e}'.
\end{aligned} \tag{15}$$

By construction, $\mathbf{x}$ is uniformly distributed over $\mathcal{V}_s$ and is independent of $\mathbf{c}$. One can also rewrite (15) as

$$\mathbf{y}' = \mathbf{B}\mathbf{c}' + \mathbf{e}' \tag{16}$$

where $\mathbf{c}' \in \Lambda_s + \mathbf{c}$. The remarkable fact in (15) and (16) is that the desired signal $\mathbf{c}$ is now translated by an unknown lattice point $\boldsymbol{\lambda} \in \Lambda_s$. However, since $\mathbf{c}$ and $\mathbf{c}' = \mathbf{c} + \boldsymbol{\lambda}$ 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 $\Lambda_s + \mathbf{c}$, rather than in the codeword $\mathbf{c}$ itself). It follows that in order to recover the information message, the decoder has to identify the coset $\Lambda_s + \mathbf{c}$ that contains $\mathbf{c}'$. This is achieved in two steps. The decoder first finds

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z} \in \mathbb{Z}^{2MT}} |\mathbf{y}' - \mathbf{B}\mathbf{G}\mathbf{z}|^2 \tag{17}$$

where $\mathbf{G}$ is the generator matrix of the channel coding lattice Λ_c (notice that (17) corresponds to applying the generalized minimum Euclidean distance lattice decoder defined in (11) to the channel output $\mathbf{y}$ with the choices $\mathbf{\Gamma} = \mathbf{F}$, $\mathbf{\Xi} = \mathbf{BG}$ and $\mathbf{a} = \mathbf{Bu}$.). Then, the decoded codeword is given by

$$\hat{\mathbf{c}} = [\mathbf{G}\hat{\mathbf{z}}] \mod \Lambda_s. \tag{18}$$

In a nutshell, lattice decoding as described above follows naturally as a consequence of the mod-Λ scheme: lattice decoding is not just a trick to make the receiver simpler, but it is an essential component of the whole construction. Finally, we note that because of the block diagonal structure of $\mathbf{H}$, $\mathbf{B}$ is also block diagonal with the $2M \times 2M$ upper triangular block $\mathbf{B}'$ repeated T times. By construction we have

$$\det\left(\mathbf{B}^{\mathsf{T}}\mathbf{B}\right) = \det\left((\mathbf{B}')^{\mathsf{T}}\mathbf{B}'\right)^T = \det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{2T}$$

(see Appendix A).

The optimality of LAST codes with the mod-Λ scheme and lattice decoding, in the limit of large T, is given by the following result.

Theorem 5 *For a fixed non-random channel matrix $\mathbf{H}^c$, the rate*

$$R_{\text{mod}}(\mathbf{H}^c, \rho) \triangleq \log \det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right) \tag{19}$$

is achievable by mod-Λ LAST coding.

Proof. We consider a sequence of nested lattices satisfying Theorem 2. Hence, the MMSE-GDFE estimation error signal

$$\mathbf{e}' = -\left[\mathbf{B} - \mathbf{FH}\right]\mathbf{x} + \mathbf{Fw} \tag{20}$$

converges in distribution (in the sense of divergence) to the noise vector $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$. This follows from Lemma 1 of Appendix A and from the fact that Λ_s is asymptotically good for MSE quantization, implying that $\mathbf{x} \rightarrow \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ as $T \rightarrow \infty$.

Intuitively, in the limit for large T, the channel (16) resulting from the mod-Λ construction is equivalent to sending a point $\mathbf{c}' \in \Lambda_c$ through a linear channel with matrix $\mathbf{B}$ plus an asymptotically Gaussian error signal $\mathbf{e}'$ independent of $\mathbf{c}'$. If $\mathbf{e}'$ was exactly Gaussian, the same steps in the proof of Theorem 3 would apply to this setup and we would see immediately that there exists a sequence of nested lattices such that, for sufficiently large T, the probability of error can be made smaller than any desired $\epsilon > 0$ provided that

$$R < \frac{1}{2}\log\det\left((\mathbf{B}')^{\mathsf{T}}\mathbf{B}'\right) = \log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right).$$

Note also that this holds for any $\mathbf{H}^c$ (also of non-full column rank), since $\mathbf{B}$ is always invertible (for any finite SNR ρ).

The only technical difficulty is that here the error $\mathbf{e}'$ is not exactly Gaussian for any finite T. In Appendix D we show that despite this problem the achievable rate is indeed given by (19), as if the estimation error was exactly Gaussian. □

Next, we consider the achievable diversity-vs-multiplexing tradeoff for LAST codes under the mod-Λ scheme. Suppose we have a family of nested lattices $\Lambda_s(\rho) \subseteq \Lambda_c(\rho)$ of fixed dimension $2MT$, indexed by ρ. As $\rho \to \infty$, the ratio $|\Lambda_c(\rho)/\Lambda_s(\rho)|$ increases as ρ^{rT} for some $r > 0$. This implies that the rate of the corresponding nested LAST codes is $\mathcal{C}_\rho$ is $R(\rho) = r \log \rho$. We shall show that there exist families of nested LAST codes for which the corresponding diversity gain is $d^\star(r)$. This is stated in the following theorem, which is the central contribution of this paper.

Theorem 6 *There exists a sequence of nested LAST codes with block length $T \geq M+N-1$ that achieves the optimal diversity-vs-multiplexing tradeoff curve $d^\star(r)$ for all $r \in [0, \min\{M,N\}]$ under the mod-Λ scheme.*

Proof. The main difficulty here is that we wish to prove the result for any fixed block length $T \geq M + N - 1$. Hence, we cannot use the sequences of nested lattices of Theorem 2, since for these sequences the coding lattice Λ_c satisfies Theorem 1 (the Minkowski-Hlawka theorem) only asymptotically for $T \to \infty$. On the other hand, Theorem 1 holds for any finite T. This motivates us to use the ensemble of lattices defined in [27] as coding lattice ensemble here. The key observation is that in order to achieve the diversity-vs-multiplexing optimal tradeoff we do not need a very "clever" shaping lattice. Indeed, any sequence of shaping lattices with finite covering efficiency (i.e., for which η_{cov}, as a function of the SNR ρ, is uniformly bounded by a constant $\beta < \infty$) can be used to achieve the optimal tradeoff. The details of the proof are given in Appendix E. Obviously, in any practical code construction one would look for a good shaping lattice, in order to achieve better power efficiency for finite ρ (recall that the diversity-vs-multiplexing tradeoff is achieved asymptotically for large ρ). □

It follows immediately from the arguments used to prove Theorems 5 and 6 that the mod-Λ construction achieves an average probability of error equals to the outage probability, assuming white Gaussian inputs, in the limit of a large block length (i.e., $T \to \infty$).

3.3 Where is the magic?

The generalized mod-Λ construction presented in the previous section has three main ingredients: 1) the nested lattice structure, 2) the random dither, and 3) the MMSE-GDFE lattice decoding. We now attempt to identify the roles of these three elements, and hence, highlight the various advantages offered by the generalized mod-Λ construction. To this end, we resort back to the *spherical* LAST codes used in Section 3.1 and characterize their performance under MMSE-GDFE lattice decoding without dithering.

Theorem 7 *There exists a sequence of spherical LAST codes with block length $T \geq M + N - 1$ that achieves the optimal diversity-vs-multiplexing tradeoff curve $d^\star(r)$ for all $r \in [0, \min\{M, N\}]$ under MMSE-GDFE lattice decoding. The coding/decoding scheme need no common randomness, i.e., achievability is obtained for a sequence of fixed codebooks.*

Proof. The main idea in the proof is to replace the random dither with the *optimal* translate and the shaping lattice with a sphere. In order to exploit the techniques used in the proof of Theorem 6, we add a dither which is uniformly distributed over the Voronoi cell of the coding lattice at the decoder. Adding such a dither cannot improve the performance of the receiver, and requires no common randomness since the dither is generated by the receiver and it is not known by the transmitter. The proof then follows in the footsteps of the proof of Theorem 6 as detailed in Appendix F. □

Theorem 7 argues that lattice coding (without random dithering and nesting) and MMSE-GDFE lattice decoding are **sufficient** to achieve the optimal diversity-vs-multiplexing tradeoff. As a corollary of Theorem 7, it is straightforward to see that spherical lattice coding with ML decoding also achieve the optimal tradeoff. In fact, ML decoding cannot be worse than minimum Euclidean distance lattice decoding. Now, we can identify the following additional advantages offered by random dithering and nested (Voronoi) coding.

1. The random dither renders the noise signal independent of the transmitted codeword. This fact along with the geometric uniformity of lattice coding (under lattice decoding) imply that the probability of error is independent of the transmitted codeword in the generalized mod-Λ construction. Hence, all the claims regarding the average probability of error extend naturally to the maximum probability of error. As it is clearly seen in the proof of Theorem 7, this is not generally true in the case of spherical lattice coding with the optimal translate, under both lattice decoding and ML decoding.

2. In practice, finding the optimal translate for spherical LAST codes may be prohibitive (especially for large dimensions). In these cases, randomizing the choice of the translate (i.e., the random dither) avoids the bad choices and saves computational power. Moreover, with the random dither, the transmitted power is also independent of the transmitted codeword.

3. While the complexity of lattice decoding is *almost*[7] independent of the shaping region, the

[7] With Voronoi codes, there is a slight increase in decoding complexity due to the last step of identifying the coset.

encoding complexity of spherical LAST codes is significantly higher than that of Voronoi LAST codes. The lack of structure in the carving region of spherical codes results in a look-up table encoder whereas encoding Voronoi codes reduces, again, to the search for the closest lattice point problem [25].

4 Numerical results

In this section, we present a selected set of numerical examples. Those examples are chosen to highlight three main points, namely: 1) the potential performance gains possible with LAST coding, 2) the gain offered by MMSE-GDFE lattice decoding over naive lattice decoding, and 3) the ability of random LAST coding with MMSE-GDFE lattice decoding to achieve the optimal diversity-vs-multiplexing tradeoff.

In order to illustrate the first point, we compare LAST coding with linear dispersion (LD) coding in Figure 1. We first observe that LD coding can be obtained as a special case of LAST coding as follows. After proper scaling and translation, the matrix codewords in an LD code can be written as [16]

$$\mathbf{S}(\mathbf{u}) = \sum_{\ell=1}^{m} \mathbf{G}_\ell u_\ell, \tag{21}$$

where $u_\ell \in \{0, ..., Q-1\}$, Q is the size of the input PAM constellation, $m = 2MT$, and $\{\mathbf{G}_\ell \in \mathbb{R}^{2M\times T}, \ell \in \{1, ..., 2MT\}\}$ are the spreading matrices of the LD code. By letting $\mathbf{g}_\ell = \text{vec}(\mathbf{G}_\ell)$, $\mathbf{G}_{LD} = [\mathbf{g_1}, ..., \mathbf{g_m}]$, and $\Lambda_m^{LD} = \{\mathbf{G}^{LD}\mathbf{u} : \mathbf{u} \in \mathbb{Z}^m\}$, we can now obtain the vector representation of the LD code as the intersection of Λ_m^{LD} with the region $\mathcal{R}^{LD}$ defined as the **image** of the m-dimensional hypercube under the mapping $\mathbf{G}^{LD}$ (i.e, $\mathcal{R}^{LD} = \{\mathbf{x} = \mathbf{G}^{LD}\mathbf{u} : \mathbf{u} \in \mathbb{R}^m, 0 \leq u_\ell \leq Q-1, \ell = \{1, ..., m\}$).

In Figure 1, we use the same generator matrix for both the LAST and the LD codes and report the performance with ML decoding. The difference in the performance can be, therefore, attributed to the the difference in the shaping region. In fact, the dependence of the shaping region in LD coding on the generator matrix of the lattice implies a fundamental limit on the achievable minimum Euclidean distance and coding gain of this class of codes (as argued in [7]). By relaxing this constraint on the shaping region, LAST coding avoids this limitation. We remark that the performance trend in Figure 1 was observed for other random choices of generator matrices (the results are not reported here for brevity).

Figure 2 illustrates the second point. In this figure, one can see that the naive application of lattice decoding allows for achieving full diversity **only** with vertical codes ($T = 1$) in symmetric configurations ($N = M$). For larger T, only by utilizing an MMSE-GDFE front end one can achieve full diversity with lattice decoding. Finally, Figures 3 and 4 validate the achievability of the optimal tradeoff with LAST coding and MMSE-GDFE lattice decoding. These figures report the performance of *random ensembles* of spherical and nested LAST codes (obtained via Construction A where $(n = 2MT, p, k)$ are the parameters of the linear code [27]) in a 2×2 MIMO system. As argued in [7], the optimality of the approach is illustrated in the constant gap between the probability of error curves and the outage probability at different SNRs and different rates (at sufficiently high SNR). Furthermore, the small gap between the performance of *random* LAST codes and the outage probability ($2-4$ dB at 10^{-5} block error rate) demonstrates that these codes rival the best ones available in the literature.

5 Conclusions

In this paper, we developed a novel framework for constructing optimal coding/decoding schemes for delay limited MIMO fading channels. In particular, we introduced the class of LAST codes. Within this class, we proposed a generalization of Erez and Zamir mod-Λ construction and proved its optimality with respect to the diversity-vs-multiplexing tradeoff. Through this generalization, we established the central role of MMSE-GDFE in approaching the fundamental limits of MIMO channels in the high SNR regime. Our results settle the open problem posed by Zheng and Tse on the existence of explicit coding constructions that achieve the optimal diversity-vs-multiplexing tradeoff. Furthermore, we prove the existence of lattice codes which are good for both AWGN channels and delay limited MIMO fading channels. The random coding arguments developed in this work can offer valuable guidelines for future works on optimal code constructions and low complexity decoding algorithms. Our current investigations explore two directions: 1) using the number theoretic tools proposed in [8] to further optimize the LAST codes (i.e., minimize the gap to the outage) and 2) developing low complexity variants of the generalized minimum Euclidean distance lattice decoder and a more precise characterization of the complexity of such decoders.

Appendices

In Appendix A, we review some known facts about MMSE-GDFE which will be needed later in the proofs. In the rest of the Appendices, we detail the proofs of our results.

A The MMSE-GDFE

Consider the real additive-noise MIMO linear channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{x}$ and $\mathbf{w}$ have mean zero, covariance $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{T}}] = \mathbb{E}[\mathbf{w}\mathbf{w}^{\mathsf{T}}] = \mathbf{I}$, and are mutually uncorrelated and where $\mathbf{H} \in \mathbb{R}^{n \times m}$.

The MIMO matched filter is given by the linear transformation defined by the matrix $\mathbf{H}^{\mathsf{T}}$, and its output is

$$\mathbf{y}_{\mathrm{mf}} = \mathbf{H}^{\mathsf{T}}\mathbf{y} = \mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{w}_{\mathrm{mf}} \tag{22}$$

$\mathbf{w}_{\mathrm{mf}}$ has covariance $\mathbf{H}^{\mathsf{T}}\mathbf{H}$.

The standard derivation of the MMSE-GDFE forward and feedback filter matrices is briefly outlined as follows. We seek a decision-feedback equalizer in the form

$$\mathbf{z} = \mathbf{F}_{\mathrm{mf}}\mathbf{y}_{\mathrm{mf}} - (\mathbf{B}_{\mathrm{mf}} - \mathbf{I})\widehat{\mathbf{x}} \tag{23}$$

where $\mathbf{F}_{\mathrm{mf}}, \mathbf{B}_{\mathrm{mf}} \in \mathbb{R}^{m\times m}$, and $\mathbf{B}_{\mathrm{mf}}$ is upper triangular and monic (i.e., it has unit diagonal elements). The vector $\widehat{\mathbf{x}}$ contains an estimate (hard-decisions) of the transmitted symbol $\mathbf{x}$ based on the equalizer output $\mathbf{z}$. Thanks to the strictly upper triangular form of $\mathbf{B}_{\mathrm{mf}} - \mathbf{I}$, (23) is recursively computable from the m-th to 1st component (going upward). Assuming $\widehat{\mathbf{x}} = \mathbf{x}$ (ideal feedback assumption), we find $\mathbf{F}_{\mathrm{mf}}$ and $\mathbf{B}_{\mathrm{mf}}$ such that the mean-square estimation error (MSE) $\mathbb{E}[|\mathbf{e}|^2]$, where $\mathbf{e} = \mathbf{z} - \mathbf{x}$, is minimized.

This can be obtained by imposing the orthogonality condition $\mathbb{E}[\mathbf{e}\mathbf{y}_{\mathrm{mf}}^{\mathsf{T}}] = \mathbf{0}$, by solving first with respect to $\mathbf{F}_{\mathrm{mf}}$ as a function of $\mathbf{B}_{\mathrm{mf}}$, and then finding the optimal $\mathbf{B}_{\mathrm{mf}}$ under the upper triangular and monic constraint.

After solving for $\mathbf{F}_{\mathrm{mf}}$, we obtain

$$\mathbf{F}_{\mathrm{mf}} = \mathbf{B}_{\mathrm{mf}}\left[\mathbf{H}^{\mathsf{T}}\mathbf{H} + \mathbf{I}\right]^{-1} = \mathbf{B}_{\mathrm{mf}}\mathbf{\Sigma}^{-1} \tag{24}$$

where we define the *system covariance matrix* $\mathbf{\Sigma} \triangleq \mathbf{H}^{\mathsf{T}}\mathbf{H} + \mathbf{I}$.

By substituting (24) into the expression of $\mathbf{e}$, we obtain

$$\mathbf{e} = \mathbf{B}_{\mathrm{mf}}\mathbf{\Sigma}^{-1}\mathbf{y}_{\mathrm{mf}} - \mathbf{B}_{\mathrm{mf}}\mathbf{x} = \mathbf{B}_{\mathrm{mf}}\mathbf{d}$$

where we let $\mathbf{d} \triangleq \mathbf{\Sigma}^{-1}\mathbf{y}_{\mathrm{mf}} - \mathbf{x}$. Since $\mathbf{B}_{\mathrm{mf}}$ is upper triangular and monic, $\mathbf{e}$ can be interpreted as a prediction error. Namely, we can write

$$e_k = d_k + \sum_{j=k+1}^{m} b_{k,j}d_j$$

Therefore, $-\sum_{j=k+1}^{m} b_{k,j}d_j$ is the linear MMSE estimate of d_k from the samples $d_{k+1}, \ldots, d_m$ (identified with the "past" of the sequence $\mathbf{d}$ with respect to the k-th component). Again by applying the orthogonality principle and using the fact that $\mathbf{B}_{\mathrm{mf}}$ must be the upper triangular we obtain that the MSE is minimized if $\mathbf{B}_{\mathrm{mf}}$ is the whitening filter for $\mathbf{d}$ (i.e., it makes the covariance matrix of $\mathbf{e}$ diagonal).

After some algebra, we get that $\mathbb{E}[\mathbf{d}\mathbf{d}^{\mathsf{T}}] = \mathbf{\Sigma}^{-1}$ (notice that $\mathbf{\Sigma}$ is always invertible). Let the Cholesky decomposition of $\mathbf{\Sigma}$ be

$$\mathbf{\Sigma} = \mathbf{B}_{\mathrm{mf}}^{\mathsf{T}} \mathbf{\Delta} \mathbf{B}_{\mathrm{mf}}$$

where $\mathbf{B}_{\mathrm{mf}}$ is upper triangular and monic and $\mathbf{\Delta}$ is diagonal with positive diagonal elements. It is immediate to check that

$$\mathbb{E}[\mathbf{e}\mathbf{e}^{\mathsf{T}}] = \mathbb{E}[\mathbf{B}_{\mathrm{mf}} \mathbf{d}\mathbf{d}^{\mathsf{T}} \mathbf{B}_{\mathrm{mf}}^{\mathsf{T}}] = \mathbf{\Delta}^{-1}$$

is diagonal, as desired. By substituting in (24), we obtain the corresponding forward filter matrix as

$$\mathbf{F}_{\mathrm{mf}} = \mathbf{\Delta}^{-1} \mathbf{B}_{\mathrm{mf}}^{-\mathsf{T}}$$

Any left-multiplication by a non-singular diagonal matrix of both $\mathbf{F}_{\mathrm{mf}}$ and $\mathbf{B}_{\mathrm{mf}}$ yields an equivalent MMSE-GDFE. In particular, we multiply by $\mathbf{\Delta}^{1/2}$ in order to make the covariance of the estimation error $\mathbf{e}$ equal to $\mathbf{I}$, and we obtain the MMSE-GDFE applied directly on the original channel output in the form

$$\mathbf{z} = \mathbf{F}\mathbf{y} - \overline{\mathcal{U}}(\mathbf{B})\widehat{\mathbf{x}}$$

where $\mathbf{B} = \mathbf{\Delta}^{1/2}\mathbf{B}_{\mathrm{mf}}$, $\overline{\mathcal{U}}(\cdot)$ takes the strictly upper triangular part of its argument, and $\mathbf{F} = \mathbf{B}^{-\mathsf{T}}\mathbf{H}^{\mathsf{T}}$. Under the ideal feedback assumption, the resulting error signal $\mathbf{e} = \mathbf{z} - \mathbf{x}$ has covariance $\mathbf{I}$.

Interestingly, we can define the augmented channel matrix

$$\widetilde{\mathbf{H}} = \left[\begin{array}{c} \mathbf{H} \\ \mathbf{I} \end{array} \right]$$

and its QR decomposition,

$$\widetilde{\mathbf{H}} = \widetilde{\mathbf{Q}}\mathbf{R}$$

where $\widetilde{\mathbf{Q}} \in \mathbb{R}^{(n+m)\times m}$ has orthonormal columns and $\mathbf{R} \in \mathbb{R}^{m\times m}$ is upper triangular with positive diagonal elements. Moreover, we denote by $\mathbf{Q} = \mathbf{H}\mathbf{R}^{-1}$ the upper $n \times m$ part of $\widetilde{\mathbf{Q}}$. Then, it is immediate to show that

$$\begin{array}{rcl} \mathbf{B} & = & \mathbf{R} \\ \mathbf{F} & = & \mathbf{Q}^{\mathsf{T}} \end{array} \tag{25}$$

Moreover, by construction we have $\mathbf{B}^{\mathsf{T}}\mathbf{B} = \widetilde{\mathbf{H}}^{\mathsf{T}}\widetilde{\mathbf{H}} = \mathbf{I} + \mathbf{H}^{\mathsf{T}}\mathbf{H}$.

Finally, we have the following

Lemma 1 *Let* $\mathbf{v} = (\mathbf{B} - \mathbf{F}\mathbf{H})\mathbf{x} + \mathbf{F}\mathbf{w}$, *where* $\mathbf{x}$ *and* $\mathbf{w}$ *are uncorrelated, with mean zero and covariance matrix* $\mathbf{I}$. *Then,* $\mathbb{E}[\mathbf{v}\mathbf{v}^{\mathsf{T}}] = \mathbf{I}$.

Proof. We have $\mathbb{E}[\mathbf{v}\mathbf{v}^{\mathsf{T}}] = (\mathbf{B} - \mathbf{F}\mathbf{H})(\mathbf{B} - \mathbf{F}\mathbf{H})^{\mathsf{T}} + \mathbf{F}\mathbf{F}^{\mathsf{T}}$. First, notice that

$$\begin{aligned} \mathbf{F}\mathbf{H} &= \mathbf{Q}^{\mathsf{T}}\mathbf{H} \\ &= \mathbf{B}^{-\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{H} \\ &= \mathbf{B}^{-\mathsf{T}}(\mathbf{H}^{\mathsf{T}}\mathbf{H} + \mathbf{I}) - \mathbf{B}^{-\mathsf{T}} \\ &= \mathbf{B} - \mathbf{B}^{-\mathsf{T}} \end{aligned} \tag{26}$$

and that

$$\mathbf{F}\mathbf{F}^{\mathsf{T}} = \mathbf{B}^{-\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{B}^{-1}$$

Hence,

$$\begin{aligned} \mathbb{E}[\mathbf{v}\mathbf{v}^{\mathsf{T}}] &= \mathbf{B}^{-\mathsf{T}}\mathbf{B}^{-1} + \mathbf{B}^{-\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{B}^{-1} \\ &= \mathbf{B}^{-\mathsf{T}}(\mathbf{I} + \mathbf{H}^{\mathsf{T}}\mathbf{H})\mathbf{B}^{-1} \\ &= \mathbf{I} \end{aligned} \tag{27}$$

□

B Proof of Theorem 3

We consider an ensemble of $2MT$-dimensional random lattices $\{\Lambda\}$ with fundamental volume V_f satisfying Theorem 1. The random lattice codebook is $\mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$, for some fixed translation vector $\mathbf{u}_0$ and where $\mathcal{R}$ is the $2MT$-dimensional sphere of radius $\sqrt{MT}$ centered in the origin. Hence, for each $\mathbf{x} \in \mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$ the input constraint $|\mathbf{x}|^2 \leq MT$ is satisfied.

Since $\mathbf{H}^c$ has rank M, the pseudoinverse of $\mathbf{H}$ is given by [31]

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathsf{T}} \tag{28}$$

and satisfies $\mathbf{H}^{\dagger}\mathbf{H} = \mathbf{I}$. Without loss of generality we consider the following decoder. In the first step, we apply the linear zero-forcing (ZF) equalizer given by $\mathbf{H}^{\dagger}$ in order to obtain

$$\mathbf{r} = \mathbf{H}^{\dagger}\mathbf{y} = \mathbf{x} + \mathbf{e}, \tag{29}$$

where $\mathbf{x} \in \Lambda$ is the transmitted point and $\mathbf{e} = \mathbf{H}^{\dagger}\mathbf{w}$, is a noise vector $\sim \mathcal{N}(\mathbf{0}, \frac{1}{2}(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1})$. In the second step, we apply the *ambiguity lattice decoder* of [27]. This decoder is defined by a decision region $\mathcal{E} \subset \mathbb{R}^{2MT}$ and outputs $\widehat{\mathbf{x}} \in \Lambda$ if $\mathbf{r} \in \mathcal{E} + \widehat{\mathbf{x}}$ and there exists no other point $\mathbf{x}' \in \Lambda$ such that $\mathbf{r} \in \mathcal{E} + \mathbf{x}'$. We define the ambiguity event $\mathcal{A}$ as the event that the received point $\mathbf{r}$ belongs to $\{\mathcal{E} + \mathbf{x}\} \cap \{\mathcal{E} + \mathbf{x}'\}$ for some pair of distinct lattice points $\mathbf{x}, \mathbf{x}' \in \Lambda$. If $\widehat{\mathbf{x}} \neq \mathbf{x}$ or $\mathcal{A}$ occur, we have error.

For given Λ and $\mathcal{E}$ we have

$$P_e(\mathcal{E}|\Lambda) \leq \Pr(\mathbf{e} \notin \mathcal{E}) + \Pr(\mathcal{A}) \tag{30}$$

By taking the expectation over the ensemble of random lattices, from Theorem 4 of [27] we obtain

$$\overline{P_e}(\mathcal{E}) \stackrel{\triangle}{=} \mathbb{E}_\Lambda[P_e(\mathcal{E}|\Lambda)] \leq \Pr(\mathbf{e} \notin \mathcal{E}) + (1+\delta)\frac{V(\mathcal{E})}{V_f} \tag{31}$$

for arbitrary $\delta > 0$.

We choose as decision region the ellipsoid defined by

$$\mathcal{E}_{T,\gamma} \stackrel{\triangle}{=} \left\{\mathbf{z} \in \mathbb{R}^{2MT} \ : \ \mathbf{z}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{z} \leq MT(1+\gamma)\right\} \tag{32}$$

It follows from standard typicality arguments that for any $\epsilon > 0$ and $\gamma > 0$ there exists $T_{\gamma,\epsilon}$ such that for all $T > T_{\gamma,\epsilon}$

$$\Pr(\mathbf{e} \notin \mathcal{E}_{T,\gamma}) < \epsilon/2 \tag{33}$$

Hence, for sufficiently large T, there exists at least a lattice $\Lambda^\star$ in the ensemble with error probability satisfying

$$P_e(\mathcal{E}_{T,\gamma}|\Lambda^\star) \leq \epsilon/2 + (1+\delta)\frac{V(\mathcal{E}_{T,\gamma})}{V_f} \tag{34}$$

For this lattice, we choose the translation vector $\mathbf{u}_0^\star$ such that (8) holds. By letting $|\mathcal{C}(\Lambda^\star, \mathbf{u}_0^\star, \mathcal{R})| = \exp(TR)$, we can write

$$P_e(\Lambda^\star, \mathcal{E}_{T,\gamma}) \leq \epsilon/2 + (1+\delta)\frac{V(\mathcal{E}_{T,\gamma})\exp(TR)}{V(\mathcal{R})} \tag{35}$$

¿From standard geometry formulas, we have

$$\begin{aligned} \frac{V(\mathcal{E}_{T,\gamma})}{V(\mathcal{R})} &= (1+\gamma)^{MT}\text{det}\left(\mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1/2} \\ &= (1+\gamma)^{MT}\left(\frac{M}{\rho}\right)^{MT} \text{det}\left((\mathbf{H}^c)^\mathsf{H}\mathbf{H}^c\right)^{-T} \end{aligned} \tag{36}$$

where we have used the definition of $\mathbf{H}$ in terms of $\mathbf{H}^c$.

The second term in the upper-bound (35) can be made smaller than $\epsilon/2$ for sufficiently large T if

$$R < \frac{1}{T}\log\frac{V(\mathcal{R})}{V(\mathcal{E}_{T,\gamma})} = M\log\rho + \log\det\left(\frac{1}{M}(\mathbf{H}^c)^\mathsf{H}\mathbf{H}^c\right) - \gamma' \tag{37}$$

where $\gamma' \to 0$ as $\gamma \to 0$. This shows the achievability of the rate $R_{\text{ld}}(\mathbf{H}^c, \rho)$ in (12) with the ambiguity decoder. The final step in the proof follows by noticing that with this choice of decision region in (32), the probability of error of the ambiguity decoder upper-bounds that of the generalized minimum Euclidean distance lattice decoder (11) with the choices $\mathbf{\Gamma} = \mathbf{I}$, $\mathbf{\Xi} = \mathbf{HG}$ and $\mathbf{a} = -\mathbf{H}\mathbf{u}_0^\star$.

C Proof of Theorem 4

We consider an ensemble of $2MT$-dimensional random lattices $\{\Lambda\}$ with fundamental volume V_f satisfying Theorem 1. The random lattice codebook is $\mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$, for some fixed translation vector $\mathbf{u}_0$ and where $\mathcal{R}$ is the $2MT$-dimensional sphere of radius $\sqrt{MT}$ centered in the origin. Hence, for each $\mathbf{x} \in \mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$ the input constraint $|\mathbf{x}|^2 \leq MT$ is satisfied.

We upper-bound the average probability of error (average over the channel and over the lattice ensemble) as

$$\overline{P_e}(\rho) \triangleq \mathbb{E}_\Lambda[P_e(\rho)] \leq \Pr(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)) + \mathbb{E}_\Lambda \left[\Pr(\text{error}, R_{\text{ld}}(\mathbf{H}^c, \rho) > R(\rho)|\Lambda)\right] \quad (38)$$

In order to compute $\Pr\left(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)\right)$, we follow in the footsteps of [10]. Denoting $R = r\log(\rho)$ and $\det\left((\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right) = \rho^{-\sum_{i=1}^M \alpha_i}$, where $\alpha_i \triangleq -\log\lambda_i/\log\rho$ and where $0 \leq \lambda_1 \leq \cdots \leq \lambda_M$ are the eigenvalues of $(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c$, we can write

$$\Pr\left(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)\right) = \Pr\left(\sum_{i=1}^M \alpha_i \geq M - r\right) \quad (39)$$

Hence,

$$\Pr\left(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)\right) = \mathbb{E}_{\boldsymbol{\alpha}}\left[1\left\{\sum_{i=1}^M \alpha_i \geq M - r\right\}\right] \quad (40)$$

Using the fact that $\{\lambda_1, \ldots, \lambda_M\}$ follow a Wishart distribution [1, 10], it is possible to show that

$$\mathbb{E}_{\boldsymbol{\alpha}}\left[1\left\{\sum_{i=1}^M \alpha_i \geq M - r\right\}\right] \doteq \int_{\mathcal{B}} \exp\left(-\log(\rho)\sum_{i=1}^M (2i - 1 + N - M)\alpha_i\right) d\boldsymbol{\alpha} \quad (41)$$

where $\mathcal{B} \subseteq \mathbb{R}^M$ is defined by $\alpha_1 \geq \cdots \geq \alpha_M \geq 0$ and by $\sum_{i=1}^M \alpha_i \geq M - r$. As a consequence of Varadhan's lemma [32], we obtain

$$\begin{aligned}
-d_c &\triangleq \lim_{\rho\to\infty} \frac{\log\Pr\left(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)\right)}{\log\rho} \\
&= \lim_{z\to\infty} \frac{1}{z}\log\int_{\mathcal{B}} \exp\left(-z\sum_{i=1}^M (2i - 1 + N - M)\alpha_i\right) d\boldsymbol{\alpha} \\
&= -\inf_{\boldsymbol{\alpha}\in\mathcal{B}} \sum_{i=1}^M (2i - 1 + N - M)\,\alpha_i \qquad (42)
\end{aligned}$$

which can be written more concisely as

$$\Pr\left(R_{\text{ld}}(\mathbf{H}^c, \rho) \leq R(\rho)\right) \doteq \rho^{-d_c}, \quad (43)$$

It is straightforward to see that the minimization in the last line of (42) is achieved for $\alpha_1 = M - r$ and $\alpha_i = 0$ for all $i > 1$, yielding $d_c = (1 + N - M)(M - r)$.

Now, let $P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)$ denote the probability of error for a given choice of Λ and rate $R(\rho)$ given that the channel matrix has determinant det$((\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c) = \rho^{-\sum_i \alpha_i}$ and let $\mathcal{B}'$ denote the region defined by $\alpha_1 \geq \cdots \geq \alpha_M \geq 0$ and by $\sum_{i=1}^M \alpha_i \leq M - r$. We have

$$\begin{array}{l} \mathbb{E}_\Lambda\left[\text{Pr}\left(\text{error}, R_{\text{ld}}(\mathbf{H}^c, \rho) > R(\rho)|\Lambda\right)\right] = \int_{\mathcal{B}'} p\left(\boldsymbol{\alpha}\right) \mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)]d\boldsymbol{\alpha} \\ \doteq \int_{\mathcal{B}'} \exp\left(-\log(\rho)\sum_{i=1}^M (2i - 1 + N - M)\alpha_i\right) \mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)]d\boldsymbol{\alpha} \end{array} \quad (44)$$

In order to bound $\mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)]$, we apply again the ambiguity decoder to the ZF channel output (29) with decision region $\mathcal{E}_{T,\gamma}$ defined in (32). Let $\mathbf{H}^{\mathsf{T}}\mathbf{H} = \mathbf{V}\mathbf{S}\mathbf{V}^{\mathsf{T}}$, with $\mathbf{V} \in \mathbb{R}^{M \times M}$ orthogonal and $\mathbf{S}$ diagonal with non-negative (positive almost surely) diagonal elements. Then, $\mathbf{e}' = \mathbf{S}^{1/2}\mathbf{V}^{\mathsf{T}}\mathbf{e}$, where $\mathbf{e}$ is defined in (29), is $\sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$. Using this fact, we obtain the Chernoff bound

$$\begin{aligned} \text{Pr}(\mathbf{e} \notin \mathcal{E}_{T,\gamma}) &= \text{Pr}\left(\mathbf{e}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{H}\mathbf{e} \geq MT(1+\gamma)\right) \\ &= \text{Pr}\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \\ &\leq \min_{\lambda \geq 0} \ \exp\left(-MT\left(\lambda(1+\gamma) + \log(1-\lambda)\right)\right) \\ &= (1+\gamma)^{MT}e^{-MT\gamma} \end{aligned} \quad (45)$$

Using the optimal translate for every lattice in the ensemble and noticing that $|\mathcal{C}_\rho| = \rho^{rT}$, we get

$$V_f \geq V(\mathcal{R})\rho^{-rT} \quad (46)$$

where $\mathcal{R}$ is the sphere of radius $\sqrt{MT}$ centered in the origin. From Theorem 4 of [27] and (36) we find, for all arbitrary $\delta > 0$,

$$\begin{aligned} \mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)] &\leq (1+\gamma)^{MT}e^{-MT\gamma} + (1+\delta)(1+\gamma)^{MT}\left(\frac{\rho}{M}\right)^{-MT}\rho^{rT}\text{det}\left((\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{-T} \\ &= (1+\gamma)^{MT}\left[e^{-MT\gamma} + c_1\rho^{-T\left(M-r-\sum_{i=1}^M \alpha_i\right)}\right] \end{aligned} \quad (47)$$

where c_1 does not depend on ρ.

Finally, we let $\gamma = \log\rho$ and we use (47) in (44). By applying again Varadhan's lemma we obtain that

$$\mathbb{E}_\Lambda\left[\text{Pr}\left(\text{error}, R_{\text{ld}}(\mathbf{H}^c, \rho) > R(\rho)|\Lambda\right)\right] \doteq \rho^{-d_n}, \quad (48)$$

where

$$d_n = \inf_{\boldsymbol{\alpha} \in \mathcal{B}'}\left\{\sum_{i=1}^M (2i - 1 + N - M - T)\alpha_i + T(M - r)\right\} \quad (49)$$

It is easily seen that if $1 + N - M \leq T$ then the minimum is obtained by letting $\alpha_1 = M - r$ and $\alpha_i = 0$ for all $i > 1$ and yields $d_n = (1 + N - M)(M - r)$, while if $1 + N - M > T$ then the minimum is obtained for $\alpha_i = 0, \;\; \forall\, i$, yielding $T(M - r)$. In the first case, the exponents of the two terms in the upper-bound (38) coincide. In the second case, the second term of (38) dominates the first term. We conclude that the exponent of the upper-bound is given by (13).

By using (43) and (48) in (38), we obtain that there exists at least a sequence of lattice codes $\{\mathcal{C}_\rho^\star\}$ in the ensemble that achieves the diversity-vs-multiplexing tradeoff given by (13) . The final statement of Theorem 4 follows by noticing that for $T = 1$ the optimal tradeoff is given by [10] $N(1 - r/M)$, that for $M = N$ coincides with (13), and that for any $N \geq M$ and $T \geq 1 + N - M$ we obtain the straight-line segment joining the points $(r = M - 1, d = N - M + 1)$ and $(r = M, d = 0)$, which coincides with $d^\star(r)$.

D Proof of Theorem 5

We consider an ensemble of $2MT$-dimensional nested lattices $\{\Lambda_s \subseteq \Lambda_c\}$ satisfying Theorem 2. Consequently, as $T \to \infty$ $\{\Lambda_c\}$ asymptotically satisfies Theorem 1. We denote by V_c and V_s the fundamental volumes of Λ_c and Λ_s, respectively. By construction, $R = \frac{1}{T} \log V_s/V_c$ and V_c is fixed (constant with T). Moreover, Λ_s has second-order moment $\sigma^2(\Lambda_s) = 1/2$, so that the input power constraint is satisfied (recall that the input $\mathbf{x}$ of the mod-Λ channel is uniformly distributed over $\mathcal{V}_s$).

Since $\mathbf{B}$ is invertible, we obtain the equivalent channel output

$$\mathbf{y}'' = \mathbf{B}^{-1}\mathbf{y}' = \mathbf{c}' + \mathbf{e}''$$

where $\mathbf{c}' \in \Lambda_c$. Then, we apply the *ambiguity lattice decoder* of [27] with decision region

$$\mathcal{E}_{T,\gamma} \stackrel{\triangle}{=} \left\{\mathbf{z} \in \mathbb{R}^{2MT} \;:\; \mathbf{z}^\mathsf{T}\mathbf{B}^\mathsf{T}\mathbf{B}\mathbf{z} \leq MT(1+\gamma)\right\} \tag{50}$$

The probability of error for given Λ_c is upper-bounded by

$$P_e(\mathcal{E}_{T,\gamma}|\Lambda_c) \leq \Pr(\mathbf{e}'' \notin \mathcal{E}_{T,\gamma}) + \Pr(\mathcal{A}) \tag{51}$$

where $\mathcal{A}$ is the ambiguity event defined as in the Proof of Theorem 3. By taking the expectation over the ensemble of random lattices, from Theorem 4 of [27] we obtain

$$\overline{P_e}(\mathcal{E}) \stackrel{\triangle}{=} \mathbb{E}_{\Lambda_c}[P_e(\mathcal{E}|\Lambda_c)] \leq \Pr(\mathbf{e}'' \notin \mathcal{E}_{T,\gamma}) + (1+\delta)\frac{V(\mathcal{E}_{T,\gamma})}{V_c} \tag{52}$$

where $\delta \to 0$ as $T \to \infty$, since by construction the sequence $\{\Lambda_c\}$ satisfies Theorem 1 for large T. By using the fact that $V_c = V_s \exp(-TR)$ and that

$$V(\mathcal{E}_{T,\gamma}) = (1+\gamma)^{MT}\text{det}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1/2}V(\mathcal{B}(\sqrt{MT}))$$

where $V(\mathcal{B}(\sqrt{MT}))$ is the volume of a $2MT$-dimensional sphere of radius $\sqrt{MT}$, we obtain

$$\Pr(\mathcal{A}) \leq (1{+}\delta) \exp\left(-T\left[\log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^\mathsf{T}\mathbf{H}^c\right) - M\log(1+\gamma) - \frac{1}{T}\log\frac{V(\mathcal{B}(\sqrt{MT}))}{V_s} - R\right]\right) \tag{53}$$

Since $\{\Lambda_s\}$ is a good sequence of lattice for MSE quantization and, by construction, the second-order moment of Λ_s is equal to $1/2$ for each T, using standard geometry and Stirling formulas it is easy to show that

$$\frac{1}{T}\log\frac{V(\mathcal{B}(\sqrt{MT}))}{V_s} \rightarrow 0$$

as $T \rightarrow \infty$. Therefore, since $\gamma > 0$ is arbitrary, we obtain the upper-bound

$$\Pr(\mathcal{A}) \leq \exp\left(-T\left[\log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^\mathsf{T}\mathbf{H}^c\right) - R - \gamma'\right]\right) \tag{54}$$

where $\gamma' \rightarrow 0$ as $T \rightarrow \infty$. We conclude that for all $\epsilon > 0$ there exists a sequence of pairs of nested lattices $\{\Lambda_s \subseteq \Lambda_c^\star\}$ for which the ambiguity probability is upper-bounded by $\epsilon/2$, for sufficiently large T, provided that

$$R < \log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^\mathsf{T}\mathbf{H}^c\right)$$

The proof is then complete if we show that $\Pr(\mathbf{e}'' \notin \mathcal{E}_{T,\gamma}) < \epsilon/2$ for arbitrary $\gamma > 0, \epsilon > 0$ and sufficiently large T. Recalling the definition of $\mathcal{E}_{T,\gamma}$, this condition can be written in the more convenient form

$$\Pr(|\mathbf{e}'|^2 \geq MT(1+\gamma)) \leq \epsilon/2 \tag{55}$$

where $\mathbf{e}' = -\left[\mathbf{B} - \mathbf{F}\mathbf{H}\right]\mathbf{x} + \mathbf{F}\mathbf{w}$ is the MMSE estimation error signal and $\mathbf{x} \sim$ Uniform over $\mathcal{V}_s$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ are statistically independent.

For reasons that will become clear later, we consider a *noisier* system by adding the noise vector $\mathbf{w}_2$ to the received signal before processing (we will determine the variance of $\mathbf{w}_2$ later). So, we replace $\mathbf{e}'$ by the sum $\mathbf{e}' + \mathbf{F}\mathbf{w}_2$ in equation (55). In order to use the Chernoff bound as in the proof of Theorem 4, we need to replace the self-noise $\mathbf{x}$ by a white Gaussian vector g (with a possibly higher variance). To this end, we follow in the footsteps of [22]. In particular, following the argument of Lemma 11 in [22], we obtain that

$$f_\mathbf{x}(\mathbf{z}) \leq \eta_{\text{cov}}^{2MT} f_\mathbf{v}(\mathbf{z}), \tag{56}$$

where $f_\mathbf{x}(\mathbf{z})$ is the pdf of $\mathbf{x}$, $f_\mathbf{v}(\mathbf{z})$ is the pdf of a random vector uniformly distributed over the covering sphere $\mathcal{B}(r_{\text{cov}})$, of radius $r_{\text{cov}} = r_{\text{cov}}(\Lambda_s)$ and where $\eta_{\text{cov}} = r_{\text{cov}}(\Lambda_s)/r_{\text{eff}}(\Lambda_s)$ denotes the covering efficiency of Λ_s.

Let $\overline{r} = \sqrt{\frac{MT+1}{MT}} r_{\rm cov}$, and let σ^2 denote the second-order moment of the sphere $\mathcal{B}(\overline{r})$. Using standard geometry formulas we obtain

$$\sigma^2 = G^* V \left(\mathcal{B}(\overline{r})\right)^{1/MT} = \frac{\overline{r}^2}{2MT+2} = \frac{r_{\rm cov}^2}{2MT} \tag{57}$$

where G^* denotes the normalized second-order moment of a sphere in $2MT$ dimensions.

Let g be a random $2MT$-dimensional vector $\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, and denote by $f_{\mathbf{g}}(\boldsymbol{\nu})$ its pdf. For $|\mathbf{z}| = \sqrt{2MT}\sigma = r_{\rm cov}$ we have

$$-\frac{1}{2MT} \log f_{\mathbf{g}}(\mathbf{z}) = \frac{1}{2} \log 2\pi e \sigma^2 \tag{58}$$

Also, for $|\mathbf{z}| = r_{\rm cov}$ we have

$$\begin{aligned} -\frac{1}{2MT} \log f_{\mathbf{v}}(\mathbf{z}) &= \frac{1}{2} \log V(\mathcal{B}(r_{\rm cov}))^{1/MT} \\ &= \frac{1}{2} \log \frac{r_{\rm cov}^2}{2(MT+1)G^*} \\ &= \frac{1}{2} \log \frac{MT\sigma^2}{(MT+1)G^*} \\ &= \frac{1}{2} \log 2\pi e \sigma^2 - \frac{1}{2} \log \left(2\pi e \left(1 + \frac{1}{MT}\right) G^*\right) \end{aligned} \tag{59}$$

By putting together (58) and (59), and from the fact that $f_{\mathbf{g}}(\mathbf{z})$ is decreasing in $|\mathbf{z}|$ while $f_{\mathbf{v}}(\mathbf{z})$ is zero for $|\mathbf{z}| > r_{\rm cov}$ and constant for $|\mathbf{z}| \leq r_{\rm cov}$, we obtain that, for all $\mathbf{z} \in \mathbb{R}^{2MT}$,

$$f_{\mathbf{v}}(\mathbf{z}) \leq \exp\left(MT \log\left(2\pi e \left(1 + \frac{1}{MT}\right) G^*\right)\right) f_{\mathbf{g}}(\mathbf{z}) \tag{60}$$

Moreover, by using the fact that $G^* = \frac{(MT!)^{1/MT}}{2\pi(MT+1)}$ and the Stirling formula $\log n! = n \log n - n + o(n)$, we obtain

$$MT \log\left(2\pi e \left(1 + \frac{1}{MT}\right) G^*\right) = o(MT)$$

therefore, the "blowing-up" factor in front of $f_{\mathbf{g}}(\mathbf{z})$ in (60) is a sub-exponentially increasing function $\exp(o(MT))$ of T.

We define the error vector

$$\widetilde{\mathbf{e}} = (\mathbf{B} - \mathbf{F}\mathbf{H})\mathbf{g} + \mathbf{F}(\mathbf{w} + \mathbf{w}_2) \tag{61}$$

where we choose $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, (\sigma^2 - 1/2)\mathbf{I})$. Since the second-order moment of the covering sphere $\mathcal{B}(r_{\rm cov})$ is given by $r_{\rm cov}^2/(2MT+2) \leq \sigma^2$, and it is certainly not smaller than the second-order moment of $\mathcal{V}_s$, which by construction is equal to $1/2$, we have that $\sigma^2 - 1/2$ is non-negative, hence the additional noise $\mathbf{w}_2$ is well-defined.

From the bounds developed above, it is immediate to conclude that

$$\Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \leq \exp(o(MT))\eta_{\text{cov}}^{2MT} \Pr\left(|\widetilde{\mathbf{e}}|^2 \geq MT(1+\gamma)\right) \tag{62}$$

Notice that $\mathbf{g} \sim \mathbf{w}+\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. From Lemma 1 we obtain that $\widetilde{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, therefore, we can use the same Chernoff bounding approach used in the proof of Theorem 4 and obtain

$$\Pr\left(|\widetilde{\mathbf{e}}|^2 \geq MT(1+\gamma)\right) \leq \left(\frac{1+\gamma}{2\sigma^2}\right)^{MT} \exp\left(-MT\left(\frac{1+\gamma}{2\sigma^2} - 1\right)\right) \tag{63}$$

By using (63) in (62) we obtain

$$\begin{aligned} \Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) &\leq \exp(o(MT))\eta_{\text{cov}}^{2MT} \left(\frac{1+\gamma}{2\sigma^2}\right)^{MT} \exp\left(-MT\left(\frac{1+\gamma}{2\sigma^2} - 1\right)\right) \\ &= \exp\left(-MT\left(\frac{1+\gamma}{2\sigma^2} - 1 - \log\frac{1+\gamma}{2\sigma^2} - 2\log\eta_{\text{cov}} - \frac{o(MT)}{MT}\right)\right) \\ &\leq \epsilon/2 \end{aligned} \tag{64}$$

where the last inequality holds for sufficiently large T, since $\eta_{\text{cov}} \to 1$ (recall that Λ_s is a good sequence of lattices for covering) and by noticing that

$$\frac{1+\gamma}{2\sigma^2} - 1 - \log\frac{1+\gamma}{2\sigma^2} > 0$$

for some arbitrary $\gamma > 0$, since the function $x - 1 - \log x$ is strictly positive for $x \geq 0$, $x \neq 1$, and since $\sigma^2 \to 1/2$ as $T \to \infty$. The fact that σ^2 tends to $1/2$ can be seen as follows:

$$\begin{aligned} \sigma^2 &= \frac{r_{\text{cov}}^2}{2MT} \\ &= \frac{\eta_{\text{cov}}^2 r_{\text{eff}}(\Lambda_s)^2}{2MT} \\ &= \frac{\eta_{\text{cov}}^2}{2MT}\left[\frac{V_s}{V(\mathcal{B}(1))}\right]^{1/MT} \\ &= \frac{\eta_{\text{cov}}^2}{2MT}\frac{\sigma^2(\Lambda_s)/G(\Lambda_s)}{V(\mathcal{B}(1))^{1/MT}} \end{aligned} \tag{65}$$

Using the fact that $\sigma^2(\Lambda_s) = 1/2$ by construction, that $G(\Lambda_s) \to \frac{1}{2\pi e}$ since $\{\Lambda_s\}$ is good for MSE quantization, and the formula

$$V(\mathcal{B}(1)) = \frac{\pi^{MT}}{\Gamma(MT+1)}$$

of the volume of the unit-radius sphere in $2MT$ dimensions, we obtain that $\lim_{T\to\infty}\sigma^2 = 1/2$. This concludes the proof. Interestingly, we remark that by replacing the self noise $\mathbf{x}$ by a Gaussian random vector $\mathbf{g}$ of slightly larger variance and adding noise $\mathbf{w}_2$ provides only a vanishing increase in error probability for large block length T.

E Proof of Theorem 6

We consider the Loeliger ensemble of mod-p lattices defined in [27] (see also [22, 28]). 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

$$\Lambda_p = \kappa \left(\mathbf{g}\mathbb{Z}_p + p\mathbb{Z}^{2MT} \right) \tag{66}$$

where $p \to \infty$, $\kappa \to 0$ is a scaling coefficient adjusted such that the fundamental volume $V_f = \kappa^{2MT} p^{2MT-1} = 1$, $\mathbb{Z}_p$ denotes the field of mod-p integers, and $\mathbf{g} \in \mathbb{Z}_p^{2MT}$ 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 $r^2_{\text{cov}} = 1/2$ in order to satisfy the input power constraint. The coding lattice is obtained as $\Lambda_c = 1/\tau\Lambda_s$, where $\tau = \lfloor \rho^{r/2M} \rfloor$ in order to satisfy the transmission rate constraint that $R(\rho) \doteq r \log \rho$. This yields the fundamental volumes

$$\begin{aligned} V_f(\Lambda_s) \triangleq V_s &= \zeta^{2MT} && (67) \\ V_f(\Lambda_c) \triangleq V_c &= \left(\frac{\zeta}{\tau}\right)^{2MT} && (68) \end{aligned}$$

Since our proof relies on self similar lattices for constructing nested codes, we need to establish that the lattices in the ensemble are *reasonably* good for both channel coding and shaping. Our first step is to expurgate the ensemble in (66) appropriately such that the remaining lattices in the expurgated ensemble satisfy an upper-bound on the covering efficiency that grows only logarithmically with ρ. Let $r_p = r_{\text{eff}}/\log(\rho)$. Then, using the uniform distribution of the lattice points and a simple union bounding argument (the same as the one used in [28]), we get

$$\Pr\left\{ (\Lambda_p + \mathcal{B}(r_p)) \text{ is a packing} \right\} \geq 1 - \left(\frac{2r_p + d}{r_{\text{eff}}} \right)^{2MT}, \tag{69}$$

where $2d = \kappa\sqrt{2MT}$ is the diagonal of the elementary cube with side κ. It then follows that

$$\begin{aligned} \Pr\left\{ \eta_{\text{pack}} = \frac{r_{\text{pack}}}{r_{\text{eff}}} \geq \frac{1}{\log(\rho)} \right\} &= \Pr\left\{ (\Lambda_p + \mathcal{B}(r_p)) \text{ is a packing} \right\} \\ &\geq \frac{(\log(\rho))^{2MT} - (2 + d/r_{\text{eff}})^{2MT}}{(\log(\rho))^{2MT}}, && (70) \end{aligned}$$

where d can be made as small as desired by letting $p \to \infty$.

This motivates us to expurgate the ensemble by removing the lattices with $\eta_{\text{pack}} < \frac{1}{\log(\rho)}$. Now, we need to upper-bound the covering efficiency of lattices in the expurgated ensemble. Without loss of generality, we consider a point "$\mathbf{a}$" in the fundamental region at a distance r_{cov} from the origin. Using the convexity of the fundamental region, we then construct a cone **inside** the fundamental region with apex at "$\mathbf{a}$" and axis passing through the origin by connecting "$\mathbf{a}$" to all the points in the projection of the packing sphere over the orthogonal subspace to the cone axis. Since the volume of this cone is smaller than or equal to the fundamental volume $V_f = 1$, one can see that $\eta_{\text{cov}} \leq \eta_{max} = O\left((\log(\rho))^{2MT-1}\right)$ for all lattices in the expurgated ensemble. In order to use the Minkowski-Hlwaka theorem with the expurgated ensemble, we will need the following relation

$$\mathbb{E}_{\Lambda_{exp}}(\chi) \leq \frac{(\log(\rho))^{2MT}}{(\log(\rho))^{2MT} - (2 + d/r_{\text{eff}})^{2MT}} \mathbb{E}_{\Lambda_p}(\chi), \tag{71}$$

where $\mathbb{E}_{\Lambda_{exp}}(.)$ is the expectation with respect to the expurgated ensemble, $\mathbb{E}_{\Lambda_p}(.)$ is the expectation with respect to the ensemble in (66)[8], and χ is a non-negative random variable (i.e., $\chi \geq 0$).

We use again the *ambiguity lattice decoder* of [27] with decision region $\mathcal{E}_{T,\gamma}$ given in (50). This way, we obtain the following upper bound on the average error probability (averaged over the expurgated code ensemble for fixed channel matrix) as a modified version of (52) where now $\delta \to 0$ as $p \to \infty$, $V_c \geq V_s \rho^{-rT}$, and equation (71) is enforced

$$\Pr(\mathcal{A}|\mathbf{H}^c) \leq (1+\delta) \frac{(\log(\rho))^{2MT}}{(\log(\rho))^{2MT} - (2 + d/r_{\text{eff}})^{2MT}} \left((1+\gamma)\eta_{\text{max}}^2\right)^{MT} \rho^{rT} \text{det}\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{-T}. \tag{72}$$

Next, we consider the first term in the union bound, given by $\Pr(|\mathbf{e}'|^2 \geq MT(1+\gamma))$. This term does not depend on the channel matrix, and as in the proof of Theorem 4 our goal is to show that

$$\Pr(|\mathbf{e}'|^2 \geq MT(1+\gamma)) \dot{\leq} \rho^{-d^\star(r)}$$

i.e., that this term can be neglected as it is exponentially vanishing with respect to the ambiguity probability term.

Since $\mathbf{e}'$ is not Gaussian, we have to resort to a bounding technique analogous to what done in the proof of Theorem 5. Again, we introduce a $2MT$-dimensional Gaussian random vector $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, 0.5\mathbf{I})$ and define the modified error vector

$$\widetilde{\mathbf{e}} = (\mathbf{B} - \mathbf{F}\mathbf{H})\mathbf{g} + \mathbf{F}\mathbf{w}. \tag{73}$$

[8] It is straightforward to see that the same relation holds if the two ensembles are scaled with the same, but arbitrary, factor.

By proceeding exactly in the same way as in Section D, we conclude that

$$\Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \leq \exp(o(MT))\eta_{\max}^{2MT} \Pr\left(|\widetilde{\mathbf{e}}|^2 \geq MT(1+\gamma)\right). \tag{74}$$

From Lemma 1 we obtain that $\widetilde{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, 0.5\mathbf{I})$, therefore, we can use the usual Chernoff bounding technique and obtain

$$\begin{aligned} \Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) &\leq \exp(o(MT))\eta_{\max}^{2MT}(1+\gamma)^{MT}\exp(-MT\gamma) \\ &= \exp\left(-MT\left(\gamma - \log(1+\gamma) - 2\log\eta_{\max} - \frac{o(MT)}{MT}\right)\right) \end{aligned} \tag{75}$$

By letting $\gamma = \log\rho$, we obtain

$$\Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \quad \dot{\leq} \quad \rho^{-MT} \tag{76}$$

For $T \geq M+N-1$, the exponent of $\Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right)$ with respect to $\log\rho$ is clearly larger than $d^\star(r)$ (whose maximum is MN). Hence, we conclude that the first term in the union bound is exponentially vanishing and can be neglected.

Having analyzed the average error probability (over the ensemble of nested LAST codes) of the ambiguity decoder, we are now ready to conclude the proof of Theorem 6. We upper-bound the average probability of error (averaged over the expurgated ensemble and over the channel) as

$$\overline{P_e}(\rho) \triangleq \mathbb{E}_\Lambda[P_e(\rho)] \leq \Pr(R_{\text{mod}}(\mathbf{H}^c,\rho) \leq R(\rho)) + \mathbb{E}_{\Lambda_c}\left[\Pr(\text{error}, R_{\text{mod}}(\mathbf{H}^c,\rho) > R(\rho)|\Lambda_c)\right] \tag{77}$$

where $R_{\text{mod}}(\mathbf{H}^c,\rho) = \log\det(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c)$. Since the event $\{R_{\text{mod}}(\mathbf{H}^c,\rho) \leq R(\rho)\}$ coincides with the information outage probability with Gaussian i.i.d. inputs, the same analysis of [10] applies here, yielding

$$\Pr(R_{\text{mod}}(\mathbf{H}^c,\rho) \leq R(\rho)) \doteq \rho^{-d^\star(r)} \tag{78}$$

We define again the normalized log-eigenvalues $\alpha_i \triangleq -\log\lambda_i/\log\rho$, where $0 \leq \lambda_1 \leq \cdots \leq \lambda_M$ are the eigenvalues of $(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c$. Following [10], we have

$$\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right) \doteq \exp\left(\log(\rho)\sum_{i=1}^{\min\{M,N\}}[1-\alpha_i]^+\right)$$

Correspondingly, the outage event can be written in terms of $\boldsymbol{\alpha}$ as

$$\mathcal{B} = \left\{\boldsymbol{\alpha} \in \mathbb{R}_+^{\min\{M,N\}} \quad : \quad \sum_i [1-\alpha_i]^+ \leq r, \quad , \alpha_1 \geq \cdots \geq \alpha_{\min\{M,N\}}\right\}$$

We let $\mathcal{B}'$ denote the complement of $\mathcal{B}$, and $P_e(R(\rho)|\alpha, \Lambda_c)$ denote the probability of error of the MMSE-GDFE lattice decoder applied to the nested LAST code formed by Λ_c, Λ_s, of rate $R(\rho) \doteq r \log \rho$, for given channel with normalized log-eigenvalues $\boldsymbol{\alpha}$. As for the second term in (77), because of what said above, we have

$$\begin{aligned} \mathbb{E}_{\Lambda_c}\left[\Pr(\text{error}, R_{\text{mod}}(\mathbf{H}^c, \rho) > R(\rho)|\Lambda_c)\right] &\doteq \int_{\mathcal{B}'} p(\boldsymbol{\alpha}) \mathbb{E}_{\Lambda_c}[P_e(R(\rho)|\alpha, \Lambda_c)] d\boldsymbol{\alpha} \\ &\dot{\leq} \int_{\mathcal{B}'} p(\boldsymbol{\alpha}) \Pr(\mathcal{A}|\boldsymbol{\alpha}) d\boldsymbol{\alpha} \end{aligned}$$

$$\doteq \int_{\mathcal{B}'} \exp\left(-\log(\rho)\left(\sum_{i=1}^{\min\{M,N\}} (2i-1+|M-N|)\alpha_i + T\left(\sum_{i=1}^{\min\{M,N\}} [1-\alpha_i]^+ - r\right)\right)\right) d\boldsymbol{\alpha} \tag{79}$$

$$\doteq \rho^{-d^\star(r)} \tag{80}$$

where we have used the explicit expression (72) for the average (over the lattice ensemble) ambiguity probability conditioned over the channel, i.e., with respect to $\boldsymbol{\alpha}$. The final result (80) follows from noticing that (79) is identical to equation (20) in [10], that is, it is equivalent (in the sense of $\doteq$) to the probability of error of random Gaussian codebooks under ML decoding. □

F Proof of Theorem 7

We reconsider LAST coding with spherical shaping region, as in the proof of Theorem 4, but we shall replace standard lattice decoder by the MMSE-GDFE lattice decoding. Consider the lattice code $\mathcal{C}(\Lambda, \mathbf{u}_0^\star, \mathcal{R})$ where $\mathcal{R} = \mathcal{B}(\sqrt{MT})$ is the $2MT$-dimensional sphere with radius $\sqrt{MT}$, such that the input power constraint is satisfied for all codewords. For each choice of Λ, we use a translation vector $\mathbf{u}_0^\star$ such that (8) is satisfied (we know that such $\mathbf{u}_0^\star$ exists, possibly not uniquely).

At the receiver, we consider the MMSE-GDFE lattice decoding defined by

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z} \in \mathbb{Z}^{2MT}} |\mathbf{F}\mathbf{y} - \mathbf{B}\mathbf{u}_0^\star - \mathbf{B}\mathbf{G}\mathbf{z}|^2 \tag{81}$$

where $\mathbf{F}, \mathbf{B}$ are the MMSE-GDFE matrices defined in Section A, and $\mathbf{G}$ is the generator matrix of Λ. If $\mathbf{G}\hat{\mathbf{z}} + \mathbf{u}_0^\star$ is not in $\mathcal{R}$, an error is declared.

As argued in previous proofs, the error probability of the above decoder is upper-bounded by the error probability of the ambiguity decoder for the lattice translate $\Lambda + \mathbf{u}_0^\star$ with decision region $\mathcal{E}_{T,\gamma}$ defined in (50) applied to the modified channel output $\mathbf{y}'' = \mathbf{B}^{-1}\mathbf{y}'$, where we define

$$\mathbf{y}' = \mathbf{B}\mathbf{c} - [\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{c} + \mathbf{F}\mathbf{w} = \mathbf{B}\mathbf{c} + \mathbf{e}' \tag{82}$$

where $\mathbf{c} \in \mathcal{C}(\Lambda, \mathbf{u}_0^\star, \mathcal{R})$.

Assuming that Λ belongs to an ensemble satisfying Theorem 1 we can upper-bound the average probability of error, where now average is both with respect to the lattice ensemble and with respect to the codewords of the lattice code, for fixed channel matrix, as

$$\begin{aligned} \overline{P_e}(\mathcal{E}_{T,\gamma}) &\triangleq \mathbb{E}_\Lambda \left[\frac{1}{|\mathcal{C}|} \sum_{\mathbf{c}\in\mathcal{C}} \Pr\left(\text{error}|\Lambda, \mathbf{c}, \mathcal{E}_{T,\gamma}\right)\right] \\ &\leq \Pr(|\mathbf{e}'|^2 \geq MT(1+\gamma)) + (1+\delta)(1+\gamma)^{MT}\text{det}\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{-T} \rho^{rT} \end{aligned} \tag{83}$$

where $\mathbf{e}'$ is distributed as $-[\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{c} + \mathbf{F}\mathbf{w}$ with $\mathbf{c}$ $\sim$Uniform over the codebook $\mathcal{C}$. It is clear from the proof of Theorems 4 and 6 that Theorem 7 holds if we can show that the first term in (83) satisfies $\Pr(|\mathbf{e}'|^2 \geq MT(1+\gamma)) \ \dot{\leq} \ \rho^{-d^\star(r)}$.

We define the modified error signal

$$\begin{aligned} \widetilde{\mathbf{e}} &= \mathbf{e}' - [\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{u} + \mathbf{F}\mathbf{w}_2 \\ &= -[\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{x} + \mathbf{F}(\mathbf{w} + \mathbf{w}_2) \end{aligned} \tag{84}$$

where $\mathbf{u}$ is uniformly distributed on the packing sphere $\mathcal{B}(r_{\text{pack}})$, where r_{pack} denotes the packing radius of Λ, and $\mathbf{w}_2$ is white Gaussian with a variance that will be specified later. By construction, $\mathbf{x}$ is uniformly distributed over the region

$$\mathcal{R}' = \bigcup_{\mathbf{c}\in\mathcal{C}} \{\mathbf{c} + \mathcal{B}(r_{\text{pack}})\} \tag{85}$$

of volume $V(\mathcal{R}') = |\mathcal{C}|V(\mathcal{B}(r_{\text{pack}}))$. This region is certainly contained in the sphere $\mathcal{B}(\sqrt{MT} + r_{\text{pack}})$. Hence, we have that

$$f_{\mathbf{x}}(\mathbf{z}) \leq \frac{V(\mathcal{B}(\sqrt{MT} + r_{\text{pack}}))}{|\mathcal{C}|V(\mathcal{B}(r_{\text{pack}}))} f_{\mathbf{v}}(\mathbf{z}), \quad \forall \ \mathbf{z} \in \mathbb{R}^{2MT}$$

where $f_{\mathbf{x}}(\mathbf{z})$ denotes the pdf of $\mathbf{x}$ and $f_{\mathbf{v}}(\mathbf{z})$ denotes the pdf of a random vector $\mathbf{v}$, uniformly distributed over the sphere $\mathcal{B}(\sqrt{MT} + r_{\text{pack}})$. Notice also that, for $|\mathcal{C}| = \exp(TR) = \rho^{rT}$, we have

$$\frac{V(\mathcal{B}(\sqrt{MT} + r_{\text{pack}}))}{|\mathcal{C}|V(\mathcal{B}(r_{\text{pack}}))} = \left(1 + \frac{\sqrt{MT}}{r_{\text{pack}}}\right)^{2MT} \rho^{-rT}$$

We can replicate the proof technique used in Sections D) and E), that consists of replacing $\mathbf{x}$ with $\mathbf{v}$ and successively replacing $\mathbf{v}$ by $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, where

$$\sigma^2 = \frac{1}{2}\left(1 + \frac{r_{\text{pack}}}{\sqrt{MT}}\right)^2 \geq \frac{1}{2} \tag{86}$$

We also choose the per-component variance of $\mathbf{w}_2$ as $\sigma^2 - 1/2$, that is non-negative.

By Lemma 1 in Section A we have that

$$(\mathbf{B} - \mathbf{F}\mathbf{H})\mathbf{g} + \mathbf{F}(\mathbf{w} + \mathbf{w}_2) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

By applying the Chernoff bound on the Gaussian tail probability $\Pr(|(\mathbf{B} - \mathbf{F}\mathbf{H})\mathbf{g} + \mathbf{F}(\mathbf{w} + \mathbf{w}_2)|^2 \geq MT(1+\gamma))$ we obtain the upper-bound

$$\begin{aligned} \Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \quad &\leq \quad \exp(o(MT)) \left(1 + \frac{\sqrt{MT}}{r_{\text{pack}}}\right)^{2MT} \rho^{-rT} \cdot \\ &\quad \exp\left(-MT\left(\frac{1+\gamma}{2\sigma^2} - 1 - \log\frac{1+\gamma}{2\sigma^2}\right)\right) \end{aligned} \tag{87}$$

Now, for each SNR ρ we select Λ in the Loeliger ensemble $\mathcal{L}_{p,2MT}$ with fundamental volume $V_f(\rho)$ and arbitrarily large p. It turns out that since the shaping region $\mathcal{R} = \mathcal{B}(\sqrt{MT})$ does not depend on ρ, in order to achieve coding rate $R(\rho) = r\log\rho$ the fundamental volume $V_f(\rho)$ must vanish as ρ^{-rT}, i.e., $V_f(\rho) \doteq \rho^{-rT}$. Moreover, since these lattices satisfy the Minkowski-Hlawka theorem (Theorem 1), it follows from Theorem 3 of [27] and Theorem 1 of [28] that for almost all lattices in the ensemble, the packing efficiency

$$\eta_{\text{pack}} = \left(\frac{V(\mathcal{B}(r_{\text{pack}}))}{V_f(\rho)}\right)^{1/2MT} = \frac{r_{\text{pack}}}{r_{\text{eff}}}$$

is lower-bounded by

$$\eta_{\text{pack}} \dot{\geq} \frac{1}{2} \tag{88}$$

for sufficiently large p. From a union bound argument similar to what was used in [28] to prove *the simultaneous goodness* of lattices, we can see that there exist lattices $\Lambda^\star$ in the ensembles, for increasing ρ, such that their ambiguity probability is upper-bounded by the second term in (83) and their packing efficiency is lower bounded by (88). For such lattices we have

$$r_{\text{pack}} \dot{\geq} \frac{1}{2}\left(\frac{V_f}{V(\mathcal{B}(1))}\right)^{1/2MT} \doteq \rho^{-r/2M}$$

We conclude that $\sigma^2 \to 1/2$ and

$$\left(1 + \frac{\sqrt{MT}}{r_{\text{pack}}}\right)^{2MT} \rho^{-rT} = O(1)$$

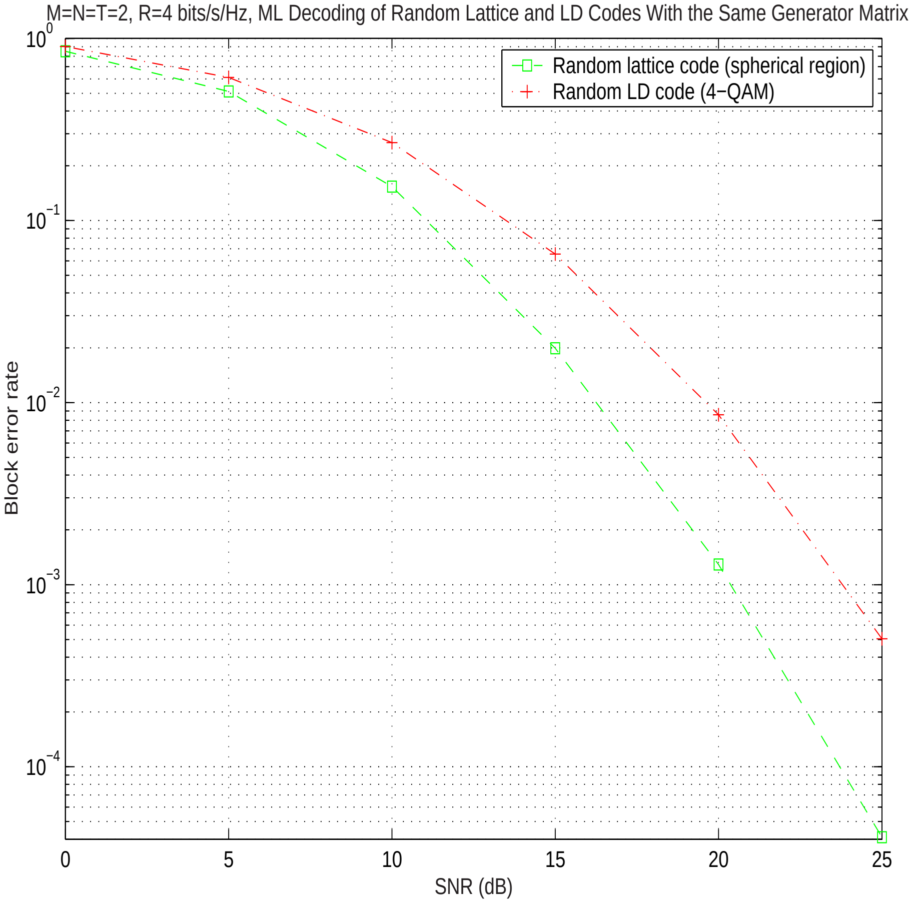

Figure 1: Random LAST code versus random LD code.

as $\rho \to \infty$.

Using these asymptotics in (87) and letting $\gamma = \log \rho$ we obtain that

$$\Pr\left(|\mathbf{e}'|^2 \geq MT(1+\gamma)\right) \;\dot{\leq}\; \rho^{-MT} \;\dot{\leq}\; \rho^{-d^\star(r)}$$

where the last inequality holds since $T \geq M + N - 1$. This concludes the proof. □

References

[1] E. Teletar. Capacity of multi-antenna Gaussian channels. *Technical Report, AT&T-Bell Labs*, June 1995.

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

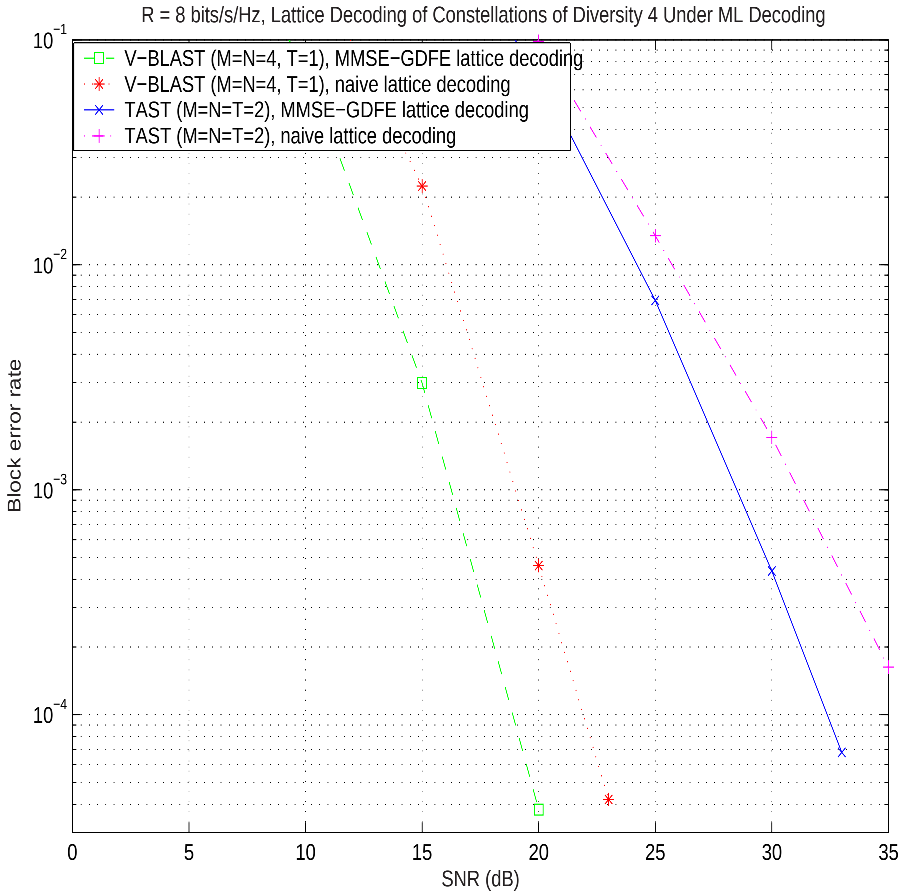

Figure 2: MMSE-GDFE lattice decoding versus naive lattice decoding.

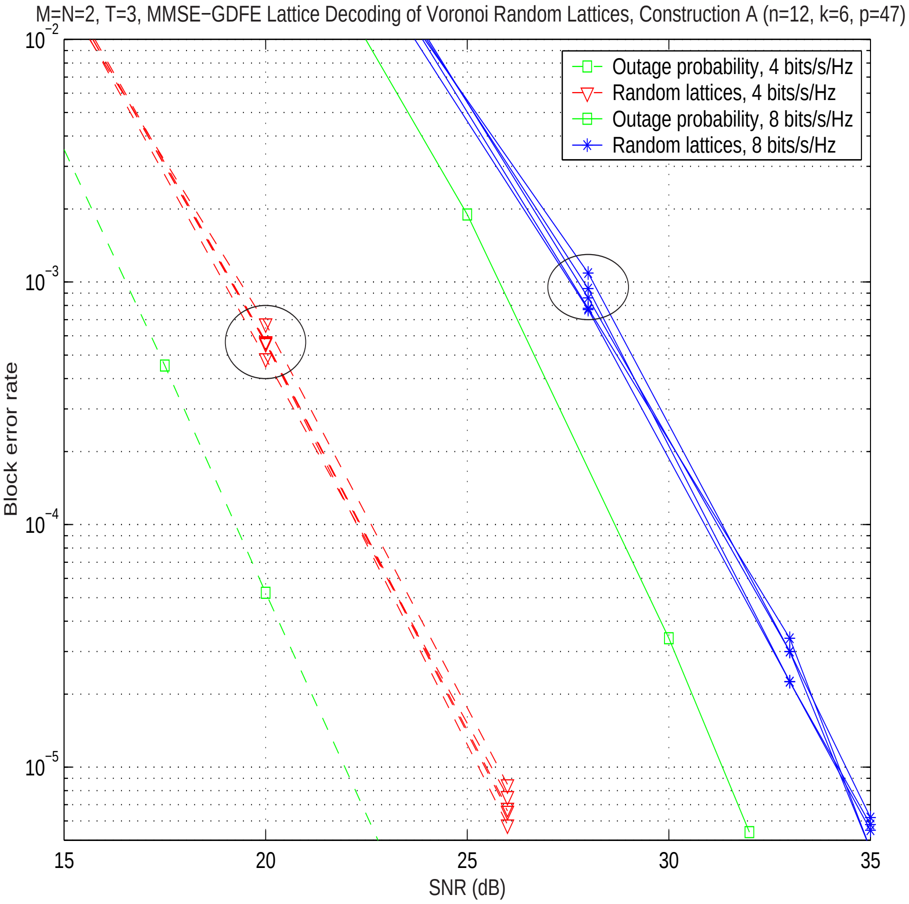

Figure 3: Random nested LAST codes with MMSE-GDFE lattice decoding achieve the optimal diversity-vs-multiplexing tradeoff.

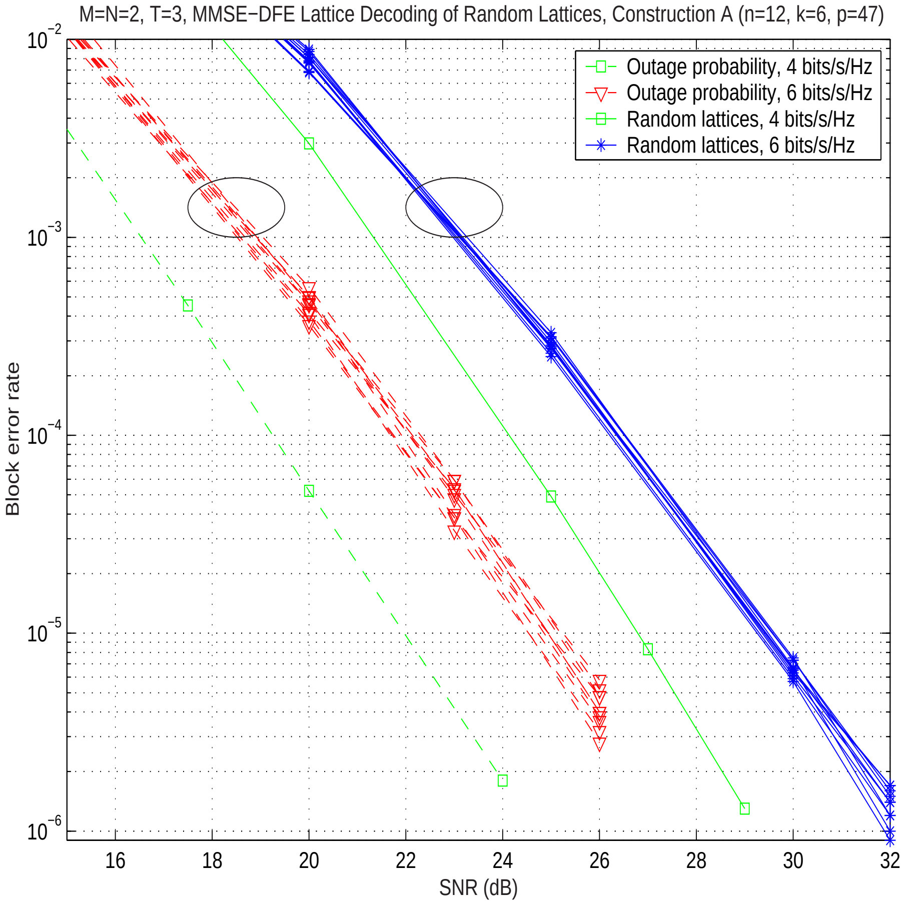

Figure 4: Random spherical LAST codes with MMSE-GDFE lattice decoding achieve the optimal diversity-vs-multiplexing tradeoff.

[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. Inform. Theory*, IT-44:744–765, March 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. *IEEE Vehicular Technology Conference*, pages 136–140, Atlanta, 1996.

[5] B. M. Hochwald, G. Caire, B. Hassibi, and T. L. Marzetta (*ed.*). Special Issue on Space-Time Transmission, Reception, Coding, and Signal Processing. *IEEE Trans. Inform. Theory*, Oct. 2003.

[6] S. Verdu and T. S. Han. A general Formula for Channel Capacity. *IEEE Trans. Inform. Theory*, IT-40:1147–1157, July 1994.

[7] M. O. Damen, H. El Gamal, and N. Beaulieu. Linear threaded algebraic space-time constellations. *IEEE Trans. Inform. Theory*, Oct. 2003

[8] H. El Gamal and M. O. Damen. Universal space-time coding. *IEEE Trans. Info. Theory*, 49:1097-1119, May 2003.

[9] M. O. Damen, H. El Gamal, and G. Caire. On maximum likelihood decoding and the search of the closest lattice point. *IEEE Trans. Info. Theory*, Oct. 2003.

[10] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. Info. Theory*, 49:1073–1096, May 2003.

[11] S. M. Alamouti. A simple transmitter diversity scheme for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16:1451 –1458, October 1998.

[12] G. J. Foschini. Layered space-time architecture for wireless communication in fading environments when using multiple antennas. *Bell Labs Tech. J.*, 2, Autumn 1996.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Info. Theory*, IT-45:1456–1467, July 1999.

[14] R. W. Heath, Jr and A. J. Paulraj. Switching between multiplexing and diversity based on constellation distance. *38th Annual Allerton Conf. on Comm. Control, and Comput.*, Monticello, IL, Sept. 30- Oct. 2, 2000.

[15] H. Yao and G. W. Wornell. Achieving the full MIMO diversity-vs-multiplexing frontier with rotation-based space-time codes. *41th Annual Allerton Conf. on Comm. Control, and Comput.,* Monticello, IL, Oct. 2-4, 2003.

[16] B. Hassibi and B. Hochwald. High rate codes that are linear in space and time. *IEEE Trans. Inform. Theory*, 48:1804–824, July 2002.

[17] N. Prasad and M. K. Varanasi, ”D-BLAST lattice codes for MIMO block Rayleigh fading channels”, *40th Annual Allerton Conf. on Comm. Control, and Comput.*, Monticello, IL, Oct., 2002.

[18] M. O. Damen, A. Chkeif, and J.-C. Belfiore, “Lattice codes decoder for space-time codes,” *IEEE Commun. Lett.*, Vol. 4, pp. 161–163, May 2000.

[19] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 261:515–534, 1982.

[20] R. De Buda. Some optimal codes have structure. *IEEE J. Select. Areas Commun.*, Vol-6:893–899, Aug. 1989.

[21] R. Urbanke and B. Rimoldi. Lattice codes can achieve capacity on the AWGN channel. *IEEE Trans. Inform. Theory*, IT-44:273–278, Jan. 1998.

[22] U. Erez and R. Zamir. Lattice decoding can achieve $\frac{1}{2}\log(1 + SNR)$ on the AWGN channel using nested codes. *submitted to IEEE Trans. Inform. Theory*, 2001.

[23] G. D. Forney, Jr. On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener. *41th Annual Allerton Conf. on Comm. Control, and Comput.*, Monticello, IL, Oct. 2-4, 2003.

[24] 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 *et al.*, ed.), 79:127. Boston: Kluwer, 1997.

[25] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1992.

[26] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest point search in lattices. *IEEE Trans. Inform. Theory,* 48:2201–2214, Aug. 2002.

[27] H-A. Loeliger. Averaging bounds for lattices and linear codes. *IEEE Trans. Inform. Theory*, 43:1767–1773, Nov. 1997.

[28] U. Erez, R. Zamir, and S. Litsyn. Lattices which are good for (almost) everything. *Proc. IT Workshop*, pp. 271-274, Paris, France, April-May 2003 (the journal version submitted to IEEE Trans. Info. Theory).

[29] C. A. Rogers. *Packing and Covering*. Cambridge University Press, Cambridge 1964.

[30] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Inform. Theory*, 42:1152-1159, July 1996.

[31] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, UK, 1985.

[32] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed., Springer, New York, 1998.

[33] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., Oxford Sciences Pub., Oxford, UK, 1979.