On the Role of MMSE in Lattice Decoding: Achieving the Optimal Diversity-vs-Multiplexing Tradeoff *

Hesham El Gamal
Ohio-State University

Giuseppe Caire
Eurecom Institut

Mohamed Oussama Damen
University of Alberta

Abstract

In this paper, we introduce the class of lattice space-time codes as a generalization of linear dispersion (LD) coding. We characterize the diversity-vs-multiplexing tradeoff achieved by random lattice coding with lattice decoding. This characterization establishes the optimality of lattice vertical codes when coupled with lattice decoding. We then generalize Erez and Zamir mod-Λ construction to multiple-input multiple-output (MIMO) channels and show that this construction achieves the optimal diversity-vs-multiplexing tradeoff under minimum mean square error (MMSE) lattice decoding. This result settles the open problem posed by Zheng and Tse on the construction of *explicit* coding and decoding schemes that achieve the optimal tradeoff.

1 Introduction

Hassibi and Hochwald coined the name linear dispersion (LD) coding to denote a class of codes that are linear over the field of complex numbers. In the LD coding framework [1], inputs drawn from a quadrature amplitude modulation (QAM) constellation are spread across time and antennas through properly constructed spreading matrices. In this work, we introduce the class of *lattice space-time codes* as a generalization of LD coding. The idea is to carve the space-time code directly from a properly constructed lattice. As shown in the sequel and [2], this generalization allows for constructing codes and decoding algorithms that approach the fundamental limits of multiple-input multiple-output (MIMO) channels.

One important feature of lattice codes is that they can be decoded by a class of efficient decoders known as *lattice decoders*. Lattice decoding disregards the boundaries of the lattice code and finds 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 the complexity, compared to maximum likelihood (ML) decoding, since 1) It avoids the need for complicated *boundary control* for certain class of lattice codes [3] and 2) It allows for using efficient preprocessing algorithms (e.g., the LLL algorithm [4]) which are known to offer significant reductions in the complexity of algorithms that search for the closest lattice point.

Recently, Zheng and Tse have established the fundamental tradeoff between multiplexing and diversity in the high signal-to-noise ratio (SNR) regime of MIMO channels [5]. In this characterization, the authors used Gaussian codebooks with ML decoding and posed the problem of explicit construction of coding/decoding schemes that realize the optimal tradeoff curve as an open problem. In this paper we investigate the diversity-vs-multiplexing tradeoff achievable by lattice coding under lattice decoding. We find an achievable diversity-vs-multiplexing tradeoff curve which coincides with the optimal tradeoff for "space-only" codes (i.e., when the block length is 1 symbol interval). We further generalize Erez and Zamir mod-Λ construction to the case of multiple-input multiple-output (MIMO) channels [6] and show that it achieves the optimal diversity-vs-multiplexing tradeoff under minimum mean square error (MMSE) lattice decoding[1]. One of the important goals of this paper is to establish the central role

*This work was funded in part by the National Science Foundation under Grant ITR 0219892.

[1] MMSE lattice decoding will be rigorously defined in the sequel.

of MMSE generalized decision feedback equalization (MMSE-GDFE) in approaching the fundamental limits of MIMO channels in the high SNR regime.

The rest of this paper is organized as follows. In Section 2, we introduce the class of lattice codes. An achievable diversity-vs-multiplexing tradeoff curve for random lattice codes with lattice decoding is obtained in Section 3. The generalization of Erez and Zamir mod-Λ construction to the MIMO case is presented in Section 4 where we establish the optimality of this scheme with respect to the tradeoff criterion. Selected numerical results that reveal some interesting insights on the performance of lattice space-time codes are presented in Section 5. Section 6 summarizes the main contributions of this work and discusses some important implications of our results.

2 Lattice space-time codes

We consider the standard quasi-static flat-fading M-transmit N-receive MIMO channel [7]. At time instant t, M modulated symbols, $\mathbf{s}_t^c \triangleq (s_{1t}^c, \dots, s_{Mt}^c)^T$ are transmitted in parallel from the M transmit antennas. [2] The corresponding received complex baseband signal is given by

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

where $\mathbf{H}^c$ is the $N \times M$ channel matrix with (i,j)-th element h_{ij}^c, representing the fading coefficient between the j-th transmit and the i-th receive antenna. These fading coefficients are assumed to be independent, identically distributed (i.i.d.) $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, constant for $t = 1, \dots, T$, where T denotes the duration of a space-time code word (block length). The noise vector $\mathbf{w}_t^c$ is temporally and spatially white with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. Assuming that the input satisfies $\mathbb{E}[|\mathbf{s}_t^c|^2] \leq M$, the parameter ρ takes on the meaning of SNR per receiver antenna. We also assume that $\mathbf{H}^c$ is perfectly known at the receiver and completely unknown at the transmitter. To simplify the presentation, we assume that $M = N$ and use the real matrix representation of the received signal

$$\mathbf{Y} = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{S} + \mathbf{W}, \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^{2M\times T}$, $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_T] \in \mathbb{R}^{2M\times T}$, $\mathbf{s}_t^{\mathsf{T}} = \left[\mathrm{Re}\{\mathbf{s}_t^c\}^{\mathsf{T}}, \mathrm{Im}\{\mathbf{s}_t^c\}^{\mathsf{T}}\right]^{\mathsf{T}}$, $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_T] \in \mathbb{R}^{2M\times T}$, $\mathbf{w}_t^{\mathsf{T}} = \left[\mathrm{Re}\{\mathbf{w}_t^c\}^{\mathsf{T}}, \mathrm{Im}\{\mathbf{w}_t^c\}^{\mathsf{T}}\right]^{\mathsf{T}}$, and

$$\mathbf{H} = \begin{bmatrix} \mathrm{Re}\{\mathbf{H}^c\} & -\mathrm{Im}\{\mathbf{H}^c\} \\ \mathrm{Im}\{\mathbf{H}^c\} & \mathrm{Re}\{\mathbf{H}^c\} \end{bmatrix} \tag{3}$$

is the $2M \times 2M$ real channel matrix. The design of space-time signals, therefore, reduces to the construction of a codebook $\mathcal{S} = \{\mathbf{S}\}$ of matrices in $\mathbb{R}^{2M\times T}$ that enjoy certain desirable properties.

We will recall here some notation from lattice theory (e.g. [8]). An m-dimensional real lattice Λ is a discrete additive sub-group 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. 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 $\mathbb{R}^m$. 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 $V_f(\Lambda) = \int_{\mathcal{V}} d\mathbf{x} = \sqrt{\det(\mathbf{G}^{\mathsf{T}}\mathbf{G})}$. For any Λ and $\mathcal{R}$, there exists $\mathbf{u}_0^\star$ such that

$$|\mathcal{C}(\Lambda, \mathbf{u}_0^\star, \mathcal{R})| \geq \frac{V(\mathcal{R})}{V_f(\Lambda)} \tag{4}$$

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

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

[2] The superscript ^{c} denotes complex quantities.

A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is *good* (or *sphere-bound achieving* [9]) if

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

For a sequence of good lattices with second-order moment σ^2, a random vector uniformly distributed over the Voronoi region $\mathcal{V}(\Lambda_m)$ converges in distribution (in the sense of divergence) to a Gaussian i.i.d. random vector with per-component variance equal to σ^2 [10].

The class of *full-rate* lattice space-time codes (STCs) is defined as follows:

Definition 1 *A full-rate lattice ST code is defined by a codebook $\mathcal{S}$ of matrices $\mathbf{S} \in \mathbb{R}^{2M\times T}$ such that $\mathcal{C} \triangleq \{vec(\mathbf{S}), \mathbf{S} \in \mathcal{S}\}$ is a lattice code of (real) dimension $m = 2MT$. The rate of $\mathcal{S}$ is given by $R = \frac{1}{T}\log|\mathcal{S}|$ bits PCU.* $\diamond$

We used the term *full-rate* in the previous 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 detailed in [2], this class of STCs represents a non-trivial generalization of the LD coding framework. This generalization is instrumental in approaching the fundamental limits of MIMO channels as argued here and in [2].

3 Achievable performance with 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 Λ that is closest (according to a suitable decoding metric) to the received point, irrespectively of whether this point is in $\mathcal{R}$ or not. We argue that the sub-optimality of lattice decoding, as compared to ML, may entail a significant loss in the achievable diversity-vs-multiplexing tradeoff in certain cases. Since the channel is AWGN, we shall consider the minimum Euclidean distance lattice decoder defined by

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}\in\mathbb{Z}^m} |\mathbf{y} - \mathbf{A}\mathbf{u}|^2\,, \tag{5}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$ and

$$\mathbf{A} = \sqrt{\frac{\rho}{M}}\begin{pmatrix} \mathbf{H} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H} \end{pmatrix}\mathbf{G} = \mathbf{H}_{eq}\mathbf{G}, \tag{6}$$

where $\mathbf{G}$ is the generator matrix of Λ.

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

Proposition 1 *Suppose that $\mathbf{H}^c$ is invertible, then the rate*

$$R_{\text{ld}}(\mathbf{H}^c, \rho) \triangleq M\log\rho + \log\,\det\left(\frac{1}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right) \tag{7}$$

is achievable by lattice space-time coding and minimum Euclidean distance lattice decoding.

Proof. We consider the ensemble of $2MT$-dimensional random lattices $\{\Lambda\}$, of the same fundamental volume V_f, generated by a distribution that satisfies the Minkowski-Hlawka Theorem [8]. 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 $\mathbf{s} \in \mathcal{C}(\Lambda, \mathbf{u}_0, \mathcal{R})$ the input constraint $|\mathbf{s}|^2 \leq MT$ is satisfied.

Let $\mathbf{H}_{eq} \in \mathbb{R}^{2MT \times 2MT}$ be the equivalent channel matrix related to $\mathbf{H}^c$ via (6) and (3). Since $\mathbf{H}^c$ is invertible, without loss of generality we consider the following decoder. In the first step, we apply the linear zero-forcing (ZF) equalizer in order to obtain

$$\mathbf{r} = \mathbf{H}_{eq}^{-1}\mathbf{y} = \mathbf{s} + \mathbf{e}, \tag{8}$$

where $\mathbf{s} \in \Lambda$ is the transmitted point and $\mathbf{e} = \mathbf{H}_{eq}^{-1}\mathbf{w}$, with $\mathbf{w} = \text{vec}(\mathbf{W})$, is a noise vector $\sim \mathcal{N}(\mathbf{0}, \frac{1}{2}(\mathbf{H}_{eq}^{\mathsf{T}}\mathbf{H}_{eq})^{-1})$. In the second step, we apply the *ambiguity lattice decoder* of [8]. 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 $\widehat{\mathbf{x}} \neq \mathbf{s}$ or $\mathcal{A}$ occurs, we have error.

For given Λ and $\mathcal{E}$ we have

$$P_e(\Lambda, \mathcal{E}) \leq \Pr(\mathbf{e} \notin \mathcal{E}) + \Pr(\mathcal{A}) \tag{9}$$

By taking the expectation over the ensemble of random lattices, from Theorem 4 in [8] we obtain

$$\overline{P_e}(\mathcal{E}) \triangleq \mathbb{E}_\Lambda[P_e(\Lambda, \mathcal{E})] \leq \Pr(\mathbf{e} \notin \mathcal{E}) + (1+\delta)\frac{V(\mathcal{E})}{V_f} \tag{10}$$

for arbitrary $\delta > 0$.

We choose as decision region the ellipsoid defined by

$$\mathcal{E}_{T,\gamma} \triangleq \left\{\mathbf{z} \in \mathbb{R}^{2MT} \ : \ \mathbf{z}^{\mathsf{T}}\mathbf{H}_{eq}^{\mathsf{T}}\mathbf{H}_{eq}\mathbf{z} \leq MT(1+\gamma)\right\} \tag{11}$$

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

$$\Pr(\mathbf{e} \notin \mathcal{E}_{T,\gamma}) < \epsilon/2 \tag{12}$$

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

$$P_e(\Lambda^\star, \mathcal{E}_{T,\gamma}) \leq \epsilon/2 + (1+\delta)\frac{V(\mathcal{E}_{T,\gamma})}{V_f} \tag{13}$$

For this lattice, we choose the translation vector $\mathbf{u}_0^\star$ such that (4) holds. By letting $|\mathcal{C}(\Lambda^\star, \mathbf{u}_0^\star, \mathcal{R})| = 2^{TR}$, we can write

$$P_e(\Lambda^\star, \mathcal{E}_{T,\gamma}) \leq \epsilon/2 + (1+\delta)\frac{V(\mathcal{E}_{T,\gamma})2^{TR}}{V(\mathcal{R})} \tag{14}$$

From standard geometry formulas, we have

$$\begin{aligned}\frac{V(\mathcal{E}_{T,\gamma})}{V(\mathcal{R})} &= (1+\gamma)^{MT}\det\left(\mathbf{H}_{eq}^{\mathsf{T}}\mathbf{H}_{eq}\right)^{-1/2} \\ &= (1+\gamma)^{MT}\left(\frac{M}{\rho}\right)^{MT}\det\left((\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{-T}\end{aligned} \tag{15}$$

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

The second term in the upper bound (14) can be made smaller than $\epsilon/2$ for sufficiently large T if

$$R < \frac{1}{T}\log\frac{V(\mathcal{R})}{V(\mathcal{E}_{T,\gamma})} = M\log\rho + \log\det\left(\frac{1}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right) - \gamma' \tag{16}$$

where $\gamma' \to 0$ as $\gamma \to 0$. This shows the achievability of the rate $R_{\text{ld}}(\mathbf{H}^c, \rho)$ in (7) with the ambiguity decoder. The final step in the proof follows by noting that with this choice of decision region in (11), the probability of error of the ambiguity decoder upperbounds that of the minimum Euclidean distance lattice decoder (5). $\square$

The largest achievable rate with Gaussian i.i.d. inputs is given by $\log\det\left(\mathbf{I}+\frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)$. We observe the loss of in terms of achievable rate of lattice coding and lattice decoding compared to general coding and ML (or any other asymptotically optimal) decoding. This is analogous to the loss of *one* in the signal-to-noise ratio entailed by lattice decoding in the standard AWGN channel (e.g., [8]). As for the case of AWGN, a converse to Proposition 1 is missing.

Next, we consider a random channel matrix $\mathbf{H}^c$ as defined in Section 2 and we derive a lower bound on the diversity-vs-multiplexing tradeoff achieved with lattice space-time coding and lattice decoding. Following [5], consider a family of lattice space-time codes $\mathcal{C}$ for fixed M and T, obtained from lattices of 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 average block error probability, where averaging is with respect to the random channel matrix $\mathbf{H}^c$). Define the multiplexing gain r and the diversity gain d of the family $\{\mathcal{C}_\rho\}$ as

$$r=\lim_{\rho\to\infty}\frac{R(\rho)}{\log\rho},\quad \text{and}\quad d=-\lim_{\rho\to\infty}\frac{\log(P_e(\rho))}{\log\rho} \tag{17}$$

In [5], the optimal tradeoff curve $d^*(r)$, yielding for each r the maximum possible d, was found for unrestricted coding and ML decoding. In particular, for any block length $T\geq 2M-1$ the optimal tradeoff is given by the piecewise linear function joining the points $(k,(M-k)^2)$ for $k=0,\ldots,M$.

For lattice space-time codes under lattice decoding we have the following result:

Proposition 2 *There exists a sequence of lattice codes that achieves a diversity advantage $d(r)=M-r$ for $r\in[0,M]$ and for any block length $T\geq 1$ under Euclidean distance lattice decoding. With $T=1$, this coincides with the optimal diversity gain with no restrictions on coding and decoding.*

Proof. We consider an ensemble of random lattice space-time codes satisfying Proposition 1. We upper bound the average probability of error (average over the channel and over the lattice ensemble) as

$$\overline{P_e}(\rho)\stackrel{\triangle}{=}\mathbb{E}_\Lambda[P_e(\rho)]\leq \Pr(R_{\rm ld}(\mathbf{H}^c,\rho)\leq R(\rho))+\mathbb{E}_\Lambda\left[\Pr(\text{error},R_{\rm ld}(\mathbf{H}^c,\rho)>R(\rho)|\Lambda)\right] \tag{18}$$

In order to compute $\Pr\left(R_{\rm ld}(\mathbf{H}^c,\rho)\leq R(\rho)\right)$, we follow in the footsteps of [5]. 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 $\alpha_i\stackrel{\triangle}{=}-\log\lambda_i/\log\rho$ and where $0\leq\lambda_1\leq\cdots\leq\lambda_M$ are the eigenvalues of $(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c$, we get

$$\Pr\left(R_{\rm ld}(\mathbf{H}^c,\rho)\leq R(\rho)\right)\doteq\rho^{-d_c}, \tag{19}$$

where

$$d_c=\inf_{\boldsymbol{\alpha}\in\mathcal{B}}\sum_{i=1}^M(2i-1)\,\alpha_i \tag{20}$$

where $\doteq$ refers to the exponential equality as defined in [5],[3] and where the region $\mathcal{B}\subseteq\mathbb{R}^M$ is defined by $\alpha_1\geq\cdots\geq\alpha_M\geq 0$ and by $\sum_{i=1}^M\alpha_i\geq M-r$. It is straightforward to see that the minimization in (20) is achieved for $\alpha_1=M-r$ and $\alpha_i=0$ for all $i>1$, yielding $d_c=M-r$.

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

$$\mathbb{E}_\Lambda\left[\Pr\left(\text{error},R_{\rm ld}(\mathbf{H}^c,\rho)>R(\rho)|\Lambda\right)\right]=\int_{\mathcal{B}'}p\left(\boldsymbol{\alpha}\right)\mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha},\Lambda\right)]d\boldsymbol{\alpha} \tag{21}$$

[3] Similarly, we shall use $\dot{\geq}$ and $\dot{\leq}$.

where $p(\boldsymbol{\alpha})$ is the joint probability density function (pdf) of $(\alpha_1, ..., \alpha_M)$. We apply again the ambiguity decoder to the ZF channel output (8) with decision region $\mathcal{E}_{T,\gamma}$ defined in (11). By noticing that $\mathbf{H}_{eq}\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$, we get the Chernoff bound

$$\begin{aligned}\Pr(\mathbf{e} \notin \mathcal{E}_{T,\gamma}) &= \Pr(|\mathbf{H}_{eq}\mathbf{e}|^2 \geq MT(1+\gamma)) \\ &\leq \min_{\lambda \geq 0} \ \exp\left(-MT\left(\lambda(1+\gamma) + \log(1-\lambda)\right)\right) \\ &= (1+\gamma)^{MT} e^{-MT\gamma} \end{aligned} \tag{22}$$

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

$$V_f \geq V(\mathcal{R})\rho^{-rT} \tag{23}$$

where $\mathcal{R}$ is the sphere of radius $\sqrt{MT}$ centered in the origin. From Theorem 4 of [8] and (15) we find, for all arbitrary $\delta > 0$,

$$\begin{aligned}\mathbb{E}_\Lambda[P_e\left(R(\rho)|\boldsymbol{\alpha}, \Lambda\right)] &\leq (1+\gamma)^{MT}e^{-MT\gamma} + (1+\delta)(1+\gamma)^{MT}\left(\frac{\rho}{M}\right)^{-MT}\rho^{rT}\text{det}\left((\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{-T} \\ &= (1+\gamma)^{MT}\left[e^{-MT\gamma} + c_1\rho^{-T\left(M-r-\sum_{i=1}^{M}\alpha_i\right)}\right] \end{aligned} \tag{24}$$

where c_1 does not depend on ρ.

Now, we let $\gamma = \log\rho$ and we use again the argument of Zheng and Tse on the dominance of the term with the highest exponent in the (RHS) of (21). We obtain

$$\mathbb{E}_\Lambda\left[\Pr\left(\text{error}, R_{\text{ld}}(\mathbf{H}^c, \rho) > R(\rho)|\Lambda\right)\right] \ \dot{\leq} \ \rho^{-d_n}, \tag{25}$$

where

$$d_n = \inf_{\boldsymbol{\alpha}\in\mathcal{B}'}\left\{\sum_{i=1}^{M}(2i-1)\alpha_i + T\left(M - r - \sum_{i=1}^{M}\alpha_i\right)\right\} \tag{26}$$

It is easily seen that $d_n = M - r$, attained again at $\alpha_1 = M - r$ and $\alpha_i = 0$ for all $i > 1$. By using (19) and (25) in (18), we obtain that there exists at least a sequence of lattice codes $\{\mathcal{C}_\rho^\star\}$ in the ensemble that achieves diversity gain $d = M - r$ with multiplexing gain r. The other statements of Proposition 2 follow from noticing that the achievable diversity gain is the same for all $T > 1$ and that the optimal diversity advantage for $T = 1$ is $d^\star(r) = M - r$ [5]. □

Proposition 2 establishes the optimality of lattice decoding, in terms of the tradeoff, for vertical (space only) coding. For applications that can tolerate larger block length ($T > 1$), the lower bound in Proposition 2 matches the optimal tradeoff curve only at the points $d = 0, r = M$ and $d = 1, r = M-1$, i.e., for very large multiplexing gain. The difference between the lattice-decoding 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, this bound still highlights the loss in performance entailed by lattice coding under lattice decoding. Importantly, we observe that the scheme considered in Proposition 2 fails to exploit block length to increase the diversity gain. The reason for this failure can be traced back to the loss in the achievable rate of Proposition 1 with respect to the optimal (under unrestricted coding and decoding) achievable rate.

4 The generalized mod-Λ construction and its optimality

In [6], Erez and Zamir showed that lattice "Voronoi" codes achieve the AWGN channel capacity under lattice decoding, if the lattice decoder is modified by including a linear MMSE estimation stage and if a dither random signal (implying common randomness at transmitter and receiver) is used. This random dither renders the MMSE estimation error signal independent of the transmitted code word (see also [12]). 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 general

MIMO channels and show that for fixed $\mathbf{H}^c$ and $T \to \infty$ our scheme achieves rates up to the *optimal* $\log\det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)$, that is strictly larger than $R_{\rm ld}(\mathbf{H}^c, \rho)$ of Proposition 1. Moreover, we show that our mod-Λ scheme achieves all points in the optimal diversity-vs-multiplexing tradeoff curve $d^*(r)$.

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

Definition 2 *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. $\diamond$

We say that a lattice space-time code is nested if the underlying lattice code is nested. If $\mathcal{C}$ is a nested lattice code, minimum Euclidean distance lattice decoding takes on the particularly appealing form

$$\widehat{\mathbf{x}} = [Q_{\Lambda_c}(\mathbf{y})] \mod \Lambda_s \tag{27}$$

where we define the lattice quantization function

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

and the modulo-lattice function

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

In practice, a closest lattice point search algorithm ([3] and references therein) can be used to find first $Q_{\Lambda_c}(\mathbf{y})$ and then obtain the modulo Λ_s "projection". With the above lattice decoder, 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 lattice 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 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 $\mathbf{c} \in \mathcal{C}$, generates a dither signal $\mathbf{u}$ with uniform distribution over $\mathcal{V}_s$ and computes

$$\mathbf{x} = [\mathbf{c} - \mathbf{u}] \mod \Lambda_s \tag{28}$$

The signal $\mathbf{x}$ is then mapped onto the space-time coding matrix $\mathbf{S} \in \mathbb{C}^{M\times T}$ according to Definition 1 and is sent through the channel (2). The vectorized (real) received signal is given by

$$\mathbf{y} = \text{vec}(\mathbf{Y}) = \mathbf{H}_{eq}\mathbf{x} + \mathbf{w} \tag{29}$$

where $\mathbf{H}_{eq}$ is defined in (6).

We replace the *scalar* scaling of [6] by the matrix multiplication by the forward filter matrix $\mathbf{F}$ of the MMSE-GDFE [13]. Moreover, instead of adding the dither signal $\mathbf{u}$ at the receiver [6], we add the dither signal filtered by the upper triangular feedback filter matrix $\mathbf{B}$ of the MMSE-GDFE.

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

$$\begin{aligned}
\mathbf{y}' &= \mathbf{F}\mathbf{y} + \mathbf{B}\mathbf{u} \\
&= \mathbf{F}\left(\mathbf{H}_{eq}(\mathbf{c} - \mathbf{u} + \boldsymbol{\lambda}) + \mathbf{w}\right) + \mathbf{B}\mathbf{u} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) - [\mathbf{B} - \mathbf{F}\mathbf{H}_{eq}]\,(\mathbf{c} - \mathbf{u} + \boldsymbol{\lambda}) + \mathbf{F}\mathbf{w} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) - [\mathbf{B} - \mathbf{F}\mathbf{H}_{eq}]\,\mathbf{x} + \mathbf{F}\mathbf{w} \\
&= \mathbf{B}(\mathbf{c} + \boldsymbol{\lambda}) + \mathbf{e}'
\end{aligned} \tag{30}$$

By construction, $\mathbf{x}$ is uniformly distributed over $\mathcal{V}_s$ and it is independent of $\mathbf{c}$. If we consider a sequence of good shaping lattices for increasing dimension T, the MMSE-GDFE estimation error signal

$$\mathbf{e}' = -[\mathbf{B} - \mathbf{F}\mathbf{H}_{eq}]\,\mathbf{x} + \mathbf{F}\mathbf{w} \tag{31}$$

converges in distribution (in the sense of divergence) to the noise vector $\mathbf{w} \sim \mathcal{N}(\mathbf{0},\frac{1}{2}\mathbf{I})$ (by using the fact that, as $T \to \infty$, $\mathbf{x} \to \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ and the properties of MMSE-GDFE equalization it is straightforward to show that $\mathbb{E}[\mathbf{e}'\mathbf{e}'^{\mathsf{T}}] \to \frac{1}{2}\mathbf{I}$). The second remarkable fact in (30) 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 $\mathbf{c}' = \mathbf{c} + \boldsymbol{\lambda}$ belongs to the same coset of Λ_s in Λ_c, this translation does not involve any loss of information. We can rewrite (30) as

$$\mathbf{y}' = \mathbf{B}\mathbf{c}' + \mathbf{e}' \tag{32}$$

where $\mathbf{c}' \in \Lambda_s + \mathbf{c}$. It follows that in order to recover the information message, the decoder has to identify the coset $\Lambda_s + \mathbf{c}$ that contains $\mathbf{c}'$. Then, lattice decoding in the form defined by (27) 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 and necessary component of the whole construction. Finally, we note that because of the block diagonal structure of $\mathbf{H}_{eq}$, also $\mathbf{B}$ is block diagonal with the $2M \times 2M$ upper triangular block $\mathbf{B}'$ repeated T times. By construction we have

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

The optimality of the above construction in the limit of large T is given by the following result.

Proposition 3 *For a fixed, non-random, channel matrix $\mathbf{H}^c$, the rate*

$$R_{\mathrm{mod}}(\mathbf{H}^c, \rho) \triangleq 2M \log \left\lfloor \det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{\frac{1}{2M}} \right\rfloor \tag{33}$$

is achievable by the mod-Λ lattice space-time coding scheme where $\lfloor . \rfloor$ denotes the rounding to the smallest integer.

Proof. We follow in the footsteps of the proof of Proposition 1 after replacing $\mathbf{H}_{eq}$ with $\mathbf{B}$. We also consider the ensemble of lattices obtained via construction A in [14] for Λ_c. It is argued in [14] that random lattices generated according to this ensemble are good for quantization with probability one as $T \to \infty$. We let the shaping lattice be an integer scaling of the coding lattice, i.e., we let $\Lambda_s = a\Lambda_c$ for $a \in \mathbb{Z}_+$. The coding rate of the nested lattice code $\mathcal{C} = \Lambda_c \cap \mathcal{V}_s$ is given by $(1/(2MT)) \log V_f(\Lambda_s)/V_f(\Lambda_c) = \log a$ bit/dimension. Hence, the coding rate of the corresponding nested lattice space-time code is $2M \log a$ bits PCU.

In the limit for large T, the channel (32) 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 $\mathbf{e}'$ which is white and Gaussian with probability one. It is now straightforward to see that the steps in the proof of Proposition 1 apply to this setup and there exists a sequence of coding lattices Λ_c such that, for sufficiently large T, the probability of error can be made smaller than any desired $\epsilon > 0$ provided that

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

Moreover, this holds for any $\mathbf{H}^c$ (even non-invertible), since $\mathbf{B}$ is always invertible (for any finite SNR ρ). The largest integer a satisfying the above inequality is

$$a = \left\lfloor \det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)^{\frac{1}{2M}} \right\rfloor$$

By using this choice of a in the expression for the rate of the nested lattice space-time code we obtain that (33) is achievable. □

We notice here that if one finds sequences of good nested lattice codes (in the sense of [6]) approaching any arbitrary nesting ratio, then for any $\mathbf{H}^c$ the optimal rate $\log\ \det\left(\mathbf{I} + \frac{\rho}{M}(\mathbf{H}^c)^{\mathsf{H}}\mathbf{H}^c\right)$ is achievable by the mod-Λ scheme. However, using self-similar nested lattices (implying integer nesting ratio) suffices to prove the following result, about the optimality of the mod-Λ scheme with respect to the diversity-vs-multiplexing tradeoff.

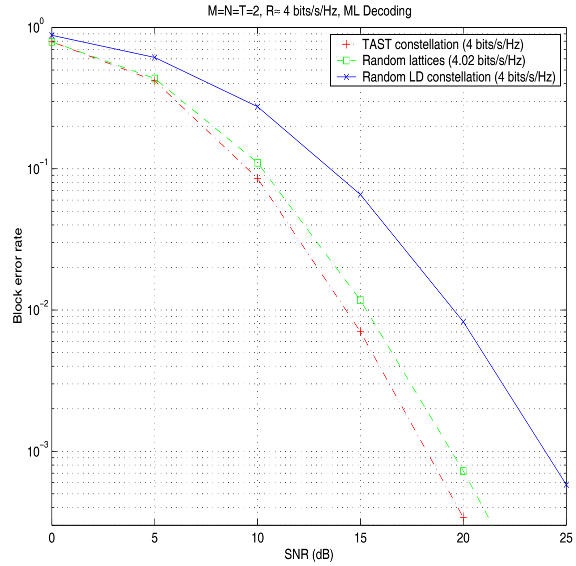

Figure 1: Random lattice versus random LD codes.

Proposition 4 *There exists a sequence of nested lattice space-codes with block length* $T \geq 2M-1$ *that achieves the optimal diversity advantage* $d^*(r)$ *for all* $r \in [0, M]$ *under the mod-*Λ *scheme.*

The proof is omitted here (and reported in [2]) owing to space limitations.

5 Numerical Results

Some representative results are given in this section (more details are given in [2]). Fig. 1 shows the performance of random lattices versus random LD codes for $M = N = T = 2$. The entries of the random lattice generator matrix are chosen to be i.i.d. zero-mean Gaussian random variables with a common unit variance. The lattice code is generated by a sphere of the appropriate radius. The same (normalized) generator matrix is used in the LD code with input QAM alphabet. We also show the performance of the optimized TAST code [11] for comparisons. Observe that a random lattice can approach the performance of an optimized LD construction [11]. The figure illustrates the usefulness of the framework proposed in this work over the LD framework [1].

Fig. 2 shows ZF-DFE (i.e., minimum distance) lattice decoding of vertical (space only) and space-time codes. As predicted by Proposition 2, ZF-DFE achieves full diversity of vertical codes, but fails to achieve full diversity (which is otherwise achieved under ML decoding) of space-time codes. This limitation on ZF-DFE lattice decoding is also true for the algorithm in [15] when used with space-time codes (see [2] for more details).

Fig. 3 illustrates the advantage of using MMSE-GDFE with lattice decoding where a simple scheme of random lattice constellations (with spherical regions) is shown to achieve the optimal diversity-vs-multiplexing tradeoff (Propositions 3, 4). We finally observe that the near-optimal performance of the MMSE-GDFE lattice decoding overcomes the obstacles encountered in digital communications when using the LLL reduction in combination with the algorithms in [2] (i.e., boundaries control), where the latter reduction is known to largely reduce the decoding complexity in quasi-static fading scenarios. Further, the MMSE-GDFE filtering (also) largely improves sub-optimal decoding algorithms using the LLL reduction methods (e.g., [15, 2]).

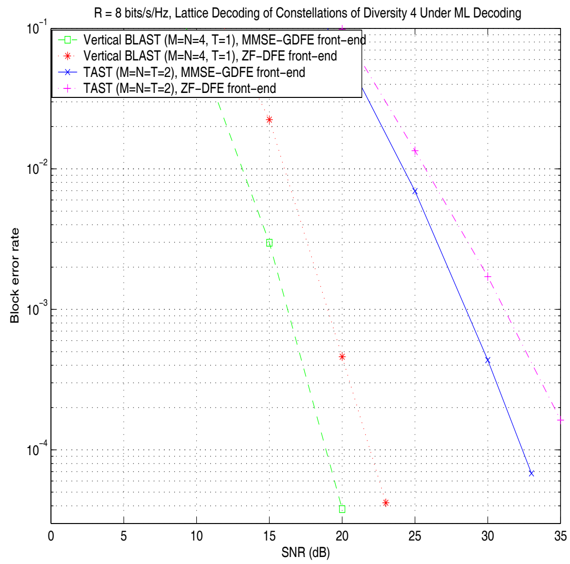

Figure 2: ZF-DFE lattice decoding.

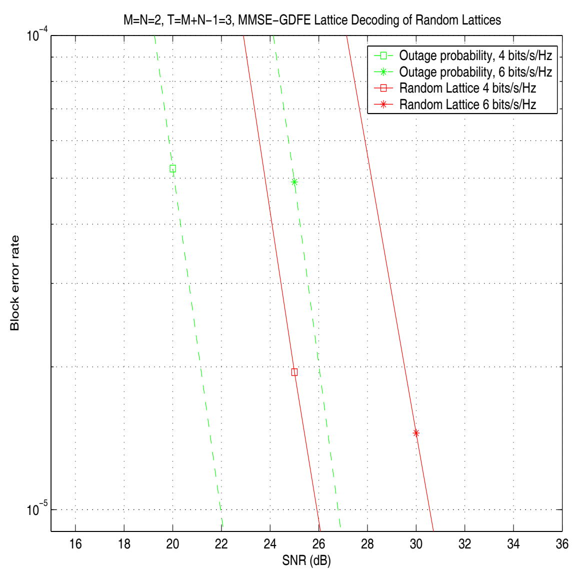

Figure 3: MMSE-GDFE lattice decoding achieve optimal diversity-vs-multiplexing tradeoff

6 Conclusions

In this paper, we introduced the class of lattice space-time coding. We characterized a lower bound on the achievable diversity-vs-multiplexing tradeoff using random lattice codes and lattice decoding. Based on this characterization, we established the optimality of vertical lattice coding with minimum distance lattice decoding in terms of the diversity-vs-multiplexing tradeoff. For space-time codes, we generalized Erez and Zamir mod$-\Lambda$ coding construction to the MIMO fading channel scenario. We further showed that this coding scheme achieves all the points on the optimal tradeoff curve, when coupled with MMSE-GDFE filtering and lattice decoding at the receiver. The theoretical claims were validated through representative simulation results.

The MMSE-GDFE filtering turns out to be an essential component for achieving optimality of the schemes considered in Propositions 3 and 4. The advantage of the MMSE-GDFE over the ZF-DFE is at low and high SNRs (where one may be tempted to expect the ZF and MMSE to have the same performance at high SNRs); an observation also made in [5] with D-BLAST schemes. This phenomenon over MIMO fading channels can be attributed to the "neutralization" of the faded eigenvalues of the MIMO channel done by the MMSE-GDFE filter (i.e., addition of $\mathbf{I}$), which cannot be done by the ZF-DFE filter, even at high SNRs when the faded eigenvalues are $\ll \text{SNR}^{-1}$. Finally, the optimality of the mod-Λ schemes for MIMO fading channels, established here for short codes ($T \geq 2M-1$), makes the proposed class of lattice space-time codes a serious candidate for achieving near-optimal performances (with low complexity) in delay-limited MIMO fading channels.

References

[1] B. Hassibi and B. Hochwald. High rate codes that are linear in space and time. IT, 48:1804–824, July 2002.

[2] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding: Approaching the fundamental limits of MIMO channels," *to be submitted to* IT, Oct. 2003.

[3] M. O. Damen, H. El Gamal, and G. Caire. On maximum likelihood decoding and the search of the closest lattice point. IT, Oct. 2003.

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

[5] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. IT, 49:1073–1096, May 2003.

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

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

[8] H-A. Loeliger. Averaging bounds for lattices and linear codes. IT, 43:1767–1773, Nov. 1997.

[9] G. D. Forney, Jr., and M. D. Trott, and S.Y. Chung, Sphere-bound-achieving coset codes and multilevel coset codes. IT, 46:820-850, May 2000.

[10] R. Zamir and M. Feder. On lattice quantization noise. IT, 42:1152-1159, July 1996.

[11] H. El Gamal and M. O. Damen. Universal space-time coding. IT, 49:1097-1119, May 2003.

[12] G. D. Forney, Jr. On the role of MMSE estimation in approaching the information-theoretic limits of linear gaussian channels: Shannon meets Wiener. *Proc. 41-th Allerton Conf.*, Urbana, Oct. 2-4, 2003.

[13] 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.

[14] 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.

[15] C. Windpassinger and R. Fischer. Low-complexity near maximum likelihood detection and precoding for MIMO systems using lattice reduction. *Proc. IT Workshop*, pp. 345-348, Paris, France, April-May 2003.