DMT Optimality of LR-Aided Linear Decoders for a General Class of Channels, Lattice Designs, and System Models

Joakim Jaldén, *Member, IEEE*, and Petros Elia

***Abstract*—This paper identifies the first general, explicit, and nonrandom MIMO encoder-decoder structures that guarantee optimality with respect to the diversity-multiplexing tradeoff (DMT), without employing a computationally expensive maximum-likelihood (ML) receiver. Specifically, the work establishes the DMT optimality of a class of regularized lattice decoders, and more importantly the DMT optimality of their lattice-reduction (LR)-aided *linear* counterparts. The results hold for all channel statistics, for all channel dimensions, and most interestingly, irrespective of the particular lattice-code applied. As a special case, it is established that the LLL-based LR-aided linear implementation of the MMSE-GDFE lattice decoder facilitates DMT optimal decoding of any lattice code at a worst-case complexity that grows at most linearly in the data rate. This represents a fundamental reduction in the decoding complexity when compared to ML decoding whose complexity is generally exponential in the rate. The results' generality lends them applicable to a plethora of pertinent communication scenarios such as quasi-static MIMO, MIMO-OFDM, ISI, cooperative-relaying, and MIMO-ARQ channels, in all of which the DMT optimality of the LR-aided linear decoder is guaranteed. The adopted approach yields insight, and motivates further study, into joint transceiver designs with an improved SNR gap to ML decoding.**

***Index Terms*—Diversity-multiplexing tradeoff, lattice decoding, lattice reduction, linear decoding, multiple-input multiple-output (MIMO), regularization, space-time coders-decoders.**

I. Introduction

The general multidimensional linear channel model

$$y = Hx + w$$

adequately represents a plethora of communication system models which utilize multidimensional transmit-receive signals for attaining increased rates and reliability in the presence of fading. Such system models include quasi-static multiple-input multiple-output (MIMO), MIMO-OFDM, ISI, amplify-and-forward (AF), decode-and-forward (DF), and MIMO automatic repeat request (ARQ) models. Each of the above models introduces its own structure on $\boldsymbol{H}$ and $\boldsymbol{x}$, its own error performance limits, and its own requirements on coding and decoding schemes. Finding general-purpose transceiver structures with (provably) good performance in these scenarios, and with a reasonable computational complexity, is challenging.

A. Background and Previous Work

Substantial amounts of work have focused on identifying performance criteria and constructing different coding schemes specifically suited to the different system models. For example in the case of the $n_{\rm T} \times n_{\rm R}$ quasi-static MIMO channel, we have seen the orthogonal space-time (ST) designs [1], [2] providing full diversity but doing so only at rates much less than those theoretically possible, codes like V-BLAST [3] providing full rate MIMO benefits but with much reduced diversity, and codes from the general linear dispersion designs [4] providing full rate benefits but no diversity guarantees for increasing spectral efficiencies.

In outage limited communications systems, the fundamental limits with respect to the spectral efficiency and decoding error probability in the high signal-to-noise ratio (SNR) limit were succinctly characterized by Zheng and Tse's diversity multiplexing tradeoff (DMT) [5]. The tradeoff incorporated several previous performance measures and has been extensively adopted ever since as a benchmark for transceiver design and analysis. The work in [5] also introduced the notion of DMT optimal designs, i.e., designs capable of achieving the fundamental DMT of the underlying channel (cf. [5] or Section II-B).

1) Coding: Towards finding DMT optimal codes, the work in [5] proved the existence of such codes for the case of the i.i.d. Rayleigh fading quasi-static MIMO channel[1] by using ensembles of random Gaussian codes over a finite coding duration, and thus reduced system model dimensionality. Although providing codes of finite length, such a construction is highly impractical given the lack of structure that would allow for practical codeword enumeration and decoding. This issue was addressed in [6] which, for the same setting, proved the existence of random ensembles of DMT optimal codes that accept a lattice structure.

Manuscript received May 25, 2009; revised May 21, 2010. Date of current version September 15, 2010. This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (Contract 216715). The work of J. Jaldén was supported by a FWF Grant N10606 (SISE) and by SSF Grant ICA08-0046. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT 2009), Seoul, Korea, June 2009.

J. Jaldén was with the Institute of Communications and Radio-Frequency Engineering, Technical University of Vienna. He is now with the ACCESS Linnaeus Center, KTH Signal Processing Lab, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: jalden@kth.se).

P. Elia is with the Mobile Communications Department, EURECOM, F-06904 Sophia Antipolis cedex, France (e-mail: elia@eurecom.fr).

Communicated by H. Bölcskei, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2059493

[1] By the quasi-static channel we explicitely refer to the flat fading, point-to-point MIMO scenario originally considered in [5] with no additional structural constraints imposed on $\boldsymbol{x}$, cf. Section VI-A.

The same work successfully identified the suitability of the lattice framework for MIMO coding problems, and its effect on issues such as that of finding efficient shaping regions for the transmitted signals. However, DMT optimal random lattice designs inherently rely on different lattices for each rate and SNR and, furthermore, do not provide deterministic means by which to identify the lattice generator matrices.

These two issues were conclusively solved in [7] and [8] which first provided practical construction criteria for DMT optimal codes for the quasi-static Rayleigh fading MIMO channel, and then explicitly constructed the first unified family of DMT optimal codes for all channel dimensions, for the same scenario. These cyclic division algebra (CDA)-based codes, which were built based on the work of [9]–[11], managed to employ for any given number of transmit antennas n_{T} *a single* lattice generator matrix which is explicitly identified. Furthermore, these codes guarantee DMT optimality for all fading statistics, due to the fact that they satisfy the *approximate universality* criterion of [12]. Other CDA codes [13], and later constructed variants of CDA-based codes [14]–[16], currently perform best among all existing ST codes over the quasi-static MIMO channel. Specifically, the perfect ST codes proposed in [13], and later extended in [14], allow for approximate universality as well as *information losslessness* (cf. [17]) for rotationally invariant ST channels. Later work in [15] employed the perfect ST code architecture, together with the lattice space-time (LAST) code framework in [6], to provide for an improved shaping region and better performance at lower values of SNR. Furthermore the work in [16] drew ST codes from subsets of CDAs that constitute *maximal orders*, which interestingly ensure a better fundamental volume of the corresponding lattice, and better energy efficiency [16]. The above DMT optimal codes, originally constructed for the quasi-static MIMO channel, also form the basis for several modified schemes that DMT optimally apply to other system models, see, e.g., [18]–[22]. More detailed expositions of a few alternative scenarios are given in Section VI, although we also note that for many alternative system models the problem of designing DMT optimal codes is still open.

The codes discussed above have to date only been shown to provide DMT optimality in the presence of an ML decoder[2], and hence decoding complexity has remained the fundamental limitation in obtaining (provably) good decoding error probability performance in a computationally efficient manner. This limitation, roughly speaking, originates from the fact that such codes must in general be drawn, due to enumerability and rate requirements, from lattices whose dimension "matches" the inherently high dimension of $\boldsymbol{H}$. Furthermore, in all but rare cases, the diversity requirements force code-channel lattices that cannot be decomposed into substantially "smaller" and simpler component lattices, without severely sacrificing rate gains. The high dimensionality, in conjunction with the high spectral efficiency that is envisioned in future telecommunications, introduce prohibitive ML decoding complexity.

2) Decoding: The current *de facto* decoding method for lattice designs is the sphere decoder (SD) [23]–[25], which is capable of obtaining the exact ML solution more efficiently than a full search of the codebook. Efficient hardware implementations of the sphere decoder have been presented for moderately sized problems, see, e.g., [26]. The complexity of obtaining the ML solution by sphere decoding is however known to scale poorly (i.e., exponentially) with the problem size [27], [28]. This said, there are also methods that limit the SD complexity at the cost of an increased probability of error, e.g., methods based on channel matrix regularization, lattice reduction, and early termination [24], [25]. We are however unaware of any previous theoretical error probability performance guarantee for a non-ML sphere decoder implementation[3].

Substantial interest has been drawn by linear receivers based on the zero-forcing (ZF) or the minimum mean square error (MMSE) criteria, as these receivers avoid exact ML solutions and allow for simple implementation (cf. [29] and references therein). An inherent limitation of ZF-based linear receivers is that ill-conditioned channel matrices lead to substantial noise amplification. This motivated the introduction of MMSE-based linear receivers which can be seen as ZF receivers that take into consideration the presence of additive noise and hence utilize a better-conditioned equivalent channel matrix. It is the case though that for ill-conditioned channel matrices, both these linear receivers, as well as receivers based on successive interference cancellation (SIC), are for the most part substantially suboptimal, as recent DMT analysis in [30] reveals.

Notable steps towards better performing efficient receivers included the introduction of lattice-reduction (LR) techniques in [31] and [32]. Motivated by the fact that ZF is optimal in the presence of orthogonal channels, the work in [31], [32] proposed the use of LR methods for better, nearly orthogonal conditioning of the equivalent channel matrix, prior to simple ZF or SIC decoding. This approach was partly validated by simulations (cf. [25]) and by analysis as in [33] which showed that LR-aided ZF decoding can achieve maximal receive diversity for fixed-rate uncoded V-BLAST. LR-aided ZF decoding is, however, not DMT optimal in general [6], [34]. The work in [24], [25], and [35] proposed lattice decoding with MMSE-GDFE preprocessing which is well suited for the case of underdetermined or singular channels. Contemporary work on LR-aided decoding in an MMSE preprocessed basis appeared in [36]. Simulation results indicated that such methods are capable of near-ML performance at a computational complexity that remains low [15], [24], [25], [36].

3) Tranceivers With Reduced Implementation Complexity: Several works focused on providing codes with reduced ML decoding complexity, mainly for the quasi-static MIMO channel. Such work includes the multigroup decodable codes based on Clifford algebras in [37], and the codes in [38] for asymmetric ($n_{\mathrm{R}} < n_{\mathrm{T}}$) quasi-static MIMO channels. Similarly motivated work in [39] identified existing 2×2 full-rate full-diversity codes for the 2×2 quasi-static MIMO channel [40]–[42], as fast decodable codes since they incur reduced sphere decoding complexity by essentially reducing the dimensionality of the

[2] A notable exception are the random LAST codes in [6], as discussed in Section I-A-III and throughout the present paper.

[3] We hasten to add that it follows by the results presented herein that the Schnorr-Euchner SD implementation with early termination is in fact DMT optimal when applied to a regularized and lattice reduced channel, cf. Section II-B.

SD search space from eight real dimensions to 6 real dimensions. This reduction is achieved by linearly combining two Alamouti style *twisted* codes, such that the corresponding QR decomposition employed in SD, yields a sparse $\boldsymbol{R}$ matrix. The sparseness property was shown to be unique to the case of $n_{\rm T} = T = 2$ where $n_{\rm T}$ and T denotes the number of transmit antennas and the coding duration, respectively, and further extensions to the 4×2 quasi-static MIMO channel came at the expense of reduced diversity [39].

Towards bridging the gap between ML and linear decoders, a hybrid transceiver was proposed in [43] for the quasi-static MIMO channel to jointly employ an ML and an *unbiased MMSE-SIC* receiver, on an infinitely long $(T \rightarrow \infty)$ D-BLAST style $n_{\rm T} \times T$ space-time spreading (STS) code with an underlying QAM constellation. This hybrid transceiver allows for *partial* reduction in decoding complexity, and provides DMT optimality with $2n_{\rm T}$-dimensional ML decoding (in every time slot). For the case where $n_{\rm R} \geq n_{\rm T}$, a pure ML receiver would generally incur a dimensionality of $2n_{\rm T}T$ real symbols.

One of the most fundamentally important steps towards establishing that DMT optimality can be achieved with computationally efficient encoders and decoders was given in [6]. In the setting of the i.i.d. Rayleigh fading quasi-static MIMO channel, it was shown that the random codes from the ensemble proposed in [6] may be DMT optimally decoded by a lattice decoder (whereby the constellation boundaries are ignored in the decoding process). This was accomplished by the inclusion of the MMSE-GDFE preprocessing step and a random lattice translate. It should, however, be noted that an exact implementation of the MMSE-GDFE lattice decoder still requires the solution to a closest-vector-problem, which is NP-hard in general [44]. Currently, except for the Alamouti transceiver structure [1] over the 2×1 quasi-static MISO channel, all known DMT optimal explicit, nonrandom, transceivers employ ML detection, and incur worst-case complexity that is exponential in the data rate.

B. Principal Results and Outline

The contribution of this work lies in the identification of a large class of scenarios where efficient variants of LR-aided linear lattice decoding, which is a generally suboptimal but computationally advantageous decoding strategy, achieve the DMT of the ML decoder. The work also presents the first explicit characterization of efficient non-ML encoder-decoder structures that meet the fundamental DMT performance limits, for very general channel statistics, dimensions, and models. DMT optimality is shown to be achieved with the smallest known complexity order among all DMT-optimal decoders that apply to general lattice designs.

As a first step towards providing computationally efficient DMT optimality, Theorem 1 in Section III-C, proves that regularized lattice decoders are DMT optimal. The proposed class of decoders employs an unconstrained lattice search in a *regularized* metric which applies an incremental penalization to lattice points further from the origin. The decoder structure includes, as a special case, the MMSE-GDFE lattice decoder [6], [24]. The DMT optimality holds irrespective of the channel's fading statistics and irrespective of the lattice design which is decoded (cf. [7]–[11], [13]–[16], [18]–[22]), as long as the lattice design and fading distribution jointly induce a (right) continuous[4] DMT curve (cf. [5]) under ML detection. Currently all known DMT curves for the system models considered herein are continuous except possibly at the maximal multiplexing gain. The result holds also when ML decoding, due to suboptimality of the code applied, does not achieve the fundamental DMT of the channel. This further strengthens the view of regularized lattice decoding as a DMT optimal *decoding* strategy.

As a second step towards computationally efficient DMT optimality, Theorem 2, in Section IV-A, extends the above result to the class of all C*-approximate* implementations of regularized lattice decoders. Two decoders are here said to be C-approximate when their minimum metrics are at a distance less than some *constant* C (cf. Section IV-A). The DMT optimality of LLL-based LR-aided linear decoders, being C-approximate decoders, is then established by Corollary 2a.

Theorem 3 in Section IV-C then considers the computational complexity of the LR-aided solutions and proves that LR-aided DMT optimal decoding is feasible at a *worst-case* complexity of $O(\log \rho)$ where ρ denotes the SNR, i.e., at a complexity which grows at most linearly in the data rate. With LLL LR worst-case complexity known to be generally unbounded [45], the upper bound is guaranteed by exploiting channel information at the receiver and rigorously relating code-channel lattices that result in high probability of error, to lattices that may induce high LR complexity. The bound quantifies, in the scale of interest, the fundamental reduction in the decoding complexity of the proposed explicit transceivers, when compared to the ML decoder which has a complexity that is generally exponential in the rate. It also resolves, in the negative, the long standing open problem of whether DMT optimality generally requires a complexity that is exponential in rate.

Section V considers different generalizations including the case of nested lattice designs, partial channel knowledge, general and possibly non-Gaussian noise characteristics, and provides a discussion of the case where the diversity multiplexing characteristic of some scenario is discontinuous and/or unknown. Section VI then shows how the result directly applies to several pertinent computationally demanding communication scenarios such as MIMO-OFDM, ISI, AF, decode-and-forward and MIMO-ARQ settings, in all of which the DMT optimality of the efficient decoders is guaranteed, again for any lattice design and fading distribution. Conclusions are provided in Section VII.

C. Notation

$\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$, respectively, denote the integer, the real, and the complex numbers. $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ denote the set of n-dimensional and $m \times n$-dimensional real vectors and matrices. Similar definitions apply to $\mathbb{Z}$ and $\mathbb{C}$. Vectors and matrices are, respectively, denoted by lower- and upper-case bold letters, i.e., $\boldsymbol{x}$ and $\boldsymbol{X}$. The identity matrix is denoted $\boldsymbol{I}$ and its size is made clear by the context. The all-zeros vector or matrix is denoted $\mathbf{0}$. $\boldsymbol{X}^{\rm T}$,

[4] A similar continuity assumption is required (although not explicitly stated) in establishing the DMT optimality of approximately universal codes, cf. [12, Th. 3.1].

$\boldsymbol{X}^{\mathrm{H}}$, and $\boldsymbol{X}^{-1}$ denotes the transpose, conjugate transpose, and inverse of a matrix $\boldsymbol{X}$. $\|\boldsymbol{x}\|$ denotes the Euclidean norm of $\boldsymbol{x}$, and $\|\boldsymbol{X}\|_{\mathrm{F}}^{2}$ the Frobenius norm of $\boldsymbol{X}$. No notational difference is made between random variables (vectors and matrices) and their realizations. The multivariate real valued Gaussian distribution with zero mean and covariance $\boldsymbol{I}$ is denoted $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

II. System Model

A. The Generic MIMO Channel

We consider a generic $n \times m$ (real) MIMO channel model

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

where $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{H} \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{w} \in \mathbb{R}^m$. The transmitted codewords $\boldsymbol{x}$ are assumed to be uniformly distributed over some codebook $\mathcal{X} \subset \mathbb{R}^n$, and statistically independent of $\boldsymbol{H}$. The noise is assumed to be i.i.d. Gaussian with unit variance, i.e., $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Under these assumptions the optimal decoder, in the sense that it minimizes the probability of codeword error, is the ML decoder given by

$$\hat{\boldsymbol{x}}_{\mathrm{ML}} = \arg\min_{\hat{\boldsymbol{x}} \in \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2. \tag{2}$$

The channel $\boldsymbol{H}$ is assumed random (i.e., fading) with a distribution parameterized by a real parameter $\rho \geq 0$. The parameter ρ will throughout be interpreted as the SNR of the channel, although this is strictly speaking not required for the analysis. We assume that one use of (1) corresponds to T uses of some underlying "physical" channel, which motivates a definition of the rate in terms of bits per channel use (bpcu) according to

$$R = \frac{1}{T}\log_2 |\mathcal{X}| \tag{3}$$

where $|\mathcal{X}|$ denotes the cardinality or size of $\mathcal{X}$. The model in (1) is known to encompass many pertinent communication scenarios (cf. [24]), and several explicit examples are provided in Section VI. The obtained results hold in the general setting unless otherwise explicitly stated.

B. The Diversity-Multiplexing Tradeoff

Following [5] we refer to a family of codes, $\mathcal{X}(\rho)$, parameterized by ρ as a *scheme* and define the *multiplexing gain* r of the scheme according to

$$r \triangleq \lim_{\rho\to\infty} \frac{R(\rho)}{\log_2 \rho} = \lim_{\rho\to\infty} \frac{1}{T}\frac{\log|\mathcal{X}(\rho)|}{\log \rho}. \tag{4}$$

As we will be interested in the system behavior as a function of the multiplexing gain r, we will use the term *design* to denote a set of schemes over some range of r. In this sense we would consider the Alamouti code [1] or V-BLAST [3] with appropriately chosen constellations as designs (cf. [5, Section VII]). We will in what follows write $\mathcal{X}_r$ to express the dependence of the codebook (or more appropriately the sequence of codebooks) on the r, while the dependence on ρ is suppressed for notational reasons. The *diversity gain* of the design under ML decoding is given, as a function of r, according to (cf. [5])

$$d_{\mathrm{ML}}(r) \triangleq -\lim_{\rho\to\infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}} \neq \boldsymbol{x})}{\log \rho} \tag{5}$$

(provided the limit exists) where $\boldsymbol{x}$ is assumed uniformly distributed over $\mathcal{X}_r$ and where $\hat{\boldsymbol{x}}_{\mathrm{ML}}$ is given by (2) for $\mathcal{X} = \mathcal{X}_r$. The expression in (5) will in general define a tradeoff between the multiplexing gain and diversity gain, particular to the design and channel at hand [5].

As shown in [5, Lemma 5] the diversity gain $d_{\mathrm{ML}}(r)$ is under the power constraint, $\mathrm{E}\left\{\|\boldsymbol{x}\|^2\right\} \leq T$, upper bounded by the outage exponent $d_{\mathrm{out}}(r)$ where

$$d_{\mathrm{out}}(r) \triangleq -\lim_{\rho\to\infty} \frac{\log \mathrm{P}(\log\det(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}) < 2Tr\log\rho)}{\log\rho}. \tag{6}$$

In the case of the i.i.d. Rayleigh fading quasi-static MIMO channel (cf. Section VI-A), $d_{\mathrm{out}}(r)$ is given by the piece-wise linear curve connecting $(k, (n_{\mathrm{R}} - k)(n_{\mathrm{T}} - k))$ for $k = 1, \ldots, \min(n_{\mathrm{T}}, n_{\mathrm{R}})$[5]. Similar results have been obtained for other fading distributions [46]. A code is said to be approximately universal [12] for the particular system model at hand if $d_{\mathrm{ML}}(r) = d_{\mathrm{out}}(r)$ under any fading distribution. For the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ quasi-static MIMO channel, approximately universal codes have been constructed for all r, n_{R}, and n_{T} provided $T \geq n_{\mathrm{T}}$[7], [8].

As frequently done in works on the DMT, we will make use of the $\doteq$ notation where $f(\rho) \doteq \rho^x$ iff (cf. [5])

$$\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log\rho} = x. \tag{7}$$

The symbols $\dot{\geq}$ and $\dot{\leq}$ are defined similarly. In this notation a scheme has multiplexing gain r if $|\mathcal{X}| \doteq \rho^{rT}$ and diversity gain d under ML decoding if $\mathrm{P}\left(\hat{\boldsymbol{x}}_{\mathrm{ML}} \neq \boldsymbol{x}\right) \doteq \rho^{-d}$.

III. Lattice Codes and Decoding

A. Lattice Designs

An n-dimensional real valued lattice Λ is the discrete additive subgroup of $\mathbb{R}^n$ given by

$$\Lambda \triangleq \{\boldsymbol{G}\boldsymbol{z} | \boldsymbol{z} \in \mathbb{Z}^n\}. \tag{8}$$

The full rank matrix $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ is referred to as the generator matrix of Λ. We will consider the class of lattice designs, defined as follows.

Definition 1 (Lattice Design):

A lattice design is defined by the pair $(\Lambda, \mathcal{R})$ where $\Lambda \subset \mathbb{R}^n$ is a lattice and $\mathcal{R}$ is a compact (i.e., closed and bounded) convex subset of $\mathbb{R}^n$, which contains $\boldsymbol{0}$ in its interior. For $r \geq 0$ the sequence of lattice codes $\mathcal{X}_r$ is given by $\mathcal{X}_r = \Lambda_r \cap \mathcal{R}$ where $\Lambda_r \triangleq \phi_r \Lambda$ and $\phi_r \triangleq \rho^{-\frac{rT}{n}}$.

As in [6], we refer to $\mathcal{R}$ as the *shaping region* of the lattice design. It is important to note that we assume that $\mathcal{R}$ and Λ are fixed and independent of ρ and that, in general, $\mathcal{R}$ has to be appropriately chosen so that the design satisfy the given power constraint, e.g., $\mathrm{E}\left\{\|\boldsymbol{x}\|^2\right\} \leq T$. This definition of a lattice design is slightly more restrictive than the definition of lattice space-time codes considered in [6], in that we require the same lattice (and shaping region) to be used for all multiplexing gains r and SNR ρ. Note, however, that while we restrict the maximum value of $\|\boldsymbol{x}\|^2$ by the shaping region, we are not restricting the analysis to short-term power constraints, as long-term power allocation policies may often be considered part of the effective channel $\boldsymbol{H}$.

It is straightforward to verify that the multiplexing gain of $\mathcal{X}_r$ is indeed r. By a principle, dating back to Gauss, stating that the number of lattice points in a large set is well approximated by the volume of the set, we have [47]

$$\begin{aligned}|\mathcal{X}_r| &= |\Lambda_r \cap \mathcal{R}| = |\phi_r \Lambda \cap \mathcal{R}| \\ &= |\Lambda \cap \phi_r^{-1}\mathcal{R}| = \frac{\phi_r^{-n} V(\mathcal{R})}{V(\mathcal{V}_\Lambda)} + o\left(\phi_r^{-n}\right) \doteq \rho^{rT} \qquad (9)\end{aligned}$$

where $V(\mathcal{R})$ and $V(\mathcal{V}_\Lambda)$ denotes the volume of the shaping region and the fundamental (Voronoi) cell of Λ, respectively.

The assumption that $\boldsymbol{G}$ is a square matrix can be made without loss of generality. To see this assume that $\boldsymbol{G} \in \mathbb{R}^{n \times k}$ where $k < n$ and note that for $\boldsymbol{x} \in \Lambda$ we have $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{z}$ for $\boldsymbol{z} \in \mathbb{Z}^k$. Write $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{G}'$ where $\boldsymbol{U} \in \mathbb{R}^{n \times k}$ has orthogonal columns and $\boldsymbol{G}' \in \mathbb{R}^{k \times k}$ is full rank, let $\boldsymbol{H}' = \boldsymbol{H}\boldsymbol{U}^{\mathrm{T}}$ and $\boldsymbol{x}' = \boldsymbol{G}'\boldsymbol{z}$. We obtain $\boldsymbol{H}\boldsymbol{x} = \boldsymbol{H}\boldsymbol{G}\boldsymbol{z} = \boldsymbol{H}'\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{G}'\boldsymbol{z} = \boldsymbol{H}'\boldsymbol{G}'\boldsymbol{z} = \boldsymbol{H}'\boldsymbol{x}'$, i.e., transmitting $\boldsymbol{x}$ over $\boldsymbol{H}$ is equivalent to transmitting $\boldsymbol{x}'$ over $\boldsymbol{H}'$. As no explicit assumption is made regarding the fading distribution of $\boldsymbol{H}$, we may equivalently consider the channel given by $\boldsymbol{H}'$, and use the square generator matrix $\boldsymbol{G}'$ in the formulation of the lattice design. The two equivalent cases naturally result in the same DMT curve. On the other hand, if $k > n$ we may extend $\boldsymbol{G} \in \mathbb{R}^{n \times k}$ to a $k \times k$ full rank matrix by the addition of $k - n$ linearly independent rows, while adding $k - n$ columns containing zeros to $\boldsymbol{H}$ in the corresponding positions, thus leaving the input-output relation of (1) unaltered.

The definition of a lattice design admits most of the codes mentioned in Section I-A-I in a straightforward manner in the sense that the code construction may be completely described by the pair $(\Lambda, \mathcal{R})$. The largest subclass of lattice codes, generally known as linear dispersion codes (cf. [4] and [6]), additionally satisfy $\boldsymbol{x} = \phi_r \sum_{i=1}^{n/2}(\boldsymbol{a}_i\alpha_i + \boldsymbol{b}_i\beta_i)$ for some fixed $\boldsymbol{a}_i, \boldsymbol{b}_i \in \mathbb{R}^n$, $i = 1, \ldots, n/2$, where α_i and β_i constitute the real and imaginary part of a complex constituent data symbol chosen from a suitable constellation, e.g., a QAM or HEX [48] constellation. The structure of the linear dispersion codes provides efficient encoding, and naturally yields a shaping region $\mathcal{R}$ in the form of an orthotope with axes aligned with the columns of the corresponding generator matrix $\boldsymbol{G} = [\boldsymbol{a}_1, \boldsymbol{b}_1, \ldots, \boldsymbol{a}_{n/2}, \boldsymbol{b}_{n/2}]$. The class of linear dispersion codes include the constructions in [7]–[11], [13], [14], [18]–[22] as well as many classical designs [1]–[3]. Also the codes with reduced decoding complexity in [39]–[43] belong to this class of codes. It is known that a better shaping gain may be achieved through a more careful design of the shaping region $\mathcal{R}$ (cf. [6] and [15]).

Before continuing, two remarks are in order. While [15] defines single lattices which provide strong lattice codes, the specific encoding strategy proposed in [15] will in general also introduce a (pseudorandom) translate of the lattice Λ_r. This is not covered by our basic definition of lattice designs which specifies the code exclusively in terms of Λ and $\mathcal{R}$. Although the results presented in the following straightforwardly extend to cover such lattice translates, we shall in the interest of notational simplicity not consider this at first. Instead, we outline the changes required to handle this generalization in Section V. Furthermore we remark that we make no assumptions regarding the optimality of the code design itself, i.e., we do not assume that $d_{\mathrm{ML}}(r) = d_{\mathrm{out}}(r)$, and consequently the results are applicable also to suboptimal designs such as, e.g., V-BLAST.

B. Lattice Decoding

The ML decoder in (2) implements a search for the codeword closest to $\boldsymbol{y}$ over $\mathcal{X}_r = \Lambda_r \cap \mathcal{R}$ [23], [24]. As in [6] and [24], we use the term *lattice decoding* to refer to an unconstrained search over Λ_r, i.e., a search where the constraint imposed by $\mathcal{R}$ is ignored by the decoder. The rationale behind such an approach is that it symmetrizes the problem and allows for the structure of the lattice to be exploited in order to reduce the computational complexity of the decoder [23]–[25].

The naive lattice decoder (cf. [6]) is obtained by simply removing the constraint imposed by $\mathcal{R}$ in the ML decoder while keeping the decision metric unaltered, i.e.

$$\hat{\boldsymbol{x}}_{\mathrm{NL}} = \arg\min_{\hat{\boldsymbol{x}} \in \Lambda_r} \|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2. \qquad (10)$$

In the event that $\hat{\boldsymbol{x}}_{\mathrm{NL}} \notin \mathcal{X}_r$ the decoder declares an error. It is known that the performance loss incurred by neglecting the codebook boundary $\mathcal{R}$ may in this case be substantial, and that the naive lattice decoder is not DMT optimal in general [6], [34]. In particular, naive lattice decoding was shown to be a strictly suboptimal decoding strategy for perfect ST codes applied to the point-to-point MIMO channel with $n_{\mathrm{T}} \geq 2$ [34]. It was further shown in [49], albeit for a fixed-rate uncoded setting, that mapping $\hat{\boldsymbol{x}}_{\mathrm{NL}}$ back to $\mathcal{X}_r \subset \Lambda_r$ by component-wise rounding does not significantly improve performance. Nonetheless, as proved in [6] for the i.i.d. Rayleigh fading quasi-static MIMO channel, the problem does not lie with lattice decoding per se, but with the naive implementation. In particular, after an appropriate alteration of the decoding metric, it was by a random coding argument shown that lattice coding and decoding is sufficient for achieving DMT optimal performance in this scenario [6].

Intuitively, as the naive lattice decoder (10) is suboptimal in terms of its diversity, it must mean that $\hat{\boldsymbol{x}}_{\mathrm{NL}} \neq \mathcal{R}$ with a probability that is large in relation to $\mathrm{P}\left(\hat{\boldsymbol{x}}_{\mathrm{ML}} \neq \boldsymbol{x}\right)$, i.e., the decoder is relatively likely to decide in favor of a codeword outside the region defined by $\mathcal{R}$. As $\mathcal{R}$ is bounded it is plausible that a regularization [50] of the decoding metric may reduce the probability of "out of region" error events, and improve the probability of error.

C. DMT Optimality of Regularized Lattice Decoding

The (general) regularized lattice decoder is given by

$$\hat{\boldsymbol{x}}_L = \arg\min_{\hat{\boldsymbol{x}}\in\Lambda_r} \|\boldsymbol{y}-\boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} \tag{11}$$

where $\|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} = \hat{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{T}\hat{\boldsymbol{x}}$ for some given positive definite matrix $\boldsymbol{T} = \boldsymbol{T}^{\mathrm{T}}$. The additive term $\|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}}$ applies an incremental penalization to lattice points further from the origin, and reduces the probability of error associated with codewords outside of the shaping region. This notion is formalized by the following theorem, which constitutes one of the main contributions of this work, and states that (11) is a DMT optimal decoding strategy for lattice designs, in a remarkably general sense. The proof is given in Section III-D.

Theorem 1: For any lattice design $(\Lambda, \mathcal{R})$, and for any fading distribution such that $d_{\mathrm{ML}}(r)$ is (right) continuous at r, the regularized lattice decoder is DMT optimal, i.e.

$$d_L(r) = d_{\mathrm{ML}}(r), \tag{12}$$

where

$$d_L(r) \triangleq -\lim_{\rho\to\infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x})}{\log\rho}, \tag{13}$$

for $\boldsymbol{x}$ uniformly distributed over $\mathcal{X}_r$, and $\hat{\boldsymbol{x}}_L$ given by (11).

Before proving Theorem 1, we remark that for $\boldsymbol{T} = \boldsymbol{I}$ the regularized decoder is equivalent to the MMSE-GDFE decoder considered in [6], if we neglect the lattice translate considered therein. In particular, the regularized lattice decoder in (11) is equivalently given by (cf. Appendix A)

$$\hat{\boldsymbol{x}}_L = \arg\min_{\hat{\boldsymbol{x}}\in\Lambda_r} \|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}\|^2 \tag{14}$$

where $\boldsymbol{F}\in\mathbb{R}^{n\times m}$ and $\boldsymbol{B}\in\mathbb{R}^{n\times n}$ are MMSE-GDFE forward and feedback filters [6]. This equivalence is interesting in light of the fact that the motivation of the MMSE-GDFE decoder in [6] was largely information theoretic in nature, while the regularization view is arguably of a more signal processing flavor. Theorem 1 thus extends the results of [6] and proves DMT optimality of MMSE-GDFE decoding for any lattice designs based on a single, fixed, generator matrix. We also note that although the specific matrix $\boldsymbol{T}$ in (11) has no effect on the diversity gain (provided $\boldsymbol{T}$ is full rank) it may significantly affect the coding gain and should in practice be chosen based on the shaping region, code, and channel statistics.

D. Proof of Theorem 1

We begin by providing the following lemma, proven in Appendix B. The purpose of the lemma is to connect the probability of ML error with the existence of a small codeword difference $\|\boldsymbol{H}(\hat{\boldsymbol{x}}_1-\hat{\boldsymbol{x}}_2)\|^2$ where $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ belong to a subset of the codebook. In essence, the lemma provides a "deep fade typical error" probability bound in line with [29, Ch. 3].

Lemma 1: Let $\mathcal{B}$ be the spherical region given by

$$\mathcal{B} \triangleq \{\boldsymbol{d}\in\mathbb{R}^n \mid \|\boldsymbol{d}\|^2 \leq \gamma\} \tag{15}$$

where the radius $\gamma > 0$ (independent of ρ) is chosen such that $\boldsymbol{d}_1 + \boldsymbol{d}_2 \in \mathcal{R}$ for any $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathcal{B}$. Let

$$\nu_r \triangleq \min_{\boldsymbol{d}\in\mathcal{B}\cap\Lambda_r:\boldsymbol{d}\neq\boldsymbol{0}} \frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2. \tag{16}$$

Then, for any $r > 0$ it holds that

$$\limsup_{\rho\to\infty} \frac{\log\mathrm{P}(\nu_r \leq 1)}{\log\rho} \leq -d_{\mathrm{ML}}(r). \tag{17}$$

The existence of the set $\mathcal{B}$ in (15) follows by the assumption that $\boldsymbol{0}$ is contained in the interior of $\mathcal{R}$. Now, let $\zeta > 0$ be given and choose $\delta > 0$ such that

$$\frac{2\zeta T}{n} > \delta > 0. \tag{18}$$

This may clearly be done for arbitrary $\zeta > 0$. We will in the following assume that $\nu_{r+\zeta} \geq 1$ and that $\|\boldsymbol{w}\|^2 \leq \rho^{\delta}$, and prove that these two conditions are sufficient for a correct decision by the regularized lattice decoder in (11), provided that ρ is sufficiently large. Hence, in order for an error to occur at large ρ, one of the assumptions must fail.

To this end, consider first the metric in (11) for the transmitted codeword $\boldsymbol{x}$, i.e.,

$$\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|^2_{\boldsymbol{T}} \leq \rho^{\delta} + c \tag{19}$$

where $\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x} = \boldsymbol{w}$ and $\|\boldsymbol{w}\|^2 \leq \rho^{\delta}$ was used, and where

$$c \triangleq \max_{\boldsymbol{r}\in\mathcal{R}} \|\boldsymbol{r}\|^2_{\boldsymbol{T}}.$$

Note that $c < \infty$ as $\mathcal{R}$ is bounded and that c is independent of the transmitted codeword $\boldsymbol{x}$ and ρ.

In order to bound the metric for $\hat{\boldsymbol{x}} \in \Lambda_r$ where $\hat{\boldsymbol{x}} \neq \boldsymbol{x}$, we note that $\nu_{r+\zeta} \geq 1$ implies

$$\frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2 \geq 1 \quad \forall\, \boldsymbol{d}\in\mathcal{B}\cap\Lambda_{r+\zeta},\ \boldsymbol{d}\neq\boldsymbol{0} \tag{20}$$

by the definition in (16). As $\Lambda_r = \rho^{\frac{\zeta T}{n}}\Lambda_{r+\zeta}$ it follows that

$$\frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2 \geq \rho^{\frac{2\zeta T}{n}} \quad \forall\, \boldsymbol{d}\in\rho^{\frac{\zeta T}{n}}\mathcal{B}\cap\Lambda_r,\ \boldsymbol{d}\neq\boldsymbol{0} \tag{21}$$

after scaling (20) by $\rho^{\frac{\zeta T}{n}}$. As $\mathcal{R}$ is bounded, and as $\zeta > 0$, it holds that $\mathcal{R} \subset \frac{1}{2}\rho^{\frac{\zeta T}{n}}\mathcal{B}$ for all $\rho \geq \rho_1$, given some sufficiently large ρ_1. This implies that $\boldsymbol{x} \in \frac{1}{2}\rho^{\frac{\zeta T}{n}}\mathcal{B}$ for $\rho \geq \rho_1$ since $\boldsymbol{x}\in\mathcal{R}$. It is important to note here that while ρ_1 may depend on ζ and $\mathcal{R}$, it can be chosen independent of the particular $\boldsymbol{x}$ transmitted.

For any $\hat{\boldsymbol{x}} \in \frac{1}{2}\rho^{\frac{\zeta T}{n}}\mathcal{B}\cap\Lambda_r$, $\hat{\boldsymbol{x}} \neq \boldsymbol{x}$, it holds that $\boldsymbol{d} = \boldsymbol{x}-\hat{\boldsymbol{x}} \in \rho^{\frac{\zeta T}{n}}\mathcal{B}\cap\Lambda_r$. By (21) we have

$$\frac{1}{4}\|\boldsymbol{H}(\boldsymbol{x}-\hat{\boldsymbol{x}})\|^2 = \frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2 \geq \rho^{\frac{2\zeta T}{n}} \tag{22}$$

where $\boldsymbol{d} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. As $\|\boldsymbol{w}\|^2 \leq \rho^\delta$ it follows by (22) and (18) that $\frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2 \gg \|\boldsymbol{w}\|^2$ for large ρ. In particular, there is some $\rho_2 \geq \rho_1$, independent of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, for which the triangle inequality implies that

$$\|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 = \|\boldsymbol{H}(\boldsymbol{x} - \hat{\boldsymbol{x}}) + \boldsymbol{w}\|^2 \geq \rho^{\frac{2\zeta T}{n}}$$

for all $\rho \geq \rho_2$. Consequently

$$\|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} \geq \rho^{\frac{2\zeta T}{n}} \tag{23}$$

for any $\hat{\boldsymbol{x}} \in \Lambda_r$ where $\hat{\boldsymbol{x}} \in \frac{1}{2}\rho^{\frac{\zeta T}{n}}\mathcal{B}$ and $\rho \geq \rho_2$.

In the case that $\hat{\boldsymbol{x}} \notin \frac{1}{2}\rho^{\frac{\zeta T}{n}}\mathcal{B}$, it follows by the definition in (15) that $\|\hat{\boldsymbol{x}}\|^2 \geq \frac{1}{4}\gamma\rho^{\frac{2\zeta T}{n}}$ which implies $\|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} \geq \frac{1}{4}\gamma\lambda_{\min}(\boldsymbol{T})\rho^{\frac{2\zeta T}{n}}$ where $\lambda_{\min}(\boldsymbol{T}) > 0$ denotes the minimum eigenvalue of $\boldsymbol{T}$. It follows that

$$\|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} \geq \frac{1}{4}\gamma\lambda_{\min}(\boldsymbol{T})\rho^{\frac{2\zeta T}{n}} \tag{24}$$

for any $\hat{\boldsymbol{x}} \notin \rho^{\frac{\zeta T}{n}}\mathcal{B}$.

Let

$$a(\rho) \triangleq \rho^\delta + c \quad \text{and} \quad b(\rho) \triangleq \min\left(1, \frac{1}{4}\gamma\lambda_{\min}(\boldsymbol{T})\right)\rho^{\frac{2\zeta T}{n}} \tag{25}$$

and note that (18) implies that there is some $\rho_3 \geq \rho_2$, again independent of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, for which $a(\rho) < b(\rho)$ for all $\rho > \rho_3$. For the transmitted codeword $\boldsymbol{x}$ we have by (19) that

$$\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|^2_{\boldsymbol{T}} \leq a(\rho).$$

For any other $\hat{\boldsymbol{x}} \in \Lambda_r$ (i.e., $\hat{\boldsymbol{x}} \in \Lambda_r \backslash \{\boldsymbol{x}\}$) it holds by (23) and (24) that

$$\|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}} \geq b(\rho) > a(\rho) \tag{26}$$

for all $\rho \geq \rho_3$. This implies that the transmitted codeword yields the minimum metric in (11), or equivalently that $\hat{\boldsymbol{x}}_L = \boldsymbol{x}$ as long as $\rho \geq \rho_3$ and under the assumptions that $\nu_{r+\zeta} \geq 1$ and $\|\boldsymbol{w}\|^2 \leq \rho^\delta$. For an error to occur when $\rho \geq \rho_3$ it is thus required that $\nu_{r+\zeta} < 1$ or $\|\boldsymbol{w}\| > \rho^\delta$.

Applying the union bound to the probability of error yields

$$\mathrm{P}(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x}) \leq \mathrm{P}(\nu_{r+\zeta} < 1) + \mathrm{P}(\|\boldsymbol{w}\| > \rho^\delta), \tag{27}$$

for $\rho \geq \rho_3$. As $\mathrm{P}\left(\|\boldsymbol{w}\| > \rho^\delta\right) \doteq \rho^{-\infty}$, due to the exponential tail of the Gaussian distribution, the second term in (27) is asymptotically irrelevant. By Lemma 1 it follows that $\mathrm{P}\left(\nu_{r+\zeta} < 1\right) \dot{\leq} \rho^{-d_{\mathrm{ML}}(r+\zeta)}$. Note here also that Lemma 1 is applicable even when $r = 0$ since it is applied at a multiplexing gain of $r + \zeta > 0$. It follows that

$$\limsup_{\rho \to \infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x})}{\log \rho} \leq -d_{\mathrm{ML}}(r + \zeta). \tag{28}$$

By observing that (28) holds for an arbitrary choice of $\zeta > 0$, we may conclude that

$$\limsup_{\rho \to \infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x})}{\log \rho} \leq -d_{\mathrm{ML}}(r) \tag{29}$$

for any $r \geq 0$, provided that

$$\lim_{\zeta \to 0^+} d_{\mathrm{ML}}(r + \zeta) = d_{\mathrm{ML}}(r),$$

i.e., provided $d_{\mathrm{ML}}(r)$ is right continuous at r. As $\mathrm{P}\left(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x}\right) \geq \mathrm{P}\left(\hat{\boldsymbol{x}}_{\mathrm{ML}} \neq \boldsymbol{x}\right)$ due to the optimality of the ML decoder it holds that

$$\liminf_{\rho \to \infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_L \neq \boldsymbol{x})}{\log \rho} \geq -d_{\mathrm{ML}}(r)$$

which combined with (29) establish the claim of Theorem 1.

E. A Geometric Example

In order to provide further intuition into the suboptimality of the naive lattice decoder, and the argument made in Section III-D it is useful to consider the example provided in Fig. 1, where Λ_r is a scaled version of the integer lattice $\mathbb{Z}^2$ and where the shaping region $\mathcal{R}$ is spherical. The image of Λ_r and $\mathcal{R}$ under the linear map induced by $\boldsymbol{H}$ are shown in Fig. 1(b). In the example, $\boldsymbol{H} \in \mathbb{R}^{2\times 2}$ is nearly rank deficient. For the illustration, $\sigma_1(\boldsymbol{H}) = 40\sigma_2(\boldsymbol{H})$ where $\sigma_i(\boldsymbol{H})$ denotes the ith singular value of $\boldsymbol{H}$.

We will in the following discussion assume that $\boldsymbol{x} = \boldsymbol{0}$ corresponds to the transmitted codeword and, for simplicity, that $\boldsymbol{T} = \boldsymbol{I}$. As seen in Fig. 1(b) no other codeword $\hat{\boldsymbol{x}} \in \mathcal{X}_r \backslash \{\boldsymbol{x}\}$ is mapped close to $\boldsymbol{H}\boldsymbol{x}$ by the linear map $\boldsymbol{H}$. Thus, the ML decoder is unlikely to make an error. However, when considering decoding to the full lattice Λ_r, the (naive) lattice decoder is likely to decide in favor of the, in Fig. 1(a), indicated codeword $\hat{\boldsymbol{x}} \in \Lambda_r$. This is a consequence of the fact that $\hat{\boldsymbol{x}}$ lies close to the space spanned by the right singular vector corresponding to the smallest singular value of $\boldsymbol{H}$ (cf. Fig. 1(a)). The closeness of $\boldsymbol{H}\hat{\boldsymbol{x}}$ to $\boldsymbol{H}\boldsymbol{x}$ illustrates the problem with the naive lattice decoder, i.e., even when no codewords in $\mathcal{X}_r$ lie close to the space corresponding to a weak singular value of $\boldsymbol{H}$ it may be likely that a "hypothetical" codeword in Λ_r does. This view is strengthened by the observation that the performance of the naive lattice decoder is often determined by the statistics of the channel's weakest eigenmode (cf. [6] and [34]), although the fixed-rate V-BLAST result in [33] provides an exception to this rule. Note also that in this particular example, mapping $\hat{\boldsymbol{x}}$ to the closest point in $\mathcal{X}_r$ will not lead to a correct decision (cf. [49]).

The intuitive argument behind the regularization in (11) is that any lattice point $\hat{\boldsymbol{x}}$ (far) outside the constellation region $\mathcal{R}$, which implies that $\|\hat{\boldsymbol{x}}\|^2$ is large, is significantly penalized by the regularized decision metric. For codewords $\hat{\boldsymbol{x}} \neq \boldsymbol{x}$ in $\mathcal{R}$ the first quadratic term in (11) will be large, unless the ML decoder is also likely to be in error. Although this heuristic argument fails for codewords $\hat{\boldsymbol{x}}$ close to the boundary of $\mathcal{R}$, this problem may be circumvented under the continuity assumption of Theorem 1 by considering a larger constellation region, corresponding to the codebooks used at a marginally higher multiplexing gain.

The effect of the regularization can also be seen in Fig. 1(c) that shows the image of Λ_r under the linear transformation of the MMSE-GDFE feedback filter $\boldsymbol{B}$ in (14), corresponding to a regularized version of $\boldsymbol{H}$. For the purpose of the illustration, we have chosen $\boldsymbol{B}$ so that it shares left and right singular vectors

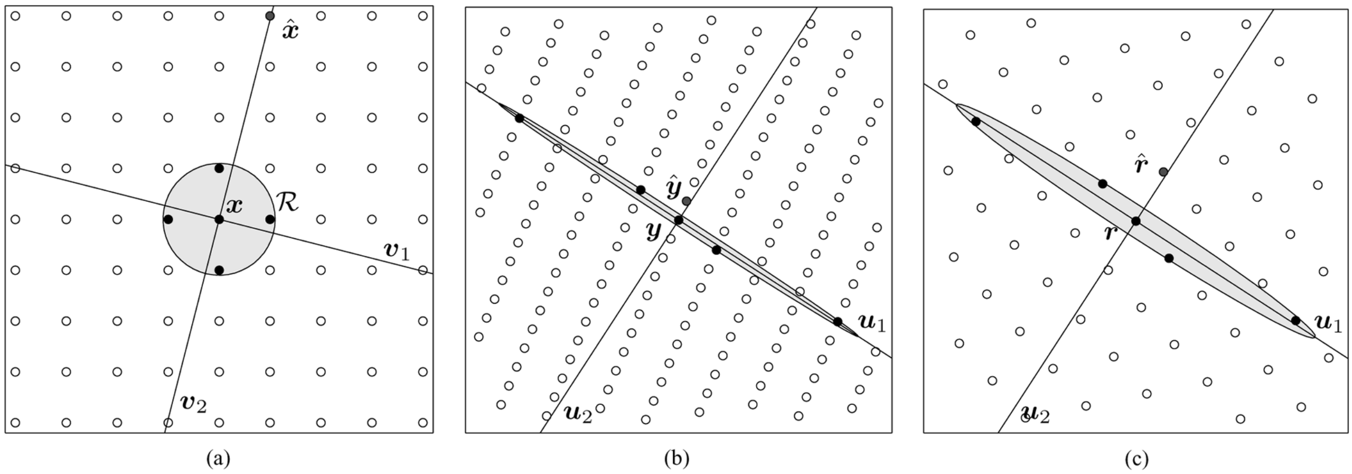

Fig. 1. Transformation of lattice Λ_r and spherical shaping region $\mathcal{R}$ under linear map induced by channel $\boldsymbol{H}$ and MMSE filter $\boldsymbol{B}$. Singular vectors of $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, where $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2)$, are shown as solid lines for reference. The matrix $\boldsymbol{B}$ is such that $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ where $\boldsymbol{B}$ shares left and right singular vectors with $\boldsymbol{H}$. Further, $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$, $\hat{\boldsymbol{y}} = \boldsymbol{H}\hat{\boldsymbol{x}}$, $\boldsymbol{r} = \boldsymbol{B}\boldsymbol{x}$, and $\hat{\boldsymbol{r}} = \boldsymbol{B}\hat{\boldsymbol{x}}$.

with $\boldsymbol{H}$. While the image of codewords inside $\mathcal{R}$ under the transformations $\boldsymbol{H}$ and $\boldsymbol{B}$ are relatively similar [cf. Fig. 1(b) and (c)], codewords outside the constellation $\mathcal{R}$ are more affected by the change from $\boldsymbol{H}$ to $\boldsymbol{B}$. Note in particular the difference between $\hat{\boldsymbol{y}} = \boldsymbol{H}\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{r}} = \boldsymbol{B}\hat{\boldsymbol{x}}$ in Fig. 1(b) and (c). Decoding to the closest lattice point in Fig. 1(c) is in this case clearly a better approximation of the ML decoder than decoding to the closest lattice point in Fig. 1(b).

IV. Computationally Efficient Decoding

A. DMT Optimality of Approximate Lattice Decoding

Obtaining $\hat{\boldsymbol{x}}_L$ in (11) still requires the minimization of a quadratic function over the discrete lattice Λ_r, a problem which is NP-hard in general, even after preprocessing [44]. This implies that even if lattice reduction techniques are used when obtaining the exact solution to (11), it is unlikely that there will be any general techniques with a (worst-case) complexity that grows sub-exponentially in the problem dimension n, unless the code itself provides a structure that simplifies decoding, such as for example in the case of orthogonal designs [1], [2]. For most high-performance lattice codes no such efficient solutions to (11) are known, which motivates the study of suboptimal implementations of the regularized lattice decoder.

The codeword $\hat{\boldsymbol{x}}_L$ is by definition the codeword which provides the minimum metric in (11). A *C-approximate solution* to (11) is any $\hat{\boldsymbol{x}} \in \Lambda_r$ which for $C > 1$ satisfies

$$\xi(\hat{\boldsymbol{x}}) \le C\xi(\hat{\boldsymbol{x}}_L) \quad \text{where} \quad \xi(\hat{\boldsymbol{x}}) = \|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{\boldsymbol{x}}\|^2_{\boldsymbol{T}}. \quad (30)$$

An algorithm that for fixed C is capable of producing a C-approximate solution to (11), for arbitrary inputs $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{H} \in \mathbb{R}^{m\times n}$, is referred to as a *C-approximation algorithm* [51]. In what follows we prove that any C-approximation algorithm for (11) is sufficient for DMT optimal decoding in the sense of Theorem 1.

Theorem 2: For any lattice design $(\Lambda, \mathcal{R})$, and fading distribution such that $d_{\mathrm{ML}}(r)$ is (right) continuous at r, all C-approximate implementations of the regularized lattice decoder are DMT optimal provided C is independent of ρ, i.e.,

$$d_{\mathrm{A}}(r) = d_{\mathrm{ML}}(r), \quad (31)$$

where

$$d_{\mathrm{A}}(r) \triangleq -\lim_{\rho\to\infty} \frac{\log \mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{A}} \neq \boldsymbol{x})}{\log \rho}, \quad (32)$$

for $\boldsymbol{x}$ uniformly distributed over $\mathcal{X}_r$, and where $\hat{\boldsymbol{x}}_{\mathrm{A}}$ is any C-approximate solution to (11).

Proof: The proof follows from the proof of Theorem 1, provided in Section III-D. In particular, consider $a(\rho)$ and $b(\rho)$ defined in (25). By the assumption in (18) it follows that

$$\lim_{\rho\to\infty} \frac{b(\rho)}{a(\rho)} = \infty.$$

We may thus select $\rho_4 \ge \rho_3$ such that $b(\rho) \ge Ca(\rho)$ for all $\rho \ge \rho_4$. As the metric for the transmitted codeword $\boldsymbol{x}$ is upper bounded by $a(\rho)$, and the metric of any other codeword is lower bounded by $b(\rho)$, it follows that when $\rho \ge \rho_4$, the only C-approximate solution to (11) is $\boldsymbol{x}$, i.e., $\hat{\boldsymbol{x}}_{\mathrm{A}} = \boldsymbol{x}$ for $\rho \ge \rho_4$, under the assumptions that $\nu_{r+\zeta} \le 1$ and $\|\boldsymbol{w}\|^2 \le \rho^{\delta}$. The remaining proof is then analogous to the proof of Theorem 1 in Section III-D. □

B. DMT Optimality of LR-aided Lattice Decoding

The existence of computationally efficient C-approximate solutions is thus of interest for DMT optimal decoding of lattice designs. Fortunately, such solutions are already known, both with respect to (11), or to the equivalent MMSE-GDFE formulation in (14). In fact, as shown in Appendix A, any C-approximate solution to (14) is also a C-approximate solution to (11). Of special interest in the communications context is Babai's nearest plane algorithm [52], which is equivalent to the LLL-based [53] LR-aided SIC solution to (14) [25], [31], [32],

[52]. The nearest plane algorithm provides a computationally efficient C_1-approximate solution (14) with $C_1 \triangleq 2^{\frac{n}{2}}$ [52]. Similarly, the LLL-based LR-aided linear solution to (14), discussed in [52] as the rounding algorithm, provides a C_2-approximate solution whith $C_2 \triangleq 1 + 2n(9/2)^{\frac{n}{2}}$. For completeness, we give the following corollary to Theorem 2.

Corollary 2a: The efficient LLL-based LR-aided linear (or SIC) implementations of the regularized lattice decoders provide DMT optimal decoding of any lattice design under the assumptions made in Theorem 1 and 2.

Proof: The corollary follows by the equivalence of the LR-aided linear decoder and the rounding algorithm in [52], or of the LR-aided SIC decoder and the nearest plane algorithm in [52], in conjunction with Theorem 2. □

Corollary 2a applies directly to the LR-aided linear implementation of the MMSE-GDFE decoder [24], [25], [35], due to the equivalence of the MMSE-GDFE decoder and the regularized decoder as outlined in Appendix A. The corollary applies also to the LR-aided MMSE-SIC decoder proposed in [36], when applied to the equivalent channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{G}\boldsymbol{s} + \boldsymbol{w}$$

where $\boldsymbol{s} \in \phi_r \mathbb{Z}^n$. Note however that in the latter case we would have $\boldsymbol{T} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}$, as opposed to $\boldsymbol{T} = \boldsymbol{I}$, reflecting a regularization of $\boldsymbol{s}$ rather than $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{s}$. In the case of perfect codes [13], where $\boldsymbol{G} = \boldsymbol{I}$, the metric of the MMSE-GDFE and the MMSE-SIC decoder coincides.

Corollary 2a applies also to a time-limited implementation of the Schnorr-Euchner (SE) sphere decoder [23], [54] operating in the LLL reduced regularized lattice, provided the sphere decoder tree-search is allowed to reach the first leaf-node. This follows as the first leaf-node found by the SE SD corresponds to the Babai-point, i.e., the solution obtained by the nearest plane algorithm (cf. [23]). Finding further candidate codewords with smaller metric can only improve the approximation ratio.

C. Decoding Complexity

Both the LR-aided SIC and linear decoders discussed above begin by LLL reducing the lattice generated by $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{G}$, where $\boldsymbol{G}$ is the generator matrix of Λ and where $\boldsymbol{B}$ is the MMSE feedback filter (cf. [6] and Appendix A), followed by a SIC or linear decoding stage in the reduced basis. Note here that by the regularization of $\boldsymbol{B}$ the matrix $\boldsymbol{M}$ is always full rank which makes the LLL algorithm applicable, regardless of the channel realization and the system dimensionality. The complexity of the decoding stage is only $O(n^2)$ [31], [32], [36] while the pre-processing relying on the LLL reduction is more complex. It is therefore relevant to consider the complexity of the LLL algorithm when applied to $\boldsymbol{M}$ in order to address the complexity of DMT optimal decoding of lattice designs. We refer the reader to [31], [32], and [36] for the implementation details of LR-aided decoders.

The LLL algorithm provides an iterative approach to lattice reduction [53]. The number K of LLL iterations required to reduce a given lattice generator matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ may be bounded according to [45], [55]

$$K \leq n^2 \log_s \kappa(\boldsymbol{M}) + n \tag{33}$$

where $s = 2/\sqrt{3}$ and where $\kappa(\boldsymbol{M})$ denotes the 2-norm condition number of $\boldsymbol{M}$. Each iteration requires $O(n^2)$ floating point operations [53]. The number of operations per iteration may, however, be reduced to $O(n)$ if only an effectively LLL-reduced basis is required, as is the case when a SIC decoder is applied in the reduced basis [56].

It is important here to note that for arbitrary $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ there is no universal upper bound on the number of iterations required to reduce $\boldsymbol{M}$ [45]. Thus, the worst-case complexity of the LLL-based LR-aided decoder is unbounded if applied to arbitrary channels. However, in order to achieve DMT optimal performance it is not required to LLL reduce every conceivable channel. To see this, consider a decoder implementation which is allowed to time-out, and declare an error, when the number of floating point operations exceeds a given threshold. Denote the time-out event $\mathcal{T}$, and note that as long as $\mathrm{P}(\mathcal{T}) \dot{\leq} \rho^{-d_{\mathrm{ML}}(r)}$ the time-limitation imposed will not reduce the diversity gain, or potential DMT optimality, of the decoder. In light of (33) we may thus limit the application of the LLL algorithm to bases $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{G}$ with bounded condition number $\kappa(\boldsymbol{M})$, or allow the decoder the option to time out, stop, and declare an error. In order to be able to provide an effective statement regarding the worst case decoding complexity under time-outs, we impose here a moderate restriction on the channels considered.

We say that a channel is *power limited* if $\mathrm{E}\left\{\|\boldsymbol{H}\|_{\mathrm{F}}^2\right\} \dot{\leq} \rho$ and note that this is required whenever we wish to interpret the parameter ρ as an average SNR at the receiver. For the class of power limited channels we may make the following statement, proven in Appendix C.

Lemma 2: For any power limited channel there is some constant $\alpha > 0$ where for $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{G}$ it holds that

$$\mathrm{P}(\kappa(\boldsymbol{M}) \geq \rho^{\alpha}) \dot{\leq} \rho^{-d_{\mathrm{ML}}(r)} \tag{34}$$

provided $d_{\mathrm{ML}}(r) < \infty$.

By applying Lemma 2, (33) and Corollary 2a, together with the previous discussion, the following statement regarding the complexity of DMT optimal decoding can thus be made. Note here that the signal space dimension n is considered fixed and is thus hidden in the big-O expression.

Theorem 3: For power limited channels, over any range of multiplexing gains r where $d_{\mathrm{ML}}(r)$ is continuous, DMT optimal decoding of any lattice design is feasible at a worst-case complexity of $O(\log \rho)$.

Proof: The theorem follows by imposing the constraint $\kappa(\boldsymbol{M}) \leq \rho^{\alpha}$ in (33), where α is chosen according to Lemma 2, and noting that such a restriction in the set of channels to which the decoder is applied does not reduce the diversity. □

Although the bound in Theorem 3 implies an increase in the LLL LR complexity for increasing SNR, this complexity only grows linearly in $\log \rho$. By comparing to (3) and (9) it may be seen that this corresponds to a linear increase in complexity as a function of the rate R at high SNR. The LLL complexity should also be put in context with the full search implementation of the ML decoder whose complexity is $|\mathcal{X}_r|$ and thus exponential in R. The worst-case sphere decoding complexity reported for fast decodable codes [39]–[42] is also exponential in R, albeit with a smaller exponent than the full search. The same holds true for the hybrid transceiver in [43] (given $n_{\mathrm{T}} \geq 2$). All such lattice-based designs may, however, be DMT optimally decoded using an LR-aided regularized lattice decoder structure with $O(\log \rho)$ complexity, potentially at some loss in coding gain, but at no diversity loss.

Finally, we note that in the case where $d_L(r) = \infty$ the statement in (34) in Lemma 2 cannot be guaranteed based on the condition that $\mathrm{E}\left\{\|\boldsymbol{H}\|^2\right\} \dot{\leq} \rho$ alone. However, for any channel statistics under which $\mathrm{P}\left(\|\boldsymbol{H}\|^2 \geq \rho^{\alpha}\right) \doteq \rho^{-\infty}$ for some sufficiently large α, Theorem 3 still applies. This includes for instance the quasi-static MIMO channel (cf. Section VI-A) under i.i.d. Rayleigh fading, or any other fading distribution with exponential tails.

D. The Search for Improved Approximation Algorithms

It is in the context of C-approximation algorithms important to note that while DMT optimality follows for any finite C, the gap in terms of SNR to the optimal implementation of (11) will in general depend on C. Thus, the loss in performance at practical SNR may be unacceptable for unduly large values of C. This motivates further study into new approximation algorithms, and code designs, that jointly yield improved approximation ratios.

Such methods may include stronger LR methods such as the *deep insertion* LLL variant [54] that is more computationally expensive but which finds better bases. Other LR approaches include methods based on the Korkine-Zolotareff bases (cf. [23]), and the algebraic lattice reduction approach in [57]. The latter method was presented for the 2×2 golden code [58] over the quasi-static MIMO channel, and approximates the channel matrix with the matrix representation of an invertible element of the maximal order of the CDA. Codes in which the ML decoder may be applied to spaces of reduced dimensionality (cf. [37], [38], as well as [39]–[42]) may benefit from a reduced gap between ML and lattice decoding due to the general dependence of the approximation constant C and the lattice dimension. This would suggest the use of transceivers based on reduced-dimensionality codes and regularized lattice decoding, as a good way to further approach ML error performance with a reduced SNR penalty. The topic of C-approximate solutions is, however, in the context of space-time decoding relatively unexplored at this stage.

V. Generalizations

In this section we consider a few straightforward generalizations in terms of the class of designs covered by the results as well as the modeling assumptions imposed in Section II.

A. Nested Lattice Designs

In the proof of Theorem 1, and in the lattice designs of Section III-A, we assume a fixed shaping region $\mathcal{R}$, applied for all ρ. This condition could, however, be relaxed in favor of a sequence of shaping regions $\mathcal{R}(\rho)$, such that $\underline{\mathcal{R}} \subset \mathcal{R}(\rho) \subset \overline{\mathcal{R}}$ for sufficiently large ρ where $\underline{\mathcal{R}}$ and $\overline{\mathcal{R}}$ are fixed "inner" and "outer" shaping regions that satisfy the conditions in Section III-A. Such an extension could be of interest for nested lattice codes [59] involving a shaping lattice $\underline{\Lambda}_r$ satisfying $\underline{\Lambda}_r \subset \Lambda_r$ where $\mathcal{R}$ is the Voronoi region of $\underline{\Lambda}_r$, i.e., $\mathcal{R} = \mathcal{V}_{\underline{\Lambda}_r}$ [6], [15], [59]. One option along this line is to let $\underline{\Lambda}_r = \omega_r \Lambda_r$ where $\omega_r \in \mathbb{N}$ is an appropriately selected integer (i.e., self-similar nesting [6]). This will in general require $\mathcal{R}$ to weakly depend on ρ, if we wish the code to be properly defined for all r and ρ. Alternatively, self-similar nested designs could also be accommodated by replacing the assumption that $\phi_r = \rho^{\frac{-rT}{n}}$ by the relaxed assumption $\phi_r \doteq \rho^{-\frac{rT}{n}}$, e.g., $\phi_r^{-1} = \lceil \rho^{\frac{rT}{n}} \rfloor$ where $\lceil \cdot \rfloor$ denotes rounding to the nearest integer. The proof given in Section III-D straightforwardly extends to cover these cases, at the expense of somewhat more cumbersome notation.

B. Random Lattice Translates (Dithering)

In [6] and [15] a random lattice translate, or dither, known to both transmitter and receiver was included in the lattice code design. The inclusion of a properly chosen random lattice translate builds upon a construction in [59] and tends to simplify the analysis of MMSE receivers by making the MMSE estimation error independent of the transmitted codeword.

In the setup considered herein we may include such a lattice translate by considering codebooks of the form $\mathcal{X}_r = (\Lambda_r + \boldsymbol{u}) \cap \mathcal{R}$ where $\boldsymbol{u}$ is the random lattice translate, possibly dependent on ρ and r. This construction allows for the inclusion of the "mod-Λ" nested lattice codes considered in [6] and [15]. Note, however, that the specific way in which the mod-Λ construction in [6] maps information messages to codewords, although important from an implementational point of view, is irrelevant to the analysis presented herein as we only consider decoding and not encoding.

The proofs of Theorem 1 and Lemma 1 only need to change in that $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \Lambda_r + \boldsymbol{u}$ replace $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \Lambda_r$ in order to establish DMT optimality of the regularized lattice decoder given by

$$\hat{\boldsymbol{x}}_L = \arg \min_{\hat{\boldsymbol{x}} \in \Lambda_r + \boldsymbol{u}} \|\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}\|^2 + \|\hat{x}\|_{\boldsymbol{T}}^2.$$

In particular, the bound in (19) holds as is, the bound in (23) applies to any $\hat{\boldsymbol{x}} \in \frac{1}{2}\rho^{\frac{\zeta T}{n}} \mathcal{B} \cap (\Lambda_r + \boldsymbol{u})$, and (24) applies to any $\hat{\boldsymbol{x}} \notin \frac{1}{2}\rho^{\frac{\zeta T}{n}} \mathcal{B}$ as before. It follows that regularized lattice decoding is DMT optimal also for designs which include arbitrary chosen random or nonrandom lattice translates. However, it also follows that no such lattice translate is required for DMT optimality. Still, as argued in [6], [15], inclusion of a lattice translate could symmetrize the code, and potentially improve the characteristics of the code at finite SNR.

C. Noise Generalizations

It is valuable to point out that Theorem 1 is only weakly dependent on the nature of the additive noise. In fact, the only parts of the proof that explicitly depend on the Gaussian assumption, is in the lower bound on the pairwise error probability (PEP) in (47) and (49), and where it is concluded that $\mathrm{P}\left(\|\boldsymbol{w}\|^2 \geq \rho^{\delta}\right) \doteq \rho^{-\infty}$ in Section III-D. Thus, for any noise statistics under which $\mathrm{P}\left(\|\boldsymbol{w}\|^2 \geq \rho^{\delta}\right) \doteq \rho^{-\infty}$ and where we may assume a nonzero lower bound on the PEP as in (49), the regularized decoder may be shown to at least match the diversity of the (mismatched) ML decoder in (2), i.e., $d_L(r) \geq d_{\mathrm{ML}}(r)$. In the case of correlated Gaussian noise, the model in (1) is generally directly applicable after absorbing a noise whitening filter into the channel matrix.

The noise generalization also proves useful when the noise component in (1) contains self interference, i.e., $\boldsymbol{w} = \boldsymbol{E}\boldsymbol{x} + \boldsymbol{v}$ for some stochastic $\boldsymbol{E} \in \mathbb{R}^{m \times n}$ and noise $\boldsymbol{v}$. This encompasses the partially coherent scenario when the receiver only knows the channel approximately, in which case $\boldsymbol{E}$ would model the channel estimation error. Under the assumption that $\|\boldsymbol{E}\|_{\mathrm{F}}^2$ is independent of $\boldsymbol{x}$ and ρ, which is typically the case when the channel is estimated using pilots of power proportional to the transmit signal power, and when $\mathrm{P}\left(\|\boldsymbol{E}\|_{\mathrm{F}}^2 \geq \rho^{\delta}\right) \doteq \rho^{-\infty}$ the previous results apply, in spite of the fact that the noise is no longer independent of the transmit signal. In particular, the lower bound of the PEP in (47) and (49) applies straightforwardly by the additive noise alone, and $\mathrm{P}\left(\|\boldsymbol{w}\|^2 \geq \rho^{\delta}\right) \doteq \rho^{-\infty}$ follows by the tail assumption on $\|\boldsymbol{E}\|_{\mathrm{F}}^2$. We also note that the argument in Section III-D does not rely on independence between $\boldsymbol{x}$ and $\boldsymbol{w}$. Thus, the regularized lattice decoder is provably good also in some scenarios involving nonperfect channel state information (CSI) at the receiver.

D. Lower Bounds on the Diversity

Finally, consider an arbitrary, continuous, lower bound on the diversity of the ML decoder, i.e., $d_{\mathrm{ML}}(r) \geq \underline{d}_{\mathrm{ML}}(r)$. It is clear that (17) holds with $\underline{d}_{\mathrm{ML}}(r)$ in place of $d_{\mathrm{ML}}(r)$. Thus, (28) and (29) also holds with $\underline{d}_{\mathrm{ML}}(r)$ in place of $d_{\mathrm{ML}}(r)$ and it follows that $d_L(r) \geq \underline{d}_{\mathrm{ML}}(r)$, i.e., that same lower bound applies to the regularized lattice decoder. Naturally, this observation may be of interest in scenarios where the diversity of the ML decoder is discontinuous and/or not explicitly known.

An important special case is where $d_{\mathrm{ML}}(r) = \infty$ over some open interval of r. The application of a sequence of continuous lower bounds may be used to establish that $d_L(r) = \infty$ over the same interval. Of special interest here is the scenario when lattice decoding of an approximately universal lattice code (e.g., [8] and [15]) is restricted to channels not in outage, in which case it follows that $d_L(r) = d_{\mathrm{ML}}(r) = \infty$. A direct application of this result is given in Section VI-D.

VI. Examples

We proceed by providing a few example scenarios to which the results developed in the previous section are applicable. The examples in Sections VI-A–VI-C are straightforward in the sense that they simply establish a distribution for $\boldsymbol{H}$ in (1), to which Theorems 1, 2 and 3 are directly applicable. The example in Section VI-D is, however, more involved.

A. The Quasi-Static MIMO Channel

The n_{T}-transmit n_{R}-receive antenna quasi-static (flat-fading) MIMO channel commonly given by (cf. [6])

$$\boldsymbol{y}_t^c = \sqrt{\rho}\boldsymbol{H}^c\boldsymbol{x}_t^c + \boldsymbol{w}_t^c, \quad t = 1, \ldots, T \tag{35}$$

where $\boldsymbol{H}^c \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ has some distribution independent of ρ, where $\boldsymbol{x}_t^c \in \mathbb{C}^{n_{\mathrm{T}}}$, $\boldsymbol{y}_t^c \in \mathbb{C}^{n_{\mathrm{R}}}$, and $\boldsymbol{w}_t^c \in \mathbb{C}^{n_{\mathrm{R}}}$, and where t denotes a time index. The channel may be rewritten in the form of (1) where $\boldsymbol{x} = \left[\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_T^{\mathrm{T}}\right]^{\mathrm{T}}$ with

$$\boldsymbol{x}_t^{\mathrm{T}} = \left[\Re\left(\boldsymbol{x}_t^c\right)^{\mathrm{T}}, \Im\left(\boldsymbol{x}_t^c\right)^{\mathrm{T}}\right]$$

and where $\Re(\cdot)$ and $\Im(\cdot)$ denotes the real and imaginary part, respectively, $\boldsymbol{w} = \left[\boldsymbol{w}_1^{\mathrm{T}}, \ldots, \boldsymbol{w}_T^{\mathrm{T}}\right]^{\mathrm{T}}$ with

$$\boldsymbol{w}_t^{\mathrm{T}} = \left[\Re\left(\boldsymbol{w}_t^c\right)^{\mathrm{T}}, \Im\left(\boldsymbol{w}_t^c\right)^{\mathrm{T}}\right]$$

and

$$\boldsymbol{H} = \sqrt{\rho}\boldsymbol{I} \otimes \begin{bmatrix} \Re(\boldsymbol{H}^c) & -\Im(\boldsymbol{H}^c) \\ \Im(\boldsymbol{H}^c) & \Re(\boldsymbol{H}^c) \end{bmatrix}. \tag{36}$$

The channel in (35) is also often written in an equivalent matrix form

$$\boldsymbol{Y}^c = \sqrt{\rho}\boldsymbol{H}^c\boldsymbol{X}^c + \boldsymbol{W}^c \tag{37}$$

where $\boldsymbol{X}^c = [\boldsymbol{x}_1^c, \ldots, \boldsymbol{x}_T^c]$ and $\boldsymbol{W}^c = [\boldsymbol{w}_1^c, \ldots, \boldsymbol{w}_T^c]$. Under the short-term average input power constraint

$$\frac{1}{|\mathcal{X}|}\sum_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{X}^c\|_{\mathrm{F}}^2 = \frac{1}{|\mathcal{X}|}\sum_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{x}\|^2 \leq T \tag{38}$$

and an appropriate scaling of $\boldsymbol{H}^c$, the parameter ρ takes on the interpretation of an average SNR per receive antenna (cf. [5] and [6]).

B. The Parallel MIMO Channel (MIMO-OFDM)

A natural extension of the quasi-static MIMO channel is the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ parallel, or MIMO-OFDM, channel. In this setting

$$\boldsymbol{Y}_l^c = \sqrt{\rho}\boldsymbol{H}_l^c\boldsymbol{X}_l^c + \boldsymbol{W}_l^c, \quad l = 1, \ldots, L \tag{39}$$

where $\boldsymbol{X}_l^c = \left[\boldsymbol{x}_{l,1}^c, \ldots, \boldsymbol{x}_{l,T}^c\right] \in \mathbb{C}^{n_{\mathrm{T}} \times T}$ denotes the complex space-time block codeword transmitted over the lth sub-channel in the T time-slots, and where $\boldsymbol{H}_l^c \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ is the channel matrix for the lth sub-channel. Similar to the flat fading quasi-static channel, it is clear by the linearity of (39) that the parallel channel can be rewritten according to (1). Coding across the parallel channels is achieved by the appropriate choice of generator matrix $\boldsymbol{G}$. For the rate definition it is conventional to consider one use of (39) as LT channel uses.

Naturally, the DMT characteristics of the parallel channel depend of the statistics of $[\boldsymbol{H}_1^c, \ldots, \boldsymbol{H}_L^c]$. In the particular case where $\boldsymbol{H}_l^c$ for $l = 1, \ldots, L$ represent the OFDM tones for a Q-tap i.i.d. Rayleigh fading channel, i.e.,

$$\boldsymbol{H}_l^c = \sum_{q=0}^{Q-1} \tilde{\boldsymbol{H}}_q^c e^{-i2\pi q \frac{l-1}{L}}, \quad l = 1, \ldots, L \tag{40}$$

where $\tilde{\boldsymbol{H}}_q^c \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$, $q = 0, \ldots, Q-1$, are stochastically independent i.i.d. Rayleigh fading taps in the time domain, the maximal diversity gain is $f_Q(r)$ where $f_Q(r)$ is given by the piecewise linear curve connecting $(k, (Q\overline{n} - k)(\underline{n} - k))$ for $k = 1, \ldots, \underline{n}$ where $\overline{n} = \max(n_{\mathrm{R}}, n_{\mathrm{T}})$ and $\underline{n} = \min(n_{\mathrm{R}}, n_{\mathrm{T}})$, respectively, [60]. Lattice designs that achieve the optimal DMT of the setting, under ML decoding, were given in [19], [61] for particular values of n_{T} and L and in [62] for the general case of n_{T}, L. We conclude that low complexity and DMT optimal decoding of these codes, when transmitted over the channel in (39) and (40), is feasible. A generalization of the scenario described above to more general selective-fading MIMO channels is found in [63].

C. The AF Relay Channel

Over the AF relay channel, one or several relays amplify and retransmit the signal received in previous time-slots, in order to aid the transmission of data from a source to destination. An initial, orthogonal, version of this scenario was in the DMT context studied in [64]. As an example, we here consider another AF protocol, namely the single-antenna single relay nonorthogonal amplify and forward (NAF) protocol proposed in [65], operating over a quasi-static channel. We omit constant transmit power scaling factors for brevity. One transmission from the source followed by a joint source relay transmission may be modeled according to (cf. [66])

$$\boldsymbol{y}_t^c = \begin{bmatrix} \sqrt{\rho} h_1^c & 0 \\ \rho b h_2^c h_3^c & \sqrt{\rho} h_1^c \end{bmatrix} \boldsymbol{x}_t^c + \begin{bmatrix} 0 \\ \sqrt{\rho} b h_3^c \end{bmatrix} w_t^c + \boldsymbol{v}_t^c \tag{41}$$

where h_1^c, h_2^c and h_3^c are the complex gains from source to destination, source to relay, and relay to destination, respectively. The term w_t^c represents the receiver noise at the relay and $\boldsymbol{v}_t^c$ the noise at the destination. The relay amplification b is in general allowed to depend on ρ and h_2 and must satisfy

$$|b|^2 \leq \frac{1}{\rho |h_2|^2 + 1} \tag{42}$$

in order to meet the relay transmit power constraint. After noise whitening (41) becomes equivalent to (35) with

$$\boldsymbol{H}^c = \begin{bmatrix} h_1^c & 0 \\ \frac{\sqrt{\rho} b h_2^c h_3^c}{\sqrt{\rho |b h_3^c|^2 + 1}} & \frac{h_1^c}{\sqrt{\rho |b h_3^c|^2 + 1}} \end{bmatrix} \tag{43}$$

where one transmission over (43) corresponds to two channel uses in the definition of the rate. As argued in [18], any approximately universal code designed for the 2×2 quasi-static MIMO channel is able to achieve a diversity gain of $d_{\mathrm{ML}}(r) = (1 - r) + (1 - 2r)^+$, under AWGN noise and i.i.d. Rayleigh fading assumptions, provided b is properly selected. This also corresponds to the maximal diversity over the class of linear AF protocols [66]. We see that the AF protocol defines a (somewhat complicated) set of channel statistics, parameterized by ρ. It follows directly by the continuity of $d_{\mathrm{ML}}(r)$ that $d_L(r) = d_{\mathrm{ML}}(r)$ over $r \in [0, 1]$.

There are several generalizations of AF protocols to more relays and different relay actions, [20], [66], [67]. General to this setting is that the particular AF protocol determines the statistics of the equivalent channel in (1), similar to (43). Lattice designs for some of these generalizations are found in [18], [20]. The application of Theorem 1, 2, and 3 is straightforward to most, if not all, lattice designs in these settings, once the AF protocol is established. Note that the lattice designs in [18], [20] provide for approximate universality over these system models, and as a result, $d_L(r)$ is optimal in these settings. Note also that even for scenarios where $d_{\mathrm{ML}}(r)$ is not known, it follows from the discussion in Section V-D that any continuous lower bound on $d_{\mathrm{ML}}(r)$ applies also to $d_L(r)$.

D. The L-round MIMO-ARQ Channel

Consider the L-round MIMO ARQ setting where, as in [68], signaling of the information across the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ quasi-static MIMO channel uses an L-round automatic retransmission request (ARQ) protocol that assumes the presence of a noiseless feedback channel conveying one bit of information per use of the feedback channel. During the lth round, an $n_{\mathrm{T}} \times T$ code-matrix $\boldsymbol{X}_l$ is transmitted where $[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_L] \in \mathcal{X} \subset \mathbb{C}^{n_{\mathrm{T}} \times LT}$, and a decoder D_l is applied to decode the fragment $[\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_l]$ [cf. (37) and (39)] corresponding to the fragmented code $[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_l] \in \mathbb{R}^{n_{\mathrm{T}} \times lT}$ with multiplexing gain $r_l = r_1 / l$. The decoder D_l either generates an acknowledgment (ACK) in which case a hard decision is made and the transmission of that message terminates, or generates a negative acknowledgment (NACK) in which case another transmission round is requested. The last decoder D_L always tries to decode the message. An error is considered only when a message is decoded erroneously. The DMT characteristics of the MIMO-ARQ channel were first considered in [68] where also the optimal DMT was obtained under two different fading models. We shall for sake of brevity only consider long-term fading where the channel $\boldsymbol{H}^c$ remains constant over the L-rounds. We show in what follows how the results obtained herein can be applied to prove DMT optimality of lattice coding and LR-aided linear decoding for the MIMO-ARQ channel, for all n_{R}, n_{T}, L and fading statistics.

To this end, let $\bar{\mathcal{A}}_1$ denote the event that a NACK is requested in the first round, and let $r_{\max} = \sup\{r | d_{\mathrm{out}}(r) > 0\}$ where $d_{\mathrm{out}}(r)$ denotes the optimal DMT for $L = 1$, i.e., in the absence of feedback. We assume that $d_{\mathrm{out}}(r)$ is continuous over $r \in [0, r_{\max})$. As in [21] we consider in parallel a fictitious system where $[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_L]$ is transmitted and where each of the decoders $\mathcal{D}_l$ operates independently on each of the fragments $[\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_l]$, $l = 1, \ldots, L$. Let $P_{e,l}(r_l)$ denote the probability of error of $\mathcal{D}_l$ in the fictitious system, and let $P_e(r)$ denote the overall probability of error at the expected or average multiplexing gain r. The work in [21] provides, based on the work in [68], the following sufficient conditions for overall DMT optimality in the MIMO-ARQ setting.

1) Equation $\mathrm{P}(\bar{\mathcal{A}}_1) \doteq \rho^{-\epsilon}$, $\epsilon > 0$
2) Equation $P_{e,l}(r_l) \,\dot{\leq}\, P_{e,L}(r_L)$, $l = 1, \ldots, L - 1$.
3) Equation $P_{e,L}(r_L) \doteq \rho^{-d_{\mathrm{out}}(r_L)}$

In brief, optimality follows from the above by observing that

$$P_{e,L}(r_L) \leq P_e(r) \leq \sum_{l=1}^{L} P_{e,l}(r_l)$$

which by the second condition implies that $R_e(r) \doteq P_{e,L}(r_L)$. Based on the first condition it may be shown that $r = r_1$ (cf. [21]) and by the third condition it follows that

$$P_e(r) \doteq \rho^{-d_{\mathrm{out}}(r_L)} = \rho^{-d_{\mathrm{out}}\left(\frac{r}{L}\right)}$$

which corresponds to the maximal ARQ diversity [68]. The reader is referred to [21] for a detailed analysis.

Now, let each $\mathcal{D}_l$ apply regularized lattice decoding, and an ACK-NACK policy similar to [21], [68] where an ACK is generated if and only if

$$\log\det(\boldsymbol{I}+\rho\boldsymbol{H}^c(\boldsymbol{H}^c)^{\mathrm{H}}) \geq \frac{x}{l}\log\rho > r_l\log\rho \qquad (44)$$

for some x such that $r_1 < x < r_{\max}$. This ACK-NACK policy is independent[5] of $r = r_1$, provided $r_1 < x$. Consider now the application of a code where each fragment code is approximately universal. Explicit lattice codes of this type are provided in [22]. As (44) implies that the decoders for $l = 1, \ldots, L-1$ are only applied to channels not in outage it follows, as explained in Section V-D, by the approximate universality of the fragment codes that $P_{e,l}(r_l) \doteq \rho^{-\infty}$, for $l = 1, \ldots, L-1$. For $l = L$ it follows directly by Theorem 1 and 2 that $P_{e,L}(r_L) \doteq \rho^{-d_{\mathrm{out}}(r_L)}$. Regarding $\bar{\mathcal{A}}_1$ it follows by (44) that $\mathrm{P}(\bar{\mathcal{A}}_1) \doteq \rho^{-d_{\mathrm{out}}(x)}$ where $d_{\mathrm{out}}(x) > 0$ as $x < r_{\max}$, establishing the DMT optimality of the regularized lattice decoder for $r \in [0, r_{\max})$ when applied to the codes proposed in [22].

We remark that the DMT optimality of lattice coding and decoding for the MIMO-ARQ channel was in fact proven already in [68], albeit under the assumption of i.i.d. Rayleigh fading and $T \geq n_{\mathrm{R}} + n_{\mathrm{T}} - 1$, using a random construction similar to [6]. The argument presented above extends this result to LR-aided linear decoding, the minimum delay setting ($T = n_{\mathrm{T}}$) and more general fading statistics.

E. Further Examples and Lattice Designs

The examples given above only constitute a subset of the scenarios to which the main results presented herein are applicable. For instance, ISI channels and generally selective fading channels [63] may be handled similarly to the parallel channel in Section VI-B. The finite rate feedback scenarios and long term power allocation policies considered in [69] are handled similarly to the MIMO-ARQ channel in Section VI-D. Dynamic decode-and-forward (DDF) protocols, where relays decode and forward a received message whenever the relevant channels are not in outage, are also handled similarly to the MIMO-ARQ channel. The results extend to cover orthogonal amplify and forward (OAF) [64] as well as orthogonal and nonorthogonal selection decode and forward (OSDF and NSDF) relay protocols [20]. Approximately universal distributed codes exist for several such cooperative protocols and scenarios, such as for example in [20] for the OSDF and NSDF protocols, and as a result the regularized lattice decoders and their LR-aided linear counterparts achieves the corresponding approximate universality in these settings as well.

Table I identifies the lattice dimensionality employed by DMT optimal implementations for different channels, as well as refers the reader to explicit descriptions of the designs[6]. The potentially very large lattice dimensions faced when decoding such designs makes reduced complexity decoders essential to the successful deployment of these designs.

TABLE I
LATTICE DIMENSIONALITY AND REFERENCES FOR EXPLICIT TRANSCEIVERS IN DIFFERENT SETTINGS

Channel	n	Lattice source
$m \times m$ MIMO	$2m^2$	[8], [13], [15]
$m \times m$, L-tone MIMO-OFDM	$2m^2L$	[19], [20], [61]
$m \times m$, m-round MIMO-ARQ	$2m^2$	[21]
$m \times m$, L-round MIMO-ARQ (AU)	$2m^2L$	[21]
m-relay OAF	$2m$	[20]
2-relay OSDF, NSDF ($r = 2$)	$32, 162$	[20]
m-relay NAF	$8(m-1)$ $8(m-1)^2$	[18] [20]
m-relay DDF, L-slots, $m > 2$	$2m^2L$	[22]

VII. Conclusion

The work presented an explicit characterization of efficient encoder-decoder structures that meet the fundamental DMT performance limits, and do so for very general channel statistics, dimensions, and models. Specifically, it proved that regularized lattice decoders in general, and the MMSE-GDFE decoder in particular, provide DMT optimal decoding in its most general form, irrespective of the particular code applied. It also established, for the first time, that computationally efficient LR-aided linear decoders are capable of achieving the entire DMT. The generality of the results obtained lends them applicable to a plethora of pertinent communication scenarios which inherently introduce nonstandard channel statistics, code-structure limitations and prohibitively high ML-decoding complexity.

In terms of information theoretic guarantees on error probability performance, the work extended prior state-of-art to a very general setting. In terms of implementability, the work covered the gap that exists, between the point of proving the existence of non-ML optimal transceivers, and the point of establishing what these transceivers are and how they can be efficiently applied. In terms of complexity guarantees, the work provides worst-case guarantees on the complexity required for DMT optimality. This is done despite the fact that the employed algorithms are generally known to have unbounded worst-case complexity.

In terms of generality over codes, dimensions and channel statistics, we observe the following: Generality with respect to the codes addresses issues of legacy, and guarantees that the efficient regularized decoder structure will maintain, in most circumstances, the ML decoder DMT performance of the existing code structure. The generality thus also applies to communication scenarios which place restrictions on the form of the codes applied.

Generality with respect to channel dimensions is pertinent to computationally demanding scenarios that involve encoding over a large number of degrees of freedom, such as multitoned OFDM, multitap ISI, as well as multiround MIMO-ARQ and multislot DDF channels. In all the above, error probability performance gains require an increasing number of rounds/slots, which in turn result in linear increases in the problem dimensionality and exponential increases in the ML decoding complexity. The same generality with respect to dimension bypasses

[5] This is a technical requirement for the application of Theorem 1 that stems from the fact that we assume the statistics of $\boldsymbol{H}$ in (1) to be independent of the multiplexing gain of the code applied. Note, however, that the independence is only required in a neighborhood of the target multiplexing gain.

[6] In the case of OAF and m-round MIMO-ARQ, DMT optimality is limited to a class of channels. All relay channels consider single-antenna nodes.

issues of channel asymmetry, as well as allows for a unified exposition of the problem.

Finally, generality with respect to fading statistics maintains the pertinent asymptotic guarantees to cases where the underlying fading and noise statistics are not entirely known, specifically to scenarios which inherently introduce hard to characterize channels such as different cooperative relaying protocols, as well as MIMO-OFDM and time-varying channels with arbitrary correlations.

In terms of practicality, the presented transceivers allow for a broad spectrum of rate-reliability-complexity guarantees that result in near-optimal transmission energy, and reduced algorithmic power consumption and delay. Under the requirement for nonexponentially complex decoders, the work also allows for these rate-reliability guarantees in the presence of reduced hardware complexity, such as for example with a minimum number of transmit and receive antennas. Furthermore, the efficient and universal applicability of the transceivers over different system models, allows for further diversification of resources over hybrid channels that near-optimally induce further gains in performance. In terms of future work, the results naturally motivate further joint study into new approximation algorithms and code designs that together yield improved approximation ratios, and better performance in the nonasymptotic regime.

Appendix A
Equivalence of the MMSE-GDFE and the Regularized Lattice Decoder

By "completion of squares" the regularized metric in (11) may be written according to

$$\begin{aligned}\|\boldsymbol{y}-\boldsymbol{H}\hat{\boldsymbol{x}}\|^2+\|\hat{\boldsymbol{x}}\|_{\boldsymbol{T}}^2 &= \hat{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\hat{\boldsymbol{x}}-2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}\hat{\boldsymbol{x}}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\\ &\quad+\hat{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{T}\hat{\boldsymbol{x}}\\ &= \hat{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\hat{\boldsymbol{x}}-2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{B}\hat{\boldsymbol{x}}\\ &\quad+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{y}+\Gamma\\ &= \|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}\|^2+\Gamma \end{aligned} \quad (45)$$

where $\boldsymbol{B}$ is any matrix for which $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+\boldsymbol{T})$, where $\boldsymbol{F}=\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}$ and where

$$\Gamma=\boldsymbol{y}^{\mathrm{T}}[\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}+\boldsymbol{T})^{-1}\boldsymbol{H}^{\mathrm{T}}]\boldsymbol{y}\geq 0.$$

As Γ does not depend on $\hat{\boldsymbol{x}}$ it may be disregarded in the optimization over $\hat{\boldsymbol{x}}$, i.e., the regularized lattice decoder may be alternatively expressed as

$$\hat{\boldsymbol{x}}_L=\arg\min_{\hat{\boldsymbol{x}}\in\Lambda_r}\|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}\|^2. \quad (46)$$

Comparing $\boldsymbol{B}$, $\boldsymbol{F}$, and (46) or (14) to the corresponding expressions in [6], establishes the equivalence of the regularized decoder and the MMSE-GDFE decoder when $\boldsymbol{T}=\boldsymbol{I}$.

Further, if $\hat{\boldsymbol{x}}_{\mathrm{A}}$ is a C-approximate solution to (46), i.e., if

$$C\|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}_L\|^2\geq\|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}_{\mathrm{A}}\|^2$$

for $C\geq 1$, it follows that

$$C(\|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}_L\|^2+\Gamma)\geq\|\boldsymbol{F}\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{x}}_{\mathrm{A}}\|^2+\Gamma$$

which by (45) implies that $\hat{\boldsymbol{x}}_{\mathrm{A}}$ is also a C-approximate solution to (11).

Appendix B
Proof of Lemma 1

The probability that any particular $\hat{\boldsymbol{x}}\neq\boldsymbol{x}$, for fixed $\boldsymbol{H}$ and $\boldsymbol{x}$, achieves a lower ML metric that the transmitted codeword $\boldsymbol{x}$ is given by the conditional pairwise error probability according to [29, Appendix A.2]

$$\begin{aligned}\mathrm{P}(\boldsymbol{x}\to\hat{\boldsymbol{x}}|\boldsymbol{x},\boldsymbol{H}) &\triangleq \mathrm{P}(\|\boldsymbol{y}-\boldsymbol{H}\hat{\boldsymbol{x}}\|^2<\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2|\boldsymbol{x},\boldsymbol{H})\\ &= Q\left(\frac{1}{2}\|\boldsymbol{H}(\hat{\boldsymbol{x}}-\boldsymbol{x})\|\right)\end{aligned} \quad (47)$$

where $Q(\cdot)$ is the Q-function [29]. The conditional probability of an ML error satisfies

$$\mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}}\neq\boldsymbol{x}|\boldsymbol{x},\boldsymbol{H})\geq\mathrm{P}(\boldsymbol{x}\to\hat{\boldsymbol{x}}|\boldsymbol{x},\boldsymbol{H}) \quad (48)$$

for arbitrary $\hat{\boldsymbol{x}}\in\mathcal{X}_r$, $\hat{\boldsymbol{x}}\neq\boldsymbol{x}$, as any pairwise error event implies that $\boldsymbol{x}$ does not attain the lowest ML metric and is thus not selected by the ML decoder.

Let $\boldsymbol{d}=\boldsymbol{d}(\boldsymbol{H})$ be the minimizer of (16), and let $\hat{\boldsymbol{x}}=\hat{\boldsymbol{x}}(\boldsymbol{H},\boldsymbol{x})=\boldsymbol{x}+\boldsymbol{d}(\boldsymbol{H})$. Thus, for any $\boldsymbol{H}$ such that $\nu_r=\frac{1}{4}\|\boldsymbol{H}\boldsymbol{d}\|^2\leq 1$ it follows by (47) that

$$\mathrm{P}(\boldsymbol{x}\to\hat{\boldsymbol{x}}|\boldsymbol{x},\boldsymbol{H})\geq Q(1)>0 \quad (49)$$

as the Q-function is decreasing in its augment. Further, as $\boldsymbol{d}\neq\boldsymbol{0}$ and $\boldsymbol{d}\in\mathcal{B}\cap\Lambda_r$ by definition [cf. (16)], it follows that $\hat{\boldsymbol{x}}\neq\boldsymbol{x}$ and $\hat{\boldsymbol{x}}\in\mathcal{X}_r$ (i.e., $\hat{\boldsymbol{x}}$ is a valid codeword) whenever $\boldsymbol{x}\in\mathcal{B}$. Consequently

$$\mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}}\neq\boldsymbol{x}|\boldsymbol{x}\in\mathcal{B},\nu_r\leq 1)\geq Q(1) \quad (50)$$

by (48) and (49) and the above discussion. Further

$$\begin{aligned}\mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}}\neq\boldsymbol{x}) &\geq \mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}}\neq\boldsymbol{x}|\boldsymbol{x}\in\mathcal{B},\nu_r\leq 1)\\ &\quad\cdot\mathrm{P}(\boldsymbol{x}\in\mathcal{B},\nu_r\leq 1)\\ &\geq Q(1)\mathrm{P}(\boldsymbol{x}\in\mathcal{B})\mathrm{P}(\nu_r\leq 1).\end{aligned} \quad (51)$$

where the last inequality follows from (50) and the independence of $\boldsymbol{x}\in\mathcal{B}$ and $\nu_r\leq 1$.

By applying the same approximation as in (9) it may, provided $r>0$, be shown (cf. [47]) that

$$\lim_{\rho\to\infty}\mathrm{P}(\boldsymbol{x}\in\mathcal{B})=\frac{V(\mathcal{B})}{V(\mathcal{R})}>0$$

when $\boldsymbol{x}$ is uniformly distributed over $\mathcal{X}_r=\mathcal{R}\cap\Lambda_r$. This implies that $\mathrm{P}(\boldsymbol{x}\in\mathcal{B})\doteq\rho^0$. It therefore follows from (51) that

$$\mathrm{P}(\hat{\boldsymbol{x}}_{\mathrm{ML}}\neq\boldsymbol{x})\,\dot{\geq}\,\mathrm{P}(\nu_r\leq 1)$$

which is equivalent to (17). □

APPENDIX C PROOF OF LEMMA 2

Assume that

$$\mathrm{P}\left(\|\boldsymbol{H}\|_{\mathrm{F}}^2 \geq \rho^x\right) \geq \rho^{-d_{\mathrm{ML}}(r)}$$

for sufficiently large ρ. It then follows that

$$\mathrm{E}\left\{\|\boldsymbol{H}\|_{\mathrm{F}}^2\right\} \geq \rho^{x-d_{\mathrm{ML}}(r)}. \tag{52}$$

Thus, if $\mathrm{E}\left\{\|\boldsymbol{H}\|_{\mathrm{F}}^2\right\} \dot{\leq} \rho$ it holds that

$$\mathrm{P}\left(\|\boldsymbol{H}\|_{\mathrm{F}}^2 \geq \rho^x\right) \dot{\leq} \rho^{-d_{\mathrm{ML}}(r)} \tag{53}$$

for any $x > d_{\mathrm{ML}}(r) + 1$. Let $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{G}$ where $\boldsymbol{G}$ is the code lattice generator and $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} + \boldsymbol{T}$ (cf. Appendix A). It holds that

$$\kappa^2(\boldsymbol{M}) = \frac{\lambda_{\max}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})}{\lambda_{\min}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})}$$

where $\lambda_{\max}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})$ and $\lambda_{\min}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})$ denotes the largest and smallest eigenvalues of $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$. Note that $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M} = \boldsymbol{G}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{G}$. As $\lambda_{\min}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}) \geq \lambda_{\min}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{G}) > 0$ and $\lambda_{\max}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}) \leq \lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{G}) + \lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{G})$ where $\lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{G}) \leq \lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})\|\boldsymbol{H}\|_{\mathrm{F}}^2$ it follows that

$$\kappa^2(\boldsymbol{M}) \leq \frac{\lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})\|\boldsymbol{H}\|_{\mathrm{F}}^2 + \lambda_{\max}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{G})}{\lambda_{\min}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{G})}. \tag{54}$$

For $\alpha > \frac{1}{2}x$ it follows by (54) that for sufficiently large ρ

$$\|\boldsymbol{H}\|_{\mathrm{F}}^2 \leq \rho^x \quad \Rightarrow \quad \kappa(\boldsymbol{M}) \leq \rho^{\alpha}.$$

Thus, by (53) it follows that

$$\mathrm{P}(\kappa(\boldsymbol{M}) \geq \rho^{\alpha}) \dot{\leq} \rho^{-d_{\mathrm{ML}}(r)}$$

for any $\alpha > \frac{1}{2}(d_{\mathrm{ML}}(r) + 1)$. □

REFERENCES

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, Jul. 1999.

[3] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp.*, Pisa, Italy, 1998.

[4] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1804–1824, Jul. 2002.

[5] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, May 2003.

[6] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 968–985, Jun. 2004.

[7] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space-time codes that achieve the diversity-multiplexing gain tradeoff," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, 2005.

[8] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3869–3884, Sep. 2006.

[9] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full-diversity, high-rate, space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2596–2616, Oct. 2003.

[10] J.-C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. IEEE Inf. Theory Worshop (ITW)*, Paris, France, Mar. 2003.

[11] T. Kiran and B. S. Rajan, "STBC-schemes with non-vanishing determinant for certain number of transmit antennas," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2984–2992, Aug. 2005.

[12] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3233–3258, Jul. 2006.

[13] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3885–3902, Sep. 2006.

[14] P. Elia, B. A. Sethuraman, and P. V. Kumar, "Perfect space-time codes for any number of transmit antennas," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3853–3868, Nov. 2007.

[15] K. R. Kumar and G. Caire, "Space-time codes from structured lattices," *IEEE Trans. Inf. Theory*, vol. 55, pp. 547–556, Feb. 2009.

[16] C. Hollanti, J. Lahtonen, K. Ranto, and R. Vehkalahti, "On the densest MIMO lattices from cyclic division algebras," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3751–3780, Aug. 2009.

[17] M. O. Damen, A. Tewfik, and J.-C. Belfiore, "A construction of a space-time code based on number theory," *IEEE Trans. Inf. Theory*, vol. 48, pp. 753–760, Mar. 2002.

[18] S. Yang and J.-C. Belfiore, "Optimal space-time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 647–663, Feb. 2007.

[19] H.-F. Lu, "Constructions of multiblock space-time coding schemes that achieve the diversity multiplexing tradeoff," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3790–3796, Aug. 2008.

[20] P. Elia, K. Vinodh, M. Anand, and P. V. Kumar, "D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols," *IEEE Trans. Inf. Theory*, vol. 55, Jul. 2009.

[21] S. A. Pawar, K. R. Kumar, P. Elia, P. V. Kumar, and B. A. Sethuraman, "Space-time codes achieving the DMD tradeoff of the MIMO-ARQ channel," *IEEE Trans. Inf. Theory*, vol. 55, Jul. 2009.

[22] P. Elia and P. V. Kumar, "Space-time codes that are approximately universal for the parallel, multiblock and cooperative DDF channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, 2009.

[23] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2201–2214, Aug. 2002.

[24] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2389–2401, Oct. 2003.

[25] A. D. Murugan, H. E. Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, pp. 933–953, Mar. 2006.

[26] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bölcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE J. Solid-State Circuits*, vol. 40, pp. 1566–1577, Jul. 2005.

[27] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, pp. 1474–1484, Apr. 2005.

[28] J. Jaldén and B. Ottersten, "On the limits of sphere decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, Sep. 2005.

[29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge , U.K.: Cambridge Univ. Press, 2005.

[30] K. K. Raj, G. Caire, and A. L. Moustakas, "The diversity-multiplexing tradeoff of linear MIMO receivers," in *Proc. IEEE Inf. Theory Worshop (ITW)*, Lake Tahoe, NV, Sep. 2007, pp. 487–492.

[31] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Conf. Commun. (GLOBECOM)*, Taipei, Taiwan, Nov. 2002.

[32] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Inf. Theory Worshop (ITW)*, Paris, France, Mar. 2003.

[33] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4801–4805, Dec. 2007.

[34] M. Taherzadeh and A. K. Khandani, "On the limitations of the naive lattice decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007.

[35] M. O. Damen, H. E. Gamal, and G. Caire, "MMSE-GDFE lattice decoding for underdetermined linear channels," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, 2004.

[36] D. Wübben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice reduction," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, Jun. 2004.

[37] S. Karmakar and B. S. Rajan, "Multigroup-decodable STBCs from Clifford algebra," *IEEE Trans. Inf. Theory*, vol. 55, pp. 223–231, Jan. 2009.

[38] C. Hollanti and K. Ranto, "Asymmetric space-time block codes for MIMO systems," in *Proc. IEEE Inf. Theory Workshop on Inf. Theory for Wireless Netw.*, Bergen, Norway, Jul. 2007.

[39] E. Biglieri, Y. Hong, and E. Viterbo, "On fast-decodable space-time block codes," *IEEE Trans. Inf. Theory*, vol. 55, pp. 524–530, Feb. 2009.

[40] O. Tirkkonen and R. Kashaev, "Combined information and performance optimization of linear MIMO modulations," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Lausanne, Switzerland, Jun. 2002.

[41] J. Paredes, A. B. Gershman, and M. Gharavi-Alkhansari, "A 2×2 space-time code with non-vanishing determinant and fast maximum likelihood decoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007.

[42] M. Samuel and M. P. Fitz, "Reducing the detection complexity by using 2×2 multi-strata space-time codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007.

[43] A. Medles and D. T. M. Slock, "Achieving the optimal diversity vs multiplexing tradeoff for MIMO flat channels with QAM space-time spreading and DFE equalization," *IEEE Trans. Inf. Theory*, vol. 52, Dec. 2006.

[44] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1212–1215, Mar. 2001.

[45] J. Jaldén, D. Seethaler, and G. Matz, "Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008.

[46] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1547–1557, Apr. 2007.

[47] U. Betke and K. Böröczky, Jr, "Asymptotic formulae for the lattice point enumerator," *Canad. J. Math.*, vol. 51, no. 2, pp. 225–249, 1999.

[48] G. D. Forney, Jr, R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, pp. 632–647, Sep. 1984.

[49] C. Studer, D. Seethaler, and H. Bölcskei, "Finite lattice-size effects in MIMO detection," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2008.

[50] A. N. Tikhonov, A. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-posed Problems*, 1st ed. New York: Springer, 1995.

[51] J. Hromkovič, *Algorithms for Hard Problems: Introduction to Combinatorial Optimization, Randomization, Approximation and Heuristics*, 2nd ed. New York: Springer, 2002.

[52] L. Babai, "On lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, Mar. 1986.

[53] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Matematische Annalen*, vol. 261, no. 4, pp. 1432–1807, Dec. 1982.

[54] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, pp. 181–191, 1994.

[55] H. Daudée and B. Vallée, "An upper bound on the average number of iterations of the LLL algorithm," *Theoret. Comput. Sci.*, vol. 123, no. 1, Jan. 1994.

[56] C. Ling and H. Howgrave-Graham, "Effective LLL reduction for lattice decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007.

[57] L. Luzzi, G. R.-B. Othman, and J.-C. Belfiore, "Algebraic reduction for space-time codes based on quaternion algebras," in *IEEE Trans. Inf. Theory*, 2008, available on arXiv:0809.3365v2 (cs.IT).

[58] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2×2 full-rate space-time code with non-vanishing determinants," *IEEE Trans. Inf. . Theory*, vol. 51, Apr. 2005.

[59] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1 + \mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, Oct. 2004.

[60] A. Medles and D. T. M. Slock, "Optimal diversity vs. multiplexing treadeoff for frequency selective MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, Sep. 2005.

[61] S. Yang, J.-C. Belfiore, and G. R.-B. Othman, "Perfect space-time block codes for parallel MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, Jul. 2006.

[62] P. Elia and P. K. Vijay, "Approximately-universal space-time codes for the parallel, multi-block and cooperative-dynamic-decode-and-forward channels," 2007 [Online]. Available: arXiv:0706.3502v2 [cs.IT]

[63] P. Coronel and H. Bölcskei, "Diversity-multiplexing tradeoff in selective-fading MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007.

[64] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[65] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1099–1109, Aug. 2004.

[66] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4152–4172, Dec. 2005.

[67] S. Yang and J.-C. Belfiore, "Towards the optimal amplify-and-forward cooperative diversity scheme," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3114–3126, Sep. 2007.

[68] H. E. Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3601–3621, Aug. 2006.

[69] T. T. Kim and M. Skoglund, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2743–2759, Aug. 2007.

Joakim Jaldén (S'03–M'08) received the M.Sc. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002 and 2007, respectively.

From July 2007 to June 2009, he held a Postdoctoral research position with the Vienna University of Technology, Vienna, Austria. He also studied at Stanford University, CA, from September 2000 to May 2002, and worked at ETH, Zürich, Switzerland, as a Visiting Researcher, from August to September 2008. In July 2009, he joined the Signal Processing Lab within the School of Electrical Engineering, KTH, Stockholm, as an Assistant Professor.

Dr. Jaldén has been awarded the IEEE Signal Processing Society's 2006 Young Author Best Paper Award for his work on MIMO communications, and the first prize in the Student Paper Contest at the 2007 International Conference on Acoustics, Speech and Signal Processing (ICASSP). He is also a recipient of the Ingvar Carlsson Award issued in 2009 by the Swedish Foundation for Strategic Research

Petros Elia received the B.Sc. degree from the Illinois Institute of Technology, Chicago, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 2006.

Since February 2008, he has been an Assistant Professor with the Department of Mobile Communications, EURECOM, Sophia Antipolis, France. He has been the recipient of the Fulbright scholarship, and the corecipient of the USC-EE Best Student Paper Award. He has also been a Visiting Scholar with the Indian Institute of Science in 2004. His research interests include information theoretic and coding theoretic aspects of wireless communications, cooperative communications, MIMO transceiver design-performance-complexity, dense network behavior, queueing theory for cross-layer aspects, and soft-biometrics.