Vanishing the gap to exact lattice search at a subexponential complexity: LR-aided regularized decoding

Arun Singh and Petros Elia
Mobile Communications Department
EURECOM, Sophia Antipolis, France
Email: {singhak, elia}@eurecom.fr

Joakim Jaldén
ACCESS Linnaeus Center, Signal Processing Lab
KTH Royal Institute of Technology, Stockholm, Sweden
Email: jalden@kth.se

***Abstract*— This work identifies the first lattice decoding solution that achieves, in the most general outage-limited MIMO setting and the high rate and high SNR limit, both a vanishing gap to the error-performance of the exact solution of regularized lattice decoding, as well as a computational complexity that is subexponential in the number of codeword bits as well as the rate. The proposed solution employs lattice reduction (LR)-aided regularized (lattice) sphere decoding and proper timeout policies.**

In light of the fact that, prior to this work, a vanishing gap was generally attributed only to full lattice searches that have exponential complexity, in conjunction with the fact that subexponential complexity was generally attributed to early-terminated (linear) solutions which have though a performance gap that can be up to exponential in dimension and/or rate, the work constitutes the first proof that subexponential complexity need not come at the cost of exponential reductions in lattice decoding error performance. Finally, these performance and complexity guarantees hold for the most general MIMO setting, for all reasonable fading statistics, all channel dimensions, and all lattice codes.

I. INTRODUCTION

The work applies to the general setting of outage limited MIMO communications, where MIMO techniques offer significant advantages in terms of increased throughput and reliability, although at the price of substantially higher computational complexity for decoding at the receivers. This complexity brings to the fore the need for efficient decoders that nicely tradeoff performance with complexity.

In terms of ML-based decoding, the use of the brute-force ML decoder, introduces a complexity that scales exponentially with the channel dimension, time, and with the total number of codeword bits. If on the other hand, a small gap to the exact ML performance is acceptable, then different branch-and-bound algorithms such as the sphere decoder (SD) have been known to require reduced computational resources. These decoding solutions require that the underlying code be linear, an assumption that we adopt here. Despite the reduced complexity of sphere decoding of such linear codes, recent work in [1] has revealed that, to achieve vanishing gap to ML solutions, even such branch and bound algorithms generally require computational resources that, albeit significantly smaller than exhaustive ML decoder in many scenarios of practical interest, again grow exponentially in the rate and the dimensionality and remain prohibitive for problems characterized by large dimensionality.

A. Lattice decoding as a computationally efficient alternative to ML-based decoding

The high complexity required by ML-based (bounded) decoding solutions, serves as further motivation for exploring other families of decoding methods. A natural alternative is lattice decoding obtained by simply removing the constellation boundaries of the ML-based search, an action that loosely speaking exploits a certain symmetry which in turn may yield faster implementations. It is the case though that even with lattice decoding, the computational complexity can be prohibitive: finding the exact solution to the lattice decoding problem is generally an NP hard problem (cf. [2]). At the same time though, the other extreme of very early terminations of lattice decoding, such as linear solutions, have been known to achieve computational efficiency at the expense though of a very sizable gap to the exact solution of the lattice decoding problem.

In this work we provide the first lattice decoding solution that achieves, in the most general outage-limited MIMO setting and the high rate and high SNR limit, both a vanishing gap to the error-performance of the exact solution of regularized lattice decoding, as well as a computational complexity that is subexponential in the total number of codeword bits as well as the rate.

This conference paper represents the shortened version of a larger work in [3] which establishes that lattice reduction (LR) is the special ingredient that allows for complexity reductions. In light of this, it is important to note that while there is a general agreement in the community that lattice reduction does reduce complexity, cf. [4], this has not yet been supported analytically in any relevant communication settings. In fact, and quite opposite to common wisdom, it was recently shown that for a fixed radius sphere decoding implementation of the naive lattice decoder [5], LR does not improve the sphere decoder complexity tail exponent. What our present work shows is that in terms of the high SNR complexity in the outage-limited MIMO setting, LR brings down the complexity from a complexity cost that is comparable to ML (bounded) sphere decoding – which is exponential in the rate [1] – to a complexity cost that is subexponential. However, due to space constraints and in order to accentuate our main positive points we shall herein restrict our attention to the proof that LR-aided regularized (lattice) sphere decoding and proper timeout policies are *sufficient* for achieving a vanishing SNR-gap to exact MMSE lattice decoding at sub-exponential decoding complexity.

B. Rate, reliability and complexity in outage-limited MIMO communications

In the high SNR regime (SNR will be henceforth denoted as ρ), a given encoder $\mathcal{X}_r$ and decoder $\mathcal{D}_r$ are said to achieve a spatial multiplexing gain r (cf. [6]) and diversity gain $d(r)$ if

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

where $R(\rho)$ denotes the transmission rate and P_e denotes the probability of codeword error.

A high SNR measure of the computational resources required by $\mathcal{X}_r, \mathcal{D}_r$ to achieve a certain rate-reliability performance, was recently introduced in [1] to be the *complexity exponent*

$$c(r)\triangleq\inf\{x| -\lim_{\rho\to\infty}\frac{\log \mathrm{P}\left(N\geq\rho^x\right)}{\log\rho}>d(r)\} \qquad (1)$$

where N is the random variable describing the total number of spent floating point operations (flops). Generally N fluctuates with underlying channel, noise and transmitted code vector. A simple operational interpretation would be that the measure asymptotically describes the (maximum) number of flops one should have at their disposal (irrespective of the codeword and the channel realization) to achieve a certain rate-reliability performance which in turn is asymptotically described by $d(r)$.

The nature of the complexity measure (N) draws from that of the reliability measure in outage limited communications, where neither average-case nor worst-case analysis is revealing of the behavior of reliability (average case error-behavior would simply state that at high SNR, the probability of error approaches 1 for all rates below ergodic capacity, whereas the worst-case behavior would simply state that the capacity is zero). Similarly the complexity measure here deviates from the classical worst-case approach, and reveals a more pertinent *effective* complexity measure. Similarly the complexity exponent aims to best capture the entirety of the decoding complexity problem. Following the same analogy with DMT, we note that in a $n_T\times n_R$ MIMO system, the probability varies between 1 and $K_1\cdot\rho^{-n_{\mathrm{T}}n_{\mathrm{R}}}$, bringing to the fore an error exponent $d(r)\in[0,n_{\mathrm{T}}n_{\mathrm{R}}]$ (K_1 is a subpolynomial function of ρ). Similarly, the fact that the complexity of any reasonable decoder can vary between 1 and $K_2 2^{RT}=K_2\cdot\rho^{rT}$ flops (again K_2 is essentially a constant), brings to the fore a complexity exponent $c(r)\in[0,rT]$.

For any simplified variant of the baseline (exact) MMSE-preprocessed lattice decoder, the performance gap can, in the high SNR regime, be quantified as

$$g_L(c)\triangleq\lim_{\rho\to\infty}\frac{P_e}{\mathrm{P}\left(\hat{\mathbf{x}}_{rld}\neq\mathbf{x}\right)}$$

where $\mathrm{P}\left(\hat{\mathbf{x}}_{rld}\neq\mathbf{x}\right)$ describes the probability of error of the exact MMSE-preprocessed lattice decoder, where P_e denotes the probability of error of the simplified decoder, and where c (i.e., $c(r)$) describes the (asymptotic rate of increase of the) computational resources required to achieve this performance gap. Generally a smaller computational complexity c implies a larger gap $g_L(c)$. The clear task has remained for some time to construct decoders that optimally traverse this tradeoff between g and c, i.e., that reduce the performance gap to the exact lattice decoding solution, with reasonable computational complexity. Equivalently, in the high SNR regime, and for $N_{\max}$ denoting the computational resources in flops required to achieve a certain gap g to the baseline exact MMSE-preprocessed lattice decoder, the above task can be described as trying to minimize

$$C_L(g)\triangleq\inf\{\lim_{\rho\to\infty}\frac{\log N_{\max}}{\log\rho} \ : \ g_L=g\}.$$

This will be achieved later on.

C. Contributions

In this work we prove that the LR-aided MMSE-preprocessed lattice decoder, implemented by a fixed-radius sphere decoder and timeout policies that occasionally abort decoding and declare an error, achieves

$$g_L(\epsilon)=1, \quad C_L(g)=0 \qquad \forall\epsilon>0, g\geq 1,$$

i.e., achieves a vanishing gap to the exact implementation of MMSE-preprocessed lattice decoding and does so with a complexity exponent that vanishes to zero, which in turn implies subexponential complexity in the sense that the complexity scales slower that any conceivable exponential function.

The complexity exponent of LR-aided regularized (non-Euclidean metric) *linear* decoding, with a timeout policy based on the cost of LR (recently shown to be DMT optimal [7]) is zero, but with a potentially huge gap to ML and also a potentially huge gap to a exact solution to the lattice decoding problem, this gap can climb to $g=2^{\kappa/2}$, where κ is the total dimensionality of the problem. This gap is covered successfully by the LR-aided regularized lattice SD decoder, which achieves a vanishing gap to the exact lattice decoding, and does so with a complexity exponent that vanishes to zero, making it the first solution to achieve a vanishing gap to the exact lattice decoding problem, with complexity that is subexponential in the number of codeword bits as well as the rate.

D. Notation

We use $\doteq$ to denote the *exponential equality*, i.e., we write $f(\rho)\doteq\rho^B$ to denote $\lim_{\rho\to\infty}\frac{\log f(\rho)}{\log\rho}=B$, and $\dot{\leq}$, $\dot{\geq}$ are similarly defined. With this notation, we can write $P_e\doteq\rho^{-d(r)}$. In this paper we use $\|(\bullet)\|$ to denote the Euclidean norm of $(\bullet)$ and $(\bullet)^H$ to denote the conjugate transpose of $(\bullet)$.

II. System Model

We consider the general $m \times n$ point-to-point multiple-input multiple-output model given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ respectively denote the transmitted codewords, the received signal vectors, and the additive white Gaussian noise with unit variance, where as stated the parameter ρ takes the role of the signal to noise ratio (SNR), and where the fading matrix $\mathbf{H} \in \mathbb{R}^{n\times m}$ is assumed to be random, with elements drawn from arbitrary statistical distributions. We consider that one use of (2) corresponds to T uses of some underlying "physical" channel. We further assume the transmitted codewords $\mathbf{x}$ to be uniformly distributed over some codebook $\mathcal{X} \in \mathbb{R}^m$, to be statistically independent of the channel $\mathbf{H}$, and to satisfy the power constraint

$$E\{\|\mathbf{x}\|^2\} \leq T. \tag{3}$$

We finally consider the rate,

$$R = \frac{1}{T}\log|\mathcal{X}|, \tag{4}$$

in bits per channel use (bpcu), where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$.

We consider linear full rate codes, specifically, for $r \geq 0$, a (sequence of) linear (lattice) code(s) $\mathcal{X}_r$ is given by $\mathcal{X}_r = \Lambda_r \cap \mathcal{R}^{'}$ where $\Lambda_r \triangleq \rho^{\frac{-rT}{\kappa}}\Lambda$ and $\Lambda \triangleq \{\mathbf{G}\mathbf{s} | \mathbf{s} \in \mathbb{Z}^\kappa\}$, where $\mathbb{Z}^\kappa$ denotes the $\kappa = \min\{m,n\}$ dimensional integer lattice, where $\mathcal{R}^{'}$ is a compact convex subset of $\mathbb{R}^\kappa$ referred to as the *shaping region* of the code, where $\mathcal{R}^{'}$ contains the all zero vector, and where the full rank matrix $\mathbf{G} \in \mathbb{R}^{m\times\kappa}$ is the generator matrix of Λ. Both $\mathcal{R}^{'}$ and $\mathbf{G}$ are taken to be independent of ρ. For the class of lattice codes considered here, the codewords take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}}\mathbf{G}\mathbf{s}, \qquad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}}\mathcal{R}, \tag{5}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}^{'}$ that preserves the code. Combining (2) and (5) yields the equivalent model

$$\mathbf{y} = \mathbf{M}_r\mathbf{s} + \mathbf{w} \tag{6}$$

where

$$\mathbf{M}_r = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{H}\mathbf{G} \in \mathbb{R}^{n\times\kappa} \tag{7}$$

is a function of the multiplexing gain r. For simplicity of notation we will, in most cases, denote $\mathbf{M}_r$ with $\mathbf{M}$.

The basic naive lattice decoder which ignores $\mathcal{R}$, while keeping the decision metric of the ML decoder unaltered, takes the form (cf. [7], also [4])

$$\hat{\mathbf{s}}_{nld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{y} - \mathbf{M}\hat{\mathbf{s}}\|^2. \tag{8}$$

As a result of neglecting the boundary region, the above decoder declares an error if $\hat{\mathbf{s}}_{nld} \notin \mathbb{S}_r^\kappa$, resulting in possible performance costs incurred by neglecting the boundary constraint. These costs motivated the use of MMSE preprocessing which essentially regularizes the decision metric to penalize vectors outside the boundary constraint $\mathbb{S}_r^\kappa$ (cf. [7]).

III. LR-aided MMSE Preprocessed Lattice Sphere Decoding Complexity

A. MMSE Preprocessed Lattice Decoding

The MMSE preprocessed lattice decoder is obtained by implementing an unconstrained search over the MMSE preprocessed lattice, and takes the form

$$\hat{\mathbf{s}}_{rld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{F}\mathbf{y} - \mathbf{R}\hat{\mathbf{s}}\|^2, \tag{9}$$

where $\mathbf{F}$ and $\mathbf{R}$ are respectively the MMSE forward and feedback filters such that $\mathbf{F} = \mathbf{R}^{-H}\mathbf{M}^H$, and where

$$\mathbf{R}^H\mathbf{R} = \mathbf{M}^H\mathbf{M} + \alpha_r^2\mathbf{I}, \tag{10}$$

for $\alpha_r = \rho^{\frac{-rT}{\kappa}}$. $\mathbf{R}$ is an upper triangular matrix (see details in Appendix II). For $\mathbf{r} \triangleq \mathbf{F}\mathbf{y}$, the preprocessed model transforms from (6) to

$$\begin{aligned} \mathbf{r} &= \mathbf{R}^{-H}\mathbf{M}^H\mathbf{M}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}^{-H}(\mathbf{R}^H\mathbf{R} - \alpha_r^2\mathbf{I})\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}\mathbf{s} - {\alpha_r}^2\mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}\mathbf{s} + \mathbf{w}' \end{aligned} \tag{11}$$

where

$$\mathbf{w}^{'} = -\alpha_r^2\mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w}, \tag{12}$$

is the equivalent noise that includes self interference and colored Gaussian noise. Consequently the corresponding MMSE preprocessed lattice decoder takes the form

$$\hat{\mathbf{s}}_{rld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2. \tag{13}$$

B. LR-aided MMSE preprocessed Lattice Sphere Decoding

Lattice reduction techniques have been typically used in the MIMO setting to provide better error performance in the presence of suboptimal decoding solutions (cf. [8] [9]). In the current setting the LR algorithm, which is employed at the receiver after the action of MMSE preprocessing, aims to properly modify the search of the MMSE preprocessed lattice decoder (cf. (13)), from

$$\hat{\mathbf{s}}_{rld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2$$

to the new

$$\tilde{\mathbf{s}}_{lr-rld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{R}\mathbf{T}\hat{\mathbf{s}}\|^2, \tag{14}$$

by accepting as input the MMSE preprocessed lattice generator matrix $\mathbf{R}$, and producing as output the matrix $\mathbf{T} \in \mathbb{Z}^{\kappa\times\kappa}$ which is unimodular meaning that it has integer coefficients and unit-norm determinant, and which is designed so that the eigenvalues of $\mathbf{R}\mathbf{T}$ are more concentrated around one value. As a result of this unimodularity, we have that $\mathbf{T}^{-1}\mathbb{Z}^\kappa = \mathbb{Z}^\kappa$, and consequently the new search in (14) corresponds to yet another *lattice* decoder, referred to as the LR-aided MMSE preprocessed lattice decoder, which operates over a generally better conditioned channel matrix $\mathbf{R}\mathbf{T}$.

Finally with sphere decoding in mind, the LR algorithm is followed by the QR decomposition[1] of the new lattice-reduced MMSE preprocessed matrix $\mathbf{RT}$, resulting in a new upper triangular model

$$\tilde{\mathbf{r}} \quad = \quad \tilde{\mathbf{R}}\tilde{\mathbf{s}} + \mathbf{w}'' \tag{15}$$

and the new LR-aided MMSE preprocessed lattice search, which accepts the application of the sphere decoder, and which takes the form

$$\tilde{\mathbf{s}}_{lr-rld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}} \right\|^2, \tag{16}$$

where $\tilde{\mathbf{Q}}\tilde{\mathbf{R}} = \mathbf{RT}$ corresponds to the QR-decomposition of $\mathbf{RT}$, where $\tilde{\mathbf{R}}$ is upper triangular, where $\tilde{\mathbf{r}} \triangleq \tilde{\mathbf{Q}}^H \mathbf{r}$, $\tilde{\mathbf{s}} = \mathbf{T}^{-1}\mathbf{s}$, and where $\mathbf{w}'' = \tilde{\mathbf{Q}}^H \mathbf{w}'$.

At the very end,

$$\hat{\mathbf{s}}_{lr-rld} = \mathbf{T}\tilde{\mathbf{s}}_{lr-rld}, \tag{17}$$

allows for calculation of the estimate of the transmitted (information) vector $\mathbf{s}$ in (6).

We note here that this (exact) solution of the LR-aided MMSE preprocessed lattice decoder defined by (16), (17), is identical to the exact solution of the MMSE preprocessed lattice decoder given by (13), in the sense that

$$\begin{aligned}
\min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2 \quad &= \quad \min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \left\|\mathbf{r} - \mathbf{RTT}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(a)}{=} \quad \min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \left\|\mathbf{r} - \tilde{\mathbf{Q}}\tilde{\mathbf{R}}\mathbf{T}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(b)}{=} \quad \min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \left\|\tilde{\mathbf{r}} - \tilde{\mathbf{R}}\mathbf{T}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&= \quad \min_{\hat{\mathbf{s}}\in\mathbf{T}^{-1}\mathbb{Z}^{\kappa}} \left\|\tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(c)}{=} \quad \min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \left\|\tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}}\right\|^2,
\end{aligned} \tag{18}$$

where (a) follows from the fact that $\tilde{\mathbf{Q}}\tilde{\mathbf{R}} = \mathbf{RT}$, (b) follows from the rotational invariance of the Euclidean norm, and (c) follows from the fact that $\mathbf{T}^{-1}\mathbb{Z}^{\kappa} = \mathbb{Z}^{\kappa}$.

The sphere decoder solves (16) by recursively enumerating all lattice vectors $\hat{\mathbf{s}} \in \mathbb{Z}^{\kappa}$ within a given sphere of radius $\xi > 0$, i.e., by identifying as candidates the vectors $\hat{\mathbf{s}}$ that satisfy

$$\|\tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}}\|^2 \leq \xi^2. \tag{19}$$

The algorithm specifically uses the upper triangular nature of $\tilde{\mathbf{R}}$ to recursively identify partial symbol vectors $\hat{\mathbf{s}}_k$, $k = 1, \cdots, \kappa$, for which

$$\|\tilde{\mathbf{r}}_k - \tilde{\mathbf{R}}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2, \tag{20}$$

where $\hat{\mathbf{s}}_k$ and $\tilde{\mathbf{r}}_k$ respectively denote the last k components of $\hat{\mathbf{s}}$ and $\tilde{\mathbf{r}}$, and where $\tilde{\mathbf{R}}_k$ denotes the $k \times k$ lower right submatrix of $\tilde{\mathbf{R}}$. Clearly any set of vectors $\hat{\mathbf{s}} \in \mathbb{Z}^{\kappa}$, with common last k components that fail to satisfy (20), may be excluded from the set of candidate vectors.

[1]A more proper statement would be that the QR decomposition is performed by the LR algorithm it self.

The enumeration of partial symbol vectors $\hat{\mathbf{s}}_k$ is equivalent to the traversal of a regular tree with κ layers – one layer per symbol component of the information vectors, such that layer k corresponds to the kth component of the transmitted information vector $\tilde{\mathbf{s}}$. There is a one-to-one correspondence between the nodes at layer k and the partial vectors $\hat{\mathbf{s}}_k$. We say that a node is visited by the sphere decoder if and only if the corresponding partial vector $\hat{\mathbf{s}}_k$ satisfies (20), i.e., there is a bijection between the visited nodes at layer k and the set

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{Z}^k \mid \|\tilde{\mathbf{r}}_k - \tilde{\mathbf{R}}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}. \tag{21}$$

The total number of visited nodes (in all layers of the tree) is given by

$$N_{SD} = \sum_{k=1}^{\kappa} N_k, \tag{22}$$

where $N_k \triangleq |\mathcal{N}_k|$ is the number of visited nodes at layer k of the search tree. The total number of visited nodes is commonly taken as a measure of the sphere decoder complexity. It is easy to show that in the scale of interest the SD complexity exponent $c(r)$ would not change if instead of considering the number of visited nodes, we considered the number of flops spent by the decoder[2].

In choosing the search radius ξ, we note that for the transmitted information vector $\mathbf{s}$, the metric in (16) satisfies

$$\|\tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}}\|^2 = \|\mathbf{w}^{''}\|^2,$$

which means that if $\|\mathbf{w}^{''}\| > \xi$, then the transmitted information vector is excluded from the search, resulting in a decoding error. As Lemma 2 will later argue, taking into consideration the self-interference and non-Gaussianity of $\mathbf{w}^{''}$, we can set $\xi = \sqrt{z \log \rho}$, for some $z > z' > d(r)$ such that

$$\mathrm{P}\left(\|\mathbf{w}^{''}\|^2 > \xi^2\right) \dot{<} \rho^{-z'},$$

which implies a vanishing probability of excluding the transmitted information vector from the search, and a vanishing degradation of performance.

The usefulness of LR techniques in improving the error performance of suboptimal MIMO lattice decoders has received substantial attention, with works like [8] and [9] introducing the approach in the context of MIMO detectors, and works like [7] rigorously revealing the approach's error-performance capabilities for a very broad setting. On the other hand, as was discussed in the introduction, apart from simulations little has been shown about the role of LR in reducing the complexity of MIMO decoders. Thus, we proceed to rigorously prove that indeed lattice reduction techniques, and specifically a proper utilization of the LLL algorithm [11], can dramatically reduce the complexity of such MIMO decoders.

[2]This stems from the fact that the total number of flops required for evaluating the bound in (20) may be upper and lower bounded by constants that are independent of ρ (cf. [10]).

C. Decoding Complexity Analysis

We are here interested in establishing the SD complexity behavior when implementing the search in (16). As in (21), we now identify the corresponding unpruned sets at layer k to be

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{Z}^k \mid \|\tilde{\mathbf{r}}_k - \tilde{\mathbf{R}}_k \hat{\mathbf{s}}_k\|^2 \leq \xi^2\}, \quad (23)$$

and in bounding the above, we first focus on understanding the statistical behavior of the LLL-reduced submatrices $\tilde{\mathbf{R}}_k$ $k = 1, \cdots, \kappa$. Towards this, and for $d_L(r-\epsilon)$ denoting the diversity gain of the exact implementation of the regularized lattice decoder at multiplexing gain $r-\epsilon$, we have the following lemma whose proof appears in Appendix I.

Lemma 1: The smallest singular values $\sigma_{min}(\tilde{\mathbf{R}}_k)$ of the submatrices $\tilde{\mathbf{R}}_k$, $k = 1, \cdots, \kappa$, after MMSE preprocessing and LLL lattice reduction, satisfy

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_k) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}, \text{ for all } r \geq \epsilon > 0. \quad (24)$$

To bound the cardinality N_k of $\mathcal{N}_k$, and eventually the total number $N_{SD} = \sum_{k=1}^{\kappa} N_k$ of lattice points visited by the SD, we proceed along the lines of the work in [1], making the proper modifications to account for MMSE preprocessing, for the removal of the bounding region, and for lattice reduction.

Towards this we see that, after removing the bounding constraint, Lemma 1 in [1] tells us that

$$N_k \triangleq |\mathcal{N}_k| \leq \prod_{i=1}^{k} \left[\sqrt{k} + \frac{2\xi}{\sigma_i(\tilde{\mathbf{R}}_k)}\right],$$

where

$$\sigma_{min}(\tilde{\mathbf{R}}_k) = \sigma_1(\tilde{\mathbf{R}}_k) \leq \cdots \leq \sigma_k(\tilde{\mathbf{R}}_k)$$

are the singular values of $\tilde{\mathbf{R}}_k$. Consequently we have that

$$N_k \leq \left[\sqrt{k} + \frac{2\xi}{\sigma_{min}(\tilde{\mathbf{R}}_k)}\right]^k. \quad (25)$$

As a result, for any $\tilde{\mathbf{R}}_k$ such that

$$\sigma_{min}(\tilde{\mathbf{R}}_k) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}, \quad (26)$$

and given that $\xi = \sqrt{z \log \rho}$ for some finite z, then

$$N_k \leq \left(\sqrt{k} + \frac{2\sqrt{z \log \rho}}{\rho^{\frac{-\epsilon T}{\kappa}}}\right)^k \doteq \rho^{\frac{\epsilon T k}{\kappa}}, \quad (27)$$

which guarantees that the total number of visited lattice points is upper bounded as

$$N_{SD} = \sum_{k=1}^{\kappa} N_k \dot{\leq} \sum_{k=1}^{\kappa} \rho^{\frac{\epsilon T k}{\kappa}} \doteq \rho^{\epsilon T}. \quad (28)$$

Consequently, directly from Lemma 1, we have that

$$\mathrm{P}\left(N_{SD} \dot{\geq} \rho^{\epsilon T}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}. \quad (29)$$

A similar approach deals with the complexity of the LLL algorithm, which is known (cf. [12]) to be generally unbounded. Specifically drawing from [7, Lemma 2], under the assumption of power-limited channels [3] (cf. [7]), and for N_{LR} denoting the number of flops spent by the LLL algorithm, one can readily conclude that

$$\mathrm{P}\left(N_{LR} \geq \gamma \log \rho\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}, \quad (30)$$

for any $\gamma > \frac{1}{2}(d(r-\epsilon))$. Consequently the overall complexity

$$N \doteq N_{SD} + N_{LR},$$

in flops, for the LR-aided MMSE preprocessed lattice sphere decoder, satisfies the following

$$\begin{aligned} \mathrm{P}\left(N \dot{\geq} \rho^{\epsilon T}\right) &\doteq \mathrm{P}\left(\{N_{SD} \dot{\geq} \rho^{\epsilon T}\} \cup \{N_{LR} \dot{\geq} \rho^{\epsilon T}\}\right) \\ &\dot{\leq} \rho^{-d_L(r-\epsilon)}. \end{aligned} \quad (31)$$

As a result, going back to (1) we see that the complexity exponent for the decoder, in the presence of appropriate timeout policies, takes the form

$$\begin{aligned} c(r) = &\inf\left\{x \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N \geq \rho^x\right)}{\log \rho} > d_L(r)\right\} \\ = &\inf\left\{\epsilon \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N \dot{\geq} \rho^{\epsilon T}\right)}{\log \rho} > d_L(r)\right\} \end{aligned}$$

and vanishes arbitrarily close to zero, resulting in a zero complexity exponent. The following theorem then holds.

Theorem 1: LR-aided MMSE preprocessed lattice sphere decoding introduces a zero complexity exponent.

D. Gap to the exact solution of MMSE preprocessed lattice decoding

We here go one step further and prove that the LR-aided regularized lattice sphere decoder and the associated timeout policies that guarantee a vanishing complexity exponent, also guarantee a vanishing gap to the error performance of the exact lattice decoding implementation. This result is motivated by potentially exponential gaps in the performance of other DMT optimal decoders (cf. [7]), where these gaps may grow with an exponential form of up to $2^{\frac{\kappa}{2}}$ (cf. [13]).

Towards establishing this gap, we recall that the exact MMSE preprocessed lattice decoder in (13) makes errors when $\hat{\mathbf{s}}_{rld} \neq \mathbf{s}$. On the other hand the LLL-reduced MMSE preprocessed lattice sphere decoder with run-time constraints, in addition to making the same errors ($\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}$), also makes errors when the run-time limit of ρ^x becomes active, i.e., when $N \geq \rho^x$, as well as when a small search radius causes $\mathcal{N}_\kappa = \emptyset$. Consequently the corresponding performance gap to the exact regularized decoder, takes the form

$$g_L(x) = \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\{\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\} \cup \{N \geq \rho^x\} \cup \{\mathcal{N}_\kappa = \emptyset\}\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)}.$$

[3] This is a moderate assumption that asks that $\mathrm{E}\left\{\|\boldsymbol{H}\|_{\mathrm{F}}^2\right\} \dot{\leq} \rho$. We note that this holds true for any telecommunications setting.

To bound the above gap, we apply the union bound and the fact that

$$\mathrm{P}\left(\mathcal{N}_\kappa = \emptyset\right) \leq \mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)$$

to get that

$$g_L(x) \leq \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)} + \lim_{\rho\to\infty} \frac{\mathrm{P}\left(N \geq \rho^x\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)} + \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)}. \quad (32)$$

Furthermore from (18) we observe that

$$\mathrm{P}\left(\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\right) = \mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right), \quad (33)$$

and from (31), and under the assumption that $d_L(r-\epsilon) > d_L(r)$, we recall that

$$\mathrm{P}\left(N \dot{\geq} \rho^{\epsilon T}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}$$

which implies that for any $x > 0$ it holds that

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(N \geq \rho^x\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)} = 0. \quad (34)$$

Finally the last term in (32) relates to the search radius ξ, and to the behavior of the noise $\mathbf{w}''$ which was shown in (12), (15) to take the form

$$\mathbf{w}'' = \tilde{\mathbf{Q}}^H\left(-\alpha_r^2\mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w}\right).$$

The following lemma, whose proof is found in Appendix II, accounts for the fact that $\mathbf{w}''$ includes self interference and colored noise, to bound the last term in (32).

Lemma 2: There exist a finite $z > d_L(r)$ for which a search radius $\xi = \sqrt{z\log\rho}$ guarantees that

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s}\right)} = 0. \quad (35)$$

Consequently combining (33), (34) and (35) gives that $g_L(x) = 1, \ \forall x > 0$. The following directly holds.

Theorem 2: LR-aided MMSE preprocessed lattice sphere decoding with a computational constraint activated at ρ^x flops, allows for a vanishing gap to the exact solution of MMSE preprocessed lattice decoding, for any $x > 0$. Equivalently the same LR-aided decoder guarantees that

$$g_L(\epsilon) = 1, \quad C_L(g) = 0 \qquad \forall\epsilon > 0, g \geq 1,$$

for all fading statistics, all MIMO scenarios, and all lattice designs.

IV. Conclusions

In the setting of outage limited MIMO communications, the work proved that a vanishing gap to the error performance of fully implemented regularized lattice decoder can be achieved at a computational complexity that is subexponential in the rate. These performance and complexity guarantees hold for the most general MIMO settings, for almost all fade statistics, all channel dimensions and all lattice codes.

The vanishing gap approach serves as an analytical refinement over basic diversity analysis which generally fails to address potentially massive gaps between theory and practice. A vanishing gap essentially guarantees that the two error curves are arbitrarily close, given a sufficiently high SNR, which is a much stronger condition than DMT optimality which only guarantees an error gap that is subexponential in ρ, and can thus be unbounded and infinite.

Appendix I
Proof for Lemma 1

For $\mathbf{R}_r^H\mathbf{R}_r = \mathbf{M}_r^H\mathbf{M}_r + \alpha_r^2\mathbf{I}$ (cf. (10))[4], it follows by the bounded orthogonality defect of LLL reduced bases that there is a constant $K_\kappa > 0$ independent of $\mathbf{R}_r$ and ρ, for which (c.f. [11] and the proof in [14])

$$\sigma_{max}(\tilde{\mathbf{R}}_r^{-1}) \leq \frac{K_\kappa}{\lambda(\mathbf{R}_r)} \quad (36)$$

where

$$\lambda(\mathbf{R}_r) \triangleq \min_{\mathbf{c}\in\mathbb{Z}^\kappa\setminus\mathbf{0}} \|\mathbf{R}_r\mathbf{c}\| \quad (37)$$

denotes the shortest vector in the lattice generated by $\mathbf{R}_r$. As a result we have that

$$\sigma_{min}(\tilde{\mathbf{R}}_r) \geq \frac{\lambda(\mathbf{R}_r)}{K_\kappa}. \quad (38)$$

Looking to lower bound $\sigma_{min}(\tilde{\mathbf{R}}_r)$, we seek a bound on $\lambda(\mathbf{R}_r)$. Towards this let $r' = r - \gamma$ for some $r \geq \gamma > 0$, in which case for $\mathbf{s}$ being the information vector corresponding to the transmitted vector, and for any $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ such that $\hat{\mathbf{s}} \neq \mathbf{s}$, it follows that

$$\begin{aligned} \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| &= \|(\mathbf{r} - \mathbf{R}_{r'}\mathbf{s}) + \mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| \\ &\leq \|\mathbf{r} - \mathbf{R}_{r'}\mathbf{s}\| + \|\mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| \end{aligned} \quad (39)$$

and

$$\begin{aligned} \|\mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| &\geq \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| - \|\mathbf{r} - \mathbf{R}_{r'}\mathbf{s}\| \\ &= \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| - \|\mathbf{w}\|. \end{aligned} \quad (40)$$

From (40) it is clear that to find a lower bound on $\lambda(\mathbf{R}_{r'})$, we need to lower bound $\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|$ for all $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ and upper bound $\|\mathbf{w}\|$. Let us, for now, assume that $\|\mathbf{w}\|^2 \leq \rho^b$. To lower bound $\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|$, we draw from the equivalence of MMSE preprocessing and the regularized metric (cf. equation (45) in [7]), and rewrite

$$\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|^2 = \|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\|^2 + \alpha_{r'}^2\|\hat{\mathbf{s}}\|^2 - c, \quad (41)$$

where $c \triangleq \mathbf{y}^H[\mathbf{I} - \mathbf{M}_{r'}^H(\mathbf{M}_{r'}^H\mathbf{M}_{r'} + \alpha_{r'}^2\mathbf{I})^{-1}\mathbf{M}_{r'}]\mathbf{y} \geq 0$. We now note that for $\hat{\mathbf{s}} = \mathbf{s}$ then $\|\mathbf{y} - \mathbf{M}_{r'}\mathbf{s}\|^2 + \alpha_{r'}^2\|\mathbf{s}\|^2 = \|\mathbf{w}\|^2 \leq \rho^b$, and since the left hand side of (41) cannot be negative, and furthermore given that c is independent of $\hat{\mathbf{s}}$, we conclude that $c \leq \rho^b$.

We will now proceed to lower bound $\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\|^2 + \alpha_{r'}^2\|\hat{\mathbf{s}}\|^2$ and then use (41) to lower bound $\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|$. Towards lower bounding $\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\|^2 + \alpha_{r'}^2\|\hat{\mathbf{s}}\|^2$ we draw from Theorem 1 in [7] and let $\mathcal{B}$ be the spherical region given by

$$\mathcal{B} \triangleq \{d \in \mathbb{R}^\kappa \mid \|\mathbf{d}\|^2 \leq \Gamma\}$$

[4] Note the transition to the notation reflecting the dependence of $\mathbf{R}$ on r

where the radius $\Gamma > 0$, which is independent of ρ, is chosen so that $\mathbf{d}_1 + \mathbf{d}_2 \in \mathcal{R}$ for any $\mathbf{d}_1, \mathbf{d}_2 \in \mathcal{B}$. The existence of the set $\mathcal{B}$ follows by the assumption that $\mathbf{0}$ is contained in the interior of $\mathcal{R}$. Now let

$$\nu_{r'} \triangleq \min_{\mathbf{d} \in \rho^{\frac{r'T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa : \mathbf{d} \neq \mathbf{0}} \frac{1}{4} \left\| \mathbf{M}_{r'} \mathbf{d} \right\|^2,$$

and for given $\gamma > \zeta > 0$ choose $b > 0$ such that

$$\frac{2\zeta T}{\kappa} > b > 0.$$

This may clearly be done for arbitrary $\zeta > 0$. We will in the following temporarily assume that $\nu_{r'+\zeta} \geq 1$ and prove that, together with $\|\mathbf{w}\|^2 \leq \rho^b$, the two conditions are sufficient for $\lambda(\tilde{\mathbf{R}}_{r'}) \dot{\geq} \rho^{\frac{\zeta T}{\kappa}}$ to hold.

In order to bound the metric for $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ where $\hat{\mathbf{s}} \neq \mathbf{s}$, we note that $\nu_{r'+\zeta} \geq 1$ implies that $\forall\ \mathbf{d} \in \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa, \mathbf{d} \neq \mathbf{0}$ it is the case that

$$\begin{aligned} \frac{1}{4} \left\| \mathbf{M}_{r'+\zeta} \mathbf{d} \right\|^2 &\geq 1 \\ \frac{1}{4} \left\| \rho^{\frac{1}{2} - \frac{(r'+\zeta)T}{\kappa}} \mathbf{H} \mathbf{G} \mathbf{d} \right\|^2 &\overset{(a)}{\geq} 1 \\ \frac{1}{4} \left\| \rho^{\frac{1}{2} - \frac{r'T}{\kappa}} \mathbf{H} \mathbf{G} \mathbf{d} \right\|^2 &\geq \rho^{\frac{2\zeta T}{\kappa}} \end{aligned} \quad (42)$$

where (a) follows from the fact that

$$\mathbf{M}_r = \rho^{\frac{1}{2} - \frac{rT}{\kappa}} \mathbf{H} \mathbf{G}.$$

Consequently

$$\frac{1}{4} \left\| \mathbf{M}_{r'} \mathbf{d} \right\|^2 \geq \rho^{\frac{2\zeta T}{\kappa}} \ \forall\ \mathbf{d} \in \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa, \mathbf{d} \neq \mathbf{0}. \quad (43)$$

As $\mathcal{R}$ is bounded, and as $\zeta > 0$, it holds that $\mathcal{R} \subset \frac{1}{2} \rho^{\frac{\zeta T}{\kappa}} \mathcal{B}$ for all $\rho \geq \rho_1$ for a sufficiently large ρ_1. This implies that

$$\mathbf{s} \in \frac{1}{2} \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \text{ for } \rho \geq \rho_1$$

since $\mathbf{s} \in \rho^{\frac{r'T}{\kappa}} \mathcal{R}$.

For $\mathbf{s}, \mathbf{d} \in \frac{1}{2} \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa$, there exists a $\hat{\mathbf{s}} \in \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa$, $\hat{\mathbf{s}} \neq \mathbf{s}$, such that $\hat{\mathbf{s}} = \mathbf{d} + \mathbf{s}$. Hence for any $\hat{\mathbf{s}} \in \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B} \cap \mathbb{Z}^\kappa$, we have from (43) that

$$\frac{1}{4} \left\| \mathbf{M}_{r'} (\hat{\mathbf{s}} - \mathbf{s}) \right\|^2 = \frac{1}{4} \left\| \mathbf{M}_{r'} \mathbf{d} \right\|^2 \geq \rho^{\frac{2\zeta T}{\kappa}}. \quad (44)$$

As $\|\mathbf{w}\|^2 \leq \rho^b$, it follows that

$$\frac{1}{4} \left\| \mathbf{M}_{r'} \mathbf{d} \right\|^2 \geq \|\mathbf{w}\|^2$$

for large ρ, and that

$$\left\| \mathbf{y} - \mathbf{M}_{r'} \hat{\mathbf{s}} \right\|^2 = \left\| \mathbf{M}_{r'} (\mathbf{s} - \hat{\mathbf{s}}) + \mathbf{w} \right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \quad (45)$$

Consequently

$$\left\| \mathbf{y} - \mathbf{M}_{r'} \hat{\mathbf{s}} \right\|^2 + \alpha_{r'}^2 \left\| \hat{\mathbf{s}} \right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \quad (46)$$

On the other hand if $\hat{\mathbf{s}} \notin \rho^{\frac{(r'+\zeta)T}{\kappa}} \mathcal{B}$, then by definition of $\mathcal{B}$ we have that

$$\alpha_{r'}^2 \left\| \hat{\mathbf{s}} \right\|^2 \geq \frac{1}{4} \Gamma \rho^{\frac{2\zeta T}{\kappa}},$$

and consequently that

$$\left\| \mathbf{y} - \mathbf{M}_{r'} \hat{\mathbf{s}} \right\|^2 + \alpha_{r'}^2 \left\| \hat{\mathbf{s}} \right\|^2 \geq \frac{1}{4} \Gamma \rho^{\frac{2\zeta T}{\kappa}}. \quad (47)$$

From (46) and (47) we then conclude that

$$\left\| \mathbf{y} - \mathbf{M}_{r'} \hat{\mathbf{s}} \right\|^2 + \alpha_{r'}^2 \left\| \hat{\mathbf{s}} \right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \quad (48)$$

Given (46) and (48), for any $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ such that $\hat{\mathbf{s}} \neq \mathbf{s}$, it is the case that

$$\left\| \mathbf{y} - \mathbf{M}_{r'} \hat{\mathbf{s}} \right\|^2 + \alpha_{r'}^2 \left\| \hat{\mathbf{s}} \right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}},$$

which combined with $c \leq \rho^b$ allows for (41) to give that

$$\left\| \mathbf{r} - \mathbf{R}_{r'} \hat{\mathbf{s}} \right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \quad (49)$$

Applying (37) and (40), we have

$$\begin{aligned} \lambda(\mathbf{R}_{r'}) &\geq \left\| \mathbf{r} - \mathbf{R}_{r'} \hat{\mathbf{s}} \right\| - \|\mathbf{w}\| \\ &\dot{\geq} \rho^{\frac{\zeta T}{\kappa}} - \rho^{\frac{b}{2}} \\ &\doteq \rho^{\frac{\zeta T}{\kappa}} \end{aligned} \quad (50)$$

where the exponential inequality follows from (49). Furthermore we know that

$$\lambda(\mathbf{R}_r) = \rho^{\frac{-\gamma T}{\kappa}} \lambda(\mathbf{R}_{r'}) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}} \quad (51)$$

where $\epsilon = \gamma - \zeta$, $r \geq \epsilon > 0$, and from (38) and (51) it follows that $\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}$.

We now note that the above implies that for $\nu_{r'+\zeta} \geq 1$ and $\|\mathbf{w}\|^2 \leq \rho^b$ then $\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}$, and thus applying the union bound yields

$$\begin{aligned} \mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) &= \mathrm{P}\left((\nu_{r'+\zeta} < 1) \cup (\|\mathbf{w}\|^2 > \rho^b)\right) \\ &\leq \mathrm{P}\left(\nu_{r'+\zeta} < 1\right) + \mathrm{P}\left(\|\mathbf{w}\|^2 > \rho^b\right). \end{aligned}$$

We know from the exponential tail of the Gaussian distribution that

$$\mathrm{P}\left(\|\mathbf{w}\|^2 > \rho^b\right) \doteq \rho^{-\infty}$$

and from Lemma 1 in [7] that

$$\mathrm{P}\left(\nu_{r'+\zeta} < 1\right) \dot{\leq} \rho^{-d_{ML}(r'+\zeta)},$$

where $d_{ML}(r'+\zeta)$ denotes the diversity gain of the ML decoder at multiplexing gain $r'+\zeta$. Hence

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_{ML}(r-\epsilon)}$$

for all $r \geq \epsilon > 0$.

The association with the singular values

$$\sigma_1(\tilde{\mathbf{R}}_{r,k}) \leq \cdots \leq \sigma_k(\tilde{\mathbf{R}}_{r,k})$$

is made using the interlacing property of singular values of sub-matrices, which gives that

$$\sigma_i(\tilde{\mathbf{R}}_{r,k}) \geq \sigma_i(\tilde{\mathbf{R}}_r),\ i \leq k = 1, \cdots, \kappa, \quad (52)$$

and for $k = 1, \cdots, \kappa$, that

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_{r,k}) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_{ML}(r-\epsilon)}.$$

From DMT optimality of the exact implementation of the regularized lattice decoder, we have that

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_{r,k}) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}.$$

This proves Lemma 1.□

Appendix II
Proof for Lemma 2

For a search radius that grows as $\xi = \sqrt{z \log \rho} \doteq \rho^0$, we first prove that

$$\mathrm{P}\left(\|\mathbf{w}^{''}\|^2 > \xi^2 \right) \doteq \rho^{-z'}$$

for $z > z' > d_L(r)$. To do this, we will need an understanding of the equivalent noise, and thus consider an equivalent representation of the MMSE preprocessed lattice decoder. Towards this let (cf. [15])

$$\mathbf{QR} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \mathbf{R} = \begin{bmatrix} \mathbf{M} \\ \alpha_r \mathbf{I} \end{bmatrix} \in \mathbb{R}^{(n+m)\times m} \quad (53)$$

be the thin QR factorization of the modified channel matrix, where

$$\mathbf{Q}_1 = \mathbf{R}^{-1}\mathbf{M} \in \mathbb{R}^{n\times m}, \qquad \mathbf{Q}_2 = \alpha_r \mathbf{R}^{-1} \in \mathbb{R}^{m\times m}$$

and where

$$\mathbf{R}^H\mathbf{R} = \mathbf{M}^H\mathbf{M} + \alpha_r^2\mathbf{I}.$$

It then follows that for $\mathbf{F} = \mathbf{Q}_1^H$, the MMSE-preprocessed lattice decoder is equivalent to lattice decoding in the presence of channel $\mathbf{R}$ and noise

$$\begin{aligned} \mathbf{w}^{'} &= -\alpha_r^2 \mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= -\alpha_r \mathbf{Q}_2^H \mathbf{s} + \mathbf{Q}_1^H \mathbf{w}. \end{aligned} \quad (54)$$

Consequently we calculate

$$\begin{aligned} & \mathrm{P}\left(\|\mathbf{w}^{'}\| > \xi \right) \\ = \; & \mathrm{P}\left(\{ \| -\alpha_r \mathbf{Q}_2^H \mathbf{s} \| + \| \mathbf{Q}_1^H \mathbf{w} \| \} > \xi \right) \\ \overset{(a)}{=} \; & \mathrm{P}\left(\{ \| -\alpha_r \mathbf{Q}^H \begin{bmatrix} \mathbf{s} \\ \mathbf{0} \end{bmatrix} \| + \| \mathbf{Q}^H \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} \| \} > \xi \right) \\ \leq \; & \mathrm{P}\left(m\{ \sup_{\mathbf{s}\in \mathbb{S}_r^\kappa} \| -\alpha_r \mathbf{s} \| + \|\mathbf{w}\| \} > \xi \right) \\ \overset{(b)}{=} \; & \mathrm{P}\left(mK + m\|\mathbf{w}\| > \xi \right) \\ = \; & \mathrm{P}\left(m\|\mathbf{w}\| > (z \log \rho)^{\frac{1}{2}} - mK \right) \\ \overset{(c)}{\leq} \; & \mathrm{P}\left(m\|\mathbf{w}\| > (z_1 \log \rho)^{\frac{1}{2}} \right) \\ = \; & \mathrm{P}\left(\|\mathbf{w}\|^2 > \frac{z_1}{m^2} \log \rho \right) \\ \overset{(d)}{=} \; & \mathrm{P}\left(\|\mathbf{w}\|^2 > z_2 \log \rho \right) \\ \doteq \; & \rho^{-z_2} \end{aligned} \quad (55)$$

where (a) follows from the MMSE preprocessed equivalent channel representation, and where the inequalities in (b), (c) and (d) follow for some fixed K and for some arbitrary z_1, z_2 satisfying $z > z_1 > z_2 > 0$ independent of ρ. Consequently

$$\mathrm{P}\left(\|\mathbf{w}^{''}\| > \xi \right) = \mathrm{P}\left(\|\tilde{\mathbf{Q}}^H \mathbf{w}^{'}\| > \xi \right) \doteq \rho^{-z'}$$

for some $0 < z' < z_2$ and we have

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(\|\mathbf{w}^{''}\| > \xi \right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{rld} \neq \mathbf{s} \right)} = \lim_{\rho\to\infty} \rho^{(d_L(r) - z')} = 0,$$

where last equality follows after choosing the search radius such that $z > z' > d_L(r)$. This proves Lemma 2.□

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 228044, from the Swedish Foundation for Strategic Research (SSF) / grant ICA08-0046, from FP7/2007-2013 grant agreement no. 257616 (CONECT), and from the Mitsubishi Electric R&D Centre Europe project Home-eNodeBS.

References

[1] J. Jaldén and P. Elia, "Sphere decoding complexity exponent for decoding full rate codes over the quasi-static MIMO channel," Feb. 2011, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1102.1265 [cs.IT].

[2] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212 – 1215, Mar. 2001.

[3] A. Singh, P. Elia, and J. Jaldén, "Achieving a vanishing performance-gap to exact lattice decoding at a subexponential complexity," Jul. 2011, submitted to *IEEE Trans. Inf. Theory*, available on arXiv:1107.4021v1 [cs.IT].

[4] 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, no. 10, pp. 2389–2401, Oct. 2003.

[5] D. Seethaler, J. Jalden, C. Studer, and H. Bölcskei, "Tail behavior of sphere-decoding complexity in random lattices," in *Proc. IEEE International Symposium on Information Theory, ISIT*, Jun. 2009, pp. 729 – 733.

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

[7] J. Jaldén and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4765–4780, Oct. 2010.

[8] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Telecommunications Conference, GLOBECOM*, Taipei, Taiwan, Nov. 2002.

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

[10] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806– 2818, Aug. 2005.

[11] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematishe Annalen*, vol. 261, no. 4, pp. 1432–11 807, Dec. 1982.

[12] J. Jaldén, D. Seethaler, and G. Matz, "Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Las Vegas, Nevada, USA, Apr. 2008.

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

[14] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL Reduction Achieves the Receive Diversity in MIMO Decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4801 –4805, Dec. 2007.

[15] 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 International Conference on Communications, ICC*, vol. 2, Jun. 2004, pp. 798 – 802.