On the Complexity Distribution of Sphere Decoding

D. Seethaler*, J. Jaldén°, C. Studer‡, and H. Bölcskei‡

* RobArt, 4020 Linz, Austria
e-mail: dominik.seethaler@gmail.com

° Royal Institute of Technology (KTH), 100 44 Stockholm, Sweden
e-mail: joakim.jalden@ee.kth.se

‡ ETH Zurich, 8092 Zurich, Switzerland
e-mail:{studerc,boelcskei}@nari.ee.ethz.ch

Abstract

We analyze the (computational) complexity distribution of sphere decoding (SD) for random infinite lattices. In particular, we show that under fairly general assumptions on the statistics of the lattice basis matrix, the tail behavior of the SD complexity distribution is fully determined by the inverse volume of the fundamental regions of the underlying lattice. Particularizing this result to $N \times M$, $N \geq M$, i.i.d. circularly symmetric complex Gaussian lattice basis matrices, we find that the corresponding complexity distribution is of Pareto-type with tail exponent given by $N - M + 1$. A more refined analysis reveals that the corresponding average complexity of SD is infinite for $N = M$ and finite for $N > M$. Finally, for i.i.d. circularly symmetric complex Gaussian lattice basis matrices, we analyze SD preprocessing techniques based on lattice-reduction (such as the LLL algorithm or layer-sorting according to the V-BLAST algorithm) and regularization. In particular, we show that lattice-reduction does not improve the tail exponent of the complexity distribution while regularization results in a SD complexity distribution with tails that decrease faster than polynomial.

Index Terms

Closest lattice point problem, sphere decoding, complexity distribution, random lattices, MIMO wireless

This work was supported in part by the STREP project No. IST-026905 (MASCOT) within the Sixth Framework Programme of the European Commission. This paper was presented in part at IEEE ISIT 2009, Seoul, South Korea, June 2009.

I. INTRODUCTION

Finding the closest lattice point in an infinite lattice is commonly referred to as the *closest lattice point* (CLP) problem (see, e.g., [1]). The sphere decoding (SD) algorithm [1]–[8] is a promising approach for solving the CLP problem efficiently. The (computational) complexity of SD, as measured in terms of the number of lattice points searched by the algorithm, depends strongly on the lattice basis matrix and is, in general, difficult to characterize analytically. However, if the lattice basis matrix is assumed *random*, the complexity of SD is random as well and one can resort to a characterization of the *complexity distribution* of SD. This approach is similar in spirit to the line of work conducted in the 1960's aimed at characterizing the complexity distribution of sequential decoding of convolutional codes [9], [10], where the randomness of complexity is a consequence of random additive noise. Previous work on the complexity of SD focused on characterizing the mean and the variance of SD complexity for i.i.d. Gaussian lattice basis matrices [11]–[14]. Characterizing and understanding the complexity distribution is important, for example, when SD is used under practically relevant run-time constraints (see, e.g., [8]). In this paper, we make a first attempt in this direction by analyzing the tail behavior of the SD complexity distribution in terms of corresponding tail exponents (i.e., polynomial decay rates). This approach allows us, among others, to characterize the impact of lattice basis matrix preprocessing–e.g., regularization or lattice-reduction–on the tail behavior of the SD complexity distribution.

The main contributions of this paper can be summarized as follows:

- Under fairly general assumptions on the statistics of the lattice basis matrix and for a large class of preprocessing methods, we prove that the tail exponent of the SD complexity distribution is given by the tail exponent of the distribution of the inverse volume of the fundamental regions of the underlying lattice. These results comprise, for example, lattice basis matrices with non-zero mean correlated complex Gaussian distributed entries as well as preprocessing through lattice-reduction (LR) (see, e.g., [1]).
- Specializing our main result to the case of $N \times M$, $N \geq M$, i.i.d. circularly symmetric complex Gaussian lattice basis matrices, we find that the complexity distribution of SD is of Pareto-type (i.e., the corresponding tails decrease polynomially) with tail exponent given by $N - M + 1$. We show that this tail exponent cannot be improved (i.e., increased) by LR

including layer-sorting (LS) (e.g., according to the V-BLAST algorithm [15]) as a special case. We find, however, that regularization of the lattice basis matrix results in a complexity distribution with faster-than-polynomially decreasing tails. Here it is important to note that solving the CLP on a regularized lattice basis matrix does, in general, not yield the solution to the original CLP.

- For i.i.d. circularly symmetric complex Gaussian lattice basis matrices, we show that the average (w.r.t. the lattice basis matrix) complexity of SD is infinite for $N = M$ and finite for $N > M$; this complements results derived in [11], [12].

Notation: We write $A_{i,j}$ for the entry in the ith row and jth column of the matrix $\mathbf{A}$ and x_i for the ith entry of the vector $\mathbf{x}$. Slightly abusing common terminology, we call an $N \times M$, $N \geq M$, matrix $\mathbf{A}$ unitary if it satisfies $\mathbf{A}^H\mathbf{A} = \mathbf{I}$, where ^{H} denotes conjugate transposition, i.e., transposition ^{T} followed by element-wise complex conjugation *, and $\mathbf{I}$ is the identity matrix of appropriate size. The inverses of $\mathbf{A}$ and $\mathbf{A}^H$ are referred to as $\mathbf{A}^{-1}$ and $\mathbf{A}^{-H}$, respectively. For a $N \times M$ matrix $\mathbf{A}$, $\mathrm{vec}(\mathbf{A}) = [\mathbf{a}_1^T \dots \mathbf{a}_M^T]^T$, where $\mathbf{a}_i$, $i = 1, \dots, M$, is the ith column of $\mathbf{A}$. The ordered eigenvalues of a positive semidefinite $M \times M$ matrix $\mathbf{A}$ are referred to as $\lambda_i(\mathbf{A})$, $i = 1, \dots, M$, with ordering such that $0 \leq \lambda_1(\mathbf{A}) \leq \dots \leq \lambda_M(\mathbf{A})$; the corresponding determinant is given by $\det(\mathbf{A}) = \prod_{i=1}^{M} \lambda_i(\mathbf{A})$. The Euclidean- and the Frobenius norm are denoted by $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$, respectively, and $|\mathcal{X}|$ refers to the cardinality of the set $\mathcal{X}$. The ceil-function is designated as $\lceil\cdot\rceil$. Furthermore, $\mathbb{C}\mathbb{Z}$ stands for the set of Gaussian integers, i.e., $\mathbb{C}\mathbb{Z} = \mathbb{Z} + \sqrt{-1}\,\mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. The lattice with $N \times M$, $N \geq M$, full-rank basis matrix $\mathbf{A}$ is defined as

$$\mathcal{L}(\mathbf{A}) = \left\{\mathbf{A}\mathbf{d} : \mathbf{d} \in (\mathbb{C}\mathbb{Z})^M\right\} \subset \mathbb{C}^N.$$

For $N = M$, the corresponding covering radius is given by the largest distance from a point in $\mathbb{C}^M$ to its closest lattice point in $\mathcal{L}(\mathbf{A})$ [16], i.e., by

$$\mu(\mathbf{A}) = \max_{\mathbf{x} \in \mathbb{C}^M} \min_{\mathbf{d} \in (\mathbb{C}\mathbb{Z})^M} \|\mathbf{x} - \mathbf{A}\mathbf{d}\|. \tag{1}$$

$\mathbb{E}\{\cdot\}$ stands for the expectation operator. We write $x \sim \chi_a$ if the RV x is χ-distributed with $a > 0$ degrees of freedom and normalized such that $\mathbb{E}\{x^2\} = a$. The probability density function (pdf) of the RV $x \sim \chi_a$ is then given by [17]

$$f_x(t) = \frac{2^{1-a/2}}{\Gamma(a/2)} t^{a-1} e^{-t^2/2}, \quad t \geq 0 \tag{2}$$

and $f_x(t) = 0$, $t < 0$, where $\Gamma(a) = \int_0^\infty y^{a-1}e^{-y}dy$ refers to the Gamma function. For the corresponding cumulative distribution function (cdf) we have $P\big[x \leq t\big] = \gamma_{a/2}(t^2/2)$, where $\gamma_a(t)$ denotes the (regularized) lower incomplete Gamma function. If $x \sim \chi_a$, the RV $y = x^2$ is χ^2-distributed with $a > 0$ degrees of freedom and cdf $P\big[y \leq t\big] = \gamma_{a/2}(t/2)$. A complex Gaussian RV x with mean μ_x and variance σ_x^2, i.e., $x - \mu_x$ is circularly symmetric Gaussian distributed, is denoted as $x \sim \mathcal{CN}(\mu_x, \sigma_x^2)$. The "little o" notation $g(x) = o(f(x))$, $x \to x_0$, stands for $\lim_{x\to x_0} g(x)/f(x) = 0$. We write $\stackrel{d}{=}$ for equality in distribution. The natural logarithm is referred to as $\log(\cdot)$. By $g(x) \stackrel{a}{\sim} f(x)$, $x \to x_0$, we mean that $\lim_{x\to x_0} g(x)/f(x) = 1$. We write $g(x) \doteq f(x)$, $x \to x_0$, to mean that $\lim_{x\to x_0} \log g(x)/\log(x) = \lim_{x\to x_0} \log f(x)/\log(x)$, assuming, of course, that the corresponding limits exist. The symbols $\dot{\leq}$ and $\dot{\geq}$ are defined analogously. We write $g(x) \doteq x^{\pm\infty}$, $x \to x_0$, and $g(x) \doteq x^0$, $x \to x_0$, respectively, for $\lim_{x\to x_0} \log g(x)/\log(x) = \pm\infty$ and $\lim_{x\to x_0} \log g(x)/\log(x) = 0$.

A. The Closest Lattice Point Problem

The CLP problem (or integer least squares problem) refers to computing

$$\widehat{\mathbf{d}} \;=\; \underset{\mathbf{d}\in(\mathbb{CZ})^M}{\arg\,\min} \|\mathbf{r} - \mathbf{H}\mathbf{d}\|^2 \tag{3}$$

for a given vector $\mathbf{r} \in \mathbb{C}^N$ and a given full-rank matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$, $N \geq M$. In words, solving (3) amounts to finding the point in the lattice $\mathcal{L}(\mathbf{H})$ that is closest (in Euclidean sense) to $\mathbf{r}$. In communications, (3) is known as the maximum-likelihood (ML) detection problem for detecting a transmitted vector $\mathbf{d}' \in (\mathbb{CZ})^M$ based on the linear model

$$\mathbf{r} \;=\; \mathbf{H}\mathbf{d}' + \mathbf{w} \tag{4}$$

with $\mathbf{H}$ known at the receiver and noise $\mathbf{w}$ having i.i.d. circularly symmetric complex Gaussian components. For example, in the case of ML detection in multiple-input multiple-output (MIMO) wireless systems with spatial multiplexing, $\mathbf{H}$ is the channel matrix, which is typically assumed to consist of i.i.d. circularly symmetric complex Gaussian components. For examples of further ML detection problems, which are of the form (3), we refer to [6], [12], [18]. We note that in the practically relevant case of $\mathbf{d}'$ in (4) being drawn from a finite subset $\mathcal{A}$ of $(\mathbb{CZ})^M$, ML detection corresponds to (3) only if $\mathcal{A}$ is relaxed to $(\mathbb{CZ})^M$. Such a relaxation step induces a performance loss (see, e.g., [19]) but is necessary if, e.g., SD in combination with LLL-based preprocessing is used.

B. Sphere Decoding

A prominent approach for solving (3) is the SD algorithm [1]–[8]. In the following, we consider Fincke-Pohst SD [2] without radius reduction and restarting as done, e.g., in [11]–[14]. The algorithm starts by computing the (unique) QR-decomposition (QRD) $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is unitary of dimension $N \times M$ and $\mathbf{R}$ is an $M \times M$ upper triangular matrix with positive real-valued elements on its main diagonal. We note that (3) can equivalently be written as

$$\widehat{\mathbf{d}} = \underset{\mathbf{d}\in(\mathbb{CZ})^M}{\arg\min} \|\mathbf{y} - \mathbf{Rd}\|^2 \tag{5}$$

where $\mathbf{y} = \mathbf{Q}^H\mathbf{r}$. Next, (5) is solved subject to a *sphere constraint* (SC), which amounts to considering only those lattice points $\mathbf{Rd}$ that lie within a hypersphere of radius ρ around $\mathbf{y}$, i.e., all $\mathbf{d}$ that satisfy

$$\|\mathbf{y} - \mathbf{Rd}\|^2 \leq \rho^2. \tag{6}$$

Here, the sphere radius ρ has to be chosen sufficiently large for the search sphere to contain at least one lattice point $\mathbf{Rd}$. Note, however, that if ρ is chosen too large, too many points will satisfy the SC and the complexity of SD will be high. As detailed next, triangularization and the corresponding SC (6) enable an efficient recursive solution of the CLP problem in (3).

Consider the length-k subvectors $\mathbf{d}_k \in (\mathbb{CZ})^k$ of $\mathbf{d}$ defined as $\mathbf{d}_k = (d_{M-k+1} \cdots d_M)^T$, $k = 1, \ldots, M$, where k is called the *layer* index. The metric $\|\mathbf{y} - \mathbf{Rd}\|^2 = \|\mathbf{y}_M - \mathbf{R}_M\mathbf{d}_M\|^2$ can be computed recursively according to

$$\|\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k\|^2 = \|\mathbf{y}_{k-1} - \mathbf{R}_{k-1}\mathbf{d}_{k-1}\|^2 + |\Delta_k(\mathbf{d}_k)|^2 \tag{7}$$

where

$$|\Delta_k(\mathbf{d}_k)|^2 = \left| y_{M-k+1} - \sum_{i=M-k+1}^{M} R_{M-k+1,i}\, d_i \right|^2 \tag{8}$$

denotes the metric update for layer k, $\mathbf{R}_k$ refers to the $k \times k$ bottom right (upper triangular) submatrix of $\mathbf{R}$ associated with $\mathbf{d}_k$, and $\mathbf{y}_k = (y_{M-k+1} \cdots y_M)^T$. Thus, with (7), a necessary condition for $\mathbf{d}$ to satisfy the SC is that any associated $\mathbf{d}_k$ satisfy the *partial SC*

$$\|\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k\|^2 \leq \rho^2. \tag{9}$$

This formulation now enables the following approach for finding all integer vectors $\mathbf{d}$ that satisfy (6) in an efficient (recursive) manner. Starting with layer $k = 1$, the condition (9) reduces to

$$|y_M - R_{M,M}d_M|^2 \leq \rho^2 \tag{10}$$

which states that the Mth component d_M of all $\mathbf{d} \in (\mathbb{CZ})^M$ satisfying (6) must be located inside a circle of radius $\rho/R_{M,M}$ and center point $y_M/R_{M,M}$. In the next step, for every $d_M \in \mathbb{CZ}$ that satisfies (10), one finds all $d_{M-1} \in \mathbb{CZ}$ such that (9) is satisfied for $k = 2$. This procedure is repeated until $k = M$. Among the lattice points delivered by the algorithm, the one with minimum $\|\mathbf{y} - \mathbf{Rd}\|^2$ constitutes the solution of (5).

II. Complexity Distribution of SD

We define the computational complexity of SD as the number of lattice points searched by the algorithm, i.e., the total number of vectors $\mathbf{d}_k \in (\mathbb{CZ})^k$, $k = 1, \ldots, M$, that satisfy the partial SCs in (9) (cf. [11], [20]). Specifically, we define the kth *layer complexity* of SD as

$$S_k = \left|\left\{\mathbf{d}_k \in (\mathbb{CZ})^k : \|\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k\|^2 \leq \rho^2\right\}\right| \tag{11}$$

with the resulting *total complexity*

$$S = \sum_{k=1}^{M} S_k. \tag{12}$$

It was shown in [7] that S is proportional (at least for the finite-lattice case) to the VLSI-implementation complexity of SD.

A. Complexity Distribution and Tail Exponents

The quantities S_k, $k = 1, \ldots, M$, and S, defined above, are functions of $\mathbf{H}$, $\mathbf{r}$, and ρ. In the following, we let $\mathbf{H}$ and $\mathbf{r}$ be random (potentially statistically dependent) and we let ρ be a deterministic, strictly positive, and finite scalar. For example, in the case of ML detection of the transmitted data vector $\mathbf{d}'$ in MIMO wireless systems with spatial multiplexing, $\mathbf{r}$ is given by (4), where the entries of $\mathbf{H}$ (the channel matrix) and $\mathbf{w}$ (the noise vector) are typically assumed i.i.d. circularly symmetric complex Gaussian. Furthermore, ρ is typically chosen as a function of the noise variance such that the SD algorithm finds $\mathbf{d}'$ with a certain (high) probability (cf. [4], [11], [21]). Since $\mathbf{H}$ and $\mathbf{r}$ are random, S_k and S are random as well and can be characterized through their respective (complementary) distributions: The layer-wise complexity distributions $\mathrm{P}[S_k \geq L]$, $k = 1, \ldots, M$, and the total complexity distribution $\mathrm{P}[S \geq L]$. While these distributions are hard to come by analytically, it turns out that the corresponding *tail exponents* ξ_k, $k = 1, \ldots, M$, and ξ, defined by

$$\mathrm{P}[S_k \geq L] \doteq L^{-\xi_k}, \quad L \to \infty$$

and

$$\mathrm{P}[S \geq L] \doteq L^{-\xi}, \quad L \to \infty$$

are amenable to an analytical characterization. We note that $S = \sum_{k=1}^{M} S_k$ implies

$$\xi = \min\{\xi_1, \ldots, \xi_M\}. \tag{13}$$

The tail exponents characterize the tail behavior of the corresponding complexity distributions in terms of polynomial decay rates in L for $L \to \infty$. We note that the tail exponents are non-negative by definition. If the tail exponent is infinity or zero, the corresponding complexity distribution decreases faster or slower than polynomial in L, respectively. In particular, for finite non-zero tail exponents, the corresponding complexity distributions are of Pareto-type meaning that they decay polynomially in L for large L.

The tail exponents can be used to draw general conclusions about the corresponding complexity distributions as follows. If the complexity distributions $\mathrm{P}[S^{(1)} \geq L]$ and $\mathrm{P}[S^{(2)} \geq L]$ have tail exponents $\xi^{(1)}$ and $\xi^{(2)}$ with $\xi^{(1)} > \xi^{(2)}$, we have $\mathrm{P}[S^{(1)} \geq L] < \mathrm{P}[S^{(2)} \geq L]$ for sufficiently large L. Similarly, if the complexity distribution $\mathrm{P}[S \geq L]$ has tail exponent ξ, we have

$$L^{-(\xi+\delta)} \leq \mathrm{P}[S \geq L] \leq L^{-(\xi-\delta)} \tag{14}$$

for any $\delta > 0$ and sufficiently large L. Larger tail exponents are, in general, desirable as they imply that the probability of the complexity being atypically large is smaller. This, for example, is advantageous in the context of MIMO detection under practically relevant run-time constraints, i.e., when limits on the number of lattice points that can be searched are imposed (see, e.g., [8]).

We conclude this section by emphasizing that the complexity tail exponent as defined above characterizes one specific characteristic of the corresponding complexity distribution and does by no means yield a full picture. For example, if $\mathrm{P}[S \geq L] = c\,L^{-1}$, with some constant $c > 0$ not depending on L, the corresponding tail exponent equals 1 irrespective of the multiplicative constant c. Furthermore, if the tail exponents of $\mathrm{P}[S^{(1)} \geq L]$ and $\mathrm{P}[S^{(2)} \geq L]$ are equal, i.e., $\xi^{(1)} = \xi^{(2)} = \xi$, no order relation between $\mathrm{P}[S^{(1)} \geq L]$ and $\mathrm{P}[S^{(2)} \geq L]$ can be inferred from ξ.

B. Main Result

Preliminaries: The complexity of SD can often be reduced by employing additional channel matrix preprocessing techniques such as lattice-reduction (LR) (see, e.g., [1]) or regularization

(see, e.g., [8], [22]). For LR, which includes layer-sorting (LS) as a special case, SD is applied to the triangularized form of the CLP problem (5) obtained by applying the QRD to $\mathbf{HT}$ instead of $\mathbf{H}$, where $\mathbf{T}$ is a unimodular matrix depending solely on $\mathbf{H}$ (see Section III-C for more details). In the remainder of the paper, any type of *preprocessing* is incorporated into the mapping from $\mathbf{r}$ to $\mathbf{y}$ and the mapping from $\mathbf{H}$ to $\mathbf{R}$. In particular, we assume that $\mathbf{y}$ is a general function of $\mathbf{r}$ and $\mathbf{H}$ and that $\mathbf{R}$ is a general function of $\mathbf{H}$ only, where both functions depend on the specific preprocessing[1]. We note that this implies that the submatrices $\mathbf{R}_k$, $k = 1, \ldots, M$, of $\mathbf{R}$ (cf. (9)) are general functions of $\mathbf{H}$ only.

Theorem 1: Consider SD with fixed ρ ($0 < \rho < \infty$) and let $\mathbf{H}$ and $\mathbf{r}$ be random (potentially statistically dependent). The corresponding kth layer complexity S_k, defined in (11), satisfies

$$\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H \mathbf{R}_k)} \geq L\right], \quad L \to \infty \tag{15}$$

provided all of the following conditions are met:

- *Statistics of* $\mathbf{H}$: There exists a constant $\beta \in \mathbb{R}$, $\beta > 0$, such that the probability density function (pdf) $f(\mathbf{H})$ of $\mathbf{H}$ satisfies the scaling property

$$f(\mathbf{H}) \geq \beta f(a\mathbf{H}) \tag{16}$$

  for all $\mathbf{H} \in \mathbb{C}^{N \times M}$ and all $a \in \mathbb{R}$, $a > 1$.

- *Statistics of* $\mathbf{H}$ *and preprocessing*: The covering radius $\mu(\mathbf{R})$ of $\mathcal{L}(\mathbf{R})$ satisfies

$$\mathrm{P}[\mu(\mathbf{R}) \geq L] \doteq L^{-\infty}, \quad L \to \infty. \tag{17}$$

- *Preprocessing*: Let $\det(\mathbf{R}_k^H \mathbf{R}_k) = g_k(\mathbf{H})$ and $\mu(\mathbf{R}) = g_\mu(\mathbf{H})$. There exist constants $\alpha_k, \alpha \in \mathbb{R}$, $\alpha_k > 0$, $\alpha > 0$, such that the functions $g_k(\mathbf{H})$ and $g_\mu(\mathbf{H})$ satisfy, respectively, the scaling properties

$$g_k(b\mathbf{H}) = b^{\alpha_k} g_k(\mathbf{H}) \tag{18}$$

  and

$$g_\mu(b\mathbf{H}) = b^{\alpha} g_\mu(\mathbf{H}) \tag{19}$$

  for all $\mathbf{H} \in \mathbb{C}^{N \times M}$ and all $b \in \mathbb{R}$, $b > 0$.

Proof: See Appendix A. □

[1]For example, consider the case that the triangularized form of the CLP problem is obtained by applying the QRD directly to $\mathbf{H}$, i.e., $\mathbf{H} = \mathbf{QR}$, as described in Section I-B. The resulting preprocessing based on direct QRD is captured in the mappings from $\mathbf{r}$ to $\mathbf{y}$ and $\mathbf{H}$ to $\mathbf{R}$ as $\mathbf{y} = \mathbf{Q}^H \mathbf{r}$ and $\mathbf{R} = \mathbf{Q}^H \mathbf{H}$, respectively.

Basic Proof Approach: We now briefly state the basic proof approach for Theorem 1. The proof is based on separately establishing the exponential lower bound $\mathrm{P}[S_k \geq L] \dot{\geq} \mathrm{P}\big[1/\det(\mathbf{R}_k^H\mathbf{R}_k) \geq L\big]$, $L \to \infty$, and the exponential upper bound $\mathrm{P}[S_k \geq L] \dot{\leq} \mathrm{P}\big[1/\det(\mathbf{R}_k^H\mathbf{R}_k) \geq L\big]$, $L \to \infty$, which then combine to $\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\big[1/\det(\mathbf{R}_k^H\mathbf{R}_k) \geq L\big]$. The starting points for establishing these bounds are the following upper and lower bounds on the layer complexity S_k (see [23, Ch. 3.2, Eq. (3.3)]):

$$\frac{V_k(\rho) - \mu(\mathbf{R}_k)A_k(\rho)}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \leq S_k \leq \frac{V_k(\rho + \mu(\mathbf{R}_k))}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \tag{20}$$

where $V_k(\rho)$ and $A_k(\rho)$ denote the volume and the surface area of a hypersphere with radius ρ in k complex-valued dimensions, i.e.,

$$V_k(\rho) = \frac{\pi^k \rho^{2k}}{k!} \tag{21}$$

and

$$A_k(\rho) = \frac{2\pi^k \rho^{2k-1}}{(k-1)!}. \tag{22}$$

We note that our proof uses all conditions of Theorem 1, i.e., (16)–(19), to establish the exponential lower bound, while the exponential upper bound only needs condition (17).

Discussion of the Theorem: Theorem 1 states that the tail exponent of $\mathrm{P}[S_k \geq L]$ is fully characterized by the tail exponent of $\mathrm{P}[1/\det(\mathbf{R}_k^H\mathbf{R}_k) \geq L]$ provided that conditions (16)–(19) are satisfied. It is immediate that the tail exponent of $\mathrm{P}[S_k \geq L]$ then depends only on the statistics of the lattice basis matrix $\mathbf{H}$ and neither on the statistics of $\mathbf{r}$ nor on the particular choice of ρ as long as $0 < \rho < \infty$. The conditions (16)–(19) constitute fairly general requirements on the statistics of the lattice basis matrix $\mathbf{H}$ and on the preprocessing method. For example, in Appendix B, it is shown that the conditions (16)–(19) are satisfied for direct QRD if the entries of $\mathbf{H}$ are jointly Gaussian with arbitrary non-singular covariance matrix and arbitrary finite mean, i.e., for $\mathbf{H}$ being a Rayleigh- or Ricean-fading MIMO channel with non-singular covariance matrix. In Section III-C, it is furthermore demonstrated that Theorem 1 also comprises preprocessing techniques based on LR such as V-BLAST LS or LR according to the LLL algorithm.

It is interesting to note that $\det(\mathbf{R}_k^H\mathbf{R}_k)$ is the volume of the fundamental regions of $\mathcal{L}(\mathbf{R}_k)$ [16]. A well-known approximation for S_k is given by [23]

$$\widehat{S}_k = \frac{V_k(\rho)}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \tag{23}$$

where $V_k(\rho)$ was defined in (21). This approximation simply counts the number of fundamental regions (each occupied by exactly one lattice point) that fit into the k-dimensional search sphere and becomes exact if averaging of S_k is performed over $\mathbf{y}_k$ assumed uniformly distributed over $\mathcal{L}(\mathbf{R}_k)$ [23]. Motivated by this insight, $\widehat{S}_k$ was used in [1] and [20] to assess the complexity of different SD variants. It immediately follows that (15) can equivalently be written as

$$\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\Big[\widehat{S}_k \geq L\Big], \quad L \to \infty \tag{24}$$

which holds for arbitrary statistics of $\mathbf{r}$ (or, equivalently, $\mathbf{y}_k$) and no averaging over $\mathbf{y}_k$ is required.

Finally, we emphasize that Theorem 1 does not depend on the specific *shape* of the search region, a k-dimensional hypersphere in the SD case. Indeed, the theorem continues to hold if the search sphere is replaced by a general bounded search region with a non-empty interior (see Appendix C). This, for example, includes a SD variant based on the l^∞-norm [14] where the induced search regions are hypercubes.

III. Tail Exponents for i.i.d. Gaussian zero-mean $\mathbf{H}$

In the remainder of the paper, we assume lattice basis matrices $\mathbf{H}$ whose entries are i.i.d. Gaussian with zero-mean and variance σ_H^2 (this model is often used in the context of MIMO detection). Specializing Theorem 1 to this case is shown below to lead to particularly simple and interesting results.

A. Tail Exponents for Direct QRD

As shown in Appendix B, all the conditions of Theorem 1 are met for direct QRD and for the entries of $\mathbf{H}$ being jointly Gaussian distributed with arbitrary non-singular covariance matrix and arbitrary finite mean. This evidently includes the i.i.d. Gaussian zero-mean case considered here and, hence, the relation

$$\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H \mathbf{R}_k)} \geq L\right], \quad L \to \infty \tag{25}$$

holds for $k = 1, \ldots, M$. Using results from [24] on the near-zero behavior of the eigenvalues of complex Wishart matrices, it is shown in Appendix D that

$$\mathrm{P}[\det(\mathbf{R}_k^H \mathbf{R}_k) \leq \epsilon] \doteq \epsilon^{N-M+1}, \quad \epsilon \to 0 \tag{26}$$

for $k = 1, \ldots, M$. For SD with direct QRD and i.i.d. Gaussian zero-mean lattice basis matrices, the result in (26) together with (25) now establishes that

$$\mathrm{P}[S_k \geq L] \doteq L^{-(N-M+1)}, \quad L \to \infty, \quad k = 1, \ldots, M. \tag{27}$$

Evidently, the corresponding total complexity then satisfies (see (13))

$$\mathrm{P}[S \geq L] \doteq L^{-(N-M+1)}, \quad L \to \infty. \tag{28}$$

We can now draw the following conclusions:

- The distributions of the individual layer and total complexities are of Pareto-type with tail exponents $\xi_k = \xi = N - M + 1$, $k = 1, \ldots, M$.
- Increasing N (i.e., the number of receive antennas in a MIMO detection context) for fixed M (i.e., the number of transmit antennas) results in improved (i.e., larger) complexity tail exponents.

Relation to Diversity Order: The complexity tail exponent $N - M + 1$ is reminiscent of the SNR exponent (or diversity order) of conventional suboptimum MIMO detection schemes such as linear and successive interference cancellation (SIC) schemes. Specifically, the error probability $P_e(\mathrm{SNR})$ as a function of the signal-to-noise ratio (SNR) of such schemes satisfies $P_e(\mathrm{SNR}) \doteq \mathrm{SNR}^{-(N-M+1)}$, $\mathrm{SNR} \to \infty$ (in contrast to optimum detection where one would get $P_e(\mathrm{SNR}) \doteq \mathrm{SNR}^{-N}$, $\mathrm{SNR} \to \infty$). In fact, this equivalence is not a coincidence as the same statistical properties of $\mathbf{H}$ are responsible for the diversity order of linear or SIC detectors and for the complexity tail exponent of SD. In particular, we note that the typical error event analysis in [25, Chapter 3] reveals that (see also [26])

$$P_e(\mathrm{SNR}) \doteq \mathrm{P}\left[R_{M,M}^2 < \frac{1}{\mathrm{SNR}}\right], \quad \mathrm{SNR} \to \infty \tag{29}$$

for the error probability of the first detected data symbol in SIC detection (which determines the overall diversity order). On the other hand, for the first layer complexity distribution of SD, it follows from Theorem 1 that

$$\mathrm{P}[S_1 \geq L] \doteq \mathrm{P}\left[\frac{1}{R_{M,M}^2} \geq L\right], \quad L \to \infty$$

which has the same structure as the right hand side (RHS) of (29). Consequently, we obtain the same exponents for the diversity order and the first layer complexity distribution. For the other layer complexity distributions no such direct link can be established.

B. Refined Analysis for Direct QRD

As already mentioned, a shortcoming of the tail exponent analysis is that the impact of multiplicative constants is not captured. As a result, for example, the tail exponent analysis does not capture the impact of the sphere radius ρ on the complexity distribution of SD. In the following, we provide a more refined analysis of the first layer complexity distribution $\mathrm{P}[S_1 \geq L]$ for direct QRD on i.i.d. Gaussian zero-mean lattice basis matrices $\mathbf{H}$. This characterization reveals, among other factors, the influence of the sphere radius ρ on $\mathrm{P}[S_1 \geq L]$ and, furthermore, allows us to derive a lower bound on the total complexity distribution $\mathrm{P}[S \geq L]$, which we use to draw interesting conclusions on the average (w.r.t. to $\mathbf{H}$ and $\mathbf{r}$) total complexity of SD complementing results derived in [11], [12].

From (20) for $k = 1$ together with $\mu(\mathbf{R}_1) = \mu(R_{M,M}) = R_{M,M}/\sqrt{2} \leq R_{M,M}$, we obtain

$$\frac{\pi\rho^2 - 2\pi R_{M,M}\rho}{R_{M,M}^2} \leq S_1 \leq \frac{\pi(\rho + R_{M,M})^2}{R_{M,M}^2}$$

which can equivalently be written as

$$\pi\left(\left(\frac{\rho}{R_{M,M}} - 1\right)^2 - 1\right) \leq S_1 \leq \pi\left(\frac{\rho}{R_{M,M}} + 1\right)^2.$$

Consequently, we obtain

$$\mathrm{P}\left[\frac{\rho^2}{R_{M,M}^2} \geq L'\right] \leq \mathrm{P}[S_1 \geq L] \leq \mathrm{P}\left[\frac{\rho^2}{R_{M,M}^2} \geq L''\right]$$

where $L' = \left(\sqrt{\frac{L}{\pi} + 1} + 1\right)^2$ and $L'' = \left(\sqrt{\frac{L}{\pi}} - 1\right)^2$. Furthermore, since $\frac{\sqrt{2}}{\sigma_H} R_{M,M} \sim \chi_{2(N-M+1)}$ [27, Lemma 2.1], we arrive at

$$\gamma_{N-M+1}\left(\frac{\rho^2}{\sigma_H^2 L'}\right) \leq \mathrm{P}[S_1 \geq L] \leq \gamma_{N-M+1}\left(\frac{\rho^2}{\sigma_H^2 L''}\right). \tag{30}$$

These upper and lower bounds together with $L' \overset{a}{\sim} L'' \overset{a}{\sim} L/\pi$, $L \to \infty$, and $V_1(\rho) = \pi\rho^2$ yield

$$\mathrm{P}[S_1 \geq L] \overset{a}{\sim} \gamma_{N-M+1}\left(\frac{V_1(\rho)}{\sigma_H^2 L}\right), \quad L \to \infty. \tag{31}$$

Next, noting that

$$\gamma_a(y) = \frac{1}{a!} y^a (1 + o(1)), \quad y \to 0$$

we obtain

$$\mathrm{P}[S_1 \geq L] \overset{a}{\sim} \frac{1}{(N-M+1)!} \left(\frac{\sigma_H^2}{V_1(\rho)} L\right)^{-(N-M+1)}, \quad L \to \infty. \tag{32}$$

Compared to $\mathrm{P}[S_1 \geq L] \doteq L^{-(N-M+1)}$, $L \to \infty$, we have thus obtained a finer picture of the first layer complexity distribution since multiplicative constants are also captured. More specifically, (32) quantifies the impact of the sphere radius ρ and shows, as expected, that $\mathrm{P}[S_1 \geq L]$ increases with increasing ρ for large L. Next, we note that (32) establishes a stronger version of (24) for the first layer. In particular, (31) together with $\mathrm{P}\Big[\widehat{S}_1 \geq L\Big] = \mathrm{P}\Big[\frac{V_1(\rho)}{R_{M,M}^2} \geq L\Big] = \gamma_{N-M+1}\Big(\frac{V_1(\rho)}{\sigma_H^2 L}\Big)$ implies that

$$\mathrm{P}[S_1 \geq L] \overset{a}{\sim} \mathrm{P}\Big[\widehat{S}_1 \geq L\Big], \quad L \to \infty.$$

We can now trivially lower-bound the distribution $\mathrm{P}[S \geq L]$ of the total complexity according to $\mathrm{P}[S \geq L] \geq \mathrm{P}[S_1 \geq L]$. Together with the lower bound $\mathrm{P}[S_1 \geq L] \geq \gamma_{N-M+1}\Big(\frac{\rho^2}{\sigma_H^2 L'}\Big)$ (see (30)) and using $L' \geq 4 + L$, $\gamma_a(x) \geq e^{-x}\frac{x^a}{a!}$ (this is a direct consequence of the series expansion of the incomplete Gamma function [28, Sec. 6.5]), and $e^{-\rho^2/\sigma_H^2(L+4)} \geq e^{-\rho^2/4\,\sigma_H^2}$, we obtain

$$\mathrm{P}[S \geq L] \geq C\,(L+4)^{-(N-M+1)} \tag{33}$$

where

$$C = \frac{1}{(N-M+1)!}\left(\frac{\rho^2}{\sigma_H^2}\right)^{N-M+1} e^{-\frac{\rho^2}{4\,\sigma_H^2}}.$$

The total complexity distribution can therefore be lower-bounded by a Pareto distribution with exponent $N-M+1$. We note that this result is similar in spirit to results obtained in the context of sequential decoding of convolutional codes over additive white Gaussian noise channels, where randomness results from random additive noise. Specifically, in [9], [10] it was shown that the complexity distribution of sequential decoding is lower-bounded by a Pareto distribution; the corresponding exponent was found to depend on the code rate and on the noise variance.

The Pareto lower bound (33) can be used to draw a number of interesting conclusions on the expected total complexity $\mathbb{E}\{S\}$, where the expectation is taken over $\mathbf{H}$ (assumed i.i.d. Gaussian and zero-mean) and $\mathbf{r}$ (arbitrary statistics). Starting from

$$\mathbb{E}\{S\} = \int_0^\infty \mathrm{P}[S \geq L]\;\mathrm{d}L \tag{34}$$

the lower bound in (33) yields

$$\mathbb{E}\{S\} \geq C \int_0^\infty (L+4)^{-(N-M+1)}\,\mathrm{d}L. \tag{35}$$

The integral on the RHS of (35) does not converge for $N = M$ as the integrand behaves as L^{-1} in this case. Consequently, $\mathbb{E}\{S\}$ is infinite for $N = M$. This complements results on the

average complexity of SD derived in [11], [12]. Specifically, [12, Corollary 1, Eqn. (5)] provides an expression for $\mathbb{E}\{S\}$ for the special case (with respect to our assumptions) of $\mathbf{r}$ given by (4) in terms of an infinite sum over incomplete Gamma functions. Based on this expression it seems hard to draw general conclusions on $\mathbb{E}\{S\}$. The lower bound in (35), however, immediately shows that this sum does not converge for $N = M$. Moreover, for $N > M$, $\mathbb{E}\{S\}$ is finite and, consequently, [12, Corollary 1, Eqn. (5)] converges in this case. To see this, we start from (34) together with the upper bound

$$\mathrm{P}[S \geq L] \leq \begin{cases} 1, & \text{for} \quad L \leq L_0 \\ L^{-(N-M+1-\delta_0)}, & \text{for} \quad L > L_0 \end{cases} \tag{36}$$

for sufficiently large L_0 and $\delta_0 < 1$ (cf. the upper bound in (14)). Inserting (36) into (34), we get

$$\mathbb{E}\{S\} \leq \int_0^{L_0} \mathrm{d}L + \int_{L_0}^{\infty} L^{-(N-M+1-\delta_0)} \, \mathrm{d}L$$

which implies $\mathbb{E}\{S\} < \infty$ for any $N > M$. We can therefore conclude that the average total complexity $\mathbb{E}\{S\}$ of SD is bounded for $N > M$ while it is unbounded for $N = M$. Along these lines, we note that the Pareto lower and upper bounds on $\mathrm{P}[S \geq L]$ according to (33) and (36), respectively, can be used to show directly that all moments $\mathbb{E}\{S^a\}$ with $a < N - M + 1$ are finite while all corresponding higher order moments with $a \geq N - M + 1$ are infinite. For example, for $N = M + 1$, $\mathbb{E}\{S\}$ is finite (as shown above) while the second moment $\mathbb{E}\{S^2\}$ (and, hence, the variance of S) is infinite.

C. Tail Exponents for LR-Based Preprocessing

Next, we analyze the tail behavior of the complexity distribution induced by LR-based preprocessing. We define LR-based preprocessing (see, e.g., [1]) as applying, prior to the QRD, the transformation $\mathbf{G} = \mathbf{HT}$, where $\mathbf{T}$ is an $M \times M$ unimodular matrix, i.e., $T_{i,j} \in \mathbb{CZ}$, $\forall i, j$, and $|\det(\mathbf{T})| = 1$. The matrix $\mathbf{T}$ can, in general, depend on $\mathbf{H}$ and is obtained, for example, through the LLL algorithm [29], which finds a basis matrix $\mathbf{G}$ of the lattice $\mathcal{L}(\mathbf{H})$ that is closer to an orthogonal matrix than $\mathbf{H}$. Another important preprocessing technique is LS (e.g., according to the V-BLAST algorithm [15]), which is just a special case of LR obtained by restricting $\mathbf{T}$ to be a permutation matrix (i.e., exactly one entry in each row and in each column of $\mathbf{T}$ is equal to one and all other entries are equal to zero).

The triangularized form of the CLP problem, when LR-based preprocessing is applied, is given by (5) with $\mathbf{R}$ and $\mathbf{y}$ replaced by $\widetilde{\mathbf{R}}$ and $\widetilde{\mathbf{y}} = \widetilde{\mathbf{Q}}^H\mathbf{r}$, respectively, where $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{R}}$ are the QR-factors of $\mathbf{G}$, i.e., $\mathbf{G} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}$. Consequently, $\mathbf{R}$ in Theorem 1 is now replaced by $\widetilde{\mathbf{R}}$. If we denote the corresponding solution of (5) as $\widetilde{\mathbf{d}}$, the final solution of (3) is $\widehat{\mathbf{d}} = \mathbf{T}\widetilde{\mathbf{d}}$.

1) Verifying the Conditions of Theorem 1 for LR-Based Preprocessing: Let us first verify the conditions of Theorem 1 for LR-based preprocessing of an i.i.d. Gaussian zero-mean lattice basis matrix $\mathbf{H}$. Condition (16) just depends on the statistics of $\mathbf{H}$ and was already verified in Appendix B-A for the more general class of lattice basis matrices that follow a correlated Ricean-fading distribution. To verify the remaining conditions, we start by noting that

$$\mathbf{QRT} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}. \tag{37}$$

Furthermore, let us write $\mathbf{RT} = \mathbf{Q}'\,\mathbf{R}'$, where $\mathbf{Q}'$ and $\mathbf{R}'$ are the QR-factors of $\mathbf{RT}$. With (37), we obtain $\mathbf{QQ}'\,\mathbf{R}' = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}$. Since $\mathbf{QQ}'$ is unitary and the QR-factors are unique, it follows that $\mathbf{QQ}' = \widetilde{\mathbf{Q}}$ and, hence, $\mathbf{R}' = \widetilde{\mathbf{R}}$, which results in

$$\mathbf{R} = \mathbf{Q}'\,\widetilde{\mathbf{R}}\mathbf{T}^{-1}. \tag{38}$$

Using (38) along with the facts that $\mathbf{T}^{-1}$ is unimodular (since $\mathbf{T}$ is unimodular) and $\mathbf{Q}'$ is unitary, it can be shown that

$$\mu(\mathbf{R}) = \mu(\widetilde{\mathbf{R}}) \tag{39}$$

and

$$\det(\mathbf{R}^H\mathbf{R}) = \det(\widetilde{\mathbf{R}}^H\widetilde{\mathbf{R}}). \tag{40}$$

Hence, LR preserves the covering radius as well as the volume of the fundamental regions associated with $\mathbf{R}$.

Since (17) and (19) hold for direct QRD (see Appendix B-B and Appendix B-C, respectively), it follows from (39) that conditions (17) and (19) are also satisfied for LR-based preprocessing. As condition (18) is satisfied for direct QRD (see Appendix B-C), it follows from (40) that condition (18) is satisfied for LR-based preprocessing for layer $k = M$. Hence, all the conditions of Theorem 1 are satisfied for LR-based preprocessing for $k = M$. As shown in Section III-C2 below, this will be sufficient to draw very general conclusions about the tail behavior of the total complexity distribution incurred by LR-based preprocessing.

For the layers $k < M$, condition (18) cannot be guaranteed if we allow $\mathbf{T}$ to be an arbitrary function of $\mathbf{H}$. It turns out, however, that all LR algorithms delivering a unimodular transformation matrix $\mathbf{T}$ that is invariant to a positive scaling of $\mathbf{H}$, i.e., $\mathbf{H}$ and $b\mathbf{H}$ for all $b \in \mathbb{R}$, $b > 0$, result in the same $\mathbf{T}$, satisfy condition (18) also for $k = 1, \ldots, M-1$. In this case, if $\widetilde{\mathbf{R}}$ denotes the lattice-reduced R-factor of $\mathbf{H}$, the lattice-reduced R-factor of $b\mathbf{H}$ is $b\widetilde{\mathbf{R}}$, as a consequence of the uniqueness of the QRD. Hence, $g_k(b\mathbf{H}) = \det(b^2\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k) = b^{2k}\det(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k) = b^{2k}g_k(\mathbf{H})$ and condition (18) is satisfied with $\alpha_k = 2k$ for $k = 1, \ldots, M$ (as is the case for direct QRD, see Appendix B-C). An LR algorithm with this property satisfies Theorem 1 for *all* $k = 1, \ldots, M$ and is referred to as a *standard* LR algorithm. All previously proposed LR algorithms in the literature that we are aware of are standard. Specifically, it can be shown, by inspection, that the LLL algorithm [29], the Seysen algorithm [30], LS according to the V-BLAST algorithm [15], and the sorted-QRD in [31] are all standard LR algorithms.

2) Tail Exponents for General LR-Based Preprocessing: As stated above, for general (i.e., not necessarily standard) LR-based preprocessing and $k = M$, the equivalence in (40) implies

$$\mathrm{P}[S_M \geq L] \doteq \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_M^H\mathbf{R}_M)} \geq L\right], \quad L \to \infty.$$

Combined with (27) this allows us to conclude that

$$\mathrm{P}[S_M \geq L] \doteq L^{-(N-M+1)}, \quad L \to \infty \tag{41}$$

or, equivalently, $\xi_M = N-M+1$ for general LR-based preprocessing. Consequently, (13) implies that the tail exponent of the total complexity distribution for general LR-based preprocessing satisfies

$$\xi \leq N - M + 1.$$

We can therefore conclude that *general LR-based preprocessing does not improve the tail exponent of the total complexity distribution* as compared to that obtained for direct QRD (cf. (28)). It is, however, important to note that LR-based preprocessing techniques typically *reduce* the computational complexity of SD, an aspect not reflected by our tail exponent result. In Section III-C4, we will see that the tail exponents associated with the layers $k < M$ can be improved by LR-based preprocessing using the LLL algorithm.

3) Tail Exponents with Layer-Sorting: Let us next specialize LR-based preprocessing to the case of standard LS[2], where $\mathbf{T}$ is a (potentially $\mathbf{H}$-dependent) permutation matrix that is invariant to positive scaling of $\mathbf{H}$. In this case, we can characterize the tail exponents of the corresponding total complexity and layer-wise complexity distributions precisely. As shown in Appendix E, for any LS strategy, we obtain

$$\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big) \leq \epsilon\Big] \doteq \epsilon^{N-M+1}, \quad \epsilon \to 0, \quad k = 1, \ldots, M$$

which is identical to the near-zero behavior obtained for direct QRD (see (26)). For standard LS strategies, Theorem 1 holds (see Section III-C1) and implies

$$\mathrm{P}\left[S_k \geq L\right] \doteq L^{-(N-M+1)}, \quad L \to \infty, \quad k = 1, \ldots, M. \tag{42}$$

Hence, $\xi_k = \xi = N - M + 1$, and we can conclude that *standard LS does not improve the layer complexity tail exponents* as compared to direct QRD (cf. (27)).

4) Tail Exponents with LLL: For LR based on the LLL algorithm, Theorem 1 implies that

$$\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\left[\frac{1}{\det(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k)} \geq L\right], \quad L \to \infty, \quad k = 1, \ldots, M. \tag{43}$$

Based on (43), it is shown in Appendix F that

$$\mathrm{P}[S_k \geq L] \dot{\leq} L^{-\frac{N}{k}}, \quad L \to \infty, \quad k = 1, \ldots, M \tag{44}$$

or, equivalently, $\xi_k \geq N/k$. Comparing with $\xi_k = N-M+1$ for direct QRD (cf. (27)) or standard LS (cf. (42)), we can conclude that LLL preprocessing improves the tail exponents at least up to layer $k \leq \lceil N/(N-M+1)\rceil - 1$. Hence, as compared to direct QRD or standard LS, LLL improves the tail exponents of the *lower* layers. However, we again note, as shown in Section III-C2, that the tail exponents of the total complexity distribution cannot be improved by LLL preprocessing. In the following, consider $N = M$. We have $\xi_k = M/k > 1$, $k = 1, \ldots, M-1$, and $\xi_M = 1$ (see (43) or (41)), which, together with (13), establishes that the tail behavior of the total complexity distribution of SD with LLL preprocessing is dominated by the tail behavior of the Mth layer complexity distribution; in particular, we have $\xi = \xi_M = 1$, as is the case with direct QRD and standard LS.

[2]LS according to the V-BLAST algorithm [15] and the sorted-QRD algorithm [31] fall into the category of standard LS algorithms.

D. Tail Exponents with Regularization

Next, we analyze the tail behavior of the regularized CLP problem

$$\overline{\mathbf{d}} \;=\; \underset{\mathbf{d}\in(\mathbb{CZ})^M}{\arg\,\min} \|\mathbf{z}-\mathbf{F}\mathbf{d}\|^2 \tag{45}$$

with the $(N+M)\times M$ regularized lattice basis matrix

$$\mathbf{F} = \begin{bmatrix} \mathbf{H} \\ \kappa\mathbf{I} \end{bmatrix} \tag{46}$$

where $\kappa \in \mathbb{R}$, $\kappa > 0$, is the regularization parameter (see, e.g., [8]) and $\mathbf{z}$ is the length-$(N+M)$ vector $\mathbf{z} = [\mathbf{r}^T\,\mathbf{0}^T]^T$. A regularization with a specific κ is, for example, obtained by minimum-mean square error (MMSE) based preprocessing (see, e.g., [6], [8], [22] or MMSE generalized decision-feedback equalization in [32]). We emphasize that the solution of the regularized CLP problem (45) will, in general, *not* equal the solution of the original CLP problem. It turns out, however, that for suitably chosen κ, regularization induces only a small performance loss while the resulting reduction in SD complexity can be significant [6], [8].

The triangularized form of the regularized CLP problem (45) is given by (5) with $\mathbf{R}$ and $\mathbf{y}$ replaced by $\overline{\mathbf{R}}$ and $\overline{\mathbf{y}} = \overline{\mathbf{Q}}^H\mathbf{z}$, respectively, where $\overline{\mathbf{Q}}$ and $\overline{\mathbf{R}}$ are the QR-factors of $\mathbf{F}$, i.e., $\mathbf{F} = \overline{\mathbf{Q}}\,\overline{\mathbf{R}}$. We again consider $\mathbf{H}$ to be zero-mean i.i.d. Gaussian and next verify that condition (17) for $\overline{\mathbf{R}}$ is satisfied. Condition (17), as shown in Appendix A-B, is sufficient to state the exponential upper bound

$$\mathrm{P}[S_k \geq L] \,\dot{\leq}\, \mathrm{P}\left[\frac{1}{\det\big(\overline{\mathbf{R}}_k^H\overline{\mathbf{R}}_k\big)} \geq L\right], \quad L\to\infty, \quad k=1,\ldots,M. \tag{47}$$

Following the steps leading from (61) to (62), we get

$$\mathrm{P}\big[\mu\big(\overline{\mathbf{R}}\big) \geq L\big] \leq \mathrm{P}[\|\mathbf{F}\|_{\mathrm{F}} \geq L]\,.$$

From (46), we have $\|\mathbf{F}\|_{\mathrm{F}}^2 = \|\mathbf{H}\|_{\mathrm{F}}^2 + \kappa^2 M$, which implies

$$\begin{aligned}\mathrm{P}\big[\mu\big(\overline{\mathbf{R}}\big) \geq L\big] &\leq \mathrm{P}\big[\|\mathbf{H}\|_{\mathrm{F}}^2 + \kappa^2 M \geq L^2\big] \\ &\doteq \mathrm{P}\big[\|\mathbf{H}\|_{\mathrm{F}} \geq L\big], \quad L\to\infty. \end{aligned} \tag{48}$$

As shown in Appendix B-B (cf. (62)–(66)), we have $\mathrm{P}\big[\|\mathbf{H}\|_{\mathrm{F}} \geq L\big] \,\dot{\leq}\, L^{-\infty}$, $L\to\infty$, for $\mathbf{H}$ correlated Gaussian distributed with arbitrary finite mean. Hence, (48) implies $\mathrm{P}\big[\mu\big(\overline{\mathbf{R}}\big) \geq L\big] \,\dot{\leq}$

$L^{-\infty}$ and condition (17) of Theorem 1 is satisfied for regularization-based preprocessing and i.i.d. Gaussian zero-mean $\mathbf{H}$, which proves (47).

Let us next analyze the RHS of (47). The interlacing theorem for bordered matrices [33, Theorem 4.3.8] implies that (cf. (77))

$$\lambda_i\Big(\overline{\mathbf{R}}_k^H\overline{\mathbf{R}}_k\Big) \geq \lambda_i\Big(\overline{\mathbf{R}}^H\overline{\mathbf{R}}\Big), \quad i=1,\ldots,k. \tag{49}$$

Furthermore, using (46), we can write $\lambda_i\Big(\overline{\mathbf{R}}^H\overline{\mathbf{R}}\Big) = \lambda_i\Big(\mathbf{F}^H\mathbf{F}\Big) = \lambda_i\big(\mathbf{H}^H\mathbf{H}\big) + \kappa^2$, which, together with (49), shows that $\lambda_i\Big(\overline{\mathbf{R}}_k^H\overline{\mathbf{R}}_k\Big) \geq \kappa^2$ and, consequently,

$$\det\big(\overline{\mathbf{R}}_k^H\overline{\mathbf{R}}_k\big) \geq \kappa^{2k} \tag{50}$$

i.e., $\det\big(\overline{\mathbf{R}}_k^H\overline{\mathbf{R}}_k\big)$ is strictly positive for any regularization parameter $\kappa > 0$. Inserting (50) into (47) therefore yields

$$\mathrm{P}[S_k \geq L] \doteq L^{-\infty}, \quad L\to\infty, \quad k=1,\ldots,M.$$

We can now conclude that the distributions of the individual layer and total complexities of SD with regularization decrease *faster than polynomial* in L. This is in stark contrast to direct QRD and LR preprocessing, which (as shown in the previous sections) have total complexity distributions that are of Pareto-type with tail exponent $N-M+1$. We conclude by noting that the performance degradation induced by regularization can be very small while the complexity improvements (as indicated by the faster-than-polynomial tail behavior) can be significant.

IV. Numerical Results

We consider SD for data detection in $N\times M$ MIMO wireless systems with spatial multiplexing, where $\mathbf{r} = \mathbf{H}\mathbf{d}' + \mathbf{w}$ (cf. (4)) with the entries of $\mathbf{H}$ and $\mathbf{w}$ assumed i.i.d. $\mathcal{CN}(0,1/M)$ and i.i.d. $\mathcal{CN}(0,\sigma^2)$, respectively, and with the transmitted vector $\mathbf{d}' \in (\mathbb{CZ})^M$. We note that the complexity of SD is random in $\mathbf{H}$ and $\mathbf{w}$ and does not depend on the particular $\mathbf{d}'$, which is due to the fact that any sums/differences of valid lattice points are again valid lattice points. Following the approach proposed in [4], [11], [21], the sphere radius ρ in (6) is chosen such that the transmitted data vector $\mathbf{d}'$ is found by the SD algorithm with probability 0.99. This is accomplished by setting

$$\mathrm{P}\big[\|\mathbf{w}\|_2 \leq \rho\big] = \gamma_N\left(\frac{\rho^2}{\sigma^2}\right) = 0.99$$

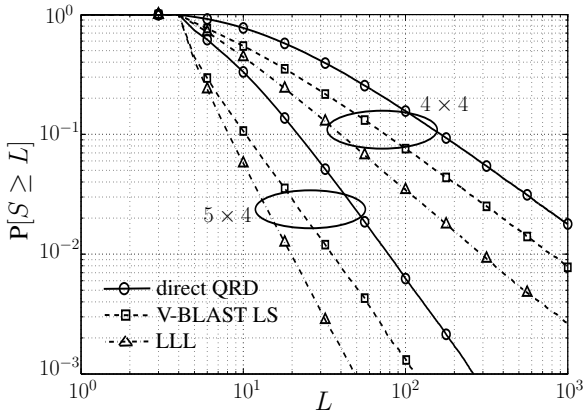

Fig. 1. Distribution of total complexity $\mathrm{P}[S \geq L]$ of SD with direct QRD, V-BLAST LS, and LLL preprocessing for 4×4 and 5×4 MIMO systems.

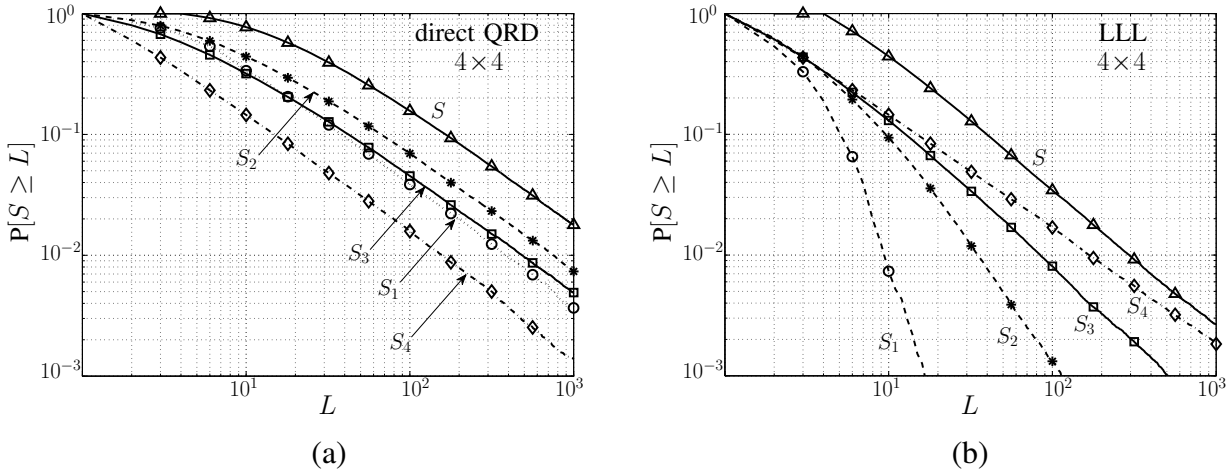

Fig. 2. Distributions of individual layer complexities $\mathrm{P}[S_k \geq L]$, $k = 1, 2, 3, 4$, and of total complexity $\mathrm{P}[S \geq L]$ of SD for a 4×4 MIMO system with (a) direct QRD and (b) LR-based preprocessing using the LLL algorithm.

and solving for ρ, i.e., $\rho = \sigma \sqrt{\gamma_N^{-1}(0.99)}$. For $1/\sigma^2$ we assume a value of $15\,\mathrm{dB}$, which results in $\rho \approx 0.5636$.

Fig. 1 shows the total complexity distribution $\mathrm{P}[S \geq L]$ in double log-scale for SD with direct QRD, V-BLAST LS [15], and with LLL preprocessing [34, with parameter $\delta = 3/4$] for 4×4 and 5×4 MIMO systems. We can draw the following conclusions:

- For direct QRD in the 4×4 case, the total complexity distribution in Fig. 1 exhibits a large-L behavior of L^{-1} as predicted by (28) for $N = M$.
- Adding one receive antenna improves the tail behavior significantly and leads to a large-L behavior of L^{-2} as predicted by (28).

- LLL preprocessing and V-BLAST LS reduce the complexity, as compared to direct QRD, but do not change the tail exponent of the total complexity distribution (see Section III-C).

For the same setup as in Fig. 1 for the 4×4 case, Fig. 2 shows the layer-wise complexity distributions $\mathrm{P}[S_k \geq L]$, $k = 1, 2, 3, 4$, and the total complexity distribution $\mathrm{P}[S \geq L]$ with (a) direct QRD and (b) LLL preprocessing, respectively. We can draw the following conclusions:

- For direct QRD, all layer-wise complexity distributions (see Fig. 2(a)) exhibit a large-L behavior of L^{-1} as predicted by (27) for $N = M$.
- With LLL preprocessing, the tail exponents of the layer-wise complexity distributions at the lower layers are improved as compared to that of direct QRD (compare Fig. 2(b) with Fig. 2(a)). However, the last layer (in this case layer 4) exhibits a large-L behavior of L^{-1} (see Fig. 2(b)) and, hence, dominates the tail behavior of the total complexity distribution (see (12) and (13)).

V. Conclusions

We analyzed the tail behavior of the (computational) complexity distribution of the sphere decoding (SD) algorithm in random infinite lattices. Our results complement and extend previous work that characterized the mean and the variance of SD complexity. In particular, we characterized the tail behavior of the complexity distribution in terms of corresponding tail exponents (i.e., polynomial decay rates). We found that the tail exponent of the SD complexity distribution is given by the tail exponent of the distribution of the inverse volume of the fundamental regions of the underlying lattice. This result was shown to hold under fairly general assumptions on SD preprocessing and on the statistics of the lattice basis matrix including, e.g., preprocessing based on lattice-reduction (LR) and the case of the lattice basis matrix being correlated Ricean distributed.

For $N \times M$ i.i.d. circularly symmetric complex Gaussian lattice basis matrices, we found that the complexity distribution of SD is of Pareto-type with tail exponent given by $N - M + 1$. This shows that increasing N (e.g., the number of receive antennas in the context of MIMO wireless communications) for given M (e.g., the number of transmit antennas) results in larger tail exponents. By means of a more refined analysis of the complexity distribution of SD, we also showed that the average complexity of SD is infinite for $N = M$ and finite for $N > M$, which complements average complexity results derived in the literature. We finally found that

the tail exponent of $N-M+1$ cannot be increased by preprocessing based on lattice-reduction including layer-sorting and the LLL algorithm as special cases while regularization results in a SD complexity distribution with tails that decrease faster than polynomial. We note, however, that lattice-reduction based preprocessing typically reduces the complexity of SD (although this is not reflected in the tail exponents of the complexity distributions) and that regularization-based preprocessing can be applied prior to lattice-reduction for further complexity reduction.

Throughout the paper, we considered the Fincke-Pohst variant of the SD algorithm with a fixed and lattice-independent choice of the sphere radius. The tools developed in this paper could turn out useful in analyzing the complexity distribution of more advanced SD approaches. Specifically, it would be interesting to understand the impact of a lattice-dependent choice of the sphere radius (as used in Schnorr-Euchner variants of the SD algorithm) on the tail behavior of the complexity distribution. Furthermore, most of our results do not capture multiplicative and additive constants that appear in the complexity distribution. A more refined analysis would clearly be desirable.

ACKNOWLEDGMENTS

The authors would like to thank C. Akçaba and P. Coronel for helpful discussions.

APPENDIX A
PROOF OF THEOREM 1

A. Exponential Lower Bound

For establishing the exponential lower bound, we use the lower bound on S_k given in (20), which, together with $\mu(\mathbf{R}_k) \leq \mu(\mathbf{R})$, $k=1,\ldots,M$, yields[3]

$$\mathrm{P}[S_k \geq L] \geq \mathrm{P}\left[\frac{V_k(\rho)-\mu(\mathbf{R})A_k(\rho)}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \geq L\right].$$

Next, consider a constant $c \in \mathbb{R}$, $c>0$, such that $V_k(\rho)-cA_k(\rho)>0$ and define $c' = V_k(\rho)-cA_k(\rho)>0$. We then have

$$\mathrm{P}[S_k \geq L] \geq \mathrm{P}\left[\mathbf{H} \in \mathcal{B}\right] \tag{51}$$

[3]Note that condition (17) implies full-rank $\mathbf{R}_k$ with probability one. Consequently, $\det(\mathbf{R}_k^H\mathbf{R}_k)>0$ and $\mu(\mathbf{R}_k)<\infty$ with probability one. However, it is straightforward to show that Theorem 1 also holds in the case where $\mathbf{R}_k$ is rank-deficient with non-zero probability, where we would have $\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\big[1/\det(\mathbf{R}_k^H\mathbf{R}_k) \geq L\big] \doteq L^0$, $L\to\infty$.

where

$$\mathcal{B} = \left\{ \mathbf{H} \colon \left(\frac{c'}{g_k(\mathbf{H})} \geq L \right) \cap \left(g_\mu(\mathbf{H}) \leq c \right) \right\}$$

with $\det(\mathbf{R}_k^H \mathbf{R}_k) = g_k(\mathbf{H})$ and $\mu(\mathbf{R}) = g_\mu(\mathbf{H})$. With property (16), we further obtain

$$\mathrm{P}[\mathbf{H} \in \mathcal{B}] = \int_{\mathbf{H} \in \mathcal{B}} f(\mathbf{H}) \mathrm{d}\mathbf{H} \geq \beta \int_{\mathbf{H} \in \mathcal{B}} f(L^\delta \mathbf{H}) \mathrm{d}\mathbf{H}$$

for all $\delta > 0$, $L > 1$, and some $\beta > 0$. Performing the change of variables $\mathbf{H}' = L^\delta \mathbf{H}$ and invoking conditions (18) and (19) yields

$$\mathrm{P}[\mathbf{H} \in \mathcal{B}] \geq \beta \, L^{-2MN\delta} \, \mathrm{P}[\mathbf{H} \in \mathcal{B}'] \tag{52}$$

where

$$\mathcal{B}' = \left\{ \mathbf{H} \colon \left(\frac{c'}{g_k(\mathbf{H})} \geq L^{1-\delta\alpha_k} \right) \cap \left(g_\mu(\mathbf{H}) \leq c L^{\delta\alpha} \right) \right\}.$$

Next, noting that for two events A_1 and A_2, by the inclusion-exclusion principle, $\mathrm{P}[A_1 \cap A_2] \geq \mathrm{P}[A_1] - \mathrm{P}[\bar{A}_2]$, where $\bar{A}_2$ denotes the complementary event of A_2, we get

$$\mathrm{P}[\mathbf{H} \in \mathcal{B}'] \geq \mathrm{P}\left[\frac{c'}{g_k(\mathbf{H})} \geq L^{1-\delta\alpha_k} \right] - \mathrm{P}\left[g_\mu(\mathbf{H}) > c L^{\delta\alpha} \right].$$

Now (17) with $\mu(\mathbf{R}) = g_\mu(\mathbf{H})$ and $\delta, \alpha > 0$ implies $\mathrm{P}\left[g_\mu(\mathbf{H}) > c L^{\delta\alpha} \right] \doteq L^{-\infty}$, $L \to \infty$, which, together with (51) and (52), yields

$$\mathrm{P}[S_k \geq L] \dot{\geq} L^{-2MN\delta} \, \mathrm{P}\left[\frac{c'}{g_k(\mathbf{H})} \geq L^{1-\delta\alpha_k} \right], \quad L \to \infty.$$

Let us write $\mathrm{P}[1/g_k(\mathbf{H}) \geq L] \doteq L^{-a}$, $L \to \infty$, for some constant $a \geq 0$. We then have $\mathrm{P}[S_k \geq L] \dot{\geq} L^{-2MN\delta - (1-\delta\alpha_k)a}$, $L \to \infty$. As this result holds for arbitrarily small values of δ, we can conclude that $\mathrm{P}[S_k \geq L] \dot{\geq} L^{-a} \doteq \mathrm{P}[1/g_k(\mathbf{H}) \geq L]$, $L \to \infty$, which establishes the exponential lower bound.

B. Exponential Upper Bound

To establish the exponential upper bound, we start from the upper bound on S_k in (20), which yields

$$\mathrm{P}[S_k \geq L] \leq \mathrm{P}\left[\frac{V_k(\rho + \mu(\mathbf{R}))}{\det(\mathbf{R}_k^H \mathbf{R}_k)} \geq L \right] \tag{53}$$

where, again, we used $\mu(\mathbf{R}_k) \leq \mu(\mathbf{R})$, $k = 1, \ldots, M$. Note that $\mathrm{P}[xy \geq L] = \mathrm{P}\big[(xy \geq L) \cap (y < L^\delta) \big] + \mathrm{P}\big[(xy \geq L) \cap (y \geq L^\delta) \big] \leq \mathrm{P}\big[x \geq L^{1-\delta} \big] + \mathrm{P}\big[y \geq L^\delta \big]$ for any two

RVs $x, y \in \mathbb{R}$ and any constant $\delta \in \mathbb{R}$, $0 < \delta < 1$. Applying this to (53) with $x = 1/\det(\mathbf{R}_k^H \mathbf{R}_k)$ and $y = V_k(\rho + \mu(\mathbf{R}))$, we get

$$\mathrm{P}[S_k \geq L] \leq \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H \mathbf{R}_k)} \geq L^{1-\delta}\right] + \mathrm{P}\big[V_k(\rho + \mu(\mathbf{R})) \geq L^{\delta}\big]. \tag{54}$$

With (21) and the binomial theorem, we can write

$$V_k\big(\rho + \mu(\mathbf{R})\big) = \frac{\pi^k}{k!} \sum_{i=0}^{2k} \binom{2k}{i} \rho^{2k-i} \mu(\mathbf{R})^i. \tag{55}$$

Next, applying the general property

$$\mathrm{P}\left[\sum_{i=1}^{M} x_i \geq L\right] \leq \sum_{i=1}^{M} \mathrm{P}\left[x_i \geq \frac{L}{M}\right] \tag{56}$$

which holds for any set of RVs x_i, $i = 1, \ldots, M$, together with (55) yields

$$\mathrm{P}\big[V_k(\rho + \mu(\mathbf{R})) \geq L^{\delta}\big] \dot{\leq} \sum_{i=0}^{2k} \mathrm{P}\big[\mu(\mathbf{R})^i \geq L^{\delta}\big], \quad L \to \infty.$$

Property (17) (for the terms corresponding to $i > 0$) and $\mathrm{P}[c'' \geq L] \doteq L^{-\infty}$, $L \to \infty$, for any constant $c'' \geq 0$ (for the term corresponding to $i = 0$) now directly imply $\mathrm{P}\big[V_k(\rho + \mu(\mathbf{R})) \geq L^{\delta}\big] \doteq L^{-\infty}$, $L \to \infty$, and, hence, by (54)

$$\mathrm{P}[S_k \geq L] \dot{\leq} \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H \mathbf{R}_k)} \geq L^{1-\delta}\right], \quad L \to \infty.$$

As before, writing $\mathrm{P}\big[1/\det(\mathbf{R}_k^H \mathbf{R}_k) \geq L\big] \doteq L^{-a}$, $L \to \infty$, for some constant $a \geq 0$, we get $\mathrm{P}[S_k \geq L] \dot{\leq} L^{-(1-\delta)a}$, $L \to \infty$. As this result holds for arbitrarily small values of δ, we can conclude that $\mathrm{P}[S_k \geq L] \dot{\leq} L^{-a} \doteq \mathrm{P}\big[1/\det(\mathbf{R}_k^H \mathbf{R}_k) \geq L\big]$, $L \to \infty$, which establishes the exponential upper bound.

APPENDIX B
THE CONDITIONS OF THEOREM 1 HOLD FOR CORRELATED RICEAN FADING AND DIRECT QRD

In the following, we show that conditions (16)–(19) are satisfied for direct QRD of a jointly complex Gaussian distributed lattice basis matrix $\mathbf{H}$ with arbitrary non-singular covariance matrix and arbitrary finite mean. With $\mathbf{h} = \mathrm{vec}(\mathbf{H}) \in \mathbb{C}^{NM}$, the pdf $f(\mathbf{h})$ of $\mathbf{h}$ is given by [35]

$$f(\mathbf{h}) = c_1 \, e^{-(\mathbf{h}-\boldsymbol{\mu})^H \mathbf{C}^{-1} (\mathbf{h}-\boldsymbol{\mu})} \tag{57}$$

where $\boldsymbol{\mu} = \mathbb{E}\{\mathbf{h}\}$, $\mathbf{C} = \mathbb{E}\{(\mathbf{h} - \boldsymbol{\mu})(\mathbf{h} - \boldsymbol{\mu})^H\}$, and $c_1 = 1/(\pi^{NM} \det(\mathbf{C}))$.

A. *Condition* (16)

We start by noting that condition (16) of Theorem 1 can equivalently be written as

$$f(\mathbf{h}) \geq \beta f(a\mathbf{h}) \tag{58}$$

for all $\mathbf{h} \in \mathbb{C}^{NM}$ and all $a \in \mathbb{R}$, $a > 1$, with some constant $\beta \in \mathbb{R}$, $\beta > 0$. Inserting (57) into (58) reveals that it suffices to show that

$$(\mathbf{h}-\boldsymbol{\mu})^H \mathbf{C}^{-1}(\mathbf{h}-\boldsymbol{\mu}) \leq (a\mathbf{h}-\boldsymbol{\mu})^H \mathbf{C}^{-1}(a\mathbf{h}-\boldsymbol{\mu}) + \beta' \tag{59}$$

for $\beta' = -\log(\beta)$. Reformulating (59) gives

$$\big((a+1)\mathbf{h}-\boldsymbol{\mu}\big)^H \mathbf{C}^{-1}\big((a+1)\mathbf{h}-\boldsymbol{\mu}\big) - \boldsymbol{\mu}^H\mathbf{C}^{-1}\boldsymbol{\mu} + \frac{a+1}{a-1}\beta' \geq 0. \tag{60}$$

Since the first term on the left hand side of (60) is always positive, it remains to show that there exists a constant β' (equivalently, a constant $\beta = e^{-\beta'}$) such that

$$\frac{a+1}{a-1}\beta' \geq \boldsymbol{\mu}^H\mathbf{C}^{-1}\boldsymbol{\mu}$$

for all $a \in \mathbb{R}$, $a > 1$. Indeed, since $(a+1)/(a-1) \geq 1$, any $\beta' \geq \boldsymbol{\mu}^H\mathbf{C}^{-1}\boldsymbol{\mu}$, or, equivalently, any $\beta \leq e^{-\boldsymbol{\mu}^H\mathbf{C}^{-1}\boldsymbol{\mu}}$ establishes condition (58), which concludes the proof.

B. *Condition* (17)

We first note that the squared covering radius $\mu^2(\mathbf{R})$ can be upper-bounded according to (see Appendix G and, for real-valued lattices, [20, Prop. 1])

$$\mu^2(\mathbf{R}) \leq \frac{1}{2}\sum_{i=1}^{M} R_{i,i}^2. \tag{61}$$

Further upper-bounding the RHS of (61) by $\sum_{i=1}^{M} R_{i,i}^2$ and noting that $R_{i,i}^2 \leq \|\mathbf{h}_i\|^2$ for direct QRD, we obtain $\mu^2(\mathbf{R}) \leq \|\mathbf{H}\|_{\mathrm{F}}^2$ and, consequently,

$$\mathrm{P}\left[\mu(\mathbf{R}) \geq L\right] \leq \mathrm{P}\big[\|\mathbf{H}\|_{\mathrm{F}} \geq L\big]. \tag{62}$$

Next, property (56) applied to the RHS of (62) results in

$$\mathrm{P}\left[\mu(\mathbf{R}) \geq L\right] \leq \sum_{i=1}^{N}\sum_{j=1}^{M} \mathrm{P}\left[|H_{i,j}| \geq \frac{L}{\sqrt{NM}}\right]. \tag{63}$$

Noting that in the case of Ricean fading, as assumed here, $H_{i,j} \stackrel{d}{=} \mu_{i,j} + \widetilde{H}_{i,j}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, where $\mu_{i,j} = \mathbb{E}\{H_{i,j}\}$ and $\widetilde{H}_{i,j} \sim \mathcal{CN}(0, \sigma_{i,j}^2)$, the terms on the RHS of (63) can be upper-bounded as

$$\mathrm{P}\left[|H_{i,j}| \geq \frac{L}{\sqrt{NM}}\right] = \mathrm{P}\left[|\mu_{i,j} + \widetilde{H}_{i,j}| \geq \frac{L}{\sqrt{NM}}\right] \leq \mathrm{P}\left[|\mu_{i,j}| + |\widetilde{H}_{i,j}| \geq \frac{L}{\sqrt{NM}}\right].$$

Again employing property (56), we get

$$\mathrm{P}\left[|H_{i,j}| \geq \frac{L}{\sqrt{NM}}\right] \leq \mathrm{P}\left[|\mu_{i,j}| \geq \frac{L}{2\sqrt{NM}}\right] + \mathrm{P}\left[|\widetilde{H}_{i,j}| \geq \frac{L}{2\sqrt{NM}}\right]. \tag{64}$$

Noting that $|\widetilde{H}_{i,j}|$ is χ_2-distributed, we obtain

$$\mathrm{P}\left[|\widetilde{H}_{i,j}| \geq \frac{L}{2\sqrt{NM}}\right] = e^{-\frac{L^2}{4NM\sigma_{i,j}^2}}. \tag{65}$$

Next, we define $\mu_{\max} = \max_{i,j}|\mu_{i,j}|$ (which, according to our finite-mean assumption, is finite) and $\sigma_{\max}^2 = \max_{i,j}\sigma_{i,j}^2$. For $L > 2\sqrt{NM}\,\mu_{\max}$, the first term on the RHS of (64) is zero, which together with (65) implies

$$\mathrm{P}\left[|H_{i,j}| \geq \frac{L}{\sqrt{NM}}\right] \leq e^{-\frac{L^2}{4NM\sigma_{\max}^2}}$$

for sufficiently large L. Hence, using (63), we obtain

$$\mathrm{P}\left[\mu(\mathbf{R}) \geq L\right] \leq NMe^{-\frac{L^2}{4NM\sigma_{\max}^2}} \tag{66}$$

for sufficiently large L. Finally, condition (17) is verified by noting that $e^{-\frac{L^2}{4NM\sigma_{\max}^2}} \doteq L^{-\infty}$, $L \to \infty$.

C. Conditions (18) and (19)

We start by noting that the R-factor of $b\mathbf{H}$ with $b \in \mathbb{R}$, $b > 0$, is given by $b\mathbf{R}$ as a direct consequence of the uniqueness of the QRD. We therefore get $g_k(b\mathbf{H}) = \det(b^2\mathbf{R}_k^H\mathbf{R}_k) = b^{2k}\det(\mathbf{R}_k^H\mathbf{R}_k) = b^{2k}g_k(\mathbf{H})$, which shows that condition (18) is satisfied with $\alpha_k = 2k$. Condition (19) can be verified as follows. By the definition of the covering radius (1), we can write

$$\begin{aligned}
\mu(b\mathbf{R}) &= \max_{\mathbf{x}\in\mathbb{C}^M} \min_{\mathbf{d}\in(\mathbb{CZ})^M} \|\mathbf{x} - b\mathbf{R}\mathbf{d}\| \\
&= b \max_{\mathbf{x}\in\mathbb{C}^M} \min_{\mathbf{d}\in(\mathbb{CZ})^M} \left\|\frac{\mathbf{x}}{b} - \mathbf{R}\mathbf{d}\right\| \\
&= b\,\mu(\mathbf{R})
\end{aligned}$$

implying that $g_\mu(b\mathbf{H}) = \mu(b\mathbf{R}) = b\,\mu(\mathbf{R}) = b\,g_\mu(\mathbf{H})$. This shows that condition (19) is satisfied with $\alpha = 1$.

APPENDIX C
THEOREM 1 FOR GENERAL SEARCH REGIONS

In the following, we show that Theorem 1 continues to hold if the search sphere around $\mathbf{y}_k$ is replaced by a general bounded search region $\mathcal{R}_k \subset \mathbb{C}^k$ with non-empty interior. Following (11), the complexity $\widetilde{S}_k$ of SD with search region $\mathcal{R}_k$ is given by

$$\widetilde{S}_k = \left|\left\{\mathbf{d}_k \in (\mathbb{C}\mathbb{Z})^k : (\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k) \in \mathcal{R}_k\right\}\right|.$$

The proof is based on separately establishing the exponential upper bound

$$\mathrm{P}\big[\widetilde{S}_k \geq L\big] \dot{\leq} \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \geq L\right], \quad L \to \infty \tag{67}$$

and the exponential lower bound

$$\mathrm{P}\big[\widetilde{S}_k \geq L\big] \dot{\geq} \mathrm{P}\left[\frac{1}{\det(\mathbf{R}_k^H\mathbf{R}_k)} \geq L\right], \quad L \to \infty \tag{68}$$

through, respectively, circumscribing and inscribing $\mathcal{R}_k$ by properly chosen hyperspheres.

We start by proving (67). Since $\mathcal{R}_k$ is bounded, it can be circumscribed by a hypersphere with finite radius $\rho^{(1)}$ implying that $\|\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k\| \leq \rho^{(1)}$ for all $(\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k) \in \mathcal{R}_k$. Consequently, if we denote by $S_k(\rho^{(1)}, \mathbf{y}_k)$ the complexity of conventional SD according to (11) with radius $\rho = \rho^{(1)}$ and center $\mathbf{y}_k$, we have $\widetilde{S}_k \leq S_k(\rho^{(1)}, \mathbf{y}_k)$ and

$$\mathrm{P}\big[\widetilde{S}_k \geq L\big] \leq \mathrm{P}\big[S_k(\rho^{(1)}, \mathbf{y}_k) \geq L\big]. \tag{69}$$

Next, we note that Theorem 1 does not depend on the particular choice of the sphere radius $\rho^{(1)}$. Hence, (15) (if the conditions of Theorem 1 are met) can be applied directly to the RHS of (69), which establishes the exponential upper bound (67). It remains to establish the corresponding exponential lower bound (68). Since $\mathcal{R}_k$ has non-empty interior, it can be inscribed by a hypersphere with strictly positive radius $\rho^{(2)}$ and any center $\mathbf{m} \in \mathbb{C}^k$ such that $\mathbf{y}_k - \mathbf{R}_k\mathbf{d}_k \in \mathcal{R}_k$ whenever $\|\mathbf{y}_k - \mathbf{m} - \mathbf{R}_k\mathbf{d}_k\| \leq \rho^{(2)}$. Consequently, $\widetilde{S}_k \geq S_k(\rho^{(2)}, \mathbf{y}_k - \mathbf{m})$, which results in

$$\mathrm{P}\big[\widetilde{S}_k \geq L\big] \geq \mathrm{P}\big[S_k(\rho^{(2)}, \mathbf{y}_k - \mathbf{m}) \geq L\big]. \tag{70}$$

Theorem 1 depends neither on the radius $\rho^{(2)}$ nor on the center $\mathbf{y}_k - \mathbf{m}$ of the search sphere and, hence, (15) (if the conditions of Theorem 1 on the statistics of the lattice basis matrix and on preprocessing are satisfied) can be applied directly to the RHS of (70), which establishes the exponential lower bound (68). The proof is concluded by combining (67) and (68).

APPENDIX D
NEAR-ZERO BEHAVIOR OF DET($\mathbf{R}_k^H\mathbf{R}_k$)

In the following, we show (26) for direct QRD and lattice basis matrices $\mathbf{H}$ whose entries are i.i.d. $\mathcal{CN}(0,\sigma_H^2)$. We first note that the nonzero entries of $\mathbf{R}$ are statistically independent with $\frac{\sqrt{2}}{\sigma_H}R_{i,i} \sim \chi_{2(N-i+1)}$ and $R_{i,j} \sim \mathcal{CN}(0,\sigma_H^2)$, for $i = 1,\ldots,M$, $j > i$ [27, Lemma 2.1]. Hence, the submatrix $\mathbf{R}_k$ has the same statistics as the matrix $\mathbf{R}_k'$, which would be obtained by QRD of a $(k+\Delta)\times k$, $\Delta = N-M$, matrix $\mathbf{H}_k$ having i.i.d. $\mathcal{CN}(0,\sigma_H^2)$ entries. This implies that the eigenvalues of $\mathbf{R}_k^H\mathbf{R}_k$ have the same statistics as the eigenvalues of $\mathbf{H}_k^H\mathbf{H}_k = \mathbf{R}_k'^H\mathbf{R}_k'$, i.e., $\lambda_i\big(\mathbf{R}_k^H\mathbf{R}_k\big) \stackrel{d}{=} \lambda_i\big(\mathbf{H}_k^H\mathbf{H}_k\big)$, $i = 1,\ldots,k$. We thus have

$$\mathrm{P}[\det(\mathbf{R}_k^H\mathbf{R}_k) \le \epsilon] = \mathrm{P}\left[\prod_{i=1}^{k}\lambda_i\big(\mathbf{H}_k^H\mathbf{H}_k\big) \le \epsilon\right] \tag{71}$$

so that the analysis of the near-zero behavior of $\det(\mathbf{R}_k^H\mathbf{R}_k)$ can be based on results in [24], which establish the near-zero behavior of the eigenvalues $\lambda_i\big(\mathbf{H}_k^H\mathbf{H}_k\big)$ for i.i.d. Gaussian zero-mean matrices $\mathbf{H}_k$.

Following [24], we start by performing the variable transformations $\lambda_i\big(\mathbf{H}_k^H\mathbf{H}_k\big) = \epsilon^{\alpha_i}$, $i = 1,\ldots,k$, and we assume $\epsilon < 1$ (note that this implies $\alpha_1 \ge \cdots \ge \alpha_k$). Setting $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_k)^T$, we can rewrite (71) as

$$\mathrm{P}[\det(\mathbf{R}_k^H\mathbf{R}_k) \le \epsilon] = \mathrm{P}[\boldsymbol{\alpha} \in \mathcal{B}] \tag{72}$$

where $\mathcal{B} = \{\boldsymbol{\alpha} : \alpha_1 + \cdots + \alpha_k \ge 1\}$. From [24, p. 1079], it follows that

$$\mathrm{P}[\det(\mathbf{R}_k^H\mathbf{R}_k) \le \epsilon] \doteq \mathrm{P}[\boldsymbol{\alpha} \in \mathcal{B}'], \quad \epsilon \to 0 \tag{73}$$

where

$$\mathcal{B}' = \{\boldsymbol{\alpha} : \alpha_1 + \cdots + \alpha_k \ge 1,\ \alpha_1 \ge \cdots \ge \alpha_k \ge 0\}.$$

Furthermore, from [24, p. 1080], we have

$$\mathrm{P}[\boldsymbol{\alpha} \in \mathcal{B}'] \doteq \epsilon^{\min_{\boldsymbol{\alpha}\in\mathcal{B}'}\sum_{i=1}^{k}(\Delta+2i-1)\alpha_i}, \quad \epsilon \to 0. \tag{74}$$

The minimum of $\sum_{i=1}^{k}(\Delta+2i-1)\alpha_i$ over $\boldsymbol{\alpha} \in \mathcal{B}'$ is achieved by setting $\alpha_1 = 1, \alpha_2 = 0,\ldots,\alpha_k = 0$, which results in

$$\min_{\boldsymbol{\alpha}\in\mathcal{B}'}\sum_{i=1}^{k}(\Delta+2i-1)\alpha_i = \Delta+1.$$

Combining this result with (73) and (74), and using $\Delta = N - M$, we obtain

$$\mathrm{P}[\det(\mathbf{R}_k^H \mathbf{R}_k) \leq \epsilon] \doteq \epsilon^{N-M+1}, \quad \epsilon \to 0 \tag{75}$$

which concludes the proof.

APPENDIX E
NEAR-ZERO BEHAVIOR OF $\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big)$ FOR LS

In the following, we show that

$$\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \leq \epsilon\Big] \doteq \epsilon^{N-M+1}, \quad \epsilon \to 0, \quad k = 1, \ldots, M \tag{76}$$

is satisfied for any LS strategy, i.e., if $\mathbf{T}$ is a permutation matrix, and for i.i.d. zero-mean Gaussian $\mathbf{H}$. Recall that $\widetilde{\mathbf{R}}_k$ refers to the $k \times k$ bottom right (upper triangular) submatrix of the R-factor $\widetilde{\mathbf{R}}$, obtained by QRD of $\mathbf{HT}$, i.e., $\mathbf{HT} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}$, where $\mathbf{T}$ is the (in general, $\mathbf{H}$-dependent) permutation matrix obtained by the LS algorithm. The proof of (76) will be accomplished by separately establishing the exponential upper bound $\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \leq \epsilon\Big] \dot{\leq} \epsilon^{N-M+1}$, $\epsilon \to 0$, and the exponential lower bound $\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \leq \epsilon\Big] \dot{\geq} \epsilon^{N-M+1}$, $\epsilon \to 0$, which then combine to (76). Both bounds are obtained by relating the near-zero behavior of $\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big)$ to the near-zero behavior of the smallest eigenvalue of $\mathbf{H}^H\mathbf{H}$.

A. Exponential Upper Bound

We first note that $\widetilde{\mathbf{R}}_k \widetilde{\mathbf{R}}_k^H$ is a principal submatrix of $\widetilde{\mathbf{R}}\widetilde{\mathbf{R}}^H$ due to the upper triangular structure of $\widetilde{\mathbf{R}}$. Furthermore, we have that $\lambda_i\big(\widetilde{\mathbf{R}}_k \widetilde{\mathbf{R}}_k^H\big) = \lambda_i\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big)$ and $\lambda_i\big(\widetilde{\mathbf{R}}\widetilde{\mathbf{R}}^H\big) = \lambda_i\big(\widetilde{\mathbf{R}}^H \widetilde{\mathbf{R}}\big)$, which, together with the interlacing theorem for bordered matrices [33, Theorem 4.3.8], implies that

$$\lambda_i\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \geq \lambda_i\big(\widetilde{\mathbf{R}}^H \widetilde{\mathbf{R}}\big), \quad i = 1, \ldots, k. \tag{77}$$

Due to $\widetilde{\mathbf{R}}^H \widetilde{\mathbf{R}} = \mathbf{T}^H \mathbf{H}^H \mathbf{H} \mathbf{T}$ (recall that $\mathbf{HT} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}$) with $\mathbf{T}$ being a permutation matrix and hence unitary, $\lambda_i\big(\widetilde{\mathbf{R}}^H \widetilde{\mathbf{R}}\big) = \lambda_i\big(\mathbf{H}^H \mathbf{H}\big)$, $i = 1, \ldots, M$. Together with (77), we obtain $\lambda_i\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \geq \lambda_i\big(\mathbf{H}^H \mathbf{H}\big)$, $i = 1, \ldots, k$, which results in

$$\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \geq \prod_{i=1}^{k} \lambda_i\big(\mathbf{H}^H \mathbf{H}\big).$$

We can therefore conclude that

$$\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le\epsilon\Big]\le\mathrm{P}\left[\prod_{i=1}^{k}\lambda_i\big(\mathbf{H}^H\mathbf{H}\big)\le\epsilon\right]. \tag{78}$$

Similar to Appendix D (cf. (72)–(75)), the near-zero behavior of the RHS can be analyzed by means of the results in [24]. Specifically, we obtain

$$\mathrm{P}\left[\prod_{i=1}^{k}\lambda_i\big(\mathbf{H}^H\mathbf{H}\big)\le\epsilon\right]\doteq\epsilon^{N-M+1},\quad\epsilon\to 0$$

which, together with (78), establishes the desired exponential upper bound.

B. Exponential Lower Bound

Evidently, we have

$$\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le\epsilon\Big]\ge\mathrm{P}\Big[\big(\lambda_1\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le\epsilon\big)\cap\big(\lambda_k\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le 1\big)\Big]. \tag{79}$$

We next derive a sufficient condition for the event

$$\big(\lambda_1\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le\epsilon\big)\cap\big(\lambda_k\big(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\big)\le 1\big) \tag{80}$$

which will imply a lower bound on the RHS of (79). This lower bound will not depend on the particular LS strategy employed, since it will be given in terms of the eigenvalues $\lambda_i(\mathbf{H}^H\mathbf{H})$, $i=1,\ldots,M$, and the corresponding eigenvectors $\mathbf{u}_i$ of $\mathbf{H}^H\mathbf{H}$.

From the decomposition

$$\big(\mathbf{H}^H\mathbf{H}\big)^{-1}=\sum_{i=1}^{M}\frac{1}{\lambda_i(\mathbf{H}^H\mathbf{H})}\mathbf{u}_i\mathbf{u}_i^H$$

we get

$$\Big[\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\Big]_{m,m}\ge\frac{1}{\lambda_1(\mathbf{H}^H\mathbf{H})}|u_{1,m}|^2$$

where $|u_{1,m}|^2=\big[\mathbf{u}_1\mathbf{u}_1^H\big]_{m,m}$. Let us now consider the event

$$\lambda_1(\mathbf{H}^H\mathbf{H})\le\nu\epsilon\quad\text{and}\quad|u_{1,m}|^2>\nu,\quad m=1,\ldots,M \tag{81}$$

for some constant $\nu<1/M$. We obtain

$$\Big[\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\Big]_{m,m}\ge\frac{1}{\epsilon},\quad m=1,\ldots,M. \tag{82}$$

The diagonal elements of $\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$ can be related to the diagonal elements of $\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}$ as follows. Let us first write $\widetilde{\mathbf{R}}$ as

$$\widetilde{\mathbf{R}} = \begin{bmatrix} \widetilde{\mathbf{R}}_{\bar{k}} & \mathbf{A} \\ \mathbf{0} & \widetilde{\mathbf{R}}_k \end{bmatrix}$$

where $\widetilde{\mathbf{R}}_{\bar{k}}$ is an $(M-k)\times(M-k)$ upper triangular matrix and $\mathbf{A}$ is an $(M-k)\times k$ full matrix. For the inverse of this partitioned matrix we obtain [33, Section 0.7.3]

$$\widetilde{\mathbf{R}}^{-1} = \begin{bmatrix} \widetilde{\mathbf{R}}_{\bar{k}}^{-1} & -\widetilde{\mathbf{R}}_{\bar{k}}^{-1}\mathbf{A}\widetilde{\mathbf{R}}_k^{-1} \\ \mathbf{0} & \widetilde{\mathbf{R}}_k^{-1} \end{bmatrix} \tag{83}$$

which implies that $\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}$ is a principal submatrix of $\left(\widetilde{\mathbf{R}}^H\widetilde{\mathbf{R}}\right)^{-1}$ obtained by deleting the first $M-k$ rows and columns of $\left(\widetilde{\mathbf{R}}^H\widetilde{\mathbf{R}}\right)^{-1}$. Since $(\mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})^{-1} = \left(\widetilde{\mathbf{R}}^H\widetilde{\mathbf{R}}\right)^{-1}$, $\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}$ is also a principal submatrix of $\mathbf{T}^{-1}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{T}^{-H}$. In particular, this results in

$$\left[\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}\right]_{i,i} = \left[\mathbf{T}^{-1}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{T}^{-H}\right]_{M-k+i,M-k+i}, \quad i=1,\ldots,k. \tag{84}$$

Furthermore, since $\mathbf{T}$ is a permutation matrix, the matrix $\mathbf{T}^{-1}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{T}^{-H}$ has the same diagonal elements (just at different positions) as the matrix $(\mathbf{H}^H\mathbf{H})^{-1}$. Therefore, (84) together with (82) implies that

$$\left[\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}\right]_{i,i} \geq \frac{1}{\epsilon}, \quad i=1,\ldots,k$$

which results in

$$\lambda_1\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right) \leq \epsilon \tag{85}$$

upon using

$$\frac{1}{\lambda_1\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)} \geq \left[\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)^{-1}\right]_{i,i}, \quad i=1,\ldots,k.$$

Consequently, the events (81) imply (85), which corresponds to the first event in (80). It now remains to establish a sufficient event for the second event in (80), i.e., for $\lambda_k\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right) \leq 1$. From the interlacing theorem for bordered matrices [33, Theorem 4.3.8] applied to $\mathbf{H}^H\mathbf{H}$ and $\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k$ (see Section E-A), we have $\lambda_M(\mathbf{H}^H\mathbf{H}) \geq \lambda_k\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right)$. Consequently, $\lambda_M(\mathbf{H}^H\mathbf{H}) \leq 1$ implies $\lambda_k\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right) \leq 1$. With (79) and all the established sufficient events for (80), we obtain

$$\mathrm{P}\Big[\det\left(\widetilde{\mathbf{R}}_k^H\widetilde{\mathbf{R}}_k\right) \leq \epsilon\Big] \geq \mathrm{P}\big[\left(\lambda_1(\mathbf{H}^H\mathbf{H}) \leq \nu\epsilon\right) \cap \left(\lambda_M\left(\mathbf{H}^H\mathbf{H}\right) \leq 1\right)\big]\, \mathrm{P}\big[|u_{1,m}|^2 > \nu, \forall m\big]$$

where we used the fact that the eigenvectors $\mathbf{u}_i$, $i=1,\ldots,M$, are statistically independent of the eigenvalues $\lambda_i(\mathbf{H}^H\mathbf{H})$, $i=1,\ldots,M$, [27, Lemma 2.6]. Since $\mathbf{u}_1$ is uniformly distributed

on the unit sphere [27, Lemma 2.6], we have $\mathrm{P}[|u_{1,m}|^2 > \nu, \forall m] > 0$ for any $\nu < 1/M$. This finally implies

$$\begin{aligned}\mathrm{P}\Big[\det\big(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k\big) \leq \epsilon\Big] &\dot{\geq} \mathrm{P}\big[\big(\lambda_1(\mathbf{H}^H\mathbf{H}) \leq \epsilon\big) \cap \big(\lambda_M\big(\mathbf{H}^H\mathbf{H}\big) \leq 1\big)\big], \quad \epsilon \to 0 \\ &\doteq \epsilon^{N-M+1}, \quad \epsilon \to 0\end{aligned}$$

where the exponential equality again follows directly from [24, p. 1080].

APPENDIX F
EXPONENTIAL UPPER BOUND FOR LLL

In the following, starting from

$$\mathrm{P}[S_k \geq L] \doteq \mathrm{P}\left[\frac{1}{\det(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k)} \geq L\right], \quad L \to \infty \tag{86}$$

we prove (44) for LR-based preprocessing using the LLL algorithm. In this case, $\widetilde{\mathbf{R}}$ is obtained by QRD of $\mathbf{G}$, i.e., $\mathbf{G} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{R}}$, with $\mathbf{G}$ denoting the LLL-reduced basis associated with $\mathbf{H}$. The diagonal elements of $\widetilde{\mathbf{R}}$ satisfy the well-known LLL conditions [29], [34]

$$\widetilde{R}_{m,m}^2 \geq c\, \widetilde{R}_{m-1,m-1}^2, \quad m = 2, \ldots, M \tag{87}$$

for some constant $c > 0$ (more precisely, $0 < c < 0.5$), which gives

$$\widetilde{R}_{m,m}^2 \geq c^{m-1}\, \widetilde{R}_{1,1}^2, \quad m = 1, \ldots, M. \tag{88}$$

Based on (88), we can now lower-bound

$$\det(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k) = \prod_{i=1}^{k} \widetilde{R}_{M-i+1,M-i+1}^2$$

according to

$$\det(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k) \geq c'\, \widetilde{R}_{1,1}^{2k} \tag{89}$$

where $c' = c^a$ with $a = \sum_{i=1}^{k}(M-i)$. Furthermore, since $\widetilde{\mathbf{R}}$ is upper triangular, $\widetilde{R}_{1,1}$ is the Euclidean norm of a basis vector associated with $\mathcal{L}(\widetilde{\mathbf{R}})$. Hence, $\widetilde{R}_{1,1} \geq \gamma(\widetilde{\mathbf{R}})$, where $\gamma(\widetilde{\mathbf{R}})$ denotes the length of the shortest nonzero vector in $\mathcal{L}(\widetilde{\mathbf{R}})$, i.e., $\gamma(\widetilde{\mathbf{R}}) = \min_{\mathbf{d} \in (\mathbb{CZ})^M, \mathbf{d} \neq \mathbf{0}} \|\widetilde{\mathbf{R}}\mathbf{d}\|$. From (38) we can conclude that $\gamma(\widetilde{\mathbf{R}}) = \gamma(\mathbf{R})$. With $\gamma(\mathbf{R}) = \gamma(\mathbf{H})$, we obtain $\widetilde{R}_{1,1} \geq \gamma(\mathbf{H})$, which, upon using (89), results in

$$\det(\widetilde{\mathbf{R}}_k^H \widetilde{\mathbf{R}}_k) \geq c'\, \gamma^{2k}(\mathbf{H}). \tag{90}$$

Inserting (90) into the RHS of (86) yields

$$\mathrm{P}[S_k \geq L] \dot{\leq} \mathrm{P}\big[\gamma^{2k}(\mathbf{H}) \leq L^{-1}\big], \quad L \to \infty. \tag{91}$$

From [36, Lemma 3] we have

$$\mathrm{P}\big[\gamma(\mathbf{H}) \leq L^{-1}\big] \leq \begin{cases} c'' L^{-2N}, & \text{for } M < N \\ c'' L^{-2N} \max\{-(-\ln L)^{N+1}, 1\}, & \text{for } M = N \end{cases}$$

with some constant c'', which, together with (91), establishes the final result (44).

APPENDIX G
UPPER BOUND (61) ON $\mu^2(\mathbf{R})$

We first define $\widehat{\mathbf{d}}_{\text{SIC}}$ as the detection result obtained by successive interference cancelation (SIC) (see, e.g., [31]). Noting that $\mathbf{d}_k = [d_{M-k+1}\, \mathbf{d}_{k-1}^T]^T$, the components of $\widehat{\mathbf{d}}_{\text{SIC}}$ are obtained by solving M scalar minimization problems according to

$$\widehat{d}_{\text{SIC},M-k+1} = \underset{d_{M-k+1} \in \mathbb{CZ}}{\arg\min} \left| \Delta_k \Big(\big[d_{M-k+1}\, \widehat{\mathbf{d}}_{\text{SIC},k-1}^T \big]^T \Big) \right|^2, \quad k = 1, \ldots, M \tag{92}$$

starting with $k = 1$, where the metric update $|\Delta_k(\mathbf{d}_k)|^2$ is defined in (8). Since $\widehat{\mathbf{d}}_{\text{SIC}}$ is a suboptimum solution to the CLP problem (5), we have

$$\mu^2(\mathbf{R}) = \max_{\mathbf{y} \in \mathbb{C}^M} \min_{\mathbf{d} \in (\mathbb{CZ})^M} \|\mathbf{y} - \mathbf{R}\mathbf{d}\|^2 \leq \max_{\mathbf{y} \in \mathbb{C}^M} \|\mathbf{y} - \mathbf{R}\widehat{\mathbf{d}}_{\text{SIC}}\|^2. \tag{93}$$

The distance $\|\mathbf{y} - \mathbf{R}\widehat{\mathbf{d}}_{\text{SIC}}\|^2$ achieved by the SIC detector can be further upper-bounded as follows. Since (92) corresponds to a simple scalar minimization problem over Gaussian integers that are scaled by $R_{M-k+1,M-k+1}$, we obtain

$$\big|\Delta_k\big(\widehat{\mathbf{d}}_{\text{SIC},k}\big)\big|^2 \leq \frac{1}{2} R_{M-k+1,M-k+1}^2, \quad k = 1, \ldots, M$$

for any $\mathbf{y}$. Consequently, we have

$$\begin{aligned} \|\mathbf{y} - \mathbf{R}\widehat{\mathbf{d}}_{\text{SIC}}\|^2 &= \sum_{k=1}^{M} \big|\Delta_k\big(\widehat{\mathbf{d}}_{\text{SIC},k}\big)\big|^2 \\ &\leq \frac{1}{2} \sum_{k=1}^{M} R_{M-k+1,M-k+1}^2 \end{aligned}$$

which, together with (93), establishes (61) and hence concludes the proof.

REFERENCES

[1] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[2] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comp.*, vol. 44, pp. 463–471, Apr. 1985.

[3] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," in *Proc. ICCS/ISITA 1992*, vol. 1, Singapore, Nov. 1992, pp. 127–131.

[4] E. Viterbo and E. Biglieri, "A universal decoding algorithm for lattice codes," in *GRETSI 14-ème Colloq.*, Juan-les-Pins, France, Sept. 1993, pp. 611–614.

[5] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1639–1642, July 1999.

[6] M. O. Damen, H. El 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. 2389–2402, Oct. 2003.

[7] 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. of Solid-State Circuits*, vol. 40, no. 7, pp. 1566–1577, July 2005.

[8] C. Studer, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Algorithms and VLSI implementation," *IEEE J. Sel. Areas Comm.*, vol. 26, no. 2, pp. 290–300, Feb. 2008.

[9] J. Savage, "The distribution of the sequential decoding computation time," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 143–147, Apr. 1966.

[10] I. M. Jacobs and E. R. Berlekamp, "A lower bound to the distribution of computation for sequential decoding," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 167–174, April 1967.

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

[12] H. Vikalo and B. Hassibi, "On the sphere decoding algorithm II. Generalizations, second-order statistics, and applications to communications," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2819–2834, Aug. 2005.

[13] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[14] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.

[15] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, pp. 14–16, Jan. 1999.

[16] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Berlin, Heidelberg, New York: Springer, 1988.

[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[18] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933–953, Mar. 2006.

[19] C. Studer, D. Seethaler, and H. Bölcskei, "Finite lattice-size effects in MIMO detection," in *Proc. 42nd Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA*, Oct. 2008, pp. 2032–2037.

[20] A. H. Banihashemi and A. K. Khandani, "On the complexity of decoding lattices using the Korkin-Zolotarev reduced basis," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 162–171, Jan. 1998.

[21] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[22] R. Böhnke, D. Wübben, V. Kühn, and K. D. Kammeyer, "Reduced complexity MMSE detection for BLAST architectures," in *Proc. IEEE Globecom 2003*, vol. 4, San Francisco, CA, Dec. 2003, pp. 2258–2262.

[23] P. M. Gruber and J. M. Wills, Eds., *Handbook of Convex Geometry*. vol. B, North Holland, Amsterdam: Elsevier, 1993.

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

[25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Boston (MA): Cambridge University Press, 2005.

[26] Y. Jiang, X. Zheng, and J. Li, "Asymptotic performance analysis of V-BLAST," in *Proc. IEEE Globecom 2005*, vol. 6, St. Louis, MO, Nov. 2005, pp. 3882–3886.

[27] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA: Now Publishers Inc. 2004.

[28] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965.

[29] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.

[30] M. Seysen, "Simultaneous reduction of a lattice basis and its reciprocal basis," *Combinatorica*, vol. 13, pp. 363–376, 1993.

[31] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K. D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *Electronics Letters*, vol. 37, no. 22, pp. 1348–1350, Oct. 2001.

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

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge (UK): Cambridge Univ. Press, 1993.

[34] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity MIMO detection," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2701–2710, July 2009.

[35] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Hoboken, New Jersey: Wiley, 2005.

[36] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4567–4582, Dec. 2007.