

Finite Lattice-Size Effects in MIMO Detection

C. Studer*, D. Seethaler[‡], and H. Bölcskei[‡]

*Integrated Systems Laboratory
ETH Zurich, 8092 Zurich, Switzerland
e-mail: studer@iis.ee.ethz.ch

[‡]Communication Technology Laboratory
ETH Zurich, 8092 Zurich, Switzerland
e-mail: {seethal,boelcskei}@nari.ee.ethz.ch

Abstract—Many powerful data detection algorithms employed in multiple-input multiple-output (MIMO) communication systems, such as sphere decoding (SD) and lattice-reduction (LR)-aided detection, were initially designed for infinite lattices. Detection in MIMO systems is, however, based on finite lattices. In this paper, we systematically study the consequences of finite lattice-size for the performance and complexity of MIMO detection algorithms formulated for infinite lattices. Specifically, we find, considering performance and complexity, that LR does not seem to offer advantages when used in conjunction with SD.

I. INTRODUCTION

The computational complexity associated with maximum-likelihood (ML) detection (MLD) in multiple-input multiple-output (MIMO) communication systems poses significant challenges for corresponding VLSI implementations [1]. Promising approaches for complexity reduction make explicit use of the structure in the problem by transforming it into an equivalent lattice decoding¹ problem, where the underlying lattice is finite. Some of the most prominent techniques for efficient lattice decoding were, however, developed for infinite lattices, e.g., sphere decoding (SD) [2] or lattice-reduction (LR)-aided detection, e.g., [3], [4].

The goal of this paper is to investigate the fundamental differences, in terms of complexity and performance, between MIMO detection on finite and on infinite lattices. We show that relaxation of the MIMO detection problem to infinite lattices followed by LR can reduce the complexity of SD at the cost of a significant performance loss in terms of error rate. This performance loss is caused by symbol estimates that do not belong to the finite lattice and hence require remapping (onto the finite lattice). We demonstrate that (near-)ML achieving remapping strategies entail a complexity which is comparable to that of SD operating directly on the finite lattice, thus rendering LR in conjunction with SD unattractive for practical applications.

Notation: Matrices are set in boldface capital letters, vectors in boldface lowercase letters. The superscripts T and H stand for transpose and conjugate transpose, respectively. We write $A_{i,j}$ for the entry in the i th row and j th column of the matrix \mathbf{A} , b_i for the i th entry of the vector $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_N]^T$, and we denote the ℓ^2 -norm of $\mathbf{b} \in \mathbb{C}^N$ as $\|\mathbf{b}\|$. \mathbf{I}_N stands for the $N \times N$ identity matrix and

This work was supported in part by the STREP project No. IST-026905 (MASCOT) within the Sixth Framework Programme (FP6) of the European Commission.

¹In the context of this paper it would be more appropriate to use the term “lattice detection”.

$\mathbf{1}_N$ denotes the N -dimensional all-ones vector. $|\mathcal{O}|$ designates the cardinality of the set \mathcal{O} and $\mathbb{C}\mathbb{Z}$ stands for the set of Gaussian integers, i.e., $\mathbb{C}\mathbb{Z} = \mathbb{Z} + \sqrt{-1}\mathbb{Z}$. The real and imaginary part of $x \in \mathbb{C}$ is denoted by $\Re\{x\}$ and $\Im\{x\}$, respectively.

II. MIMO DETECTION AS LATTICE DECODING

Consider a coherent MIMO system, i.e., the receiver has perfect channel state information (CSI) while the transmitter does not have CSI, with M transmit and M receive antennas². The bit-stream to be transmitted is mapped to M -dimensional transmit symbol vectors $\mathbf{s} \in \mathcal{O}^M$, where \mathcal{O} corresponds to the underlying scalar QAM constellation. The associated complex baseband input-output relation is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (1)$$

where \mathbf{H} stands for the $M \times M$ channel matrix, \mathbf{y} is the M -dimensional received signal vector, and \mathbf{n} is an i.i.d. circularly symmetric complex Gaussian distributed noise vector of dimension M with variance N_o per complex entry. The signal-to-noise ratio is defined as $\text{SNR} = E_s/N_o$, where E_s denotes the average signal power per receive antenna.

A. Transformation to Lattices

In order to see that the MIMO detection problem is equivalent to a lattice decoding problem, we start by mapping the elements $s \in \mathcal{O}$ to elements $x \in \mathbb{C}\mathbb{Z}$ using the transformation $x = as + c$. The constants $a, c \in \mathbb{R}$ with $a > 0$ and $c > 0$ are independent of s and are chosen such that $x \in \mathcal{X} \subset \mathbb{C}\mathbb{Z}$ with $|\mathcal{X}| = |\mathcal{O}|$ and

$$\mathcal{X} = \{x \in \mathbb{C}\mathbb{Z} \mid (k_{\min} \leq \Re\{x\} \leq k_{\max}) \wedge (k_{\min} \leq \Im\{x\} \leq k_{\max})\} \quad (2)$$

where $k_{\min}, k_{\max} \in \mathbb{Z}$. Note that (2) can be used for square QAM constellations to check whether $x' \in \mathbb{C}\mathbb{Z}$ is in \mathcal{X} by performing separate boundary checks for the real and imaginary part of x' . In the case of non-square QAM constellations the boundary checks take a slightly more complicated form. The transmit vectors $\mathbf{s} \in \mathcal{O}^M$ can be mapped to vectors $\mathbf{x} \in \mathcal{X}^M \subset (\mathbb{C}\mathbb{Z})^M$ according to

$$\mathbf{x} = \mathbf{a}\mathbf{s} + \mathbf{c} \quad (3)$$

²The restriction to the number of transmit antennas being equal to the number of receive antennas is made for simplicity of exposition. The results are also valid for the number of receive antennas being larger than the number of transmit antennas.

where $\mathbf{c} = c\mathbf{1}_M$. The inverse transformation associated with (3) is given by $\mathbf{s} = a^{-1}(\mathbf{x} - \mathbf{c})$. The input-output relation (1) can now be transformed into

$$\mathbf{r} = \mathbf{G}\mathbf{x} + \mathbf{n} \quad (4)$$

where $\mathbf{G} = a^{-1}\mathbf{H}$ and $\mathbf{r} = \mathbf{y} + \mathbf{G}\mathbf{c}$ is a translated version of the received vector \mathbf{y} . The essence of the transformation of (1) into (4) is that now the received vector \mathbf{r} can be interpreted as a lattice point $\mathbf{u} \in \mathcal{L}(\mathbf{G})$ that has been translated by the additive Gaussian noise vector \mathbf{n} . Here,

$$\mathcal{L}(\mathbf{G}) \triangleq \left\{ \mathbf{G}\mathbf{x} \mid \mathbf{x} \in \mathcal{X}^M \right\} \quad (5)$$

denotes the finite lattice generated by \mathbf{G} . In the remainder of the paper, we shall work with the input-output relation (4) exclusively.

B. ML Detection

MLD in MIMO systems computes the estimate

$$\hat{\mathbf{u}}^{\text{ML}} = D^{\text{ML}}(\mathbf{r}) = \arg \min_{\mathbf{u} \in \mathcal{L}(\mathbf{G})} \|\mathbf{r} - \mathbf{u}\| \quad (6)$$

which amounts to solving a closest-vector problem (CVP) in the finite lattice $\mathcal{L}(\mathbf{G})$. Since each lattice point in $\mathcal{L}(\mathbf{G})$ is associated with a transmit vector in \mathcal{X}^M according to the relation $\mathbf{u} = \mathbf{G}\mathbf{x}$, the ML-estimate $\hat{\mathbf{u}}^{\text{ML}}$ obtained by solving (6) can be transformed into³ $\hat{\mathbf{x}}^{\text{ML}} = \mathbf{G}^{-1}\hat{\mathbf{u}}^{\text{ML}}$, which upon inversion of (3) yields the ML-estimate $\hat{\mathbf{s}}^{\text{ML}}$. Solving (6) through an exhaustive search over all lattice points in $\mathcal{L}(\mathbf{G})$ typically results in high computational complexity as $|\mathcal{L}(\mathbf{G})| = |\mathcal{X}|^M$ may be large.

C. Relaxation and Lattice Reduction

A promising approach to reducing the computational complexity associated with solving (6) is to relax the finite lattice in (5) to the infinite lattice⁴

$$\underline{\mathcal{L}}(\mathbf{G}) \triangleq \left\{ \mathbf{G}\underline{\mathbf{x}} \mid \underline{\mathbf{x}} \in (\mathbb{C}\mathbb{Z})^M \right\} \quad (7)$$

and to solve a relaxed version of (6) by computing

$$\underline{\hat{\mathbf{u}}}^{\text{ML}} = \underline{D}^{\text{ML}}(\mathbf{r}) = \arg \min_{\underline{\mathbf{u}} \in \underline{\mathcal{L}}(\mathbf{G})} \|\mathbf{r} - \underline{\mathbf{u}}\| \quad (8)$$

i.e., by searching over the infinite lattice generated by \mathbf{G} . This search can be carried out more efficiently by using algorithms from lattice theory, e.g., SD [2] possibly in conjunction with LR through the LLL algorithm [5]. As the resulting estimate $\underline{\hat{\mathbf{u}}}^{\text{ML}}$ is not necessarily in $\mathcal{L}(\mathbf{G})$, remapping of $\underline{\hat{\mathbf{u}}}^{\text{ML}}$ onto the finite lattice $\mathcal{L}(\mathbf{G})$ is required whenever $\underline{\hat{\mathbf{u}}}^{\text{ML}} \notin \mathcal{L}(\mathbf{G})$. In the remainder of the paper, the term ‘‘MLD’’ always refers to the finite-lattice MLD problem (6), whereas ‘‘relaxed MLD’’ is used to refer to the infinite-lattice MLD problem (8).

³Note that we implicitly assume that \mathbf{G} has full rank; this is satisfied, for example, with probability 1, if the entries of \mathbf{G} are i.i.d. circularly symmetric complex Gaussian distributed. Furthermore, in practice \mathbf{G}^{-1} does not have to be computed explicitly as \mathbf{u} in (6) can be replaced by $\mathbf{G}\mathbf{x}$ and the minimization can be performed over $\mathbf{x} \in \mathcal{X}^M$.

⁴In the remainder of the paper, underlined quantities always refer to the infinite-lattice case.

The main motivation for relaxation resides in the fact that LR can be applied only to infinite lattices. LR computes an equivalent and ‘‘more orthogonal’’ basis (for the lattice $\underline{\mathcal{L}}(\mathbf{G})$) with the generator matrix $\mathbf{B} = \mathbf{G}\mathbf{T}$, where \mathbf{T} is an $M \times M$ unimodular matrix, i.e., $|\det(\mathbf{T})| = 1$ with $T_{i,j} \in \mathbb{C}\mathbb{Z}$ ($\forall i, j$). Thanks to the unimodularity of \mathbf{T} , we have $\underline{\mathcal{L}}(\mathbf{B}) = \underline{\mathcal{L}}(\mathbf{G})$. It is important to note that this equivalence holds only for infinite lattices, in contrast to finite lattices where $\mathcal{L}(\mathbf{B}) \neq \mathcal{L}(\mathbf{G})$, in general. We therefore conclude that relaxation to an infinite lattice is essential for the application of LR to MIMO detection.

We emphasize that LR, in general, reduces the complexity of a subsequent SD step and results in (often significant) performance improvements when followed by sub-optimal detectors. In particular, in [6] it is shown that LR followed by linear detection achieves full diversity.

D. Schnorr-Euchner Sphere Decoding

SD was initially designed for the efficient solution of CVPs on infinite lattices [2], but has also proven very efficient for MLD [7], [8]. In the following, we briefly review the main ingredients of Schnorr-Euchner SD (SESD) with radius reduction for the finite-lattice case [9].

The algorithm starts by performing a QR decomposition (QRD) of \mathbf{G} according to $\mathbf{G} = \mathbf{Q}\mathbf{R}$, where the $M \times M$ matrix \mathbf{Q} is unitary and the upper-triangular $M \times M$ matrix \mathbf{R} has real-valued non-negative entries on its main diagonal. Left-multiplying (4) by \mathbf{Q}^H leads to the modified input-output relation $\tilde{\mathbf{r}} = \mathbf{R}\mathbf{x} + \tilde{\mathbf{n}}$, where $\tilde{\mathbf{r}} = \mathbf{Q}^H\mathbf{r}$ and $\tilde{\mathbf{n}} = \mathbf{Q}^H\mathbf{n}$ is again (thanks to \mathbf{Q} being unitary) i.i.d. circularly symmetric complex Gaussian distributed with variance N_o per complex entry. We arrange the partial symbol vectors (PSVs) $\mathbf{x}^{(i)} = [x_i \ x_{i+1} \ \dots \ x_M]^T$ in a tree that has its root just above level $i = M$ and leaves on level $i = 1$; the leaves correspond to vectors $\mathbf{x} \in \mathcal{X}^M$. Each PSV is associated with a partial distance (PD) according to

$$d(\mathbf{x}^{(i)}) = \|\tilde{\mathbf{r}}^{(i)} - \mathbf{R}_i\mathbf{x}^{(i)}\|, \quad i = 1, 2, \dots, M \quad (9)$$

where $\tilde{\mathbf{r}}^{(i)} = [\tilde{r}_i \ \tilde{r}_{i+1} \ \dots \ \tilde{r}_M]^T$ and \mathbf{R}_i contains the lower-right $(M - i + 1) \times (M - i + 1)$ -dimensional sub-block of \mathbf{R} . The MLD problem has therefore been transformed into a weighted tree-search problem, where PSVs and PDs are associated with nodes in the tree. For brevity, we shall often say ‘‘the node $\mathbf{x}^{(i)}$ ’’ meaning the node corresponding to the PSV $\mathbf{x}^{(i)}$. Each path through the tree (from the root node down to a leaf node) corresponds to a symbol vector $\mathbf{x} \in \mathcal{X}^M$. Note that in the infinite-lattice case, the tree-search problem needs to be relaxed to vectors $\underline{\mathbf{x}} \in (\mathbb{C}\mathbb{Z})^M$. The solution of (6) is given by the leaf associated with the smallest metric.

The essence of SD is to constrain the search in the tree through a sphere constraint (SC), realized by searching only those nodes that satisfy $d(\mathbf{x}^{(i)}) \leq r$, where r is the search radius associated with a hypersphere centered in the received point. SESD refers to conducting the tree search in a depth-first manner and visiting the children of a given node in ascending order of their PDs [9]. Since the PDs $d(\mathbf{x}^{(i)})$ are non-decreasing as a function of $i = M, M - 1, \dots, 1$, a node $\mathbf{x}^{(i)}$ violating the SC can be pruned along with

the entire subtree emanating from that node. The algorithm initializes $r = \infty$ and performs the update $r \leftarrow d(\mathbf{x})$ whenever a valid leaf node \mathbf{x} has been found. This approach is known in the literature as radius reduction [1], [9].

The complexity measure employed in this paper is given by the number of nodes visited by SEDS in the tree search, including the leaf nodes, but excluding the root node. This complexity measure is directly related to the VLSI implementation complexity for the finite-lattice case⁵ [1]. The complexity required for preprocessing (i.e., QRD or LR) will be ignored throughout the paper. We note, however, that in latency-critical applications, the preprocessing complexity can be critical.

In the remainder of the paper, “SESD”, “relaxed SEDS”, and “LR-aided SEDS” refer to SEDS on finite lattices, on infinite lattices, and to LR followed by SEDS on infinite lattices, respectively.

III. COMPLEXITY OF SEDS AND OF RELAXED SEDS

We next analyze the differences in the complexity behavior of SEDS and of relaxed SEDS.

A. Low-SNR Regime

Finite-lattice case: SEDS visits a node $\mathbf{x}^{(i)}$ if the SC $d(\mathbf{x}^{(i)}) \leq r$ is satisfied. Since the minimum search radius of SEDS (guaranteeing that the corresponding hypersphere contains at least one lattice point) is given by the Euclidean distance between the received vector and the ML solution, we can conclude that SEDS visits at least all nodes $\mathbf{x}^{(i)}$ satisfying

$$d(\mathbf{x}^{(i)}) \leq \min_{\mathbf{x} \in \mathcal{X}^M} \|\tilde{\mathbf{r}} - \mathbf{R}\mathbf{x}\| = r_{\min}. \quad (10)$$

Thanks to radius reduction, the final (with respect to the repeated tree-traversals as described above) search radius of SEDS corresponds to r_{\min} . We can therefore conclude that the genie-aided choice $r = r_{\min}$ for the search radius initialization, indeed, yields a lower bound on the number of nodes visited by SEDS. Denoting the transmitted data vector as $\mathbf{x}' \in \mathcal{X}^M$, we obtain a lower bound on r_{\min} as follows:

$$\begin{aligned} r_{\min} &\stackrel{\text{a)}}{=} \min_{\mathbf{x} \in \mathcal{X}^M} \|\tilde{\mathbf{n}} + \mathbf{R}(\mathbf{x}' - \mathbf{x})\| \geq \\ &\stackrel{\text{b)}}{\geq} \min_{\mathbf{x} \in \mathcal{X}^M} \left(\|\tilde{\mathbf{n}}\| - \|\mathbf{R}(\mathbf{x} - \mathbf{x}')\| \right) = \\ &= \|\tilde{\mathbf{n}}\| - \max_{\mathbf{x} \in \mathcal{X}^M} \|\mathbf{R}(\mathbf{x} - \mathbf{x}')\| \geq \\ &\stackrel{\text{c)}}{\geq} \|\tilde{\mathbf{n}}\| - \sigma_{\max}(\mathbf{R})K \end{aligned} \quad (11)$$

where a) results from $\tilde{\mathbf{r}} = \mathbf{R}\mathbf{x}' + \tilde{\mathbf{n}}$, b) follows from the inverse triangle inequality, and c) is a consequence of the Rayleigh-Ritz theorem [10] ($\sigma_{\max}(\mathbf{R})$ denotes the largest singular value of \mathbf{R}) and the fact that the constant $K = \max_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}^M} \|\mathbf{x} - \mathbf{x}'\|$ is finite in the finite-lattice case. From (11) we can infer that SEDS visits at least all nodes $\mathbf{x}^{(i)}$ satisfying

$$d(\mathbf{x}^{(i)}) \leq \|\tilde{\mathbf{n}}\| - \sigma_{\max}(\mathbf{R})K. \quad (12)$$

⁵We expect this complexity measure to be equally relevant for the infinite-lattice case.

Note that setting the search radius equal to the right-hand side (RHS) of (12) will no longer guarantee that the corresponding hypersphere contains at least one lattice point. This is, however, irrelevant for the point we want to make next.

Defining $\mathbf{v} = \tilde{\mathbf{n}}/\sqrt{N_o}$, we can rewrite (12) as

$$\left\| \mathbf{v}^{(i)} + \frac{\mathbf{R}_i(\mathbf{x}'^{(i)} - \mathbf{x}^{(i)})}{\sqrt{N_o}} \right\| \leq \|\mathbf{v}\| - \frac{\sigma_{\max}(\mathbf{R})K}{\sqrt{N_o}}. \quad (13)$$

We can now see that for $N_o \rightarrow \infty$ and hence for $\text{SNR} \rightarrow 0$, the condition in (13) reduces to $\|\mathbf{v}^{(i)}\| \leq \|\mathbf{v}\|$, which is trivially satisfied for $i > 1$ for every realization of $\tilde{\mathbf{n}}$ and for given \mathbf{R} . In this case, SEDS visits at least all nodes down to and including the level just above the leaf level. An intuitive explanation for this effect is that the noise can shift the transmit vector arbitrarily far away from the finite lattice and hence, the RHS of (11) and consequently r_{\min} can become arbitrarily large, which inhibits efficient tree pruning.

Infinite-lattice case: The fundamental difference between the complexity of SEDS and of relaxed SEDS is due to the fact that in the infinite-lattice case, the maximum distance between the received vector \mathbf{r} and the nearest lattice point is bounded for a given (full rank) \mathbf{R} . The complexity of relaxed SEDS can therefore be lower than that of SEDS. The downside is that relaxed SEDS is not guaranteed to find the ML solution, which necessarily has to be a point in the finite lattice $\mathcal{L}(\mathbf{G})$. Specifically, if relaxed SEDS delivers a lattice point outside $\mathcal{L}(\mathbf{G})$, this point needs to be remapped onto the finite lattice. Corresponding remapping approaches are described in Section IV. In the following, we derive an analytic upper bound on the complexity of relaxed SEDS.

The number of nodes $\underline{\mathbf{x}}^{(i)}$ on tree level i that satisfy the SC with radius r is denoted by $\underline{\mathcal{C}}_i(r)$ and corresponds to the number of lattice points in $\underline{\mathcal{L}}(\mathbf{R}_i)$ that are within a $2(M - i + 1)$ -dimensional (real-valued) hypersphere of radius r centered at $\mathbf{r}^{(i)}$. This number can be upper-bounded by dividing the volume of a $2(M - i + 1)$ -dimensional hypersphere of radius $r + \mu_i$ by the volume of a Voronoi cell in $\underline{\mathcal{L}}(\mathbf{R}_i)$ [11], where μ_i corresponds to the covering radius of the lattice $\underline{\mathcal{L}}(\mathbf{R}_i)$ [12]. More specifically, we have [13]

$$\underline{\mathcal{C}}_i(r) \leq \frac{V_{(M-i+1)}(r + \mu_i)}{\text{Vol}(\underline{\mathcal{L}}(\mathbf{R}_i))} \quad (14)$$

where $V_k(r) = \frac{\pi^k}{k!} r^{2k}$ denotes the volume of a $2k$ -dimensional hypersphere with radius r and the volume of a Voronoi cell of $\underline{\mathcal{L}}(\mathbf{R}_i)$ is $\text{Vol}(\underline{\mathcal{L}}(\mathbf{R}_i)) = \det(\mathbf{R}_i^H \mathbf{R}_i)$.

In order to derive an upper bound on the total complexity $\underline{\mathcal{C}} = \sum_{i=1}^M \underline{\mathcal{C}}_i(r)$ of relaxed SEDS, we consider the case where only the first radius update is performed. The first leaf node found by relaxed SEDS corresponds to the Babai point $\hat{\underline{\mathbf{x}}}^{\text{B}}$ [14] and hence the radius after the first update from the initial value $r = \infty$ is given by $d(\hat{\underline{\mathbf{x}}}^{\text{B}})$, which can be upper-bounded as [9], [14]

$$d(\hat{\underline{\mathbf{x}}}^{\text{B}}) = \|\tilde{\mathbf{r}} - \mathbf{R}\hat{\underline{\mathbf{x}}}^{\text{B}}\| \leq \sqrt{\frac{1}{2} \sum_{j=1}^M R_{j,j}^2} = \beta. \quad (15)$$

Since the covering radius μ_1 is also upper-bounded by β [11], [15] and, more generally, μ_i can be upper-bounded

as $\mu_i \leq \sqrt{\frac{1}{2} \sum_{j=i}^M R_{j,j}^2} = \gamma_i$, we have $r + \mu_i \leq \beta + \gamma_i$, which yields an upper bound on the total complexity of relaxed SESD according to

$$\underline{C} \leq \sum_{i=1}^M \frac{V_{(M-i+1)}(\beta + \gamma_i)}{\det(\mathbf{R}_i^H \mathbf{R}_i)}. \quad (16)$$

B. High-SNR Regime

In the high-SNR regime, relaxed SESD is likely to return the ML solution. Moreover, if $\|\tilde{\mathbf{n}}\|$ is sufficiently small, relaxed SESD and SESD find the same leaf nodes, which will belong to the finite lattice $\mathcal{L}(\mathbf{G})$. Consequently, relaxed SESD and SESD will operate with the same search radii and hence the same SCs. However, the two detectors will, in general, not exhibit the same complexity, as the SESD additionally (to the SC) takes into account the finite alphabet nature of $\mathcal{X} \subset \mathbb{C}\mathbb{Z}$, whereas relaxed SESD visits *all* points in $\mathbb{C}\mathbb{Z}$ satisfying the SC. Therefore, in the high-SNR regime, SESD tends to result in smaller complexity than relaxed SESD.

C. Numerical Complexity Assessment

A simple numerical result serves to demonstrate the potential complexity savings of relaxed SESD over SESD in the low-SNR regime. We assume a MIMO channel with $\mathbf{R} = \mathbf{I}_M$ and $M = 4$. Counting all the nodes down to and including the level just above the leaf level and noting that only one leaf node is visited by SESD if \mathbf{R} is a diagonal matrix, leads to the corresponding complexity

$$C = \frac{|\mathcal{X}|^M - 1}{|\mathcal{X}| - 1}$$

of SESD for $\text{SNR} \rightarrow 0$ (see the discussion in the last paragraph of Section III-A). For a 16-QAM alphabet, this leads to $C = 4369$ nodes. The total complexity of relaxed SESD is upper-bounded by (16), which can be refined by noting that relaxed SESD visits only one leaf node in the case of $\mathbf{R} = \mathbf{I}_M$. Consequently, we have

$$\underline{C} \leq 1 + \sum_{i=2}^M V_{(M-i+1)} \left(\sqrt{\frac{M}{2}} + \sqrt{\frac{M-i+1}{2}} \right)$$

which shows that at most 1928 nodes will be visited, irrespectively of the SNR. We can therefore conclude that relaxation can result in a significant complexity reduction at low SNR values.

Fig. 1 shows the complexity⁶ corresponding to SESD (which yields ML performance) and to relaxed SESD (denoted by R-SESD). Consistent with the observations made in Sections III-A and III-B, we see that relaxation leads to complexity savings in the low-SNR regime, but results in higher complexity in the high-SNR regime. We can furthermore see

⁶All simulation results are for an uncoded $M = 4$ MIMO system with 16-QAM symbol constellation using Gray labeling. The entries of \mathbf{H} are i.i.d. circularly symmetric complex Gaussian distributed with unit variance. For (relaxed) SESD, we use sorted QR-decomposition (SQRD) according to [16], for LR, a complex-valued variant of the LLL algorithm with $\delta = 3/4$ and SQRD preprocessing as described in [17], [18]. All complexity and performance simulations are averaged over 640,000 channel realizations and for each channel realization a single noise realization has been generated.

that relaxation followed by LR (denoted by LR-SESD) leads to lower complexity than SESD for all SNR values.

The complexity reductions of R-SESD and LR-SESD over SESD come, however, at a significant performance loss in terms of error-rate and in the case of LR-SESD also at increased computational complexity in the preprocessing stage caused by the need for LR (realized through the LLL algorithm [5], for example). For naive lattice decoding, i.e., estimates $\hat{\mathbf{u}}^{\text{ML}} \notin \mathcal{L}(\mathbf{G})$ are simply discarded and an error is declared, Fig. 2 shows a 3 dB SNR loss of LR-SESD compared to SESD (MLD). We emphasize that the performance of R-SESD, not shown in Fig. 2, is identical to that of LR-SESD; the associated complexities are, however, different (cf. Fig. 1). Note that naive lattice decoding was shown to achieve full diversity, while leading to an unbounded SNR gap (growing logarithmically in SNR) for MIMO systems with an equal number of transmit and receive antennas [19]. In the next section, we discuss several approaches to the mitigation of this performance loss, i.e., we consider algorithms for remapping estimates $\hat{\mathbf{u}}^{\text{ML}} \notin \mathcal{L}(\mathbf{G})$ onto $\mathcal{L}(\mathbf{G})$.

IV. REMAPPING METHODS

If the estimate delivered by relaxed SESD satisfies $\hat{\mathbf{u}}^{\text{ML}} \in \mathcal{L}(\mathbf{G})$, the lattice point $\hat{\mathbf{u}}^{\text{ML}}$ corresponds to the ML solution (6), i.e., $\hat{\mathbf{u}}^{\text{ML}} = \hat{\mathbf{u}}^{\text{ML}}$. If $\hat{\mathbf{u}}^{\text{ML}} \notin \mathcal{L}(\mathbf{G})$ and the performance of naive lattice decoding is not sufficient, the solution delivered by relaxed SESD needs to be remapped onto the finite lattice $\mathcal{L}(\mathbf{G})$. Since all constellation points $\mathbf{x} \in \mathcal{X}^M$ (corresponding to elements $\mathbf{u} \in \mathcal{L}(\mathbf{G})$ through the relation⁷ $\mathbf{u} = \mathbf{G}\mathbf{x}$) satisfy (2), determining whether $\hat{\mathbf{u}}^{\text{ML}}$ is in the finite lattice $\mathcal{L}(\mathbf{G})$ can be performed efficiently through element-wise boundary checks of the real and imaginary parts of $\hat{\mathbf{x}}^{\text{ML}}$. In the following, we briefly review an existing remapping method known as quantization and describe novel remapping approaches that realize different performance/complexity trade-offs.

A. Quantization

The standard remapping approach known in the literature is quantization and amounts to remapping $\hat{\mathbf{x}}^{\text{ML}} \notin \mathcal{X}^M$ to the closest (in Euclidean distance) symbol vector $\hat{\mathbf{x}}_q \in \mathcal{X}^M$ according to [4], [18]

$$\hat{\mathbf{x}}_q = \arg \min_{\mathbf{x} \in \mathcal{X}^M} \|\hat{\mathbf{x}}^{\text{ML}} - \mathbf{x}\|. \quad (17)$$

As quantization corresponds to element-wise slicing of the entries of $\hat{\mathbf{x}}^{\text{ML}}$ to the nearest candidates in \mathcal{X} , it requires low computational complexity. The resulting estimate $\hat{\mathbf{x}}_q$ does, however, not necessarily correspond to the ML solution (6) and hence leads to a (potentially significant) performance loss.

Fig. 2 shows the performance obtained by LR-aided SESD followed by remapping through quantization (denoted by LR-SESD, quant.). The numerical results indicate a slight improvement over naive lattice decoding, but still show a

⁷In the remainder of the paper, we treat vectors $\mathbf{x} \in (\mathbb{C}\mathbb{Z})^M$ and $\mathbf{u} \in \mathcal{L}(\mathbf{G})$ as well as $\mathbf{x} \in \mathcal{X}^M$ and $\mathbf{u} \in \mathcal{L}(\mathbf{G})$ as interchangeable as they are related through the one-to-one mappings $\mathbf{u} = \mathbf{G}\mathbf{x}$ and $\mathbf{u} = \mathbf{G}\mathbf{x}$, respectively. Of course, this assumes that \mathbf{G} has full rank.

2.5 dB SNR loss compared to ML performance. Finally, Fig. 2 shows the performance obtained by LR-aided successive interference cancellation (SIC) on the infinite lattice with remapping through naive lattice decoding (denoted by LR-SIC, naive). The resulting performance loss, compared to relaxed SEDS with remapping through naive lattice decoding, is low (approximately 0.25 dB SNR), whereas the complexity of LR-SIC is significantly lower than that of LR-SESD.

B. Optimal Lattice Remapping

The discussion above brings out the importance of remapping. It is therefore natural to ask for the optimum lattice remapping strategy in the sense of maximizing the probability of the result of remapping to correspond to the transmitted symbol vector given the result of the relaxed detector $\hat{\mathbf{u}}$. Let $\mathbf{u}' \in \mathcal{L}(\mathbf{G})$ be the transmitted vector and assume that the relaxed detector yields

$$\underline{D}(\mathbf{r}) = \underline{D}(\mathbf{u}' + \mathbf{n}) = \hat{\mathbf{u}}, \quad \hat{\mathbf{u}} \in \mathcal{L}(\mathbf{G}) \setminus \mathcal{L}(\mathbf{G}).$$

Optimal lattice remapping according to the criterion described above now amounts to identifying the lattice point $\mathbf{u} \in \mathcal{L}(\mathbf{G})$ which maximizes $P[\mathbf{u}' = \mathbf{u} | \underline{D}(\mathbf{r}) = \hat{\mathbf{u}}]$, i.e.,

$$\hat{\mathbf{u}}_{\text{opt}} = \arg \max_{\mathbf{u} \in \mathcal{L}(\mathbf{G})} P[\mathbf{u}' = \mathbf{u} | \underline{D}(\mathbf{r}) = \hat{\mathbf{u}}].$$

We emphasize that the optimum remapping rule depends on $\underline{D}(\mathbf{r})$ and hence, on the relaxed MIMO detector employed. Using Bayes's theorem and assuming that all transmit vectors are equally likely, we get

$$\hat{\mathbf{u}}_{\text{opt}} = \arg \max_{\mathbf{u} \in \mathcal{L}(\mathbf{G})} P[\underline{D}(\mathbf{r}) = \hat{\mathbf{u}} | \mathbf{u}' = \mathbf{u}] \quad (18)$$

which amounts to solving a finite-lattice MLD problem (6) based on the relaxed estimate $\hat{\mathbf{u}}$ instead of the received vector \mathbf{r} . We stress that relaxed MLD followed by optimum lattice remapping according to (18) will, in general, not achieve ML performance as remapping is conducted based on $\hat{\mathbf{u}}$ rather than on \mathbf{r} . For the sake of simplicity of exposition, in the following, we write $P[\underline{D}(\mathbf{r}) = \hat{\mathbf{u}} | \mathbf{u}]$ instead of $P[\underline{D}(\mathbf{r}) = \hat{\mathbf{u}} | \mathbf{u}' = \mathbf{u}]$.

Closest-vector remapping: The optimal remapping rule in (18) can, in general, not be expressed in closed form, as it depends on the decision region of the relaxed MIMO detector. We can, however, obtain a simple remapping strategy that achieves near-optimal performance by computing an approximation of (18) as follows. We start by noting that

$$P[\underline{D}(\mathbf{r}) = \hat{\mathbf{u}} | \mathbf{u}] = \int_{\mathcal{D}(\hat{\mathbf{u}})} f_G(\mathbf{w} - \mathbf{u}) d\mathbf{w} \quad (19)$$

where

$$f_G(\mathbf{w}) = (\pi N_o)^{-M} \exp(-N_o^{-1} \|\mathbf{w}\|^2)$$

is the joint probability density function of a multi-variate complex Gaussian with i.i.d. circularly symmetric components each of which has variance N_o and

$$\mathcal{D}(\hat{\mathbf{u}}) = \left\{ \mathbf{w} \in \mathbb{C}^M \mid \underline{D}(\mathbf{w}) = \hat{\mathbf{u}} \right\}$$

denotes the decision region of the relaxed MIMO detector employed. Next, we consider optimal remapping for relaxed

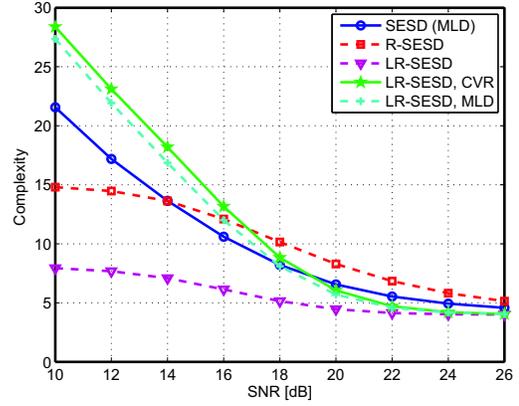


Fig. 1. Complexity of SEDS (MLD), relaxed SEDS (R-SESD), and LR-aided SEDS (LR-SESD).

MLD, i.e., $\underline{D}(\mathbf{r}) = \underline{D}^{\text{ML}}(\mathbf{r})$ as defined in (8) and $\hat{\mathbf{u}} = \hat{\mathbf{u}}^{\text{ML}}$, and note that in this case (19) formally corresponds to the probability of mistakenly decoding the transmitted point \mathbf{u} for $\hat{\mathbf{u}}^{\text{ML}}$. This probability can be upper-bounded as

$$P[\underline{D}^{\text{ML}}(\mathbf{r}) = \hat{\mathbf{u}}^{\text{ML}} | \mathbf{u}] \leq Q\left(\frac{\|\hat{\mathbf{u}}^{\text{ML}} - \mathbf{u}\|}{\sqrt{2N_o}}\right) \quad (20)$$

where $Q(a) = (2\pi)^{-\frac{1}{2}} \int_a^\infty \exp(-\frac{x^2}{2}) dx$. Replacing the function to be maximized over in (18) by the upper bound in (20) leads to a novel remapping rule, which we refer to as closest-vector remapping (CVR) and which is given by

$$\hat{\mathbf{u}}_{\text{cv}} = \arg \min_{\mathbf{u} \in \mathcal{L}(\mathbf{G})} \|\hat{\mathbf{u}}^{\text{ML}} - \mathbf{u}\|. \quad (21)$$

The solution of (21) corresponds to the point $\hat{\mathbf{u}}_{\text{cv}} \in \mathcal{L}(\mathbf{G})$ that is closest (in Euclidean distance) to the lattice point $\hat{\mathbf{u}}^{\text{ML}} \in \mathcal{L}(\mathbf{G}) \setminus \mathcal{L}(\mathbf{G})$. We emphasize that CVR amounts to solving a finite-lattice CVP and can, hence, be carried out by (non-relaxed) SEDS with the relaxed ML-estimate $\hat{\mathbf{u}}^{\text{ML}}$ taking the role of the received point. Note that CVR according to (21) exhibits structural similarities to quantization according to (17). Formally, quantization can be seen as CVR with $\mathbf{G} = \mathbf{I}_M$, which amounts to ignoring the structure of the lattice generated by \mathbf{G} .

Numerical performance and complexity results: Fig. 2 compares the performance of relaxed SEDS followed by CVR to that obtained by relaxed SEDS followed by quantization. We observe that CVR significantly outperforms quantization (and naive lattice decoding) and yields close-to-ML performance. This improvement comes, however, at the cost of having to solve a finite-lattice CVP, namely (21), in contrast to simple element-wise slicing in the case of quantization and to simply discarding estimates that do not belong to $\mathcal{L}(\mathbf{G})$ in the case of naive lattice decoding, both of which exhibit significantly smaller complexity.

We next compare the overall complexity (i.e., the total complexity of both SEDS runs) of LR-aided SEDS followed by CVR (denoted by LR-SESD, CVR) to that of SEDS operating directly on the finite lattice, i.e., SEDS (MLD). Fig. 1 shows that LR-SESD with CVR outperforms SEDS (MLD) *only* in the very-high-SNR regime with the corresponding complexity savings being minor. In addition, this comparison favors LR-SESD with CVR as the additional complexity associated with LR has been neglected.

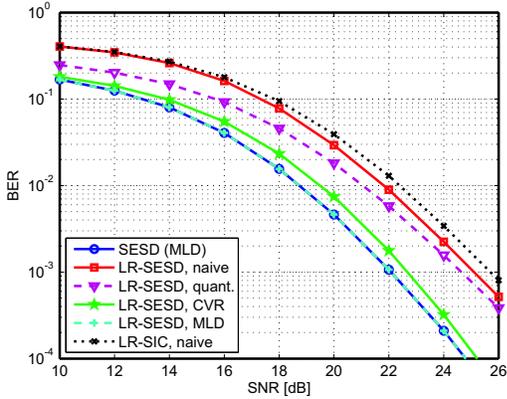


Fig. 2. Bit error rate (BER) performance comparison of SEDS (MLD), LR-SESD, and LR-SIC.

C. Two-Stage Detection

The discussion above leads us to the question of whether using $\underline{D}(\mathbf{r}) = \hat{\mathbf{u}}$ for remapping instead of \mathbf{r} is sensible. After all, by mapping \mathbf{r} onto $\hat{\mathbf{u}}$ we are potentially discarding useful information. In the following, we show that, indeed, performing remapping based on \mathbf{r} rather than on $\hat{\mathbf{u}}$ leads to a performance improvement (compared to, e.g., CVR) without requiring additional complexity.

To this end, we propose a new detector structure, which consists of two stages. The first stage corresponds to relaxation from $\mathcal{L}(\mathbf{G})$ to $\underline{\mathcal{L}}(\mathbf{G})$ (possibly followed by LR) and using a relaxed MIMO detector. The second stage is invoked *only* if the first stage yields a result that does *not* belong to the finite lattice $\mathcal{L}(\mathbf{G})$ and corresponds to a MIMO detector operating on the finite lattice. This second-stage detector *uses the received vector* \mathbf{r} instead of $\hat{\mathbf{u}}$. We emphasize that if both stages employ MLD, the two-stage detector achieves ML performance, since in the case of $\hat{\mathbf{u}}^{\text{ML}} \notin \mathcal{L}(\mathbf{G})$, the second stage performs non-relaxed MLD on the basis of \mathbf{r} .

Numerical performance and complexity results: Fig. 1 shows the overall complexity of two-stage detection (i.e., the total complexity across the two stages) using LR-aided SEDS in the first stage and SEDS in the second stage (denoted by LR-SESD, MLD). Again, the complexity associated with LR is ignored. Consulting the corresponding performance results in Fig. 2, we can conclude that two-stage detection employing SEDS in both stages and LR in the first stage outperforms LR-SESD with CVR both in terms of complexity *and* performance and, hence, the performance loss associated with LR-SESD followed by CVR (compared to ML performance) must result from the fact that $\hat{\mathbf{u}}^{\text{ML}}$ rather than \mathbf{r} is used for remapping. We can furthermore conclude that in the high-SNR regime the two-stage approach exhibits slightly lower complexity than SEDS, while realizing the same (i.e., ML) performance. We finally note that the computational complexity of LR and the complexity of checking whether $\hat{\mathbf{u}} \in \mathcal{L}(\mathbf{G})$ has been neglected in this comparison.

V. CONCLUSIONS

We demonstrated that relaxation from finite to infinite lattices combined with lattice reduction (LR) can reduce the (tree-search) complexity of Schnorr-Euchner sphere decoding (SESD), but generally leads to a significant performance loss. This performance loss is caused by the fact that the relaxed

detector does not necessarily deliver estimates that belong to the finite lattice and hence does not realize ML performance. Remapping methods that yield (close-to) ML performance necessitate solving finite-lattice closest vector problems (CVPs). Hence, if ML performance is required, using SEDS directly on the finite lattice seems to require smaller complexity, as remapping is avoided and only one (finite-lattice) CVP needs to be solved. We proposed an ML-optimal two-stage detector which was shown to require lower complexity than SEDS on the finite lattice provided i) the SNR is high enough and ii) the complexity corresponding to LR and to checking whether estimates correspond to the finite lattice is neglected. Generally speaking, we conclude that relaxation and LR are better suited for use in conjunction with low-complexity detectors (such as linear detection or SIC) with overall sub-optimal performance.

REFERENCES

- [1] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bölcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 7, pp. 1566–1577, July 2005.
- [2] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. of Computation*, vol. 44, pp. 463–471, Apr. 1985.
- [3] H. Yao and G. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE GLOBECOM*, vol. 1, Nov. 2002, pp. 424–428.
- [4] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Inf. Theory Workshop*, Mar. 2003, pp. 345–348.
- [5] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, no. 4, pp. 515–534, Dec. 1982.
- [6] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. on Inf. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec. 2007.
- [7] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," Master's thesis, Chinese University of Hong Kong, Dept. of Information Engineering, June 1991.
- [8] E. Viterbo and E. Biglieri, "A universal decoding algorithm for lattice codes," *Colloque sur le Traitement du Signal et ses Applications (GRETSI)*, vol. 14, pp. 611–614, Sept. 1993.
- [9] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. on Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.
- [10] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985.
- [11] A. H. Banihashemi and A. K. Khandani, "On the complexity of decoding lattices using the Korkin-Zolotarev reduced basis," *IEEE Trans. on Inf. Theory*, vol. 44, no. 1, pp. 162–171, Jan. 1998.
- [12] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. Springer, 1999.
- [13] J. M. Wills, "Bounds for the lattice point enumerator," *Geometriae Dedicata*, vol. 40, no. 2, pp. 237–244, Nov. 1991.
- [14] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, Mar. 1986.
- [15] R. Kannan, "Minkowski's convex body theorem and integer programming," *Math. of Operations Research*, vol. 12, no. 3, pp. 415–440, Aug. 1987.
- [16] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K.-D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *IEE Electronics Letters*, vol. 37, no. 22, pp. 1348–1350, Oct. 2001.
- [17] Y. H. Gan and W. H. Mow, "Complex lattice reduction algorithms for low-complexity MIMO detection," in *Proc. IEEE GLOBECOM*, Nov. 2005, pp. 2953–2957.
- [18] D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer, "MMSE-based lattice-reduction for near-ML detection of MIMO systems," in *Proc. ITG/IEEE Workshop on Smart Antennas*, Mar. 2004, pp. 106–113.
- [19] M. Taherzadeh and A. K. Mahmoud, "On the limitations of the naive lattice decoding," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2007, pp. 201–204.