Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2007, Article ID 84265, 13 pages
doi:10.1155/2007/84265

Research Article

Efficient Near Maximum-Likelihood Detection for Underdetermined MIMO Antenna Systems Using a Geometrical Approach

Kai-Kit Wong[1] and Arogyaswami Paulraj[2]

[1] *Adastral Park Research Campus, University College London, Martlesham IP5 2BS, UK*
[2] *Information Systems Laboratory, Stanford University, Stanford, CA 94305, USA*

Received 9 January 2007; Revised 21 May 2007; Accepted 10 October 2007

Recommended by P. Djurić

Maximum-likelihood (ML) detection is guaranteed to yield minimum probability of erroneous detection and is thus of great importance for both multiuser detection and space-time decoding. For multiple-input multiple-output (MIMO) antenna systems where the number of receive antennas is at least the number of signals multiplexed in the spatial domain, ML detection can be done efficiently using sphere decoding. Suboptimal detectors are also well known to have reasonable performance at low complexity. It is, nevertheless, much less understood for obtaining good detection at affordable complexity if there are less receive antennas than transmitted signals (i.e., underdetermined MIMO systems). In this paper, our aim is to develop an effcient detection strategy that can achieve near ML performance for underdetermined MIMO systems. Our method is based on the geometrical understanding that the ML point happens to be a point that is "close" to the decoding hyperplane in all directions. The fact that such proximity-close points are much less is used to devise a decoding method that promises to greatly reduce the decoding complexity while achieving near ML performance. An average-case complexity analysis based on Gaussian approximation is also given.

1. INTRODUCTION

One approach to achieve high rate in wireless channels is to form a multiple-input multiple-output (MIMO) channel by employing multiple antennas at both ends [1, 2]. The merit is that multiple signal streams can be accommodated simultaneously so that a higher rate can be supported. But the challenge is that optimal performance comes with maximum-likelihood (ML) detection/decoding complexity, which is known to grow exponentially with the number of signals coexisted in the same radio channel.

For an (N_t, N_r) antenna system, where $N_t \leq N_r$, ML detection can be obtained in a relatively cheap way by spherical detection or widely known as sphere decoding [3–8] (the notation (N_t, N_r) is used to denote a MIMO system which has N_t transmit antennas and N_r receive antennas). In sphere decoding, the channel is rotated onto a space such that the spatially multiplexed signals are only sequentially dependent. A computationally efficient algorithm is thus available to search the signal points that fall inside the decoding hypersphere for a given radius. And, the ML detection can then be efficiently obtained by solving the dual spherical search problem with a judicious choice of radius. Some recent improvements on sphere decoding can be found in [9–11] to further reduce its complexity.

In practice, however, it is not likely to have a sufficient number of receive antennas for decoupling the spatial signals (especially in the downlink where the same frequency may be reused for support of multiple users as suggested by information theory [12, 13]) (extension from detection to soft decoding is possible using techniques such as that in [14]). It is therefore important to consider the asymmetric case, where $N_t > N_r$. In this underdetermined MIMO setting, the immediate difficulty is that the existing low-complexity detectors such as zeroforcing (ZF) or minimum-mean-square-error (MMSE) receivers would not work, not to mention the performance degradation suffered from the suboptimal detectors due to receiver linearization.

The main difficulty for efficient ML detection of an underdetermined MIMO system lies in the detection problem of an $(N_t - N_r + 1, 1)$ system where the decision of a multisymbol vector needs to be made based only on a single received

signal. Further, the inefficiency of detecting the multiple-input single-output (MISO) problem will propagate to the upper decoding layers, leading to exponential average complexity. There were a few attempts that looked at generalizing the sphere decoder to cope with overloaded detection [15, 16] and they will be used as benchmarks in this paper.

Focusing on making hard detection, this paper aims to devising an efficient detection scheme for underdetermined MIMO antenna systems (i.e., with $N_t > N_r$). Knowing that the bottleneck of realizing efficient detection lies in the MISO detection in the first decoding layer, we exploit the geometrical representation of the MISO ML detection, and then develop an efficient detection algorithm to achieve near ML performance at much lower complexity. In particular, we observe that an ML solution, which is the lattice point minimizing the Euclidean distance from the decoding hyperplane, happens to be the point that first sees the hyperplane (or regarded as "close") in *all* coordinate axes. The fact that the points, jointly closed in all axes, are rare permits an efficient implementation of ML detection. In light of this interpretation, we will refer to the proposed detection scheme as *planar detection*.

This paper has made the following contributions.

(i) An efficient MIMO detector, which is applicable for underdetermined MIMO systems to achieve exact ML performance for real-valued signal constellations and near ML performance for general complex modulations, is proposed. Numerical results will show that the proposed detector requires much lower average complexity than the recently published techniques [15, 16], especially for high-level modulations.

(ii) Making a few approximations, we derive analytically the average-case complexity exponent, $e_c \triangleq \log_N \mathcal{C}(N, \mathrm{SNR})$, foran ideal realization of the proposed planar detector, where $\mathcal{C}(N, \mathrm{SNR})$ is the expected computational complexity in the number of elementary calculations averaged over many independent channel instantiations and the transmit lattices.

The remainder of the paper is organized as follows. Section 2 describes the channel model for an underdetermined MIMO antenna system and introduces the ML detection problem. In Section 3, we present the proposed planar detection algorithm for MISO-ML detection. Section 4 is dedicated to the derivation of the average-case complexity exponent for an ideal realization of a planar detector. Simulation results will be given in Section 5, and we conclude the paper in Section 6.

2. PROBLEM FORMULATION AND CHANNEL MODEL

2.1. System model

The system of a MIMO antenna network, where N_t antennas are located at the transmitter and N_r antennas are located at the receiver, can be written as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{s}} + \tilde{\mathbf{n}}, \tag{1}$$

where $\tilde{\mathbf{y}} = [\tilde{y}_1\ \tilde{y}_2\ \cdots\ \tilde{y}_{N_r}]^T$ is the received signal vector, $\tilde{\mathbf{H}} = [\tilde{h}_{i,j}]_{i=1,\ldots,N_r; j=1,\ldots,N_t}$ is the MIMO channel matrix, $\tilde{\mathbf{s}} = [\tilde{s}_1\ \tilde{s}_2\ \cdots\ \tilde{s}_{N_t}]^T$ is the transmitted symbol vector, and $\tilde{\mathbf{n}}$ is the noise vector whose elements are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance of $N_0/2$ per dimension, that is, $\mathcal{N}(0, N_0/2)$. Note that this paper focuses on overloaded asymmetric MIMO antenna systems where $N_t > N_r$, $\tilde{\mathbf{H}}$ is a fat matrix and constitutes an underdetermined system.

Assuming that channel state information (CSI) is unavailable at the transmitter, $\tilde{\mathbf{s}}$ consists of only data or coded data without signal preprocessing. To further simplify our discussion, we assume that the transmitter is sending independent symbols across antennas (i.e., purely uncoded spatial multiplexing). Therefore, $\tilde{s}_n \in \tilde{\mathcal{Q}}$, where $\tilde{\mathcal{Q}}$ is the discrete alphabet set (e.g., $\tilde{Q} = \{-1-j, -1+j, 1-j, 1+j\}$) if 4-QAM (quadrature amplitude modulation) is considered.

ML detection is realized by finding s that minimizes

$$\tilde{\mathbf{s}}_{\mathrm{ML}} = \arg\min_{s\in\tilde{\mathcal{Q}}^{N_t}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\tilde{\mathbf{s}}\|^2. \tag{2}$$

For efficient detection, it is often advantageous to convert the above complex system model into an equivalent real-valued representation as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3}$$

where

$$\begin{aligned}
\mathbf{y} &= \left[\mathrm{Re}\{\tilde{\mathbf{y}}\}^T\ \ \mathrm{Im}\{\tilde{\mathbf{y}}\}^T\right]^T,\\
\mathbf{H} &= \begin{bmatrix} \mathrm{Re}\{\tilde{\mathbf{H}}\} & -\mathrm{Im}\{\tilde{\mathbf{H}}\} \\ \mathrm{Im}\{\tilde{\mathbf{H}}\} & \mathrm{Re}\{\tilde{\mathbf{H}}\} \end{bmatrix},\\
\mathbf{s} &= \left[\mathrm{Re}\{\tilde{\mathbf{s}}\}^T\ \ \mathrm{Im}\{\tilde{\mathbf{s}}\}^T\right]^T,\\
\mathbf{n} &= \left[\mathrm{Re}\{\tilde{\mathbf{n}}\}^T\ \ \mathrm{Im}\{\tilde{\mathbf{n}}\}^T\right]^T,
\end{aligned} \tag{4}$$

where the superscript T denotes transposition. Note that $\mathbf{H}$ is now of dimensions $N \times M$, where $M = 2N_t$ and $N = 2N_r$. As a consequence, our detection problem becomes

$$\mathbf{s}_{\mathrm{ML}} = \arg\min_{s\in\mathcal{Q}^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2, \tag{5}$$

where $\mathcal{Q} = \{-1, 1\}$ if 4-QAM is used and $\mathcal{Q} = \{-3, -1, 1, 3\}$ if 16-QAM is used.

2.2. Spherical search for MIMO-ML detection

To efficiently perform (5) [and hence (2)], it is proposed in [3–8] to solve a dual search problem that finds all the lattice points satisfying

$$\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 \le C^2 \tag{6}$$

for a given radius $C(> 0)$. The one that is the closest to the center of the hypersphere, $\mathbf{y}$, will be the ML point giving the smallest Euclidean distance. The complexity advantage of this search can be described as follows.

By structuring the channel using QR decomposition, that is, $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is an upper triangular matrix, (6) can be rewritten as

$$\|\check{\mathbf{y}} - \mathbf{Rs}\|^2 \le C^2, \tag{7}$$

$$\begin{aligned} &\left(R_{1,1}s_1 + R_{1,2}s_2 + \cdots + R_{1,M}s_M - \check{y}_1^2\right)^2 \\ &+\left(R_{2,2}s_2 + R_{2,3}s_3 + \cdots + R_{2,M}s_M - \check{y}_2^2\right)^2 \\ \Longleftrightarrow \quad &\qquad\vdots \qquad\qquad\qquad \le C^2 \\ &+\left(R_{N-1,N-1}s_{N-1} + R_{N-1,M}s_M - \check{y}_{N-1}^2\right)^2 \\ &+\left(R_{N,N}s_N + \cdots + R_{N,M}s_M - \check{y}_N^2\right)^2, \end{aligned} \tag{8}$$

where $\check{\mathbf{y}} \triangleq \mathbf{Q}^T\mathbf{y}$.

If $M = N$ or the channel is square, searching the points that satisfy (7) can be realized by finding the points satisfying all of the following inequalities:

$$\left(\check{y}_\ell - \sum_{k=\ell}^{N} R_{\ell,k}s_k\right)^2 \le C_\ell^2, \quad \text{for } \ell = 1, 2, \ldots, N, \tag{9}$$

where

$$C_\ell = \begin{cases} \sqrt{C_{\ell+1}^2 - \left(\check{y}_{\ell+1} - \sum_{k=\ell+1}^{N} R_{\ell+1,k}s_k\right)^2} & \ell \neq N, \\ C & \ell = N. \end{cases} \tag{10}$$

The search can be done efficiently by examining the following inequalities in a sequential manner:

$$\frac{\check{y}_\ell - \sum_{k=\ell+1}^{N} R_{\ell,k}s_k - C_\ell}{R_{\ell,\ell}} \le s_\ell \le \frac{\check{y}_\ell - \sum_{k=\ell+1}^{N} R_{\ell,k}s_k + C_\ell}{R_{\ell,\ell}}. \tag{11}$$

The complexity advantage lies in the fact that when $\ell = M$, the search only involves one transmitted symbol, that is,

$$\frac{\check{y}_M - C}{R_{M,M}} \le s_M \le \frac{\check{y}_M + C}{R_{M,M}}. \tag{12}$$

After s_M is known, this information will be carried forward to the upper layer (i.e., $\ell = M - 1$) and the search will then involve only one undetermined symbol s_{M-1} (since s_M has been found). This decoding strategy continues until all the symbol sequences that satisfy the inequalities are found.

Unfortunately, if $M > N$ or the channel is fat, then at the Nth layer, we will have

$$\left[\check{y}_N - \left(R_{N,N}s_N + \cdots + R_{N,M}s_M\right)\right]^2 \le C^2 \tag{13}$$

that involves $M - N + 1$ symbols for detection. It becomes the complexity bottleneck of the detection problem. Recognizing that it has a similar form as the detection problem of a real-valued MISO system. In the sequel, our effort will first be spent onto developing an efficient detection algorithm for a MISO system, and then extend the results for an underdetermined MIMO system.

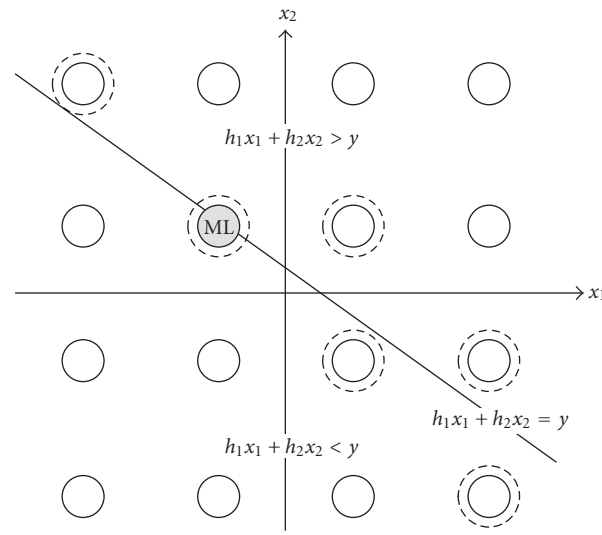


Figure 1: A diagram showing the decoding line when $\mathcal{Q} = \{-3, -1, 1, 3\}$ and $K = 2$.

3. PLANAR DETECTION

3.1. Geometrical interpretation and algorithm

For a spatially multiplexed MISO system where the inputs are independent symbols, $x_k \in \mathcal{Q}$, which take values from the alphabet set $\mathcal{Q}$, the received signal can be written as

$$y = h_1x_1 + h_2x_2 + \cdots + h_Kx_K + \eta, \tag{14}$$

where h_n denotes the effective channel response from the kth transmitter input to the receiver, and $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$ denotes the noise.

ML detection aims to find the values, $x_1, x_2, \ldots, x_K$, jointly or the vector $\mathbf{x} = [x_1\ x_2\ \ldots\ x_K]^T$ that minimizes the following:

$$\mathbf{x}_{\mathrm{ML}} = \arg \min_{x_1, x_2, \ldots, x_K} (h_1x_1 + h_2x_2 + \cdots + h_Kx_K - y)^2. \tag{15}$$

Geometrically, it means to find the point $(x_1, x_2, \ldots, x_K) \in \mathcal{Q}^K$ that is the closest to the decoding hyperplane (see Figure 1 for $\mathcal{Q} = \{-3, -1, 1, 3\}$ and $K = 2$),

$$\mathcal{P} : h_1x_1 + h_2x_2 + \cdots + h_Kx_K = y. \tag{16}$$

The following definition and theorem link the proximity closeness of a point to the ML detection (15), which will later permit an efficient implementation of ML detection.

Definition 1. Closeness of a poin. Assuming for the sake of simplicity that the channels are real and positive, that is, $h_1, h_2, \ldots \geq 0$, a point $(x_1, x_2, \ldots, x_K)$ is said to be close in coordinate axis or direction k (or abbreviated as d-k) if x_k is the first point that sees the decoding hyperplane from the half-space $h_1x_1 + \cdots + h_Kx_K > y$ or $h_1x_1 + \cdots + h_Kx_K < y$, that is,

$$x_k \in \left\{x : \min_{x\in\mathcal{Q}} x \text{ s.t. } x > \frac{y - \sum_{m\neq k} h_m x_m}{h_k}\right\} \cup \left\{x : \max_{x\in\mathcal{Q}} x \text{ s.t. } x < \frac{y - \sum_{m\neq k} h_m x_m}{h_k}\right\}. \tag{17}$$

It should be noted that the closeness property of a point in one particular direction is conditioning on the coordinates in other directions. In other words, if some coordinates change, the closeness property may be destroyed.

Theorem 2. *Joint closeness of the ML point. The point $(x_1, x_2, \ldots, x_K) \in \mathcal{Q}^K$ that corresponds to the ML detection must be close in all directions (or close in d-$1, 2, \ldots, K$).*

Proof. Assuming (proof by contradiction) that the ML point $(x_1, x_2, \ldots, x_K)$ is not close in some direction, say k, so x_k does not satisfy (17). Then, there will exist a coordinate x_k^* that satisfies (17) and produces another point $(x_1, x_2, \ldots, x_k^*, \ldots, x_K)$, which is now close in d-k. This new point will give a smaller distance from the decoding hyperplane (16). Thus, we have a contradiction and the proof is complete. □

The joint closeness property serves as a necessary condition for the ML detection (15). The fact that the number of points that are jointly close in all directions are scarce becomes the key to the design of an efficient detection algorithm. In Figure 1, we illustrate an example of a decoding line if $\mathcal{Q} = \{-3, -1, 1, 3\}$. In this example, there are six lattice points that are close in d-$1, 2$. As compared to the total of 16 possible lattice points, if we search only the closed points, it can significantly reduce the decoding complexity while still achieving the exact ML performance. To do so, we propose the following *planar detection* algorithm. (To supplement the following description, the pseudocode of the algorithm is also given in Algorithm 1.)

Planar Detection algorithm

(1) Initialize the sets $\mathrm{C}\cdot\mathrm{x} = \mathrm{C}\cdot\mathrm{d} = \emptyset$ and the dimension index $k = 1$. Arbitrarily choose a lattice point $\mathbf{x} \in \mathcal{Q}^K$ and store it in the set $\mathrm{B}\cdot\mathrm{x} = \{\mathbf{x}\}$. Compute

$$\Delta y(x) = \mathbf{h}^T\mathbf{x} - y = h_1x_1 + h_2x_2 + \cdots + h_Kx_K - y, \tag{18}$$

which is stored in the set $\mathrm{B}\cdot\mathrm{d} = \{\Delta y(\mathbf{x})\}$.

(2) If $\mathrm{B}\cdot\mathrm{x} = \emptyset$, then go to Step 6. Otherwise, find the two coordinates that make the first element of $\mathrm{B}\cdot\mathrm{x}$, denoted as $e_1[\mathrm{B}\cdot\mathrm{x}]$, close in d-k. This is done by computing

$$\begin{aligned} \hat{a}_k &= \left\{x : \min_{x\in\mathcal{Q}} x \text{ s.t. } x > x_B\right\}, \\ \check{a}_k &= \left\{x : \max_{x\in\mathcal{Q}} x \text{ s.t. } x < x_B\right\}, \end{aligned} \tag{19}$$

where

$$x_B = x_k - \frac{e_1[\mathrm{B}\cdot\mathrm{d}]}{h_k}. \tag{20}$$

If $\hat{a}_k \neq \emptyset$, then produce a new point $\hat{\mathbf{x}} = \mathbf{x}$ but with $\hat{x}_k = \hat{a}_k$ and compute

$$\Delta y(\hat{\mathbf{x}}) = h_k(\hat{x}_k - x_B). \tag{21}$$

Similarly, we will have also $\check{\mathbf{x}}$ with $\check{x}_k = \check{a}_k$ and $\Delta y(\check{\mathbf{x}})$ if $\check{a} \neq \emptyset$.

(3) If $\hat{x}_k = x_k$, the point $\mathbf{x}$ is close in d-$1, 2, \ldots, k$. Then do the followings.
   (i) If $k < K$, update $k := k + 1$, $\mathrm{B}\cdot\mathrm{x} := \{\mathrm{B}\cdot\mathrm{x}, \check{\mathbf{x}}\}$ and $\mathrm{B}\cdot\mathrm{d} := \{\mathrm{B}\cdot\mathrm{d}, \Delta y(\check{\mathbf{x}})\}$.
   (ii) If $k = K$, erase $\mathbf{x}$ from the set $\mathrm{B}\cdot\mathrm{x}$ and reset $k = 1$. The point $\mathbf{x}$ is a point jointly closed in all dimensions and a candidate for ML detection (15). The point is stored in $\mathrm{C}\cdot\mathrm{x}$ and its Euclidean distance from the decoding hyperplane is stored in $\mathrm{C}\cdot\mathrm{d}$.
   (iii) Go back to Step 2.

(4) If $\check{x}_k = x_k$, the point $\mathbf{x}$ is close in d-$1, 2, \ldots, k$. Then do the followings.
   (i) If $k < K$, update $k := k + 1$, $\mathrm{B}\cdot\mathrm{x} := \{\mathrm{B}\cdot\mathrm{x}, \hat{x}\}$, and $\mathrm{B}\cdot\mathrm{d} := \{\mathrm{B}\cdot\mathrm{d}, \Delta y(\hat{\mathbf{x}})\}$.
   (ii) If $k = K$, erase $\mathbf{x}$ from the set $\mathrm{B}\cdot\mathrm{x}$ and reset $k = 1$. The point $\mathbf{x}$ is a point jointly closed in all dimensions and a candidate for ML detection (15). The point is stored in $\mathrm{C}\cdot\mathrm{x}$ and its Euclidean distance from the decoding hyperplane is stored in $\mathrm{C}\cdot\mathrm{d}$.
   (iii) Go back to Step 2.

(5) Erase the first elements of the sets $\mathrm{B}\cdot\mathrm{x}$ and $\mathrm{B}\cdot\mathrm{d}$. Reset $k = 1$ and also update $\mathrm{B}\cdot\mathrm{x} := \{\mathrm{B}\cdot\mathrm{x}, \hat{\mathbf{x}}, \check{\mathbf{x}}\}$, and $\mathrm{B}\cdot\mathrm{d} := \{\mathrm{B}\cdot\mathrm{d}, \Delta y(\hat{\mathbf{x}}), \Delta y(\check{\mathbf{x}})\}$. Then, go back to Step 2.

(6) All of the points jointly close in all directions are located and stored in $\mathrm{C}\cdot\mathrm{x}$ with their Euclidean distances stored in $\mathrm{C}\cdot\mathrm{d}$. Among these points, the one that gives the smallest $\Delta y^2(\mathbf{x})$ is the ML detection point for the MISO system (15).

In the above algorithm, we have assumed that h_1, $h_2, \ldots \geq 0$. However, if some channels are negative, one can still apply the above planar detection algorithm for the channels, $|h_1|, |h_2|, \ldots, |h_K|$, and then obtain the corresponding ML solution by flipping the signs of the symbols for the negative channels.

Planar Detection $(y, \mathbf{h}, \mathcal{Q})$
comment: Initialize the starting lattice point and its Euclidean distance

$$\begin{cases} \mathbf{x} \in \mathcal{Q}^K \text{ is arbitraily chosen} \\ d = \mathbf{h}^T\mathbf{x} - y \end{cases} \quad \begin{cases} \mathrm{B}\cdot\mathrm{x} &= \{\mathbf{x}\} \\ \mathrm{B}\cdot\mathrm{d} &= \{d\} \\ \mathrm{B}\cdot\mathrm{s} &= \{\mathbf{0}\} \end{cases} \text{ and } \begin{matrix} \mathrm{C}\cdot\mathrm{x} = \varnothing \\ \mathrm{C}\cdot\mathrm{d} = \varnothing \end{matrix}$$

comment: The WHILE loop proceeds from one point to another to identify the closed points
while $\mathrm{B}\cdot\mathrm{x} \neq \varnothing$

do
- $I = \max\{\text{find}(e_1[\mathrm{B}\cdot\mathrm{s}] = 1)\}$
- if $I = \varnothing$
  - then $I = 0$
- if $I = K$
  - then
    - if $e_1[\mathrm{B}\cdot\mathrm{x}] \notin \mathrm{C}\cdot\mathrm{x}$
      - then $\begin{cases} \mathrm{C}\cdot\mathrm{x} &= \{\mathrm{C}\cdot\mathrm{x}, e_1[\mathrm{B}\cdot\mathrm{x}]\} \\ \mathrm{C}\cdot\mathrm{d} &= \{\mathrm{C}\cdot\mathrm{d}, e_1[\mathrm{B}\cdot\mathrm{d}]\} \end{cases}$
    - output $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}, \text{ERASE}(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}))$
  - else
    - $J = I + 1$
    - output $(\mathrm{P}, \mathrm{d}_p, N_p, \text{FINDING CLOSED POINTS}\,(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathbf{h}, J, \mathcal{Q}))$
    - if $N_p = 1$
      - then
        - if $e_1[\mathrm{B}\cdot\mathrm{x}](J) = e_1[\mathrm{P}](J)$
          - then $e_1[\mathrm{B}\cdot\mathrm{s}](J) = 1$
        - output $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}, \text{ERASE}\,(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}))$
      - else if $N_p = 2$
        - then
          - if $e_1[\mathrm{B}\cdot\mathrm{x}](J) = e_1[\mathrm{P}](J)$ or $e_1[\mathrm{B}\cdot\mathrm{x}](J) = e_2[\mathrm{P}](J)$
            - then
              - $e_1[\mathrm{B}\cdot\mathrm{s}](J) = 1$
              - if $e_1[\mathrm{B}\cdot\mathrm{x}](J) = e_1[\mathrm{P}](J)$
                - then $\begin{cases} \mathrm{B}\cdot\mathrm{x} &= \{\mathrm{B}\cdot\mathrm{x}, e_2[\mathrm{P}]\} \\ \mathrm{B}\cdot\mathrm{d} &= \{\mathrm{B}\cdot\mathrm{d}, e_2[\mathrm{d}_p]\} \\ \mathrm{B}\cdot\mathrm{s} &= \{\mathrm{B}\cdot\mathrm{s}, \mathbf{0}\} \end{cases}$
                - else $\begin{cases} \mathrm{B}\cdot\mathrm{x} &= \{\mathrm{B}\cdot\mathrm{x}, e_1[\mathrm{P}]\} \\ \mathrm{B}\cdot\mathrm{d} &= \{\mathrm{B}\cdot\mathrm{d}, e_1[\mathrm{d}_p]\} \\ \mathrm{B}\cdot\mathrm{s} &= \{\mathrm{B}\cdot\mathrm{s}, \mathbf{0}\} \end{cases}$
            - else
              - output $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}, \text{ERASE}\,(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s}))$
              - $\begin{cases} \mathrm{B}\cdot\mathrm{x} &= \{\mathrm{B}\cdot\mathrm{x}, e_1[\mathrm{P}], e_2[\mathrm{P}]\} \\ \mathrm{B}\cdot\mathrm{d} &= \{\mathrm{B}\cdot\mathrm{d}, e_1[\mathrm{d}_p] e_2[\mathrm{d}_p]\} \\ \mathrm{B}\cdot\mathrm{s} &= \{\mathrm{B}\cdot\mathrm{s}, \mathbf{0}, \mathbf{0}\} \end{cases}$

returne $(\mathrm{C}\cdot\mathrm{x}, \mathrm{C}\cdot\mathrm{d})$

Algorithm 1: The pseudocode of the Planar Detection algorithm where $\varnothing$ is an empty set, $e_i[\mathrm{A}]$ denotes the ith element of a given set A, and $e_i[\mathrm{A}](j)$ denotes the jth entry of the ith element of A. The algorithms of finding closed points and erase are shown in Algorithms 2 and 3, respectively.

This algorithm guarantees to obtain the ML lattice point of the real-valued MISO system (15) and this can be seen from the fact that the algorithm performs the search exhaustively but along the decoding hyperplane $\mathcal{P}$. As such, the worse-case complexity occurs when the decoding hyperplane cuts the middle of the lattice space, but much reduction in the average-case complexity is anticipated as the decoding hyperplane may cut the edges of or when the cut is away from the lattice space. A more detailed analysis will be given in Section 4. Figure 2 shows an example (with $K = 2$ and $\mathcal{Q} = \{-3, -1, 1, 3\}$) of the decoding procedure for planar detection. As can be seen in this example, eight points are actually visited instead of six points because some extra points have to be visited in order to browse along the decoding hyperplane $\mathcal{P}$.

Because of the exhaustive nature of the algorithm, it is necessary to have another set, $\mathrm{V}\cdot\mathrm{x}$, which stores all the visited lattice points. This set is important to avoid repeated visits of the same point through the search. In this way, convergence will be ensured. Due to the limited space of the paper, the inclusion of the set $\mathrm{V}\cdot\mathrm{x}$ is omitted in both the algorithm description and the pseudocode.

Finding Closed Points $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathbf{h}, J, \mathcal{Q})$

$\mathrm{P} = \varnothing$

$\mathrm{d}_p = \varnothing$

$x_{\mathrm{B}} = e_1[\mathrm{B}\cdot\mathrm{x}] - \dfrac{e_1[\mathrm{B}\cdot\mathrm{d}]}{h_J}$

$\hat{a} = \min\{\mathcal{Q}(\mathrm{find}(\mathcal{Q} > x_{\mathrm{B}}))\}$

$\check{a} = \max\{\mathcal{Q}(\mathrm{find}(\mathcal{Q} \le x_{\mathrm{B}}))\}$

if $\hat{a} \neq \varnothing$

then
$\mathrm{T} = e_1[\mathrm{B}\cdot\mathrm{x}]$
$\mathrm{T}(J) = \hat{a}$
$\mathrm{d}_t = h_J(\hat{a} - x_{\mathrm{B}})$
$\mathrm{P} = \{\mathrm{P}, \mathrm{T}\}$
$\mathrm{d}_p = \{\mathrm{d}_p, \mathrm{d}_t\}$

if $\check{a} \neq \varnothing$

then
$\mathrm{T} = e_1[\mathrm{B}\cdot\mathrm{x}]$
$\mathrm{T}(J) = \check{a}$
$\mathrm{d}_t = h_J(\check{a} - x_{\mathrm{B}})$
$\mathrm{P} = \{\mathrm{P}, \mathrm{T}\}$
$\mathrm{d}_p = \{\mathrm{d}_p, \mathrm{d}_t\}$

if $\hat{a} = \varnothing$ and $\check{a} = \varnothing$

then $N_p = 0$

else if $\{\hat{a} = \varnothing$ and $\check{a} \neq \varnothing\}$ or $\{\hat{a} \neq \varnothing$ and $\check{a} = \varnothing\}$

then $N_p = 1$

else if $\hat{a} \neq \varnothing$ and $\check{a} \neq \varnothing$

then $N_p = 2$

return $(\mathrm{P}, \mathrm{d}_p, N_p)$

Algorithm 2: The pseudocode of the finding closed points algorithm.

Erase $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s})$

$\mathrm{B}\cdot\mathrm{x} = \mathrm{B}\cdot\mathrm{x} \setminus e_1[\mathrm{B}\cdot\mathrm{x}]$

$\mathrm{B}\cdot\mathrm{d} = \mathrm{B}\cdot\mathrm{d} \setminus e_1[\mathrm{B}\cdot\mathrm{d}]$

$\mathrm{B}\cdot\mathrm{s} = \mathrm{B}\cdot\mathrm{s} \setminus e_1[\mathrm{B}\cdot\mathrm{s}]$

return $(\mathrm{B}\cdot\mathrm{x}, \mathrm{B}\cdot\mathrm{d}, \mathrm{B}\cdot\mathrm{s})$

Algorithm 3: The pseudocode of the erase algorithm.

3.2. Incorporating "good" points

In this section, we integrate the planar detection algorithm above for use of efficient detection of an undetermined MIMO system. Recall that for MIMO systems with $M > N$, after QR decomposition of the channel, we have to find the lattice points that satisfy all of the following inequalities:

$$\left(\check{y}_N - \sum_{k=N}^{M} R_{N,k} s_k\right)^2 \le C^2, \tag{22}$$

$$\vdots$$

$$\sum_{j=1}^{N}\left(\check{y}_j - \sum_{k=j}^{M} R_{j,k} s_k\right)^2 \le C^2. \tag{23}$$

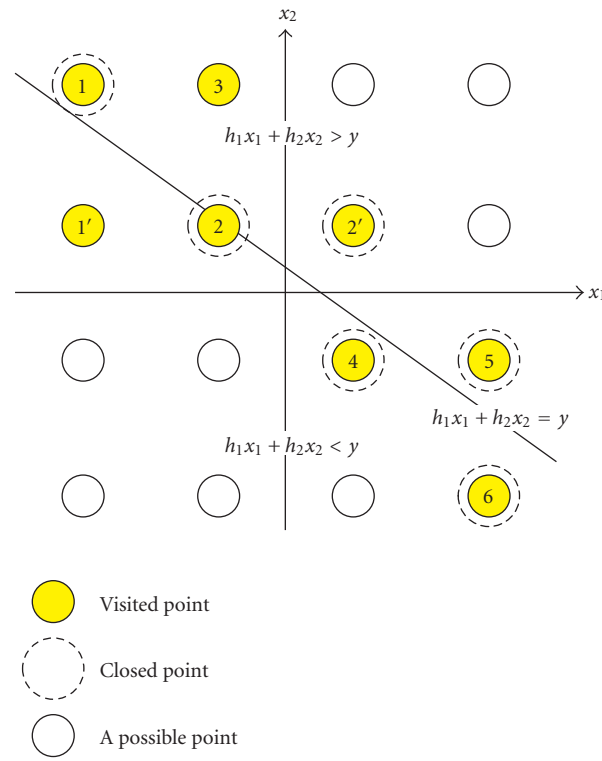


Figure 2: A diagram showing the decoding sequence of planar detection.

Note that (22) is the weakest necessary condition for the spherical search (7) while (23) is regarded as the strongest necessary (indeed sufficient) condition for the ML search. As described previously, complexity saving is realized by examining from the weakest condition to the strongest condition. At a reasonable signal-to-noise ratio (SNR), it is of high probability that the ML solution, which minimizes the Euclidean distance from $\check{\mathbf{y}}$ overall, is also the ML solution of (22). For this reason, instead of performing an overall search of the inequalities sequentially, we can concentrate on (22) and have the hyperplane

$$\mathcal{P} : R_{N,N} s_N + R_{N,N+1} s_{N+1} + \cdots + R_{N,M} s_M = \check{y}_N. \tag{24}$$

Apparently, this is the same as (16) but with $h_k = R_{N,N+k-1}$, $x_k = s_{N+k-1}$ for $k = 1, 2, \ldots, K$ and $y = \check{y}_N$. Therefore, the proposed planar detection strategy can be readily used to identify the closed points to the decoding hyperplane (24). With these candidates of $s_N, s_{N+1}, \ldots, s_M$, we can then sequentially determine $\{s_\ell\}_{\ell=1}^{N-1}$ using (11) for a given C.

The decoding method is very efficient as it avoids exhaustively searching all the possibilities for $s_N, \ldots, s_M$. However, due to the fact that the ML solution may not be one of the closed points obtained by planar detection, there is chance that the ML point is mistakenly discarded or pruned at the Nth decoding layer, and thus ML performance is no longer guaranteed. This probability of miss depends on the level of modulation being used and the operating SNR (for details, see the numerical results in Section 5).

Let

$$\mathrm{C}\cdot\mathrm{x} = \{\mathbf{s}_{N\leftrightarrow M} : \mathbf{s}_{N\leftrightarrow M} \text{ is close in } d-N,\ldots,M\}, \tag{25}$$

where $\mathbf{s}_{N\leftrightarrow M} \triangleq [s_N\ s_{N+1}\ \cdots\ s_M]^T$ and let

$$\mathrm{C}\cdot\mathrm{d} = \{\Delta y(\mathbf{s}_{N\leftrightarrow M}) : \mathbf{s}_{N\leftrightarrow M} \in \mathrm{C}\cdot\mathrm{x}\} \tag{26}$$

denote the two sets that contain all the closed points and their corresponding distances from the decoding hyperplane, obtained from planar detection. In order to regain the performance, we need to avoid or reduce as much as possible the probability that the ML point is being mistakenly pruned because the subcomponents of $\mathbf{s}_{\mathrm{ML}}$ are not in C·x.

The performance can be mostly recovered by slightly enlarging the set of the closed points. In this paper, this is done by adding some extra points that are potentially "good" (i.e., likely to be the ML point). The good points are found by visiting the neighboring points of some of the closed points in C·x. First of all, we obtain the set $\mathrm{C}\cdot\mathrm{x}^{(J)}(\subset \mathrm{C}\cdot\mathrm{x})$ that has J-closed points that have the smallest distances $|\Delta y|$. Then, for each point in $\mathrm{C}\cdot\mathrm{x}^{(J)}$, we alternate one symbol per each dimension to obtain new (good) points (see Figure 3), which are contained in the set

$$\mathrm{G}\cdot\mathrm{x} = \left\{\mathbf{z}\in\mathcal{Q}^K : z_\ell = \begin{cases}\min\left(s_{N\leftrightarrow M,\ell}+d, \max\limits_{s\in\mathcal{Q}} s\right)\\ \max\left(s_{N\leftrightarrow M,\ell}-d, \min\limits_{s\in\mathcal{Q}} s\right)\end{cases} \forall\, \mathbf{s}_{N\leftrightarrow M}\in\mathrm{C}\cdot\mathrm{x}^{(J)}\right\}, \tag{27}$$

where d is the separation of every adjacent constellations. As a result, the points that are carried forward to the upper layers are now in an enlarged set $\mathrm{C}\cdot\mathrm{x}^+ = \mathrm{C}\cdot\mathrm{x}\bigcup\mathrm{G}\cdot\mathrm{x}$. J is a design parameter, which can tradeoff the performance and complexity of the detector. It is hoped that $\mathrm{C}\cdot\mathrm{x}^+$ is now large enough to include the ML point, but small enough to keep the complexity low.

3.3. *Increasing radius for efficient detection*

Another key techniques to reduce the decoding complexity are sorting and shrinking radii [17, 18]. Sorting when used will navigate the search so that the ML point is more likely to be visited earlier. This is implemented through ordering of the lattice points. Specifically, we sort the possible lattice coordinates, s_ℓ, [defined by the intervals (11)] in descending order, according to their distance from the middle of the interval, that is,

$$\left| s_\ell - \left(\frac{\check{y}_\ell - \sum_{k=\ell+1}^{N} R_{\ell,k}s_k}{R_{\ell,\ell}}\right)\right|^2. \tag{28}$$

The shrinking radius approach is to miniaturize iteratively the sphere in order that the ML search can be done more efficiently. However, it is important to note that for fat MIMO channels, an increasing radius approach (i.e., starting the search with a very small radius and increasing it progressively if no point is found) turns out to be much more efficient than the shrinking radius approach [17] because of the fact that the shrinking radius approach is efficient only when the miniaturization of the decoding sphere can be done rapidly. But for an underdetermined MIMO system, we deal with in this paper, the confidence level of visiting the lattice points leading to the radius close to $C_{\min}$ is low. As a consequence, an increasing radius approach will be used instead in the detection.

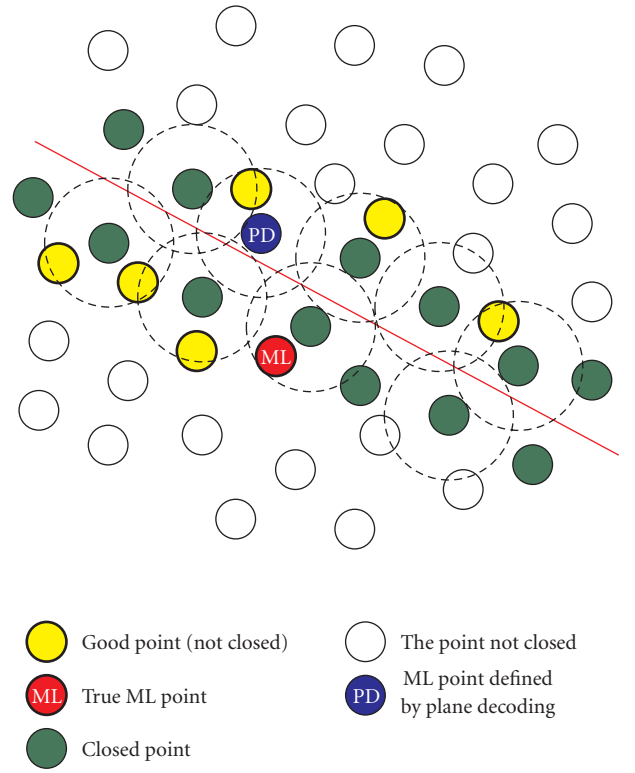


Figure 3: A diagram illustrating the good points obtained by including the neighboring points of some closed points.

To do so, note that

$$\frac{C_{\min}^2}{(N_0/2)} = \frac{\|\mathbf{n}\|^2}{(N_0/2)} \sim \chi_N^2 \tag{29}$$

is Chi-square distributed with N degrees of freedom which has the cumulative distribution function (cdf) given by

$$F_N(r) = \frac{\gamma(N/2, r/2)}{\Gamma(N/2)}, \tag{30}$$

where $\gamma(\cdot,\cdot)$ denotes the incomplete Gamma function and $\Gamma(\cdot)$ is the Gamma function. In other words, $C_{\min}$ is a random variable and $C_{\min}^2 \sim (N_0/2)\chi_N^2$. Therefore, to ensure the efficiency of the detector, the spherical search starts by choosing

$$C = \sqrt{\left(\frac{N_0}{2}\right)F_N^{-1}(i\Delta)}, \tag{31}$$

where $i = 1$ and Δ is judiciously set (e.g., typically $\Delta = 0.1$). If no point can be found, then C is updated by $i := i+1$. Since C is kept small at the beginning of the search, the detection complexity is minimized.

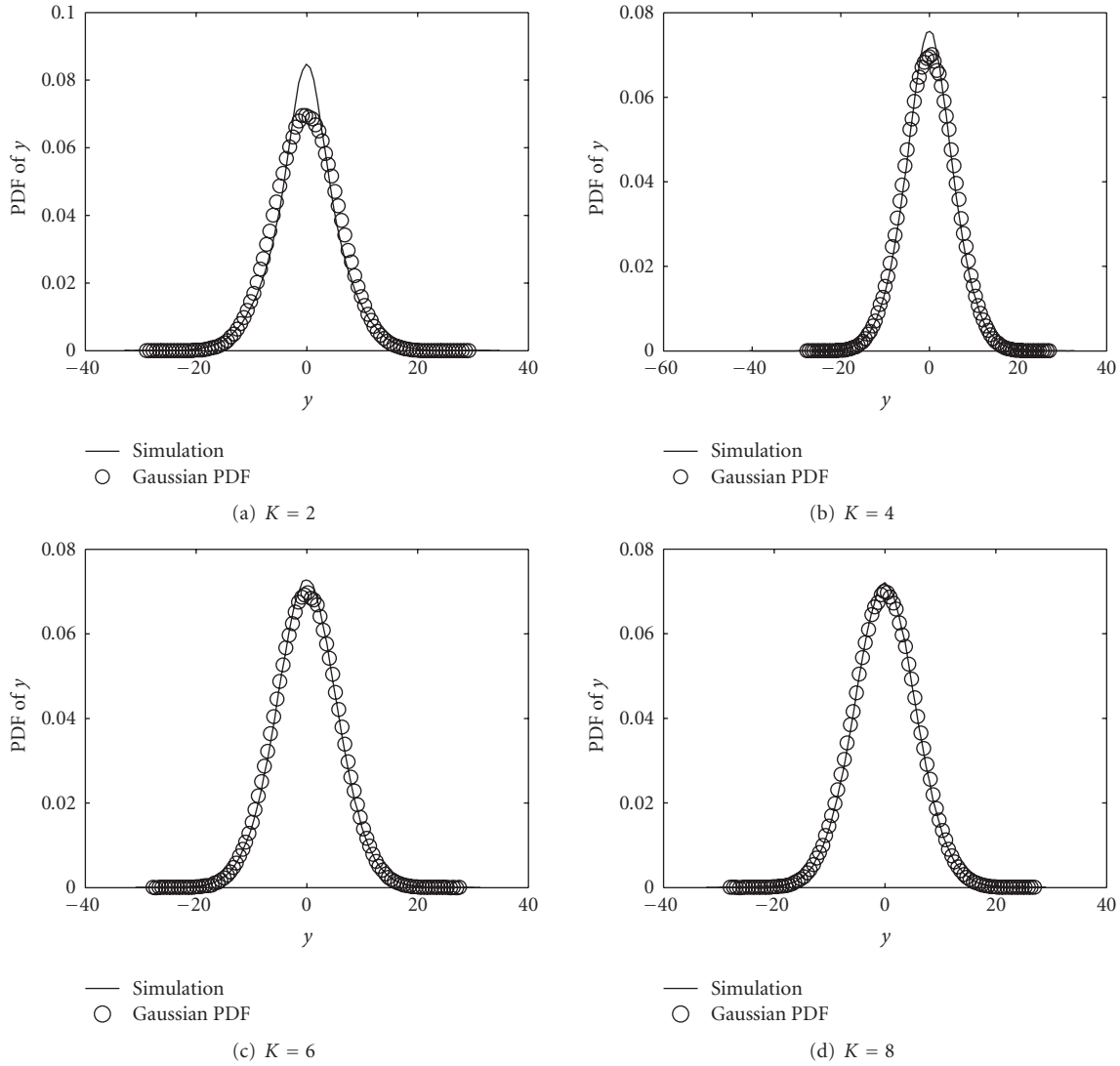


Figure 4: Comparison of the actual (or simulated) and Gaussian distributions for 4-PAM at SNR of 10 dB with various K.

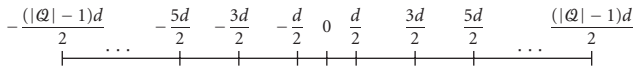


Figure 5: Signal constellations for $|\mathcal{Q}|$-PAM.

4. FIRST LOOK AT AVERAGE-CASE COMPLEXITY FOR A LARGE K

To gain understanding on the average-case complexity involved, this section is devoted to give the first look at the expected computational complexity of planar detection averaged over independent fading channels and the transmit sequences $\mathbf{x} \in \mathcal{Q}^K$ assuming an ideal implementation, that is, all the visited points during detection are jointly close or overheads are ignored. In particular, our analysis also relies on the assumption that K is large so that the distribution of the received signal can be worked out though numerical results in Figure 4 show that the approximation is surprisingly good for small K. The complexity exponent to be derived will serve as an average-case complexity measure for an ideal realization of planar detection and will provide a reasonable estimate on the average complexity for actual implementation of planar detection. To make the analysis succinct, the overall complexity involving both planar and spherical detections for a given radius C is not considered although we believe that it is possible with the help of the works in [6, 7].

Consider pulse-amplitude modulation (PAM) with $|\mathcal{Q}|$ constellations as shown in Figure 5, that is,

$$\mathcal{Q} = \{\pm 0.5d, \pm 1.5d, \ldots, \pm 0.5(|\mathcal{Q}|-1)d\}, \tag{32}$$

where d is the separation of every adjacent constellations. As a result, the average transmit energy per dimension, $\sigma_x^2 \triangleq \mathrm{E}[x_k^2]$, is given by

$$\sigma_x^2 = \frac{2}{|\mathcal{Q}|} \sum_{\ell=1}^{|\mathcal{Q}|/2} \left(\ell - \frac{1}{2}\right)^2 d^2. \tag{33}$$

Denoting $d_c = (|\mathcal{Q}| - 1)d$ as the distance between the two endpoints of $\mathcal{Q}$ and then simplifying (33), we have

$$\sigma_x^2 = \frac{d_c^2}{12}\left(\frac{|\mathcal{Q}|+1}{|\mathcal{Q}|-1}\right). \tag{34}$$

Note that if $x_k \in \mathcal{Q}$ [defined in (32)], then $\mathbf{x} \in \mathcal{Q}^K$ will form the lattice space of a hypercube. Nevertheless, for simplicity sake, we will approximate it by the lattice space of a hypersphere and this approximation greatly simplifies the calculation of the intersection between the decoding hyperplane and the lattice space later on. To preserve the same transmit energy, we set

$$d_c^K = \frac{\pi^{K/2}}{\mathcal{I}(K)} R^K, \tag{35}$$

where

$$\mathcal{I}(K) = \begin{cases} \left(\frac{K}{2}\right)! & \text{if } K \text{ is even,} \\ \frac{\sqrt{\pi}}{2^K}\frac{K!}{((K-1)/2)!} & \text{if } K \text{ is odd,} \end{cases} \tag{36}$$

and R denotes the radius of the lattice hypersphere

$$\mathcal{S} \;:\; x_1^2 + x_2^2 + \cdots + x_K^2 \le R^2, \quad \mathbf{x} \in \mathcal{Q}^K. \tag{37}$$

For an ideal implementation of planar detection, the average-case complexity allows the following expansion:

$$\mathcal{C}(K, \mathrm{SNR}) \approx 4\mathrm{E}[\mathcal{N}_{\mathrm{close}}], \tag{38}$$

where $\mathcal{N}_{\mathrm{close}}$ denotes the number of lattice points jointly close in all directions and we have used the fact that only four elementary computations are required for each visited point [see (20) and (21)]. To know the complexity $\mathcal{C}$, it requires the estimation of the average number of $\mathcal{N}_{\mathrm{close}}$.

For a given channel state $(h_1, h_2, \ldots, h_K)$ and the transmit lattice $\mathbf{x}$, the number of jointly close points, $\mathcal{N}_{\mathrm{close}}$, depends greatly on the intersectional area (or volume if $K > 3$) between the lattice space and the decoding hyperplane. This can be exemplified in Figure 6 where the number of jointly close points is shown for the decoding line $2x_1 + 3x_2 = y$ with various y. As can be seen, $\mathcal{N}_{\mathrm{close}}$ is, by and large, proportional to the sectional length between the decoding line and the lattice space. Also, it should be noted that $\mathcal{N}_{\mathrm{close}}$ depends largely on y (i.e., the shortest distance of the decoding line from the origin) but not the channel state. Therefore, for $K = 2$, we have

$$\mathcal{N}_{\mathrm{close}} \approx \left(\frac{N_{\max} - 1}{\mathcal{L}_{\max}}\right)\mathcal{L}(y) + 1, \tag{39}$$

where $\mathcal{L}(y)$ denotes the sectional length as a function of y and $N_{\max}$ denotes the maximal possible number of jointly close points. Equation (39) is written in a form such that when $\mathcal{L}(y) = 0$, then $\mathcal{N}_{\mathrm{close}} = 1$ while if $\mathcal{L}(y) = \mathcal{L}_{\max}$, then $\mathcal{N}_{\mathrm{close}} = N_{\max}$. In this example, $N_{\max} = 6$ which occurs when the decoding line intersects with the stepwise border

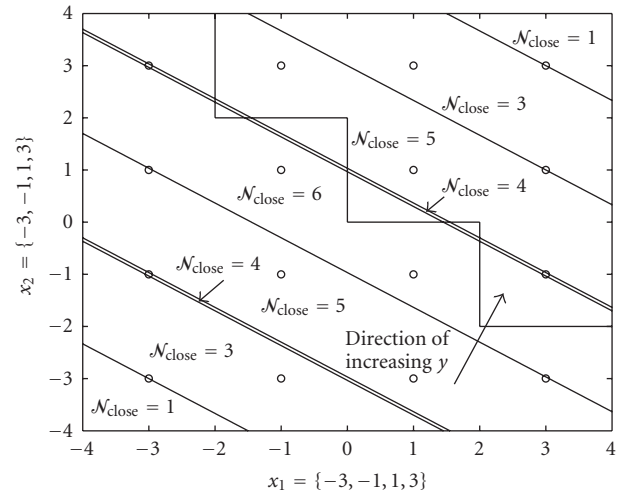


Figure 6: An example showing how $\mathcal{N}_{\mathrm{close}}$ depends on y assuming the decoding line $2x_1 + 3x_2 = y$ and $\mathcal{Q} = \{-3, -1, 1, 3\}$. The regions where the decoding line lies are labelled with the corresponding values of $\mathcal{N}_{\mathrm{close}}$.

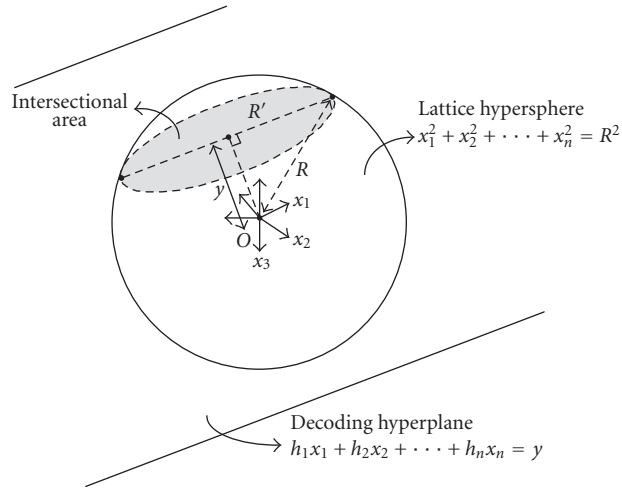


Figure 7: Geometry of the decoding hyperplane and the lattice hypersphere.

lines shown in Figure 6. For a general $|\mathcal{Q}|$-PAM, it can be easily seen that $N_{\max} \approx 2|\mathcal{Q}|$ if $K = 2$.

This idea is illustrated in Figure 7 and is generalized here for $K > 2$ so that

$$\mathcal{N}_{\mathrm{close}} = \left(\frac{N_{\max} - 1}{\mathcal{V}_{\max}}\right)\mathcal{V}(y) + 1, \tag{40}$$

where $\mathcal{V}$ denotesthe intersectional area or volume. Because the intersection of the decoding hyperplane $\mathcal{P}$ and the lattice hypersphere $\mathcal{S}$ is another hypersphere, the intersectional volume can be found as

$$\mathcal{V}(y) = \begin{cases} \frac{\pi^{(K-1)/2}}{\mathcal{I}(K-1)}\left(R^2 - y^2\right)^{(K-1)/2} & \text{if } |y| \le R, \\ 0 & \text{if } |y| > R. \end{cases} \tag{41}$$

Table 1: The average number of flops for planar detection with $J = 0$. The numbers in brackets indicate the required complexity for a BF detector.

(N_t, N_r)	4-QAM	16-QAM	64-QAM
(2,1)	191.5 (352)	335.2 (5.6k)	737.1 (90.1k)
(3,1)	368.7 (1.9k)	2.27k (123k)	25.5k (7.86M)
(3,2)	759.1 (4.35k)	1.26k (279k)	2.9k (17.8M)
(4,3)	1.87k (8.1k)	3.35k (516k)	10.15k (33M)

Table 2: The average number of flops for planar detection with $J = 10$.

(N_t, N_r)	4-QAM	16-QAM	64-QAM
(2,1)	210.6	426.5	863.0
(3,1)	429.0	2.448k	25.8k
(3,2)	786.0	1.425k	3.366k
(4,3)	1.91k	3.723k	11.89k

Table 3: The average number of flops for planar detection with $J = 20$.

(N_t, N_r)	4-QAM	16-QAM	64-QAM
(2,1)	210.6	471.4	947.6
(3,1)	453.5	2.56k	25.9k
(3,2)	786.0	1.47k	3.48k
(4,3)	1.91k	3.77k	12.13k

Table 4: The average number of flops for DLSD.

(N_t, N_r)	4-QAM	16-QAM	64-QAM
(2,1)	381.9	1.95k	8.01k
(3,1)	1.08k	17.7k	0.699M
(3,2)	1.0k	3.52k	20.8k
(4,3)	1.99k	4.65k	28.2k

Moreover, we know that $\mathcal{V}_{\max} = \mathcal{V}(0)$ and $N_{\max} \approx 2|\mathcal{Q}|^{K-1}$. Equation (40) can therefore be expressed as

$$\mathcal{N}_{\text{close}} = \begin{cases} (2|\mathcal{Q}|^{K-1} - 1)\left[1 - \left(\frac{y}{R}\right)^2\right]^{(K-1)/2} + 1 & \text{if } |y| \le R, \\ 1 & \text{if } |y| > R. \end{cases} \tag{42}$$

To find the expected value of $\mathcal{N}_{\text{close}}$, we need the probability distribution of y. Recognizing that

$$y = h_1 x_1^{(t)} + h_2 x_2^{(t)} + \cdots + h_K x_K^{(t)} + \eta, \tag{43}$$

where $(x_1^{(t)}, x_2^{(t)}, \ldots, x_K^{(t)})$ is the actual lattice point being transmitted, a random model for the channel, and the transmit lattice is required. In this paper, we assume that h_k's are i.i.d. zero-mean unit-variance Gaussian random variables, that is, $\mathrm{E}[h_k] = 0$ but $\mathrm{E}[h_k^2] = 1 \forall k$. Likewise, $x_k^{(t)}$'s are i.i.d. uniform distributed discrete random variables from the set

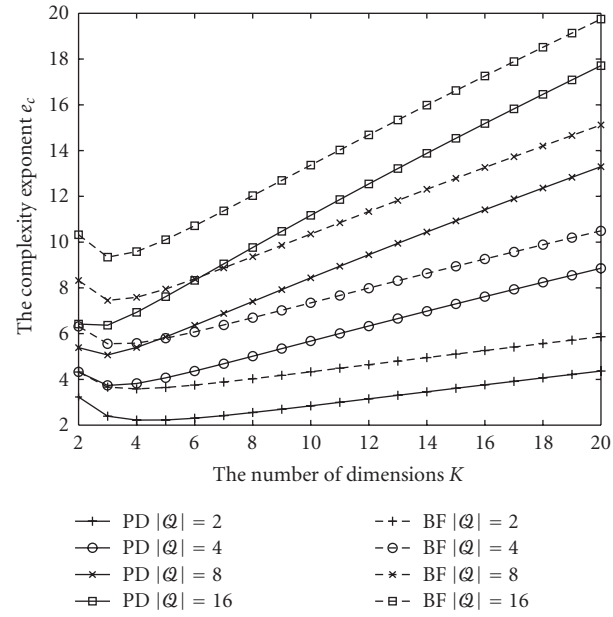


Figure 8: The complexity exponent as a function of K for various $|\mathcal{Q}|$ and SNR = 15 dB. PD refers to planar detection and BF refers to exhaustive ML detection.

$\mathcal{Q}$. If K is large, for example, $K > 3$, by the central limit theorem, y will be nearly Gaussian with

$$\begin{aligned} \mathrm{E}[y] &= 0, \\ \mathrm{E}[y^2] &= K\sigma_x^2 + \sigma_\eta^2 \end{aligned} \tag{44}$$

so that

$$f_Y(y) = \frac{1}{\sqrt{2\pi(K\sigma_x^2 + \sigma_\eta^2)}} e^{-y^2/2(K\sigma_x^2+\sigma_\eta^2)}. \tag{45}$$

Although the Gaussian approximation is analytically correct only at asymptotically large K, results in Figure 4 show that the approximation is surprisingly good even when K is small.

As a consequence, we get

$$\begin{aligned} \mathrm{E}[\mathcal{N}_{\text{close}}] &= \int_R^\infty f_Y(y)dy + \int_{-\infty}^{-R} f_Y(y)dy \\ &\quad + \int_{-R}^{R} \left\{ (2|\mathcal{Q}|^{K-1} - 1)\left[1 - \left(\frac{y}{R}\right)^2\right]^{(K-1)/2} + 1 \right\} \\ &\quad \times f_Y(y)dy \\ &= 1 + \frac{2|\mathcal{Q}|^{K-1} - 1}{\sqrt{2\pi}} \\ &\quad \times \int_{-R/\sqrt{K\sigma_x^2+\sigma_\eta^2}}^{R/\sqrt{K\sigma_x^2+\sigma_\eta^2}} \left[1 - \left(\frac{K\sigma_x^2 + \sigma_\eta^2}{R^2}\right) y^2\right]^{(K-1)/2} \\ &\quad \times e^{-y^2/2} dy. \end{aligned} \tag{46}$$

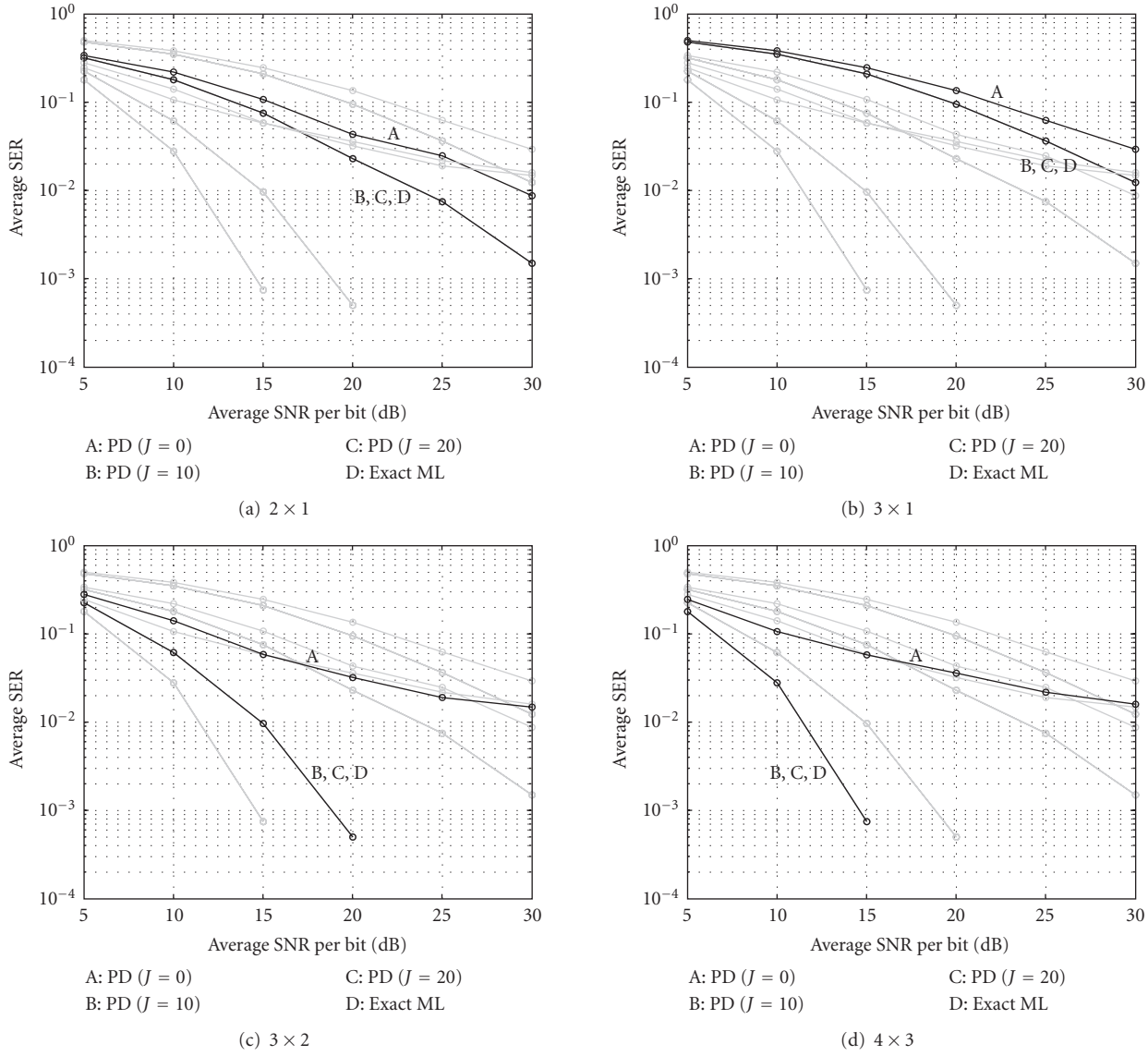


Figure 9: The average SER performance versus the average SNR per bit for 4-QAM modulation.

After some mathematical manipulations, it can further be simplified as

$$\mathrm{E}[\mathcal{N}_{\mathrm{close}}] = 1 + \frac{2|\mathcal{Q}|^{K-1} - 1}{\sqrt{2\pi}} \left(\frac{R}{\sqrt{K\sigma_x^2 + \sigma_\eta^2}} \right) \times \int_{-\pi/2}^{\pi/2} \cos^K \theta e^{-(R^2/2(K\sigma_x^2+\sigma_\eta^2))\sin^2\theta} d\theta. \tag{47}$$

Now, define

$$\mathrm{SNR} \triangleq \frac{\sum_{k=1}^{K} \mathrm{E}[x_k^2]}{\mathrm{E}[\eta^2]} = \frac{K\sigma_x^2}{\sigma_\eta^2} \tag{48}$$

and note that $R = \mu\sigma_x$ where

$$\mu = \sqrt[K]{\mathcal{A}(K)} \sqrt{\frac{12}{\pi}} \sqrt{\frac{|\mathcal{Q}| - 1}{|\mathcal{Q}| + 1}}. \tag{49}$$

As such, we have also

$$\frac{R}{\sqrt{K\sigma_x^2 + \sigma_\eta^2}} = \frac{\mu}{\sqrt{K}} \frac{\sqrt{\mathrm{SNR}}}{\sqrt{\mathrm{SNR} + 1}}. \tag{50}$$

To summarize, we can now rewrite $\mathcal{C}(K, \mathrm{SNR})$ as

$$\mathcal{C}(K, \mathrm{SNR}) = 4\left[1 + \frac{2|\mathcal{Q}|^{K-1} - 1}{\sqrt{2\pi}} \cdot \frac{\mu}{\sqrt{K}} \cdot \frac{\sqrt{\mathrm{SNR}}}{\sqrt{\mathrm{SNR} + 1}} \times \int_{-\pi/2}^{\pi/2} \cos^K \theta e^{-(\mu^2/2K)(\mathrm{SNR}/(\mathrm{SNR}+1))\sin^2\theta} d\theta \right] \tag{51}$$

with μ previously defined in (49). Additionally, the complexity exponent, e_c, can be readily found as

$$e_c = \frac{\log \mathcal{C}(K, \mathrm{SNR})}{\log K}. \tag{52}$$

It is often useful to consider e_c for complexity comparisons because if e_c approaches to a constant, then the complexity is polynomial, but if e_c grows like $K/\log K$, then the complexity is exponential. For brute-force (BF) ML detection of a MISO system, $\mathcal{C}(K, \text{SNR}) = |\mathcal{Q}|^K(2K+1)$ and the complexity exponent e_c will grow like $K/\log K$. Therefore, the complexity is exponential.

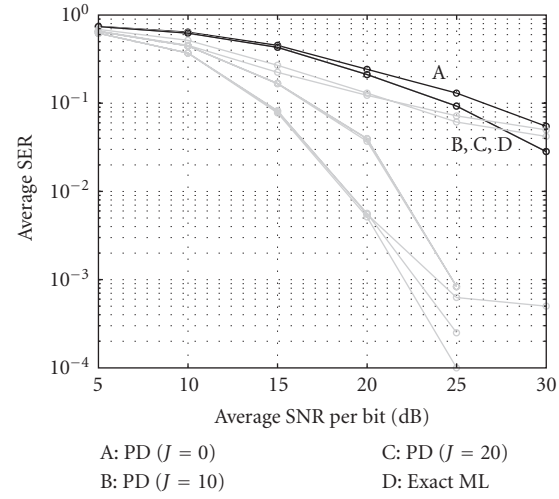
5. SIMULATION RESULTS

Computer simulations have been done to evaluate the average detection complexity and error performance. In the simulations, quasistatic flat fading channels have been assumed. To illustrate the complexity reduction, we provide the average number of floating point operations (flops) measured over many independent channel realizations at average SNR of 20 dB. In addition, results for the average symbol error rate (SER) performances for various average SNR are also provided.

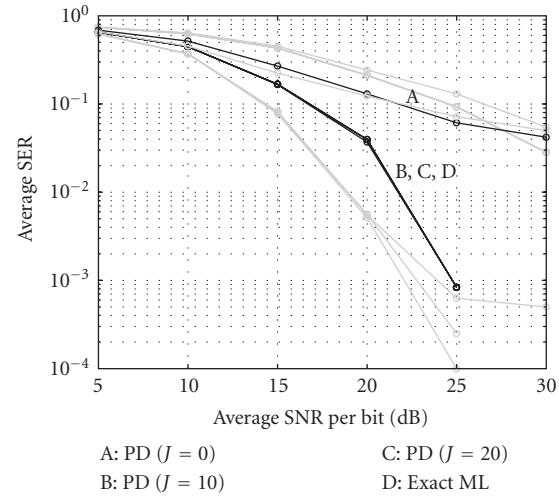
Table 1 reveals the average number of flops at average SNR of 20 dB for MIMO-ML detection using planar detection with $J = 0$ (the parameter used to design the set for good points) and a BF search [whose number of flops is $|\mathcal{Q}|^{N_t}(8N_t + 6)$ for MISO channels and $2|\mathcal{Q}|^{N_t}N_r(4N_t + 5)$ for MIMO channels]. In Table 1, the numbers in $(\cdot)$ indicate the results for BF-ML detection. Results in Tables 2 and 3 are, respectively, the measured average flops for planar-sphere detection when $J = 10$ and $J = 20$. Results conclude that significant reduction in complexity is possible especially for larger MIMO or higher level modulation. Moreover, only a mild increase in complexity is required for a larger J for improved SER (which will become apparent in the results of Figures 9 and 10).

Results showing the average complexity of the double-layer sphere decoder (DLSD) in [16] are also provided in Table 4 for comparison. A close observation of the results in this table indicates that for slightly overloaded systems [i.e., (3,2) and (4,3) 4-QAM systems], the proposed scheme and DLSD require similar average complexity; though DLSD, in general, has slightly greater average flops. However, for more overloaded systems such as (3,1) and higher-level modulations, we observe significant complexity savings of using the proposed scheme as compared to DLSD. In particular, results reveal that the proposed scheme with $J = 20$ requires only 14.5% of the complexity required by DLSD for a (3,1) 16-QAM system. Much complexity reduction is seen for a 64-QAM system where the proposed scheme with $J = 20$ requires only 11.8% and 3.7% of the complexity of DLSD, respectively, for (2,1) and (3,1) settings.

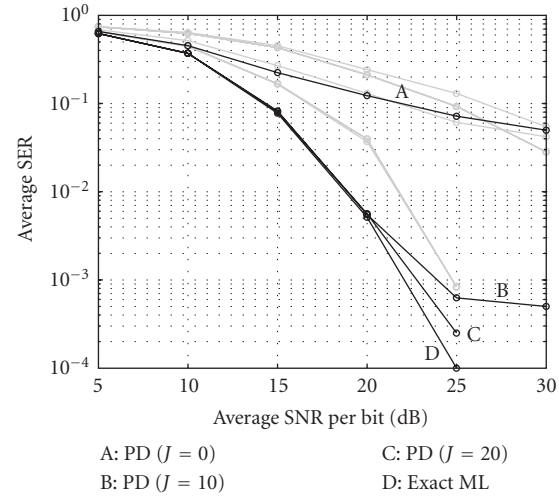
Using the expression derived in (52) [and (51)], simulations have also been conducted to evaluate the average-case complexity of planar detection for real-valued MISO systems for various size of signal constellations $|\mathcal{Q}|$ and various number of dimensions K. Complexity exponent results are plotted in Figure 8 for comparisons. In this figure, SNR = 15 dB has been assumed, but it should be noted that the results are insensitive to the SNR. Results show that the complexity of planar detection is still exponential as it grows linearly with K. This is however not surprising because there is no chan-



(a) 2×1



(b) 3×2



(c) 4×3

Figure 10: The average SER performance versus the average SNR per bit for 16-QAM modulation.

nel structure that can be exploited to simplify the detection (note that in a square MIMO, sphere decoding simplifies the detection by rotating the channel into a triangular structure). Havingthis in mind, planar detection in fact significantly reduces the complexity as compared toa BF-ML detector, the only choice for MISO ML. In particular, planar detection achieves a complexity reduction from e_c to $e_c - 2$ irrespective of the signal constellation size $|\mathcal{Q}|$ and the SNR without compromising the ML performance.

Figures 9 and 10 show the average SER results using planar-sphere detection with various J for 4-QAM and 16-QAM, respectively. Results for exact ML detection are provided for comparison. As we can see, near ML performance can be achieved for both 4-QAM and 16-QAM. For 4-QAM, $J = 0$ is sufficient while $J = 20$ is required for 16-QAM. It tells us that the set of closed points defined by the bottom layer is likely to contain the ML point so that optimal performance is preserved. For 16-QAM, this is not the case and SER degradation is observed. However, the ML point tends to be in the proximity of the closed points. Thus, the decoding performance improves with J.

6. CONCLUSION

This paper has presented a novel efficient detection algorithm that can achieve near ML performance for underdetermined MIMO channels. The complexity advantage arises from a better geometrical understanding of the ML-detection problem, which tells us that the ML point at the bottom decoding layer appears to be a point that is close to the decoding hyperplane in all coordinate axes. The fact that these points are much less is used to prune the decoding tree in order to reduce the complexity. Although the performance is not ML-guaranteed, the error performance can be largely recovered by visiting some extra points, which are potentially "good." Simulation results have demonstrated that significant reduction in complexityis obtained while achieving near ML performance. Average-case complexity for planar detection has also been derived with some approximations. The expression derived can serve as a complexity measure for an ideal realization of planar detection and provide a reasonable estimate on the average complexity for an actual implementation of a planar detector.

ACKNOWLEDGMENTS

This work was supported in part by Wai-Sun Leung Fellowships from the University of Hong Kong and in part by the Engineering and Physical Science Research Council (EPSRC) under Grant EP/E022308/1, United Kingdom.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[3] U. Fincke and M. Phost, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, 1985.

[4] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1591–1600, 1994.

[5] O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Communications Letters*, vol. 4, no. 5, pp. 161–163, 2000.

[6] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm—I: expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2806–2818, 2005.

[7] H. Vikalo and B. Hassibi, "On the sphere-decoding algorithm—II: generalizations, second-order statistics, and applications to communications," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2819–2834, 2005.

[8] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[9] N. D. Sidiropoulos and Z.-Q. Luo, "A semidefinite relaxation approach to MIMO detection for high-order QAM constellations," *IEEE Signal Processing Letters*, vol. 13, no. 9, pp. 525–528, 2006.

[10] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best Sphere decoding for MIMO detection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 491–503, 2006.

[11] Y. Xie, Q. Li, and C. N. Georghiades, "On some near optimal low complexity detectors for MIMO fading channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1182–1186, 2007.

[12] D. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *Proceedings of the DIMACS Workshop on Signal Processing for Wireless Transmission*, Series Discrete Math. and Theoretical Computer Science, American Mathematical Society, Piscataway, NJ, USA, October 2002.

[13] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[14] H. Vikalo, B. Hassibi, and T. Kailath, "Iterative decoding for MIMO channels via modified sphere decoding," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 2299–2311, 2004.

[15] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Generalized sphere decoder for asymmetrical space-time communication architecture," *Electronics Letters*, vol. 36, no. 2, pp. 166–167, 2000.

[16] Z. Yang, C. Liu, and J. He, "A new approach for fast generalized sphere decoding in MIMO Systems," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 41–44, 2005.

[17] A. M. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *Proceedings of IEEE International Conference on Communications (ICC '02)*, vol. 1, pp. 460–464, New York, NY, USA, April-May 2002.

[18] K.-K. Wong and A. Paulraj, "On the decoding order of MIMO maximum-likelihood sphere decoder: linear and non-linear receivers," in *Proceedings of IEEE 59th Vehicular Technology Conference (VTC '04)*, vol. 2, pp. 698–702, Milan, Italy, May 2004.