

# Efficient Joint Maximum-Likelihood Channel Estimation and Signal Detection

Haris Vikalo, Babak Hassibi, and Petre Stoica

**Abstract**—In wireless communication systems, channel state information is often assumed to be available at the receiver. Traditionally, a training sequence is used to obtain the estimate of the channel. Alternatively, the channel can be identified using known properties of the transmitted signal. However, the computational effort required to find the joint ML solution to the symbol detection and channel estimation problem increases exponentially with the dimension of the problem. To significantly reduce this computational effort, we formulate the joint ML estimation and detection as an integer least-squares problem, and show that for a wide range of signal-to-noise ratios (SNR) and problem dimensions it can be solved via sphere decoding with expected complexity comparable to the complexity of heuristic techniques.

**Index Terms**—Integer least-squares problem, sphere decoding, wireless communications, multiple-antenna systems, expected complexity, joint detection and estimation.

## I. INTRODUCTION

THE pursuit for high-speed data services has resulted in a tremendous amount of research activity in the wireless communications community. To obtain high reliability of the transmission, particular attention has been paid to the design of receivers (see, e.g., [1], [2], and the references therein).

In the system design, one often assumes knowledge of the channel coefficients at the receiver. These coefficients are typically obtained by sending a training sequence, thus sacrificing a fraction of the transmission rate. On the other hand, in practical systems, due to rapid changes of the channel and/or limited resources, training and channel tracking may be infeasible. One possible remedy is to differentially encode the transmitted data and thus eliminate the need for channel knowledge. Another one is to exploit known properties of the transmitted data to learn the channel blindly – for instance, one can exploit the fact that the transmitted data belong to a finite alphabet.

We consider a problem of joint maximum-likelihood (ML) channel estimation and signal detection in a wireless communication system with a block fading channel model, where the channel is constant for some interval  $T$ , after which it changes

to an independent value held constant for another interval  $T$ , and so on. The received signal can be written as

$$\mathbf{X} = \mathbf{H}\mathbf{S}^* + \mathbf{W} = \mathbf{H} \begin{bmatrix} \mathbf{S}_\tau^* & \mathbf{S}_d^* \end{bmatrix} + \mathbf{W}, \quad (1)$$

where  $\mathbf{H}$  denotes the  $N \times M$  channel matrix, and where  $(\cdot)^*$  denotes a conjugate transpose. Matrix  $\mathbf{S}_\tau$  is the  $T_\tau \times M$  matrix of training symbols, while the matrix  $\mathbf{S}_d$  is  $T_d \times M$  matrix of data symbols, and  $T = T_\tau + T_d$ . We assume that the components of  $\mathbf{S}_\tau$  and  $\mathbf{S}_d$  are elements from a PSK constellation and, therefore, are unitary, i.e.,

$$\begin{aligned} |s_\tau^{ij}|^2 &= 1, & 1 \leq i \leq T_\tau, 1 \leq j \leq M, \\ |s_d^{ij}|^2 &= 1, & 1 \leq i \leq T_d, 1 \leq j \leq M, \end{aligned} \quad (2)$$

where  $s_\tau^{ij}$  denotes the  $(i, j)$  entry in  $\mathbf{S}_\tau$ , and  $s_d^{ij}$  denotes the  $(i, j)$  entry in  $\mathbf{S}_d$ . Furthermore, the  $N \times T$  matrix  $\mathbf{W}$  in (1) is an additive noise matrix whose elements are assumed to be independent, identically distributed (iid) complex Gaussian random variables  $\mathcal{CN}(0, \sigma^2)$ .

In (1),  $N$  denotes the number of receive antennas. Depending on  $M$ , model (1) describes one of the following practical scenarios:

- 1)  $M = 1$  describes a single-input multi-output (SIMO), single-user (SU) system.
- 2)  $M > 1$  describes either a single-user multi-input multi-output (MIMO) SU system or a SIMO multi-user (MU) system. The former has been extensively studied (see, e.g., [3]). Therefore, in this paper we focus on the latter case, i.e., we focus on SIMO MU systems.

The joint ML channel estimation and signal detection problem can be stated as follows:

$$\min_{\mathbf{H}, \mathbf{S}} \|\mathbf{X} - \mathbf{H}\mathbf{S}^*\|^2, \quad (3)$$

Problem (3) is a mixed optimization problem: it is a least-squares problem in  $\mathbf{H}$  and an *integer least-squares* problem in  $\mathbf{S}$ . The solution to the integer least-squares problems may be found by an exhaustive search over the entire symbol space. However, the complexity of the exhaustive search is exponential (since the number of elements in the symbol space is  $|\mathcal{S}|^{MT_d}$ , where  $|\mathcal{S}|$  denotes number of symbols in the PSK constellation) and often infeasible in practice. Therefore, low-complexity heuristic techniques, usually iterating between the detection of  $\mathbf{S}$  and the estimation of  $\mathbf{H}$ , are often employed (see, e.g., [4], [5] and the references therein). On the other hand, in communication applications, the sphere decoding [6] is recognized as a technique for solving integer least-squares problems with often low expected complexity (see [7], [8]).

Manuscript received August 23, 2004; revised March 30, 2005; accepted May 17, 2005. The associate editor coordinating the review of this paper and approving it for publication was C. Xiao. This work was supported in part by the NSF under grant no. CCR-0133818, by the Office of Naval Research under grant no. N00014-02-1-0578, by Caltech's Lee Center for Advanced Networking, and by the Swedish Science Council (VR).

H. Vikalo and B. Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA (e-mail: hvikalo@systems.caltech.edu; hassibi@systems.caltech.edu).

P. Stoica is with Department of Information Technology, Uppsala University, Uppsala, Sweden (e-mail: peter.stoica@it.uu.se).

Digital Object Identifier 10.1109/TWC.2006.04559

In this paper, we show how the sphere decoding algorithm can be employed to find the solution of (3), the joint estimation and detection problem corresponding to a realistic and challenging communication scenario where the realization of the channel is unknown. Furthermore, we perform statistical analysis of the objective function of the aforementioned optimization problem at high signal-to-noise ratio (SNR), and obtain results which imply that the expected complexity results obtained in [7], [8] also hold for the scheme under consideration. Hence we conclude that the joint ML channel estimation and symbol detection problem (3) can be solved via sphere decoding with low expected complexity over a wide range of system parameters. We should remark that the basic sphere decoding algorithm performs the closest point search in a rectangular lattice and the available expected complexity results of [7], [8] assume the same. Therefore, when discussing complexity, we will assume that the entries of  $\mathbf{S}$  belong to a QPSK constellation. However, sphere decoding can easily be modified and used for detection of symbols coming from general PSK-modulation schemes [9].

Note that the model (1) includes both the blind and the training-based scheme – the blind scheme model is obtained by simply setting  $T_\tau = 0$ . In this paper, we consider both schemes. We start by considering a single user system with a single transmit antenna ( $M = 1$ ) but multiple receive antennas ( $N \geq 1$ , the SIMO MU case), for which we solve (3) for any  $T_\tau \geq 0$ . This solution is presented in Section II, while the complexity of the sphere decoding algorithm for solving (3) is discussed in Section III; the results therein imply low expected complexity of the algorithm over a wide range of system parameters. Section IV presents results for  $M > 1$  (the SIMO MU case). In Section V, we briefly discuss the use of soft sphere decoding for joint detection and decoding in systems employing channel codes. Simulation results are presented in Section VI, while the summary and conclusion are in Section VII.

Some of the results in this paper were originally presented in [10].

## II. SINGLE USER CASE

For  $M = 1$ , model (1) can be written as

$$\mathbf{X} = \mathbf{h}\mathbf{s}^* + \mathbf{W} = \mathbf{h} \begin{bmatrix} \mathbf{s}_\tau^* & \mathbf{s}_d^* \end{bmatrix} + \mathbf{W}, \quad (4)$$

where  $\mathbf{h}$  denotes the  $N \times 1$  channel vector,  $\mathbf{s}_\tau$  is the  $T_\tau \times 1$  vector of training symbols,  $\mathbf{s}_d$  is the  $T_d \times 1$  vector of data symbols, and  $\mathbf{W}$  is the  $N \times T$  noise matrix. The optimization problem that needs to be solved is

$$\min_{\mathbf{h}, \mathbf{s}_d} \|\mathbf{X} - \mathbf{h}\mathbf{s}^*\|^2. \quad (5)$$

In this section, we show how to solve (5) for any  $T_\tau$ , while the blind solution follows by setting  $T_\tau = 0$ .

For any given  $\mathbf{s}$ , the channel  $\hat{\mathbf{h}}$  that minimizes (5) is given by

$$\hat{\mathbf{h}} = \mathbf{X}\mathbf{s} / \|\mathbf{s}\|^2 = \frac{1}{T} \mathbf{X}\mathbf{s}, \quad (6)$$

where we used the assumption that the entries of  $\mathbf{s}$  have unit magnitude. Substituting (6) in (5) gives

$$\begin{aligned} \|\mathbf{X} - \mathbf{h}\mathbf{s}^*\|^2 &= \|\mathbf{X}(\mathbf{I} - \frac{1}{T}\mathbf{s}\mathbf{s}^*)\|^2 \\ &= \text{tr}[\mathbf{X}(\mathbf{I} - \frac{1}{T}\mathbf{s}\mathbf{s}^*)\mathbf{X}^*] \\ &= \text{tr}(\mathbf{X}\mathbf{X}^*) - \frac{1}{T}\mathbf{s}^*\mathbf{X}^*\mathbf{X}\mathbf{s} \end{aligned} \quad (7)$$

Hence the optimization (5) is equivalent to solving

$$\max_{\mathbf{s}_d} \mathbf{s}^*\mathbf{X}^*\mathbf{X}\mathbf{s}. \quad (8)$$

The optimization problem (8) is over vectors  $\mathbf{s}$  with discrete values (in particular, the components of  $\mathbf{s}$  are complex exponentials with imaginary exponents that are scaled integers); one way to solve it is via an exhaustive search over entire symbol space. However, exhaustive search is computationally inefficient and often cannot be used in practice. On the other hand, the sphere decoding algorithm efficiently solves optimization problems over integers. However, the sphere decoding algorithm *minimizes* an objective function over integer-valued vectors, and thus we need to express (8) accordingly. To this end, let  $\hat{\lambda} = \lambda_{\max}(\mathbf{X}^*\mathbf{X})$  denote the maximum eigenvalue of  $\mathbf{X}^*\mathbf{X}$ , and let  $\rho > \hat{\lambda}$  (for instance, we can choose  $\rho = \text{tr}(\mathbf{X}^*\mathbf{X})$ ). The problem (8) is then equivalent to

$$\min_{\mathbf{s}_d} \mathbf{s}^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{s}. \quad (9)$$

Observing the structure of the data vector in the model (4), the problem (9) can be written as

$$\min_{\mathbf{s}_d} \begin{bmatrix} \mathbf{s}_\tau^* & \mathbf{s}_d^* \end{bmatrix} (\rho\mathbf{I} - \mathbf{X}^*\mathbf{X}) \begin{bmatrix} \mathbf{s}_\tau \\ \mathbf{s}_d \end{bmatrix}. \quad (10)$$

For simplicity, we will find it useful to define

$$\rho\mathbf{I} - \mathbf{X}^*\mathbf{X} = \mathbf{\Gamma}.$$

Recall that we chose  $\rho$  such that  $\rho > \lambda_{\max}(\mathbf{X}^*\mathbf{X})$ . Therefore, matrix  $\mathbf{\Gamma}$  is positive definite by construction. Partition  $\mathbf{\Gamma}$  as

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{11} & -\mathbf{\Gamma}_{12} \\ -\mathbf{\Gamma}_{12}^* & \mathbf{\Gamma}_{22} \end{bmatrix},$$

where the dimensions of the matrices  $\mathbf{\Gamma}_{11}$ ,  $\mathbf{\Gamma}_{12}$ , and  $\mathbf{\Gamma}_{22}$ , are  $T_\tau \times T_\tau$ ,  $T_\tau \times T_d$ , and  $T_d \times T_d$ , respectively. Note that since  $\mathbf{\Gamma}$  is positive definite, so is  $\mathbf{\Gamma}_2$ . Therefore, we can write

$$\begin{aligned} \begin{bmatrix} \mathbf{s}_\tau^* & \mathbf{s}_d^* \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_{11} & -\mathbf{\Gamma}_{12} \\ -\mathbf{\Gamma}_{12}^* & \mathbf{\Gamma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{s}_\tau \\ \mathbf{s}_d \end{bmatrix} \\ &= \mathbf{s}_\tau^*\mathbf{\Gamma}_{11}\mathbf{s}_\tau - \mathbf{s}_\tau^*\mathbf{\Gamma}_{12}\mathbf{s}_d - \mathbf{s}_d^*\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau + \mathbf{s}_d^*\mathbf{\Gamma}_{22}\mathbf{s}_d \\ &= (\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau)^*\mathbf{\Gamma}_{22}(\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau) \\ &\quad + \mathbf{s}_\tau^*(\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*)\mathbf{s}_\tau \end{aligned}$$

Since  $\mathbf{s}_\tau^*(\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*)\mathbf{s}_\tau$  does not depend on  $\mathbf{s}_d$ , the optimization (10) becomes

$$\begin{aligned} \min_{\mathbf{s}_d} (\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau)^*\mathbf{\Gamma}_{22}(\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau) \\ = \min_{\mathbf{s}_d} \|\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau\|_{\mathbf{\Gamma}_{22}}^2. \end{aligned} \quad (11)$$

The optimization (11) corresponds to the joint ML channel estimation and signal detection problem for a general  $T_\tau \geq 0$ .

Clearly, when solving the problem blindly we have  $T_\tau = 0$ , and (11) takes the simpler form,

$$\min_{\mathbf{s}_d} \mathbf{s}_d^* (\rho \mathbf{I} - \mathbf{X}^* \mathbf{X}) \mathbf{s}_d.$$

The problem (11) is an integer least-squares optimization. [Note that, formally speaking, the entries of the unknown vector  $\mathbf{s}_d$  are not integer numbers; however, they do belong to a discrete set and thus (11) can be treated as an integer least-squares problem.] Since  $\mathbf{\Gamma}_{22}$  is positive definite, it has a Cholesky factorization of the form  $\mathbf{\Gamma}_{22} = \mathbf{R}^* \mathbf{R}$ , where  $\mathbf{R}$  is an upper-triangular matrix. Thus the sphere decoding algorithm of Fincke and Pohst [6] can be applied to solve (11). We should note that the original sphere decoding algorithm of [6] deals with real-valued vectors and matrices. To solve the complex-valued integer least-squares problem (11), we express all complex quantities via their real and imaginary components, and then solve the real-valued equivalent to (11). Alternatively, we may use the complex sphere decoding algorithm of [9] to solve (11) directly.

Rather than exhaustively searching over the entire symbol space, the sphere decoding algorithm performs a limited search inside a sphere of an appropriately chosen radius  $r$ , i.e., finds the point  $\mathbf{s}_d$  that minimizes (11) among all points  $\mathbf{s}_d$  that satisfy

$$\|\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1} \mathbf{\Gamma}_{12}^* \mathbf{s}_\tau\|_{\mathbf{\Gamma}_{22}}^2 \leq r^2. \quad (12)$$

The closest symbol point to  $\mathbf{\Gamma}_{22}^{-1} \mathbf{\Gamma}_{12}^* \mathbf{s}_\tau$  inside the sphere is the solution to (11). The choice of the radius  $r$  in (12) will be discussed in Section III.

The sphere decoding algorithm breaks down condition (12) into a set of  $T_d$  conditions that the components of the vector  $\mathbf{s}_d$  need to satisfy. The algorithm can be related to nulling and canceling (see, e.g., [11]) where, after a component of the vector  $\mathbf{s}_d$  that satisfies (12) is found, its contribution to  $\|\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1} \mathbf{\Gamma}_{12}^* \mathbf{s}_\tau\|_{\mathbf{\Gamma}_{22}}^2$  is subtracted. However, unlike in nulling and canceling, in sphere decoding the components of  $\mathbf{s}_d$  are not fixed until an entire vector  $\mathbf{s}_d$  which satisfies (12) is found. In particular, denote  $\hat{\mathbf{s}}_d = \mathbf{\Gamma}_{22}^{-1} \mathbf{\Gamma}_{12}^* \mathbf{s}_\tau$  and notice that we can expand the left-hand side of (12) to write (13), as shown at the top of the next page.

A necessary condition that (13) holds is that

$$|\mathbf{R}_{T_d, T_d}(\mathbf{s}_{d, T_d} - \hat{\mathbf{s}}_{d, T_d})|^2 \leq r^2. \quad (14)$$

Now, for any  $\mathbf{s}_{d, T_d}$  which satisfies (14), we state a stronger necessary condition in the equation shown after (13).

From this equation we find all possible  $\mathbf{s}_{d, T_d-1}$  such that (12) holds, and so on. The procedure continues until all components of  $\mathbf{s}_d$  which satisfy (12) are found. We omit any further details for brevity and refer the interested reader to [6], or to [7], [8] and the references therein. Furthermore, for an application of complex sphere decoding, relevant for detection of symbols coming from PSK constellations, we refer the reader to [9].

### III. COMPLEXITY OF THE ALGORITHM AND THE CHOICE OF SEARCH RADIUS

Choice of the search radius is crucial for the computational complexity of the sphere decoding algorithm. If the radius is

too small, we may not find any points inside the sphere and thus may have to repeat the search with a larger radius. On the other hand, if the radius is too big, the algorithm has to go through many points which requires a significant computation time.

A simple choice of the radius could be based on a heuristic solution to (8). For instance, an obvious heuristic for solving (8) consists of finding the eigenvector corresponding to the largest eigenvalue of  $\mathbf{X}^* \mathbf{X}$  and then projecting it onto the symbol space (i.e., rounding each entry of that eigenvector). This heuristic can be exploited as a starting point for the sphere decoding search – the norm corresponding to the heuristic solution can be used as the search radius. However, we cannot say much about the complexity of the sphere decoding algorithm corresponding to this deterministic choice of the search radius.

Alternatively, when the distribution of the objective norm function is known, one can choose the radius according to that distribution. For instance, in [7] the objective function that is being minimized by sphere decoding has a chi-square distribution, and the search radius is chosen using the cumulative density function of that distribution. In particular, the radius is chosen in such a way that the probability of finding a point inside the sphere is very high. Furthermore, the expected complexity of the sphere decoding algorithm for such a choice of the radius is found and shown to be comparable to the complexity of heuristic techniques over a wide range of SNR and problem dimensions.

In this section, we show that the objective function in (9) also has chi-square distribution at high SNRs. [Note that, for simplicity of the derivation, we will assume  $T_\tau = 0$ .] This distribution suggests a probabilistic choice of the search radius, while the expected complexity results of [7], [8] imply practical feasibility of the sphere decoding algorithm employed for solving (9). We start by proving the following lemma.

*Lemma 1:* At high SNR,  $\frac{1}{T\hat{\sigma}^2} \mathbf{s}^* (\rho \mathbf{I} - \mathbf{X}^* \mathbf{X}) \mathbf{s}$  is chi-square distributed with  $2(T-1)$  degrees of freedom, where  $\rho$  denotes the maximum eigenvalue of  $\mathbf{X}^* \mathbf{X}$ , and where  $\hat{\sigma}^2$  denotes an estimate of  $\sigma^2$ .<sup>1</sup>

**Proof:** Consider the eigenvalue decomposition of  $\mathbf{X}^* \mathbf{X}$ ,

$$\mathbf{X}^* \mathbf{X} = \begin{bmatrix} \hat{\mathbf{u}} & \hat{\mathbf{G}} \end{bmatrix} \begin{bmatrix} \hat{\lambda} & 0 \\ 0 & \hat{\mathbf{\Lambda}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}^* \\ \hat{\mathbf{G}}^* \end{bmatrix}, \quad (15)$$

where  $\hat{\lambda}$  denotes the largest eigenvalue, and where the diagonal elements of  $\hat{\mathbf{\Lambda}}$  are the remaining  $T-1$  eigenvalues in decreasing order. Taking  $\rho = \hat{\lambda}$ , we can write the objective function of the minimization (9) as

$$\begin{aligned} \mathbf{s}^* (\rho \mathbf{I} - \mathbf{X}^* \mathbf{X}) \mathbf{s} &= \mathbf{s}^* (\hat{\lambda} \mathbf{I} - \hat{\mathbf{u}} \hat{\mathbf{u}}^* \hat{\lambda} - \hat{\mathbf{G}} \hat{\mathbf{\Lambda}} \hat{\mathbf{G}}^*) \mathbf{s} \\ &= \mathbf{s}^* [\hat{\lambda} (\mathbf{I} - \hat{\mathbf{u}} \hat{\mathbf{u}}^*) - \hat{\mathbf{G}} \hat{\mathbf{\Lambda}} \hat{\mathbf{G}}^*] \mathbf{s} \\ &= \mathbf{s}^* \hat{\mathbf{G}} (\hat{\lambda} \mathbf{I} - \hat{\mathbf{\Lambda}}) \hat{\mathbf{G}}^* \mathbf{s}, \end{aligned} \quad (16)$$

where in going from the second to the third line above, we

<sup>1</sup>This estimate can be obtained (see, e.g., [12] and the references therein) as the mean of the  $(N-1)$  smallest eigenvalues of  $\mathbf{X} \mathbf{X}^*$  (or, alternatively, the smallest  $(N-1)$  non-zero eigenvalues of  $\mathbf{X}^* \mathbf{X}$ ).

$$|\mathbf{R}_{T_d, T_d}(s_{d, T_d} - \hat{s}_{d, T_d})|^2 + |\mathbf{R}_{T_d-1, T_d-1}(s_{d, T_d-1} - \hat{s}_{d, T_d-1}) - \mathbf{R}_{T_d-1, T_d}(s_{d, T_d} - \hat{s}_{d, T_d})| + \dots \leq r^2 \quad (13)$$

$$|\mathbf{R}_{T_d-1, T_d-1}(s_{d, T_d-1} - \hat{s}_{d, T_d-1}) - \mathbf{R}_{T_d-1, T_d}(s_{d, T_d} - \hat{s}_{d, T_d})|^2 \leq r^2 - |\mathbf{R}_{T_d, T_d}(s_{d, T_d} - \hat{s}_{d, T_d})|^2$$

used the fact that  $\mathbf{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^* = \hat{\mathbf{G}}\hat{\mathbf{G}}^*$ . Furthermore,

$$\begin{aligned} \mathbf{X}^*\mathbf{X} &= \|\mathbf{h}\|^2\mathbf{s}\mathbf{s}^* + \mathbf{s}\mathbf{h}^*\mathbf{W} + \mathbf{W}^*\mathbf{h}\mathbf{s}^* + \mathbf{W}^*\mathbf{W} \quad (17) \\ &= \hat{\lambda}\hat{\mathbf{u}}\hat{\mathbf{u}}^* + \hat{\mathbf{G}}\hat{\mathbf{G}}^*. \quad (18) \end{aligned}$$

Note that, using (17), we obtain

$$\begin{aligned} \hat{\mathbf{G}}^*(\mathbf{X}^*\mathbf{X})\mathbf{s} &= \hat{\mathbf{G}}^*(\|\mathbf{h}\|^2\mathbf{s}\mathbf{s}^* + \mathbf{s}\mathbf{h}^*\mathbf{W} + \mathbf{W}^*\mathbf{h}\mathbf{s}^* + \mathbf{W}^*\mathbf{W})\mathbf{s} \\ &= \lambda(\hat{\mathbf{G}}^*\mathbf{s}) + (\hat{\mathbf{G}}^*\mathbf{s})\mathbf{h}^*\mathbf{W}\mathbf{s} + \hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{h}T + \hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{W}\mathbf{s}. \quad (19) \end{aligned}$$

On the other hand, from (18),

$$\hat{\mathbf{G}}^*(\mathbf{X}^*\mathbf{X})\mathbf{s} = \hat{\lambda}\underbrace{\hat{\mathbf{G}}^*\hat{\mathbf{u}}}_{=\mathbf{0}}\hat{\mathbf{u}}\mathbf{s} + \underbrace{\hat{\mathbf{G}}^*\hat{\mathbf{G}}}_{=\mathbf{I}}\hat{\mathbf{A}}\hat{\mathbf{G}}^*\mathbf{s} = \hat{\mathbf{A}}(\hat{\mathbf{G}}^*\mathbf{s}). \quad (20)$$

Combining (19) and (20) leads to

$$\lambda(\hat{\mathbf{G}}^*\mathbf{s}) + (\hat{\mathbf{G}}^*\mathbf{s})\mathbf{h}^*\mathbf{W}\mathbf{s} + \hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{h}T + \hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{W}\mathbf{s} = \hat{\mathbf{A}}(\hat{\mathbf{G}}^*\mathbf{s}). \quad (21)$$

Let us analyze expression (21) at high SNR. Let  $\lambda$ ,  $\mathbf{A}$ ,  $\mathbf{u}$ , and  $\mathbf{G}$  denote the values of  $\hat{\lambda}$ ,  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{u}}$ , and  $\hat{\mathbf{G}}$ , respectively, in the limit of  $SNR \rightarrow \infty$  (i.e.,  $\sigma^2 \rightarrow 0$ ). Note that, since  $\mathbf{X} = \mathbf{h}\mathbf{s}^* + \mathbf{W}$ , in the high SNR regime we can write

$$\mathbf{X}^*\mathbf{X} \approx \|\mathbf{h}\|^2 T \begin{pmatrix} \mathbf{s} \\ \sqrt{T} \end{pmatrix} \begin{pmatrix} \mathbf{s}^* \\ \sqrt{T} \end{pmatrix}. \quad (22)$$

By comparing expressions (15) and (22) we obtain

$$\lambda = T\|\mathbf{h}\|^2, \quad \mathbf{A} = \mathbf{0}_{(T-1) \times (T-1)}, \quad \mathbf{u} = \frac{\mathbf{s}}{\sqrt{T}}. \quad (23)$$

It is easy to see from (23) that, at high SNR,  $\hat{\mathbf{G}}^*\mathbf{s} \rightarrow \sqrt{T}\mathbf{G}^*\mathbf{u} = \mathbf{0}$ . Thus  $(\hat{\mathbf{G}}^*\mathbf{s})\mathbf{h}^*\mathbf{W}\mathbf{s}$  and  $\hat{\mathbf{A}}(\hat{\mathbf{G}}^*\mathbf{s})$  can be treated as the higher-order terms which, as  $\sigma^2 \rightarrow 0$ , diminish much faster than the other terms in (23). By neglecting the higher-order terms in (21), we obtain that, at high SNR,

$$\lambda(\hat{\mathbf{G}}^*\mathbf{s}) + \hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{h}T \approx 0 \Rightarrow \lambda(\hat{\mathbf{G}}^*\mathbf{s}) \approx -\hat{\mathbf{G}}^*\mathbf{W}^*\mathbf{h}T,$$

which, in the SNR regime under consideration, we can further write as  $\lambda(\hat{\mathbf{G}}^*\mathbf{s}) \approx -\mathbf{G}^*\mathbf{W}^*\mathbf{h}T$ . Therefore, at high SNR,  $(\hat{\mathbf{G}}^*\mathbf{s})$  is circular Gaussian with zero mean. To find its variance, note that

$$T\mathbf{G}^*\mathbf{W}^*\mathbf{h} = T\mathbf{G}^* \begin{bmatrix} \mathbf{W}_1^*\mathbf{h} \\ \vdots \\ \mathbf{W}_N^*\mathbf{h} \end{bmatrix},$$

where  $\mathbf{W}_k$  denotes the  $k^{\text{th}}$  column of  $\mathbf{W}$ . Also, note that

$$E[\mathbf{W}_k^*\mathbf{h}\mathbf{h}^*\mathbf{W}_l] = E[\mathbf{h}^*\mathbf{W}_l\mathbf{W}_k^*\mathbf{h}] = \sigma^2\|\mathbf{h}\|^2\delta_{k,l},$$

and thus

$$\begin{aligned} \text{cov}(\hat{\mathbf{G}}^*\mathbf{s}) &= \frac{T^2}{\lambda^2} \text{cov}(\mathbf{G}^*\mathbf{W}^*\mathbf{h}) \\ &= \frac{T^2}{\lambda^2} \|\mathbf{h}\|^2 \sigma^2 \mathbf{I} = \frac{T}{\lambda} \sigma^2 \mathbf{I} \\ &= T\sigma^2(\lambda\mathbf{I} - \mathbf{A})^{-1}, \end{aligned}$$

where we used the fact that  $\lambda = T\|\mathbf{h}\|^2$ , and where  $\mathbf{A} = \mathbf{0}$  was inserted for convenience. Therefore,  $(\hat{\mathbf{G}}^*\mathbf{s})$  is circular Gaussian with zero mean and covariance  $T\sigma^2(\lambda\mathbf{I} - \mathbf{A})^{-1}$ , i.e.,

$$(\hat{\mathbf{G}}^*\mathbf{s}) \sim \mathcal{CN}[0, T\sigma^2(\lambda\mathbf{I} - \mathbf{A})^{-1}].$$

Once again we recall that, since at high SNR,  $\lambda \approx \hat{\lambda}$  and  $\mathbf{A} \approx \hat{\mathbf{A}}$ ; therefore, it approximately holds that

$$(\hat{\mathbf{G}}^*\mathbf{s}) \sim \mathcal{CN}[0, T\sigma^2(\hat{\lambda}\mathbf{I} - \hat{\mathbf{A}})^{-1}].$$

Therefore, we have shown that the scaled term in (16),  $\frac{1}{T\sigma^2}(\mathbf{s}^*\hat{\mathbf{G}})(\hat{\lambda}\mathbf{I} - \hat{\mathbf{A}})(\hat{\mathbf{G}}^*\mathbf{s})$ , is chi-square distributed with  $2(T-1)$  degrees of freedom, and hence so is  $\frac{1}{T\sigma^2}\mathbf{s}^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{s}$ .  $\square$

We use the result of Lemma 1 to choose the search radius for sphere decoding. Recall that the probability density function of the chi-square distribution is given by the normalized incomplete gamma function. Therefore, we can choose the search radius

$$r^2 = \alpha(T-1)\hat{\sigma}^2 \quad (24)$$

so that

$$\int_0^{\alpha(T-1)} \gamma(t, 2(T-1)) dt = \int_0^{\alpha(T-1)} \frac{t^{T-2}}{\Gamma(T-1)} e^{-t} dt = 1 - \epsilon, \quad (25)$$

where  $\gamma(\cdot, \cdot)$  denotes the normalized incomplete gamma function. Furthermore,  $1 - \epsilon$  is set close to 1, say, 0.99. This means that the optimal point is inside the sphere with probability  $1 - \epsilon$ . If the algorithm does not find any points inside the sphere, radius is increased and the search is repeated with the new radius.

We should further remark that in order to ensure  $\rho\mathbf{I} - \mathbf{X}^*\mathbf{X} > 0$  and prevent any possible numerical problems in applications, we suggest a choice of  $\rho > \hat{\lambda}$ , as already stated in Section 2. Then the search radius should be modified and it may be chosen as  $r' = (\rho - \hat{\lambda}) + r$ , where  $r$  is obtained as implied by (25).

So far in this section we assumed  $T_\tau = 0$ , for which the search radius is simply the  $r^2$  given by (24). To select the radius for  $T_\tau > 0$ , recall that the objective in (11) can be written as

$$\begin{aligned} \|\mathbf{s}_d - \mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*\mathbf{s}_\tau\|_{\mathbf{\Gamma}_{22}}^2 \\ = \mathbf{s}^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{s} - \mathbf{s}_\tau^*(\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*)\mathbf{s}_\tau \end{aligned}$$

Hence we choose the radius of the sphere decoding algorithm for solving (11) as

$$r_\tau^2 = r^2 - \mathbf{s}_\tau^*(\mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^*)\mathbf{s}_\tau. \quad (26)$$

#### IV. MULTI-USER CASE

The relevant optimization problem now is the one in (3). For any  $\mathbf{S}$ , the optimal channel estimate  $\hat{\mathbf{H}}$  is obtained as

$$\hat{\mathbf{H}} = \mathbf{X}\mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1},$$

and thus we have

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{S}} \|\mathbf{X} - \mathbf{H}\mathbf{S}^*\|^2 &= \min_{\mathbf{S}} \|\mathbf{X} - \mathbf{X}\mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\|^2 \\ &= \min_{\mathbf{S}} \text{tr} \{ \mathbf{X}(\mathbf{I} - \mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*)\mathbf{X}^* \} \\ &= \text{tr}(\mathbf{X}\mathbf{X}') \\ &\quad - \max_{\mathbf{S}} \text{tr} \{ (\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}^*\mathbf{X}\mathbf{S} \} \end{aligned}$$

If the columns of  $\mathbf{S}$  are approximately orthogonal, i.e., if  $\mathbf{S}^*\mathbf{S} \approx T\mathbf{I}$ , it follows that

$$\begin{aligned} &\arg \max_{\mathbf{S}} \text{tr} \{ (\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{X}^*\mathbf{X}\mathbf{S} \} \\ &\approx \arg \max_{\mathbf{S}_1, \dots, \mathbf{S}_M} \sum_{k=1}^M \mathbf{S}_k^*(\mathbf{X}^*\mathbf{X})\mathbf{S}_k, \end{aligned}$$

where  $\mathbf{S}_k$  denotes the  $k^{\text{th}}$  column in matrix  $\mathbf{S}$ . Furthermore,

$$\max_{\mathbf{S}_1, \dots, \mathbf{S}_M} \sum_{k=1}^M \mathbf{S}_k^*(\mathbf{X}^*\mathbf{X})\mathbf{S}_k = \min_{\mathbf{S}_1, \dots, \mathbf{S}_M} \sum_{k=1}^M \mathbf{S}_k^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{S}_k,$$

and thus the optimization (3) is equivalent to

$$\min_{\mathbf{S}_1, \dots, \mathbf{S}_M} \sum_{k=1}^M \mathbf{S}_k^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{S}_k, \quad \text{s.t. } \mathbf{S}^*\mathbf{S} = T\mathbf{I}. \quad (27)$$

Without the constraint  $|s^{ij}|^2 = 1$  on the entries of  $\mathbf{S}$ , (27) is a challenging optimization problem over Stiefel manifold (see, e.g., [15]). The additional alphabet constraint makes it even more difficult. We propose an efficient approximate solution and illustrate the procedure for  $M = 2$ .

We start by using sphere decoding to solve

$$\min_{\mathbf{S}_1} \mathbf{S}_1^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{S}_1. \quad (28)$$

Denote the solution to the above optimization by  $\hat{\mathbf{S}}_1$  and define

$$\mathbf{X}^{(2)} = \mathbf{X}^{(1)} - \hat{\mathbf{H}}_1\hat{\mathbf{S}}_1^* = \mathbf{X} - \frac{1}{T}\mathbf{X}\hat{\mathbf{S}}_1\hat{\mathbf{S}}_1^*, \quad (29)$$

where  $\mathbf{X}^{(1)} = \mathbf{X}$ , and where  $\hat{\mathbf{H}}_1$  denotes the ML estimate of the first column of the matrix  $\hat{\mathbf{H}}$ , computed according to (6). In other words,  $\mathbf{X}^{(2)}$  is obtained by subtracting (canceling) the contribution of the decoded first user from the received signal  $\mathbf{X}$ . Then we use sphere decoding to solve

$$\min_{\mathbf{S}_2} \mathbf{S}_2^*(\rho\mathbf{I} - (\mathbf{X}^{(2)})^*\mathbf{X}^{(2)})\mathbf{S}_2. \quad (30)$$

The following points are worthy of mention. Note that in the above discussion we assumed a general  $T_\tau \geq 0$ ; clearly, optimizations (28) and (30) can be solved as in the  $M = 1$  scenario discussed in Section 2. Furthermore, as stated in Section 1, the model with  $M > 1$  that we focus on is the SIMO MU system. This, for instance, describes an uplink in a cellular system, where each user (transmitter) has a single antenna, while the base station (receiver) has multiple antennas (whose number is denoted by  $N$  in our model). In such systems, to

ensure that the above orthogonality assumption holds we need to impose some additional coding.

The validity of the previous algorithm is predicated on assumption that the users transmit orthogonal symbol vectors. However, for  $M = 2$  user case and BPSK symbols which we study in Example 2 of Section VI, we demonstrate that the above algorithm performs well even if the orthogonality condition is removed. On the other hand, for  $M \geq 3$ , we may in principle proceed in a similar fashion – in particular, we may perform sphere decoding to detect the next user, subtract its contribution from the received signal, and repeat the previous steps. However, the aforementioned orthogonality assumption becomes rather important as the number of users increases and the interference starts limiting achievable performance, and we indeed need some form of coding to provide the orthogonality.

Finally, we note that the choice of search radius for the multi-user case is not obvious, since the results of Theorem 1 hold for the single-user case and cannot be applied directly here. Some heuristics for the choice of the radius in a 2-user case are proposed in Example 2 of Section VI. Furthermore, note that in practice more sophisticated versions of sphere decoding are used, in which the search radius is reset every time a point inside the sphere is found. There, a quick heuristic technique may be employed to determine an initial radius, and such radius initialization can be used here as well.

#### V. REMARKS ON SOFT DECODING OF CHANNEL CODES AND ON FURTHER DECREASE IN THE COMPLEXITY

So far, we have assumed that the users send uncoded data. On the other hand, if the users employ channel codes, we can use soft sphere decoding techniques for iterative decoding. To this end, one can use the modified Fincke-Pohst maximum a posteriori (FP-MAP) algorithm proposed in [17], or the list sphere decoding proposed in [9]. Both modifications exploit the fact that sphere decoding finds *all* points that satisfy the sphere constraint, e.g., in the single-user case (say, with  $T_\tau = 0$ ), finds all vectors  $\mathbf{s}$  such that

$$\|\mathbf{X} - \mathbf{h}\mathbf{s}\|^2 \leq r^2. \quad (31)$$

For the FP-MAP algorithm, instead of (31), the relevant condition is  $\|\mathbf{X} - \mathbf{h}\mathbf{s}\|^2 \leq r^2 + \sum \log p(\mathbf{s}_i)$ , where  $p(\mathbf{s}_i)$  is the *a priori* information about the  $i^{\text{th}}$  component of  $\mathbf{s}$ . Since the additional term  $\sum \log p(\mathbf{s}_i)$  does not couple components of the symbol vector, the sphere decoding algorithm only needs to be modified by adding an appropriate term to each of the necessary conditions described in Section 2. For instance, condition (14) now becomes  $|\mathbf{R}_{T_d, T_d}(\mathbf{s}_{d, T_d} - \hat{\mathbf{s}}_{d, T_d})|^2 \leq r^2 + \log(p_{T_d})$ . Remaining conditions are adjusted similarly.

By an appropriate choice of the radius, one can guarantee that the algorithm will return a number of points, which are then used to make a *soft* decision about the components of the vector  $\mathbf{s}$ . Assuming that the transmitter, prior to modulation, employs channel codes (such as convolutional, turbo, LDPC, etc.), the soft decisions made by the modified sphere decoder are then iterated with the soft decisions made by the channel decoder. For brevity, we omit any details and refer the reader to [9], [17].

On another note, in this paper we have been primarily concerned with posing the problem of joint ML channel

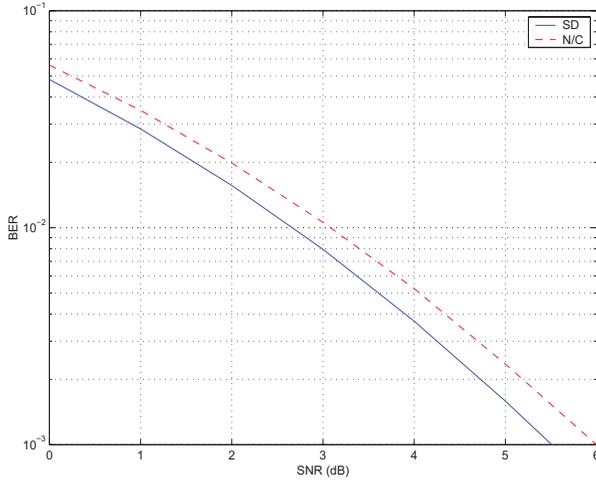


Fig. 1. BER performance comparison of the sphere decoding and N/C algorithms,  $M = 1$ ,  $N = 6$ ,  $T = 10$ , 4-PSK.

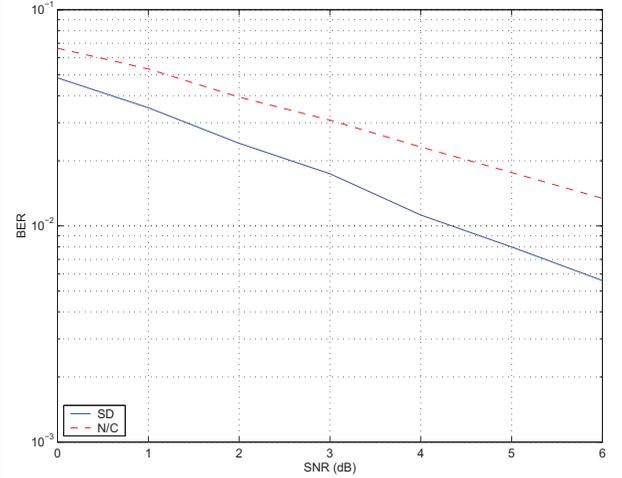


Fig. 3. BER performance comparison of sphere decoding and nulling and canceling algorithms,  $M = 2$ ,  $N = 4$ ,  $T = 10$ , 2-PSK.

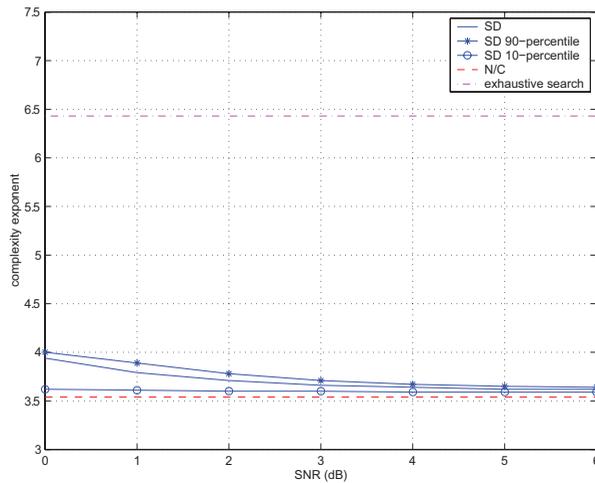


Fig. 2. Complexity exponents of sphere decoding (expected, 90-percentile, and 10-percentile), N/C, and exhaustive search,  $M = 1$ ,  $N = 6$ ,  $T = 10$ , 4-PSK.

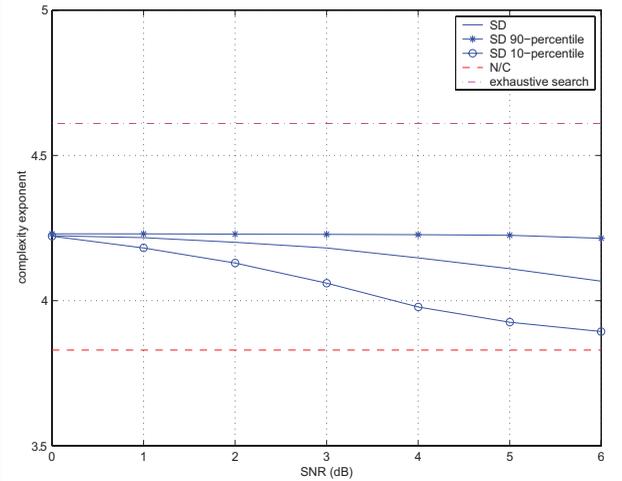


Fig. 4. Complexity exponents of sphere decoding (expected, 90-percentile, and 10-percentile), N/C, and exhaustive search,  $M = 2$ ,  $N = 4$ ,  $T = 10$ , 2-PSK.

estimation and signal detection in a form to which standard sphere decoding can be applied, and establishing a relation with the existing analytical treatment of the complexity of sphere decoding. There are many practical realizations of sphere decoding that, in terms of computational complexity, perform even better than the standard sphere decoding algorithm of [6]. For instance, the Schnorr-Euchner version of sphere decoding presented in [13], and further analyzed in [14], the radius scheduling suboptimal scheme introduced in [16], etc. Applying those techniques to further lower the complexity of sphere decoding in the context of joint ML estimation and detection may be of interest.

## VI. SIMULATION RESULTS

In this section, we give several examples that illustrate the performance and the corresponding complexity of the sphere decoding algorithm in the previously described scenarios.

**Example 1 [Single-user system]:** Consider a single user which employs one antenna, while the receiver employs  $N = 6$  antennas. The data is transmitted in blocks of length  $T = 10$

(in principle, we consider a blind scheme but embed one known symbol to remove phase ambiguity inherent to the blind schemes). We compare the BER performance of the sphere decoding algorithm and the nulling and canceling (N/C) algorithm. Simulation results are obtained by performing Monte Carlo runs in which  $\mathbf{h}$  and  $\mathbf{W}$  are varied. The vector  $\mathbf{h}$  is comprised of iid complex Gaussian random variables  $\mathcal{CN}(0, 1)$ . Components of the symbol vector  $\mathbf{s}$  are chosen from an 4-PSK (i.e., 4-QAM) constellation.

As shown in Fig. 1, the sphere decoding algorithm outperforms the N/C algorithm over the considered range of SNR. Fig. 2 shows the complexity exponent, defined as

$$e = \log_{2MT_d} F,$$

where  $F$  denotes the total number of operations required to detect a complex-valued vector  $\mathbf{s}$  (in particular, we empirically count the flops required to find the solution, including the operations that may occur due to possible radius increase). It is evident that the computational burden of sphere decoding is not significantly larger than that of nulling and canceling.

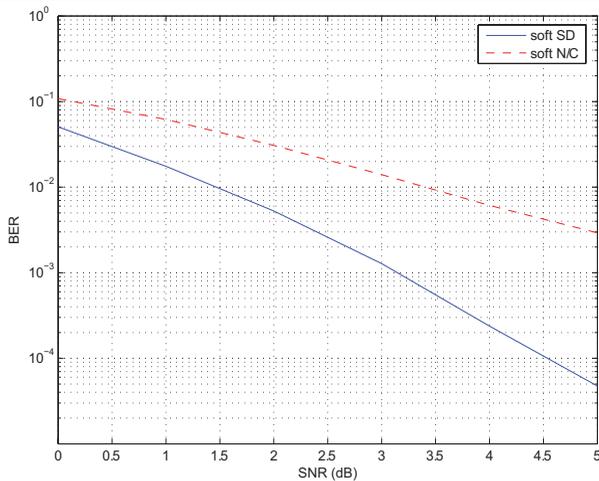


Fig. 5. BER comparison of soft sphere decoding and soft nulling and canceling,  $M = 1$ ,  $N = 4$ ,  $T = 10$ , 4-PSK,  $R = 1/2$  convolutional code of length 1000.

**Example 2 [Multi-user system]:** Consider a system with  $M = 2$  users, each one employing a single antenna, while the receiver employs  $N = 4$  antennas. The channel matrix  $\mathbf{H}$  is comprised of iid complex Gaussian random variables  $\mathcal{CN}(0, 1)$ . The data is transmitted in blocks of length  $T = 10$  (as in the single user case, one known symbol is embedded in each user's symbol sequence in order to remove phase ambiguity). The users' symbol sequences are generated randomly and independently, i.e., we do not impose orthogonality criteria on the users.

We compare the BER performance of the sphere decoding algorithm and the N/C algorithm. As shown in Fig. 3, the sphere decoding algorithm outperforms the N/C algorithm, particularly for high SNR. Fig. 4 shows the complexity exponent. The complexity exponents of sphere decoding and the N/C algorithms are comparable over the range of SNRs of interest. The complexity exponent of the scenario where exhaustive search is used in place of sphere decoding is also shown. The search radius for the sphere decoding algorithm is chosen as  $r^2 = 2T + 2\alpha(T - 1)\hat{\sigma}^2$ , i.e., it is obtained by adding the norm of each users signal,  $T$ , to the value of the search radius suggested by Theorem 1. The simulation results indicate that this particular choice of radius enables finding a point inside the sphere with high probability, as required by the algorithm.

**Example 3 [Single-user system, coded data]:** We consider again the single user system with  $N = 4$  receive antennas. The information bit sequence comprising 500 information bits is encoded by a rate  $R = 1/2$ , (7, 5) convolutional code encoded by a systematic encoder with feedback, corresponding to the generator matrix  $G(D) = [1 \quad (1 + D^2)/(1 + D + D^2)]$ . The coded sequence is modulated by means of simple Gray mapping onto a 4-PSK modulation scheme and then transmitted in blocks of  $T = 10$ . On the receiver side, the FP-MAP algorithm of [17] is used for soft detection.

Fig. 5 compares the performance of the FP-MAP with that of the nulling and canceling algorithm with soft outputs. For each entry in a transmitted symbol vector, the N/C algorithm with soft outputs cancels the previously decoded symbols and

obtains the soft information using the distribution of the noise. The soft N/C algorithm is similar to the soft MMSE equalizer of [19], and has been employed in a multi-user context in [18]. Both schemes perform a single iteration between the soft detector and the soft convolutional decoder. The soft sphere decoding algorithm significantly outperforms the soft N/C algorithm over the considered range of SNR.

## VII. SUMMARY AND CONCLUSION

We considered the joint ML channel estimation and signal detection problem for single-input multiple-output wireless channels. To reduce the computational effort, we posed the problem so that it can be solved via the use of sphere decoding. It was shown that the algorithm, when applied to the problem herein, has expected complexity comparable to the complexity of heuristic techniques over a wide range of system parameters. We treated both single-user and multi-user scenarios. For the latter, we proposed an heuristic algorithm based on nulling and canceling. Simulations illustrated the performance of the algorithm and its complexity. When compared with the best available heuristics, the sphere decoding provided much better performance at a comparable complexity, over a wide range of SNR.

There are several directions for future work and possible extensions of the current work. For instance, applying more powerful variations of sphere decoding (such as Schnorr-Euchner and radius scheduling techniques) to further lower the complexity of sphere decoding in the context of joint ML estimation and detection is of interest.

## REFERENCES

- [1] A. Paulraj, D. Gore, and R. Nabar, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.
- [2] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.
- [3] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041-2052, Dec. 2000.
- [4] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266-1270, Aug. 2003.
- [5] P. Bohlin, "Iterative Least Square Techniques with Applications to Adaptive Antennas and CDMA Systems." master's thesis, Chalmers University of Technology, 2001.
- [6] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463-471, Apr. 1985.
- [7] B. Hassibi and H. Vikalo, "On sphere decoding algorithm. I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2819-2834, Aug. 2003.
- [8] H. Vikalo and B. Hassibi, "On sphere decoding algorithm. II. Generalizations, second-order statistics, and applications to communications," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2819-2834, Aug. 2003.
- [9] B. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389-399, Mar. 2003.
- [10] P. Stoica, H. Vikalo, and B. Hassibi, "On joint ML channel estimation and signal detection for SIMO channels," in *Proc. IEEE ICASSP*, Apr. 2003, vol. 4, pp. 13-16.
- [11] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.
- [12] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Signal Processing*, vol. 37, no. 5, pp. 720-741, May 1989.
- [13] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181-191, 1994.

- [14] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2201-2214, Aug. 2002.
- [15] B. Hassibi and B. M. Hochwald, "Cayley differential unitary space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1485-1503, June 2002.
- [16] R. Gowaikar and B. Hassibi, "Efficient maximum-likelihood decoding via statistical pruning," submitted to *IEEE Trans. Inform. Theory*, 2003.
- [17] H. Vikalo, B. Hassibi, and T. Kailath, "Iterative decoding for MIMO channels via modified sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2299-2311, Nov. 2004.
- [18] P. D. Alexander, A. J. Grant, and M. C. Reed, "Iterative detection in code-division multiple-access with error control coding," *European Trans. on Tel.*, vol. 9, pp. 419-425, Sep. 1998.
- [19] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a-priori information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673-683, Mar. 2002.



**Haris Vikalo** was born in Tuzla, Bosnia and Herzegovina. He received the B.S. degree from the University of Zagreb, Croatia, in 1995, the M.S. degree from Lehigh University, Bethlehem, PA, in 1997, and the Ph.D. degree from Stanford University, Stanford, CA, in 2003, all in electrical engineering.

He held a short-term appointment at Bell Laboratories, Murray Hill, NJ, in the summer of 1999. From January 2003 to July 2003 he was a Postdoctoral Researcher, and since July 2003 he has been an Associate Scientist at the California Institute of

Technology. His research interests include wireless communications, signal processing, estimation, and genomic signal processing.



**Babak Hassibi** was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. and Ph.D. degrees from Stanford University in 1993 and 1996, respectively, all in electrical engineering.

From October 1996 to October 1998 he was a research associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ.

Since January 2001 he has been with the department of electrical engineering at the California Institute of Technology, Pasadena, CA., where he is currently an associate professor. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linköping University, Sweden. His research interests include wireless communications, robust estimation and control, adaptive signal processing and linear algebra. He is the coauthor of the books *Indefinite Quadratic Estimation and Control: A Unified Approach to  $H^2$  and  $H^\infty$  Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000). He is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council, the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering and the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE). He has been a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY special issue on "space-time transmission, reception, coding and signal processing" and is currently an Associate Editor for Communications of the IEEE TRANSACTIONS ON INFORMATION THEORY.

**Petre Stoica** is Professor of Systems Modeling with the Information Technology Department of Uppsala University in Sweden; more details about him are available at <http://user.it.uu.se/ps/ps.html>.