AN EXPONENTIAL LOWER BOUND ON THE EXPECTED COMPLEXITY OF SPHERE DECODING

*Joakim Jaldén**, *Björn Ottersten*

The Department of Signals, Sensors & Systems
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden

ABSTRACT

The sphere decoding algorithm is an efficient algorithm used to solve the maximum likelihood detection problem in several digital communication systems. The sphere decoding algorithm has previously been claimed to have polynomial expected complexity. While it is true that the algorithm has an expected complexity comparable to that of other polynomial time algorithms for problems of moderate size it is a misconception that the expected number of operations asymptotically grow as a polynomial function of the problem size. In order to illustrate this point we derive an exponential lower bound on the expected complexity of the sphere decoder.

1. INTRODUCTION

The sphere decoding algorithm has received much interest recently since it has been shown to very efficiently solve several otherwise computationally hard optimization problems that arise in digital communications [1, 2, 3].

The efficiency of the sphere decoder has previously been explained by polynomial expected complexity [4, 5]. The general detection problem is NP-hard and it is known that the worst case complexity is exponential but it is also known that the worst case complexity of the sphere decoder does not represent the practical performance of the algorithm. Instead, the average or expected performance is comparable to that of polynomial time algorithms for many problems of practical relevance. However, to attribute the efficiency of the algorithm to polynomial expected complexity is unfortunate since it is not true under the common definition of polynomial complexity [6].

To be precise, let $C(m,\rho)$ be the expected complexity of sphere decoding where m is the number of symbols jointly detected and where ρ is some finite signal to noise ratio (SNR). The sphere decoder is said to be of polynomial expected, or average, complexity for some SNR, ρ, if there exist a polynomial function $p(m)$ such that

$$C(m,\rho) \leq p(m) \quad \forall\, m \geq 1. \tag{1}$$

It is shown herein that for any ρ, no matter how large, there can never exist such a polynomial function. This is done by deriving an exponential lower bound on $C(m,\rho)$. The class of detection problems considered herein include problems previously claimed to be solvable with polynomial expected complexity, se for example [4].

Our main result and contribution is given by Theorem 1.

*jalden@s3.kth.se

2. PROBLEM DEFINITION

Let $\mathbf{s}$ be a vector of independent symbols, $s_0,\ldots,s_{m-1}$, drawn from a finite complex constellation $\mathcal{S} \subset \mathbb{C}$. That is, $\mathbf{s} \in \mathcal{S}^m$, where

$$\mathcal{S}^m = \underbrace{\mathcal{S} \times \ldots \times \mathcal{S}}_{m}.$$

In this paper we consider the detection of such a symbol vector, $\bar{\mathbf{s}}$, sent across a general MIMO channel

$$\mathbf{x} = \mathbf{H}\bar{\mathbf{s}} + \mathbf{v} \tag{2}$$

where $\mathbf{H} \in \mathbb{C}^{n\times m}$, $n \geq m$, is a random channel matrix known to the receiver and where $\mathbf{v} \in \mathbb{C}^n$ is a zero mean complex Gaussian noise vector with a variance of σ^2 per element. The noise, $\mathbf{v}$, is assumed independent of $\mathbf{H}$. Herein $\bar{\mathbf{s}}$ will always denote the symbol vector actually sent across the channel to differentiate between it and an arbitrary symbol vector $\mathbf{s}$. The set $\mathcal{S}$ could for instance be, but is not limited to, a PAM, QAM or PSK constellation. Also, (2) is applicable to a wide variety of systems, e.g. multiple antenna systems, multi carrier systems and certain classes of space time block codes and CDMA systems.

It is well known that the maximum likelihood (ML) estimate of $\bar{\mathbf{s}}$ is given by

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \underset{\mathbf{s}\in\mathcal{S}^m}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2, \tag{3}$$

a problem known to be NP-hard for general $\mathbf{H}$ and $\mathbf{x}$ [7].

Herein, some additional assumptions will be made about the system defined by (2). The symbol vectors $\bar{\mathbf{s}}$ are assumed independent of $\mathbf{H}$ and $\mathbf{v}$ and uniformly distributed on $\mathcal{S}^m$. This implies that the symbols within the symbol vector are statistically independent. The channel, $\mathbf{H}$, and symbol vector, $\bar{\mathbf{s}}$, are assumed to, for some ρ_s, satisfy

$$\frac{\mathrm{E}\left\{\|\mathbf{h}_i\bar{s}_i\|^2\right\}}{\sigma^2} \leq \rho_s \quad i = 1,\ldots,m \tag{4}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. The bound in (4) ensures that the SNR per symbol has an upper bound or specifically, that each symbol is transmitted with finite energy. This is of course satisfied for most systems of practical interest. It is however important to explicitly state when the asymptotic behavior of a system is considered. Note that under the assumptions on $\bar{\mathbf{s}}$ (4) implies

$$\frac{\mathrm{E}\left\{\|\mathbf{H}\bar{\mathbf{s}}\|^2\right\}}{\mathrm{E}\left\{\|\mathbf{v}\|^2\right\}} \triangleq \rho \leq \rho_s. \tag{5}$$

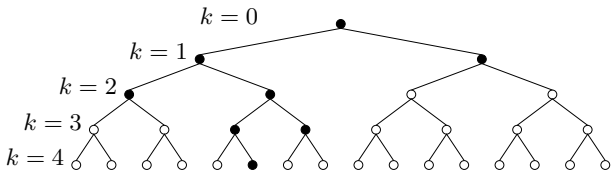

Fig. 1. Search tree for a problem of size $m = 4$ and symbol set of cardinality $|\mathcal{S}| = 2$. Nodes visited by the sphere decoder are shown in black.

In much of the sphere decoding literature $\mathbf{H}$, $\mathbf{s}$, $\mathbf{v}$ and $\mathcal{S}$ are assumed to be real valued and all results herein hold under these assumptions. Since it was recently shown [8] how to adapt the sphere decoder to complex constellations we will however consider this, more general, case.

3. THE SPHERE DECODER

The ML detector (3) can alternatively be written as

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \underset{\mathbf{s}\in\mathcal{S}^m}{\operatorname{argmin}} \|\mathbf{Q}^*\mathbf{x} - \mathbf{R}\mathbf{s}\|^2, \tag{6}$$

where $\mathbf{QR} = \mathbf{H}$ is the QR factorization of $\mathbf{H}$, that is $\mathbf{Q} \in \mathbb{C}^{n\times m}$ is a matrix with orthonormal columns and $\mathbf{R} \in \mathbb{C}^{m\times m}$ is upper triangular. The complexity of solving (6) or (3) by searching over all $\mathbf{s} \in \mathcal{S}^m$ grows as $|\mathcal{S}|^m$ where $|\mathcal{S}|$ is the cardinality, or size, of the set $\mathcal{S}$. Instead, the sphere decoder solves (6) by searching only over those $\mathbf{s} \in \mathcal{S}^m$ that satisfy

$$\|\mathbf{Q}^*\mathbf{x} - \mathbf{R}\mathbf{s}\|^2 = \|\mathbf{p}\|^2 \leq r^2 \tag{7}$$

for some r^2. If r^2 is chosen such that at least one $\mathbf{s} \in \mathcal{S}^m$ satisfies (7) then this strategy will yield the ML solution.

A set of necessary conditions for (7) to be satisfied is that

$$\sum_{i=m-k}^{m-1} |p_i|^2 \leq r^2 \qquad k = 1, \ldots, m \tag{8}$$

where p_i is the ith component of $\mathbf{p} = \mathbf{Q}^*\mathbf{x} - \mathbf{R}\mathbf{s}$. Due to the upper triangular structure of $\mathbf{R}$ the sum of (8), for some k, only depends on s_i for $i \geq m - k$. Thus, if for some combination of $s_{m-k}, \ldots, s_{m-1}$ (8) is violated all symbol vectors, $\mathbf{s}$, sharing this combination of the k last symbols can be excluded from the search.

The sphere decoding algorithm is a systematic way of testing (8). The set of possible symbols $\mathbf{s} \in \mathcal{S}^m$ can be illustrated by a tree where each path from the top to the bottom correspond to a specific symbol vector, $\mathbf{s}$, see Figure 1. The algorithm proceeds down the tree by making decisions about the symbols, s_i, starting with $i = m - 1$ until (8) is violated at some depth k. The algorithm then updates previous choices of s_i and proceeds down the tree again until all possible combinations not violating (8) have been tried. If the bottom of the tree is reached it is known that the symbol vector, $\mathbf{s}$, corresponding to the current path satisfies (7) and $\mathbf{s}$ is saved before the algorithm continues. The complexity of the sphere decoding algorithm is proportional to the number of nodes visited by the algorithm.

Note that the algorithm is sensitive to the choice of r^2. If r^2 is chosen to small no symbol vectors, $\mathbf{s}$, will satisfy (7) which is unacceptable. If r^2 is chosen to large the algorithm will visit to many nodes and be inefficient. Since it is not the purpose of this paper to discuss how to optimally choose r^2 we will simply assume that it is done in a way which ensures that (7) is satisfied for $\mathbf{s} = \bar{\mathbf{s}}$ with high probability. An example of such a strategy is given in [4] and [8]. To be precise, we assume that

$$\mathrm{E}\left\{\|\mathbf{Q}^*\mathbf{x} - \mathbf{R}\bar{\mathbf{s}}\|^2\right\} = \mathrm{E}\left\{\|\mathbf{Q}^*\mathbf{v}\|^2\right\} = \sigma^2 m \leq r^2. \tag{9}$$

The vector $\mathbf{Q}^*\mathbf{v} \in \mathbb{C}^m$ is a Gaussian vector with variance σ^2 per element due to the orthogonality of $\mathbf{Q}$.

4. COMPLEXITY

Let N be the number of nodes visited in the search tree for fixed $\mathbf{H}$ and $\mathbf{x}$. Let the expected complexity, C, be

$$C = \mathrm{E}_{\mathbf{H},\mathbf{x}}\{N\} \tag{10}$$

The object herein is to show that the expected complexity, or expected number of nodes visited, grows exponentially in m. This is formalized in the following theorem.

Theorem 1 *Assume that $\bar{\mathbf{s}}$ is drawn uniformly from the finite set $\mathcal{S}^m$. Assume that $\mathbf{H}$ and $\bar{\mathbf{s}}$ satisfy (4). Then a lower bound on the expected complexity, as defined in (10), is given by*

$$C(m) \geq \frac{|\mathcal{S}|^{\eta m} - 1}{|\mathcal{S}| - 1} \quad , \quad \eta = \frac{1}{4\rho_s + 2}. \tag{11}$$

Proof: The proof is given in Section 4.2.

Note that by Theorem 1 the expected complexity, $C(m)$, of the algorithm is lower bounded by an exponential function in m. This proves that the expected complexity can not be polynomial.

4.1. Expected Complexity

To prove the theorem, it is convenient to first introduce a lemma which formulates the expected complexity, C, in an alternative way. It does so in terms of the search depth along randomly chosen paths through the tree.

Lemma 1 *Given $\mathbf{H}$, $\mathbf{x}$ and some $\mathbf{s} \in \mathcal{S}^m$, let d be the depth of the path corresponding to $\mathbf{s}$. That is,*

$$d = \sup\{k \mid \sum_{i=m-k}^{m-1} |p_i|^2 \leq r^2\} \tag{12}$$

where $\mathbf{p} = \mathbf{Q}^*\mathbf{x} - \mathbf{R}\mathbf{s}$. Then the expected complexity, C, given by (10) can be written as

$$C = \frac{\mathrm{E}_{\mathbf{H},\mathbf{x},\mathbf{s}}\left\{|\mathcal{S}|^{d+1}\right\} - 1}{|\mathcal{S}| - 1} \tag{13}$$

if $\mathbf{s}$ is uniformly distributed over $\mathcal{S}^m$.

Proof: For fixed $\mathbf{H}$ and $\bar{\mathbf{s}}$, view the search depth, d, as a function of $\mathbf{s}$, that is $d = d(\mathbf{s})$. Also, let $I(s_{m-k}, \ldots, s_m)$ be an indicator function equal to 1 if

$$\sum_{i=m-k}^{m-1} |p_i|^2 \leq r^2 \tag{14}$$

and 0 otherwise. In other words, I indicates if a particular node is visited by the algorithm. For notational simplicity, let

$$\sum_{i,j} \cdot \qquad \text{denote} \qquad \sum_{(s_i,\ldots,s_j)\in\mathcal{S}^{j-i+1}} \cdot$$

for $j \geq i$ and interpret

$$\sum_{i,j} \cdot \qquad \text{as} \qquad 1$$

for $j < i$. Then, by using the indicator function and enumerating all nodes from depth 0 to m the number of nodes visited by the algorithm can be written as

$$\begin{aligned}
N &= \sum_{k=0}^{m} \sum_{m-k,m-1} I(s_{m-k},\ldots,s_{m-1}) \\
&= \sum_{k=0}^{m} \sum_{0,m-k-1} |\mathcal{S}|^{-(m-k)} \sum_{m-k,m-1} I(s_{m-k},\ldots,s_{m-1}) \\
&= |\mathcal{S}|^{-m} \sum_{k=0}^{m} \sum_{\mathbf{s}\in\mathcal{S}^m} |\mathcal{S}|^k I(s_{m-k},\ldots,s_{m-1}) \\
&= \sum_{\mathbf{s}\in\mathcal{S}^m} |\mathcal{S}|^{-m} \sum_{k=0}^{m} |\mathcal{S}|^k I(s_{m-k},\ldots,s_{m-1}) \\
&= \sum_{\mathbf{s}\in\mathcal{S}^m} |\mathcal{S}|^{-m} \sum_{k=0}^{d(\mathbf{s})} |\mathcal{S}|^k \\
&= \sum_{\mathbf{s}\in\mathcal{S}^m} |\mathcal{S}|^{-m} \left(|\mathcal{S}|^{d(\mathbf{s})+1} - 1\right)/(|\mathcal{S}| - 1). \qquad (15)
\end{aligned}$$

The last line equals equals

$$N = \mathrm{E}_{\mathbf{s}} \left\{ \frac{|\mathcal{S}|^{d(\mathbf{s})+1} - 1}{|\mathcal{S}| - 1} \right\} \qquad (16)$$

if $\mathbf{s}$ is uniformly distributed on $\mathcal{S}^m$. The expected complexity, C, can thus be written as

$$C = \mathrm{E}_{\mathbf{H},\mathbf{x}} \{N\} = \frac{\mathrm{E}_{\mathbf{H},\mathbf{x},\mathbf{s}} \left\{ |\mathcal{S}|^{d+1} \right\} - 1}{|\mathcal{S}| - 1} \qquad (17)$$

which concludes the proof. ■

4.2. Proof of Theorem 1

By Jensens inequality [9] the expected value in Lemma 1 can be lower bounded by

$$\mathrm{E}\left\{ |\mathcal{S}|^{d+1} \right\} \geq |\mathcal{S}|^{\mathrm{E}\{d\}+1} \qquad (18)$$

since $|\mathcal{S}|^x$ is a convex function of x. To prove Theorem 1 it must be shown that the expected value of d grows linearly with m, that is

$$\mathrm{E}\{d\} \geq \eta m - 1.$$

For notational purposes, let $\mathbf{\Phi}_k$ be a diagonal matrix with the k last diagonal elements equal to 1 and the remaining elements equal to 0. The sum of (8) can then be written as

$$\sum_{i=m-k}^{m-1} |p_i|^2 = \|\mathbf{\Phi}_k \mathbf{p}\|^2 \qquad (19)$$

and the probability that d is strictly smaller than some k is given by

$$\Pr\{d < k\} = \Pr\left\{ ||\mathbf{\Phi}_k \mathbf{p}||^2 > r^2 \right\}. \qquad (20)$$

The Markov inequality [9] upper bounds this probability as

$$\Pr\left\{ ||\mathbf{\Phi}_k \mathbf{p}||^2 > r^2 \right\} \leq \frac{\mathrm{E}\left\{ ||\mathbf{\Phi}_k \mathbf{p}||^2 \right\}}{r^2}. \qquad (21)$$

The probability that the depth d is at least as large as k can thus be lower bounded by

$$\Pr\{d \geq k\} \geq 1 - \frac{\mathrm{E}\left\{ ||\mathbf{\Phi}_k \mathbf{p}||^2 \right\}}{r^2}. \qquad (22)$$

Since $\mathbf{R}$, $\bar{\mathbf{s}}$, $\mathbf{s}$, and $\mathbf{v}$ are assumed independent, the expected value can be computed as

$$\begin{aligned}
\mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{p}\|^2 \right\} &= \mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{R}(\bar{\mathbf{s}} - \mathbf{s}) + \mathbf{\Phi}_k \mathbf{Q}^* \mathbf{v}\|^2 \right\} \\
&= \sum_{i=m-k}^{m-1} \mathrm{E}\left\{ |\bar{s}_i - s_i|^2 \right\} \mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{r}_i\|^2 \right\} + k\sigma^2. \qquad (23)
\end{aligned}$$

Since both $\bar{s}_i$ and s_i are assumed independent and uniformly distributed on $\mathcal{S}$,

$$\mathrm{E}\left\{ |\bar{s}_i - s_i|^2 \right\} \leq 2\mathrm{E}\left\{ |\bar{s}_i|^2 \right\}. \qquad (24)$$

The above holds with equality if $\mathrm{E}\{\bar{s}_i\} = 0$ which is commonly the case. Since $\mathbf{r}_i = \mathbf{Q}^* \mathbf{h}_i$

$$\mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{r}_i\|^2 \right\} \leq \mathrm{E}\left\{ \|\mathbf{h}_i\|^2 \right\}. \qquad (25)$$

Using

$$\mathrm{E}\left\{ \|\mathbf{h}_i \bar{s}_i\|^2 \right\} = \mathrm{E}\left\{ \|\mathbf{h}_i\|^2 \right\} \mathrm{E}\left\{ |\bar{s}_i|^2 \right\}. \qquad (26)$$

together with (24) and (25) yields

$$\mathrm{E}\left\{ |\bar{s}_i - s_i|^2 \right\} \mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{r}_i\|^2 \right\} \leq 2\rho_s \sigma^2. \qquad (27)$$

The expression of (23) can thus be bounded by

$$\mathrm{E}\left\{ \|\mathbf{\Phi}_k \mathbf{p}\|^2 \right\} \leq k\sigma^2(2\rho_s + 1). \qquad (28)$$

Using the additional assumption about r^2 in (9), that is

$$r^2 \geq \sigma^2 m,$$

and (22) the probability that d is greater than or equal to k can be bounded as

$$\Pr\{d \geq k\} \geq 1 - \frac{k(2\rho_s + 1)}{m}. \qquad (29)$$

Let a be given by

$$a = \left\lfloor \frac{m}{2\rho_s + 1} \right\rfloor, \qquad (30)$$

then

$$\frac{m}{2\rho_s + 1} \geq a \geq \frac{m}{2\rho_s + 1} - 1 \qquad (31)$$

and

$$\Pr\{d \geq k\} \geq 1 - \frac{k}{a} = \Pr\{\nu \geq k\} \qquad k = 0,\ldots,a \qquad (32)$$

for an integer valued random variable ν with probability mass function

$$\Pr\{\nu = k\} = \frac{1}{a} \qquad \text{for} \qquad k = 0,\ldots,a-1. \qquad (33)$$

Due to (32), the expected value of d can be bounded by

$$\begin{aligned} \mathrm{E}\{d\} &\geq \mathrm{E}\{\nu\} = \frac{1}{2}(a-1) \\ &\geq \frac{1}{2}\left(\frac{m}{2\rho_s+1} - 2\right) = \frac{m}{4\rho_s+2} - 1. \end{aligned} \tag{34}$$

Letting

$$\eta = \frac{1}{4\rho_s+2}$$

and using (18), (34) with Lemma 1 concludes the proof. ■

5. COMMENTS

The result derived herein is of a theoretical nature and is in general not a good approximation of the complexity for most detection problems. A more informative measure of the expected complexity would be the rate at which the expected complexity tends to infinity. Such a result would for instance be

$$C(m,\rho) \doteq |\mathcal{S}|^{\gamma m} \tag{35}$$

where the symbol $\doteq$ means "equal to the first order in the exponent", that is

$$\gamma = \lim_{m\to\infty} \frac{1}{m} \log_{|\mathcal{S}|} C(m,\rho). \tag{36}$$

From the result given by Theorem 1 it is known that $\gamma \geq \eta$. It is however unlikely that this inequality is tight for any relevant system. How to compute γ is not within the scope of this paper but the topic of [10]. However, to illustrate this point, the relation of η, for $\rho_s = \rho$, and γ are shown in Figure 2 for the case where $\mathcal{S} = \{+1, 1\}$ and where both $\mathbf{H} \in \mathbb{R}^{m\times m}$ and $\mathbf{v} \in \mathbb{R}^m$ consist of i.i.d. real valued normally distributed elements.

There are several known improvements to the sphere decoding algorithm that are not included in this work. These include adaptively updating the radius r^2 and permuting the symbol vector $\mathbf{s}$ in order to improve the efficiency of the algorithm [11]. Although not proved by the result herein we believe that this does not results in polynomial time versions of the sphere decoder. These improvements may however further decrease the rate γ.

6. CONCLUSIONS

We show herein that the sphere decoder is of exponential expected complexity contrary to previous claims. This is done this by deriving an exponential lower bound on the expected complexity which hold under very general assumptions about the communication system.

It is important to realize what is shown herein and what is not. The sphere decoder could for some SNR be more efficient that a polynomial time algorithm for all problem sizes of practical interest without contradicting the fact that it is of exponential expected complexity. What is stated by the complexity result is that there will always be some problem size where the polynomial time algorithm is more efficient than the sphere decoder. However, if problems of that size are to be considered interesting is of course dependent on the specific application. It is nevertheless important to realize that there is a fundamental difference in terms of complexity between the sphere decoder and polynomial complexity alternatives.

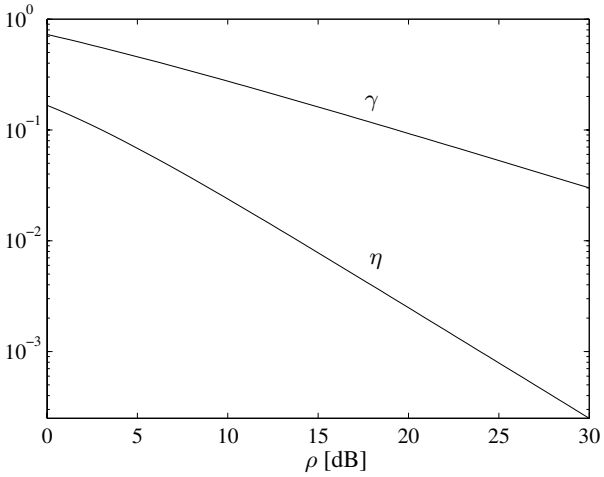

Fig. 2. Comparison of γ and η as a function of ρ

7. REFERENCES

[1] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channel," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.

[2] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, June 2002.

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

[4] B. Hassibi and H. Vikalo, "On the expected complexity of integer least-squares problems," in *Proc. IEEE ICASSP'02*, May 2002, vol. 2, pp. 1497–1500.

[5] O. Damen, K. Abed-Meraim, and M. S. Lemdani, "Further results on the sphere decoder," in *Proc. ISIT'01*, June 2001, p. 333.

[6] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*, Wilie-Interscience, 1988.

[7] S. Verdú, "Computational complexity of multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

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

[9] S. Ross, *A First Course in Probability*, Macmillan Publishing Company, second edition, 1984.

[10] J. Jaldén and B. Ottersten, "On the expected complexity of sphere decoding in digital communications," *Submitted to IEEE Trans. Sig. Proc.*, 2003.

[11] A. M. Chan and I. Lee, "A new reduced complexity sphere decoder for multiple antenna systems," in *Proc. ICC'02*, Apr. 2002.