COMPRESSED SENSING K-BEST DETECTION FOR SPARSE MULTI-USER COMMUNICATIONS

B. Knoop, F. Monsees†, C. Bockelmann†, D. Peters-Drolshagen*, St. Paul*, A. Dekorsy†*

* Institute of Electrodynamics and Microelectronics (ITEM), University of Bremen, Germany
Email: {knoop, peters, paul}@me.uni-bremen.de

† Department of Communications Engineering, University of Bremen, Germany
Email: {monsees, bockelmann, dekorsy}@ant.uni-bremen.de

ABSTRACT

Machine-type communications are quite often of very low data rate and of sporadic nature and therefore not well-suited for nowadays high data rate cellular communication systems. Since signaling overhead must be reasonable in relation to message size, research towards joint activity and data estimation was initiated. When the detection of sporadic multi-user signals is modeled as a sparse vector recovery problem, signaling concerning node activity can be avoided as it was demonstrated in previous works.

In this paper we show how well-known K-Best detection can be modified to approximately solve this finite alphabet Compressed Sensing problem. We also demonstrate that this approach is robust against parameter variations and even works in cases where fewer measurements than unknown sources are available.

Index Terms— K-Best algorithm, multi-user detection, sparse signal processing, Compressed Sensing

1. INTRODUCTION

Machine-type communications (MTC) is one of the big emerging fields for future communication systems. Nowadays high data rate systems such as LTE were designed for human-based traffic without having MTC in mind [1]. Unlike human-driven communications, MTC refers to traffic between two autonomous entities without human interaction. It is quite often of very low data rate and of sporadic nature. As the ratio of signaling to payload renders communication inefficient when data packets are very small, signaling overhead should be kept at a minimum level. Due to these properties previous approaches for integrating MTC traffic into existing communication systems have been unsuccessful, which calls for novel techniques.

In an MTC uplink scenario where sensor nodes sporadically transmit data to a central aggregation node, signaling concerning node activity can be avoided by facilitating joint activity and data detection. This detection problem can be cast as estimating a sparse multi-user signal, where inactive nodes are modeled as transmitting zeros while active nodes transmit symbols taken from a finite modulation alphabet [2].

Recent papers have approached this type of detection problem by utilizing CS algorithms. For example, in [3, 4] Orthogonal Matching Pursuit or Group OMP was used for joint data and activity detection. However, the discrete nature of the finite set of modulation symbols allows for the use of non-linear tree search algorithms, first and foremost Sphere Decoding (SD). In terms of symbol error rate (SER), SD achieves the maximum a posteriori (MAP) performance [2,5], albeit the algorithmic complexity can only be upper bounded to be exponential with the problem size [6]. As a much simpler alternative, Successive Interference Cancellation (SIC) was investigated in [7]. The paper shows that the SER performance reduces drastically whenever the system load changes due to higher user activity, which increases the interference and reduces the sparsity in the system. CS algorithms and generalized SD even work for underdetermined systems, where SIC fails to detect the sparse multi-user source signal.

K-Best detection can be interpreted as a trade-off between SD and SIC with an adjustable parameter K which determines performance and complexity of the detector [8]. The main advantage is that K-Best detection allows for a constant throughput which is not the case for SD.

The goal of this paper is to incorporate the expertise of previous works in this field and close the gap between theoretical analysis and practical implementation towards a robust and flexible algorithm for joint activity and data detection. We show that K-Best detection works with a reasonably low K even for underdetermined systems: It can achieve good symbol error rate performance with a complexity far less than MAP detection and is robust against unpredictable changes of user activity in the system.

This work was funded by the German Research Foundation (DFG) under grants PA 438/8-1 and DE 759/3-1.

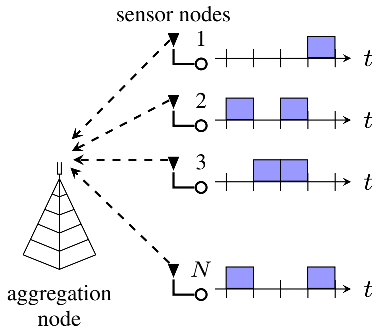

Fig. 1. Machine-to-machine uplink scenario involving N sporadically active sensor nodes.

2. SYSTEM MODEL

Fig. 1 shows the MTC multi-user uplink system assumed in this work. A set of N nodes transmits data to a central aggregation node for further processing, each node being only sporadically active. In other words, just a random subset of nodes transmits data at a specific time instance, as indicated. We assume that the detector has probabilistic but not instantaneous knowledge about the activity pattern.

The detection model can be summarized by the canonical I/O relationship in symbol clock

$$\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{w}\,, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the vector of M observations at some time instance and $\mathbf{T} \in \mathbb{R}^{M\times N}$ summarizes the medium access and channels. In this work we assume randomly coded data symbols which means that the transmit nodes precode data symbols via zero-mean Gaussian sequences. This method is very similar to CDMA and was proposed in [9] as medium access for CS-based detection. Gaussian sequences facilitate a better user separation compared to traditional pseudo-noise sequences which also coincides with the notion of random CS measurement matrices. We assume the superposition of additive white Gaussian noise with zero mean and variance σ_n^2, symbolized by $\mathbf{w}$. The transmission is of low data rate, occupying only a narrow bandwidth, which justifies the assumption of a frequency-flat fading channel.

The multi-user vector $\mathbf{x}$ contains the modulation symbols from the transmit nodes and the nth entry corresponds to the modulation symbol of node n. We employ a homogeneous activity model parametrized by the activity probability p_a, i.e. each node transmits data with probability p_a. Active nodes transmit symbols of the modulation alphabet $\mathcal{A}$ and the other nodes can be modeled as “transmitting” a zero symbol, since they are inactive (cp. [2, 10]). The application of this model makes $\Pr(x_n = 0) = 1 - p_a$ and $\Pr(x_n \in \mathcal{A}) = p_a$. If p_a is sufficiently small, the multi-user vector $\mathbf{x}$ is a sparse vector containing a considerable number of zero symbols.

2.1. Sparsity-MAP Detector

The main idea is to detect the multi-user signal $\mathbf{x} \in \mathcal{A}_0^N$ with respect to the zero-augmented symbol alphabet $\mathcal{A}_0 = \mathcal{A} \cup \{0\}$, which can be interpreted as joint activity and data estimation. The sparsity-aware MAP (S-MAP) optimization problem associated with (1) is [2]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{A}_0^N} \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0\,, \tag{2}$$

where $\|\cdot\|_0 = \#\{i : x_i \neq 0\}$ is the ℓ_0-(pseudo) norm, i.e. the total number of non-zero elements or the cardinality of the support. The penalty term

$$\lambda = 2\sigma_n^2 \log\left(\frac{1-p_a}{p_a/|\mathcal{A}|}\right) \tag{3}$$

reflects the a priori statistics of the data vector and scales with the noise power σ_n^2, as derived in [11]. Subsequently, λ is used as a system-specific regularization parameter.

In the following we restrict ourselves to constant modulus data, i.e. $|x_n| = 1$, $\forall\, x_n \in \mathcal{A}$ (cp. [10]). Hence, any phase-shift keying modulation scheme may be chosen for $\mathcal{A}$. Further, we assume, without loss of generality, $\lambda > 0$ to simplify evaluation. It can be derived from (3) that this second restriction is fulfilled as long as $p_a \leq |\mathcal{A}|/(|\mathcal{A}|+1)$ holds.

2.2. Underdetermined Systems

In the following we focus on scenarios where the number of nodes N is greater than the spreading sequence length M, i.e. we assume (1) to be underdetermined. With $M < N$, the system is overloaded in the classical sense for better resource efficiency. This relates the detection problem (2) to the theory of Compressed Sensing, since the multi-user source signal $\mathbf{x}$ is of sparse nature. In fact, (2) is the finite alphabet variant of the unconstrained Lagrangian form of the convex quadratic problem known as Basis Pursuit Denoising (BPDN),

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2 < \varepsilon\,, \tag{4}$$

where $\mathbf{T}$ may not be arbitrary, but must fulfill e.g. the restricted isometry property (RIP) for some reconstruction guarantees [12]. Note that the non-convex ℓ_0-norm is relaxed to the convex ℓ_1-norm.

The available prior knowledge about the source vector $\mathbf{x}$ can be utilized to regularize the underdetermined system of equations in (2). For instance, the constant modulus assumption allows us to substitute $\|\mathbf{x}\|_0 = \|\mathbf{x}\|_1 = \|\mathbf{x}\|_2^2$, [10]. Now we can include the regularization term into the least squares metric and reformulate the minimization problem as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{A}_0^N} \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2 + \|\sqrt{\lambda}\cdot\mathbf{x}\|_2^2 \tag{5}$$

$$= \arg\min_{\mathbf{x}\in\mathcal{A}_0^N} \left\| \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_N \end{bmatrix} - \begin{bmatrix} \mathbf{T} \\ \sqrt{\lambda}\cdot\mathbf{I}_N \end{bmatrix} \mathbf{x} \right\|_2^2 \tag{6}$$

$$= \arg\min_{\mathbf{x}\in\mathcal{A}_0^N} \|\tilde{\mathbf{y}} - \tilde{\mathbf{T}}\mathbf{x}\|_2^2\,. \tag{7}$$

function K-BEST($\hat{\mathbf{y}}, \mathbf{R}, \mathcal{A}_0, K$)
    for $n \leftarrow N, ..., 1$ **do**
        for $k \leftarrow 1, ..., K$ **do** $\triangleright$ K paths
            Retrieve history of k^{th} path $\left(\hat{x}_{n+1}, ..., \hat{x}_N\right)_k$
            for all $\hat{x}_n \in \mathcal{A}_0$ **do**
                $d_\ell \leftarrow \sum_{i=n}^{N} \left| \hat{y}_i - \sum_{j=i}^{N} r_{ij} \hat{x}_j \right|^2$
            end for
        end for
        Sort all $K \cdot |\mathcal{A}_0|$ hypotheses by $\mathbf{d}$
        Select K paths with the least metrics
        for $k' \leftarrow 1, ..., K$ **do** $\triangleright$ K new paths
            Store updated $\left(\hat{x}_n, ..., \hat{x}_N\right)_{k'}$ of k'^{th} path
        end for
    end for
    return $\hat{\mathbf{x}}$
end function

Fig. 2. Simplified structure of the K-Best detection algorithm

Here, $\tilde{\mathbf{T}}$ is the regularized $(M+N) \times N$ matrix containing the prior knowledge about the source data λ as well as the channel $\mathbf{T}$. If $\lambda > 0$ holds, $\tilde{\mathbf{T}}$ has full rank. However, solving (7) may lead to numeric difficulties when λ approaches zero.

3. SPARSE K-BEST DETECTION

It has been shown in [13] that the reformulated minimization problem (7) can be solved by tree search algorithms. The system matrix $\tilde{\mathbf{T}}$ has to be triangularized using some type of QR decomposition. In this setup we use sorted QR decomposition (SQRD), which permutes the detection order such that data of nodes with a high post-detection SNR are estimated first [14, 15]. The essential idea of SQRD is to decompose the system matrix such that $\tilde{\mathbf{T}} = \mathbf{QR\Theta}$, with $\mathbf{\Theta}$ being a permutation matrix, which determines the ordering of the subsequent detection. With SQRD, (7) becomes

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{A}_0^N} \|\hat{\mathbf{y}} - \mathbf{R\Theta x}\|_2^2 \, , \tag{8}$$

where $\hat{\mathbf{y}} = \mathbf{Q}^T \tilde{\mathbf{y}}$. $\mathbf{R}$ is an upper-triangular matrix of size $N \times N$ and has full rank as long as $\lambda > 0$ holds.

A Sphere Decoder solves (8) by a depth-first tree search and is guaranteed to find the optimal solution, but there is no guarantee that it terminates within polynomial time [6]. And another major disadvantage, besides its data-dependent runtime, is the impossibility to parallelize computations efficiently [16].

K-Best detection was proposed to improve on these aspects [8]. The algorithm sacrifices optimality but for large values of the parameter K it is able to nearly achieve optimal performance. K-Best performs a breadth-first search, and only retains those K hypotheses or search paths with the smallest associated partial metric per iteration. Thus, the algorithm operates on the search tree *unidirectionally*, which leads to a constant runtime and makes parallelization and pipelining of computations possible. This is favorable in regard to a future implementation in hardware.

Table 1. Simulation Parameters

Number of Nodes/Users	$N = 20$
Spreading Gain	M (variable)
Channel Type	AWGN
Modulation Type	BPSK
Activity Probability	$p_a = 0.2$
Detection Model	Zero-augmented alphabet

3.1. Complexity Analysis

Figure 2 outlines the structure of the K-Best algorithm. It descends on the search tree iteratively, starting from N (the root level of the tree) down to 1 (the level with leafs). Within this loop, computations can basically be grouped into two parts. First, the partial accumulated metrics of all child nodes have to be evaluated for each of the K nodes, which in fact represent whole paths from the root of the tree down to the current level. The partial metric d depends on the previous nodes of that path (detected symbols $\hat{x}_i$, with $i = n+1, ..., N$) and the child node itself (x_n). All partial metrics of the $K \cdot |\mathcal{A}_0|$ considered paths ("hypotheses") are stored in a vector $\mathbf{d}$ of length $K \cdot |\mathcal{A}_0|$ (index ℓ). Second, the path hypotheses are sorted and only those K paths with the least associated partial metric are selected and stored for the next iteration.

The computational complexity of the K-Best algorithm is therefore

$$C_{\text{K-Best}} = \mathcal{O}\{N(K|\mathcal{A}_0|C_{\text{pm}} + C_{\text{sort}})\} \, , \tag{9}$$

with C_{pm} modeling the necessary operations for the partial metric computations and C_{sort} the complexity of the sorting algorithm. In [7], the authors showed that $C_{\text{pm}} = \mathcal{O}(N^2)$ and, if Odd-Even Mergesort is employed, $C_{\text{sort}} = \mathcal{O}(K|\mathcal{A}_0|\log^2(K|\mathcal{A}_0|))$. Consequently, K-Best detection is of polynomial complexity in the number of users or transmit sensor nodes, and scales merely loglinearly with the number of search paths K. An improvement of the algorithm, called K^+-Best, further reduces the sorting complexity [17]. Since only the smallest K partial path metrics are of interest for subsequent processing, a completely sorted list of all metrics is not required. Hence, the sorting can algorithmically be simplified.

4. SIMULATION RESULTS

We investigate the performance of sparse K-Best detection on an exemplary random coding uplink system [18]. Because the columns of $\mathbf{T}$ contain random Gaussian sequences of length

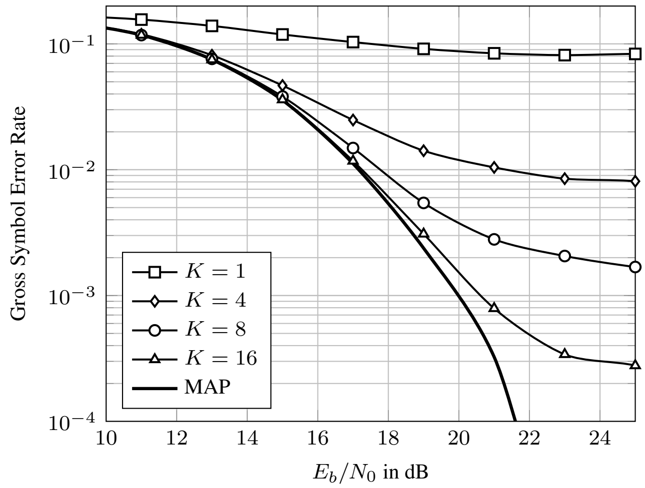

Fig. 3. Gross SER of zero-augmented BPSK and sparse K-Best detection for an underdetermined scenario.

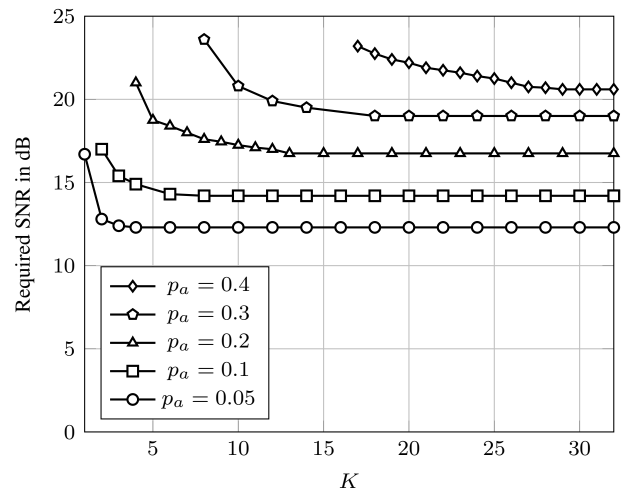

Fig. 4. Required SNR for target Gross SER of 10^{-2} with varying K.

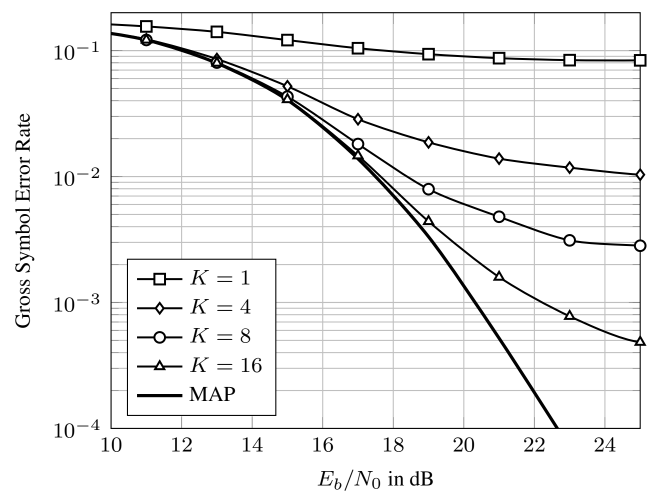

Fig. 5. Performance of K-Best detection with a mismatch between true prior and postulated prior.

M for each of the N nodes, the dimension of $\mathbf{T}$ is $M \times N$. M also signifies the number of measurements. The remaining simulation parameters are summarized in Table 1.

The first simulation setup uses a spreading length of $M = 10$. This system is clearly underdetermined as the ratio of observations to nodes is $1/2$. Fig. 3 shows the gross symbol error rate (GSER) over the zero-augmented BPSK symbol alphabet $\mathcal{A}_0 = \{0, -1, 1\}$. The S-MAP performance is used as a benchmark. It can be seen that K-Best detection with very low K results in huge performance losses. Exemplary, for $K = 1$ (SIC detection) reliable detection is infeasible. Increasing K to higher values reduces the performance loss significantly and with $K = 16$ the gap between S-MAP performance and K-Best is small in the mid SNR range. At higher SNR values, K-Best detection shows an error floor which is due to the limited coverage of the search within the search tree. In contrast to this, the S-MAP detector can achieve better performance but only with non-deterministic throughput and latency.

Setting $K = 16$ results in a computational complexity of $N \cdot K \cdot |\mathcal{A}_0| = 960$ processed hypotheses, which is still feasible from a practical perspective. In comparison, the computational complexity of S-MAP detection is not fixed and can only be upper bounded to $|\mathcal{A}_0|^N \approx 3.5 \cdot 10^9$ which is prohibitively high. Moreover, the number of processed nodes is known beforehand with K-Best detection, hence it allows for the design of parallel hardware structures with *constant* throughput and latency at the physical layer.

In the next setup we consider the robustness of K-Best detection for different user activities with an exemplary target gross symbol error rate of GSER $\leq 10^{-2}$. Fig. 4 plots the required SNR as a function of the parameter K. At low activity probabilities, the required SNR is constant for a wide range of K and an increase in K does not automatically lead to performance gains. For higher user activities, the GSER requirement can only be fulfilled with larger values of K, e.g. at $p_a = 0.4$ a $K \geq 16$ is necessary. This is due to the error floor behavior as discussed above.

Finally, we look at the robustness of K-Best detection in scenarios where user activity is to some extend random, yielding a mismatch between the actual activity rate and the activity probability assumed by the detector. In practice, the user activity has to be estimated by the detector or the detector has to be pre-adjusted according to some long term user activity statistics. We model this mismatch Δ_p as uniformly random $\Delta_p \sim \mathcal{U}\left(-\Delta_{p,\max}, \Delta_{p,\max}\right)$, i.e. the node activity probability in the system is $p = p_a + \Delta_p$, where p_a is the activity probability assumed at the detector and p is the true activity probability of the transmit nodes. Fig. 5 shows the performance of K-Best detection averaged over $10\,000$ model realizations when the mismatch is up to $50\,\%$. Then, the activity probability assumed at the detector is $p_a = 0.2$ and $\Delta_{p,\max} = p_a/2 = 0.1$. The results are similar to the results of the first scenario without activity mismatch (Fig. 3) but a performance loss is observable. However, we see that with $K = 8$ we are still able to achieve a GSER $< 10^{-2}$. Therefore we can conclude that K-Best detection is to some extent robust against parameter mismatch.

5. CONCLUSION

This work applies the K-Best algorithm to the joint activity and data detection problem of sparse machine-type communications. The main advantage of this algorithm compared to Sphere Decoding is its polynomial complexity and fixed latency which is better suited for practical implementation. We successfully showed that K-Best detection nearly achieves maximum a posteriori performance with a reasonably small search parameter K. At the cost of a slightly increased K, it even allows for reliable detection in a Compressed Sensing problem, i.e. in underdetermined multi-user systems where the number of nodes in the system is higher than the available resources. The detector is robust against activity parameter mismatch and one choice of K may serve a wide SNR and p_a region.

REFERENCES

[1] S. Lien, K. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *Communications Magazine, IEEE*, vol. 49, no. 4, pp. 66 –74, Apr. 2011.

[2] H. Zhu and G.B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454 – 465, Feb. 2011.

[3] H. F. Schepker and A. Dekorsy, "Sparse Multi-User Detection for CDMA transmission using greedy algorithms," in *8th International Symposium on Wireless Communication Systems (ISWCS)*, Aachen, Germany, Nov. 2011, pp. 291 –295.

[4] A.K. Fletcher, S. Rangan, and V.K. Goyal, "A sparsity detection framework for on-off random access channels," in *IEEE International Symposium on Information Theory, 2009. ISIT 2009*, June 2009, pp. 169–173.

[5] F. Monsees, C. Bockelmann, D. Wübben, and A. Dekorsy, "Compressed sensing bayes risk minimization for under-determined systems via sphere detection," in *IEEE 77th Vehicular Technology Conference (VTC2013-Spring)*, Dresden, Germany, Jun 2013.

[6] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *Signal Processing, IEEE Transactions on*, vol. 53, no. 4, pp. 1474–1484, 2005.

[7] B. Knoop, F. Monsees, C. Bockelmann, D. Wübben, S. Paul, and A. Dekorsy, "Sparsity-aware successive interference cancellation with practical constraints," *ITG Fachbericht Workshop on Smart Antennas 2013*, 2013.

[8] K. Wong, C. Tsui, R.S.-K. Cheng, and W. Mow, "A VLSI architecture of a k-best lattice decoding algorithm for MIMO channels," in *IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002*, 2002, vol. 3, pp. III–273–III–276 vol.3.

[9] M.S. Asif, W. Mantzel, and J. Romberg, "Channel protection: Random coding meets sparse channels," in *IEEE Information Theory Workshop, 2009. ITW 2009*, oct 2009, pp. 348 –352.

[10] Zhi Tian, G. Leus, and V. Lottici, "Detection of sparse signals under finite-alphabet constraints," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, Apr. 2009, pp. 2349 –2352.

[11] B. Knoop, T. Wiegand, and S. Paul, "Low-Complexity and Approximative Sphere Decoding of Sparse Signals," in *Asolimar Conference on Signals, Systems and Computers*, Pacific Grove USA, Nov 2012.

[12] Y. C. Eldar and G. Kutyniok, *Compressed Sensing Theory and Applications*, Cambridge University Press, Cambridge, 2012.

[13] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389 – 2402, oct. 2003.

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

[15] Qingwei Li and Zhongfeng Wang, "Improved k-best sphere decoding algorithms for MIMO systems," in *2006 IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings*, May 2006, pp. 4 pp.–1162.

[16] A. Burg, M. Borgmann, M. Wenk, and M. Zellwegerand W. Fichtnerand H. Bölcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 7, pp. 1566–1577, July 2005.

[17] N. Heidmann, T. Wiegand, and S. Paul, "Architecture and FPGA-implementation of a high throughput K^+-Best detector," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2011*, 2011, pp. 1–6.

[18] S. Verdú, *Multiuser Detection*, Cambridge Univ. Press, Cambridge, U.K., Nov. 1998.