Enabling Sphere Decoding for SCMA

Monirosharieh Vameghestahbanati, Ebrahim Bedeer, Member, IEEE, Ian Marsland, Member, IEEE, Ramy H. Gohary, Senior Member, IEEE, Halim Yanikomeroglu, Fellow, IEEE

Abstract—In this paper, we propose a reduced-complexity optimal modified sphere decoding (MSD) detection scheme for SCMA. As SCMA systems are characterized by a number of resource elements (REs) that are less than the number of the supported users, the channel matrix is rank-deficient, and sphere decoding (SD) cannot be directly applied. Inspired by the Tikhonov regularization, we formulate a new full-rank detection problem that is equivalent to the original rank-deficient detection problem for constellation points with constant modulus and an important subset of non-constant modulus constellations. By exploiting the SCMA structure, the computational complexity of MSD is reduced compared with the conventional SD. We also employ list MSD to facilitate channel coding. Simulation results demonstrate that in uncoded SCMA systems the proposed MSD achieves the performance of the optimal maximum likelihood (ML) detection. Additionally, the proposed MSD benefits from a lower average complexity compared with MPA.

Index Terms—Sparse code multiple access (SCMA), modified list sphere decoding (MSD), maximum likelihood (ML).

I. INTRODUCTION

The need to accommodate diverse types of users and applications necessitates more efficient ways to use the spectrum in 5G systems. Sparse code multiple access (SCMA) is a non-orthogonal multidimensional codebook-based configuration that can cope with the requirements of 5G systems. In essence, SCMA is a generalization of low density signature (LDS) signaling, whereby a sparse signature matrix is used to reduce the complexity of the detector at the receiver.

One of the main challenges in the design of SCMA systems is to overcome the complexity of the receiver that decodes the data generated from all active users. Inspired by the sparsity of the SCMA codewords, [1] uses a near-optimal message passing algorithm (MPA) to detect the SCMA symbols. An MPA-based algorithm was proposed in [2] that reduces the detection complexity by assigning larger weight factors to those codewords with larger probabilities. However, this improvement results in a degradation in the block error rate at high signal-to-noise ratio (SNR) regimes. In [3], a hybrid of list sphere decoding (SD) and MPA is used to reduce the complexity of SCMA detection. However, the SD detection problem is defined per resource elements (RE) and not on the entire block, and thus does not result in an optimal solution.

In this paper, we investigate the detection problem of SCMA systems, and propose a reduced-complexity optimal modified SD (MSD) detection scheme. Due to the non-orthogonal nature of SCMA systems, the number of REs is less than the number of active users. As such, the channel matrix is rank-deficient, and SD cannot be directly applied. To tackle this issue, we use Tikhonov regularization [4] to formulate a new full-rank detection problem that is equivalent to the original rank-deficient detection problem for constellation points with constant modulus, and an important subset of non-constant modulus constellations. The complexity of the proposed MSD scheme is reduced by exploiting the sparsity of SCMA codebooks, and the fact that each user spreads the same information bits over a few REs alleviates the need for expanding all tree branches. We also use list MSD to provide soft-outputs to be used with channel coding. Simulation results confirm that in uncoded scenarios the proposed MSD scheme achieves the performance of the optimal maximum likelihood (ML) detection. Furthermore, the proposed MSD benefits from a lower average complexity compared with MPA.

II. SYSTEM MODEL

Consider an uplink SCMA system with $K$ users and $N$ orthogonal resource elements (REs), where $N < K$, and each user is connected to $d_u \ll N$ REs only. In an $M$-ary signal constellation, each $L_M = \log_2 M$ bits is mapped to a $d_u$-dimensional complex constellation symbol, $x_k = (x_{1,k}, \ldots, x_{d_u,k})^T$ that is selected from a $d_u$-dimensional complex codebook $X_k$ of size $M$, and defined within a constellation set, $X_k \subseteq \mathbb{C}^{d_u}$. The $N \times d_u$ binary mapping matrix of user $k$ is denoted by $S_k$, where $s_{n,l} = 1$, $n \in \{1, \ldots, N\}$ and $l \in \{1, \ldots, d_u\}$, if and only if the $l$th symbol of user $k$ occupies resource $n$. We assume each user consists of $d_v$ layers that is connected to one RE only. The total number of layers is $K' = d_v K$, and the mapping matrix of the SCMA code, $S = [S_1 \ldots S_K]$, is then an $N \times K'$ matrix with only one non-zero element in each of its columns. The set of layers occupying resource $n$ is specified by the position of $1$s in the $n$th row of $S$, and is represented by $\mathcal{F}_n = \{k' | s_{n,k'} = 1\}$, $k' \in \{1, \ldots, K'\}$, with cardinality $d_f = |\mathcal{F}_n|$. As we will

1The indicator matrix $P$ is an $N \times K$ matrix with each of its columns defined as $p_k = \text{diag} (S_k S_k^T)$. Note that [5] uses $P$ to apply SD on each RE independently, which does achieve the ML performance. Also, using $P$ to apply SD on all REs jointly is very challenging and may not be possible. In contrast, $S$ allows us to apply SD on all REs jointly, and achieves the ML performance. The increase in dimensionality from $K$ to $K'$ due to using $S$ instead of $P$ will be compensated in the proposed algorithm.
discuss later in Section III-A and V, to simplify the detector, it
is very advantageous for $S$ to be an upper-triangular matrix, so
the first $N$ columns of $S$ constitutes an identity matrix. This
can be achieved in scenarios with static resource allocation
(RA) by assigning the first $d_v$ REs to the first user, the second $d_v$
REs to the second user, and so on. In scenarios with dynamic RA, it is easy to make $S$ an upper-triangular
matrix during the RA phase by relabelling the REs and users,
provided that there exists $N/d_v$ orthogonal users.

In an uplink transmission scenario over Rayleigh frequency
flat fading contaminated by additive white Gaussian noise
(AWGN), the $N \times 1$ received signal vector is represented by

$$ \mathbf{y} = \sum_{k=1}^{K} S_k \mathbf{H}_k \mathbf{x}_k + \mathbf{w}, $$

where $\mathbf{H}_k = \text{diag}(h_{1,k}, \ldots, h_{d_v,k})$ is a $d_v \times d_v$
diagonal matrix containing the complex channel gains for the $d_v$
REs used by user $k$, $\mathbf{x} = [x_1^T, \ldots, x_K^T]^T = (x_1, \ldots, x_K)^T$ is a $K'$-dimensional vector containing all transmitted symbols of
all users, and $\mathbf{w} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ is the $\mathcal{N}$-dimensional
complex Gaussian ambient noise. Moreover, $\mathbf{G} = (g_1, \ldots, g_K)$,
$g_k = S_k \mathbf{H}_k$, is the $N \times K'$ effective channel gain matrix.

After the reception of $\mathbf{y}$, a multiuser detection technique is
employed to recover each user’s codeword $\mathbf{x}_k$. The optimal
ML detection for SCMA transmitted codewords is given by

$$ \hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{X}} \| \mathbf{y} - \mathbf{G} \mathbf{x} \|^2, $$

where $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_{K'})^T$, denotes the detected symbols, and
$\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$. $\mathcal{X} \subseteq \mathbb{C}^{K'}$, contains the constellation
set of all users. Since the ML implementation is prohibitively
complex, we propose a reduced-complexity MSD detection
scheme that is able to achieve the optimal ML performance.

### III. MSD Detection Scheme

In this section, we develop a reduced complexity SD detection
scheme that is based on a modified tree search method, and
is capable of achieving the ML performance.

#### A. Problem Formulation

It is clear that (2) represents an under-determined system;
thus, SD cannot be directly applied. To overcome this issue,
and inspired by Tikhonov regularization [4], we rewrite $\mathbf{G}$
in (1) as,

$$ \tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G}^{(1)}_{N \times N} & \mathbf{G}^{(2)}_{N \times (K'-N)} \end{bmatrix} \quad \text{for } \quad K' = d_v = 2, \
\text{and } \quad M = 4. $$

As such, the ML detection problem is

$$ \hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{X}} \left( \| \mathbf{y} - \tilde{\mathbf{G}} \mathbf{x} \|^2 - \| \mathbf{x}^{(2)} \|^2 \right), $$

which represents well-defined systems of equations, and is
equivalent to the ML detection problem in (2) for constellation
points with constant modulus, i.e., $\| x^{(2)} \|^2$ is a constant (e.g.,
4-QAM or the codebook in [5]). Since $\mathbf{S}$ is an upper-triangular
matrix, $\tilde{\mathbf{G}}$ is upper-triangular and MSD can then be directly applied.
That is, the choice of $\mathbf{S}$ alleviates the need of QR
factorization prior to SD as in e.g., (6), which will substantially
reduce the computational complexity of the proposed MSD.

#### B. Modified Tree Search Method

The proposed MSD detection scheme can be visualized by a
search over a tree with $K'$ layers. The tree search is performed
in descending order from the last layer down to the first layer,
wherein the layers with the indices $k' \in \{(k - 1) d_v + 1, k d_v\}$
correspond to user $k$. Since each user spreads the same
information bits over $d_v$ REs, there exists up to $M$ branches for
layers with indices $k' = k d_v$, $k \in \{1, \ldots, K\}$, and only one
branch for layers with indices $k' \in \{(k - 1) d_v + 1, \ldots, k d_v - 1\}$. The modified tree search method is illustrated in Fig. 1
for a scenario with $K = 6$, $d_v = 2$, and $M = 4$.

#### C. MSD Extension to Non-constant Modulus Constellations

As mentioned in Section III-A, (5) is equivalent to (2) for constellation
points with a constant modulus, e.g., 4-QAM. However, as long as the SCMA codebook can be written as $\mathbf{x} = \Omega \mathbf{x}'$, where the entries of $\mathbf{x}'$ are taken from a
constant modulus constellation and $\Omega$ is a block diagonal
matrix, then the decoding algorithm can readily be employed
the same way as for the constant modulus case. This condition
applies for most good SCMA codebooks, e.g., [7] or 16-QAM.
For illustrative purposes, consider an $M$-QAM constellation,
$\nu = [\nu_1, \ldots, \nu_M]$, with $M = 4^n$, that can be constructed
from $4$-QAM constellations. Let $V'$ represent the $m \times 4^n$
matrix resulting from the Cartesian product of $m$ tuples of $4$-
QAM constellations, and let $v'_k, k \in \{1, \ldots, 4^n\}$, denote the $i$th column of $V'$. Each $M$-QAM constellation point, $\nu_i$, is
$\nu_i = \sum_{j=1}^{m} 2^{j-1} u'_i$. In a similar vein, the $K'$-dimensional vector $\mathbf{x}$, containing the transmitted symbols of all layers can be decomposed as $\mathbf{x} = \Omega \mathbf{x}'$, where $\Omega = \text{diag}(\omega_1, \ldots, \omega_{K'})$ is a $K' \times K'$ block diagonal matrix, and $\omega_{K'} = [2^{m-1}, \ldots, 1]$. Further, $\mathbf{x}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_{K'})^T$ is an $m K'$-dimensional vector, and the entries of $\mathbf{x}'_k$ are chosen from the $4$-QAM constellation.

The received signal vector can be re-written

$$ \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left( \| \mathbf{y} - \tilde{\mathbf{G}} \mathbf{x} \|^2 - \| \mathbf{x}^{(2)} \|^2 \right), $$

which represents well-defined systems of equations, and is
equivalent to the ML detection problem in (2) for constellation
points with constant modulus, i.e., $\| x^{(2)} \|^2$ is a constant (e.g.,
4-QAM or the codebook in [5]). Since $\mathbf{S}$ is an upper-triangular
matrix, $\tilde{\mathbf{G}}$ is upper-triangular and MSD can then be directly applied.
That is, the choice of $\mathbf{S}$ alleviates the need of QR
factorization prior to SD as in e.g., (6), which will substantially
reduce the computational complexity of the proposed MSD.

![Fig. 1: Modified tree search method for $K = 6$, $d_v = 2$, and $M = 4$.](image)

On the occasion that there are not $N/d_v$ orthogonal active users in
the system, we can relabel the layers rather than the users and some minor
modifications to the SD algorithm will be needed.
as \( y = G' x' + w \), where \( G' = G \Omega \). The MSD detection scheme then runs the same way as the constant modulus case.

IV. COMPLEXITY DISCUSSION

The principles of the SD algorithm necessitate the detected symbols, \( \hat{x} \), falls within a hypersphere of radius \( d \) by ensuring that the following is held \([6]\):

\[
d^2 \geq \sum_{i=1}^{K'} |y_i - \sum_{j=1}^{K'} \tilde{g}_{i,j} \hat{x}_j|^2, \tag{6}
\]

where \( \tilde{g}_{i,j} \) denotes each element of \( \tilde{G} \). From the choice of \( S \) to be an upper-triangular matrix and from Section III-A the following are observed for \( G \), which introduce a reduction in the number of operations involved in detecting the transmitted symbols: Firstly, since the first \( N \) columns of \( S \) form an identity matrix, \( \mathcal{G}_{1 \times N} \) is a full-rank matrix with non-zero diagonal elements, i.e., \( \tilde{g}_{i,j} \neq 0, j = 1, \ldots, K' \). Secondly, due to the sparsity of SCMA codebooks, the number of non-zero elements in the first \( N \) rows of \( G \) is \( d_f \), and the position of those non-zero elements is determined by the position of ones in the corresponding row of the mapping matrix, \( S \). Thus, only \( d_f \) layers are involved in detecting each symbol corresponding to the first \( N \) rows of \( G \). That is, from \([6]\), to detect the symbol corresponding to the first \( N \) layers, i.e., \( k' \in \{1, \ldots, N\} \), MSD selects a constellation point, \( \hat{x}_{k'} \), that satisfies

\[
d^2 \geq d_f^2 + |y_{k'} - \sum_{j \in \mathcal{F}_{1',k'}} \tilde{g}_{k',j} \hat{x}_j|^2, \tag{7}
\]

where \( \mathcal{F}_{1',k'} = \{i|\tilde{g}_{k',i} = 1\} \). Thirdly, due to the presence of \( I \) in \([3]\), the last \( K' - N \) rows of \( G \) will have only one non-zero element, which is on the diagonal. That is, for the last \( K' - N \) layers, i.e., \( k' \in \{N + 1, \ldots, K'\} \),

\[
d^2 \geq d_f^2 + |y_{k'} - \tilde{g}_{k',k'} \hat{x}_{k'}|^2. \tag{8}
\]

where \( d_f^2 = \sum_{i=k'+1}^{K'} |y_i - \sum_{j=i}^{K'} \tilde{g}_{i,j} \hat{x}_j|^2 \), when \( k' \in \{1, \ldots, K' - 1\} \), and \( d_f^2 = 0 \) when \( k' = K' \). Let \( N_{c_1} \) and \( N_{c_2} \) denote the average number of visited layers \([6]\), \([8]\)–\([10]\) for \( k' \in \{1, \ldots, N\} \), and \( k' \in \{N + 1, \ldots, K'\} \), respectively. The average complexity of MSD based on \([7]\)–\([8]\), and the complexity of log-MPA \([11]\) using max operation \([11]\) is provided in Table I Note from \([6]\), unlike \([7]\) and \([8]\), the detection of each symbol involves the contribution of all other symbols. Further, the conventional SD branches up to \( M \) possibility for all the \( K' \) layers, whereas from Section III-B we branch up to \( M \) possibilities only for \( K \) layers, and branch only 1 possibility for the \( K' - K \) layers. This suggests a substantial reduction in the complexity of the proposed MSD compared with the conventional SD\([3]\).

V. LIST MSD

To use channel coding soft outputs are required; we employ list MSD that is based on the list SD \([11]\) to provide soft outputs. Let \( N_c \) denotes the length of codewords output from the channel encoder. The codewords are partitioned into \( N_c/L_M \) digital symbols of \( L_M \) bits each. Let \( c_k = (c_{k,1}, \ldots, c_{k,L_M}) \).

As a result, the dimensionality increase due to using \( S \) instead of \( P \) is compensated with the MSD scheme.

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed MLD detection scheme over AWGN and Rayleigh fading channels, where we assume each user observes the same channel coefficients over \( d_v \) resources. We provide performance comparisons among different detection schemes for uplink SCMA systems with \( K = 6 \), \( N = 4 \), and \( d_v = 2 \).

In Fig. \([3]\) we compare the BER performance of uncoded SCMA systems that differ in their multiuser detection techniques, and operate over the AWGN channel, with the 4-ary codebook in \([5]\). We observe that MSD achieves the ML performance. Moreover, MPA approaches ML with increasing number of iterations, \( N_i \). In particular, MPA converges to
TABLE I: Average complexity of MSD and MPA

|                | MSD                                                                 | log-MPA                                                                 |
|----------------|----------------------------------------------------------------------|------------------------------------------------------------------------|
| Real Summators | $(4d_f + 2)N_{v_1} + 2N_{v_2}$                                      | $M N d_f \left( M^{d_f - 1} (4d_f - 2 + N_1 (2 + \frac{1}{M})) + N_1 (2 - \frac{1}{M}) + 5 \right)$ |
| Real Multipliers| $(4d_f + 2)N_{v_1} + 2N_{v_2}$                                      | $M N d_f (4d_f M^{d_f - 1} + 5$                                       |
| exp/log Operations | $0$                                                                  | $M N d_f N_1 (M^{d_f - 1} + 1) + 1$                                   |

The complexity of MSD is reduced compared with the conventional SD based on the sparse structure of SCMA codebooks that does not require to expand all tree branches. In addition, in order to use channel coding, list MSD was employed. Simulation results show that in uncoded SCMA systems the proposed MSD scheme achieves the performance of the optimal ML detection. Also, the proposed MSD benefits from a lower average complexity compared with MPA.

REFERENCES

[1] H. Nikopour and H. Baligh, “Sparse code multiple access,” in *Proc. IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2013, pp. 332–336.

[2] D. Wei, Y. Han, S. Zhang, and L. Liu, “Weighted message passing algorithm for SCMA,” in *Proc. International Conference on Wireless Communications & Signal Processing (WCSP)*, Oct. 2015.

[3] F. Wei and W. Chen, “Low complexity iterative receiver design for sparse code multiple access,” *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.

[4] G. H. Golub, P. C. Hansen, and D. P. O’Leary, “Tikhonov regularization and total least squares,” *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–194, 1999.

[5] Altera Innovate Asia FPGA Design Contest. (2015) 5G Algorithm Innovation Competition. [Online]. Available: http://www.innovateasia.com/5g/en/gp2.html.

[6] B. Hassibi and H. Vikalo, “On the sphere-decoding algorithm I. Expected complexity,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.

[7] M. Tahirzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, “SCMA codebook design,” in *Proc. IEEE Vehicular Technology Conference (VTC)*, Sep. 2014.

[8] T. Cui, S. Han, and C. Tellambura, “Probability-distribution-based node pruning for sphere decoding,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1586–1596, May 2013.

[9] P. Wang and T. Le-Ngoc, “A low-complexity generalized sphere decoding approach for underdetermined linear communication systems: performance and complexity evaluation,” *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3376–3388, Nov. 2009.

[10] M. M. Mansour, S. P. Alex, and L. M. A. Jalloul, “Reduced complexity soft-output MIMO sphere detectors–part I: algorithmic optimizations,” *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5505–5520, Nov. 2014.

[11] B. M. 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.

[12] Evolved Universal Terrestrial Radio Access Network, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network,” 2011.