

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Quantum Search Aided Multi-User Detection for Sparse Code Multiple Access

WENJING YE¹, WEI CHEN¹, (Senior Member, IEEE), XIN GUO², CHEN SUN², (Senior Member, IEEE), AND LAJOS HANZO³, (Fellow, IEEE)

¹Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. (e-mail: yewj16@mails.tsinghua.edu.cn; wchen@tsinghua.edu.cn)

²Sony China Research Laboratory, Sony (China) Ltd., Beijing 100028, China (e-mail: xin.april.guo@outlook.com; chen.sun@sony.com)

³School of ECS, University of Southampton, SO17 1BJ, UK (email: lh@ecs.soton.ac.uk)

Corresponding author:

The work of Wenjing Ye and Wei Chen is supported by the Beijing Natural Science Foundation (4191001). The work of L. Hanzo is supported by the EPSRC projects EP/N004558/1, EP/PO34284/1, by the Royal Society's GRFC Grant as well as by the European Research Council's Advanced Fellow Grant QuantCom.

ABSTRACT With the explosive proliferation of mobile devices, the scarcity of communication resources has emerged as a critical issue in the Next Generation (NG) communication systems. To overcome this, Non-Orthogonal Multiple Access (NOMA) has been shown to provide a beneficial spectral efficiency gain, while supporting massive connectivity. Sparse Code Multiple Access (SCMA) is one of the NOMA schemes, but its application is hampered by its high-complexity multi-user detection. Fortunately, quantum search techniques were shown to substantially reduce the complexity of multi-user detectors (MUD). However, the quantum search aided multiuser detection of SCMA is an open problem at the time of writing. Hence we conceive a pair of quantum search aided MUDs for SCMA, namely, a Quantum-assisted Message Passing Algorithm (Q-MPA) based MUD and a Quantum-assisted Sphere Decoder based MPA (QSD-MPA) MUD.

INDEX TERMS sparse code multiple access, multiuser detection, quantum search algorithm, message passing algorithm, sphere decoder.

I. INTRODUCTION

Given the explosive proliferation of mobile devices, communication resources are becoming scarce commodities. To circumvent this problem, Non-Orthogonal Multiple Access (NOMA) has been employed for enhancing the spectral efficiency in the face of massive connectivity. In the family of NOMA schemes, Sparse Code Multiple Access (SCMA) has been proposed as a promising scheme with superior link level performance [1]- [5]. However, the implementation of SCMA is challenging owing to its excessive Multi-User Detection (MUD) complexity.

To circumvent this, sophisticated classical methods have been conceived for reducing the complexity of SCMA MUD. The basic method relies on the classical Message Passing Algorithm (MPA) based detector that achieves near-optimal Bit Error Rate (BER) performance [6], [7]. However, the classical MPA still exhibits an exponentially growing complexity with the number of users. Hence compelling techniques have been put forward for further reducing the MUD

complexity. A plausible approach is to directly simplify the MPA calculation [8]- [11]. The max-log approximation of [8] and the Look Up Table (LUT) based technique of [9] can readily simplify the MPA at a negligible BER performance erosion. The Partial Marginalization technique based MPA (PM-MPA) of [10] can be adjusted to strike an attractive BER vs complexity trade-off. In [11], the MUD complexity was reduced by a technique that selects the edges of the classical MPA based on an adaptive Gaussian approximation. To further reduce the complexity of the classical MPA, low-complexity codebook design techniques were advocated in [12], [13]. Specifically, designing SCMA codebooks relying on the minimum projections technique before invoking the MPA process (ProjMPA) may result in a reduced effective codebook size [12]. Thus, the ProjMPA imposes a much lower complexity than the classical MPA. Moreover, a high-dimensional codebook design technique was proposed in [13] for reducing the MUD complexity per symbol. Additionally, sphere decoder based techniques can also be applied in

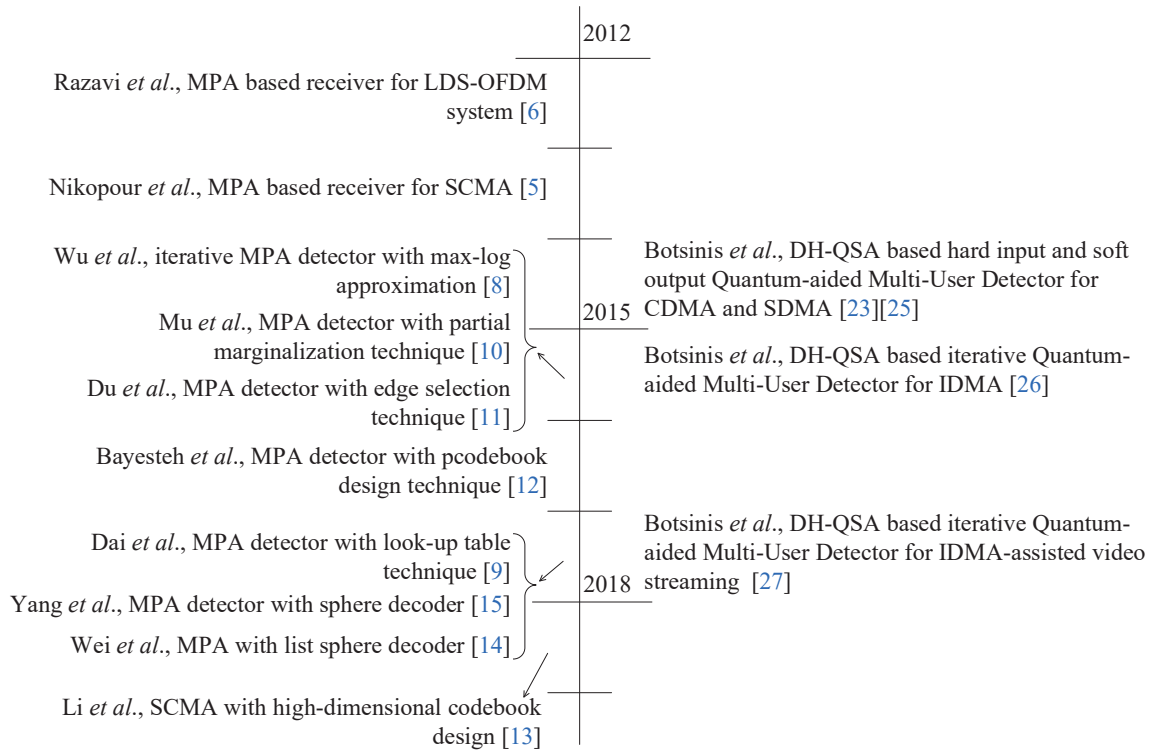


FIGURE 1: Selected contributions on SCMA detector (left) and quantum-assisted multi-user detection (right).

MUD-aided SCMA. Specifically, invoking the List Sphere Decoder (LSD) of [14] or the Sphere Decoder (SD) of [15] before the classical MPA also reduces the detection complexity at a marginal BER performance erosion.

As a future solution, the emerging quantum computing technology promises to reduce the complexity of classical search problems, as detailed in [16]. Hence, significant benefits may be expected in the field of communications by invoking quantum computing at the base station, where its complexity and price may be deemed acceptable in the future. Hence in this treatise, we investigate the benefit of Quantum Search Algorithms (QSA) in reducing the complexity of MUD.

In an unsorted database, QSAs are capable of finding desired solutions at a lower complexity than the classical exhaustive search algorithms [17]- [20]. Thus, QSAs can be chosen for accelerating the solutions of numerous classical search problems as detailed in [21]. The earliest quantum search based MUD was conceived for Code Division Multiple Access (CDMA) in [22], where a Hard Input Hard Output (HIHO) quantum-assisted MUD was shown to approach the optimal Maximum Likelihood (ML) BER performance at a reduced complexity. Recent contributions on quantum-assisted MUD focus on Space Division Multiple Access (SDMA) [23]- [25], CDMA, and Interleave Division Multiple Access (IDMA) [26]- [27]. For Space Division Multiple Access (SDMA) and CDMA, quantum-assisted MUDs with optimal and sub-optimal BER performance were

considered [23]- [25]. Promising quantum-assisted MUDs were designed for multi-carrier IDMA systems and multi-layered video streaming in [26], [27]. As a benefit of the intrinsic quantum parallelism and quantum superposition exploited by the QSAs, these contributions demonstrated significant complexity reductions compared to their classical counterparts. Fig. 1 summarizes recent contributions on low-complexity SCMA MUD and quantum-assisted MUD. However, reduced-complexity quantum-assisted MUDs have not been designed for SCMA.

Against this background, we intrinsically amalgamate QSAs with the classical MPA detection schemes for conceiving low-complexity MUDs for SCMA. Specifically, we design a pair of reduced-complexity QSA-aided classical MPAs, namely, the Quantum-assisted Message Passing Algorithm (Q-MPA) and the Quantum-assisted Sphere Decoder based MPA (QSD-MPA). Specifically, the Q-MPA is conceived by applying QSAs for accelerating the message updating in the MPA. Our theoretical analysis and simulation results show that the Q-MPA requires a lower number of Cost-Function (CF) evaluations than the classical MPA. Additionally, our simulation results demonstrate that the Q-MPA based SCMA MUDs only suffer from a modest BER performance erosion compared to the classical MPA. Furthermore, we use the ProjMPA technique of [12] for further reducing the complexity of the Q-MPA. We will also demonstrate that the QSD-MPA can further reduce the MUD complexity by harnessing the classical SD-MPA. In the QSD-

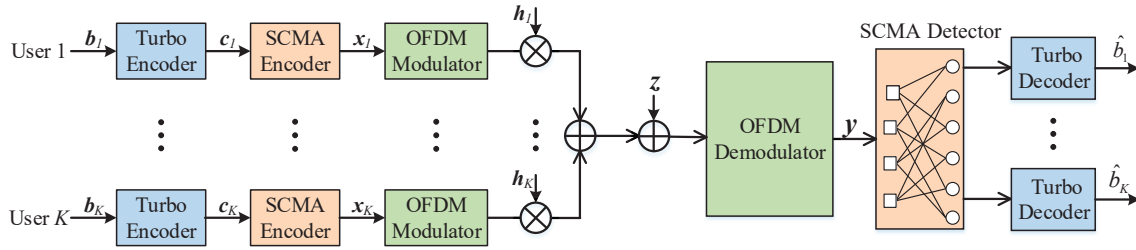


FIGURE 2: Uplink SCMA with turbo coding.

MPA, the sphere decoding algorithm can be formulated as a search algorithm. To efficiently solve the resultant search problem, we present a Quantum-assisted Sphere Decoder (QSD) based on the Boyer-Brassard-Hoyer-Tapp Quantum Search Algorithm (BBHT-QSA) [19]. Even if the number of legitimate codeword combinations is unknown, the QSD is capable of finding all the desired legitimate codeword combinations. Let us now denote the number of legitimate codeword combinations by S . When S is smaller than a certain critical value, the QSD requires a lower number of CF Evaluations (CFE) than the classical SD. By invoking the QSD before the MPA, we arrive at our new QSD-MPA MUD scheme. The theoretical analysis and simulation results show that the QSD-MPA has a lower number of CFEs than the SD-MPA, provided that the number of solutions is smaller than a critical value. Moreover, our simulation results show that the QSD-MPA approaches the BER performance of the SD-MPA at a reduced complexity. Thus, the Q-MPA and the QSD-MPA are potential candidates for reduced-complexity SCMA MUD.

The rest of this paper is organized as follows. Section II presents the uplink SCMA system model. Based on the classical MUDs and QSAs, our Q-MPA and the QSD-MPA are conceived in Sections III and IV, respectively. In Section V, our numerical results characterizing the Q-MPA and the QSD-MPA are presented. Finally, our conclusions are offered in Section VI.

II. SYSTEM MODEL

In this section, we first present the system model for the SCMA uplink along with the SCMA codeword structure.

A. UPLINK SCMA MODEL

We are interested in a non-orthogonal multiple access system, in which K ($K > 0$) users transmit to a common Base Station (BS) simultaneously in the shared spectrum. Its block diagram is shown in Fig. 2. The bit stream of the user k emanating from the source is denoted by b_k , which is entered into the turbo encoder. The output of the turbo encoder is divided into code blocks, where each block consists of R coded bits. A code block is denoted by $c_k = (c_{k,1}, c_{k,2}, \dots, c_{k,R})^T$, which is forwarded to the SCMA encoder. In the SCMA

encoder, c_k is mapped into an N -dimensional SCMA codeword denoted by $x_k = (x_{k,1}, x_{k,2}, \dots, x_{k,N})^T$ that is selected from the SCMA codebook \mathcal{X}_k of size $M = 2^R$. Finally, the SCMA codewords are transmitted by the Orthogonal Frequency Division Multiplexing (OFDM) modulator.

The modulated symbols are transmitted through the wireless channels by using N OFDM sub-carriers. Let $h_{k,n}$ denote the channel coefficient of the link between the BS and user k on the n -th sub-carrier. The channel vector between user k and the BS is given by $h_k = (h_{k,1}, h_{k,2}, \dots, h_{k,N})^T$. In addition, let z_n denote the Additive White Gaussian Noise (AWGN) imposed on the n -th sub-carrier with variance σ^2 . The noise imposed on the transmitted signal is denoted by $z = (z_1, z_2, \dots, z_N)^T$.

At the BS side, the received signal vector is denoted by $y = (y_1, y_2, \dots, y_N)^T$, in which y_n is the received signal at the n -th sub-carrier. As shown in Fig. 3(a), the signals received from all users at the same codeword location are transmitted on the same OFDM sub-carrier. Then, the received signal y is given by

$$y = \sum_{k=1}^{K} \text{diag}\{h_k\}x_k + z. \quad (1)$$

When y is received, the information bits will be restored by using both the SCMA Detector of Fig. 2 and the turbo decoder.

B. SCMA CODEWORD STRUCTURE

In this subsection, we characterize the codeword structure by using a mapping matrix and the factor graph concept. The mapping matrix and the factor graph are denoted by F and $\mathcal{F}(\mathcal{V}, \mathcal{N})$, respectively. The mapping matrix F is a $(K \times N)$ -element matrix, where the element at the k -th row and n -th column of F is denoted by $F_{k,n}$. If the n -th element of k -th user's codeword is non-zero, then we have $F_{k,n} = 1$, otherwise $F_{k,n} = 0$. Furthermore, the factor graph $\mathcal{F}(\mathcal{V}, \mathcal{N})$ is a bipartite graph with N Function Nodes (FN) and K Variable Nodes (VN). On the factor graph, the n -th FN is connected to the k -th VN via an edge if $F_{k,n} = 1$. Moreover, the set $\phi(n) = \{k : F_{k,n} = 1\}$ is defined as the collection of neighboring nodes of FN n . Similarly, the neighbor nodes of VN k are defined as $\psi(k) = \{n : F_{k,n} = 1\}$ that has a

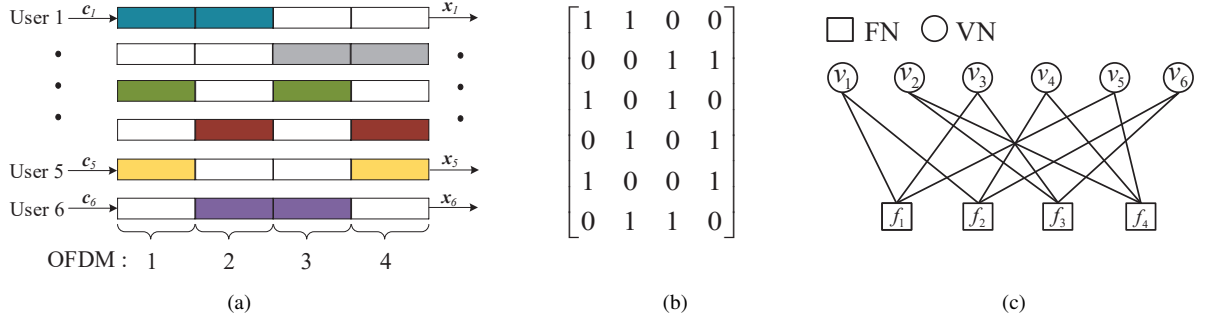


FIGURE 3: The structure of SCMA codewords. (a) The colored blocks represent the non-zero VN elements in the SCMA codewords, while the blank blocks represent the zero elements. In this example, the parameters are set as $K = 6$ and $N = 4$. (b) The mapping matrix. The 1-element and 0-element represent the non-zero element and zero element in SCMA codewords, respectively. (c) The factor graph. In this example, the degree of $\phi(n)$ and $\psi(k)$ are $d_f = 3$ and $d_v = 2$, respectively.

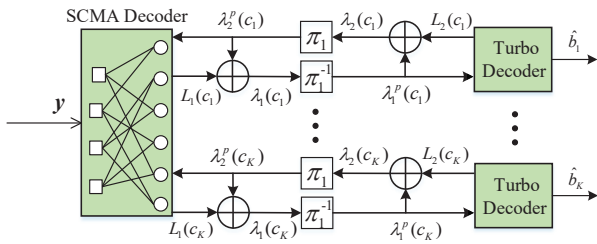


FIGURE 4: The iterative structure of the multi-user detector.

degree of $|\psi(k)| = d_v$. Additionally, the sets $\phi(n)$ and $\psi(k)$ represent the users colliding on the n -th OFDM sub-carrier and the OFDM subcarriers used to transmit the k -th user's data, respectively. The mapping matrix and factor graph are shown in Fig. 3(b) and Fig. 3(c), respectively.

III. THE QUANTUM-ASSISTED MPA MUD

In this section, we introduce a Q-MPA based iterative MUD. Based on the classical MPA, we then propose a Q-MPA detection scheme and conceive its reduced complexity counterpart.

A. THE ITERATIVE STRUCTURE OF MULTI-USER DETECTOR

In the MUD, the extrinsic information is exchanged between the SCMA detector and the turbo decoder. In the SCMA detector, the *a posteriori* Log-Likelihood Ratio (LLR) denoted by L_1 of each channel coded bit is calculated as

$$L_1(c_{k,r}) = \log \frac{\Pr\{c_{k,r} = 1|y\}}{\Pr\{c_{k,r} = 0|y\}}. \quad (2)$$

By exploiting Bayes' law, we have

$$\begin{aligned} L_1(c_{k,r}) &= \log \frac{\Pr\{y|c_{k,r} = 1\}}{\Pr\{y|c_{k,r} = 0\}} + \log \frac{\Pr\{c_{k,r} = 1\}}{\Pr\{c_{k,r} = 0\}} \\ &= \lambda_1(c_{k,r}) + \lambda_2^p(c_{k,r}), \end{aligned} \quad (3)$$

where $\lambda_1(c_{k,r})$ is the extrinsic information delivered by the SCMA detector, while $\lambda_2^p(c_{k,r})$ is the *a priori* LLR delivered by the turbo decoder. In the turbo decoder, the *a posteriori* LLR denoted by L_2 is also calculated similarly

$$L_2(c_{k,r}) = \lambda_2(c_{k,r}) + \lambda_1^p(c_{k,r}), \quad (4)$$

where $\lambda_1^p(c_{k,r})$ represents the *a priori* LLR gleaned from the SCMA detector. Using the Q-MPA, we can obtain the value of $\lambda_1(c_{k,r})$ in the SCMA detector at the cost of a reduced number of CF evaluations.

B. THE QUANTUM-ASSISTED MPA

In this section, we present a Q-MPA method along with its complexity analysis. The Q-MPA is conceived by using Durr-Hoyer Quantum Search Algorithm (DH-QSA) to reduce the complexity of the classical MPA. Readers can refer to the appendix for details about the classical MPA and DH-QSA.

1) The Procedure of Q-MPA

The Q-MPA consists of message updating in FNs, message updating in VNs, and the LLR computation. For the first two steps, the messages are updated iteratively between the FNs and the VNs on the factor graph. The message sent from the n -th FN to the k -th VN at the j -th iteration is denoted by $l_{n \rightarrow k}^j(x_k)$, where x_k is the SCMA codeword of user k . Similarly, the message sent from the k -th VN to the n -th FN at the j -th iteration is denoted by $l_{n \leftarrow k}^j(x_k)$. When $j = 0$, the initial messages are set to 0. We denote the maximum number of MPA iterations by J . Then the messages will be updated iteratively until $j = J$ is reached and the LLR will be computed by using the results of the iterative process.

We first describe the message updating in FNs. In this step, a pair of equivalent CFs, i.e., $f_{n,k}^c$ and $f_{n,k}^q$ are defined in the classical processing phase and the quantum processing phase, respectively. When the codeword of user k is x_k^m , the CF $f_{n,k}^c$

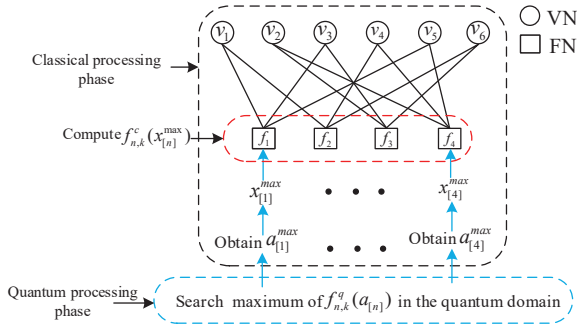


FIGURE 5: The diagram of Q-MPA. The results from quantum search will be used to update messages in the FNs. Then, the messages are updated between the FNs and VNs as the classical MPA.

is given by

$$f_{n,k}^c(x_{[n]}) = -\frac{1}{2\sigma^2} \|y_n - \sum_{l \in \phi(n)} h_{l,n} x_{l,n}\|^2 + \sum_{l \in \phi(n) \setminus k} l_{n \leftarrow l}^{j-1}(x_l), \quad (5)$$

where $x_{[n]} = \{x_{l,n} : l \in \phi(n)\}$ is a codeword combination at the n -th sub-carrier. Then, we formulate the equivalent CF $f_{n,k}^q$ in the quantum domain. Let us denote the m -th SCMA codeword in the codebook \mathcal{X}_k by x_k^m . We define $\mathcal{X}_{k,n} = (x_{k,n}^1, x_{k,n}^2, \dots, x_{k,n}^M)^T$, where $x_{k,n}^m$ is the n -th element of x_k^m . Moreover, we define a vector $L_{n \leftarrow k}^j$ given by

$$L_{n \leftarrow k}^j = [l_{n \leftarrow k}^j(x_k^1), l_{n \leftarrow k}^j(x_k^2), \dots, l_{n \leftarrow k}^j(x_k^M)]^T. \quad (6)$$

Equivalent to $f_{n,k}^c$, the CF $f_{n,k}^q$ is given by

$$f_{n,k}^q(a_{[n]}) = -\frac{1}{2\sigma^2} \|y_n - h_{k,n} x_{k,n} - \sum_{l \in \phi(n) \setminus k} h_{l,n} a_l \mathcal{X}_{l,n}\|^2 + \sum_{l \in \phi(n) \setminus k} a_l L_{n \leftarrow l}^{j-1}, \quad (7)$$

where $a_{[n]} = \{a_l : l \in \phi(n)\}$, a_l is a M -dimensional vector denoted by $a_l = (a_{l,1}, a_{l,2}, \dots, a_{l,M})$, $a_{l,m}$ belongs to $\{0, 1\}$ and $\sum_{m=1}^M a_{l,m} = 1$. Moreover, we denote a_l associated with $a_{l,m} = 1$ by a_l^m , where a_l^m belongs to $\{a_l^1, a_l^2, \dots, a_l^M\}$. If $a_l = a_l^m$, then we have $x_l = x_l^m$. As a result, if $a_{[n]} = \{a_l^{m_l} : l \in \phi(n), 1 \leq m_l \leq M\}$, then a codeword combination is formulated as $x_{[n]} = \{x_{l,n}^{m_l} : l \in \phi(n)\}$.

The message updating in FNs contains the quantum processing phase and the message processing phase. In the quantum processing phase, the entry $x_{[n]}^{max}$ that maximizes $f_{n,k}^c$ is found by a modified DH-QSA. Specifically, the DH QSA of [20] utilizes $f_{n,k}^q$ and $f_{n,k}^c$ as the CFs in the quantum domain and the classical domain, respectively. A threshold value δ is initialized randomly. Then the specific input $a_{[n]}^s$ that satisfies $f_{n,k}^q(a_{[n]}^s) > \delta$ is termed as a solution. Then, the DH-QSA iteratively employs the BBHT-QSA for finding the solutions. When a result is found by the BBHT-QSA, a

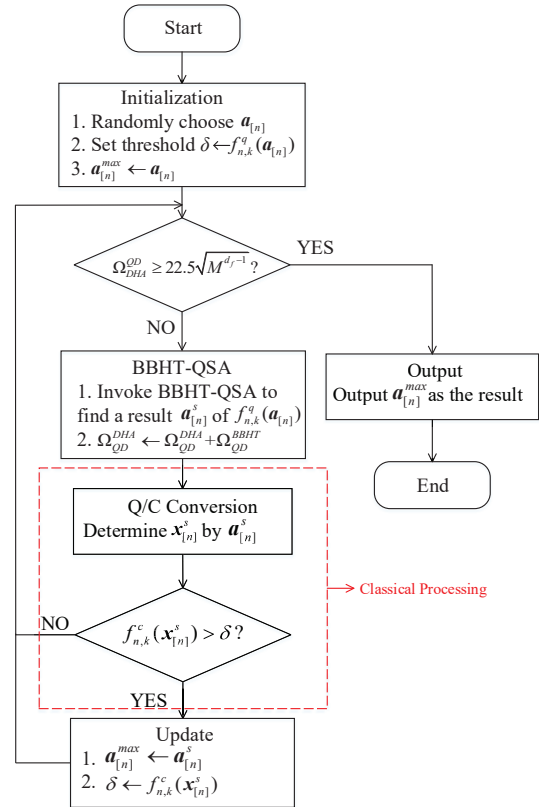


FIGURE 6: Flowchart of Algorithm 1. The Q/C Conversion step converts the quantum information $a_{[n]}^s$ output by the BBHT-QSA into classical information $x_{[n]}^s$. Then the value $f_{n,k}^c(x_{[n]}^s)$ will be computed in the classical domain.

codeword combination $x_{[n]}^s$ is determined by $a_{[n]}^s$. The CF value $f_{n,k}^c(x_{[n]}^s)$ is computed in the classical domain to check if it is a solution. Since $f_{n,k}^q$ and $f_{n,k}^c$ are equivalent, the CF value $f_{n,k}^c(x_{[n]}^s)$ is equivalent to $f_{n,k}^q(a_{[n]}^s)$. The reason for this modification is to avoid introducing extra complexity in the classical domain. If $a_{[n]}^s$ is a solution, the threshold value δ is updated to be $f_{n,k}^c(x_{[n]}^s)$. The DH-QSA will terminate when no solution can be found or the maximum number of applying Grover operators has been exhausted. Finally, the $x_{[n]}^{max}$ will be found by the DH-QSA. The DH-QSA is summarized in Algorithm 1 and its flowchart is also given in Fig. 6. Interested readers might like to refer to Appendix B for further details on the basis of quantum computing and QSA.

In the classical processing phase, the particular codeword combination $x_{[n]}^{max}$ that maximizes $f_{n,k}^c$ is found by the DH-QSA. Then, the FN updates the message $l_{n \rightarrow k}^j(x_k)$ in the classical domain as

$$l_{n \rightarrow k}^j(x_k) = f_{n,k}^c(x_{[n]}^{max}). \quad (8)$$

The message updating in FNs is illustrated in Fig. 5. The messages obtained are then forwarded to the neighbor VNs.

Algorithm 1 The Modified DH-QSA for Q-MPA

- 1: Set $\Omega_{DHA}^{QD} \leftarrow 0$. Randomly choose $a_{[n]}$, set threshold $\delta \leftarrow f_{n,k}^q(a_{[n]})$, $a_{[n]}^{max} \leftarrow a_{[n]}$.
- 2: The BBHT-QSA is employed to find a solution $a_{[n]}^s$ that satisfies $f_{n,k}^q(a_{[n]}^s) > \delta$. Obtaining a solution $a_{[n]}^s$ and Ω_{BBHT}^{QD} from the BBHT-QSA.
- 3: $\Omega_{DHA}^{QD} \leftarrow \Omega_{DHA}^{QD} + \Omega_{BBHT}^{QD}$.
- 4: Determine $x_{[n]}^s$ by $a_{[n]}^s$.
- 5: **if** $f_{n,k}^s(x_{[n]}^s) \leq \delta$ or $\Omega_{DHA}^{QD} \geq 22.5\sqrt{M^{d_f-1}}$ **then**
- 6: Output $a_{[n]}^{max}$ and exit.
- 7: **else**
- 8: Set threshold $\delta \leftarrow f_{n,k}^q(a_{[n]}^s)$, $a_{[n]}^{max} \leftarrow a_{[n]}^s$, go to step 2.
- 9: **end if**

Let us now describe the message updating processes in VNs. Since x_k belongs to $\{x_k^1, x_k^2, \dots, x_k^M\}$, the messages are updated in the VNs as follows:

$$l_{n \leftarrow k}^j(x_k) = \sum_{u \in \psi(k) \setminus n} l_{u \rightarrow k}^j(x_k). \quad (9)$$

Then, the messages obtained are forwarded to the neighbor FNs. The above message updating steps between the FNs and VNs will be continuous until the maximum number of iterations is reached.

Finally, the LLRs are computed when $j = J$ is reached. Specifically, the SCMA detector calculates the LLR of $c_{k,r}$ as

$$\lambda_1(c_{k,r}) = \max_{x_k \in \mathcal{X}^1(c_{k,r})} L(x_k) - \max_{x_k \in \mathcal{X}^0(c_{k,r})} L(x_k) - \lambda_2^p(c_{k,r}), \quad (10)$$

where we have

$$L(x_k) = \sum_{n \in \psi(k)} l_{n \rightarrow k}^J(x_k). \quad (11)$$

Both of the message updating step in the VNs and the LLR computation step are carried out in the classical processing phase. In this way, the LLR of (2) can be obtained by the Q-MPA. The Q-MPA is summarized in Algorithm 2.

2) Complexity Analysis of the Q-MPA

Since the Q-MPA consists of the quantum processing phase and the classical processing phase, the total complexity of the Q-MPA is constituted by the sum of the complexity in the quantum domain and in the classical domain. Specifically, the complexity in the quantum domain is determined by the number of Grover operator applications, while the complexity in the classical domain is determined by the number of classical CFEs in (5). In this paper, we use the idealized simplifying assumption that one Grover operator application in the quantum domain is equivalent to one CFE in the classical domain. Moreover, we characterize the complexity of the Q-MPA by the total Number of CFEs (NCFE), which is denoted by Ω_{QMMPA} . In Theorem 1, we present the NCFE of the Q-MPA.

Algorithm 2 Quantum-assisted MPA

- 1: **Initialization:**
 $l_{n \rightarrow k}^0(x_k) \leftarrow 0$, $l_{n \leftarrow k}^0(x_k) \leftarrow 0$, $L_{n \leftarrow k}^0(x_k) \leftarrow [0, 0, \dots, 0]^T$, $\Omega_{QMMPA} \leftarrow 0$.
- 2: **Iterative Updating:**
- 3: **for** $j \leq J$ **do**
- 4: **for** $n \leq N$, $k \leq K$ **do**
- 5: **for** $m \leq M$ **do**
- 6: Invoke Algorithm 1 to find the index $x_{[n]}^{max}$ that maximizes $f_{n,k}^c(x_{[n]})$, obtaining Ω_{DHA} .
- 7: $\Omega_{QMMPA} \leftarrow \Omega_{QMMPA} + \Omega_{DHA}$.
- 8: Update $l_{n \rightarrow k}^j(x_k^m) \leftarrow f_{n,k}^c(x_{[n]}^{max})$.
- 9: Update $l_{n \leftarrow k}^j(x_k^m)$ as (9). Store the messages in vector $L_{n \leftarrow k}^j(x_k)$ as (6).
- 10: **end for**
- 11: **end for**
- 12: **end for**
- 13: **Calculate LLR:**
 Compute $\lambda_1(k, r)$ as (10).

Theorem 1: The total NCFE of the Q-MPA is $O(M^{(d_f+1)/2})$.

Proof 1: In a single DHA search, let Ω_{CD} and Ω_{QD} denote the NCFE in the classical domain and the complexity in the quantum domain, respectively. According to (7), there are M^{d_f-1} codeword combinations for each CF. As a result, the size of the search space is M^{d_f-1} for each DHA search. According to the DHA introduced in Appendix B-D, the upper-bound and lower-bound of Ω_{QD} are given by

$$4.5\sqrt{M^{d_f-1}} \leq \Omega_{QD} \leq 22.5\sqrt{M^{d_f-1}}. \quad (12)$$

By contrast, the NCFE in the classical domain is on the order of $\Omega_{CD} = O(\log \sqrt{M^{d_f-1}})$. Let us denote the total NCFE of a single DHA search by Ω_{DHA} , which is given by

$$\Omega_{DHA} = \Omega_{QD} + \Omega_{CD} = O(\sqrt{M^{d_f-1}}). \quad (13)$$

Since there are M SCMA codewords for each message updating in the FNs, the total NCFE of the Q-MPA is given by

$$\Omega_{QMMPA} = O(M\Omega) = O(M^{(d_f+1)/2}). \quad (14)$$

Therefore, the NCFE of the Q-MPA is accelerated to be lower than the NCFE of the classical MPA given by $O(M^{d_f})$.

Due to the random search of the QSAs, the NCFE of a single quantum search is not constant. We define the average NCFE as the mean of Ω_{QMMPA} . To evaluate the NCFE of the Q-MPA more accurately, we use the average NCFE as the metric of characterizing the Q-MPA.

C. PROJMPA TECHNIQUE BASED Q-MPA

Although we have seen that the NCFE can be reduced by Q-MPA, it can still be further reduced by using other low-complexity classical techniques. Since the NCFE of the Q-MPA increases both with the codebook size M and with the degree of the FNs d_f , it can be further mitigated by reducing

TABLE 1: NCFE Comparison of Different Receivers

Receiver Type	Number of CFEs
Classical MPA	$O(M^{d_f})$
ProjMPA	$O(M_p^{d_f})$
Q-MPA	$O(\sqrt{M^{d_f+1}})$
ProjQMPA	$O(\sqrt{M_p^{d_f+1}})$

M or d_f . In this subsection, we hence further reduce the NCFE of the Q-MPA by invoking the classical ProjMPA technique of [12].

The SCMA codebook size can be viewed as the number of projections of the SCMA codebook over each non-zero element in the SCMA codewords [12]. By designing the SCMA codebook to have the minimum number of projections, the effective number of codebook entries is reduced from M to M_p ($M_p < M$). When a 16-point SCMA codebook is employed, the effective codebook size can be reduced to as few as 9 points [12]. Therefore, we can use the SCMA codebook having a reduced number of projections for reducing the NCFE of the Q-MPA. For simplicity, we refer to this technique as ProjQMPA.

In the ProjQMPA, the number of CF entries is reduced to $M_p^{d_f-1}$. As a result, the NCFE of each DH-QSA application in the Q-MPA becomes $O(M_p^{(d_f-1)/2})$ by using (13). Since there are M_p different possible codewords for each message in the FNs, the total NCFE of the ProjQMPA is given by $O(M_p^{(d_f+1)/2})$. Compared to the Q-MPA having the NCFE order of $O(M^{(d_f+1)/2})$, the ProjQMPA further reduces the NCFE. Furthermore, the theoretical NCFE of the Q-MPA based MUDs is given in Table I. Observe from Table I that the Q-MPA based receivers all exhibit lower NCFE than their classical counterparts. Thus, the Q-MPA can be beneficially amalgamated with the family of classical SCMA MUDs for further reducing the MUD complexity.

IV. THE QUANTUM-ASSISTED SPHERE DECODER BASED MPA MUD

In each iteration of the classical MPA, each message updating in the FNs requires M^{d_f} CFEs. To reduce the complexity of the classical MPA, the SD-MPA of [15] was proposed to find all solutions of 5 within a hypersphere by using a classical SD. Specifically, a CF is defined to measure the Euclidean distance between the received signal and the codeword combinations. Though significant complexity reduction has been achieved by the SD-MPA, the SD still requires CFEs of all codeword combinations. The SD is shown in Fig. 7. Thus, the NCFE of SD-MPA can be further reduced by using QSAs.

In this section, we present a QSD-MPA method along with its complexity analysis. Based on the classical SD-MPA, we first present a Quantum-assisted Sphere Decoder (QSD) by using the BBHT-QSA to accelerate the classical SD. Then, the classical MPA adopted in the QSD-MPA is introduced. The complexity analysis of the QSD-MPA is given at the end

of this section. Readers might like to refer to Appendix A and Appendix B for details of the classical SD-MPA and of the BBHT-QSA, respectively.

A. THE QSD-MPA

The QSD-MPA consists of two steps, namely, the QSD step and the classical MPA step. In the QSD, all the codeword combinations within a given hypersphere are chosen as legitimate codeword combinations. In the MPA, the legitimate codeword combinations are used for updating messages between the FNs and VNs on the factor graph. In this subsection, we describe the QSD step and the MPA step, respectively.

1) The QSD

In the QSD, all the legitimate codeword combinations within a given hypersphere have to be found. To achieve this, a CF is defined for quantifying the distance between each codeword combination and the received signal. A codeword combination at the n -th sub-carrier is denoted by $x_{[n]} = \{x_{l,n} : l \in \phi(n)\}$. At the n -th sub-carrier, the CF of the QSD is defined as the Euclidean distance between the codeword combination considered and the received signal, as

$$f(x_{[n]}) = \|y_n - \sum_{l \in \phi(n)} h_{l,n} x_{l,n}\|, \quad (15)$$

where y_n is the received signal. Given the radius of the hypersphere Δ , a codeword combination $x_{[n]}$ is legitimate if it satisfies

$$f(x_{[n]}) \leq \Delta. \quad (16)$$

For simplicity, a legitimate codeword combination is termed as a solution for the CF f . The number of solutions is denoted by S . When S is unknown *a priori*, the QSD finds all the solutions by using the following algorithm.

We assume that all the codeword combinations form an unsorted database with Q elements, where we have $Q = M^{d_f}$ at the beginning. An iterative quantum-assisted algorithm based on the BBHT-QSA is adopted to search through the database for all solutions. Specifically, the QSD iteratively employs the BBHT-QSA for finding the legitimate codeword combinations one by one. Since the BBHT-QSA can find a specific solution in an unsorted database even when the number of solutions is unknown *a priori*, the QSD can find all solutions in the database after a certain number of iterations. When a result $x_{[n]}^o$ is obtained by a specific BBHT-QSA iteration, the CF value $f(x_{[n]}^o)$ is computed in the classical domain to check, whether it is legitimate. If $f(x_{[n]}^o) \leq \Delta$, then $x_{[n]}^o$ will be output as a solution and it will be removed from the database. As a result, both the number of remaining solutions S and the number of remaining elements Q in the database are reduced by 1. However, if $f(x_{[n]}^o) > \Delta$, we define this occurrence as a failure of the BBHT-QSA. In this case, the number of remaining solutions S and the number of remaining elements Q in the database will not change. Furthermore, we derive a variable V to represent the number

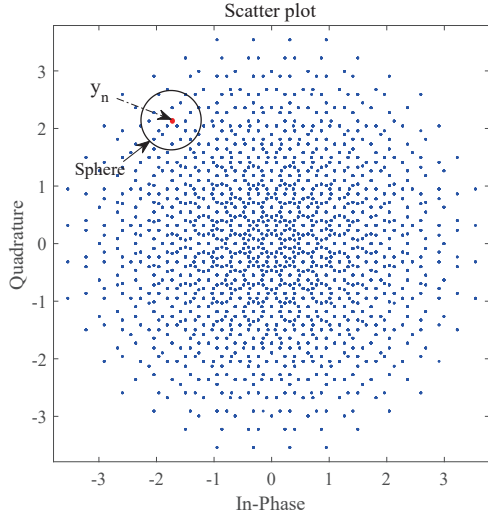


FIGURE 7: The diagram of sphere decoding. Each point represents a codeword combination at n -th sub-carrier. The legitimate codeword combinations are the points within the sphere.

of immediately consecutive failures. If failures occur twice in succession, we have $V > 1$. In this case, the QSD is believed to have obtained all solutions in the database. Therefore, the BBHT-QSA will be invoked iteratively until we reach $V > 1$. By using the above algorithm, all the legitimate codeword combinations and its CF values can be obtained. The QSD is summarized in the QSD step in Algorithm 3.

2) The MPA

In the MPA, the messages are iteratively updated between the FNs and VNs on the factor graph. However, only the legitimate codeword combinations obtained by the QSD will be used to update the messages. The message forwarded from the n -th FN to the k -th VN at the j -th iteration is denoted by $\mathcal{L}_{n \rightarrow k}^j(x_k)$, where x_k is the codeword of user k . Similarly, the message passed on from the k -th VN to the n -th FN at the j -th iteration is denoted by $\mathcal{L}_{n \leftarrow k}^j(x_k)$. If $x_{[n]}$ is a legitimate codeword combination, the FN utilizes the pre-stored CF value $f(x_{[n]})$ to update messages. Specifically, the message $\mathcal{L}_{n \rightarrow k}^j(x_k)$ is given by

$$\mathcal{L}_{n \rightarrow k}^j(x_k) = \max_{x_l: l \in \phi(n) \setminus k} \left\{ -\frac{1}{2\sigma^2} f^2(x_{[n]}) + \sum_{l \in \phi(n) \setminus k} \mathcal{L}_{n \leftarrow l}^{j-1}(x_l) \right\} \quad (17)$$

In the k -th VN, the message $\mathcal{L}_{n \leftarrow k}^j(x_k)$ is updated as

$$\mathcal{L}_{n \leftarrow k}^j(x_k) = \sum_{u \in \psi(k) \setminus n} \mathcal{L}_{u \rightarrow k}^j(x_k). \quad (18)$$

When the maximum number of MPA iterations is reached, the LLR of each coded bit is computed by using (27). The QSD-MPA is summarized in Algorithm 3.

Algorithm 3 QSD-MPA

- 1: **The QSD step:**
 $\Omega_{QSD} \leftarrow 0, \Omega_{QSD}^{QD} \leftarrow 0, \Omega_{QSD}^{CD} \leftarrow 0, V \leftarrow 0, Q \leftarrow M^{d_f}$ and set the sphere radius Δ .
- 2: **while** $V < 2$ **do**
- 3: Invoke the BBHT-QSA to search the Q possible codeword combinations. Obtaining a result $x_{[n]}^o, \Omega_{BBHT}^{QD}$, and Ω_{BBHT}^{CD} .
- 4: $\Omega_{QSD}^{QD} \leftarrow \Omega_{QSD}^{QD} + \Omega_{BBHT}^{QD}, \Omega_{QSD}^{CD} \leftarrow \Omega_{QSD}^{CD} + \Omega_{BBHT}^{CD}$.
- 5: **if** $f(x_{[n]}^o) > \Delta$ **then**
- 6: $V \leftarrow V + 1;$
- 7: **continue.**
- 8: **else**
- 9: Output $x_{[n]}^o$ as a legitimate codeword combination, store the value of $f(x_{[n]}^o)$ in a database, and delete $x_{[n]}^o$ from database;
- 10: $Q \leftarrow Q - 1, V \leftarrow 0$.
- 11: **end if**
- 12: **end while**
- 13: $\Omega_{QSD} \leftarrow \Omega_{QSD}^{QD} + \Omega_{QSD}^{CD}$.
- 14: **The MPA step:**
- 15: **for** $j \leq J$ **do**
- 16: Compute $\mathcal{L}_{n \rightarrow k}^j(x_k)$ as (17);
- 17: Compute $\mathcal{L}_{n \leftarrow k}^j(x_k)$ as (26).
- 18: **end for**
- 19: Compute $\lambda_1(k, r)$ as (27).

B. COMPLEXITY ANALYSIS OF THE QSD-MPA

We use the complexity of the classical SD-MPA as the benchmark. Since the QSD-MPA has a QSD step and a MPA step, the total complexity of the QSD-MPA is contributed to both by the QSD step and by the MPA step. We first consider the complexity of the MPA. According to the introduction of the QSD-MPA, the QSD-MPA and the SD-MPA have the same MPA step. Therefore, the complexity of the MPA step in the QSD-MPA is the same as in the SD-MPA. Moreover, the MPA step only adds a modest contribution to the overall complexity because of the low-complexity additions and multiplications.

Then, we consider the complexity of the QSD step. Again, we characterize the complexity of the QSD by the NCFE. Let us denote the total number of codeword combinations and the number of *legitimate* codeword combinations by Q and S , respectively. Since the classical SD-MPA finds the legitimate codeword combinations by exhaustively searching through all codeword combinations, the NCFE of the classical SD-MPA is Q , or equivalently, M^{d_f} . However, a lower NCFE can be imposed by the QSD-MPA, if S is less than a critical value. The total NCFE of the QSD is contributed to by the NCFE in the classical domain and that in the quantum domain. The NCFE in the quantum domain and that in the classical domain are determined by the number of Grover operator applications and by the number of CFEs, respectively.

Again we assume that one Grover operator application in the quantum domain is equivalent to one CFE in the classical domain. For simplicity, the NCFE in the quantum domain and that in the classical domain are denoted by Ω_{QSD}^{QD} and Ω_{QSD}^{CD} , respectively. Additionally, we use Ω_{QSD} to denote the overall NCFE of the QSD. As shown in Algorithm 3, Ω_{QSD}^{CD} and Ω_{QSD}^{QD} are obtained by considering all the NCFE contributions of each BBHT-QSA in the corresponding domain. In Theorem 2, we present the overall NCFE of the QSD step.

Theorem 2: The NCFE of the QSD is upper bounded by $\Omega_{QSD} = O(\sqrt{SM^{d_f-1}})$.

Proof 2: After a certain number of iterations of the BBHT-QSA, we denote the number of solutions remaining in the database by s ($0 < s \leq S$). In this case, the NCFE of the BBHT-QSA in the classical domain and the quantum domain by Ω_{BBHT}^{CD} and Ω_{BBHT}^{QD} , respectively. The NCFE Ω_{BBHT}^{QD} is upper bounded by

$$\Omega_{BBHT}^{QD} \leq 4.5\sqrt{\frac{Q}{s}}. \quad (19)$$

If $s = 0$, the BBHT-QSA requires $4.5\sqrt{Q_3}$ number of Grover operator applications. Since S is unknown *a priori*, the total NCFE of the QSD in the quantum domain is given by

$$\begin{aligned} \Omega_{QSD}^{QD} &\leq \sum_{s=0}^{S-1} 4.5\sqrt{\frac{Q-s}{S-s}} + O(\sqrt{Q}) \leq 4.5\sqrt{Q} \sum_{s=1}^S \frac{1}{\sqrt{s}} \\ &+ O(\sqrt{Q}), \end{aligned} \quad (20)$$

where the $O(\sqrt{Q})$ represents the NCFE of the failures of the BBHT-QSA. For the first item at the right side of (20), we have

$$4.5 \sum_{s=1}^S \frac{1}{\sqrt{s}} \leq 4.5 \sum_{s=1}^S 2(\sqrt{s} - \sqrt{s-1}) < 9\sqrt{S}. \quad (21)$$

According (20) and (21), the NCFE Ω_{QSD}^{QD} is $O(\sqrt{QS})$. Meanwhile, Ω_{BBHT}^{CD} is on the order of $\log_{\lambda}(\sqrt{Q})$. Note that $Q = M^{d_f-1}$. Thus, the NCFE of the QSD is given by

$$\Omega_{QSD} = \Omega_{QSD}^{CD} + \Omega_{QSD}^{QD} = O(\sqrt{SM^{d_f-1}}). \quad (22)$$

According to Theorem 2, the QSD has lower NCFE than the classical SD when S is lower than a critical value. Therefore, the QSD-MPA can be used for reducing the MUD complexity of SCMA.

According to the BBHT-QSA, each BBHT-QSA iteration may apply different number of Grover operator applications. To compare the NCFE of the QSD and the classical SD, we use the average NCFE denoted by $\tilde{\Omega}_{QSD}$.

V. NUMERICAL RESULTS

In this section, we evaluate the BER performance vs NCFE of the quantum-assisted MUD schemes.

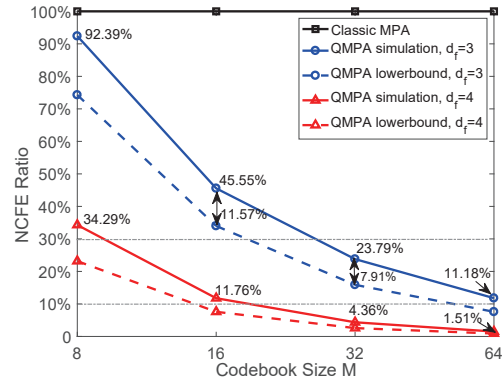


FIGURE 8: NCFE comparison between the classical MPA and the Q-MPA.

A. PERFORMANCE VS NCFE

Due to the random search of QSAs, simulations are required to evaluate the NCFE of our quantum-assisted MUDs. We proceed by our NCFE comparison between the quantum-assisted schemes and their classical counterparts in this section.

1) NCFE of the Q-MPA based MUDs

We first present our NCFE comparison between the classical MPA and the Q-MPA, where the average NCFE curves and theoretical NCFE lower-bound are portrayed. To compare the NCFE of the Q-MPA and the classical MPA, the NCFE ratio R_1 is defined as

$$R_1 = \frac{\tilde{\Omega}_{QMPA}}{M^{d_f}}, \quad (23)$$

where M^{d_f} is the NCFE of the classical MPA. The NCFE of the Q-MPA is then characterized by the NCFE ratio R_1 . In Fig. 8, the Y-axis represents the NCFE ratio R_1 and the X-axis the SCMA codebook size M . Our simulation results show that the Q-MPA exhibits 45.55% of the classical MPA NCFE for $M = 16$ and $d_f = 3$. Additionally, the average NCFE is only 11.57% higher than the theoretical NCFE lower-bound. Furthermore, the NCFE ratio curves decrease significantly as either M or d_f becomes higher. The lower NCFE imposed by the Q-MPA is because of applying fewer Grover operators in the quantum domain. Therefore, it can be concluded that the Q-MPA significantly reduces the NCFE of the classical MPA, especially when M and d_f are large.

Moreover, we compare the NCFE of the ProjQMPA and of the classical ProjMPA. In this case, the 16-point SCMA codebook is reduced to as few as 9 effective projections. When $d_f = 3$, our simulation results show that the ProjQMPA exhibits an average NCFE 80.42% compared to the classical ProjMPA. Thus, the complexity of the Q-MPA can be reduced by combining it with the classical ProjMPA technique.

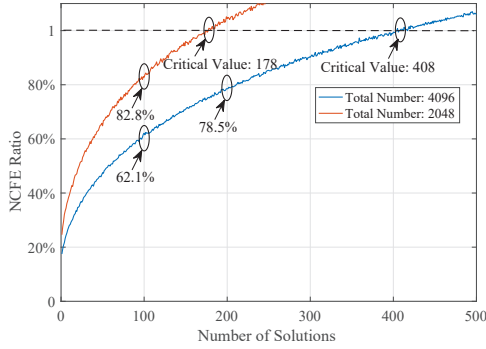


FIGURE 9: NCFE ratio vs the number of solutions.

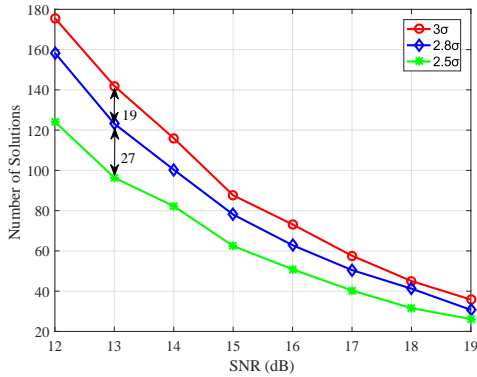


FIGURE 10: Average number of solutions vs SNR.

2) NCFE of the QSD-MPA MUD

In this subsection, the NCFE of the QSD-MPA is presented. We first characterize the NCFE of the QSD for different number of solutions. The NCFE ratio R_2 is defined as

$$R_2 = \frac{\tilde{\Omega}_{QSD}}{Q}, \quad (24)$$

where Q represents the NCFE of the classical exhaustive search, while $\tilde{\Omega}_{QSD}$ is the average NCFE of the QSD. The NCFE ratio curves of the database sizes on $Q = 4096$ and $Q = 2048$ are shown in Fig. 9. The simulation results demonstrate that the average NCFE monotonically increases with the number of solutions S . As a result, critical values exist on the NCFE ratio curves. Specifically, the critical value is 408 for $Q = 4096$ and 178 for $Q = 2048$. Explicitly, the MUD NCFE can be reduced if $R_2 \leq 1$. When $S = 200$ and $Q = 4096$, the NCFE ratio is $R_2 = 78.5\%$. Furthermore, the NCFE can be further reduced if S is a smaller number. The results are consistent with Theorem 2. Therefore, it can be concluded that the QSD reduces the NCFE of the classical SD, provided that S is below the critical value of Fig. 9.

In the following QSD-MPA based simulations, we adopt the 16-point SCMA codebook of [12] and the SCMA codeword structure shown in Fig. 3. As a result, the number of codeword combinations in each OFDM sub-carrier is $Q =$

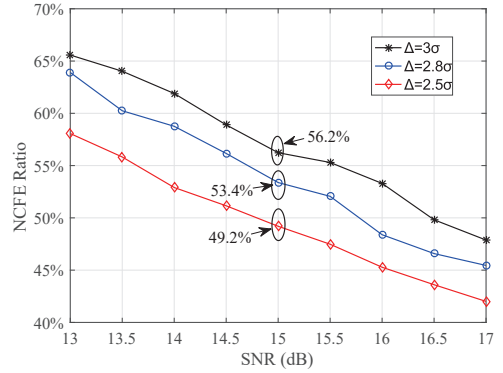


FIGURE 11: NCFE ratio vs SNR.

4096. Additionally, we set the radii of the QSD to $\Delta = 3\sigma$, $\Delta = 2.8\sigma$, and $\Delta = 2.5\sigma$, respectively. In the QSD, the number of legitimate codeword combinations may vary depending on the uncertainty of the received signal. We hereby give the simulation results that describe the average number of legitimate codeword combinations corresponding to different SNRs. As shown in Fig. 10, the results confirm that only a small number of codeword combinations are within the SNR-dependent hypersphere. Specifically, the number of solutions is only 142, when $\Delta = 3\sigma$ and the SNR is 13 dB. Moreover, the number of legitimate codeword combinations decreases with the radius of the hypersphere. According to the NCFE of the QSD, we can see that the average numbers are within the reduced-number region. Therefore, the QSD has a lower average NCFE than the classical SD.

To evaluate the average NCFE of the QSD-MPA more intuitively, we also demonstrate the NCFE vs SNRs. Since the MPA step of the QSD-MPA is the same as that of the classical MPA, we only consider the NCFE of the QSD and that of the classical SD in this simulation. We adopt the same parameters as in the simulations of Fig. 10. As shown in Fig. 11, the results show that the average NCFE of the QSD method is lower than that of the classical SD method. Specifically, when the SNR is 15 dB, the NCFE ratio is $R_2 = 56.7\%$, 53.4% , and 49.2% when $\Delta = 3\sigma$, $\Delta = 2.8\sigma$, and $\Delta = 2.5\sigma$, respectively. According to the simulation results of the QSD, the NCFE reduction is due to the sphere radius. Additionally, we can observe from the results that the NCFE ratio R_2 decreases as the SNR increases. In conclusion, the QSD-MPA detection scheme is capable of reducing the complexity of the classical MPA.

B. BER PERFORMANCE

In this section, we evaluate the BER performance of the quantum-assisted MUD schemes. Specifically, BER performance comparisons are provided between the Q-MPA and the classical MPA, the ProjQMPA and the classical ProjMPA, the QSD-MPA and the classical SD-MPA, respectively. In our simulations, the basic parameters are set as follows: the number of users is $K = 6$, the number of OFDM subcarriers

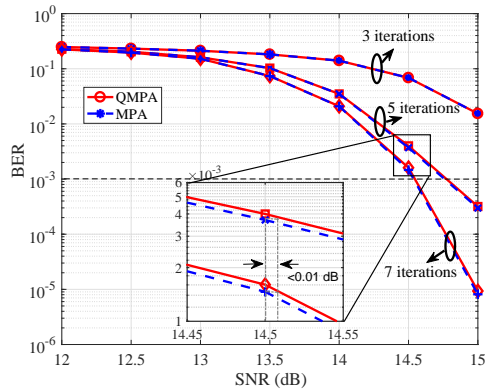


FIGURE 12: BER comparison between the classical MPA and the Q-MPA with turbo coding.

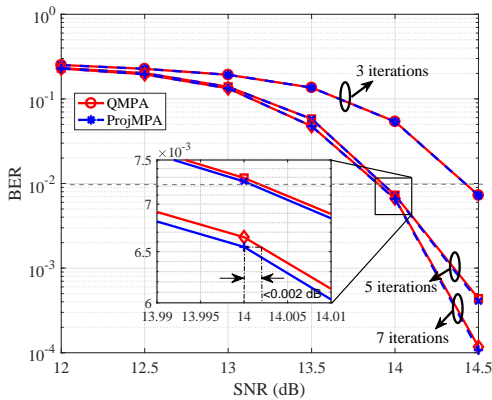


FIGURE 13: BER comparison between the classical ProjQMPA and the ProjMPA.

is $N = 4$, the degrees of the factor graph are $d_f = 3$ and $d_v = 2$. We adopt the SCMA codeword structure shown in Fig. 3. Additionally, we employ a turbo code having an interleaver length of 1024 bits and coding rate of $1/3$ as our channel code.

We first present the BER performance of the Q-MPA under 3, 5, and 7 MPA iterations. In this simulation, the 16-point SCMA codebook of [12] is employed. As shown in Fig. 12, the simulation curves of the Q-MPA basically overlap with MPA curves when the SNR is lower than 14 dB. Moreover, the Q-MPA suffers less than 0.01 dB SNR penalty at 14.5 dB. In conclusion, the Q-MPA suffers only negligible BER performance degradation, which is due to the fact that the success probability of DH-QSA is slightly less than 1.

In Fig. 13, we compare the BER performance of the ProjQMPA and of the classical ProjMPA. In this case, the number of SCMA codewords is reduced from 16 to 9 projections per non-zero OFDM sub-carrier by using the low number of projections technique. Compared to the classical MPA, the ProjQMPA achieves an even closer BER match with its classical counterpart. It can be concluded From Fig. 12 and Fig. 13 that the Q-MPA based SCMA MUDs only

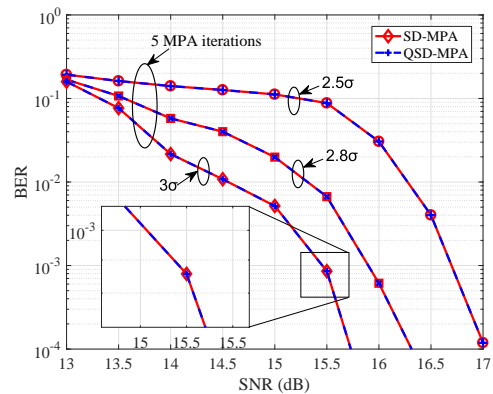


FIGURE 14: BER comparison between the QSD-MPA and the classical SD-MPA method.

suffer slight BER performance erosion.

Additionally, we give the BER performance of our QSD-MPA detection scheme. In this simulation, the 16-point SCMA codebook of [12] is adopted. As a result, the number of possible codeword combinations at a single sub-carrier is 4096. Moreover, we set the radii of the QSD as $\Delta = 2.5\sigma$, $\Delta = 2.8\sigma$, and $\Delta = 3\sigma$. As shown in Fig. 14, the BER curves of the QSD-MPA overlap with the corresponding BER curves of the classical SD-MPA. Therefore, the simulation results indicate that the QSD-MPA achieves a similar BER performance as the classical SD-MPA. The sustained BER performance of the QSD-MPA MUD is a benefit of the QSD finding all the legitimate codeword combinations within the hypersphere with near-certainty. Thus, we can conclude that the QSD-MPA achieves a similar BER as the classical SD-MPA at a lower complexity.

VI. CONCLUSIONS

In this paper, we conceived two quantum-assisted MUDs for SCMA by amalgamating QSAs and classical MUDs. Specifically, we conceived a Q-MPA scheme and a QSD-MPA scheme for reducing the complexity of the classical MPA scheme in two different ways. The Q-MPA was conceived by adopting the DH-QSA to accelerate the maximization search process of the classical MPA. By contrast, the QSD-MPA was conceived by invoking the QSD before the classical MPA. We characterize the complexity of both schemes by the total NCFE in the quantum domain and in the classical domain. Our simulation results show that the Q-MPA is capable of reducing the NCFE substantially at the cost of a negligible BER performance loss. Furthermore, the NCFE can be further reduced by using low-complexity classical techniques for shrinking the search space of the Q-MPA. Additionally, our simulation results show that the QSD-MPA can reduce the NCFE for MUD without BER performance degradation compared to the classical SD-MPA scheme. Therefore, the Q-MPA and the QSD-MPA can be used for low-complexity SCMA detection, albeit it still requires substantial research on QSAs.

APPENDIX A THE CLASSICAL MPA AND THE CLASSICAL SD-MPA

In this Appendix, we first describe the classical MPA. Then, the classical SD-MPA MUD is introduced.

A.1 CLASSICAL MPA

The classical MPA can be divided into message updating step in the FNs, message updating step in the VNs, and the LLR computation step. For the first two steps, messages are iteratively updated between the FNs and VNs on the factor graph. When the SCMA codeword of user k is x_k , the message sent from the n -th FN to the k -th VN at the j -th iteration is denoted by $\mathcal{L}_{n \rightarrow k}^j(x_k)$. Similarly, the message sent from the k -th VN to the n -th FN at the j -th iteration is denoted by $\mathcal{L}_{n \leftarrow k}^j(x_k)$. When $j = 0$, all the initial messages are set to 0. In the n -th FN, the message $\mathcal{L}_{n \rightarrow k}^j(x_k)$ is updated as

$$\mathcal{L}_{n \rightarrow k}^j(x_k) = \max_{x_l: l \in \phi(n) \setminus k} \left\{ -\frac{1}{2\sigma^2} \|y_n - \sum_{l \in \phi(n)} h_{l,n} x_{l,n}\|^2 + \sum_{l \in \phi(n) \setminus k} \mathcal{L}_{n \leftarrow l}^{j-1}(x_l) \right\}, \tag{25}$$

where y_n is the received signal at the n -th sub-carrier, and $\phi(n) \setminus k$ represents the neighbor VNs excluding the k -th VN. In the k -th VN, the message $\mathcal{L}_{n \leftarrow k}^j(x_k)$ is updated as

$$\mathcal{L}_{n \leftarrow k}^j(x_k) = \sum_{u \in \psi(k) \setminus n} \mathcal{L}_{u \rightarrow k}^j(x_k). \tag{26}$$

The maximum number of MPA iterations is denoted by J and when $j = J$ is reached, the LLR computation step is executed. Let us define $\mathcal{X}^1(c_{k,r}) = \{c_{k,r} = 1 | x_k \in \mathcal{X}\}$ and $\mathcal{X}^0(c_{k,r}) = \{c_{k,r} = 0 | x_k \in \mathcal{X}\}$. By using the results on the factor graph, the extrinsic information of (2) is obtained by

$$\lambda_1(c_{k,r}) = \max_{x_k \in \mathcal{X}^1(c_{k,r})} L(x_k) - \max_{x_k \in \mathcal{X}^0(c_{k,r})} L(x_k) - \lambda_2^p(c_{k,r}), \tag{27}$$

where

$$L(x_k) = \sum_{n \in \psi(k)} \mathcal{L}_{n \rightarrow k}^J(x_k). \tag{28}$$

In the SCMA detector, the LLR of each coded bit can thus be obtained.

In Eq.(25), the message updating operation requires searching through the codeword combinations, which results in an exponentially increasing complexity order of $O(M^{d_f})$. However, (26) will introduce only a modest complexity because of the simple multiplications and additions. Thus, the message updating in the FNs imposes the highest complexity contribution of the classical MPA.

A.2 CLASSICAL SPHERE DECODER BASED MPA

The classical SD-MPA was proposed in [15] for reducing the complexity of the classical MPA. The SD-MPA is divided

Algorithm 4 BBHT Quantum Search Algorithm

- 1: Set value $\delta_2, \lambda \leftarrow 6/5, m \leftarrow 1, \Omega_{BBHT} \leftarrow 0, \Omega_{BBHT}^{CD} \leftarrow 0, \Omega_{BBHT}^{QD} \leftarrow 0$.
- 2: Initiate the state of quantum system in $|\varphi\rangle$ as (30).
- 3: Uniformly pick $\omega \in \{1, 2, \dots, m\}$.
- 4: Apply ω number of Grover operators on the system, obtain the final state $|x_o\rangle = \mathcal{G}^\omega |x\rangle$.
- 5: Measure the final state $|x_o\rangle$ and obtain the output $|q\rangle$.
- 6: Compute $f_2(q)$ in the classical domain.
- 7: $\Omega_{BBHT}^{QD} \leftarrow \Omega_{BBHT}^{QD} + \omega, \Omega_{BBHT}^{CD} = \Omega_{BBHT}^{CD} + 1$.
- 8: **if** $f_2(q) = \delta_2$ or $\Omega_{BBHT}^{QD} \geq 4.5\sqrt{Q_2}$ **then**
- 9: $x_s \leftarrow q$, output x_s and exit.
- 10: **else**
- 11: $m \leftarrow \lfloor \lambda m, \sqrt{Q_2} \rfloor$, go to step 3.
- 12: **end if**
- 13: $\Omega_{BBHT} \leftarrow \Omega_{BBHT}^{QD} + \Omega_{BBHT}^{CD}$.

into the SD step and the MPA step. In the SD step, only the codeword combinations falling within a given hypersphere are selected as legitimate codeword combinations. The codeword combination in the n -th sub-carrier is denoted by $x_{[n]} = \{x_{l,n} : l \in \phi(n)\}$. Codeword combination $x_{[n]}$ is legitimate, if it satisfies

$$\|y_n - \sum_{l \in \phi(n)} h_{l,n} x_{l,n}\| \leq \Delta, \tag{29}$$

where y_n is the signal received by the n -th sub-carrier, while Δ is the radius of the given hypersphere. The left side of (29) is defined as the CF for the SD step. By computing and pre-storing the CF values of all the codeword combinations, the legitimate codeword combinations are obtained. In the MPA, the CF values of legitimate codeword combinations are used for updating the messages in the FNs.

Since there are a total M^{d_f} of codeword combinations at each sub-carrier, the SD requires computing and pre-storing M^{d_f} CF values. In the MPA step, only legitimate codeword combinations are considered. The SD-MPA has a significantly lower complexity than the classical MPA. Furthermore, a BER vs complexity trade-off can be struck by dynamically changing the sphere radius Δ according to the noise power σ .

APPENDIX B QUANTUM COMPUTING AND QUANTUM SEARCH ALGORITHMS

In this Appendix, we provide a rudimentary introduction to QSAs. First, the basis of quantum computing is introduced. Then, we describe three popular QSAs, namely, the Grover’s QSA, the BBHT-QSA, and the DH-QSA.

B.1 BASIS OF QUANTUM COMPUTING

In quantum computing, the basic information storage unit is *qubit* [17]. A single qubit is in a superposed state denoted by $|\rho\rangle$ that can be expressed as $|\rho\rangle = \alpha |0\rangle + \beta |1\rangle$, where $|\alpha|^2 + |\beta|^2 = 1$ and $\alpha, \beta \in \mathbb{C}$. Once a *measurement*

or *observation* of a qubit is performed, the quantum state $|\rho\rangle$ will collapse into $|0\rangle$ with probability $|\alpha|^2$ or $|1\rangle$ with probability $|\beta|^2$. To store more information in qubits, large quantum systems can be created by using the *entanglement* of qubits. Let us consider for example a quantum system relying on a pair of two entangled qubits $|\rho_1\rangle$ and $|\rho_2\rangle$. Then the state of the quantum system can be expressed as $|\rho_1\rangle|\rho_2\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$, where we have $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. Furthermore, the quantum state of a quantum system is manipulated by unitary quantum *operators* U , where $U^{-1} = U^\dagger$ [17]. By designing appropriate QSAs that are represented by unitary operators, the quantum state can be beneficially manipulated to achieve specific tasks.

B.2 GROVER'S QUANTUM SEARCH ALGORITHM

Given a value δ_1 and an unsorted database having Q_1 elements that represent the inputs of a CF f_1 , Grover's QSA finds an element q_s in the database that satisfies $f_1(q_s) = \delta_1$. The specific element q_s that satisfies $f_1(q_s) = \delta_1$ is termed as a solution. Provided that the number of solutions is known *a priori*, Grover's QSA succeeds in finding the specific entries constituting the solutions with a probability of almost 100%.

Grover's QSA prepares a quantum system that has $\log_2 Q_1$ qubits in $|0\rangle$ state. The Hadamard gate H is applied to initiate the system with an equiprobable state $|\varphi\rangle$. The resultant state $|\varphi\rangle$ is given by

$$|\varphi\rangle = H|0\rangle = \sum_{q=0}^{Q_1-1} |q\rangle, \quad (30)$$

where computational basis state $|q\rangle$ represents an element q in the database. Then, Grover operator $\mathcal{G} = HP_0HO$ is applied repeatedly to $|\varphi\rangle$, where $P_0 = 2|0\rangle\langle 0| - I$ represents a conditional phase shift operator associated with every computational basis state except for state $|0\rangle$ which is subjected to a phase shift of -1 , and O is the so-called Oracle [17] [18]. To elaborate a little further, the Oracle O evaluates f_1 for all inputs in parallel and recognizes the solutions sought in the database. If we have $f_1(q) = \delta_1$, the Oracle maps $|q\rangle$ to $-|q\rangle$, otherwise $|q\rangle$ remains unchanged. Since the complexity of the Oracle depends on the specific application [17], we assume that each application of the Grover operator corresponds to one CFE [18] [21]. Assuming that the number of solutions is S_1 , the optimal number of applying Grover operators to $|\varphi\rangle$ is $\Omega_{opt} = \lfloor \pi/4\sqrt{Q_1/S_1} \rfloor$. It follows that the probability of successfully finding a solution is $P_{opt} = \sin^2[(2\Omega_{opt} + 1)\theta]$, where $\theta = \arcsin(\sqrt{S_1/Q_1})$ [18].

B.3 BBHT QUANTUM SEARCH ALGORITHM

Given a value δ_2 and an unsorted database having Q_2 elements that represent the legitimate inputs of a CF f_2 , an element x_s that satisfies $f_2(x_s) = \delta_2$ is termed as a solution. When the number of solutions is unknown *a priori*, the BBHT-QSA is capable of finding a solution with

Algorithm 5 Durr-Hoyer Quantum Search Algorithm

- 1: Set $\Omega \leftarrow 0$. Randomly choose $q_0 \in \{0, 1, \dots, Q_3 - 1\}$, set threshold $\delta_3 \leftarrow f_3(q_0)$, $x_{max} \leftarrow q_0$.
- 2: The BBHT-QSA is employed to find a solution x_s that satisfies $f_3(x_s) > \delta_3$. Obtaining a solution x_s and Ω_{BBHT} from the BBHT-QSA.
- 3: $\Omega_{DHA} \leftarrow \Omega_{DHA} + \Omega_{BBHT}$.
- 4: **if** $f_3(x_s) \leq \delta_3$ or $\Omega_{DHA}^{QD} \geq 22.5\sqrt{Q_3}$ **then**
- 5: Output x_{max} and exit.
- 6: **else**
- 7: Set threshold $\delta_3 \leftarrow f_3(x_s)$, $x_{max} \leftarrow x_s$, go to step 2.
- 8: **end if**

a probability of almost 100%. As shown in Algorithm 4, the BBHT-QSA applies a pseudo-random number ω of Grover operators \mathcal{G} to the initial state $|\varphi\rangle$ in (30). The process of applying Grover operator will be repeated until a legitimate solution x_s is found. Let us define the number of CFEs in the quantum domain by Ω_{BBHT}^{QD} and the number of CFEs in the classical domain by Ω_{BBHT}^{CD} . The overall complexity of the BBHT-QSA is given by

$$\Omega_{BBHT} = \Omega_{BBHT}^{CD} + \Omega_{BBHT}^{QD}. \quad (31)$$

The number of solutions is denoted by S_2 . When $\lambda = 6/5$ and $S_2 > 0$, at most $\Omega_{QD}^{max} = 4.5\sqrt{Q_2/S_2}$ Grover operators are used by the BBHT-QSA before finding a solution [19], otherwise $S_2 = 0$.

B.4 DURR-HOYER QUANTUM SEARCH ALGORITHM

Given an unsorted database having Q_3 elements that represent the legitimate inputs of a CF f_3 , the DH-QSA succeeds in finding the specific entry denoted by x_{max} that maximizes f_3 with a probability of almost 100% [20] [24]. The DH-QSA randomly selects a threshold δ_3 at the beginning. Then, the BBHT-QSA is employed for finding an element having a higher CF value in the database. To achieve this, the Oracle in the BBHT-QSA only marks an element $|x\rangle$ that satisfies $f_3(x) > \delta_3$ as a solution. Once a solution x_s is obtained by the BBHT-QSA, the threshold δ_3 and x_{max} will be updated as $\delta_3 \leftarrow f_3(x_s)$ and $x_{max} \leftarrow x_s$, respectively. In DH-QSA, the BBHT-QSA will be invoked repeatedly until no solution x that satisfies $f_3(x) > \delta_3$ can be found. Let us denote the total number of applying Grover operator in the DH-QSA by Ω_{DHA}^{QD} . The DH-QSA succeeds in finding x_{max} by using at least $4.5\sqrt{Q_3}$ Grover operators, and at most $22.5\sqrt{Q_3}$ Grover operators. Thus, the lower-bound and upper-bound of Ω_{DHA}^{QD} are given by

$$4.5\sqrt{Q_3} \leq \Omega_{DHA}^{QD} \leq 22.5\sqrt{Q_3}. \quad (32)$$

Additionally, the complexity Ω_{DHA}^{CD} in the classical domain is on the order of $O(\log Q_3)$. Thus, the complexity of the DH-QSA is $O(\sqrt{Q_3})$. The DH-QSA is detailed in Algorithm 5.

REFERENCES

- [1] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74-81, Sept. 2015.
- [2] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185-191, Feb. 2017.
- [3] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017.
- [4] Y. Liu, L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1474-1487, March 2018.
- [5] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 332-336, Sept. 2013.
- [6] R. Razavi, M. AL-Imari, M. A. Imran, R. Hoshyar, and D. Chen, "On receiver design for uplink low density signature OFDM (LDS-OFDM)," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3499-3508, Nov. 2012.
- [7] F. R. Kschischang, B. J. Frey, and H. -A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001.
- [8] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 2918-2923, June 2015.
- [9] J. Dai, K. Niu, C. Dong, and J. Lin, "Improved message passing algorithms for sparse code multiple access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9986-9999, Nov. 2017.
- [10] H. Mu, Z. Ma, M. Alhaji, P. Fan, and D. Chen, "A fixed low complexity message pass algorithm detector for up-link SCMA system," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 585-588, Dec. 2015.
- [11] Y. Du, B. Dong, Z. Chen, J. Fang, P. Gao, and Z. Liu, "Low-complexity detector in sparse code multiple access systems," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1812-1815, Sept. 2016.
- [12] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for SCMA detection," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6, Dec. 2015.
- [13] L. Li, Z. Ma, P. Z. Fan, and L. Hanzo, "High-dimensional codebook design for the SCMA down link," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10118-10122, Oct. 2018.
- [14] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 621-634, Feb. 2017.
- [15] L. Yang, X. Ma, and Y. Siu, "Low complexity MPA detector based on sphere decoding for SCMA," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1855-1858, Aug. 2017.
- [16] P. Botsinis, D. Alanis, Z. Babar, H. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, "Quantum algorithms for wireless communications," *IEEE Communications Tutorials & Survey*, to be published. Doi: 10.1109/COMST.2018.2882385.
- [17] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge, U.K: Cambridge University Press, 2002.
- [18] L. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annual ACM Symposium on Theory of Computing*, ACM, May 1996.
- [19] M. Boyer, G. Brassard, P. Hoyer, and A. Tapp, "Tight bounds on quantum searching," *Fortschritte der Physik: Progress of Physics*, vol. 46, no. 4-5, pp. 493-505, April 1998.
- [20] C. Durr and P. Hoyer, "A quantum algorithm for finding the minimum," *arXiv preprint quant-ph/9607014v2*, Jan. 1999.
- [21] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1853-1888, May 2012.
- [22] S. Imre and F. Balazs, "Non-coherent multi-user detection based on quantum search," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 283-287 vol. 1, Aug. 2002.
- [23] P. Botsinis, S. X. Ng, and L. Hanzo, "Fixed-complexity quantum-assisted multi-user detection for CDMA and SDMA," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 990-1000, March 2014.
- [24] P. Botsinis, D. Alanis, Z. Babar, S. X. Ng, and L. Hanzo, "Coherent versus non-coherent quantum-assisted solutions in wireless systems," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 144-153, Dec. 2017.
- [25] P. Botsinis, D. Alanis, S. X. Ng, and L. Hanzo, "Low-complexity soft-output quantum-assisted multiuser detection for direct-sequence spreading and slow subcarrier-hopping aided SDMA-OFDM systems," *IEEE Access*, vol. 2, pp. 451-472, May 2014.
- [26] P. Botsinis, D. Alanis, Z. Babar, S. X. Ng, and L. Hanzo, "Iterative quantum-assisted multi-user detection for multi-carrier interleaved division multiple access systems," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3713-3727, Oct. 2015.
- [27] P. Botsinis, Y. Huo, D. Alanis, Z. Babar, S. X. Ng, and L. Hanzo, "Quantum search-aided multi-user detection of IDMA-assisted multi-layered video streaming," *IEEE Access*, vol. 5, pp. 23233-23255, July 2017.



WENJING YE received the B.S. degree in school of optical and electronic information from Huazhong University of Science and Technology, Wuhan, China, in 2016. He is currently pursuing the master's degree in electronic engineering from Tsinghua University, Beijing, China. His research interests include quantum computing and wireless communications.



WEI CHEN (S'05-M'07-SM'13) received his BS and Ph.D. degrees (both with the highest honors) from Tsinghua University in 2002 and 2007. From 2005 until 2007, he was also a visiting PhD student at the Hong Kong University of Science & Technology. Since 2007, he has been on the faculty at Tsinghua University, where he is a tenured full Professor, the director of degree office of Tsinghua University, and a university council member. During 2014 to 2016, he served as a deputy head of Department of Electronic Engineering. He visited Princeton University, Telecom ParisTech, and University of Southampton in 2016, 2014, and 2010 respectively. His research interests are in the areas of communication theory, stochastic optimization, and statistical learning. Dr. Chen is a Cheung Kong young scholar and a member of national program for special support for eminent professionals, also known as 10,000- talent program. He has also been supported by the national 973 youth project, the NSFC excellent young investigator project, the new century talent program of Ministry of Education, and the Beijing nova program. He received the IEEE Marconi Prize Paper Award and the IEEE Comsoc Asia Pacific Board Best Young Researcher Award in 2009 and 2011, respectively. He serves as an editor of the *IEEE TRANSACTIONS ON COMMUNICATIONS* and has served as a TPC co-chair of IEEE VTC-Spring, 2011, as well as, symposium co-chairs for IEEE ICC and Globecom. Dr. Chen is a recipient of the National May 1st Labor Medal and the China Youth May 4th Medal.



XIN GUO received the B.S. and Ph.D. degrees in computer science from the University of Science and Technology of China, Anhui, in 2003 and 2008, respectively. She is currently a Researcher with the Sony China Research Laboratory, Beijing. Her research interests include wireless communications and networks.



CHEN SUN (S'02-M'05-SM'12) received the PhD degree in electrical engineering from Nanyang Technological University, Singapore, in 2005. From August 2004 to May 2008, he was a researcher at ATR Wave Engineering Laboratories, Japan working on adaptive beamforming and direction finding algorithms of parasitic array antennas as well as theoretical analysis of cooperative wireless networks. In June 2008, he joined the National Institute of Information and Communications Technology (NICT), Japan, as expert researcher working on distributed sensing and dynamic spectrum access in TV white space. He is currently director of Wireless Network Research Department at Sony China Research Lab, Beijing. He served as the technical editor of the IEEE 1900.6 Standard in 2011 and the rapporteur of the European Telecommunications Standards Institute Reconfigurable Radio Systems EN 301 144 from 2013 to 2015. He received the IEEE Standards Association Working Group Chair Award for leadership in 2011 and the IEEE 802.11af Outstanding Contributions Award in 2014. He currently serves as the technical editor of the IEEE 802.19.1a Working Group.



LAJOS HANZO FEng, F'04, FIET, Fellow of EURASIP, received his 5-year degree in electronics in 1976 and his doctorate in 1983 from the Technical University of Budapest. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 119 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1800+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. He is a former Editor-in-Chief of the IEEE Press and a former Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to <http://www-mobile.ecs.soton.ac.uk>.

...