IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 April 2024, accepted 7 June 2024, date of publication 12 June 2024, date of current version 20 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3413084

RESEARCH ARTICLE

Quantum Speedup for Multiuser Detection With Optimized Parameters in Grover Adaptive Search

MASAYA NORIMOTO, (Graduate Student Member, IEEE), TAKU MIKURIYA, (Graduate Student Member, IEEE), AND NAOKI ISHIKAWA, (Senior Member, IEEE)
Faculty of Engineering, Yokohama National University, Yokohama 240-8501, Japan
Corresponding author: Naoki Ishikawa (ishikawa-naoki-fr@ynu.ac.jp)

This work was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 22H01484.

ABSTRACT Maximum-likelihood multiuser detection incurs a large computational complexity, and its low-complexity detection scheme suffers from a performance loss, where this tradeoff is inevitable and inherent in a classical computer. In this paper, we use the Grover adaptive search (GAS) to break the tradeoff, which is a quantum exhaustive search algorithm guaranteed to obtain the optimal solution, achieving a quadratic speedup. Specifically, we design two specific parameters of GAS to achieve the optimal performance with a reduced complexity: the initial threshold and the number of Grover rotations. The initial threshold of GAS can be optimized using a solution of semi-definite programming, and it is possible to calculate the distribution of the number of solutions smaller than the initial threshold in advance, which depends on instantaneous channel coefficients. In addition, we analyze the number of quantum gates required for GAS and show that the gate count can be reduced by bypassing the higher-order terms in the objective function, leading to a reduced circuit runtime. Our analysis and simulation results demonstrate that the proposed approach achieves the same performance as the optimal maximum-likelihood detection while reducing the query complexity of GAS, implying that the large constant overhead of quadratic speedup can be further reduced.

INDEX TERMS Grover adaptive search, multiple-input multiple-output, multiuser detection, quadratic speedup, semi-definite programming.

I. INTRODUCTION

Efficient multiuser detection schemes have been proposed in the context of both multiple-input multiple-output (MIMO) and non-orthogonal multiple access (NOMA) systems [1], [2], [3]. The optimal one is the maximum-likelihood detection (MLD), and the suboptimal ones are classified into linear and nonlinear detectors. Specifically, linear detectors include zero-forcing (ZF) and minimum mean square error (MMSE) detectors, and nonlinear detectors include successive interference cancellation, semidefinite programming (SDP) [4], lattice reduction, metaheuristics, machine learning, etc. In any case, the optimal detectors incur a large computational complexity, and the suboptimal ones suffer from a performance loss. This fundamental tradeoff cannot be overcome as long as relying on classical computation.

After the eventual end of Moore's law,[1] the performance of classical computation is expected to be saturated as the miniaturization of transistors reaches its physical limit, and the development of quantum computation is crucial to overcome this limitation. Assuming fault-tolerant quantum computation

The associate editor coordinating the review of this manuscript and approving it for publication was Pietro Savazzi.

[1]Note that the international roadmap for devices and systems has stated in 2022 that it is not surprising that Moore's law will continue for the next 10 years [5].

(FTQC), quantum algorithms with the promise of quantum speedup include Shor's algorithm for factoring [6], Grover's algorithm for searching an unsorted database [7], and the Harrow-Hassidim-Lloyd algorithm for solving linear systems of equations [8]. Grover's algorithm can find a single solution from a database of size N in $O(\sqrt{N})$ queries, while the classical exhaustive search requires $O(N)$ time, which is referred to as a quadratic speedup. Note that the well-known quantum approximate optimization algorithm (QAOA) [9] is a heuristic algorithm that assumes a noisy intermediate-scale quantum (NISQ) device and coarsely approximates a unitary time evolution operator corresponding to a time-dependent Hamiltonian by a product of unitary operators. Unfortunately, it has been proven that QAOA is unlikely to outperform classical computation as long as noise exists [10]. Another well-known approach, quantum annealing (QA) [11], is an analog counterpart of QAOA that tries to find the ground state of the time-dependent Hamiltonian similar to QAOA, but it also suffers from the same limitation [10]. In contrast to QAOA [9] and QA [11], Grover adaptive search (GAS) [12] is a quantum exhaustive algorithm that guarantees the optimality of an obtained solution with a quadratic speedup, because it can randomly sample solutions smaller than a certain threshold with a certain probability, where the tricky part is how to design the threshold and the number of Grover rotations depending on the number of solutions.

To break the fundamental tradeoff between complexity and performance, pioneering researchers have attempted to apply the quantum computing techniques to wireless communications [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. This is also because there are fundamental mathematical similarities between quantum computation and wireless communication. For example, the Hadamard gate creating an equal superposition state in quantum computation is equivalent to the spreading code in code-division multiple access (CDMA) systems, where multiuser symbols are superimposed with an equal weight. Another example is a design of noncoherent space-time codes [13] extending the quantum error correction. In the application of QAOA to MLD [14], [15], it has been confirmed that the angle parameters of QAOA depend on the statistical behavior of wireless channel coefficients and signal-to-noise ratio (SNR). In addition, the applications of QA to MLD have been studied in large-scale or massive MIMO [16], [17] and NOMA [18] systems. In the context of GAS, an early-stopping strategy [19] for CDMA is proposed to reduce the number of Grover rotations. In the authors' previous studies [20], [21], the initial threshold of GAS is designed using the statistical property of objective function fluctuating depending on noise and channel coefficients.

Against this background, we apply GAS to multiuser detection, aiming to break the fundamental tradeoff between complexity and performance. The major contributions of this paper are threefold.

1) We analyze the number of quantum gates required for GAS, where two approaches are considered: one is to map the detection problem to a higher-order unconstrained binary optimization (HUBO) problem directly, and the other is to map the problem to a quadratic unconstrained binary optimization (QUBO) problem with postprocessing.
2) We design the initial threshold by a rough estimate of SDP, and the number of Grover rotations is then designed using the distribution of the number of solutions smaller than the threshold. The distribution is obtained by Monte Carlo simulations in advance, offline, and can be used for arbitrary channel coefficients and SNR, assuming perfect channel estimation.
3) We verify that a quadratic speedup can be achieved even if the minimum number of Grover rotations is increased, which is supported by our analysis and simulations. The proposed approach is shown to be effective even if the transmit power from each user terminal (UT) is different and it achieves the same performance as the optimal MLD while reducing the query complexity of GAS to the minimum in all the considered approaches.

The idea of limiting the number of solutions with a strict initial threshold is similar to the classic sphere decoding [23]. In the sphere decoding, the detection complexity is reduced at the sacrifice of a slight performance loss. This is inevitable as long as classical computation is used, however, if we rely on quantum computation as in this paper, we can achieve the optimal performance while reducing the complexity, which may be a breakthrough in wireless communications.

The remainder of this paper is organized as follows. Section II reviews the GAS algorithm, and Section III describes the system model assumed in this paper. Section IV describes the conventional quantum-assisted multiuser detection, and Section V analyzes the number of quantum gates required for GAS. Section VI designs the initial threshold and the number of Grover rotations with an analysis of quadratic speedup. Section VII presents simulation results, and Section VIII concludes the paper.

II. GROVER ADAPTIVE SEARCH

Grover's algorithm [7] is a general framework for searching for a solution from an unsorted database, and it has been extended to the case where the number of solutions is unknown [24], to the case where the minimum value is searched [25], and to global optimization [26]. In these conventional studies [7], [19], [24], [25], [26], the Grover oracle that marks the solutions of interest has been regarded as a black box, and the studies have been conducted on the premise that the oracle can be constructed efficiently in the long term. This issue was solved by Gilliam et al. in [12], which is based on the idea of the quantum adder [27], and it constructs a quantum circuit that corresponds to a QUBO or HUBO problem of

$$\begin{aligned} \min \quad & E(\mathbf{b}) \\ \text{s.t.} \quad & \mathbf{b} \in \mathbb{B}^n, \end{aligned} \tag{1}$$

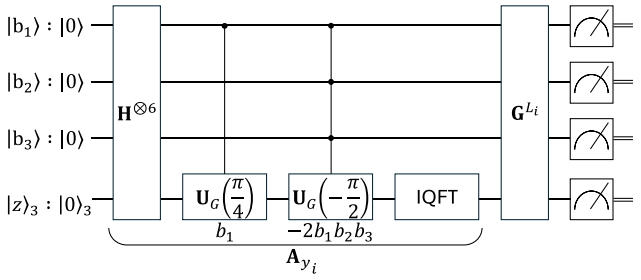


FIGURE 1. A quantum circuit of GAS corresponding to $b_1 - 2\,b_1b_2b_3$.

where $\mathbf{b} = [b_1 \ \cdots \ b_n]^{\mathrm{T}} \in \mathbb{B}^n$ is a binary vector, and $E(\mathbf{b})$ is an objective function that may contain higher-order terms such as $E(\mathbf{b}) = b_1 - 2\,b_1b_2b_3$. In the original GAS [12], the objective function $E(\mathbf{b})$ is assumed to have integer coefficients, and in the case of real coefficients, it is mentioned that the probability distribution follows the Fejér distribution, a superposition of approximated integers close to the original real value. In this state, the algorithm will not work correctly and the solution cannot be obtained. Then, a modification [20] of the algorithm was proposed to deal with real coefficients while causing a slight complexity overhead in classical computation.

The quantum circuit of GAS [12] is composed of a state preparation operator $\mathbf{A}_{y_i}$ and a Grover operator $\mathbf{G}^{L_i}$ as shown in Fig. 1. Here, $n = 3$ qubits are used to represent the binary vector $\mathbf{b} \in \mathbb{B}^n$, and $m = 3$ qubits are used for output register $|z\rangle$ to represent the objective function value $E(\mathbf{b})$. The variable y_i is a threshold for ith trial and it is a constant term subtracted from the objective function $E(\mathbf{b})$, which is adjusted for each trial in an adaptive manner, limiting the number of solutions corresponding to $E(\mathbf{b}) - y_i < 0$. The Grover operator $\mathbf{G}$ is applied L_i times, and the design of L_i depends on the number of solutions, which is unknown in advance.

The preparation operator $\mathbf{A}_{y_i}$ corresponds to a quantum circuit that calculates the objective function value $E(\mathbf{b})$ [12], where a technique similar to the quantum adder relying on the inverse quantum Fourier transform (IQFT) [27] is used. As shown in Fig. 1, each term having a coefficient $a \in \mathbb{R}$ corresponds to [12]

$$\mathbf{U}_G(\theta) = \mathbf{R}(2^{m-1}\theta) \otimes \mathbf{R}(2^{m-2}\theta) \otimes \cdots \otimes \mathbf{R}(2^0\theta) \tag{2}$$

with a coefficient $\theta = 2\pi a/2^m$, a phase gate $\mathbf{R}(\theta) = \mathrm{diag}(1, e^{\mathrm{j}\theta})$, and the imaginary number j. A coefficient multiplied by a binary variable is represented as a controlled-R (CR) gate, where the phase gate is controlled by the corresponding qubit. Similarly, a coefficient multiplied by k binary variables is represented as a $\mathrm{C}^k\mathrm{R}$ gate, which is controlled by k qubits.

The Grover operator $\mathbf{G}$ is composed of an oracle operator $\mathbf{O}$ and a Grover diffusion operator $\mathbf{D}$ as $\mathbf{G} = \mathbf{A}_{y_i}\mathbf{D}\mathbf{A}_{y_i}^{\mathrm{H}}\mathbf{O}$ [12]. The oracle $\mathbf{O}$ is an operator that flips the phase of states of interest that correspond to solutions. Given a threshold y_i, since GAS performs global minimization, the states of interest are those that satisfy $E(\mathbf{b}) < y_i$. That is, the oracle is constructed so that only the states satisfying $E(\mathbf{b}) - y_i < 0$ are flipped. The objective function value is encoded in the two's complement, and the first qubit represents the sign. Then, the oracle $\mathbf{O}$ can be implemented by applying a Z gate to the first qubit. The Grover diffusion operator $\mathbf{D}$ is given by [7]

$$D_{i,j} = \begin{cases} 0 & (i \neq j) \\ 1 & (i = j = 0) \\ -1 & (i = j \neq 0). \end{cases} \tag{3}$$

In summary, the oracle $\mathbf{O}$ flips the phase of the states of interest after the state preparation operator $\mathbf{A}_{y_i}$, and the operator $\mathbf{A}_{y_i}\mathbf{D}\mathbf{A}_{y_i}^{\mathrm{H}}$ amplifies only the states of interest, which correspond to a set of binary vectors $\mathbf{b}$ satisfying $E(\mathbf{b}) - y_i < 0$.

How should we design the number of Grover rotations L_i? The probability of success in obtaining states $E(\mathbf{b}) - y_i < 0$ is [19]

$$P_{\mathrm{success}} = \sin^2\left((2L_i + 1)\arcsin\sqrt{\frac{N_s}{N}}\right), \tag{4}$$

which depends on the size of the search space $N = 2^n$ and the number of solutions N_s smaller than the threshold y_i. Maximizing the success probability (4) yields the optimal number of Grover rotations [19]

$$L_{\mathrm{opt}} = \left\lfloor \frac{\pi}{4}\sqrt{\frac{N}{N_{\mathrm{s}}}} \right\rfloor, \tag{5}$$

which promises a quadratic speedup for obtaining the optimal solution.

The original GAS [12], [20] is summarized in Algorithm 1. Generally, the number of solutions N_s is unknown in advance, and the number of Grover rotations L_i is randomly selected from a set $\{0, \ldots, \lceil k - 1 \rceil\}$, where k is a parameter that

Algorithm 1 Original GAS [12], [20]

Input: $E : \mathbb{B}^n \to \mathbb{R}$, $\lambda = 8/7$
Output: $\mathbf{b}_i$

1: Uniformly sample $\mathbf{b}_0 \in \mathbb{B}^n$ and set $y_0 = \dot{y} = E(\mathbf{b}_0)$.
2: Set $k = 1$ and $i = 0$.
3: **repeat**
4: Randomly select the rotation count L_i from the set $\{0, \ldots, \lceil k - 1 \rceil\}$.
5: Evaluate $\mathbf{G}^{L_i}\mathbf{A}_{y_i}|0\rangle_{n+m}$, and obtain $\mathbf{b}$ and y.
6: Calculate $y = E(\mathbf{b})$ on a classical computer to obtain an exact value.
7: **if** $y < y_i$ **then**
8: $\mathbf{b}_{i+1} = \mathbf{b}$, $y_{i+1} = y$, and $k = 1$.
9: **else**
10: $\mathbf{b}_{i+1} = \mathbf{b}_i$, $y_{i+1} = y_i$, and $k = \min(\lambda k, \sqrt{2^n})$.
11: **end if**
12: $i = i + 1$.
13: **until** a termination condition is met.

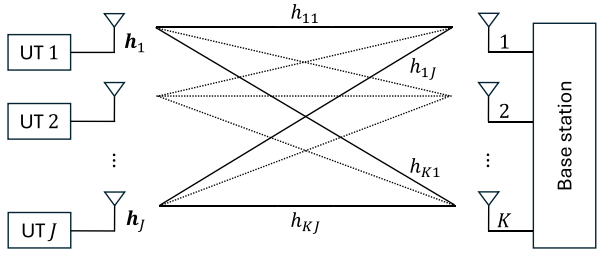


FIGURE 2. System model with J UTs and K receive antennas.

increases with the number of iterations i [24]. Specifically, once the algorithm fails to update the minimum value, the parameter k is multiplied by $\lambda > 1$. The parameter λ can be set to any value in the range of $1 < \lambda < 4/3$ [19], [24], and the IBM quantum computing simulator Qiskit [28] used in this paper sets $\lambda = 8/7$.

III. SYSTEM MODEL

We consider a narrowband system model with J UTs, each of which has a single transmit antenna, and a base station with K receive antennas, as shown in Fig. 2. A set of amplitude and phase shift keying (APSK) symbols is denoted by $\mathcal{C}$ with the constellation size $L_c = |\mathcal{C}|$. The transmitted symbol from jth user is denoted by $s_j \in \mathcal{C}$, and the power allocation ratio is denoted by ρ_j, which subsumes both multiuser MIMO and power-domain NOMA systems. Overall, the received symbol vector $\mathbf{r} \in \mathbb{C}^K$ is expressed as

$$\mathbf{r} = \sum_{j=1}^{J} \mathbf{h}_j \sqrt{\rho_j} s_j + \sigma_v \mathbf{v}, \tag{6}$$

where $\mathbf{h}_j \in \mathbb{C}^{K\times 1}$ is a channel vector with $\mathcal{CN}(0,1)$, and $\mathbf{v} \in \mathbb{C}^{K\times 1}$ is the additive white Gaussian noise (AWGN) with $\mathcal{CN}(0,1)$. The average received signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = 10 \cdot \log_{10}(\sum_j \rho_j/(J\sigma_v^2))$. The optimal performance is achieved by the MLD and is defined as a minimization problem of

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left\| \mathbf{r} - \sum_{j=1}^{J} \mathbf{h}_j \sqrt{\rho_j} s_j \right\|_{\mathrm{F}}^2 = \arg\min_{\mathbf{s}} \|\mathbf{r} - \mathbf{H}_{\mathrm{c}}\mathbf{s}\|_{\mathrm{F}}^2, \tag{7}$$

where we have a symbol vector $\mathbf{s} = [s_1 \ \cdots \ s_J]^{\mathrm{T}}$ and a channel matrix $\mathbf{H}_{\mathrm{c}} = [\sqrt{\rho_1}\mathbf{h}_1 \ \cdots \ \sqrt{\rho_J}\mathbf{h}_J] \in \mathbb{C}^{K\times J}$.

In general, ρ_j is assumed to be constant in multiuser MIMO scenarios. This is different in power-domain NOMA scenarios, and the successive interference cancellation is typically used. But, this detection technique induces an error floor in the bit error rate (BER) performance in uplink scenarios, and MLD has been considered to overcome the error floor [2], [3], [29].

IV. CONVENTIONAL QUANTUM-ASSISTED MLD

In quantum-assisted MLD, various techniques have been proposed, where they have a common approach: the detection problem is formulated as a binary optimization problem. This section briefly reviews such conventional quantum-assisted MLD techniques. Both GAS and QAOA support QUBO and HUBO problems, while QA supports only QUBO problems. This difference yields different approaches to MLD.

A. GAS FOR MLD

MLD can be formulated as a HUBO problem [20]. Let $\mathbf{b} = [\mathbf{b}_1^{\mathrm{T}} \ \cdots \ \mathbf{b}_J^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{B}^{J\log_2(L_c)}$ be a bit sequence to transmit. The problem can be formulated as [20]

$$\begin{aligned} &\min \quad E(\mathbf{b}) \\ &\text{s.t.} \quad \mathbf{b}_j \in \mathbb{B}^{\log_2 L_c}, \ \forall\, j = \{1, \cdots, J\} \end{aligned} \tag{8}$$

Here, we have the objective function of MLD [20]

$$E(\mathbf{b}) = \left\| \mathbf{r} - \sum_{j=1}^{J} \mathbf{h}_j \sqrt{\rho_j} s_j^{(L_c)}(\mathbf{b}_j) \right\|_{\mathrm{F}}^2, \tag{9}$$

where $s_j^{(L_c)}(\mathbf{b}_j) \in \mathcal{C}$ is a function of binary variables that maps a bit sequence $\mathbf{b}_j$ to a standard APSK symbol in $\mathcal{C}$. The mapping is specified in the 5G NR standard [30]. For example, in the QPSK case ($L_c = 4$), we have [30]

$$s_j^{(4)}(\mathbf{b}_j) = \frac{1}{\sqrt{2}}[(1-2b_{j,1}) + \mathrm{j}(1-2b_{j,2})]. \tag{10}$$

Similarly, we have [30]

$$\begin{aligned} s_j^{(16)}(\mathbf{b}_j) = &\frac{1}{\sqrt{10}}(1-2b_{j,1})[2-(1-2b_{j,3})] \\ &+ \frac{\mathrm{j}}{\sqrt{10}}(1-2b_{j,2})[2-(1-2b_{j,4})] \end{aligned} \tag{11}$$

in the 16-QAM case ($L_c = 16$), and [30]

$$\begin{aligned} s_j^{(64)}(\mathbf{b}_j) = &\frac{1}{\sqrt{42}}(1-2b_{j,1})[4-(1-2b_{j,3})[2-(1-2b_{j,5})]] \\ &+ \frac{\mathrm{j}}{\sqrt{42}}(1-2b_{j,2})[4-(1-2b_{j,4})[2-(1-2b_{j,6})]] \end{aligned} \tag{12}$$

in the 64-QAM case ($L_c = 64$). Later, the notation $(\cdot)^{(L_c)}$ is omitted for simplicity if the constellation size L_c is clear from the context.

The original GAS is initiated by a random solution, and the corresponding objective function value is used as an initial threshold [12]

$$\dot{y} = E(\dot{\mathbf{b}}) \tag{13}$$

and $\dot{\mathbf{b}} \in \mathbb{B}^{J\log_2(L_c)}$ is a random bit sequence. In the pioneering study [19], the output of classic linear detectors, such as ZF and MMSE, is used as an initial solution. The ZF weight matrix is given by

$$\mathbf{W} = \begin{cases} (\mathbf{H}_{\mathrm{c}}^{\mathrm{H}}\mathbf{H}_{\mathrm{c}})^{-1}\mathbf{H}_{\mathrm{c}}^{\mathrm{H}} & (J \le K) \\ \mathbf{H}_{\mathrm{c}}^{\mathrm{H}}(\mathbf{H}_{\mathrm{c}}\mathbf{H}_{\mathrm{c}}^{\mathrm{H}})^{-1} & (J > K) \end{cases} \tag{14}$$

and the MMSE weight matrix is given by

$$\mathbf{W} = \begin{cases} (\mathbf{H}_c^{\mathrm{H}}\mathbf{H}_c + \sigma_v^2\mathbf{I})^{-1}\mathbf{H}_c^{\mathrm{H}} & (J \le K) \\ \mathbf{H}_c^{\mathrm{H}}(\mathbf{H}_c\mathbf{H}_c^{\mathrm{H}} + \sigma_v^2\mathbf{I})^{-1} & (J > K) \end{cases}. \tag{15}$$

Here, the initial threshold by MMSE is given by

$$\tilde{y} = E(M^{-1}(\mathbf{W}\mathbf{r})), \tag{16}$$

where $M^{-1}(\cdot)$ is a mapping from J symbols to a sequence of $J\log_2(L_c)$ bits.

In the conventional study [20], [21], the initial threshold is designed by the distribution of the minimum of objective function (9), $E_{\min}$. If the symbol detection is correct, the only remaining term is the AWGN coefficients, and the minimum objective function value $E_{\min}$ is given by [20] and [21]

$$E_{\min} = \underbrace{\sigma_v^2}_{\text{known}} \sum_{u=1}^{K} \underbrace{|v_u|^2}_{\text{unknown}}, \tag{17}$$

where the noise variance σ_v^2 is known and the noise coefficients v_u for $u = 1, \cdots, K$ are unknown in advance. Since the noise coefficients v_u are assumed to follow a complex Gaussian distribution, $E_{\min}$ follows the Erlang distribution [20], [21]

$$f(\tilde{y}) = \frac{\gamma^K \tilde{y}^{K-1} e^{-\gamma\tilde{y}}}{(K-1)!} \tag{18}$$

with $\gamma = 1/\sigma_v^2$. The corresponding cumulative distribution function (CDF) is given by [20] and [21]

$$F(\tilde{y}) = \Pr[E_{\min} \le \tilde{y}] = 1 - e^{-\gamma\tilde{y}} \sum_{u=0}^{K-1} \frac{(\gamma\tilde{y})^u}{u!}. \tag{19}$$

Then, the probability that the initial threshold becomes smaller than the minimum objective function value $E_{\min}$ is given by [20], [21]

$$P = \Pr[E_{\min} > \tilde{y}] = e^{-\gamma\tilde{y}} \sum_{u=0}^{K-1} \frac{(\gamma\tilde{y})^u}{u!}, \tag{20}$$

and the improved initial threshold $\tilde{y}$ can be calculated by (20) for a certain small probability P.

B. QA FOR MLD

GAS supports HUBO problems, while QA supports only QUBO problems. Since the objective function of MLD is a quadratic function of symbols, the mapping from a bit sequence to a symbol $s_j(\mathbf{b}_j)$ must be linear in QA. However, the mapping (11) is quadratic and (12) is cubic, and they cannot be formulated as QUBO problems. Even if the objective function is cubic or higher, it can be converted to a quadratic function by adding auxiliary binary variables. In this case, the search space size is doubled for each auxiliary variable, which is unrealistic in terms of complexity.

To circumvent this escalating complexity, the authors in [16] invented a novel transformation that yields a linear mapping in the context of QA, termed *QuAMax transformation*. This transformation consists of three steps. First, the standard Gray-coded symbols are transmitted. Second, the receiver uses a linear mapping to obtain non-Gray-coded symbols. For example, in the 16-QAM case, we have [16]

$$s_j^{(16)'}(\mathbf{b}_j) = (4b_{j,1} + 2b_{j,2} - 3)/\sqrt{10} + (4b_{j,3} + 2b_{j,4} - 3)\mathrm{j}/\sqrt{10} \tag{21}$$

and in the 64-QAM case, we have [17]

$$s_j^{(64)'}(\mathbf{b}_j) = (8b_{j,1} + 4b_{j,2} + 2b_{j,3} - 7)/\sqrt{42} + (8b_{j,4} + 4b_{j,5} + 2b_{j,6} - 7)\mathrm{j}/\sqrt{42}. \tag{22}$$

Finally, the receiver performs a simple bit conversion to obtain a bit sequence that corresponds to the Gray-coded symbols. For example, in the 16-QAM case, we have $\hat{b}_0 = b_0$, $\hat{b}_1 = b_0 \oplus b_1$, $\hat{b}_2 = b_1 \oplus b_2$, and $\hat{b}_3 = b_2 \oplus b_3$ [16]. This postprocessing requires polynomial complexity, which is negligible against the detection complexity.

V. ANALYSIS OF QUANTUM GATE COUNT

In the conventional studies using Grover-based algorithms, the construction of a quantum oracle is assumed to be solved in the long term, and the specific complexity in terms of the number of gates has not been analyzed in detail. This section analyzes the impact of Gray coding on the number of quantum gates. Here, we focus on the number of T gates, which is an important metric when assuming surface-code-based quantum computation. We analyze the number of quantum gates required in the state preparation operator $\mathbf{A}_{y_i}$ of GAS, which dominates the total number of quantum gates.

The number of gates can be calculated by counting the number of terms in the objective function $E(\mathbf{b})$ of (9), where the mapping $s_j^{(L_c)}(\mathbf{b}_j)$ or $s_j^{(L_c)'}(\mathbf{b}_j)$ is directly substituted. Expanding the objective function (9) yields (23), as shown at the bottom of the next page, and the terms can be classified into four groups, as indicated in (23). The first group is a collection of constant terms that represent the amplitudes of received symbols. The second group is also a collection of terms that represent the amplitudes of original symbols, and they are constant in the BPSK and QPSK cases. The third group consists of independent symbols, while the fourth group consists of the interactions between different symbols.

By substituting the mapping $s_j^{(L_c)}(\mathbf{b}_j)$ or $s_j^{(L_c)'}(\mathbf{b}_j)$ into (23), one can calculate the number of terms corresponding to each order, where the terms $|s_j(\mathbf{b}_j)|^2$, $s_j(\mathbf{b}_j)$, and $s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k)$ are considered. Then, the number of H, R, controlled R (CR), and $\mathrm{C}^k\mathrm{R}$ for $k \ge 2$ gates can be calculated accordingly. In the following, we show the number of gates in the Gray-coded and non-Gray-coded 16-QAM cases as examples.

a: GRAY-CODED 16-QAM

We consider the Gray-coded 16-QAM case leading to a HUBO formulation, where the number of bits is n and the number of users is $J = n/\log_2(16) = n/4$. Out of J symbols,

the number of interactions between different two symbols $s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k)$ is $\binom{J}{2}$. From (11), since we have

$$s_j(\mathbf{b}_j) = (1 - 2b_{j,1} + 2b_{j,3} - 4b_{j,1}b_{j,3})/\sqrt{10} + (1 - 2b_{j,2} + 2b_{j,4} - 4b_{j,2}b_{j,4})\mathrm{j}/\sqrt{10} \quad (24)$$

and $s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k)$ yields four 4th-order terms, the total number of 4th-order terms is simply calculated as

$$4\binom{J}{2} = 4\frac{J(J-1)}{2} = \frac{n(n-4)}{8}. \quad (25)$$

Similarly, $s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k)$ yields 16 3rd-order terms, and the total number of 3rd-order terms is calculated as $n(n-4)/2$. The number of 2nd-order terms is 16 for each pair of symbols and two for each symbol, and the total number of 2nd-order terms is $n(n-4)/2 + n/2 = n(n-3)/2$. The number of 1st-order terms is nm. Then, the number of corresponding C^4R, C^3R, C^2R, and CR gates are $n(n-4)/8$, $n(n-4)/2$, $n(n-3)/2$, and nm, respectively.

b: NON-GRAY-CODED 16-QAM

Next, in the non-Gray-coded 16-QAM case leading to a QUBO formulation, we can analyze the number of gates in the same manner as the previous Gray-coded case. Using (21), the maximum order of the objective function is suppressed to two, and it can be formulated as a QUBO problem. The number of 2nd-order terms is 16 for each pair of symbols, and $|s_j(\mathbf{b}_j)|^2$ is simply expressed as

$$|s_j(\mathbf{b}_j)|^2 = \frac{1}{10}(4b_{j,1} + 2b_{j,2} - 3)^2 + \frac{1}{10}(4b_{j,3} + 2b_{j,4} - 3)^2, \quad (26)$$

which yields $n(n-4)/2 + n/2 = n(n-3)/2$ 2nd-order terms in total. The number of 1st-order terms is the same as the Gray-coded case, nm.

Similarly, we analyzed the number of gates required for the Gray-coded and non-Gray-coded 64-QAM cases. The results are summarized in Table 1. Since m phase gates $\mathbf{R}(\theta)$ are required to represent a coefficient, the number of phase gates is m times the number of terms. In the Gray-coded case, 16-QAM and 64-QAM require C^4R and C^6R, and 256-QAM and 1024-QAM require C^8R and $C^{10}R$, respectively. By contrast, in the non-Gray-coded case, C^2R is sufficient for supporting all constellation sizes, and the total number of gates can be reduced significantly.

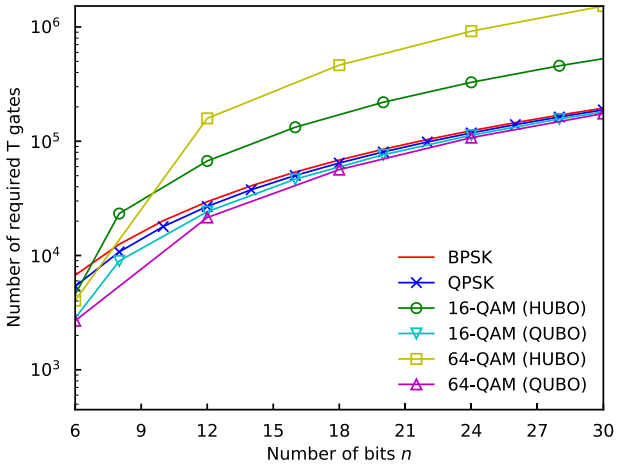


FIGURE 3. Estimated number of T gates required by each formulation with $J = n/\log_2 L_c$.

In surface-code-based quantum computation, the number of T gates determines the complexity of a quantum circuit. Fig. 3 shows the number of T gates required by each formulation, where the number of qubits required for encoding $E(\mathbf{b})$ is $m = 16$ and the number of binary variables is increased from $n = 6$ to 30. The number of T gates does not depend on the number of receive antennas K at the base station. Here, we assume that the phase gate $\mathbf{R}(\theta)$ can be directly implemented in quantum hardware and is not decomposed further. A CR gate is decomposed into CNOT and single-qubit unitary gates including $\mathbf{R}(\theta)$. For $k \leq 2$, a C^kR gate is decomposed into CR gates, which yields $14(n-1)$ T gates using $n-1$ auxiliary qubits initialized by zero [31]. Note that, for $n = 1$, no auxiliary qubits are required. As shown in Fig. 3, in the BPSK and QPSK cases, the number of T gates is constant regardless of the use of Gray coding. In the Gray-coded 16-QAM and 64-QAM cases, the number of T gates increased significantly due to the

$$\begin{aligned} E(\mathbf{b}) &= \sum_{u=1}^{K} |r_u - \sqrt{\rho_1}h_{u1}s_1(\mathbf{b}_1) - \sqrt{\rho_2}h_{u2}s_2(\mathbf{b}_2) - \cdots - \sqrt{\rho_J}h_{uJ}s_J(\mathbf{b}_J)|^2 \\ &= \sum_{u=1}^{K} \Bigg\{ \underbrace{|r_u|^2}_{\text{1. Constant terms}} + \underbrace{\sum_{j=1}^{J} \rho_j |h_{uj}|^2 |s_j(\mathbf{b}_j)|^2}_{\text{2. Amplitudes of symbols}} - \underbrace{\sum_{j=1}^{J} 2\rho_j \Re(r_u h_{uj}^*)\Re(s_j(\mathbf{b}_j)) - \sum_{j=1}^{J} 2\rho_j \Im(r_u h_{uj}^*)\Im(s_j(\mathbf{b}_j))}_{\text{3. Independent symbols}} \Bigg\} \\ &\quad + \underbrace{\sum_{j=1}^{J}\sum_{k=j+1}^{J} 2\sqrt{\rho_j}\sqrt{\rho_k}\Re(h_{uj}h_{uk}^*)\Re(s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k)) + \sum_{j=1}^{J}\sum_{k=j+1}^{J} 2\sqrt{\rho_j}\sqrt{\rho_k}\Im(h_{uj}h_{uk}^*)\Im(s_j^*(\mathbf{b}_j)s_k(\mathbf{b}_k))}_{\text{4. Interactions between different symbols}} \end{aligned} \quad (23)$$

TABLE 1. Summary of quantum gate analysis, where n is the number of bits in $\mathbf{b}$ and m is the number of qubits for encoding $E(\mathbf{b})$.

Gate	16-QAM HUBO		QUBO		64-QAM HUBO		QUBO	
H	$n+m$	$=O(n+m)$	$n+m$	$=O(n+m)$	$n+m$	$=O(n+m)$	$n+m$	$=O(n+m)$
R	m	$=O(m)$	m	$=O(m)$	m	$=O(m)$	m	$=O(m)$
CR	nm	$=O(nm)$	nm	$=O(nm)$	nm	$=O(nm)$	nm	$=O(nm)$
C^2R	$n(n-3)m/2$	$=O(n^2m)$	$n(n-3)m/2$	$=O(n^2m)$	$n(n-4)m/2$	$=O(n^2m)$	$n(n-4)m/2$	$=O(n^2m)$
C^3R	$n(n-4)m/2$	$=O(n^2m)$	0		$n(n-6)m+nm/3$	$=O(n^2m)$	0	
C^4R	$n(n-4)m/8$	$=O(n^2m)$	0		$5n(n-6)m/6$	$=O(n^2m)$	0	
C^5R	0		0		$n(n-6)m/3$	$=O(n^2m)$	0	
C^6R	0		0		$n(n-6)m/18$	$=O(n^2m)$	0	
IQFT	1		1		1		1	

TABLE 2. Summary of initial thresholds.

	Random $\dot{y}$ [12]	MMSE $\bar{y}$ [19]	Analytical $\tilde{y}$ [20]	SDP $\hat{y}$
Complexity	$O(1)$	$O(n^{2.376})$	$O(1)$	$O(n^{4.5}\log(\epsilon^{-1}))$

higher-order terms. By contrast, in the non-Gray-coded case, the number of T gates was reduced compared to the BPSK and QPSK cases. This is because the number of users J decreases as the constellation size L_c increases given the same number of qubits n.

VI. PROPOSED PARAMETER OPTIMIZATION IN GAS

This section proposes techniques for optimizing the initial threshold and the number of Grover rotations to reduce the query complexity of GAS.

A. DESIGN OF INITIAL THRESHOLD

SDP-based symbol detection for MIMO systems is widely recognized to outperform linear detectors [4], and it can be used as an initial threshold for GAS. SDP can be solved efficiently, and the worst-case complexity is polynomial [4]

$$O(\max\{\bar{m}, n\}^4 n^{1/2}\log(1/\epsilon)) \tag{27}$$

with respect to the number of binary variables n and the number of constraints $\bar{m}$, where ϵ is the precision of solution. Although various efficient methods have been proposed to obtain data symbols from the obtained positive semidefinite matrix, we use the simple quantization method [32] to obtain the initial threshold $\hat{y}$. At low SNRs, the SDP-based detection may fail, and the threshold $\hat{y}$ may become too large. To avoid this, we consider combining the initial threshold $\hat{y}$ with the threshold $\tilde{y}$ of (20), which is obtained from the minimum value distribution, and using $\min(\tilde{y}, \hat{y})$ as the initial threshold.

The major drawback of this approach is that the complexity of solving SDP is not negligible. Specifically, the complexity of the conventional random initial threshold is $O(1)$, the MMSE-based threshold is $O(J^{2.376}) = O((n/L_c)^{2.376})$,[2] and the analytical threshold of (20) is $O(1)$, while that of the SDP-based threshold is $O(n^{4.5}\log(\epsilon^{-1}))$. This complexity overhead is much smaller than the original complexity $O(L_c^{n/L_c})$. Moreover, the SDP solver itself is also shown to be accelerated by quantum computation [34], which may reduce the complexity overhead. Table 2 summarizes the complexity of each initial threshold where L_c is regarded as a constant for simplicity.

[2]In general, the complexity of MMSE is recognized as $O(J^3)$, but it can be reduced to $O(J^{2.376})$ using the Coppersmith-Winograd algorithm [33].

B. DESIGN OF GROVER ROTATIONS

When using the proposed initial threshold $\hat{y}$ or $\min(\tilde{y}, \hat{y})$, the number of solutions N_s to be amplified is likely to be smaller than that of the random initial threshold $\dot{y}$. That is, the optimal number of Grover rotations L_{opt} is also likely to become larger than that of $\dot{y}$ according to the inverse relationship of (5). In such a case, where the statistical information of N_s is known in advance, how should we design the number of Grover rotations? This case is considered to be an intermediate between the totally unknown random case [24] and the case with the full knowledge of N_s where the quantum counting algorithm [35], [36] is applied.[3]

We are interested in the states where the objective function (9) is smaller than a certain threshold y_0

$$E(\mathbf{b}) = \left\| \mathbf{r} - \sum_{j=1}^{J} \mathbf{h}_j \sqrt{\rho_j} s_j(\mathbf{b}_j) \right\|_{\text{F}}^2 < y_0, \tag{28}$$

which indicates that the detection problem is to find a symbol that is within the radius $\sqrt{y_0}$ from the received symbol as shown in Fig. 4. In Fig. 4, we show the case of $J = 2$ and QPSK, where the black circles represent the candidate superimposed symbols. From this relationship, the number of states smaller than the initial threshold y_0 can be determined by the transmit symbols, channel coefficients, the power ratio of UTs, and noise. From this complex relationship, it is difficult to derive the theoretical expression of the probability distribution of the number of solutions $p(N_s)$ for a given threshold y_0.

Instead, we investigate the number of solutions N_s through Monte Carlo simulations. In the detection problem, since the random variables $\mathbf{h}_j$ and $\mathbf{v}$ are included in the objective function, $p(N_s)$ is expected to follow a certain probability distribution. For example, Fig. 5 shows the CDF of the number of solutions N_s smaller than the threshold $\min(\tilde{y}, \hat{y})$, where the constellation size is $L_c = 16$ and power ratio of

[3]The quantum counting algorithm [35], [36] requires $O(\sqrt{N})$ queries.

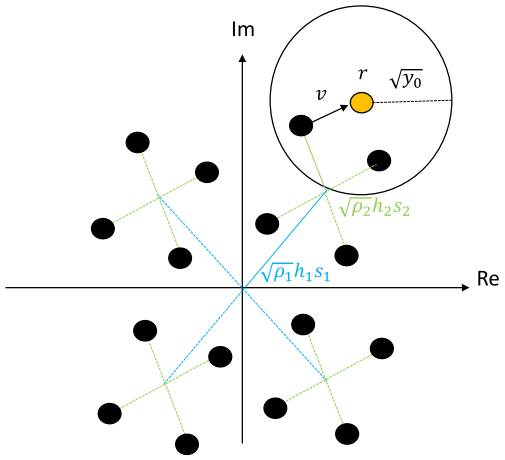


FIGURE 4. A relationship between the number of states smaller than the initial threshold y_0 and a received symbol r in multiuser detection.

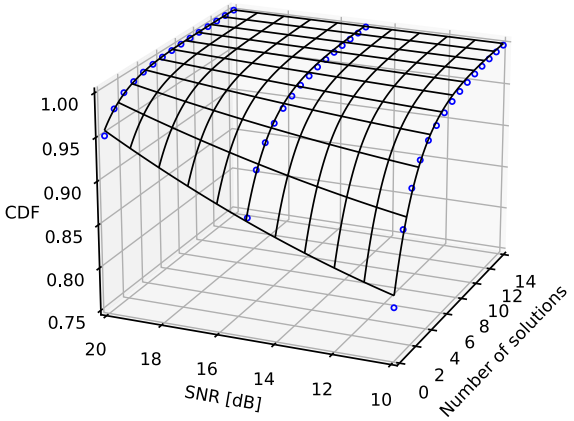
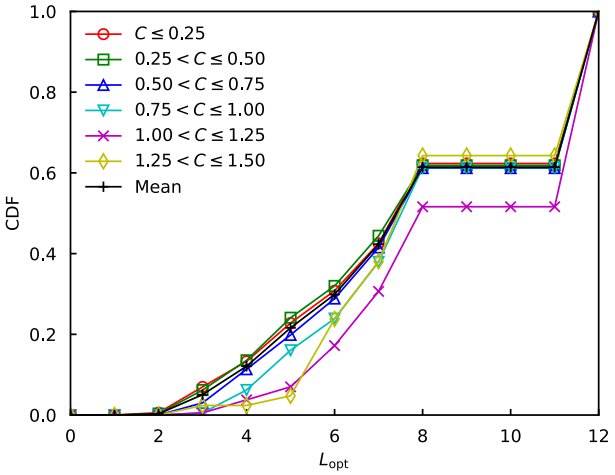
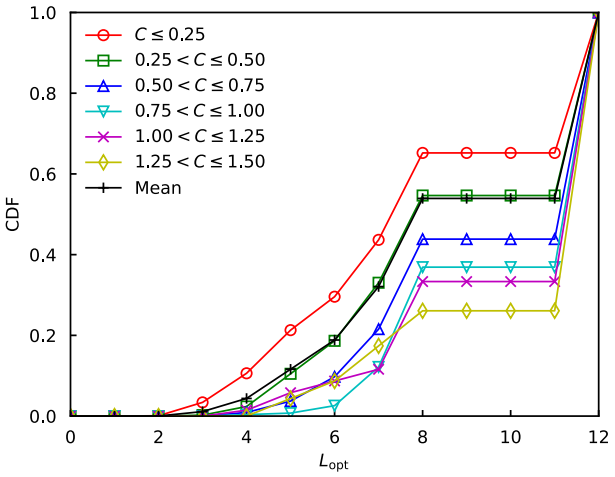


FIGURE 5. CDF of the number of solutions N_s that are smaller than the proposed initial threshold upon changing SNR.



(a) SNR = 15dB.



(b) SNR = 20dB.

FIGURE 6. CDF of the optimal number of Grover rotations corresponding to the Frobenius norm of the channel matrix ($\rho_1 = 2\rho_2$).

$J = 2$ UTs is $\rho_1 = 2\rho_2$. The blue circles represent the simulation results, and the CDF has a continuous and certain trend as shown in Fig. 5. Then, we fit the CDF with the following heuristic function of six parameters.

$$P(N_s) = c_1 + (1 - c_2^{-c_3 \cdot N_s})(1 - c_1) + c_4 \cdot \mathrm{SNR}^2 \cdot c_5^{-c_6 \cdot N_s}, \tag{29}$$

which satisfies $P(N_s) \to 1$ as $N_s \to \infty$. In the case of Fig. 5, the obtained parameters were $c_1 = 0.7747, c_2 = 1.3339, c_3 = 1.3203, c_4 = 0.0005, c_5 = 1.0051, c_6 = 70.592$. The black grid shows the fitted function, which can be used to obtain complementary values of the CDF of N_s at any SNR with sufficient accuracy.

The conventional analytical threshold $\tilde{y}$ of (20) is derived for a given average received SNR. In practical communication scenarios, coherent detection is assumed in general, and the instantaneous channel coefficients are estimated and tracked periodically, which can be used to determine the number of Grover rotations in a more appropriate manner. The reliability of multiuser detection is mainly determined by $R = \mathrm{rank}(\mathbf{H}_c) \le \min(J, K)$ singular values $\sigma_1, \cdots, \sigma_R$ of the channel matrix $\mathbf{H}_c \in \mathbb{C}^{J \times K}$, and we use its Frobenius norm

$$C \equiv \frac{1}{J \cdot K} \|\mathbf{H}_c\|_F^2 = \frac{1}{J \cdot K} \sum_{i=1}^{R} \sigma_i^2 \tag{30}$$

as a key indicator for the Grover rotations. Note that C obviously follows the Erlang distribution as with the analysis of (17) if the power allocations ratios are identical.

Figs. 6(a) and (b) show the CDF of the optimal Grover rotations L_{opt} for the proposed initial threshold $\min(\tilde{y}, \hat{y})$. For simple visualization, we divided the Frobenius norm $\|\mathbf{H}_c\|^2$ into multiple intervals and show the average value in each interval. As shown in Fig. 6, the probability of the optimal number of Grover rotations L_{opt} being 1 to 3 was negligible regardless of SNR and the Frobenius norm. When SNR = 20dB, the probability of L_{opt} being large increased as the Frobenius norm increased. By contrast, when SNR = 15dB, the Frobenius norm did not affect L_{opt} significantly.

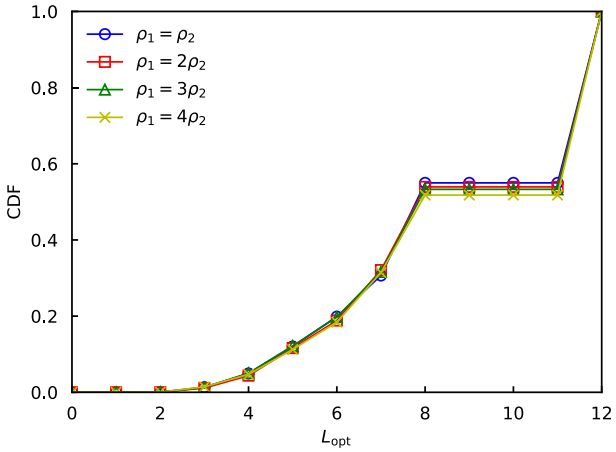


FIGURE 7. Impact of power allocation on the optimal number of Grover rotations with SNR = 20dB.

This is because the SDP-based threshold $\hat{y}$ was smaller than the analytical threshold $\tilde{y}$ in most cases.

Additionally, Fig. 7 investigates the impact of power allocation on the optimal number of Grover rotations. Even if the power allocation ratio for each UT is different, the CDF of the optimal number of Grover rotations is not significantly different. That is, the parameter optimization of GAS will not be significantly affected by the specific power allocation method.

As discussed in Figs. 6 and 7, the number of solutions N_s and the optimal number of Grover rotations L_{opt} have a certain bias when using the proposed initial threshold $\min(\tilde{y}, \hat{y})$. In the original GAS, as summarized in Algorithm 1, the number of Grover rotations L_i is drawn from a uniform distribution $[0, \lceil k - 1 \rceil]$. Here, we propose to ignore the unlikely values of $L = 0$ to $L_{\min} - 1$ and to narrow the range of the uniform distribution to $[L_{\min}, \lceil k - 1 \rceil]$. As the Grover iterations proceed, the optimal value of $L_{\min}$ is expected to increase, but it is clear that the value becomes larger than $L_{\min}$. We present the proof that the query complexity is not worsened even if $L_{\min} > 0$ in Appendix A, which indicates that this modification does not affect the optimality of the obtained solution. For example, in the case of Fig. 6, we divide $C \equiv \|\mathbf{H}_{\text{c}}\|_{\text{F}}^2/(J \cdot K)$ into multiple intervals and set $L_{\min}$ as

$$L_{\min} = \begin{cases} 4 & C < 0.25 \\ 5 & 0.25 \le C < 0.50 \\ 6 & 0.50 \le C < 1.00 \\ 7 & 1.00 \le C \end{cases}. \tag{31}$$

It can be expected that the rough boundaries here will absorb the negative effects of channel estimation errors typically encountered in practical communication scenarios.

The modified GAS is summarized in Algorithm 2, where the initial threshold is set to $\min(\tilde{y}, \hat{y})$ and the number of Grover rotations is drawn from the uniform distribution $[L_{\min}, \lceil k - 1 \rceil]$.

Algorithm 2 Proposed GAS

Input: $E : \mathbb{B}^n \to \mathbb{R}$, $\lambda = 8/7$, $L_{\min}$, $\min(\tilde{y}, \hat{y})$
Output: $\mathbf{b}$

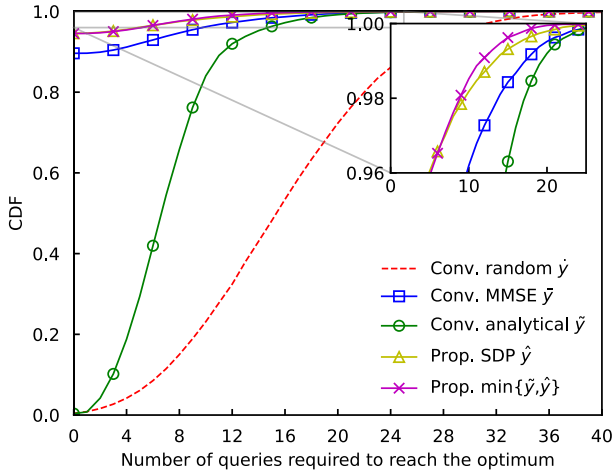
1: Set $y_0 = \min(\tilde{y}, \hat{y})$.
2: Set $k = L_{\min} + 1$ and $i = 0$.
3: **repeat**
4: Randomly select the rotation count L_i from the set $\{L_{\min}, \ldots, \lceil k - 1 \rceil\}$.
5: Evaluate $\mathbf{G}^{L_i}\mathbf{A}_{y_i}|0\rangle_{n+m}$, and obtain $\mathbf{b}$ and y.
6: Calculate $y = E(\mathbf{b})$ on a classical computer to obtain an exact value.
7: **if** $y < y_i$ **then**
8: $\mathbf{b}_{i+1} = \mathbf{b}$, $y_{i+1} = y$, and $k = L_{\min} + 1$.
9: **else**
10: $\mathbf{b}_{i+1} = \mathbf{b}_i$, $y_{i+1} = y_i$, and $k = \min(\lambda k, \sqrt{2^n})$.
11: **end if**
12: $i = i + 1$.
13: **until** a termination condition is met.

VII. PERFORMANCE COMPARISONS

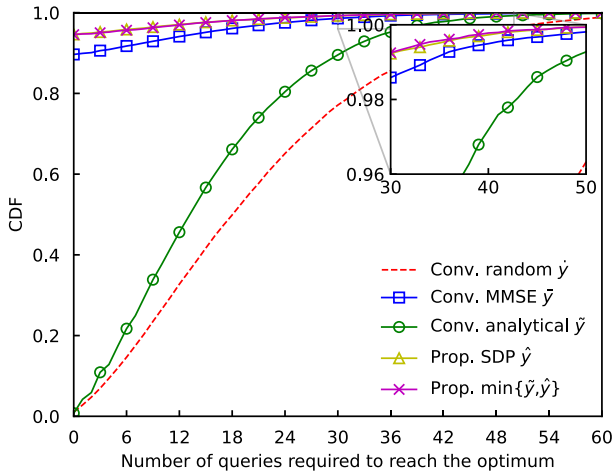
In our performance comparisons, we consider the multiuser detection problem of the uplink PD-NOMA system with $J = 2$ UTs, $K = 2$ antennas at the base station, and $L_{\text{c}} = 16$-QAM [2], [3], [29]. Here, it is not necessary to consider a case where the number of antennas at the base station is sufficiently larger than the number of UTs because in such case efficient linear equalizers such as ZF and MMSE are available, indicating that quantum computing is not required. The SNR is fixed to 20dB, and the power allocation ratio for each UT is set to $\rho_1 = 2\rho_2$. We assumed a sufficiently large number of qubits m and ignored the effect of approximation errors induced by real-valued coefficients. The probability P used in the threshold $\tilde{y}$ of (20) was fixed to 10^{-4}. To solve SDP and obtain $\hat{y}$, we used the modeling tool CVXPY [37] and the solver MOSEK [38].

As performance metrics, we use the standard query complexity as well as the number of total iterations i in Algorithms 1 and 2 until the optimal solution is obtained. The former is referred to as query complexity in the quantum domain (QD), and the latter is referred to as query complexity in the classical domain (CD) [19]. The CDFs of the query complexities in the QD and CD are compared, and the curve in the upper left indicates that the optimal solution is obtained earlier. Note that all the considered schemes can obtain the optimal solution and the achievable bit error ratios are exactly the same.

First, we compared the convergence performance of GAS with different initial thresholds. Specifically, we considered the conventional initial thresholds (1) random $\dot{y}$ [12], (2) MMSE-based $\bar{y}$ [19], and (3) analytical $\tilde{y}$ [20], as well as the proposed (4) SDP-based $\hat{y}$ and (5) $\min(\tilde{y}, \hat{y})$. As shown in Fig. 8, the convergence performance of the random initial threshold $\dot{y}$ was the worst, and the analytical threshold $\tilde{y}$ outperformed it in both the CD and QD. The

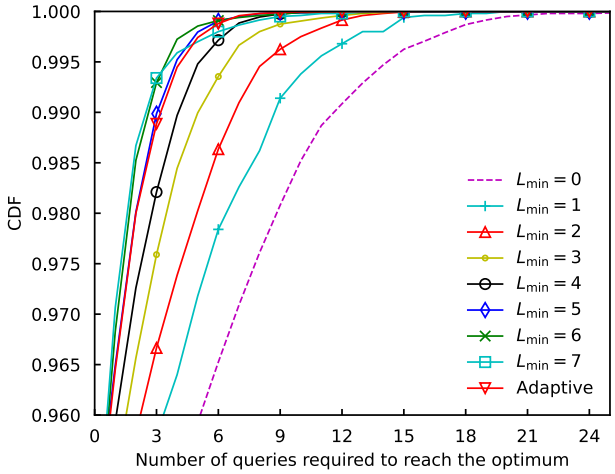


(a) CD

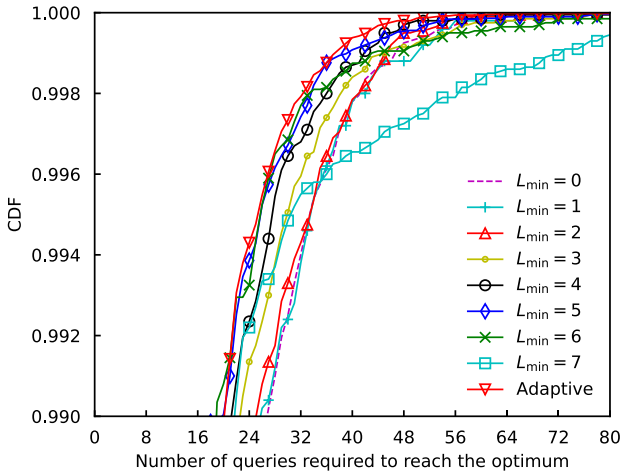


(b) QD

FIGURE 8. Comparison of different initial thresholds in GAS.



(a) CD



(b) QD

FIGURE 9. Comparison of different $L_{\min}$ values and the adaptive $L_{\min}$ of (31) using the proposed initial threshold.

MMSE-based threshold $\bar{y}$ exhibited a large value from the first iteration, which indicates that good solutions were obtained from the beginning. The proposed SDP-based threshold $\hat{y}$ outperformed the MMSE-based one, and the combination of $\tilde{y}$ and $\hat{y}$ achieved the best performance among all the considered thresholds. This is because the SDP-based threshold $\hat{y}$ sometimes fails to obtain a good solution, and the use of the analytical threshold $\tilde{y}$ works as a backup to improve the convergence performance.

Next, Fig. 9 shows the performance comparison when the minimum number of Grover rotations $L_{\min}$ was changed, where the proposed initial threshold $\min(\tilde{y}, \hat{y})$ was considered since it achieved the best performance in Fig. 8. Setting $L_{\min}$ to an appropriate value significantly improved the query complexity in CD and also yielded a slight improvement in QD. The $L_{\min} = 5$ exhibited good performances in both CD and QD. In addition, the proposed adaptive setting of $L_{\min}$ (31) according to the Frobenius norm C achieved a good balance in both CD and QD. This suggests that the heuristic setting of (31) can be considered sufficient in practice in the specific scenario.

VIII. CONCLUSION

In this paper, we addressed the multiuser detection problem in the uplink PD-NOMA system. To mitigate the escalating complexity when using MLD, we used GAS to solve the problem. Here, the mapping of the objective function to a quantum circuit of GAS is crucial for the feasibility on a quantum computer. Then, we analyzed the T gate count of different mapping methods and clarified that the one with quadratic formulation is the most efficient. We also proposed two optimized parameters for GAS: the initial threshold $\min(\tilde{y}, \hat{y})$ and the minimum number of Grover rotations $L_{\min}$. The former is a combination of the analytical threshold $\tilde{y}$ and the SDP-based threshold $\hat{y}$, and the latter is set to a heuristic adaptive value depending on the Frobenius norm of an instantaneous channel matrix. The introduction of $L_{\min} > 0$ was proven not to affect the achievement of

quadratic speedup. Our simulation results indicate that the proposed optimized GAS can achieve a good balance in both CD and QD when assuming perfect channel estimation. The calculation of $\hat{y}$ requires the preprocessing of SDP, which is a potential drawback of the proposed approach. However, it is negligible when compared to the overall complexity.

APPENDIX
ANALYSIS OF QUERY COMPLEXITY

We analyze the query complexity of GAS using the optimized parameter, $L_{\min} > 0$, based on [24]. The probability of obtaining the desired state when choosing the number of Grover rotations L_i from a uniform distribution $[L_{\min}, k-1]$ is given as

$$\begin{aligned} P_k &= \sum_{L_i=L_{\min}}^{k-1} \frac{1}{k-L_{\min}} \sin^2((2L_i+1)\theta)) \\ &= \frac{1}{2(k-L_{\min})} \sum_{L_i=L_{\min}}^{k-1} \{1-\cos((2L_i+1)2\theta)\} \\ &= \frac{1}{2} - \frac{\sin(2k\theta - 2L_{\min}\theta)\cdot\cos(2k\theta+2L_{\min}\theta)}{2(k-L_{\min})\sin(2\theta)} \\ &= \frac{1}{2} - \frac{\sin(4k\theta)+\sin(-4L_{\min}\theta)}{4(k-L_{\min})\sin(2\theta)} \end{aligned} \tag{32}$$

with angle θ such that $\sin^2\theta = N_s/N$. For an integer $k \geq L_{\min} + 2/\sin(2\theta)$, the relationship

$$\frac{\sin(4k\theta)+\sin(-4L_{\min}\theta)}{4(k-L_{\min})\sin(2\theta)} \leq \frac{1}{2(k-L_{\min})\sin(2\theta)} \leq \frac{1}{4} \tag{33}$$

yields a lower-bound of the probability $P_k \geq 1/2 - 1/4 = 1/4$. The value of k_0 at which $P_{k_0} = 1/4$ is

$$k_0 = L_{\min} + \frac{N}{\sqrt{(N-N_s)N_s}} < L_{\min} + \sqrt{\frac{N}{N_s}}. \tag{34}$$

As described in Algorithms 1 and 2, k is updated at each iteration as $k = \min(\lambda k, \sqrt{2^n})$. The average total number of Grover rotations, until k exceeds k_0, is the same as [24], i.e.,

$$\frac{1}{2}\sum_{i=1}^{\lceil \log_\lambda k_0 \rceil} \lambda^{s-1} < \frac{1}{2}\frac{\lambda}{\lambda-1}k_0. \tag{35}$$

After k exceeds k_0, the desired state can be found with at least $1/4$ probability per iteration, and the average total number of Grover rotations can be calculated as

$$\begin{aligned} &\frac{1}{2}\sum_{u=0}^{\infty}\frac{3^u}{4^{u+1}}\lambda^{u+\lceil\log_\lambda \tilde{k}_0\rceil}(L_{\min}+1) < \frac{\lambda(L_{\min}+1)}{8-6\lambda}\tilde{k}_0 \\ &= \frac{\lambda}{8-6\lambda}k_0 \end{aligned} \tag{36}$$

using $\tilde{k}_0 = k_0/(L_{\min}+1)$. Since λ is constant, the overall query complexity is upper bounded by $\lambda/(\lambda-1)/2\ k_0 + \lambda/(8-6\lambda)k_0 = O(k_0) = O(\sqrt{N/N_s})$. That is, if $L_{\min}$ is sufficiently small compared to $\sqrt{N/N_s}$, the query complexity is of the same order as that of the conventional algorithm with $L_{\min} = 0$. Therefore, the introduction of $L_{\min} > 0$ was proven not to affect the achievement of quadratic speedup.

REFERENCES

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013, pp. 1–5.
[2] J. S. Yeom, H. S. Jang, K. S. Ko, and B. C. Jung, "BER performance of uplink NOMA with joint maximum-likelihood detector," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10295–10300, Oct. 2019.
[3] H. Semira, F. Kara, H. Kaya, and H. Yanikomeroglu, "Multi-user joint maximum-likelihood detection in uplink NOMA-IoT networks: Removing the error floor," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2459–2463, Nov. 2021.
[4] Z.-Q. Luo, W.-K. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
[5] *International Roadmap for Devices and Systems: Executive Summary*, IRDS, Hyderabad, India, May 2022.
[6] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Nov. 1994, pp. 124–134.
[7] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput.*, 1996, pp. 1–14.
[8] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett.*, vol. 103, no. 15, pp. 1–15, Oct. 2009.
[9] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014, *arXiv:1411.4028*.
[10] D. Stilck França and R. García-Patrón, "Limitations of optimization algorithms on noisy quantum devices," *Nature Phys.*, vol. 17, no. 11, pp. 1221–1227, Nov. 2021.
[11] T. Kadowaki and H. Nishimori, "Quantum annealing in the transverse ising model," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 58, no. 5, pp. 5355–5363, Nov. 1998.
[12] A. Gilliam, S. Woerner, and C. Gonciulea, "Grover adaptive search for constrained polynomial binary optimization," *Quantum*, vol. 5, p. 428, Apr. 2021.
[13] T. C. Cuvelier, S. A. Lanham, B. R. L. Cour, and R. W. Heath, "Quantum codes in classical communication: A space-time block code from quantum error correction," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2383–2412, Oct. 2021.
[14] J. Cui, Y. Xiong, S. X. Ng, and L. Hanzo, "Quantum approximate optimization algorithm based maximum likelihood detection," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5386–5400, Aug. 2022.
[15] B. Gulbahar, "Maximum-likelihood detection with QAOA for massive MIMO and sherrington-kirkpatrick model with local field at infinite size," *IEEE Trans. Wireless Commun.*, early access, Sep. 2024.
[16] M. Kim, D. Venturelli, and K. Jamieson, "Leveraging quantum annealing for large MIMO processing in centralized radio access networks," in *Proc. ACM Special Interest Group Data Commun.*, NY, NY, USA, Aug. 2019, pp. 241–255.
[17] Z. I. Tabi, Á. Marosits, Z. Kallus, P. Vaderna, I. Gódor, and Z. Zimborás, "Evaluation of quantum annealer performance via the massive MIMO problem," *IEEE Access*, vol. 9, pp. 131658–131671, 2021.
[18] R. C. Kizilirmak, "Quantum annealing approach to NOMA signal detection," in *Proc. 12th Int. Symp. Commun. Syst., Netw. Digit. Signal Process. (CSNDSP)*, Jul. 2020, pp. 1–5.
[19] P. Botsinis, S. X. Ng, and L. Hanzo, "Fixed-complexity quantum-assisted multi-user detection for CDMA and SDMA," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 990–1000, Mar. 2014.
[20] M. Norimoto, R. Mori, and N. Ishikawa, "Quantum algorithm for higher-order unconstrained binary optimization and MIMO maximum likelihood detection," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1926–1939, Apr. 2023.
[21] M. Norimoto and N. Ishikawa, "Grover adaptive search for joint maximum-likelihood detection of power-domain non-orthogonal multiple access," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2023, pp. 1–5.

[22] Y. Sano, M. Norimoto, and N. Ishikawa, "Qubit reduction and quantum speedup for wireless channel assignment problem," *IEEE Trans. Quantum Eng.*, vol. 4, pp. 1–12, Jul. 2023.
[23] C. Xu, N. Ishikawa, R. Rajashekar, S. Sugiura, R. G. Maunder, Z. Wang, L.-L. Yang, and L. Hanzo, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178246–178299, 2019.
[24] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, "Tight bounds on quantum searching," *Fortschritte Der Physik*, vol. 46, nos. 4–5, pp. 493–505, Jun. 1998.
[25] C. Durr and P. Hoyer, "A quantum algorithm for finding the minimum," 1996, *arXiv: quant-ph/9607014*.
[26] D. Bulger, W. P. Baritompa, and G. R. Wood, "Implementing pure adaptive search with Grover's quantum algorithm," *J. Optim. Theory Appl.*, vol. 116, no. 3, pp. 517–529, Mar. 2003.
[27] T. G. Draper, "Addition on a quantum computer," 2000, *arXiv: quant-ph/0008033*.
[28] D. S. Steiger, T. Häner, and M. Troyer, "ProjectQ: An open source software framework for quantum computing," *Quantum*, vol. 2, p. 49, Jan. 2018.
[29] H. Semira, F. Kara, H. Kaya, and H. Yanikomeroglu, "Error performance analysis of multiuser detection in uplink-NOMA with adaptive M-QAM," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1654–1658, Aug. 2022.
[30] *5G; NR; Physical Channels and Modulation*, Standard TS 38.211, V15.2.0, 3GPP, 2018.
[31] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Phys. Rev. A*, vol. 52, no. 5, pp. 3457–3467, Nov. 1995.
[32] A. Wiesel, Y. C. Eldar, and S. Shamai, "Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 653–656, Sep. 2005.
[33] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *J. Symbolic Comput.*, vol. 9, no. 3, pp. 251–280, Mar. 1990.
[34] F. G. S. L. Brandão, A. Kalev, T. Li, C. Y.-Y. Lin, K. M. Svore, and X. Wu, "Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning," *Int. Colloq. Automata, Lang., Program.*, vol. 132, pp. 27:1–27:14, Jul. 2019. [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.27
[35] G. Brassard, P. Hoyer, and A. Tapp, "Quantum counting," 1998, *arXiv:quant-ph/9805082*.
[36] A. Miranskyy, "Using quantum computers to speed up dynamic testing of software," 2022, *arXiv:2209.04860*.
[37] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, 2016.
[38] *MOSEK Optimizer API for Python 10.1.21*, 2023. [Online]. Available: https://docs.mosek.com/9.0/faq/faq.html#miscellanea

MASAYA NORIMOTO (Graduate Student Member, IEEE) received the B.E. degree and the M.E. degree in engineering science from Yokohama National University, Kanagawa, Japan, in 2022 and 2024, respectively. He received the IEEE VTS Tokyo/Japan Chapter Young Researcher's Encouragement Award in 2023. His research interests include quantum algorithms and wireless communications.

TAKU MIKURIYA (Graduate Student Member, IEEE) received the B.E. degree from Yokohama National University, Kanagawa, Japan, in 2024, where he is currently pursuing the M.E. degree with the Graduate School of Engineering Science. His research interests include quantum algorithms and optimization theory.

NAOKI ISHIKAWA (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2014, 2015, and 2017, respectively. In 2015, he was an Academic Visitor with the School of Electronics and Computer Science, University of Southampton, U.K. From 2016 to 2017, he was a Research Fellow with the Japan Society for the Promotion of Science. From 2017 to 2020, he was an Assistant Professor with the Graduate School of Information Sciences, Hiroshima City University, Japan. He is currently an Associate Professor with the Faculty of Engineering, Yokohama National University, Kanagawa, Japan. His research interests include quantum algorithms and wireless communications. He was certified as an Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS, in 2017 and 2021.