IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 6, 2019, accepted October 28, 2019, date of publication November 4, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950901

An Enhanced Tabu Search Based Receiver for Full-Spreading NOMA Systems

INSIK JUNG[1], (Student Member, IEEE), HYUNSOO KIM[1], (Student Member, IEEE), JINKYO JEONG[1], (Student Member, IEEE), SOOYONG CHOI[2], (Member, IEEE), AND DAESIK HONG[1], (Senior Member, IEEE)

[1]Information and Telecommunication Laboratory, Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea
[2]Advanced Communication Laboratory, Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Daesik Hong (daesikh@yonsei.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea Government (MSIT) (Development on the core technologies of transmission, modulation and coding with low-power and low-complexity for massive connectivity in the IoT environment) under Grant 2016-0-00181-004, and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) under Grant 2018R1A2A1A05021029.

ABSTRACT Full-spreading non-orthogonal multiple access (FS-NOMA) is one category of the candidate technologies designed to support massive connectivity in wireless communication systems. Before it can handle the massive volume of user connections, it is important for the FS-NOMA to develop a receiver that successfully decodes target data from non-orthogonally overlapped receiving signals. However, the decoding performance of conventional interference-cancellation (IC)-based receivers is far from optimal because of error-propagation problems. To improve the decoding performance, we propose a novel FS-NOMA receiver based on the tabu-search (TS) algorithm which is a sort of machine-learning algorithm. Specifically, a novel TS mechanism and a diversification scheme are proposed to overcome the inherent adverse conditions of FS-NOMA systems which lead the TS algorithm to local optima. Simulation results demonstrate that the proposed TS-based receiver has decoding performance that is superior to that of the conventional IC-based receiver. The results also show that the proposed receiver accommodates a higher number of user connections with a given packet drop rate threshold.

INDEX TERMS Non-orthogonal multiple access, full-spreading NOMA, massive connectivity, tabu-search.

I. INTRODUCTION

It has been forecasted from both industry and academy that the 'massive connectivity' will play a pivotal role in future wireless networks [1]–[4]. In industrial fields, general electric (GE) has estimated that 50 billion connected devices could create $10 trillion in monetary value for smart factories [5]. In cellular networks, the 3rd generation partnership project (3GPP) standardization group has defined massive machine type communications (mMTC) as one of the core scenarios of 5-th generation (5G) communication systems [6], [7].

Within the context of extending connectivity, non-orthogonal multiple access (NOMA) technologies have been investigated in the literature [8], [9]. Motivated by the fact that the connectivity of the conventional orthogonal multiple access (OMA) systems is restricted by the available number of orthogonal resources, NOMA systems intentionally allow multiple devices to access the same resource. NTT Docomo introduced power domain NOMA (PD-NOMA) as an initial attempt at non-orthogonal exploitation of communication resources [10]. The PD-NOMA focuses on extending the capacity region by manipulating the power of the transmitting devices [11]. Meanwhile, interleave-division multiple access (IDMA) has been investigated as a way to improve spectral efficiency by exploiting the diversity derived from the interleaving operation [12]. In particular, it has recently been revealed that IDMA with the elementary signal estimator (ESE) receiver can achieve capacity if the infinite length channel code is applied [13]. With an eye toward focusing more on extending connectivity, code domain non-orthogonal multiple access (CD-NOMA) technologies have been newly proposed not only in academia but also in corporate research [14]–[16]. The CD-NOMA

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

technologies employ multiple-access signatures based on spreading sequences designed to steer the mutual interference among superposed devices [17]–[20].

The CD-NOMA technology can be divided into two main categories according to the multiple access signature used to control multi-user interferences (MUI): sparse-spreading NOMA (SS-NOMA) and full-spreading NOMA (FS-NOMA) [15], [16], [21]. SS-NOMA reduces the number of superposed devices per single resource by adopting a sparse signature design [18], [20]. In contrast, FS-NOMA exploits low-correlation sequences for MUI mitigation [17].

The two types of CD-NOMA technologies are looking at different direction These two types of CD-NOMA technologies are looking in different directions for solutions on how to support massive connection. Specifically, SS-NOMA is suitable for scheduling-based systems where the user activities and the corresponding signatures are perfectly controlled by the base station. If the required control information is available, a near-optimal receiver based on a message-passing algorithm (MPA) can be applied with low computational complexity, thanks to its sparse signatures [20], [22]. On the other hand, FS-NOMA is profitable for grant-free systems. Since the mutual correlation of user signatures is low, the base station can easily detect the activeness and the signature of the devices without additional control signaling [23]–[25]. This overhead reduction further enhances resource efficiency in mMTC scenarios where uplink devices have small-sized data packets. However, research on advanced receivers for the FS-NOMA systems is still in its infancy.

There have been a few related works of designing non-linear receivers for the FS-NOMA systems. Yuan *et al.* [17] proposed a multi-user receiver based on a successive interference cancellation (SIC) algorithm. However, the SIC receiver has a chronic error propagation problem which degrades multi-user detection performance. In addition, this propagation occurs more frequently in NOMA systems since the initial linear solutions are inaccurate. To cope with this error propagation phenomenon, Eid *et al.* [26] proposed a decoding method involving signal-to-interference-plus-noise-ratio (SINR)-ordered SIC as a way to reduce the error occurrence of the earlier stream's decoding process. Parallel interference cancellation (PIC) receivers have also been proposed to iteratively improve the decoding performance [24], [25]. However, the approaches in [17], [24]–[26] for FS-NOMA systems are far from achieving the optimal symbol-level detection performance because of the error propagation problem. The well-known receiver to achieve optimal symbol-level detection performance is maximum-likelihood (ML) receivers. However, the ML receivers require prohibitively high complexity, making their implementation impractical [27], [28]. More research is therefore needed to develop a near-optimal receiver for FS-NOMA systems that has bearable complexity, which will in turn increase the possible number of user connections in practical systems.

Meanwhile, there have been numerous advanced receivers proposed for multi-input-multi-output non-orthogonal multiple access (MIMO-NOMA) systems to help achieve the theoretical capacity [13], [29]–[35]. If we treat the multiple antennas of the MIMO-NOMA systems as the resources to be spreaded in FS-NOMA systems, the signal model of both systems are similar. In [13] and [33], novel capacity-achieving iterative linear-minimum-mean-squared-error (LMMSE) receivers for the MIMO-NOMA system are investigated. Specifically, the authors proposed a coded parallel-interference-cancellation (PIC) method. Thanks to the joint design combining a multi-user code, a detector and a decoder, the detection performance of the proposed receiver is within a mere 0.8dB of achieving the capacity limit of the MIMO-NOMA system. In Liu *et al.* [29] investigated the achievable rate of the approximate message passing (AMP) receiver for coded systems in the MIMO environment. The authors showed that AMP-based receivers can approach the constrained capacity limit of the MIMO systems.

Several attempts have been made to alleviate the computational complexity of the iterative MMSE-based methods. In particular, in massive multi-user MIMO (MU-MIMO) systems, Gaussian message passing (GMP)-based iterative detection schemes have been studied for underloaded systems [30] and overloaded systems [31], [32], respectively. The authors of [30]–[32] proved that the proposed schemes always converge to the LMMSE performance with low complexity. In [36], concise expectation propagation-based MPA (EP-MPA) was proposed as a way to solve non-linear problems more efficiently compared to the GMP algorithms. On the other hand, there has been active research work based on the AMP methods such as orthogonal AMP (OAMP) [37], vector AMP (VAMP) [38]–[40], and bilinear generalized AMP (BiGAMP) [38], [39], etc.

However, the receivers for the MIMO-NOMA systems target access scenarios that are different from those that FS-NOMA systems target. While the FS-NOMA systems consider grant-free scenarios where only constrained information is given to the base stations, the MIMO-NOMA receivers proposed in [13], [29]–[40] require information about the user activities and the scheduling grant. Moreover, the MIMO-NOMA receivers are designed under the assumption that code lengths are infinite, with the goal of achieving the theoretical capacity. However, this assumption is not valid for FS-NOMA systems, since it exploits short length packets in order to comply with the mMTC scenario [41], [42].

Meta-heuristic algorithms in artificial intelligence have been studied as a way to solve problems where the brute-force search complexity is extremely high [43]–[46]. Among them, the tabu search (TS) algorithm shows near-optimal performance with low complexity when used to resolve several combinatorial optimization problems [47], [48]. In particular, the TS algorithm starts with a rough initial solution and conducts a local neighborhood search in a successive manner as it tries to find the global optimal solution. In this paper, motivated by the fact that the FS-NOMA detection problem

is one of those sorts of combinatorial optimization problems whose solution spaces are composed of quantized symbol candidates, we propose a novel FS-NOMA receiver based on the tabu search algorithm to obtain near-optimal performance with low complexity.

However, FS-NOMA systems have inherent conditions that are not conducive to the use of the tabu search approach. First, the initial solution is likely to be located very far from the global optimal solution due to the MUI caused by the non-orthogonality. Furthermore, the receiver needs to explore a broad solution space because of the user overloading, since the dimension of the solution vector increases as the number of users increases. The aforementioned factors create a problem where the TS algorithm ends before reaching the global optimal solution. To find a solution that is closer to the global optimum, we propose improving the TS algorithm-based FS-NOMA receiver in the following two ways:

- Since the initial solution vector is inaccurate, we propose a restart algorithm that performs the TS algorithm with another initial solution to 'diversify' the searching area if the current solution vector is judged not to be the global optimal.
- TS prohibits recently conducted 'moves' (the direction from a current solution to the next solution) from leading the algorithm in unvisited directions. However, this move-based prohibition cannot prevent the repetition of solutions because there are multiple moves which bound to the same solution vector. Therefore, we propose a dual-control-based prohibition scheme that considers both moves and solutions together.

The remainder of this paper is organized as follows: in section II, an uplink FS-NOMA system model and the transceiver structure is introduced. In section III, the fundamentals of TS algorithm for the FS-NOMA multi-user detection problem is introduced. Thereafter, the dual-control-based TS algorithm and the diversification scheme is proposed. Simulation results are presented in section IV and the conclusion is provided in section V.

Throughout this paper, the following notations will be used. Lowercase letters a, bold lowercase letters $\mathbf{a}$ and bold uppercase letters $\mathbf{A}$ denote scalars, column vectors and matrices, respectively. Also, calligraphic letters $\mathcal{A}$ represent sets. The superscripts $(\cdot)^H$ denotes the matrix Hermitian and $(\cdot)^T$ is matrix transpose. $\mathcal{A}\backslash k$ means the set $\mathcal{A}$ with element k being excluded.

II. UPLINK FS-NOMA SYSTEM MODEL AND THE TRANSCEIVER STRUCTURE

Let us consider the uplink FS-NOMA system with K users and M orthogonal resource blocks (PRBs) as shown in Fig. 1. In NOMA systems, normally $K > M$ and we define the user overloading factor ρ as:

$$\rho = K/M \times 100(\%). \tag{1}$$

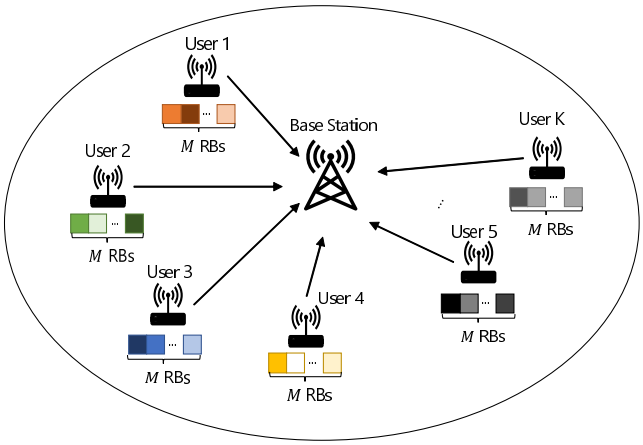


FIGURE 1. Uplink FS-NOMA system model.

For each user, the input block of the channel encoder $\mathbf{u}_k$ includes N_{I} information bits and N_{CRC} cyclic redundancy check (CRC) bits. After the channel encoding with a code rate R, the coded block is denoted by $\mathbf{b}_k$. Subsequently, the coded block is modulated with a modulation alphabet $\mathcal{A} = \{\dashv_\infty, \dashv_\in, \ldots, \dashv_Q\}$ whose cardinality $|\mathcal{A}|$ is Q. Note that each user now has $J = (N_I + N_{\mathrm{CRC}})/R/\log_2 Q$ modulated symbols $\mathbf{x}_k = [x_{k,1}, \ldots, x_{k,J}]^T$ to transmit. To clarify the explanation regarding symbol-level operations, let us consider the symbol vector $\mathbf{x} = [x_1, \ldots, x_K]^T$ which contains one modulated symbol per user.

For the k-th user, the modulated symbol x_k is spreaded into M dimensional sequence $\mathbf{c}_k = [c_1, \ldots, c_M]^T$. The spreading sequences are randomly selected from the predefined codebook, whose elements are determined by the target NOMA system's own signature. For example, multi-user shared access (MUSA) systems exploit lattice-shaped complex sequences [17] and non-orthogonal coded access (NOCA) systems [24] adopt long-term evolution (LTE) Zad-off-chu sequences. Afterwards, all the K users' transmitting vectors are multiplexed over M shared PREs.

Taking into account the M-dimensional wireless channel vector $\mathbf{h}_k = [h_1, \ldots, h_M]^T$ of the k-th user, we derive the M by 1 received signal vector $\mathbf{y}$ as:

$$\mathbf{y} = \sum_{k=1}^{K} (\mathbf{h}_k \circ \mathbf{c}_k) x_k + \mathbf{n} = (\mathbf{H} \circ \mathbf{C})\mathbf{x} + \mathbf{n} = \mathbf{G}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\circ$ denotes an element-wise product operator and $\mathbf{n}$ is the noise vector whose entries are modeled as $i.i.d$ with zero mean and standard deviation σ. Also, note that we adopt the equivalent matrix $\mathbf{G}$ that reflects the effect of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$ and sequence $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K]$ concurrently.

On the receiver side, the M-dimensional data symbol vector $\hat{\mathbf{x}}$ is reconstructed from the FS-NOMA receiver. For example, conventional FS-NOMA receivers use LMMSE based successive interference cancellation (LMMSE-SIC) or LMMSE-PIC. Based on the estimated symbol vector,

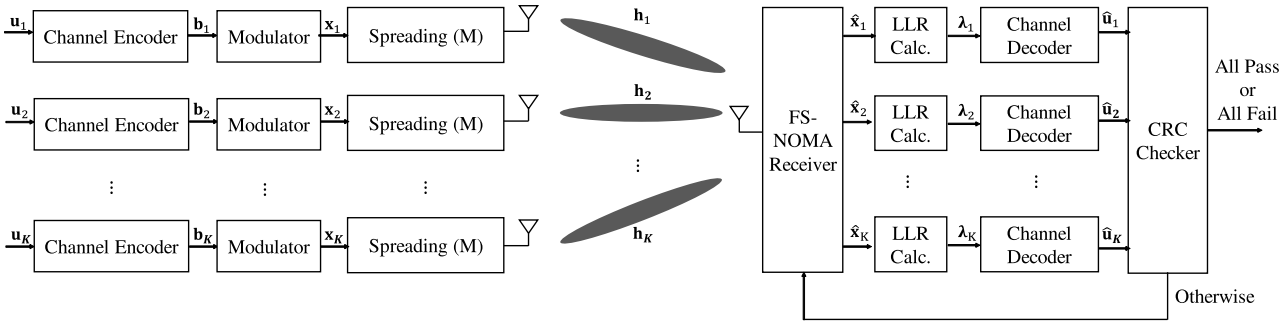


FIGURE 2. Block diagram of full-spreading non-orthogonal multiple access technologies.

the log-likelihood ratios (LLRs) of the k-th user's bits $\lambda_k = [\lambda_{1,k}, \ldots, \lambda_{\log_2 Q,k}]$ need to be computed.

The LLRs of the user bits should be derived from the hard-decisioned detection result of the FS-NOMA receiver. The maximum a posteriori probability (MAP) value $\lambda_{p,k}$ is shown as below:

$$\lambda_{p,k} = \ln\left(\frac{P(b_{k,p}=1|\mathbf{y})}{P(b_{k,p}=0|\mathbf{y})}\right), \tag{3}$$

where $b_{k,p}$ is the p-th bit of the user k to be transferred into LLR. From the hard-decisioned detection result $\hat{\mathbf{x}}$, the $\lambda_{p,k}$ can be approximated as:

$$\lambda_{p,k}(\hat{\mathbf{x}}) = -\frac{1}{2\sigma^2} \cdot (||\mathbf{y} - \mathbf{G}(\hat{\mathbf{x}} \leftarrow (b_{k,p}=1))||^2 - ||\mathbf{y} - \mathbf{G}(\hat{\mathbf{x}} \leftarrow (b_{k,p}=0))||^2), \tag{4}$$

where the $\varphi(\hat{\mathbf{x}})$ is the cost function of the arbitrary vector $\hat{\mathbf{x}}$ which can be written as:

$$\varphi(\hat{\mathbf{x}}) = ||\mathbf{y} - \mathbf{G}\hat{\mathbf{x}}||^2. \tag{5}$$

The derivation process for (4) is provided in Appendix A. In (4), the operation $\leftarrow (b_{k,p} = 0)$, will be referred to as 'bit substitution'. The operation $\leftarrow (b_{k,p} = 0)$ substitutes the k-th user's p-th bit into 0 and generates the symbol again, reflecting the modified bit. For $\leftarrow (b_{k,p} = 1)$, a similar operation is performed with 1. The detailed bit substitution process is also described in Appendix A. If the entire received block of user k is converted to the LLR, the channel decoder estimates $\hat{\mathbf{u}}_k = [\hat{u}_1, \ldots, \hat{u}_{N_I+N_{\mathrm{CRC}}}]$.

We will consider the iterative detection scheme in [24], [17] for separating multi-user signals as shown in Fig. 1. Every user block $\hat{\mathbf{u}}_k$ is individually checked by the CRC. If a decoded block of a user k passes the CRC, the receiver judges that decoding is successful and eliminates the k-th user's effect from the received multi-user signal. After eliminating all successful user signals, the receiving procedure goes back to the FS-NOMA receiver with only the users who failed to pass the CRC. The iterative receiver works until all users pass the CRC or all users fail the CRC. The transceiver structure is summarized in Fig. 2.

III. PROPOSED ENHANCED TABU SEARCH (E-TS) BASED FS-NOMA RECEIVER

In this section, we propose a novel TS-based FS-NOMA receiver. Specifically, we first present the fundamental concepts (neighborhood, move, and tabu) of the tabu search algorithm in FS-NOMA systems. We then propose an enhanced tabu search (e-TS) algorithm which exploits a novel tabu strategy called 'dual control' and a diversification method. Based on the e-TS algorithm being proposed here, we perform symbol level decoding, which takes place in the block identified as "FS-NOMA receiver" in Fig. 2.

The purpose of the symbol-level detector is to find a symbol vector $\hat{\mathbf{x}}$ whose cost function (5) is the smallest among the candidate symbol vectors. In this context, the global optimal solution is defined as Definition 1.

Definition 1 (Global Optimal Solution): The solution vector $\mathbf{x}^*$ is a globally optimal solution if

$$\varphi(\mathbf{x}^*) \leq \varphi(\mathbf{x}) \tag{6}$$

for all K-tuple solution vector $\mathbf{x}$ where $\mathbf{x}(j) \in \mathcal{A}$ for $j = 1, 2, \ldots, K$.

Since the complexity of the ML detector was proven to be an NP-hard [28], our attempt focused on obtaining a near-optimal solution with low complexity based on the e-TS algorithm proposed below.

A. FUNDAMENTALS OF THE TABU SEARCH ALGORITHM

The basic idea of TS-based FS-NOMA decoding can be described as follows: The algorithm starts with an initial solution vector. Like other receivers for FS-NOMA, the TS algorithm selects one of the linear solutions (e.g., LMMSE, zero-forcing (ZF), etc.) as an initial vector. From the initial solution, we define its 'neighborhood' vectors based on the Euclidean distance. Thereafter, the algorithm 'moves' to the best vector among the neighborhoods. To avoid falling into local optimal solutions, the TS algorithm adopts the 'tabu' scheme when it commits moves. Specifically, the preceding moves are stored in the memory to check whether the current move was already conducted in the past iterations. These procedures are repeated until the stopping criterion is satisfied.

The factors that are essential to implementing the tabu search algorithm are described below. In particular,

we provide the precise definitions for the neighborhood, move, and the tabu scheme, respectively.

1) NEIGHBORHOOD

We begin by defining the neighborhoods of a K-dimensional arbitrary solution vector $\hat{\mathbf{x}} = [\hat{x}_1, \cdots, \hat{x}_K]^T$. As the name suggests, neighborhoods are the candidate solution vectors whose Euclidean distances from the current solution vector are small. We start by defining symbol neighborhoods and then proceed to define a vector neighborhood based on the symbol neighbors.

Let us define the symbol neighborhood for an arbitrary symbol $a_q \in \mathcal{A}$, $q = 1, 2, \ldots, Q$, based on the Euclidean distance. Specifically, we define a set of symbols $\mathbf{s}_{a_q}$ whose elements are the symbols with the N smallest Euclidean distance from the symbol a_q. We can then proceed to define the vector neighbors of the arbitrary solution vector. We define a vector neighbor by replacing the symbol of one user in the solution vector with its symbol neighbor. Since the solution vector $\hat{\mathbf{x}}$ contains K users and each user's symbol has N symbol neighbors, there are KN possible vector neighbors. We express all of the vector neighbors of the solution vector $\mathbf{x}$ in matrix form:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \cdots & \underbrace{\mathbf{v}_{(k-1)K+n}}_{k^{th}\ UE's\ n^{th}\ vector\ neighbor} & \cdots \mathbf{v}_{KN} \end{bmatrix},$$

$$where\ \mathbf{v}_{(k-1)N+n}(j) = \begin{cases} x_k & for\ j \neq k \\ \mathbf{s}_{x_k}(n) & for\ j = k. \end{cases} \tag{7}$$

Based on the definition of the vector neighborhoods, local optimal solution can be defined, as provided in Definition 2.

Definition 2 (Local Optimal Solution): Let $\mathcal{N}(\mathbf{x})$ be a neighborhood function, whose output is a set of all vector neighbors $\mathbf{v}_p$ $(p = 1, 2, \ldots, KN)$ derived from the K-tuple input vector $\mathbf{x}$. A solution vector $\mathbf{x}^*$ is locally optimal if

$$\varphi(\mathbf{x}^*) \leq \varphi(\mathbf{v}_p), \tag{8}$$

for all $\mathbf{v}_p \in \mathcal{N}(\mathbf{x})$.

2) MOVE AND TABU

A 'MOVE' is an operation where the algorithm determines the solution vector for the next iteration from among the vector neighbors of the current solution vector. The definition of 'MOVE' is described in Definition 1.

Definition 3 (MOVE): In each iteration, the next solution vector is selected from among the $\mathbf{x}$'s vector neighborhoods, which are the column vectors of the matrix $\mathbf{V}$. As a reminder, each vector neighbor actually replaces a single user's symbol into one of the symbol neighbors. Thus, we define the MOVE as (k, n, q) if the algorithm selects the n-th symbol neighborhood of the k-th user's current symbol a_q. An example of the MOVE operation is given in Fig. 3.

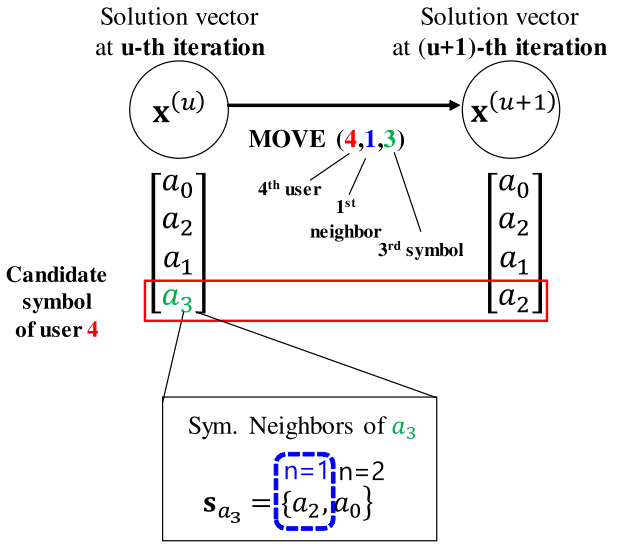


FIGURE 3. Example of the 'MOVE' operation explained in Definition 3. In this example, we show the MOVE (4,1,3) when the number of users $K = 4$, the number of symbol neighbor $N = 2$ and the modulation alphabet $\mathcal{A} = \{a_0, a_1, a_2, a_3\}$.

Let us define the set of MOVEs $\mathcal{V}$ whose elements indicate MOVEs which are mapped 1-to-1 to the vector neighbors of the current solution vector. In each iteration, the TS algorithm only evaluates and selects the next MOVE among the elements of the set $\mathcal{V}$.

The most straightforward way to determine a MOVE is to select a next solution whose likelihood cost is the smallest among the vector neighbors. However, the algorithm cannot escape local optimal solutions since they have a smaller likelihood cost than any other adjacent solution vectors. To overcome this problem, the TS algorithm exploits the 'TABU' mechanism to escape these local optimal solutions. Specifically, the algorithm memorizes the past MOVEs at the memory and prohibits those MOVEs for P subsequent iterations. Note that P is a predefined parameter which indicates the number of iterations to be prohibited for the memorized MOVE.

The TABU mechanism is performed based on a TABU MOVE matrix $\mathbf{T}_m$. The TABU MOVE matrix $\mathbf{T}_m$ is an $N \times KQ$ matrix that contains all the possible MOVEs. The element of the $\mathbf{T}_m$ denotes the number of iterations to be banned for the MOVE. Accordingly, the TABU MOVE matrix could be written as:

$$\mathbf{T}_m = \begin{bmatrix} t_{1,1} & \cdots & t_{1,(k-1)Q+1} & \cdots & t_{1,kQ} & \cdots & t_{1,KQ} \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ t_{N,1} & \cdots & t_{N,(k-1)Q+1} & \cdots & t_{N,kQ} & \cdots & t_{N,KQ} \end{bmatrix}.$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{MOVEs\ from\ k^{th}\ UE's\ symbol\ candidates}$$

(9)

Note that KQ columns indicates the symbol candidates of the K users and N rows indicates their corresponding neighbors. In other words, a TABU value for the MOVE (k, n, q) is

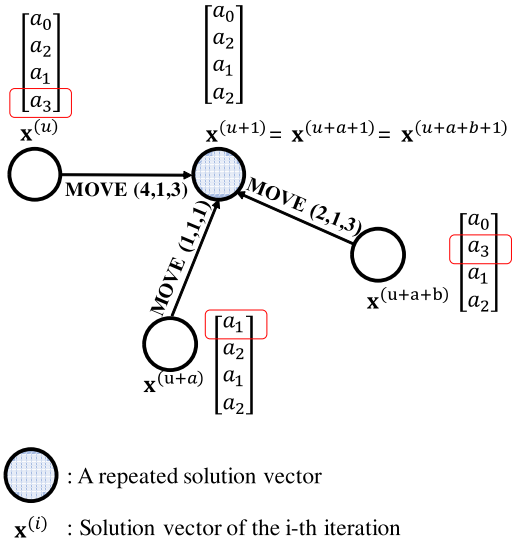


FIGURE 4. Solution repetition problem of the MOVE-based TABU mechanism. Note that the different moves (4,1,3), (1,1,1) and (2,1,3) lead the algorithm toward the same solution vector $\mathbf{x}^{(u+1)}$.

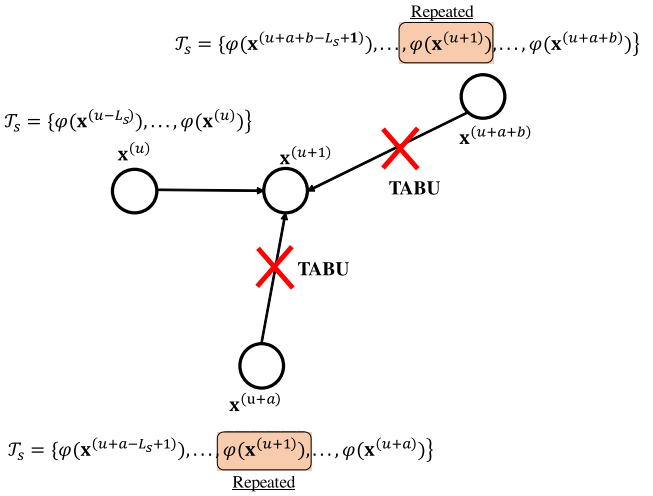


FIGURE 5. Example of the solution-controlled tabu mechanism. The algorithm exploits the past solution lists $\mathcal{T}_s$ to avoid MOVEs which bound to the same solution vector.

represented on the $(n, (k-1)Q+q)$-th entry of the $\mathbf{T}_m$ matrix.

B. PROPOSED DUAL-CONTROL BASED TABU SEARCH ALGORITHM

Note that the conventional tabu mechanism prohibits past 'MOVEs', which can be considered the 'direction' from one solution to another solution. However, as shown in Fig. 4, this strategy cannot prevent the algorithm from finding solution vectors that were already selected in past iterations. This is because there are several different MOVEs which are heading to the same solution vectors. If the same solution vector were repeated, the likelihood costs of its vector neighbors would have to be calculated multiple times, which is a waste of computation.

To overcome this drawback to the MOVE based tabu strategy, we propose a novel tabu method based on dual controlling. In particular, the proposed scheme utilizes the history of past solutions as well as the past MOVEs to determine the next solution vector. To avoid repetitive solutions, the proposed scheme additionally exploits a short-term memory $\mathcal{T}_s$ which is defined as follows: The set of TABU solutions $\mathcal{T}_s$ contains the solution vectors of past iterations. Specifically, the TABU solution set memorizes the likelihood cost of the past solution vectors. The set size cannot exceed the predefined amount L_s and the elements are updated in a first-in-first-out (FIFO) manner. Thus, the solution vector at the u-th iteration is memorized as

$$\mathcal{T}_s(mod(u, L_s)) = \varphi(\mathbf{x}^{(u)}), \tag{10}$$

where the operator $mod(a, b)$ returns a remainder which is obtained by dividing a by b. Consequently, as depicted in Fig. 5, we store the recent solution vectors in the memory so that the algorithm avoids solution repetitions.

The operation of the proposed dual-controlled TS algorithm consists of five processes: neighborhood evaluation, the best MOVE selection, TABU with dual controlling, updating parameters and checking the stopping criteria. The process is summarized in full in Fig. 6.

The parameters used in the proposed algorithm are as follows:

- $\mathbf{x}_b^{(u)}$: The best solution vector that has the smallest ML cost until the u-th iteration of the algorithm.
- cnt: The number of iterations with unchanged best solution vector.
- $\mathbf{V}^{(u)}$: Vector neighbor matrix of the k-th solution vector $\mathbf{x}^{(u)}$.
- $\mathcal{V}^{(u)}$: the set of neighboring MOVEs for the u-th solution vector.

The algorithm starts with an initial solution $\mathbf{x}^{(0)}$. Before the algorithm begins, we set the following initial conditions: $\mathbf{x}_b^{(0)} = \mathbf{x}^{(0)}, \mathbf{T}_m = 0_{[N\times KQ]}$, $\mathcal{T}_s = \phi$, $cnt = 0$. Without loss of generality, let us consider the u-th iteration.

(Step 1) Constructing Vector Neighborhoods: From the current solution vector $\mathbf{x}^{(u)}$, we calculate the neighborhood matrix $\mathbf{V}^{(u)}$. Based on the $\mathbf{V}^{(u)}$, the set of neighboring MOVEs $\mathcal{V}^{(u)}$ is generated.

(Step 2) The Best MOVE Section: In this step, we determine the best MOVE (k', n', q') whose likelihood cost is the smallest among the set of neighboring MOVEs, $\mathcal{V}^{(u)}$:

$$(k', n', q') = \underset{(k,n,q)\in\mathcal{V}^{(u)}}{\arg\min}\ \varphi(\mathbf{v}^{(u)}_{(k-1)N+n}). \tag{11}$$

If all MOVEs are prohibited so that $\mathcal{V}^{(u)}$ is empty, select the MOVE whose prohibited amount is the smallest from among the vector neighborhoods.

(Step 3) TABU with Dual Control: In this step, the selected MOVE (k', n', q') is checked by the proposed TABU scheme based on the dual controlling. The dual

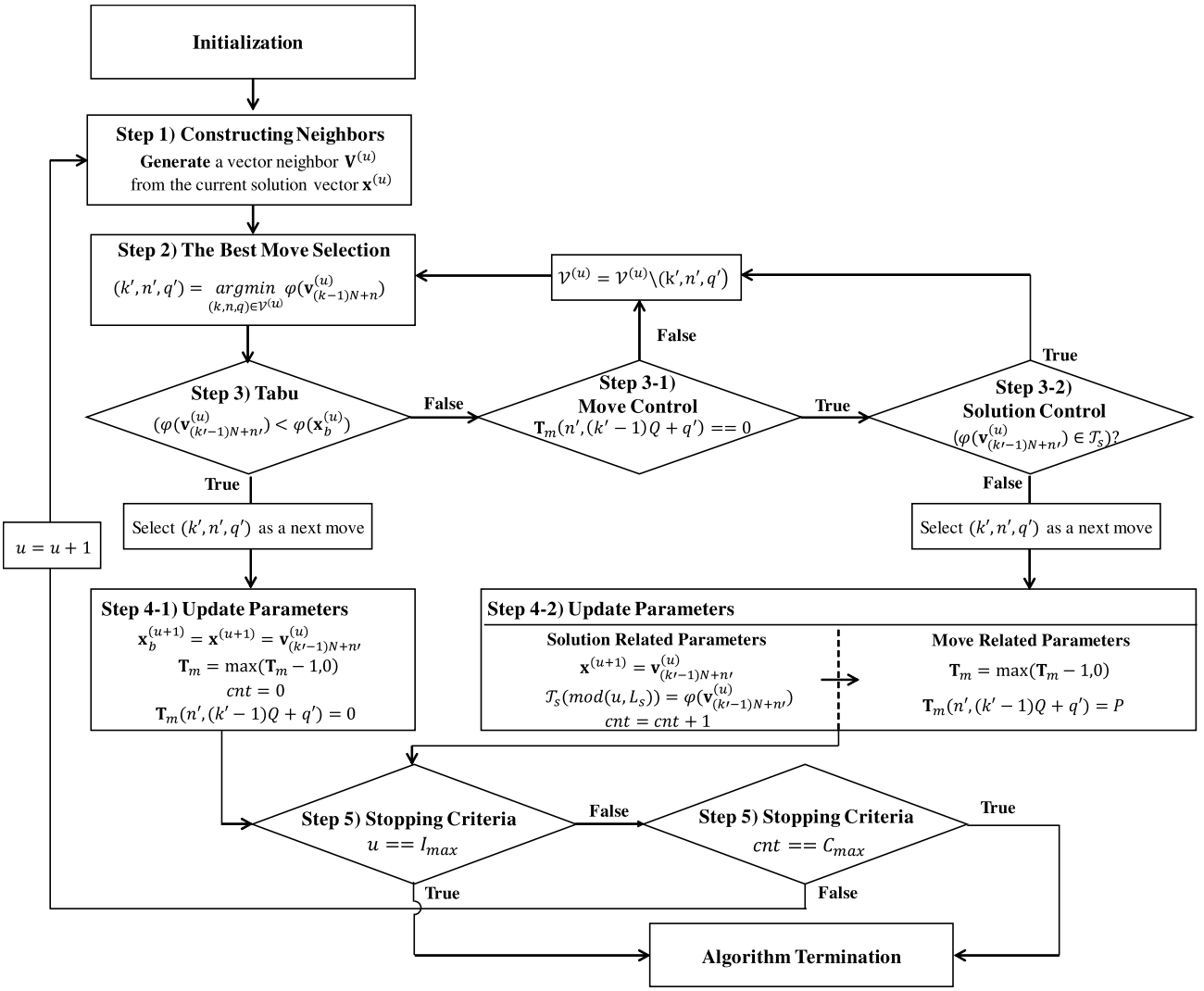


FIGURE 6. Low chart summarizing the entire procedure of the proposed dual-controlled tabu search algorithm in section III-B. Rectangular-shaped blocks denote 'process' operations and diamond-shaped blocks depict 'decision' operations.

controlling scheme consists of two sub-steps: 3-1) prohibiting repetitive MOVEs, and 3-2) prohibiting repetitive solutions.

Meanwhile, if the best solution vector up to the current iteration can be updated so that (12) is satisfied, the MOVE is not prohibited and the algorithm proceeds directly to step 4-1:

$$\varphi(\mathbf{v}^{(u)}_{(k'-1)N+n'}) < \varphi(x^{(u)}_b). \tag{12}$$

(Step 3-1) Prohibiting Repetitive MOVEs: Check whether the MOVE (k', n', q') was conducted within the recent P iterations. If the following condition is satisfied, the MOVE is not prohibited since it has not been performed recently, and the algorithm proceeds to step 3-2:

$$\mathbf{T}_m(n', (k'-1)Q + q') = 0, \tag{13}$$

Otherwise, exclude the repetitive MOVE (k', n', q') from the neighborhoods by setting $\mathcal{V}^{(u)} = \mathcal{V}^{(u)} \backslash (k', n', q')$ and go back to step 2. Note that the $\backslash$ operator erases the element from the set.

(Step 3-2) Prohibiting Repetitive Solutions: Check whether the MOVE (k', n', q') leads the algorithm to a vector neighbor that has been selected recently. If $\mathbf{v}^{(u)}_{(k'-1)N+n'} \in \mathcal{T}_s$, conduct $\mathcal{V}^{(u)} = \mathcal{V}^{(u)} \backslash (k', n', q')$ and the algorithm goes back to step 2. Otherwise, the algorithm proceeds to step 4-2.

(Step 4) Updating Parameters: From the next MOVE (k', n', q') and next solution vector $\mathbf{v}^{(u)}_{(k'-1)N+n'}$, parameters are updated in two different ways: 4-1) When the best solution vector can be updated, 4-2) When the best solution vector cannot be updated.

The common operations are as follows:

$$\begin{aligned} \mathbf{x}^{(u+1)} &= \mathbf{v}^{(u)}_{(k'-1)N+n'}, \\ \mathcal{T}_s(mod(u, L_s)) &= \varphi(\mathbf{v}^{(u)}_{(k'-1)N+n'}), \\ \mathbf{T}_m &= \max(\mathbf{T}_m - 1, 0). \end{aligned} \tag{14}$$

(Step 4-1) The Best Solution Vector Can be Updated: Since the best solution is being replaced, the algorithm judges that the current direction is an option and doesn't forbid the MOVE. In addition, the TABU amount of the MOVE is set to 0 even though the MOVE was prohibited in the recent iterations. In this regard, the parameters are updated in the following way:

$$\begin{aligned} \mathbf{x}_b^{(u+1)} &= \mathbf{v}_{(k'-1)N+n'}^{(u)}, \\ \mathbf{T}_m(n', (k'-1)Q+q') &= 0, \\ cnt &= 0. \end{aligned} \tag{15}$$

(Step 4-2) The Best Solution Vector Cannot be Updated: In this case, the algorithm prohibits the MOVE for the next P iterations to explore different directions. The parameters are updated as:

$$\begin{aligned} \mathbf{T}_m(n', (k'-1)Q+q') &= P, \\ cnt &= cnt+1. \end{aligned} \tag{16}$$

(Step 5) Checking the Stopping Criteria: We employ two parameters to determine the termination of the TS algorithm: I_{max} and C_{max}. If the iteration number u becomes equal to I_{max}, the algorithm definitively terminates. Otherwise, $u = u+1$ and then check to determine whether cnt is equal to C_{max}. If so, terminate the algorithm. If not, go back to step 1.

C. DIVERSIFICATION STRATEGY

In this subsection, we propose a diversification scheme to further improve the decoding performance. The goal of the proposed diversification is to perturb the algorithm to the unvisited area in the solution space, thereby increasing the possibility of finding a solution vector which is closer to the global optimal solution. To accomplish this goal, we first establish the conditions under which the algorithm requires diversification and then determine which solution vector to use to perturb the algorithm.

To determine if the current solution requires diversification, we will utilize the information from the TS algorithm in the previous section. Specifically, we first check whether or not the best solution has been changed for C_{max} iterations (which means $cnt = C_{max}$). If so, we judge that the algorithm satisfies the first condition for the diversification scheme because the condition implies that it is difficult to find a better solution near the current search area.

After this, we examine the likelihood cost of the current best solution vector. If the following condition is fulfilled, we perform diversification:

$$\varphi(\mathbf{x}_b^{(u)}) > d_{th}, \tag{17}$$

where d_{th} is a predefined threshold that strikes a good balance between performance and computational complexity. We determine the value of d_{th} as follows: The purpose of (17) is to ascertain whether the current solution is the same as the transmitted vector or not. Thus, let us assume that $\mathbf{x}_b^{(u)}$ is the right solution vector; the distribution of $\varphi(\mathbf{x}_b^{(u)})$ then becomes the square sum of the received M size noise vector distribution whose elements follow $i.i.d$ complex gaussian distribution with variance σ^2. In other words, $\varphi(\mathbf{x}_b^{(u)})$ follows the Gamma distribution $\Gamma(M, \sigma^2)$ with mean $M\sigma^2$ and variance $M\sigma^4$. We consider the current solution to be incorrect if the likelihood cost deviates from the mean value by more than the standard deviation, i.e.:

$$d_{th} = M\sigma^2 + \sqrt{M}\sigma^2. \tag{18}$$

If (17) is satisfied, the algorithm judges that the current solution is not the global optimum and that diversification is needed.

In FS-NOMA systems, we can easily obtain linear solutions from the equivalent matrix $\mathbf{G}$ with low computational complexity. Inspired by the fact that these linear solutions have a greater chance of being located near the global optimal solution than random solutions, we can exploit the linear solutions for the diversification point to ensure the fundamental quality of the starting search space. Based on these insights, we provide the detailed operations in Algorithm 1 below.

Algorithm 1 Diversification Algorithm

Input:
  Initial solution: $\mathbf{x}_{ini}$
  Initial solution for a diversification: $\mathbf{x}_{div}$
Output:
  Detected symbol vector: $\mathbf{x}_{fin}$
run the proposed TS algorithm in III-B
if ($\varphi(\mathbf{x}_b^{(u+1)}) < d_{th}$ and $u+1 < I_{max}$) **then**
  $\mathbf{x}_{fin} = \mathbf{x}_b^{(u+1)}$
else
  Re-initialize parameters
    $\mathbf{x}^{(u+1)} = \mathbf{x}_{div}$, $\mathbf{T}_m = 0_{[N\times KQ]}$, $\mathcal{T}_s = \phi$, $cnt = 0$
  run the proposed TS algorithm in section III-B
  $\mathbf{x}_{fin} = \mathbf{x}_b^{(u+1)}$
end if

D. COMPLEXITY ANALYSIS

In this subsection, we analyze the complexity of the proposed e-TS based FS-NOMA receiver. The number of floating-point operations (FLOPs) is used to evaluate the complexity. Note that a complex multiplication needs six real operations while a complex summation needs two real operations. Also, we consider the computational cost of the comparison operator as a two real operation [49]. The complexities of LMMSE-SIC and ML receivers are used for comparison.

Before the analysis begins, we first compute the cost of several operations which will be repeatedly used in this section. For a LMMSE receiver, the computation cost of the equalizer $\mathbf{W}_M$ must be calculated:

$$\mathbf{W}_M = (\mathbf{G}^H\mathbf{G} + \frac{1}{SNR}\mathbf{I})^{-1}\mathbf{G}^H. \tag{19}$$

For the FLOPs for matrix operation, we refer to [50]. The matrix multiplication and summation operations in the round brackets require $8K^2M + 2KM$ FLOPs. Afterwards, $4K^3 +$

$8K^2$ FLOPs are used for the inverse operation. Multiplying $\mathbf{G}^H$ requires an additional $8K^2M - 2KM$ FLOPs. The equalization consumes $8KM - 2K$ FLOPs. The total cost for a LMMSE receiver is $4K^3 + 8(2K^2 + K)M + 8K^2 - 2K$. For a ZF receiver, the computational cost of the equalizer $\mathbf{W}_Z$ must be calculated:

$$\mathbf{W}_Z = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H. \tag{20}$$

Similarly, the computational cost of a ZF receiver is $4K^3 + 8(2K^2 + K)M + 6K^2 - 2K(M + 1)$. The operation to find the maximum or minimum element from N_t real-valued elements needs $2(N_t - 1)$ FLOPs.

We first analyze the amount of computation required for each step of the proposed e-TS algorithm. In step 1, calculating the likelihood costs of the vector-neighborhood matrix $\mathbf{V}^{(u)}$ involves floating point computations. The number of FLOPs for a likelihood cost calculation (9) is $8KM + 8M - 2$. Since $\mathbf{V}^{(u)}$ consists of KN elements, step 1 requires $8K^2MN + 8KMN - 2KN$ FLOPs. In step 2, the argmin operation with KN elements consumes $3(KN - 1)$ FLOPs. In step 3, calculating (9) requires 3 FLOPs. In the following sub-steps, step 3-1 requires 3 FLOPs to examine (10) and step 3-2 involves $3L_s$ FLOPs. In step 4, the cost of updating the TABU MOVE matrix $\mathbf{T}_m$ is included. Specifically, NKQ FLOPs are needed to compute $\mathbf{T}_m - 1$, and $3NKQ$ FLOPs are needed to operate the max operation in (11). Also, in step 4-2, updating the variable *cnt* imposes 1 FLOP.

The complexity of the proposed algorithm is not deterministic. To be specific, in every iteration, the algorithm goes through different sub-steps based on which conditions are met. Thus, we derive the lower and upper bound for the complexity of the proposed method, respectively.

Lower Bound: In this case, we consider the total number of iterations of the algorithm to be C_{max}, because it is the smallest number of iterations needed for the algorithm to terminate. During the iterations, each step proceeds as follows:

- *Initial Solution:* This operation is conducted only once at the beginning of the algorithm. Obtaining the LMMSE solution requires $4K^3 + 8(2K^2 + K)M + 8K^2 - 2K$ FLOPs.
- *Step 1:* This step is performed once for every iteration. Thus, a total of $C_{max}KN(8KM + 8M)$ FLOPs are required.
- *Steps 2 and 3:* These steps are performed once for every iteration. We consider the situation where the algorithm always selects next vectors whose MOVE and corresponding solution are not prohibited. Thus, both step 2 and step 3 are not repeated so that a total of $C_{max}(2KN + 2L_s + 2)$ FLOPs are needed for these procedures.
- *Step 4:* This step is performed once for every iteration. The best solution should not be replaced so that the algorithm ends in C_{max} iterations. Thus, not step 4-1 but step 4-2 is always performed. The total cost for this procedure is $C_{max}(3NKQ - 1)$.
- *Step 5:* This step is performed once for every iteration. For C_{max} iterations, 5 FLOPs (2 comparison and 1 real addition) are computed per iteration. For one iteration, 4 FLOPs (2 comparison) are required.
- *Diversification:* In this case, we consider a situation where the algorithm always fails to satisfy (14) and does not perform diversification.

Upper Bound: When it comes to the upper bound case, the algorithm works until u reaches the maximum iteration quantity I_{max}.

- *Initial Solution:* Same as for the lower bound case above.
- *Step 1:* Same as for the lower bound case above.
- *Steps 2 and 3:* Note that if the MOVE selected in step 2 is prohibited by the TABU test in step 3, the algorithm goes back to step 2. In this regard, we consider the worst case where step 2 and step 3 are repeated KN times (until the set $\mathcal{V}$ is empty) in a every single iteration. Thus, (18) FLOPs are consumed in a single iteration:

$$\sum_{i=1}^{KN}(\underbrace{2(i-1)}_{Step2} + \underbrace{2(L_s+2)}_{Step3}) = K^2N^2 + (2L_s+3)KN \tag{21}$$

- *Step 4:* As a worst-case scenario, we consider that step 4-2 is always processed. Thus, a total of $C_{max}(3KNQ+1)$ FLOPs are needed.
- *Step 5:* For I_{max} iterations, 5 FLOPs (2 comparison and 1 real addition) are computed per iteration. For one iteration, 2 FLOPs (1 comparison) are required.
- *Diversification:* As a worst case, we consider a situation where the algorithm always satisfies (14). The threshold test requires 2 FLOPs. Furthermore, the ZF solution should be calculated to perturb the algorithm. This perturbation requires $4K^3+8(2K^2+K)M+6K^2-2K(M+1)$ FLOPs.

For an LMMSE-SIC receiver, there are K streams to decode K users' symbols. Let us denote the k-th stream as being when there are k remaining users at the receiver. In this k-th stream, the following four operations are performed: LMMSE equalization, SINR calculation, finding a user with max SINR and interference cancellation. The LMMSE equalization involves $4k^3+8(2k^2+k)M+8k^2-2k$ FLOPs. Calculating k users' SINRs consumes $8k^2M+2kM+4k^3+8k^2+2k$ FLOPs [51]. Finding a max SINR user requires $2(k-1)$ FLOPs and interference cancellation uses $8M$ FLOPs. The total FLOPs for an LMMSE-SIC receiver can be derived as:

$$\underbrace{24M+10}_{1^{st}\ stream} + \sum_{i=2}^{K}\underbrace{\left(8i^3+24i^2M+16i^2+10iM+2i+8M\right)}_{2^{nd}-K^{th}\ stream}. \tag{22}$$

Note that only LMMSE equalization is processed for the 1-st stream of the LMMSE-SIC receiver.

Let us now address the complexity of the ML receiver. Notice that there are Q^K symbol vector candidates.

TABLE 1. Complexity comparison of different decoding algorithms in terms of FLOPs.

Receiver Type	Total Number of Floating Point Operations (FLOPs)
e-TS(Lower Bound)	$C_{\max}\left(8K^2MN+3KNQ+8MKN+2KN+2L_s+6\right)+4K^3+16K^2M+8K^2+8KM-2K-1$
e-TS(Upper Bound)	$I_{\max}\left(8K^2MN+K^2N^2+K(2L_s+3)N+8KMN+3KNQ+6\right)+8K^3+32K^2M+14K^2+14KM-4K-3$
LMMSE-SIC	$8K^3M+17K^2M+17KM-18M+2K^4+\frac{28K^3}{3}+11K^2+\frac{5K}{3}-14$
ML	$Q^K(8KM+8M+2)-2$

Calculating the likelihood cost of the Q^K symbol candidates involves $Q^K(8KM+8M)$ FLOPs. In addition, finding the minimum element from the likelihood costs of the Q^K symbol candidates involves $2(Q^K-1)$ FLOPs.

We summarize the total decoding complexity of the above receivers in Table 1. Unlike the ML receiver, the complexity of the proposed e-TS receiver does not exponentially increase as the number of users (K) or the modulation alphabet size (Q) increases.

IV. SIMULATION RESULTS

For the FS-NOMA parameter settings, we considered the system evaluation methodology in the 3GPP new radio (NR) standard and several related works. The proposed TS-based FS-NOMA receiver was evaluated at both the link and system levels under the simulation environment described below.

For link-level evaluations, we employed a cyclic prefix-based orthogonal frequency division multiplexing (CP-OFDM) waveform. As used previously in Yuan's work [17], the spreading length is set to four. The elements of the sequences are randomly selected from the set $\mathcal{C} = \{1+1i, 1-1i, -1+1i, -1-1i, 1, -1, 1i, -1i, 0\}$, which is one of the common complex-type sequence types of FS-NOMA systems [16], [17], [23]. In order to reflect the user overloading feature, we consider totals of 6, 8, 10, 12 and 14 users, which correspond to 150, 200, 250, 300% overloaded systems, respectively. The wireless channel model is the Rayleigh frequency selective channel, and we assume that the base station knows the perfect channel state information. Also, QPSK is used for symbol modulation and 1/2 rate Turbo coding is employed as a forward-error-correcting code [16], [17]. Regarding the transmitted block model, we drew from the recent documents from 3GPP meetings which discuss the evaluating environment of uplink NOMA systems for mMTC scenarios [41], [42], [52]. Specifically, we consider 20 bytes of information bits as a single transmitted block for every user, including 16 bits for the CRC bits.

For system-level evaluations, the following factors are considered over and above the aforementioned link-level parameters. We considered single-cell network with uplink devices whose positions follow a Poisson distribution. We assumed a carrier frequency of 700MHz, a transmit power for users of 23dBm, and a total system bandwidth of 10MHz. The pathloss model follows a general system level assumption of 3GPP TR 38.802 [53]. Also, we reflect a traffic model with fixed packet size with a Poisson arrival with fixed average arrival rate [17], [53]. Once a UE generates a packet to transmit, it transmits its packet to the closest subframe. As previously assumed in Yuan's work, the 4RB option is adopted for data transmission [17]. The detailed simulation parameters are listed in Table 2.

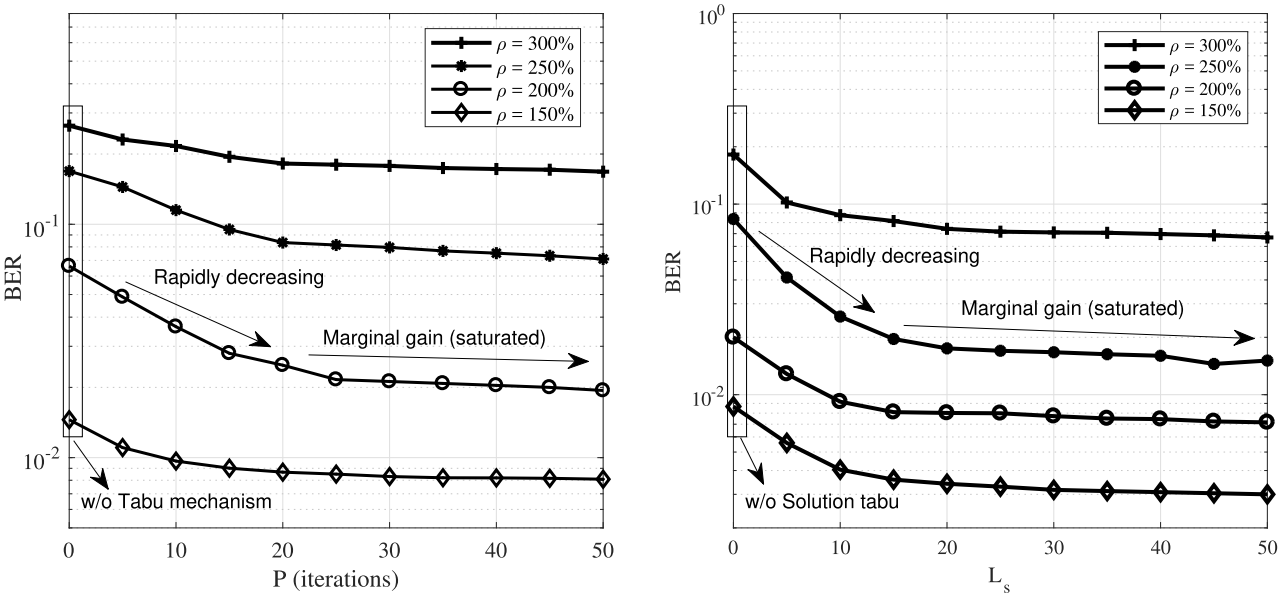
TABLE 2. FS-NOMA system parameters.

FS-NOMA System Parameters			
Link-level Parameters		**System-level Parameters**	
Parameters	**Values**	**Parameters**	**Values**
Waveform	OFDM	Center frequency (f_c)	700MHz
Spreading factor	4	Cell coverage	500m
# of TX antenna	1	Noise spectral density	$-174dBm/Hz$
#of RX antenna	1	Traffic model	Poisson with average arrival rate
Number of users	6, 8, 10, 12 (ρ= 150, 200, 250, 300)	Packet size	20 bytes
Modulation	QPSK	# of resource blocks (RB)	4RB
Channel coding	Turbo 1/2	Small scale fading	Rayleigh
# of CRC bits	16 bits	Pathloss model	$52.44+20\log_{10}(d)+20\log_{10}(f_c)$
Channel model	Rayleigh	UE transmit power	23dBm
Sequence elements	Random selection from the set below: {1 + 1i, 1 - 1i, - 1 + 1i, - 1 - 1i, 1, - 1, 1i, - 1i, 0}	UE distribution	Poisson distribution

Tabu-search related parameters are configured as following:

- For $\rho = 150\%$, $I_{max} = 100$, $C_{max} = 40$.
- For $\rho = 200\%$, $I_{max} = 500$, $C_{max} = 200$.
- For $\rho = 250\%$, $I_{max} = 2000$, $C_{max} = 800$.
- For $\rho >= 300\%$, $I_{max} = 10\rho$, $C_{max} = 4\rho$.

Note that the initial solution is set as $\mathbf{x}^{(0)} = \hat{\mathbf{x}}_{LMMSE}$ and the solution for diversification is set as $\mathbf{x}_{div} = \hat{\mathbf{x}}_{ZF}$. Also, $\hat{\mathbf{x}}_{LMMSE}$ and $\hat{\mathbf{x}}_{ZF}$ denote the LMMSE solution and ZF solution,



(a) Effect of P, SNR = 10dB

(b) Effect of L_s, SNR = 10dB

FIGURE 7. **Effect of the dual-control parameters (P and L_S) on the BER performance of the proposed e-TS algorithm. The overloading amounts considered in this figure are 150, 200, 250, and 300 %.**

respectively. The diversification threshold is set according to (15).

In the sections below, we will evaluate the performance of the proposed FS-NOMA receiver based on the TS algorithm. As a first step, link level-simulation is performed. We first observe the effect of the proposed dual-control parameters on the BER performance. We then compare the different decoding algorithms, including the proposed e-TS based receiver in terms of BER performance and computational complexity. Finally, we perform a system-level evaluation to observe whether the proposed receiver can increase the possible number of connected users in a cell coverage.

To compare the performances of the proposed receiver with the conventional receiver for FS-NOMA, we used LMMSE-SIC [15]–[17], [23]. In addition, a comparison of the performance of the FS-NOMA receiver with a direct application of TS algorithm is also compared although it has not been investigated in the literature. As a link-level BER lower bound, we exploited the ML-based receiver.

A. EFFECT OF DUAL-CONTROL PARAMETERS ON UNCODED BER PERFORMANCES

In this subsection, we evaluate the uncoded performance of the proposed e-TS algorithm according to the dual-control parameters. Fig. 7 shows the BER as the number of prohibited MOVEs (P) and solutions (L_s) increases. To observe the effect of the parameters, the SNR is fixed at 10dB. We also show the results with various user overloading factors, i.e., 150, 200, 250 and 300. These simulation results will be used to attempt to determine the appropriate value of the dual control parameters.

Fig. 7(a) clearly shows that the BER performance improves as the number of prohibited MOVEs P increases. This is because as P becomes larger, the TS algorithm can avoid repetitive MOVEs for a greater number of iterations. In particular, we can see that the BER performances increase as the value of P increases so long as P is smaller than 20. However, when $P > 20$, the gain is marginal. This is because of the TABU criterion when the algorithm determines next MOVEs. Setting a very large P value means that as the algorithm progresses, most of the elements of the TABU MOVE matrix $\mathbf{T}_m$ are non-zero so that the MOVEs are forbidden. Meanwhile, we should remind the reader that the proposed e-TS algorithm 'cancels' the prohibition if the selected MOVE leads the algorithm to the best solution vector up to the current iteration. Thus, if P exceeds some threshold value, only MOVEs which have renewed a best solution vector within recent iterations are not prohibited regardless of how the precise amount of P is. Thus, we can judge that the value of the parameter P should be greater than or equal to twenty.

Fig. 7(b) shows the effect of solution prohibition when the prohibiting amount of MOVEs P is fixed at 20. As the number of solution prohibitions L_s increases, we can observe that the BER performance improves. This is because the e-TS algorithm can prevent the repetitive solutions for a longer period. Regardless of the overloading amount, the BER rapidly decreases until $L_s = 15$, after which saturation is reached. Unlike the MOVE prohibition, the solution prohibition requires more computational complexity to increase the number of prohibitions. Specifically, when the algorithm performs the step 3-2, the number of previous solution vectors is included in the solution prohibition list $\mathcal{T}_s$ as the number

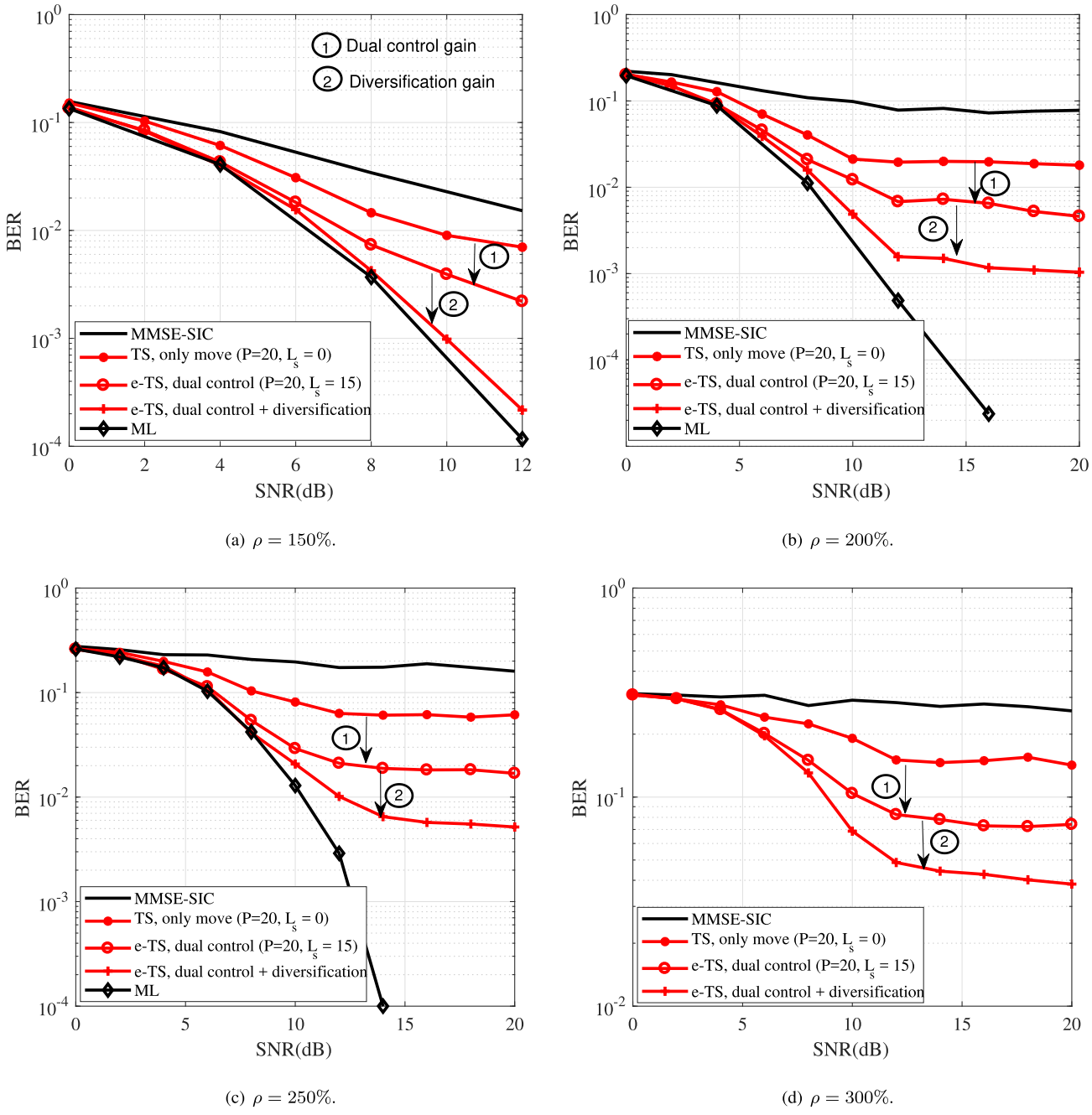


FIGURE 8. Effect of the proposed e-TS algorithm to the BER performance of the symbol-level receiver.

of solution prohibitions increases. From this result, we can conclude that setting $L_s = 15$ strikes a good balance for the algorithm between decoding accuracy and complexity.

B. UNCODED BER PERFORMANCE OF THE PROPOSED E-TS ALGORITHM

In this subsection, we evaluate and compare the BER performance of the proposed e-TS based FS-NOMA receiver with conventional LMMSE-SIC receivers. In addition, the performance of the ML-based receiver is provided as a lower bound. We provide three TS-based FS-NOMA performances as follows: 1) TS with move control only, 2) TS with dual control, 3) TS with dual control and diversification. Using these three cases, we can observe the individual gain of the proposed dual-control scheme and the diversification scheme, respectively. For all cases, we set the number of prohibited MOVEs to $P = 20$. For cases 2) and 3), we set the number of prohibited solutions $L_s = 15$. Note that the ML performance of the 300% overloaded system is not given because of the prohibitively high complexity required in simulating it.

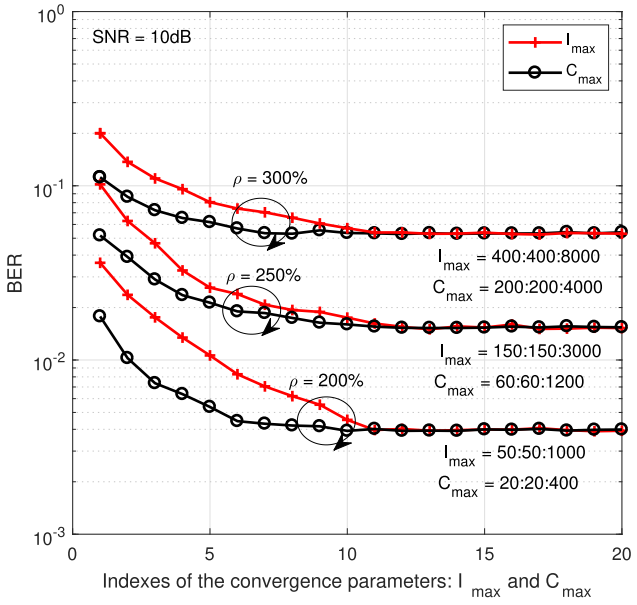


FIGURE 9. Convergence result of the proposed e-TS algorithm with respect to the parameters I_{max} and C_{max}.

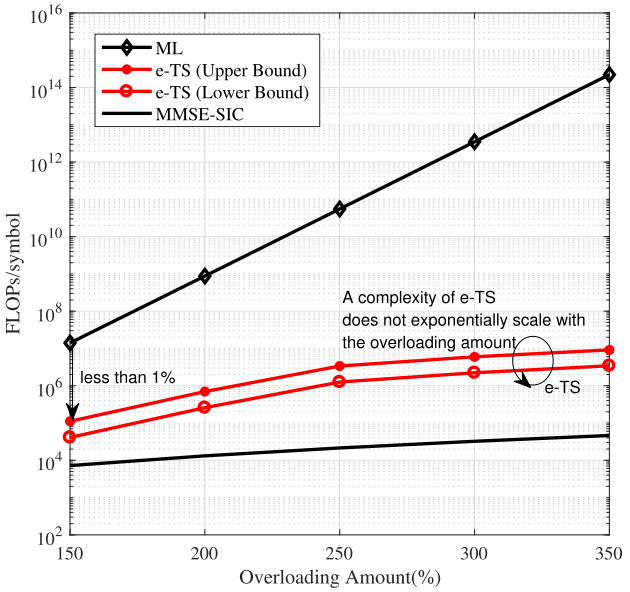


FIGURE 10. Per-symbol computational complexity of the proposed algorithm in terms of the number of FLOPs as a function of the amount of overloading.

Fig. 8 shows the BER results for the FS-NOMA receivers. The sub-figures (a), (b), (c) and (d) denote the results for 150%, 200%, 250% and 300% overloaded systems, respectively. The BER performance of the tabu search-based receivers outperforms the conventional LMMSE-SIC-based receiver even with the TS algorithm that only prohibits MOVEs. In particular, the LMMSE-SIC receivers are unable to reach a BER of 10^{-2} until an SNR of 20dB in Fig. 8, because of the severe error propagation problem. Conversely, the proposed e-TS algorithm-based receiver has near-ML performance in the low SNR regime of the 150, 200, 250% overloaded systems. Notice that the BER performance of the TS with MOVE only is very far from the ML performance. In this context, we can conclude that the proposed dual-control mechanism and the diversification scheme have a great effect on the BER performance. Meanwhile, the BER performances of the e-TS based receivers are saturated in the high SNR regime. This is because the quality of the initial solution vector does not improve as the SNR increases due to the inherent multi-user-interferences that FS-NOMA systems have.

C. CONVERGENCE PROPERTY OF THE PROPOSED E-TS RECEIVER

In this subsection, we observe the convergence property of the proposed e-TS algorithm. In particular, we present simulation results on the BER performance of the e-TS receiver with regard to the parameters I_{max} and C_{max}. In particular, the convergence tendency of the e-TS receiver is provided for 200%, 250% and 300% overloaded FS-NOMA systems, respectively.

Fig. 9 depicts the convergence result of the proposed e-TS algorithm with respect to the control parameters I_{max} and C_{max}. Throughout Fig. 9 we show the algorithm convergence for those two control parameters, individually. Note that the range of each set of parameter values is set to be sufficiently wide so that we can sufficiently observe the BER convergence tendency. Also, the simulations were conducted under a fixed SNR, 10dB.

For the attempt to observe the I_{max} convergence, the C_{max} value is fixed to a sufficiently large value so as to simulate the individual convergence characteristic of the I_{max} regardless of C_{max}. Since C_{max} stands for the number of iterations where the best solution vector remains unchanged, the convergence result of I_{max} is not affected by the value of C_{max} if the condition $C_{max} = I_{max}$ is satisfied. Meanwhile, when observing the convergence property of C_{max}, we exploited the convergence result of I_{max}. Specifically, we first fix the I_{max} value which is already converged, and then observe the convergence of C_{max}. As shown in Fig. 9, I_{max} values for C_{max} convergence observation are set to 800, 2400 and 6000 for 200%, 250% and 300% overloaded system, respectively.

Fig. 9 shows the BER performance with respect to the parameters I_{max} and C_{max}. We can see that the BER converges as the value of the parameters increases. Specifically, the proposed e-TS receiver converges with regard to the I_{max} after it is set to be higher than 800, 2400 and 6000 for the 200%, 250% and 300% overloaded systems, respectively. For C_{max}, the e-TS algorithm converges after the value is higher than 140, 600 and 1600 for the 200%, 250% and 300% overloaded FS-NOMA systems, respectively.

D. COMPUTATIONAL COMPLEXITY

Fig. 10 shows the numerical complexity results for the FS-NOMA receivers derived in section 3-D. Specifically, we present the minimum and maximum amounts of

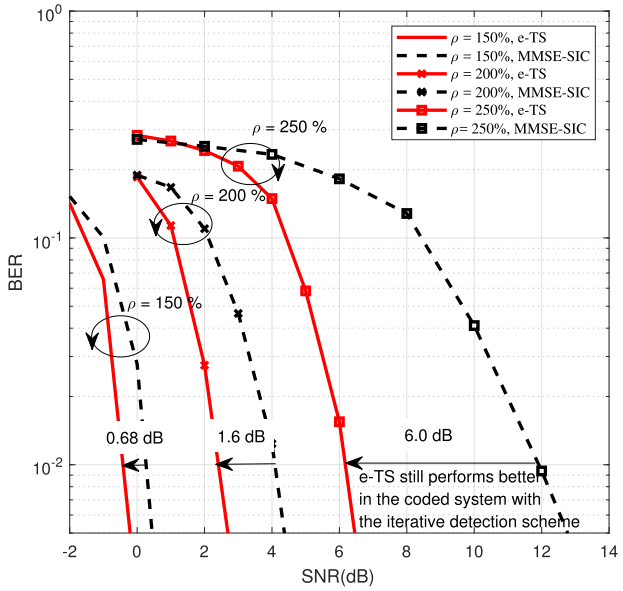


FIGURE 11. Turbo-coded BER of the iterative CRC-IC FS-NOMA receivers.

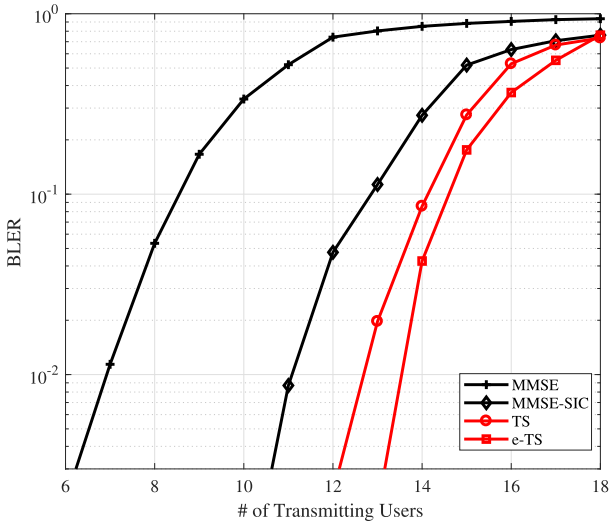


FIGURE 12. BLER curve as the number of active users increases.

per-symbol FLOPs that the proposed receiver requires and compare the results with those for the other receivers (LMMSE-SIC and ML). The e-TS parameters employed in this simulation are the same as the values that we used to evaluate the symbol-level performance in this section. We can observe that the proposed e-TS receiver requires less than 1% of the FLOPs the ML receiver requires, even in the worst case. At the same time, the computational complexity of the proposed method is similar to that of the LMMSE-SIC when user overloading is 150%. This is remarkable since the BER performance of the proposed e-TS is far better than that for LMMSE-SIC.

E. PERFORMANCE OF THE ITERATIVE CRC-IC RECEIVER WITH THE PROPOSED E-TS-BASED FS-NOMA RECEIVER

In this subsection, we present the performance of the proposed e-TS based FS-NOMA receiver with the iterative detection scheme presented in section 2. The e-TS algorithm in this subsection includes both the dual-control and the diversification schemes.

Fig. 11 shows the BER performance of the e-TS and the conventional LMMSE-SIC with diverse overloading factors. Similar to the uncoded BER cases, we can observe that the proposed e-TS receivers are superior to the LMMSE-SIC receivers in terms of the coded BER performance. For example, in the 150% overloaded system, LMMSE-SIC requires an SNR of 0.45 dB to achieve a BER of 10^{-2}, whereas the proposed e-TS achieves the same BER at -0.23 dB, which amounts to a 0.68dB gain in terms of SNR. Note that this performance gap increases as the overloading of the system becomes harsher. The explanation for this tendency lies in the error propagation problem associated with LMMSE-SIC. In highly overloaded systems, the initial LMMSE solutions suffer from unaffordable inter-user-interference which results in erroneous detection. Thus, the frequency of the error propagation dramatically increases. In contrast, the proposed e-TS algorithm is free of any error propagation problem. Although e-TS also degrades as the overloading ratio increases because of the inaccurate initial solution and the increased search dimension, it is less sensitive than LMMSE-SIC.

F. PACKET DROP RATE VERSUS CONNECTED USER DENSITY

In this subsection, the performance of the FS-NOMA receivers is evaluated through system level simulations. The target is to compare the proposed e-TS-based FS-NOMA receiver with other receivers. We first observe a BLER curve for the FS-NOMA receivers (both proposed and conventional schemes) as the number of the superposed users increases. We then evaluate a packet drop rate for the FS-NOMA receivers as the number of connected users in a cell coverage area increases. The detailed parameters are presented in Table 2.

Fig. 12 depicts the average block error rate (BLER) results when the number of transmitting users is 8 to 16. The transmitted block is composed of a 20-byte information block (N_{I}) and the 16 CRC bits (N_{CRC}) that are one of the possible transmitting setups discussed in mMTC for evaluating system performance [52]. We can observe that the proposed e-TS based receiver provides a better average BLER for a given number of transmitting users when compared to the LMMSE, LMMSE-SIC and TS receivers. For example, when the number of transmitting users is 13, the BLER of the proposed e-TS based receiver is only 0.22%, 1.6% and 9.14 % of the conventional LMMSE, LMMSE-SIC and TS receivers, respectively.

The next step is to observe the packet drop rate of the FS-NOMA receivers in a grant-free transmission scenario [17], [23], [24], [54]. We assume that every connected user generates a 20 bytes of transmitting packet through a

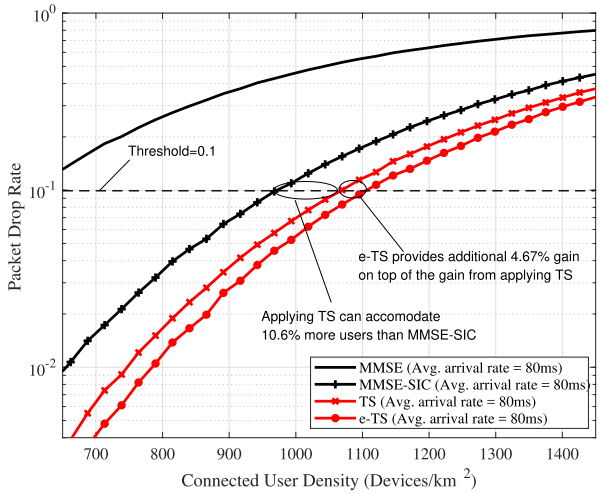


FIGURE 13. Packet drop rate of the FS-NOMA receivers according to various connected user densities when a packet arrival rate is 80ms.

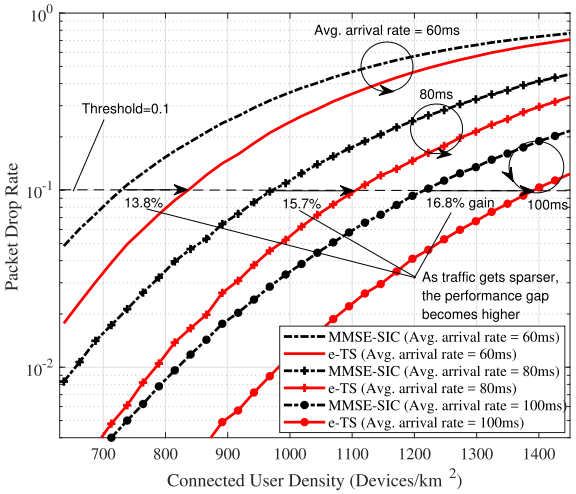


FIGURE 14. Effect of traffic arrival rate on the packet drop rate performances of FS-NOMA receivers.

Poisson arrival process with a certain average period. Also, the 4RB option is given for grant-free access per subframe with a 1ms duration [17]. Once a packet is generated, the user transmits the created packet to the nearest subframe without performing the scheduling grant procedure. We simulated 100000 subframes to observe the packet drop rates of the FS-NOMA receivers.

Fig. 13 shows the average packet drop rate as the connected user density increases. A traffic arrival rate is given as 80ms. When the packet drop rate is 0.1, the connected user densities are 602.3, 967.7, 1070 and 1120 $(devices/km^2)$ for the LMMSE, LMMSE-SIC, TS and e-TS, respectively. In other words, the proposed e-TS-based receiver can extend the possible number of connected users in a cell coverage compared to the conventional receivers. Notice that the proposed e-TS-based receiver can increase 15.74% of the number of connected users than the conventional representative receiver, LMMSE-SIC. We can also observe that the connected user density of the e-TS-based receiver is 4.67% higher than the TS-based receiver, which demonstrates that the effect of the proposed dual control and the diversification strategy are also valid in the system-level evaluations.

The effect of traffic arrival rate on the packet drop rate performance is shown in the Fig. 14. We consider three types of arrival rates;60ms, 80ms and 100ms, to examine how the rate of packet arrival affects the packet drop rates. As the arrival rate increases, we can observe that the packet drop rates of the both proposed and conventional receivers degrade. This is because as the arrival becomes denser, the number of transmitting users is likely to be higher as well.

From the Fig. 14, we also can observe how the performance gap between the two receivers changes catching up with the variation of the arrival rate. For a given packet drop rate, the gap between LMMSE-SIC and e-TS gets farther as the traffic gets sparser, in terms of connected user density. For example, when the target packet drop rate is 0.1, the performance gaps are 13.8%, 15.74% and 16.75% when the packet arrival rates are 60, 80 and 100ms, respectively. We can interpret this tendency by observing the BLER performances when there are too many transmitting users in the cell. As shown in the Fig. 12, BLER gap of the LMMSE-SIC and the e-TS becomes narrower as the number of transmitting users increases. Thus, if the packet arrival rate increases, the packet drop rate gap become narrower since the number of transmitting user statistically increases.

Throughout this section, the various performances of the proposed e-TS-based FS-NOMA receiver is evaluated in both link-level and system-level. In a link level, we first found a proper value of P and L_s to operate the e-TS algorithm. From the simulation results, we found that the BER performance is saturated when the values of P and L_s are higher than a certain value. Subsequently, we observed that the BER of the proposed e-TS-based receiver outperforms the conventional receivers in both uncoded and coded systems. Furthermore, we demonstrated that the complexity of the proposed receiver is moderate compared to the ML receiver. In a system level, we presented the packet drop rate of the receivers in a uplink grant-free FS-NOMA system with a Poisson traffic model. We observed that the proposed e-TS receiver can accommodate 15.74% more devices than the conventional LMMSE-SIC receiver when a packet drop rate threshold is 0.1.

V. CONCLUSION

In this paper, we investigated a novel multi-user receiver based on the tabu-search algorithm for FS-NOMA systems. Considering the broad search space that FS-NOMA systems have, we improved the TS algorithm from the following two perspectives: 1) a dual-controlled TABU method to quickly escape from the local optimal solutions 2) a diversification method to explore unvisited region of the solution space. As a

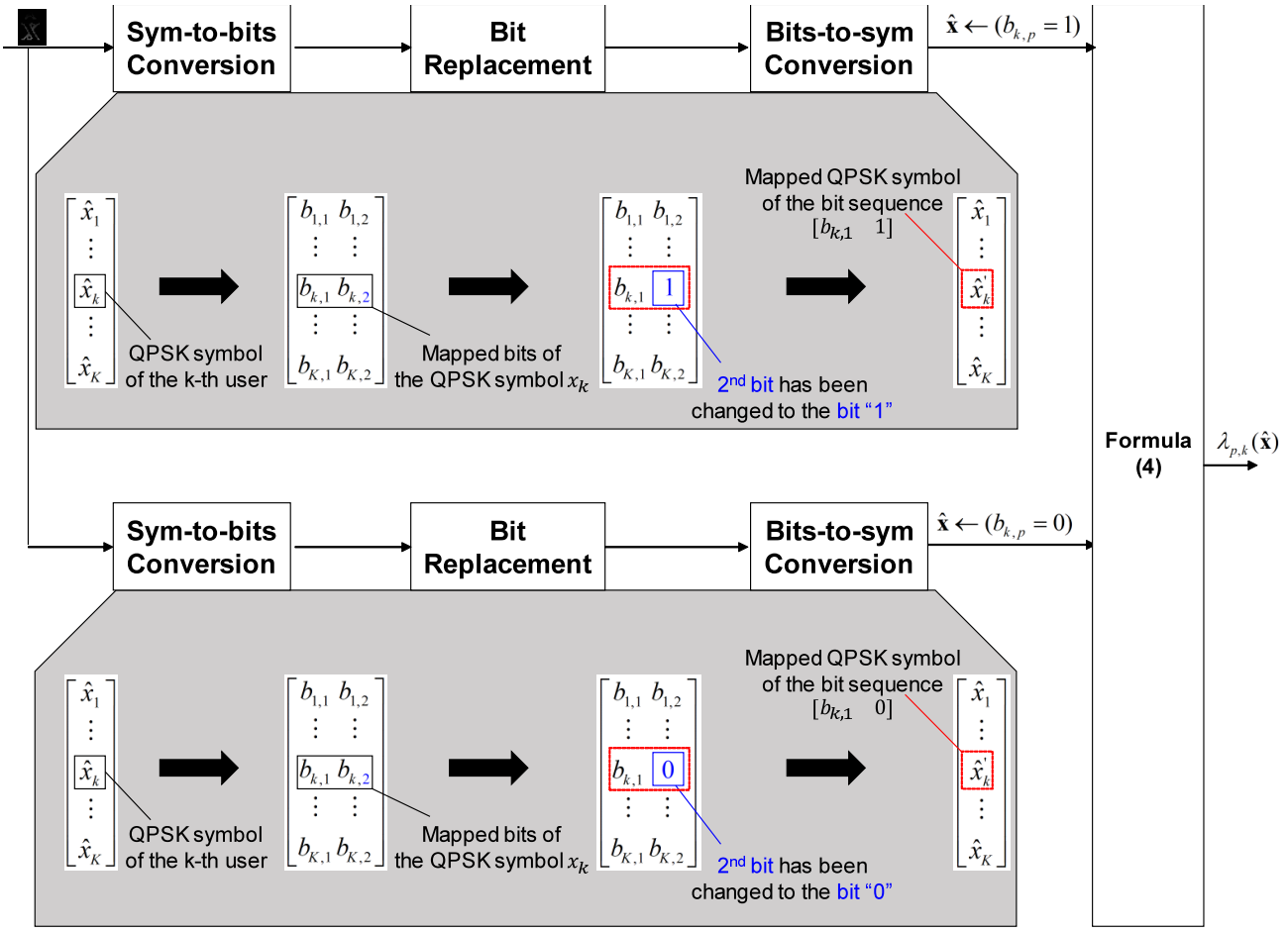


FIGURE 15. An example of the LLR-converting operation based on the derived formula (4).

result, our proposed receiver showed far better decoding performance than the conventional LMMSE-SIC receiver. Based on the superior decoding performance, an FS-NOMA system with the proposed receiver can serve a greater number of users, which in turn will be of value in supporting massive connectivity. In addition, we demonstrated that feasibility of the proposed e-TS-based receiver by analyzing the computational complexity. Possible topics for the future include an advanced receiver design to improve the BER in the high-SNR regime and extending our work to asynchronous environments.

APPENDIX
LLR CALCULATION METHOD

In this appendix, we first provide the method for approximating the LLR converting operation (4). We then describe an example of the actual operation of (4) when QPSK modulation is considered.

The goal of the LLR conversion is to maximize a posteriori probability (MAP) when a hard-decision-output is given. Let us define the log-likelihood ratio of the transmitted bit as shown [55]:

$$\lambda_{p,k} = \ln\left(\frac{P(b_{k,p}=1|\mathbf{y})}{P(b_{k,p}=0|\mathbf{y})}\right), \tag{23}$$

where $b_{k,p}$ denotes p-th bit of the k-th user's transmitted symbol. We assume that each user exploits an interleaver so that all the bits in $\mathbf{x}$ can be considered as statistically independent. By Bayes' theorem, (23) can be represented as

$$\lambda_{p,k} = \ln\left(\frac{P(b_{k,p}=1)}{P(b_{k,p}=0)}\right) + \ln\left(\frac{P(\mathbf{y}|b_{k,p}=1)}{P(\mathbf{y}|b_{k,p}=0)}\right). \tag{24}$$

Note that the conditional probability $P(\mathbf{y}|b_{k,p}=1)$ can be expressed as the summation of conditional probabilities of candidate solutions whose $b_{k,p}$ is 1. Then, (24) can be rewritten as

$$\lambda_{p,k} = \ln\left(\frac{P(b_{k,p}=1)}{P(b_{k,p}=0)}\right) + \ln\left(\frac{\sum_{\mathbf{x}\in\mathcal{X}_{b_{k,p}=1}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{X}_{b_{k,p}=0}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}\right), \tag{25}$$

where $\mathcal{X}_{b_{k,p}=1}$ is a set of the all of K-tuple vectors whose $b_{k,p}$ is 1.We then separate the denominator and numerator of the (25) into two parts each, as presented in (26) , as shown at the top of the next page, where $\mathbf{x}_{-k}$, $(k > 0)$ denotes a $(K-1)$-tuple solution vector whose k-th element is excluded from the original vector $\mathbf{x}$. Let us assume that the hard-decision result $\hat{\mathbf{x}}_{-k}$ is error-free. Also, we consider the *a priori* probability of

$$\lambda_{p,k} = \ln(\frac{(b_{k,p}=1)}{(b_{k,p}=0)}) + \ln(\frac{\sum\limits_{\mathbf{x}_{-k}\neq\hat{\mathbf{x}}_{-k},\mathbf{x}\in\mathcal{X}_{b_{k,p}=1}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) + \sum\limits_{\mathbf{x}_{-k}=\hat{\mathbf{x}}_{-k},\mathbf{x}\in\mathcal{X}_{b_{k,p}=1}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{\sum\limits_{\mathbf{x}_{-k}\neq\hat{\mathbf{x}}_{-k},\mathbf{x}\in\mathcal{X}_{b_{k,p}=0}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) + \sum\limits_{\mathbf{x}_{-k}=\hat{\mathbf{x}}_{-k},\mathbf{x}\in\mathcal{X}_{b_{k,p}=0}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}), \tag{26}$$

$$\lambda_{p,k}(\hat{\mathbf{x}}) = -\frac{1}{2\sigma^2}\cdot(||\mathbf{y}-\mathbf{G}(\hat{\mathbf{x}}\leftarrow(b_{k,p}=1))||^2 - ||\mathbf{y}-\mathbf{G}(\hat{\mathbf{x}}\leftarrow(b_{k,p}=0))||^2). \tag{29}$$

the bits "0" and "1" being equally distributed. Then, we can easily notice that the symbols are equally distributed in quadrature-amplitude modulation (QAM) scheme, too [56]. Considering that values of $P(\mathbf{x})$ are all the same, the LLR value becomes

$$\lambda_{p,k}(\hat{\mathbf{x}}) = \ln\left(\frac{P(\mathbf{y}|\hat{\mathbf{x}}\leftarrow(b_{k,p}=1))P(\hat{\mathbf{x}}\leftarrow(b_{k,p}=1))}{P(\mathbf{y}|\hat{\mathbf{x}}\leftarrow(b_{k,p}=0))P(\hat{\mathbf{x}}\leftarrow(b_{k,p}=0))}\right). \tag{27}$$

Note that the likelihood function $P(\mathbf{y}|\mathbf{x})$ can be obtained as

$$P(\mathbf{y}|\mathbf{x}) = \frac{\exp[-\frac{1}{2\sigma^2}\cdot||\mathbf{y}-\mathbf{G}\mathbf{x}||^2]}{(2\pi\sigma^2)^M}. \tag{28}$$

After substituting (28) into (27), the LLR value becomes (29), as shown at the top of this page.

The LLR calculation operation performed in way described in the (4) is as follows. To be specific, let us calculate the LLR of the p-th bit of the user k from $\hat{\mathbf{x}}$, which is the output of the e-TS receiver. The cost functions of the following two vectors should be obtained first: $\hat{\mathbf{x}}\leftarrow(b_{k,p}=0)$ and $\hat{\mathbf{x}}\leftarrow(b_{k,p}=1)$. The operation $\mathbf{x}\leftarrow(b_{k,p}=1)$ is referred to as a 'bit substitution' for the sake of clarity. The detailed explanation of the bit substitution operation is as follows: As an example, QPSK modulation is assumed, which denotes that a symbol is mapped to 2 bits. Without loss of generality, we present the case where the 2-nd bit of the k-th user is substituted into bit '1'. The bit substitution goes through three steps as shown in Fig. 15. As a first step, each element (QPSK symbol) of given input vector $\mathbf{x}$ is converted to the corresponding bits. We then substitute bit '1' into the 2-nd bit of the user k. Finally, the k-th user's symbol is substituted into the QPSK symbol x'_k which is mapped to the bits $[b_{k,1}\ \ 1]$. Likewise, the operation $\mathbf{x}\leftarrow(b_{k,p}=0)$ can be performed in a similar manner. Once the result of the bit substitutions is reflected in (4), bit LLR calculation is finished.

REFERENCES

[1] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The future of industrial communication: Automation networks in the era of the Internet of Things and industry 4.0," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 17–27, Mar. 2017.

[2] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-based smart cities: Recent advances and challenges," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 16–24, Sep. 2017.

[3] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[4] S. Mumtaz, A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, and J. Rodriguez, "Massive Internet of Things for industrial applications: Addressing wireless IIoT connectivity challenges and ecosystem fragmentation," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 28–33, Mar. 2017.

[5] H. Kodesh, "Dzone guide to the Internet of Things," General Electr. (GE) Softw., Boston, MA, USA, Tech. Rep. 09, 2015.

[6] H. Tullberg, P. Popovski, Z. Li, M. A. Uusitalo, A. Hoglund, O. Bulakci, M. Fallgren, and J. F. Monserrat, "The METIS 5G system concept: Meeting the 5G requirements," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 132–139, Dec. 2016.

[7] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive Machine-type Communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[8] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.

[9] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[10] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 611–615.

[11] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[12] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[13] L. Liu, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving MIMO-NOMA: Iterative LMMSE detection," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1758–1773, Apr. 2019.

[14] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Veh. Technol. Mag.*, vol. 13, no. 4, pp. 40–50, Dec. 2018.

[15] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 20, pp. 2294–2323, 3rd Quart., 2018.

[16] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive study and comparison on 5G NOMA schemes," *IEEE Access*, vol. 6, pp. 18511–18519, 2018.

[17] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for Internet of Things," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

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

[19] Y. Du, B. H. Dong, Z. Chen, J. Fang, and L. Yang, "Shuffled multiuser detection schemes for uplink sparse code multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1231–1234, Jun. 2016.

[20] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—A novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.

[21] H. Kim, Y.-G. Lim, C.-B. Chae, and D. Hong, "Multiple access for 5G new radio: Categorization, evaluation, and challenges," 2017, *arXiv:1703.09042*. [Online]. Available: https://arxiv.org/abs/1703.09042

[22] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2918–2923.

[23] Z. Yuan, C. Yan, Y. Yuan, and W. Li, "Blind multiple user detection for grant-free MUSA without reference signal," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.

[24] Q. Wang, Z. Zhao, D. Miao, Y. Zhang, J. Sun, M. Liu, and Z. Zhong, "Non-orthogonal coded access for contention-based transmission in 5G," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–6.

[25] Z. Zhao, D. Miao, Y. Zhang, J. Sun, H. Li, and K. Pedersen, "Uplink contention based transmission with non-orthogonal spreading," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–6.

[26] E. M. Eid, M. M. Fouda, A. S. T. Eldien, and M. M. Tantawy, "Performance analysis of MUSA with different spreading codes using ordered SIC," in *Proc. 12th Int. Conf. Comput. Eng. Syst. (ICCES)*, 2017, pp. 101–106.

[27] C. Sankaran and A. Ephremides, "Solving a class of optimum multiuser detection problems with polynomial complexity," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1958–1961, Sep. 1998.

[28] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212–1215, Mar. 2001.

[29] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems," 2019, *arXiv:1901.09559*, [Online]. Available: https://arxiv.org/abs/1901.09559

[30] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, "Convergence analysis and assurance for Gaussian message passing iterative detector in massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6487–6501, Sep. 2016.

[31] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and C. Huang, "Gaussian message passing for overloaded massive MIMO-NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 210–226, Jan. 2019.

[32] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and C. Huang, "Gaussian message passing iterative detection for MIMO-NOMA systems with massive access," in *Proc. IEEE Global Commun. Conf. GLOBECOM)*, Dec. 2016, pp. 1–6.

[33] L. Liu, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving iterative LMMSE detection for MIMO-NOMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[34] Y. Chi, L. Liu, G. Song, C. Yuen, Y. L. Guan, and Y. Li, "Practical MIMO-NOMA: Low complexity and capacity-approaching solution," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6251–6264, Sep. 2018.

[35] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[36] L. Liu, Y. Li, C. Huang, C. Yuen, and Y. L. Guan, "A new insight into GAMP and AMP," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8264–8269, Aug. 2019.

[37] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[38] X. Meng and J. Zhu, "Bilinear adaptive generalized vector approximate message passing," *IEEE Access*, vol. 7, pp. 4807–4815, 2019.

[39] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[40] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[41] *Evaluation Methodology Metrics for NoMA*, document R1-1805006, Ericksson, 3GPP, 2018.

[42] *Preliminary LLS Evaluation Results for mMTC Scenario*, document R1-1803665, Huawei, 3GPP, 2018.

[43] R. Syahputra, "Distribution network optimization based on genetic algorithm," *J. Elect. Technol.*, vol. 1, no. 1, pp. 1–9, 2017.

[44] M. Dorigo and T. Stützle, "Ant colony optimization: Overview and recent advances," in *Handbook of Metaheuristics*. Berlin, Germany: Springer, 2019, pp. 311–351.

[45] P. J. Van Laarhoven and E. H. Aarts, *Simulated Annealing: Theory and Applications*. Springer, 1987, pp. 7–15.

[46] S. Choi and T.-W. Lee, "A negentropy minimization approach to adaptive equalization for digital communication systems," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 928–936, Jul. 2004.

[47] F. Glover and M. Laguna, "Tabu search," in *Handbook of Combinatorial Optimization*. Springer, 1998, pp. 2093–2229.

[48] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[49] R. P. Brent and P. Zimmermann, *Modern Computer Arithmetic*, vol. 18. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[50] R. Hunger, "Floating point operations matrix-vector calculus," Inst. Circuit Theory Signal Process., Munich Univ. Technol., Munich, Germany, Tech. Rep., 2005.

[51] A. H. Mehana and A. Nosratinia, "Diversity of MMSE MIMO receivers," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6788–6805, Nov. 2012.

[52] *Summary of offline Discussion on LLS Clarifications*, document R1-1805682, ZTE, 3GPP, 2018.

[53] *Study on New Radio Access Technology, Physical Layer Aspects (Release 14)*, document TR 38.912, 2017.

[54] M. Cheng, Y. Wu, Y. Li, Y. Chen, and L. Zhang, "PHY abstraction and system evaluation for SCMA with UL grant-free transmission," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.

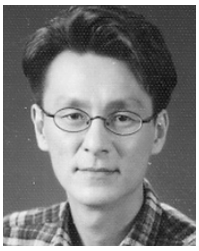
[55] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

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

INSIK JUNG (S'14) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2014, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. His research interests include B5G and 6G wireless communications, NOMA, meta-heuristics, and mobile edge computing.

HYUNSOO KIM (S'12) received the B.S. degree from the School of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea, in 2012, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include new waveform, non-orthogonal multiple access, licensed assisted access, and full-duplex. He also received the Scholarship of National Research Foundation of Korea during his B.S. degree.

JINKYO JEONG (S'15) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. His research interests include 5G and beyond 5G wireless communications, new waveform design, full-duplex systems, machine learning, and mobile edge computing.



SOOYONG CHOI (S'98–M'04) was born in Seoul, South Korea, in 1971. He received the B.S.E.E., M.S.E.E., and Ph.D. degrees from Yonsei University, Seoul, in 1995, 1997, and 2001, respectively. In 2001, he was a Postdoctoral Researcher with the IT Research Group, Yonsei University. His work focused on proposing and planning for the fourth-generation communication systems. From 2002 to 2004, he was a Postdoctoral Fellow with the University of California at San Diego, La Jolla, CA, USA. From 2004 to 2005, he was a Researcher with Oklahoma State University, Stillwater, OK, USA. From 2005 to 2016, he was an Assistant Professor and a Professor with the Department of Electrical and Electronic Engineering, Yonsei University. Since 2016, he has been a Professor with the School of Electrical and Electronic Engineering. His research interests include adaptive signal processing techniques (equalization and detection), massive MIMO systems for wireless communication, and new waveforms and multicarrier transmission techniques for future wireless communication systems.

DAESIK HONG (S'86–M'90–SM'05) received the B.S. and M.S. degrees in electronics engineering from Yonsei University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from the School of EE, Purdue University, West Lafayette, IN, USA, in 1990.

In 1991, he joined Yonsei University, where he is currently a Professor with the School of Electrical and Electronic Engineering and the Dean with the College of Engineering. He also served as the Vice President for Research Affairs and the President fro Industry-Academic Cooperation Foundation, Yonsei University, from 2010 to 2011. He also served as the Chief Executive Officer (CEO) for Yonsei Technology Holding Company, in 2011, and the President for the Institute of Electronics and Information Engineers (IEIE), in 2017. His current research interests include future wireless communication, including 5G and 6G systems, OFDM (A) and multicarrier communication, multihop and relay-based communication, in-band full-duplex, cognitive radio, and energy harvesting. He was appointed as the Underwood/Avison Distinguished Professor with Yonsei University, in 2009. He received the Best Teacher Award from Yonsei University, in 2005, 2011, and 2012. He was also a recipient of the HaeDong Outstanding Research Awards of the Korean Institute of Communications and Information Sciences (KICS), in 2006, and the IEIE, in 2009. He has been serving as the Chair for Samsung-Yonsei Research Center for Mobile Intelligent Terminals. He has served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2006 to 2011, and the IEEE WIRELESS COMMUNICATIONS LETTERS, from 2011 to 2016. More information about his research is available online: http://mirinae.yonsei.ac.kr.