Power-Level-Design-Aware Scalable Framework for Throughput Analysis of GF-NOMA in mMTC

Takeshi Hirai, *Member, IEEE*, Rei Oda, and Naoki Wakamiya, *Member, IEEE*

***Abstract*—This article proposes a scalable framework to analyse the throughput of the grant-free power-domain nonorthogonal multiple access (GF-NOMA) and presents the achievable performance in the optimized offered load at each power level (called per-level offered load) by using our framework. Our analytical model reflects packet errors caused by *power collisions*, characterized by GF-NOMA, based on the power level design guaranteeing the required signal-to-interference-and-noise ratio (SINR). This key idea enables analyzing the throughput of a large-scale GF-NOMA system more accurately than the existing analytical models. Also, this key idea enables optimizing the per-level offered load rather than an uniform one in typical optimization problems related to the throughput: the throughput maximization or energy minimization problem with a throughput condition. Our analytical results highlight some key insights into designing future access control methods in GF-NOMA. First, our analytical model achieves an approximation error of only 0.4% for the exact throughput obtained by the exhaustive search at the five power levels. The existing analytical model provides an approximation error of 25%. Next, our proposed framework highlights that the optimal per-level offered load restrictively improves the throughput above the optimally uniform per-level offered load. Finally, our proposed framework discovers a 27% more energy-efficient per-level offered load than the existing framework at the five power levels while providing higher throughput than the optimally uniform per-level offered load.**

***Index Terms*—Channel inversion, grant-free (GF), massive machine-type communications (mMTC), per-level offered load, power collision, power-domain nonorthogonal multiple access (NOMA), successive interference cancellation (SIC).**

I. Introduction

5G/6G uplink cellular networks require the massive machine-type communications (mMTC) [1], [2], [3]. A typical mMTC scenario has the following three characteristics.

1) In mMTC, a huge number of users connect to a base station (BS), although each user sporadically transmits small packets. This characteristic requires achieving enough high system throughput (simply called throughput) to accommodate such users.
2) The mMTC requires such users to access resources in a distributed manner to reduce signaling overheads for the connected BS. Such an access protocol is called a grant-free (GF) access protocol, like ALOHA.
3) Such users are typically the battery-powered sensors. Based on the above characteristics, mMTC needs enough high throughput while suppressing enough energy consumption in a GF manner.

To increase the throughput in mMTC, the GF access meets the power-domain nonorthogonal multiple access (NOMA) with the *channel inversion* (channel-inverted NOMA), called GF power-domain nonorthogonal multiple access (GF-NOMA) [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. In GF-NOMA, each user selects a predesigned power level and calculates its transmission power by the target-received power value associated with its selected level by *channel inversion*. The BS uses the successive interference cancelation (SIC) technique to decode a superposed signal made by arriving packets transmitted by the several users at each power level. Each power level is designed to guarantee a required signal-to-interference-and-noise ratio (SINR) at each SIC iteration at no *power collisions* where multiple packets arrive at a power level.

The throughput of GF-NOMA suffers from three types of packet errors (named collision errors, lower-power-level errors, and higher-power-level errors) due to *power collisions*. A power collision occurs at a power level arrived by multiple packets, called a collided level. A power collision involves collision errors at the collided level. Then, this power collision necessarily causes packet errors at lower power levels than the collided level due to unsuccessful SIC. This type of the packet errors are named the lower-power-level errors. Also, such a power collision may destroy the balance of the SINR guaranteed by the predesigned power level, and thus, packet errors occur at higher power levels than the collided level. This type of the packet errors are named the higher-power-level errors. Thus, these packet errors depend on the packet arrival or offered load at each power level (called the per-level offered load).

The throughput has been actively analysed in some related works [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] to design the GF-NOMA systems. Some related works [4], [5], [6], [7], [8], [9], [10], [11], [12] analysed the throughput in the small-scale

Manuscript received 5 March 2024; revised 29 April 2024; accepted 2 May 2024. Date of publication 14 May 2024; date of current version 23 August 2024. This work was supported Grants-in-Aid for Scientific Research KAKENHI under Grant 23K16868; in part by the Support Center for Advanced Telecommunications Technology Research; and in part by the MIC/SCOPE under Grant JP235006102. *(Corresponding author: Takeshi Hirai.)*

The authors are with the Graduate School of Information Science and Technology, Osaka University, Suita 565-0821, Osaka, Japan (e-mail: hirai.takeshi.ist@osaka-u.ac.jp; r-oda@ist.osaka-u.ac.jp; wakamiya.naoki.ist@osaka-u.ac.jp).

Digital Object Identifier 10.1109/JIOT.2024.3400996

TABLE I
RELATED WORKS ABOUT THROUGHPUT ANALYSIS OF GF-NOMA

Limitation			Related Works
Scalability (Number of Power Levels)	**Accurately Modeled Packet Errors**	**Per-Level Offered Load**	
Only small-Scale (Two and three)	All errors		[4]–[12]
Large-Scale (General)	Collision errors		[13] [14]
	Lower-power-level and collision errors		[15]–[18]
	All packet errors	Uniform	Only our previous work [19]
		Uniform and non-uniform	**Only this work**

GF-NOMA systems. These works focused on the two or three levels, and thus, the analytical models counted all the type of packet errors, analysing the exact throughput. Other related works [13], [14], [15], [16], [17], [18], i.e., state-of-the-art works, built the throughput analytical model in the large-scale GF-NOMA systems, supporting the general number of the power levels. These works mainly supported the collision errors and/or lower-power-level errors with packet arrivals at each power level to simplify the analytical models. In particular, References [17] and [18] have roughly modeled the higher-power-level errors in the analytical models in addition to the collision errors and lower-power-level errors. Some of these works [15], [16] optimized the offered load by using the analytical models. The other related works [20] and [21] evaluated the throughput through the Monte Calro simulations.

However, these related works [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] have been unsuitable for analyzing the throughput in the large-scale GF-NOMA systems by a tradeoff between the scalability and analysis accuracy. The related works [4], [5], [6], [7], [8], [9], [10], [11], [12] have supported only the small-scale GF-NOMA systems and thus limited the scalability for the number of power levels. The related works [13], [14], [15], [16], [17], [18] have unsuitably reflected the higher-power-level errors in the analytical models to overlook the characteristics of the power level design, even in related works [17], [18], especially in the large-scale GF-NOMA systems. As a result, these related works [13], [14], [15], [16], [17], [18] underestimated or overestimated the throughput and thus involved underestimating or overestimating the optimal offered load. The related works [20] and [21] have consumed much time to obtain the throughput without the impacts of the randomness, especially in the large-scale GF-NOMA systems.

To overcome the limitations in these related works [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], this article proposes a scalable and accurate analytical framework for the throughput of GF-NOMA and presents the achievable performance based on this proposed framework in mMTC. Our key idea for the proposed analytical model is to suitably approximate the event probability of the higher-power-level errors focusing on the power level design in general per-level offered loads, extended from our previous work [19]. This idea enables analyzing the throughput of a large-scale GF-NOMA system more accurately than the existing analytical models [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] and optimizing the design of GF-NOMA. Also, this article optimizes the per-level offered load as examples of using our analytical framework: the throughput optimization problem and the energy minimization problem under satisfying the required throughput. To solve these complex optimization problems, this article builds an optimization framework based on a meta-heuristic particle swarm optimization (PSO) technique. In summary, this article provides the following contributions and findings.

1) This article proposes a scalable and accurate analytical throughput model based on the power level design reflecting higher-power-level errors more suitably than the existing analytical models in Section IV.
2) This article presents that the proposed analytical model follows throughput trends more accurately than the existing analytical models, even in the large-scale GF-NOMA systems, in Section V-B.
3) This article highlights that a nonuniform per-level offered load maximizes the throughput in Section VII-B and minimizes the energy consumption in GF-NOMA, including a large-scale system, by using our analytical framework in Section VII-C.

This article is organized as follows. The next section shows the differences between this work and related work. Section III describes the system model for our analytical framework. Section IV proposes an analytical model. Section V presents the validation of our model and analytical results of throughput trends of GF-NOMA on our analytical model. Section VI describes the optimization problems based on our analytical model. Section VII shows the optimization results. The final section concludes this article.

II. RELATED WORK

This section highlights the novelty of this article compared with the related works analyzing the performance of channel-inverted GF-NOMA [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], summarized in Table I, and other related works [20], [21], [22], [23], [24], [25]. The following paragraphs briefly describe the related works.

The first group of related works [4], [5], [6], [7], [8], [9], [10], [11], [12] focused on the performance analysis in the small-scale GF-NOMA systems with a specific number of levels, like two and three levels. These works contained all packet errors to enumerate the combinations of packet arrivals at each level. Reference [4] analysed the performance of GF-NOMA with the uniform power selection schemes with an

online backoff mechanism. Some works [5] and [6] focused on the nonuniform power selection mechanisms depending on the channel gains. References [7] and [8] optimized power selection schemes in GF-NOMA. The other works [9], [10], and [11] revealed an optimal power selection scheme by a game theory tool in some conditions, such as the fading channels. In each work, selecting a level is associated with a strategy in a game. These analytical models [4], [5], [6], [7], [8], [9], [10], [11], [12] have been available for only two or three power levels, i.e., small-scale GF-NOMA systems. However, the performance characteristics in the large-scale GF-NOMA systems have remained unclear in these works.

Another group of related works [13], [14], [15], [16], [17], [18] have extended the analytical models to support the general number of levels than the three levels, i.e., large-scale GF-NOMA systems. Related works [13], [14] proposed an analytical model containing collision errors to simplify the model. Some works [15], [16] reflected the lower-power-level errors to the above model to improve this model. Other works [17], [18] partially contained the higher-power-level errors in addition to the collision errors and lower-power-level errors. Reference [17] assumed that a power collision at a level caused packet errors in all the higher power levels than the level. Reference [18] proposed that a power collision at a collided level impacted as many higher-power levels as the number of collided packets at the collided level. However, these works have modeled packet errors unsuitably enough to analyse and optimize the performance with the throughput of GF-NOMA.

The other works focused on the simulation-based throughput evaluations [20], [21]. Reference [20] proposed geographically selecting power levels to improve the throughput of the GF-NOMA with channel inversion and evaluated the throughput by the simulations. Reference [21] focused on reinforcement learning for GF-NOMA with only two power levels. However, these simulation-based optimization frameworks spend more time designing system parameters than the analysis-based ones, especially in the large-scale GF-NOMA systems.

Unlike these related works [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], our work proposes a scalable and accurate analytical model containing all patterns of packet errors suitably enough to formulate the throughput in GF-NOMA. Note that, our previous work [19] proposed an analytical model for only uniform per-level offered loads. This article significantly extends this previous work to nonuniform per-level offered loads and proposes an optimization framework to analyse the optimal per-level offered load based on the analytical model. Before the analysis and optimization, the next paragraph describes the system model of our target GF-NOMA.

Note that target GF-NOMA facilitates analyzing the throughput, and thus, this article focuses on the GF-NOMA rather than the NOMA without power controls [22], [23], [24] or code-domain NOMA [25].

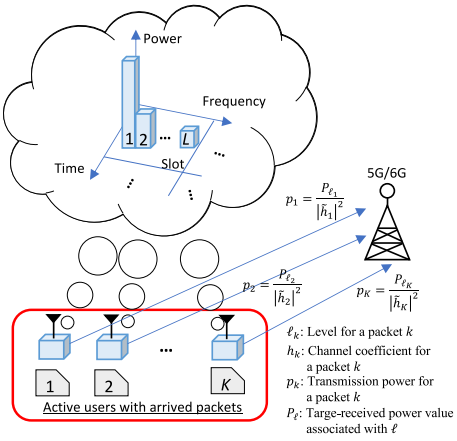

Fig. 1. Overview of the target GF-NOMA with L power levels for K packets.

TABLE II
NOTATIONS THROUGHOUT THIS ARTICLE

Notation	Definition
p_k	Transmission power of a user transmitting a packet k
P_ℓ	Target received power associated with a power level ℓ
η	Additive white Gaussian noise (AWGN) per slot with the spectrum density of N_0
h_k	Channel coefficient between a user transmitting k to its connected BS
ℓ_k	Power level selected by k
$\ell^{(c)}$	Power level where a power collision occurs
$\gamma_k^{(i)}$	SINR of k at the ith SIC iteration
Γ	Required SINR to satisfy the transmission rate R
K_ℓ	Number of arrived packets at ℓ
$\mathcal{V}_i$	Set of remaining packet signals at the ith SIC iteration
ψ_ℓ	Selection probability of ℓ
$\boldsymbol{\psi}$	Vector of selection probability at each power level
λ_ℓ	Average packet arrival rate or offered load at ℓ
$\boldsymbol{\lambda}$	Vector of per-level offered load

III. SYSTEM MODEL

Our analysis focuses on a typical GF-NOMA system [15], [16], [17], [18], [19], shown in Fig. 1. Let us consider K users become active at a slot. Each slot has L predesigned power levels, as shown in Fig. 1. K is a random variable. Such users, called active users, transmit their packets to a BS at the slot assuming an ALOHA-like protocol, namely the number of arrived packets per slot is also K. The following paragraphs describe the detailed procedures of the target GF-NOMA. The parameters used in this article are summarized in Table I.

Each active user selects a power level ℓ from L predesigned power levels according to a selection scheme. Let us consider that an user has a packet k and selects a power level ℓ_k to transmit k. This user calculates its transmission power, denoted as p_k, from its selected level ℓ_k according to *channel inversion* [13] as follows:

$$p_k = \frac{P_{\ell_k}}{|\tilde{h}_k|^2} \tag{1}$$

where P_ℓ is defined as the target-received power value associated with a power level ℓ, and $\tilde{h}_k$ is defined as its

estimated h_k, i.e., its channel coefficient between the user transmitting k and its connected BS. Note that this article assumes that the users can select all the levels to analyse the fundamental performance of GF-NOMA. This coefficient is estimated through the downlink reference signals periodically advertised by the BS in each slot. P_ℓ at each level ℓ is designed for the BS to guarantee the required SINR, denoted as Γ, as follows:

$$P_\ell = \Gamma(\Gamma+1)^{L-\ell}N_0 \tag{2}$$

where N_0 is defined as the spectral density of the additive white Gaussian noise (AWGN) over the slot, denoted as η. Note that a smaller ℓ in (2) is associated with a larger target-received power value based on the decoding order as follows:

$$P_1 > P_2 > \cdots > P_L. \tag{3}$$

Let us consider that each user selects ℓ at a probability, denoted as ψ_ℓ. Then, a selection scheme of power levels is expressed as follows:

$$\boldsymbol{\psi} = \left(\psi_1\ \psi_2 \cdots \psi_L\right) \tag{4}$$

where $\sum_{\ell=1}^{L}\psi_\ell = 1$.

Through the above protocol, the BS receives the following superposed signal at a slot:

$$y^{(0)} = \sum_{k\in\mathcal{V}_0}\sqrt{p_k}h_k s_k + \eta \tag{5}$$

where s_k is defined as the signal of a packet k, and $\mathcal{V}_i$ is defined as a set of remaining packet signals at the ith iteration, namely $\mathcal{V}_0$ is $\mathcal{V}_i$ at $i=0$ and thus contains all the transmitted packets at the slot. $\|\mathcal{V}_0\| = K$. The BS uses SIC, repeating the decoding and interference cancelation steps, to decode multiple signals in the superposed signal. At the ith decoding step, the BS experiences the following SINR, denoted as $\gamma_k^{(i)}$, for a packet k:

$$\gamma_k^{(i)} = \frac{p_k|h_k|^2}{\sum_{\hat{k}\in\mathcal{V}_i}p_{\hat{k}}|h_{\hat{k}}|^2 + N_0} \tag{6}$$

where $\mathcal{V}_i = \mathcal{V}_{i-1}\setminus\{k\}$. Each packet is delivered within a slot and requires a common transmission rate, denoted as R, based on the Shannon capacity as follows:

$$\log_2\left(1+\gamma_k^{(i)}\right) \geq R \Longleftrightarrow \gamma_k^{(i)} \geq 2^R - 1 = \Gamma. \tag{7}$$

Given the BS correctly decodes k satisfying (7), the BS cancels the replica of the decoded signal by using $\tilde{h}_k$ from the remaining superposed signal at the ith interference cancelation step as follows:

$$\begin{aligned} y^{(i)} &= y^{(i-1)} - \sqrt{p_k}\tilde{h}_k s_k \\ &\approx \sum_{\hat{k}\in\mathcal{V}_i}\sqrt{p_{\hat{k}}}h_{\hat{k}}s_{\hat{k}} + \eta. \end{aligned} \tag{8}$$

Based on the above procedures, the power level design of (2) enables the BS to satisfy (7) at no power collisions. Here, this article assumes that each user k perfectly estimates its channel coefficient, i.e., $\tilde{h}_k = h_k$, and thus, $p_k|h_k|^2 = P_\ell$, namely this article assumes the perfect SIC as the first step to analyse the key characteristics by per-level offered loads and power collisions. Then, the SINR at each iteration equals Γ at no power collisions. This condition provides the maximum interference, denoted as $I_\ell^{(+)}$, for successfully decoding a packet at a level ℓ as follows:

$$\begin{aligned} I_\ell^{(+)} &= \sum_{\hat{\ell}=\ell+1}^{L} P_{\hat{\ell}} = \sum_{\hat{\ell}=\ell+1}^{L}\Gamma(\Gamma+1)^{L-\hat{\ell}}N_0 \\ &= (\Gamma+1)^{L-\ell}N_0 - N_0 = \frac{P_\ell}{\Gamma} - N_0. \end{aligned} \tag{9}$$

By this characteristic, simply discussing per-level offered loads enables analyzing the throughput.

Based on the ALOHA-like protocol, the number of arrived packets at a slot, i.e., K, can be approximated by using the Poisson distribution with an average packet arrival rate or offered load, denoted as λ, under the typical mMTC use-cases [15], [16], [17], [18]. The probability of $K=m$ under λ is presented by using the following probability mass function, denoted as $q(m;\lambda)$:

$$\Pr(K=m) = q(m;\lambda) = \frac{e^{-\lambda}\lambda^m}{m!}. \tag{10}$$

Given λ and a selection scheme defined in (4), i.e., $\boldsymbol{\psi}$, a per-level offered load, expressed as $\boldsymbol{\lambda} = \left(\lambda_1\ \lambda_2 \cdots \lambda_L\right)$, has the offered load at each ℓ, denoted as λ_ℓ, as follows:

$$\lambda_\ell = \psi_\ell\lambda. \tag{11}$$

Here, λ_ℓ and λ have the following relationship:

$$\sum_{\ell=1}^{L}\lambda_\ell = \sum_{\ell=1}^{L}\psi_\ell\lambda = \lambda. \tag{12}$$

Based on the per-level offered load, i.e., $\boldsymbol{\lambda}$, the number of arrived packets at ℓ, denoted as K_ℓ, also follows an independent Poisson random variable with λ_ℓ:

$$\Pr(K_\ell = m) \ = q(m;\lambda_\ell) = \frac{e^{-\lambda_\ell}{\lambda_\ell}^m}{m!}. \tag{13}$$

In this model, the following section discusses our analytical model with a per-level offered load.

IV. Power Level Design-Based Analytical Throughput Model With Per-Level Offered Load

This section describes the proposed analytical model expressing the three patterns of packet errors due to power collisions and derives closed-form expressions for the throughput of GF-NOMA on the system model in Section III. Our analytical model focuses on the power level design of GF-NOMA, i.e., (2) and (9), to formulate the event probability of the higher-power-level errors. First, we discuss the packet error patterns at a power level and their event probabilities. Based on the probabilities, we formulate the normalized throughput, i.e., the average number of successfully decoded packets per slot, denoted as $T(\boldsymbol{\lambda})$, at the BS, given a per-level offered load [15], [16], [17], [18].

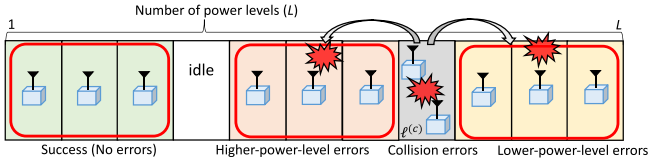

Fig. 2. Example of arrived packets at L power levels in a slot and three patterns of packet errors due to a power collision.

A. Three Patterns of Packet Errors

At a slot, a transmitted packet k may experience the following three packet error patterns: 1) collision errors; 2) lower-power-level errors; and 3) higher-power-level errors. Causing these patterns depends on the number of arrived packets at each power level at the slot. Here, K_ℓ is the number of arrived packets at ℓ, and then, the vector of the numbers is expressed as $\boldsymbol{\kappa} = \left(K_1\ K_2\ \cdots\ K_L\right)$. All the vectors are included in a set $\mathcal{K} = \{\boldsymbol{\kappa} | \boldsymbol{\kappa} \in \{0, 1, \ldots, \}^L\}$.

1) Collision Errors: Given that the two or more packets are transmitted at a power level ℓ, these packets experience collision errors. Then, ℓ is defined as a collided level, denoted as $\ell^{(c)}$, as shown in the gray level in Fig. 2. The set of $\boldsymbol{\kappa}$ causing this error pattern at a level ℓ, denoted as $\mathcal{C}_\ell$, is written as follows:

$$\mathcal{C}_\ell = \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, K_\ell \geq 2\}. \tag{14}$$

2) Lower-Power-Level Errors: Given that a power collision occurs, packet errors occur at all the lower power levels ℓ than a collided level $\ell^{(c)}$, i.e., $\ell = \ell^{(c)}+1, \ell^{(c)}+2, \ldots, L$, as shown in the yellow levels in Fig. 2. Each power collision causes SIC to unsuccessfully cancel the interference by packets in $\ell^{(c)}$ from the mixed signal. Thus, the lower-power-level errors necessarily occur with a power collision. This set of $\boldsymbol{\kappa}$ causing this error pattern at ℓ, denoted as $\mathcal{L}_\ell$, is formulated as follows:

$$\mathcal{L}_\ell = \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, K_\ell = 1, {}^\exists\ell^{(c)} < \ell, K_{\ell^{(c)}} \geq 2\}. \tag{15}$$

3) Higher-Power-Level Errors: A power collision may prevent the BS from satisfying the SINR condition in (7) for packets occupying higher power levels ℓ than a collided level $\ell^{(c)}$, i.e., $\ell < \ell^{(c)}$. These packets transmitted at $\ell < \ell^{(c)}$ experience the higher-power-level errors. Such a power collision destroys to guarantee the SINR in the predesigned power levels in (2). The set of $\boldsymbol{\kappa}$ causing this error pattern at ℓ, denoted as $\mathcal{H}_\ell$, is formulated as follows:

$$\mathcal{H}_\ell = \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, K_\ell = 1, {}^\exists\ell^{(c)} > \ell, K_{\ell^{(c)}} \geq 2, \gamma^{(i_k)}_{k(\ell)} < \Gamma\} \tag{16}$$

where $k(\ell)$ is the packet occupying a level ℓ at the i_kth SIC iteration in (6) at the decoding step for k. $\mathcal{H}_\ell$ includes too many $\boldsymbol{\kappa}$ to be exactly counted. For example, as the number of arrived packets at each level increases, packet errors at ℓ caused by the interference yielded from $\ell^{(c)}$ increase. This characteristic may fail to satisfy the SINR condition at much higher power levels than $\ell^{(c)}$.

The proposed analytical model approximates the higher-power level errors by using the following two sets: 1) $\mathcal{H}^-_\ell$ and 2) $\mathcal{H}^+_\ell$. These sets satisfy the following expression:

$$\mathcal{H}^-_\ell \subset \mathcal{H}_\ell \subset \mathcal{H}^+_\ell. \tag{17}$$

Designing these sets provides a tradeoff between the computational complexity and the accuracy of approximating the throughput. To this end, we focus on the power level design in (2) and (9), as described in the following paragraphs.

The set $\mathcal{H}^-_\ell$ is defined to contain only $\boldsymbol{\kappa}$ where a packet error necessarily occurs at a level ℓ higher than $\ell^{(c)}$. For such a $\boldsymbol{\kappa}$, one packet arrived at a level ℓ undergoes a packet error, given that a packet arrives at all the levels from $\ell+1$ to $\ell^{(c)}-1$. An example is shown in the orange levels in Fig. 2. In this figure, the collided level is occupied by $K_{\ell^{(c)}} \geq 2$ packets. Given $K_{\ell^{(c)}} \geq 2$ packets, a packet at ℓ undergoes the following interference power in the SINR based on (9):

$$\begin{aligned} I_\ell &= \sum_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} P_{\hat{\ell}} + K_{\ell^{(c)}} P_{\ell^{(c)}} = \sum_{\hat{\ell}=\ell+1}^{L} P_{\hat{\ell}} + \left(K_{\ell^{(c)}} - 1\right) P_{\ell^{(c)}} \\ &= I^{(+)}_\ell + \left(K_{\ell^{(c)}} - 1\right) P_{\ell^{(c)}}. \end{aligned} \tag{18}$$

$I_\ell > I^{(+)}_\ell$ and, this packet experiences a higher-power-level error. From the above discussions, $\mathcal{H}^-_\ell$ includes the following $\boldsymbol{\kappa}$:

$$\mathcal{H}^-_\ell = \left\{ \boldsymbol{\kappa} \in \mathcal{K} \,|\, \ell \leq {}^\forall\hat{\ell} < \min_{\substack{\ell^{(c)} > \ell \\ K_{\ell^{(c)}} \geq 2}} \ell^{(c)}, K_{\hat{\ell}} = 1 \right\}. \tag{19}$$

This set $\mathcal{H}^+_\ell$ is defined to include at least all the potential higher-power-level errors. The key characteristic to formulate such potential errors is that the power level design in (2) brings $(\Gamma+1) P_{\ell^{(c)}} = P_{\ell^{(c)}-1}$, namely accumulating $K_{\ell^{(c)}} = \Gamma+1$ packets provides the same interference power as $K_{\ell^{(c)}+1} = 1$ packet. This characteristic may prevent $\boldsymbol{\kappa}$ with $K_{\ell^{(c)}} > \Gamma + 1$ from satisfying the SINR condition; more specifically, no packets arrive at a level denoted as $\hat{\ell}$ where $\ell < \hat{\ell} < \ell^{(c)}$ and $K_{\ell^{(c)}} > \Gamma+1$, but a packet at ℓ may experience a higher-power-level error. An example is shown in Fig. 3. Then, the interference power for the level ℓ is written under $\ell^{(c)} = \ell+3$ as follows:

$$\begin{aligned} I_\ell &= P_{\ell+1} + (\Gamma + 1 + 2)\ P_{\ell^{(c)}} \\ &= P_{\ell+1} + P_{\ell+2} + 2\,P_{\ell^{(c)}} = \sum_{\tilde{\ell}=\ell+1}^{\ell^{(c)}} P_{\tilde{\ell}} + P_{\ell^{(c)}} > I^{(+)}_\ell. \end{aligned} \tag{20}$$

From the above discussions, this set contains the following $\boldsymbol{\kappa}$ with $K_{\ell^{(c)}} \geq \Gamma+1$ in addition to $\boldsymbol{\kappa} \in \mathcal{H}^-_\ell$:

$$\begin{aligned} \mathcal{H}^+_\ell &= \mathcal{H}^-_\ell \cup \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, \boldsymbol{\kappa} \notin \mathcal{H}^-_\ell, \\ &\qquad K_\ell = 1, {}^\exists\ell^{(c)} < \ell, K_{\ell^{(c)}} \geq \Gamma + 1\} \\ &\subset \mathcal{H}^-_\ell \cup \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, \boldsymbol{\kappa} \notin \mathcal{H}^-_\ell, \\ &\qquad K_\ell = 1, {}^\exists\ell^{(c)} < \ell, K_{\ell^{(c)}} \geq \lfloor \Gamma + 1 \rfloor\} \\ &= \{\boldsymbol{\kappa} \in \mathcal{K} \,|\, K_\ell = 1, {}^\exists\ell^{(c)} < \ell, K_{\ell^{(c)}} \geq \lfloor \Gamma + 1 \rfloor\} \\ &\qquad \cup \{\boldsymbol{\kappa} \in \mathcal{H}^-_\ell \,|\, {}^\forall\ell^{(c)} < \ell, K_{\ell^{(c)}} < \lfloor \Gamma + 1 \rfloor\}. \end{aligned} \tag{21}$$

B. Event Probability of Each Pattern of Packet Errors

This section discusses the event probability of each pattern at a level ℓ under the Poisson packet arrivals in (13). Here, $\boldsymbol{\kappa}$ depends on the per-level offered load, i.e., $\boldsymbol{\lambda}$, based on (13).

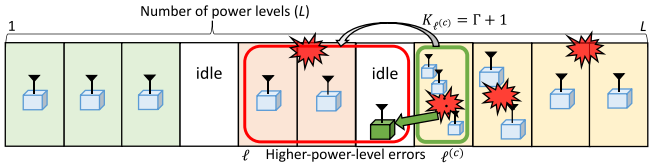

Fig. 3. Example of the upper-approximated set of higher-power-level errors $\mathcal{H}_\ell^+$ at using L power levels.

1) Event Probability of Collision Error: Collision errors occur under $K_\ell \geq 2$ based on (14). The event probability, denoted as $\Pr(\mathcal{C}_\ell)$, is written as follows:

$$\begin{aligned}\Pr(\mathcal{C}_\ell) &= \Pr(K_\ell \geq 2) = 1 - \Pr(K_\ell < 2) \\ &= 1 - \big(q(0;\lambda_\ell) + q(1;\lambda_\ell)\big). \end{aligned} \tag{22}$$

2) Event Probability of Lower-Power-Level Error: Lower-power-level errors occur under (15). This event probability is formulated as $\Pr(\mathcal{L}_\ell|K_\ell = 1)\Pr(K_\ell = 1)$, where $\Pr(\mathcal{L}_\ell|K_\ell = 1)$ is the following conditional event probability:

$$\begin{aligned}\Pr(\mathcal{L}_\ell|K_\ell = 1) &= 1 - \prod_{\hat{\ell}=1}^{\ell-1} \Pr\big(K_{\hat{\ell}} < 2\big) \\ &= 1 - \prod_{\hat{\ell}=1}^{\ell-1} \big(q\big(0;\lambda_{\hat{\ell}}\big) + q\big(1;\lambda_{\hat{\ell}}\big)\big) \\ &= 1 - \prod_{\hat{\ell}=1}^{\ell-1} \left(\frac{e^{-\lambda_{\hat{\ell}}}\lambda_\ell{}^0}{0!} + \frac{e^{-\lambda_{\hat{\ell}}}\lambda_{\hat{\ell}}}{1!}\right) \\ &= 1 - \prod_{\hat{\ell}=1}^{\ell-1} e^{-\lambda_{\hat{\ell}}}\big(1 + \lambda_{\hat{\ell}}\big) \\ &= 1 - \left(e^{-\sum_{\hat{\ell}=1}^{\ell-1}\lambda_{\hat{\ell}}}\right) \prod_{\hat{\ell}=1}^{\ell-1} \big(1 + \lambda_{\hat{\ell}}\big). \end{aligned} \tag{23}$$

3) Event Probability of Higher-Power Level Error: This event probability is formulated as $\Pr(\mathcal{H}_\ell|K_\ell = 1)\ \Pr(K_\ell = 1)$, where $\Pr(\mathcal{H}_\ell|K_\ell = 1)$ is the conditional probability expressed by the lower and upper-approximated probability based on (17) as follows:

$$\Pr\big(\mathcal{H}_\ell^-|K_\ell = 1\big)\ \leq \Pr(\mathcal{H}_\ell|K_\ell = 1)\ \leq \Pr\big(\mathcal{H}_\ell^+|K_\ell = 1\big). \tag{24}$$

The lower-approximated one is presented from (19) as (24).

The upper-approximated one is written from (21) as (25), shown at the bottom of the page. The first term shows the probability where higher power levels $\hat{\ell}$ than the level ℓ contain $K_{\hat{\ell}} \geq \lfloor \Gamma + 1 \rfloor$. The second term presents the event probability of $\boldsymbol{\kappa} \in \mathcal{H}_\ell^-$ under $K_{\hat{\ell}} < \lfloor \Gamma + 1 \rfloor$.

Note that $\Pr(K_\ell < \lfloor \Gamma + 1 \rfloor)$ in (26), shown at the bottom of the page, is presented as follows:

$$\Pr(K_\ell < \lfloor \Gamma + 1 \rfloor) = \sum_{K_\ell=0}^{\lfloor \Gamma+1-1 \rfloor} q(K_\ell;\lambda_\ell) = \sum_{K_\ell=0}^{\lfloor \Gamma \rfloor} q(K_\ell;\lambda_\ell) \tag{27}$$

and $\Pr(2 \leq K_\ell < \lfloor \Gamma + 1 \rfloor)$ in (26) is formulated as follows:

$$\begin{aligned}\Pr(2 \leq K_\ell < \lfloor \Gamma + 1 \rfloor) &= \sum_{K_\ell=2}^{\lfloor \Gamma+1-1 \rfloor} q(K_\ell;\lambda_\ell) \\ &= \sum_{K_\ell=2}^{\lfloor \Gamma \rfloor} q(K_\ell;\lambda_\ell). \end{aligned} \tag{28}$$

C. Normalized Throughput and Expectation of Packet Errors

The normalized throughput, i.e., $T(\boldsymbol{\lambda})$, is written as the summation of the expectations of the number of successfully received packets at each level, denoted as $T_\ell(\boldsymbol{\lambda})$, as follows:

$$T(\boldsymbol{\lambda}) = \sum_{\ell=1}^{L} T_\ell(\boldsymbol{\lambda}). \tag{29}$$

Let us denote $\mathcal{S}_\ell$ as the set of $\boldsymbol{\kappa}$ where successfully received packets at a level ℓ, and $N_{\mathcal{S}_\ell}$ as a random variable of the number of successfully received packets at ℓ. Each BS successfully receives at most a packet at each level, given only a packet arrives at the level ℓ, and the arrived packet experiences no lower-power-level and higher-power-level errors. Thus, this set is formulated as follows:

$$\mathcal{S}_\ell = \big\{\boldsymbol{\kappa} \in \mathcal{K}\,|\,K_\ell = 1, \overline{\mathcal{L}}_\ell \wedge \overline{\mathcal{H}}_\ell\big\}. \tag{30}$$

From (30), $T(\boldsymbol{\lambda})$ is presented as follows:

$$T(\boldsymbol{\lambda}) = \sum_{\ell=1}^{L} T_\ell(\boldsymbol{\lambda}) = \sum_{\ell=1}^{L} \mathbb{E}\big[N_{\mathcal{S}_\ell}\big] = \sum_{\ell=1}^{L} 1 \cdot \Pr(\mathcal{S}_\ell) \tag{31}$$

where $\Pr(\mathcal{S}_\ell)$, which is equal to T_ℓ based on (30), is written from (30) as follows:

$$\begin{aligned}\Pr(\mathcal{S}_\ell) &= \big(\Pr\big(\overline{\mathcal{H}}_\ell|K_\ell = 1\big)\big)\Pr(K_\ell = 1)\big(\Pr\big(\overline{\mathcal{L}}_\ell|K_\ell = 1\big)\big) \\ &= (1 - \Pr(\mathcal{H}_\ell|K_\ell = 1))q(1;\lambda_\ell)(1 - \Pr(\mathcal{L}_\ell|K_\ell = 1)). \end{aligned} \tag{32}$$

In (32), this article focuses on the lower or upper-approximated sets of the higher-power-level errors, i.e., $\mathcal{H}^\pm$,

$$\Pr\big(\mathcal{H}_\ell^-|K_\ell = 1\big) = \sum_{\ell^{(c)}=\ell+1}^{L} \left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} \Pr(K_{\hat{\ell}} = 1)\right)\Pr(K_{\ell^{(c)}} \geq 2)\right) = \sum_{\ell^{(c)}=\ell+1}^{L} \left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} q(1;\lambda_{\hat{\ell}})\right)\big(1 - q(0;\lambda_{\ell^{(c)}}) - q(1;\lambda_{\ell^{(c)}})\big)\right) \tag{25}$$

$$\begin{aligned}&\Pr\big(\mathcal{H}_\ell^+|K_\ell = 1\big) \\ &= \left(1 - \prod_{\hat{\ell}=\ell+1}^{L} \Pr\big(K_{\hat{\ell}} < \lfloor \Gamma + 1 \rfloor\big)\right) + \sum_{\ell^{(c)}=\ell+1}^{L} \left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} \Pr\big(K_{\hat{\ell}} = 1\big)\right)\Pr(2 \leq K_{\ell^{(c)}} < \lfloor \Gamma + 1 \rfloor)\left(\prod_{\tilde{\ell}=\ell^{(c)}+1}^{L} \Pr\big(K_{\tilde{\ell}} < \lfloor \Gamma + 1 \rfloor\big)\right)\right) \\ &= \left(1 - \prod_{\hat{\ell}=\ell+1}^{L} \Pr\big(K_{\hat{\ell}} < \lfloor \Gamma + 1 \rfloor\big)\right) + \sum_{\ell^{(c)}=\ell+1}^{L} \left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} q\big(1;\lambda_{\hat{\ell}}\big)\right)\Pr(2 \leq K_{\ell^{(c)}} < \lfloor \Gamma + 1 \rfloor)\left(\prod_{\tilde{\ell}=\ell^{(c)}+1}^{L} \Pr\big(K_{\tilde{\ell}} < \lfloor \Gamma + 1 \rfloor\big)\right)\right). \end{aligned} \tag{26}$$

and then, the throughput is also approximated. Let us introduce the upper or lower-approximated throughput, defined as $T^{\pm}(\boldsymbol{\lambda})$, respectively. $T^{\pm}(\boldsymbol{\lambda})$ has the following relationship for the exact throughput:

$$T^{-}(\boldsymbol{\lambda}) \le T(\boldsymbol{\lambda}) \le T^{+}(\boldsymbol{\lambda}). \tag{33}$$

Here, $\Pr(\mathcal{H}_\ell|K_\ell = 1)$ in (32) is approximated to $\Pr(\mathcal{H}_\ell^{\mp}|K_\ell = 1)$ for $T_\ell^{\pm}(\boldsymbol{\lambda})$, and thus, $T^{\pm}(\boldsymbol{\lambda})$ is presented by using $T_\ell^{\pm}(\boldsymbol{\lambda})$ as follows:

$$\begin{aligned} T^{\pm}(\boldsymbol{\lambda}) &= \sum_{\ell=1}^{L} T_\ell^{\pm}(\boldsymbol{\lambda}) \\ &= \sum_{\ell=1}^{L} \big(1 - \Pr\big(\mathcal{H}_\ell^{\mp}|K_\ell = 1\big)\big) q(1;\lambda_\ell)\big(1 - \Pr(\mathcal{L}_\ell|K_\ell = 1)\big). \end{aligned} \tag{34}$$

Additionally, we discuss the expectations of the number of packets experiencing collision errors, lower-power-level errors, and higher-power-level errors, including the lower and upper-approximated ones. The numbers of these packet errors are denoted as $N_{\mathcal{C}}$, $N_{\mathcal{L}}$, and $N_{\mathcal{H}^{\pm}}$, respectively. Also, $N_{\mathcal{C}_\ell}$, $N_{\mathcal{L}_\ell}$, and $N_{\mathcal{H}_\ell^{\pm}}$ are denoted as these numbers at ℓ. The expectations of $N_{\mathcal{C}}$, denoted as $\mathbb{E}[N_{\mathcal{C}}]$, is written from (22) by using $\mathbb{E}[N_{\mathcal{C}_\ell}]$ as follows:

$$\begin{aligned} \mathbb{E}[N_{\mathcal{C}}] &= \sum_{\ell=1}^{L} \mathbb{E}\big[N_{\mathcal{C}_\ell}\big] = \sum_{\ell=1}^{L} \sum_{K_\ell=2}^{\infty} K_\ell \cdot \Pr(K_\ell) \\ &= \sum_{\ell=1}^{L} \Big(\mathbb{E}[K_\ell] - \big(0 \cdot \Pr(K_\ell = 0) \ + 1 \cdot \Pr(K_\ell = 1)\big) \Big) \\ &= \sum_{\ell=1}^{L} \big(\lambda_\ell - \lambda_\ell e^{-\lambda_\ell}\big). \end{aligned} \tag{35}$$

The expectation of $N_{\mathcal{L}}$, denoted as $\mathbb{E}[N_{\mathcal{L}}]$, is written from (23) by using $\mathbb{E}[N_{\mathcal{L}_\ell}]$ as follows:

$$\mathbb{E}[N_{\mathcal{L}}] = \sum_{\ell=1}^{L} \mathbb{E}\big[N_{\mathcal{L}_\ell}\big] = \sum_{\ell=1}^{L} \Pr(\mathcal{L}_\ell|K_\ell = 1) q(1;\lambda_\ell). \tag{36}$$

The expectation of $N_{\mathcal{H}^{\pm}}$, denoted as $\mathbb{E}[N_{\mathcal{H}^{\pm}}]$, is written from (26) and (25) by using $\mathbb{E}[N_{\mathcal{H}_\ell^{\pm}}]$ as follows:

$$\mathbb{E}[N_{\mathcal{H}^{\pm}}] = \sum_{\ell=1}^{L} \mathbb{E}\Big[N_{\mathcal{H}_\ell^{\pm}}\Big] = \sum_{\ell=1}^{L} \Pr\big(\mathcal{H}_\ell^{\pm}|K_\ell = 1\big) q(1;\lambda_\ell). \tag{37}$$

D. Special Case: Uniform Selection Scheme

This section discusses closed-form expressions of the throughput in the uniform selection scheme of power levels as a special case of the above analytical model. In this case, the selection scheme is $\boldsymbol{\psi} = \big(\frac{1}{L}\ \frac{1}{L} \cdots \frac{1}{L}\big)$, and thus, the offered load at ℓ, i.e., λ_ℓ, shows the same value and then, λ_ℓ is presented as follows:

$$\lambda_\ell = \hat{\lambda} = \frac{\lambda}{L}. \tag{38}$$

The uniform per-level offered load, denoted as $\boldsymbol{\lambda}_u$, is presented as $\boldsymbol{\lambda}_u = \big(\hat{\lambda}\ \hat{\lambda} \cdots \hat{\lambda}\big)$. Equation (38) enables simplifying some expressions. First, the probability of each error pattern is rewritten. The lower-power-level errors occur at the following probability transformed from (23):

$$\begin{aligned} \Pr\Big(\mathcal{L}_\ell|K_\ell = 1, \hat{\lambda}\Big) &= 1 - \prod_{\hat{\ell}=1}^{\ell-1} \Pr\big(K_{\hat{\ell}} < 2\big) \\ &= 1 - \Big(q\Big(0;\hat{\lambda}\Big) + q\Big(1;\hat{\lambda}\Big)\Big)^{\ell-1}. \end{aligned} \tag{39}$$

The higher-power-level errors occur at the lower-approximated probability transformed from (25) as follows:

$$\begin{aligned} &\Pr\Big(\mathcal{H}_\ell^{-}|K_\ell = 1, \hat{\lambda}\Big) \\ &= \sum_{\ell^{(c)}=\ell+1}^{L} \left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1} \Pr\big(K_{\hat{\ell}} = 1\big) \right) \Pr\big(K_{\ell^{(c)}} \ge 2\big) \right) \\ &= \Big(1 - q(0;\hat{\lambda}) - q(1;\hat{\lambda})\Big) \sum_{\ell^{(c)}=\ell+1}^{L} q\Big(1;\hat{\lambda}\Big)^{\ell^{(c)}-\ell-1} \\ &= \Big(1 - q(0;\hat{\lambda}) - q(1;\hat{\lambda})\Big) \frac{1 - q\Big(1;\hat{\lambda}\Big)^{L-\ell}}{1 - q\Big(1;\hat{\lambda}\Big)}. \end{aligned} \tag{40}$$

The higher-power-level errors occur at the upper-approximated probability shown in (41), at the bottom of the next page, transformed from (26). Then, the throughput with the uniform selection scheme, denoted as $T_u^{\pm}(\hat{\lambda})$, is presented from (34) as follows:

$$\begin{aligned} T_u^{\pm}\Big(\hat{\lambda}\Big) = \sum_{\ell=1}^{L} &\Big(1 - \Pr\Big(\mathcal{H}_\ell^{\mp}|K_\ell = 1, \hat{\lambda}\Big)\Big) \\ &\times q\Big(1;\hat{\lambda}\Big)\Big(1 - \Pr\Big(\mathcal{L}_\ell|K_\ell = 1, \hat{\lambda}\Big)\Big). \end{aligned} \tag{42}$$

E. Characteristics of Analysis Accuracy and Throughput

1) Analysis Accuracy: The proposed analytical model has the following two approximations: 1) approximating the sets of higher-power-level errors by using $\mathcal{H}_\ell^{\pm}$ and 2) approximating packet arrivals by using the Poisson distribution. For the former one, the upper-approximated throughput, i.e., $T^{+}(\boldsymbol{\lambda})$, is expected to show better approximations than the lower-approximated one $T^{-}(\boldsymbol{\lambda})$ because $\Pr(K_{\tilde{\ell}} < \lfloor \Gamma + 1 \rfloor)$ observed in $T^{-}(\boldsymbol{\lambda})$ is more rarely than $q(0;\lambda)$ and $q(1;\lambda)$ observed in $T^{+}(\boldsymbol{\lambda})$. For the latter one, the Poisson approximations are expected to be accurate enough for the analysis from the related works [13], [14], [15], [16], [17], [18].

The proposed model shows better approximations than the existing analytical model [18] that approximates the throughput most accurately in the other existing models [15], [16], [17]. Our proposed model contains all the error patterns more suitably than [18]. The upper-approximated throughput in [18], denoted as $\hat{T}^{+}(\boldsymbol{\lambda})$, is written as follows:

$$\begin{aligned} \hat{T}^{+}(\boldsymbol{\lambda}) &= \sum_{\ell=1}^{L} \Big(q(1;\lambda_\ell) - \Pr(\mathcal{L}_\ell|K_\ell = 1) q(1;\lambda_\ell) \Big) \\ &= \sum_{\ell=1}^{L} \lambda_\ell e^{-\lambda_\ell} \Big(e^{-\sum_{\hat{\ell}=1}^{\ell-1} \lambda_{\hat{\ell}}} \Big) \prod_{\hat{\ell}=1}^{\ell-1} \big(1 + \lambda_{\hat{\ell}}\big). \end{aligned} \tag{43}$$

In the special case of the uniform selection scheme, $\hat{T}_u^+(\hat{\lambda})$ is reformulated by (38) as follows:

$$\begin{aligned}\hat{T}_u^+\left(\hat{\lambda}\right) &= \sum_{\ell=1}^{L} q\left(1;\hat{\lambda}\right)\left(q\left(0;\hat{\lambda}\right)+q\left(1;\hat{\lambda}\right)\right)^{\ell-1} \\ &= \sum_{\ell=1}^{L} \hat{\lambda}e^{-\hat{\lambda}}\left(\left(1+\hat{\lambda}\right)e^{-\hat{\lambda}}\right)^{\ell-1} \\ &= \sum_{\ell=1}^{L} \hat{\lambda}\left(1+\hat{\lambda}\right)^{\ell-1}e^{-\ell\hat{\lambda}}.\end{aligned} \tag{44}$$

$\hat{T}^+(\boldsymbol{\lambda})$ reflects the impacts of lower-power-level errors but includes no impacts of the higher-power-level errors. Thus, $\hat{T}^+(\boldsymbol{\lambda})$ overestimates the exact throughput more significantly than $T^+(\boldsymbol{\lambda})$.

The lower-approximated throughput in [18], denoted as $\hat{T}^-(\boldsymbol{\lambda})$, is based on an upper-approximated set of higher-power-level errors. Reference [18] assumed that given that a power collision occurs at $\ell^{(c)}$, higher-power-level errors occur at $K_{\ell^{(c)}}$ levels higher than $\ell^{(c)}$. In other words, Reference [18] assumed that $K_{\ell^{(c)}}$ packet arrivals require the same number of the vacant levels as the higher-power-level errors. Then, $\hat{T}^-(\boldsymbol{\lambda})$ is formulated in (45), shown at the bottom of the page, as well as (34). At the special case, i.e., GF-NOMA using the uniform power selection scheme, $\hat{T}^-(\boldsymbol{\lambda})$ is transformed by (38) to (46), shown at the bottom of the page. As discussed in Section IV-A, power collisions with $K_{\ell^{(c)}} \leq \lfloor \Gamma+1 \rfloor$ only requires a vacant level to experience no higher-power-level errors, and thus, $\hat{T}^-(\boldsymbol{\lambda})$ underestimates the exact throughput more significantly than $T^-(\boldsymbol{\lambda})$. At power collisions with $K_{\ell^{(c)}} > \lfloor \Gamma+1 \rfloor$, $T^-(\boldsymbol{\lambda})$ is expected to overestimate the exact throughput more significantly than $\hat{T}^-(\boldsymbol{\lambda})$. This event is expected to occur much less frequently than the above event around the peak throughput, and thus, the proposed model is expected to show more accurate throughput than the existing model. In Section V-B, we validate these characteristics by the quantitative evaluations.

2) Throughput Versus Per-Level Offered Load: First, this section qualitatively discusses the characteristics of the normalized throughput of per-level offered load yielded by the uniform selection scheme, i.e., $\boldsymbol{\lambda}_u$ discussed in Section IV-D, based on the three packet errors. The throughput shows an unimodal trend at a larger $\hat{\lambda}$. As $\hat{\lambda}$ increases, the number of occupied power levels increases. Also, at a larger $\hat{\lambda}$, collision errors at ℓ occur more frequently as expressed in (22). Such a larger $\hat{\lambda}$ causes the lower-power-level errors and higher-power-level errors more frequently based on (23) and (24). As a result, as $\hat{\lambda}$ increases, the throughput increases and then decreases. Additionally, the uniform selection scheme experiences higher throughput at a higher power level. A higher power level experiences higher-power-level errors more frequently, and a lower power level experiences lower-power-level errors more frequently. In particular, the first power level experiences no lower-power-level errors, and the Lth power level experiences no higher-power-level errors. Note that $\Pr(\mathcal{H}_\ell|K_\ell=1)$ is more minor than $\Pr(\mathcal{L}_\ell|K_\ell=1)$ based on the definitions in (15) and (16). As a result, these characteristics allow a lower power level to experience higher throughput.

Next, this section qualitatively discusses the throughput in per-level offered loads obtained by the nonuniform selection schemes. Such per-level offered loads provide different balances of event probabilities of the packet errors from the uniform per-level offered load. Along to λ_ℓ, power collisions at a level ℓ occur monotonically, involving the collision errors. The two other packet errors occur based on the power collisions. As a typical selection scheme, let us consider a higher selection probability at a lower power level, called a rising selection scheme, namely $\psi_\ell < \psi_{\tilde{\ell}}$, where $\ell < \tilde{\ell}$. The per-level offered load obtained by this scheme causes power collisions at a lower power level more frequently. As a result, such a per-level offered load causes higher-power-level errors more frequently than the uniform one. Also, this per-level offered load causes a higher power level to experience collision errors less frequently. This characteristic allows such a per-level offered load to experience the lower-power-level

$$\begin{aligned}\Pr\left(\mathcal{H}_\ell^+|K_\ell=1,\hat{\lambda}\right) &= \left(1-\prod_{\hat{\ell}=\ell+1}^{L}\Pr\left(K_{\hat{\ell}}<\lfloor\Gamma+1\rfloor\right)\right)+\sum_{\ell^{(c)}=\ell+1}^{L}\left(\left(\prod_{\hat{\ell}=\ell+1}^{\ell^{(c)}-1}\Pr\left(K_{\hat{\ell}}=1\right)\right)\Pr(2\leq K_{\ell^{(c)}}<\lfloor\Gamma+1\rfloor)\left(\prod_{\tilde{\ell}=\ell^{(c)}+1}^{L}\Pr\left(K_{\tilde{\ell}}<\lfloor\Gamma+1\rfloor\right)\right)\right) \\ &= \left(1-\Pr\left(K_{\hat{\ell}}<\Gamma+1\right)^{L-\ell}\right)+\sum_{\ell^{(c)}=\ell+1}^{L}\left(q\left(1;\hat{\lambda}\right)^{\ell^{(c)}-\ell-1}\Pr(2\leq K_{\ell^{(c)}}<\lfloor\Gamma+1\rfloor)\Pr\left(K_{\tilde{\ell}}<\Gamma+1\right)^{L-\ell^{(c)}}\right) \\ &= \left(1-\Pr\left(K_{\hat{\ell}}<\Gamma+1\right)^{L-\ell}\right)+\Pr(2\leq K_{\ell^{(c)}}<\lfloor\Gamma+1\rfloor)\frac{\Pr\left(K_{\tilde{\ell}}<\Gamma+1\right)^{L-\ell}-q\left(1;\hat{\lambda}\right)^{L-\ell}}{\Pr\left(K_{\tilde{\ell}}<\Gamma+1\right)-q\left(1;\hat{\lambda}\right)}\end{aligned} \tag{41}$$

$$\hat{T}^-(\boldsymbol{\lambda})=\sum_{\ell=1}^{L}\lambda_\ell e^{-\lambda_\ell}\left(e^{-\sum_{\hat{\ell}=1}^{\ell-1}\lambda_{\hat{\ell}}}\right)\prod_{\hat{\ell}=1}^{\ell-1}\left(1+\lambda_{\hat{\ell}}\right)\left(\prod_{\hat{\ell}=\ell+1}^{L}e^{-\lambda_{\hat{\ell}}}(1+\lambda_{\hat{\ell}})+\sum_{\ell^{(c)}=\ell+2}^{L}\left(q\left(\ell^{(c)};\lambda_{\ell^{(c)}}\right)\prod_{\ell_w=\ell+1}^{\ell^{(c)}-1}e^{-\lambda_{\ell_w}}\prod_{\hat{\ell}=\ell^{(c)}+1}^{L}e^{-\lambda_{\hat{\ell}}}\left(1+\lambda_{\hat{\ell}}\right)\right)\right) \tag{45}$$

$$\hat{T}_u^-\left(\hat{\lambda}\right)=\sum_{\ell=1}^{L}\hat{\lambda}\left(1+\hat{\lambda}\right)^{\ell-1}e^{-\ell\hat{\lambda}}\left(e^{-\hat{\lambda}(L-\ell)}(1+\hat{\lambda})^{(L-\ell)}+\sum_{\ell^{(c)}=\ell+2}^{L}\left(q(\ell^{(c)};\hat{\lambda})e^{-\hat{\lambda}(L-\ell-1)}(1+\hat{\lambda})^{(L-\ell^{(c)})}\right)\right) \tag{46}$$

errors less frequently than the uniform one. In addition to these trends, decreasing λ_ℓ involves reducing the probability, where a packet is correctly received at ℓ. Balancing these characteristics depends on the throughput. In contrast, using a higher selection probability at a higher power level, called a sloping selection scheme, shows the inverse characteristics, namely the lower-power-level errors are dominant to the throughput.

V. Numerical Result

A. Analysis Parameters

This section describes the analytical results by our analytical model proposed in Section IV. To validate the analysis accuracy of the proposed model, we compared the analytical results with the following two models. The first one was the exhaustive search, providing the exact throughput under the Poisson packet arrivals. This comparison supports validating the approximation accuracy of the event probability of higher-power-level errors. The second one was the Monte Carlo simulation assuming no Poisson packet arrivals to validate the Poisson approximations, discussed in Section IV-E, in each simulation, 1000 users randomly selected their power levels, given a per-level offered load. The following metric of approximation errors was introduced as an error percentage:

$$\epsilon = \left| \frac{F(\boldsymbol{\lambda})}{T(\boldsymbol{\lambda})} - 1 \right| \times 100\% \tag{47}$$

where $F(\cdot) \in \{T^{\pm}(\cdot), \hat{T}^{\pm}(\cdot)\}$, and $T(\boldsymbol{\lambda})$ is the throughput obtained by the exhaustive search or Monte Carlo simulation. Each plot was averaged over 10^6 samples. The required SINR, i.e., Γ, was set to $4 \approx 6$ dB [16].

We analysed the throughput with the per-level offered loads yielded by the two typical nonuniform selection schemes in addition to the uniform selection scheme. The first scheme was the following sloping selection scheme:

$$\psi_\ell = \frac{L - \ell + 1}{\sum_{\hat{\ell}=1}^{L} \left(L - \hat{\ell} + 1 \right)}. \tag{48}$$

The second scheme was the following rising selection scheme:

$$\psi_\ell = \frac{\ell}{\sum_{\hat{\ell}=1}^{L} \hat{\ell}}. \tag{49}$$

Fig. 4 shows the selection probability in these schemes at $L = 5$, compared with the uniform selection scheme. As shown in Fig. 4, (48), and (49) realized the nonuniform selection schemes. Also, we compared the analysis accuracy of the proposed analytical model with the existing analytical model [16], [17], [18], as discussed in Section IV-E.

B. Analysis Accuracy

This section describes the analysis accuracy of the proposed model in the uniform, rising, and sloping selection schemes of the power levels. Fig. 5 shows the normalized throughput in these selection schemes at $L = 5, 10, 15$. The horizontal axis are the average packet arrival rate (offered load), i.e., λ, and the vertical axis are the normalized throughput. Fig. 6 shows the expectations of the number of packet errors in the proposed model with increasing the offered load (i.e., average packet arrival rate), i.e., λ, at $L = 5$ in these selection schemes, respectively. The horizontal axis are the offered load and the vertical axis are the expectations of the number of packet errors.

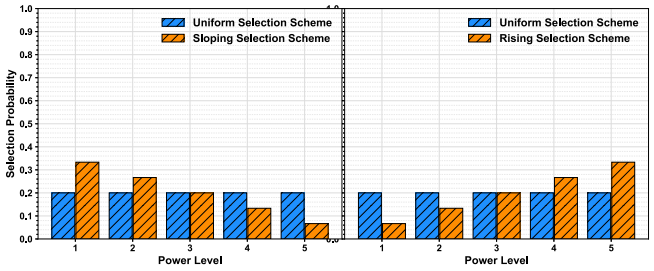

Fig. 4. Selection probability in the sloping scheme (left) in (48) and rising scheme in (49) (right) at $L = 5$, compared with the uniform selection scheme, respectively.

First, we describe the analysis accuracy in the per-level offered load yielded by the uniform selection scheme. The left graph in Fig. 5 highlights that the proposed lower-approximated model achieved underestimating the exact throughput by only 0.4% at $\lambda = L = 5$, the upper-approximated throughput was only 0.01% higher than the exact throughput. Thus, the proposed model achieved the error percentage, i.e., ϵ, within 0.4%. At the same λ, the lower-approximated and upper-approximated ones in the existing model [16], [17], [18] showed ϵ of 25% and 51%, respectively. These results highlighted that the proposed model formulated the GF-NOMA throughput more accurately than the existing model, focusing on modeling higher-power-level errors, as discussed in Section IV-E. This graph also emphasizes that the proposed model showed enough accurate throughput at a large number of power levels. At $L = 15$, the proposed lower-approximated model provided ϵ of only 0.001% for the Monte Carlo simulation at $\lambda = 5$, providing the peak throughput. At the same parameters, the existing lower-approximated one provided 27% less throughput than the Monte Carlo simulation.

1) Sloping Selection Scheme: Second, we describe those in the sloping selection scheme. The middle graph in Fig. 5 highlights that the proposed model achieved ϵ of only 0.08% for the exhaustive search at $\lambda = L = 5$; in particular, the upper-approximated throughput was only 0.003% higher than the exact throughput. At the same λ, the lower-approximated and upper-approximated ones in the existing model showed ϵ of 11% and 48% for the exhaustive search, respectively. At $L = 15$, the proposed lower-approximated throughput provided only 0.05% lower throughput than the Monte Carlo simulation at $\lambda = 4$, providing the peak throughput; at the same parameters, the existing lower-approximated one provided 13% less throughput than the Monte Carlo simulation.

2) Rising Selection Scheme: Finally, we focus on the analytical results with the rising selection scheme. The right graph in Fig. 5 highlights that this selection scheme provided less accurate throughput at a larger offered load more remarkably

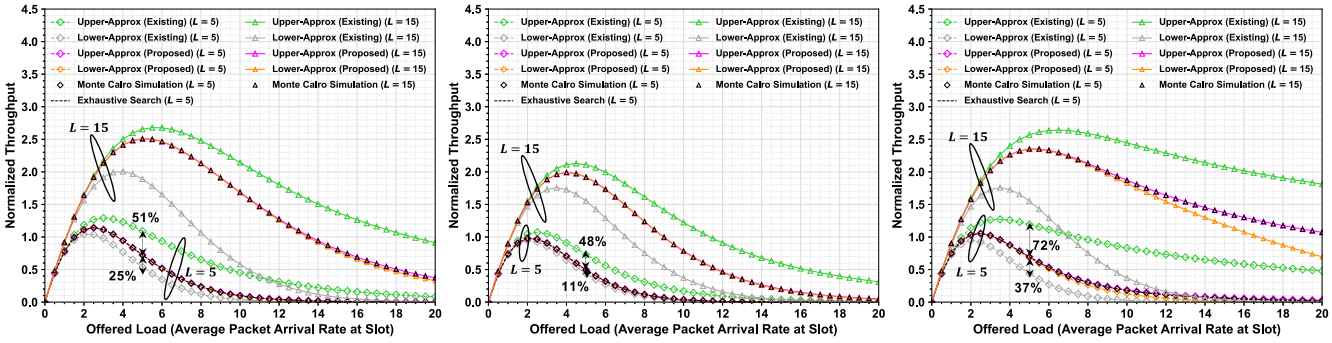

Fig. 5. Normalized throughput with increasing the offered load, i.e., λ, at $L = 5, 15$ in the uniform (left), sloping (middle), and rising (right) selection schemes, respectively.

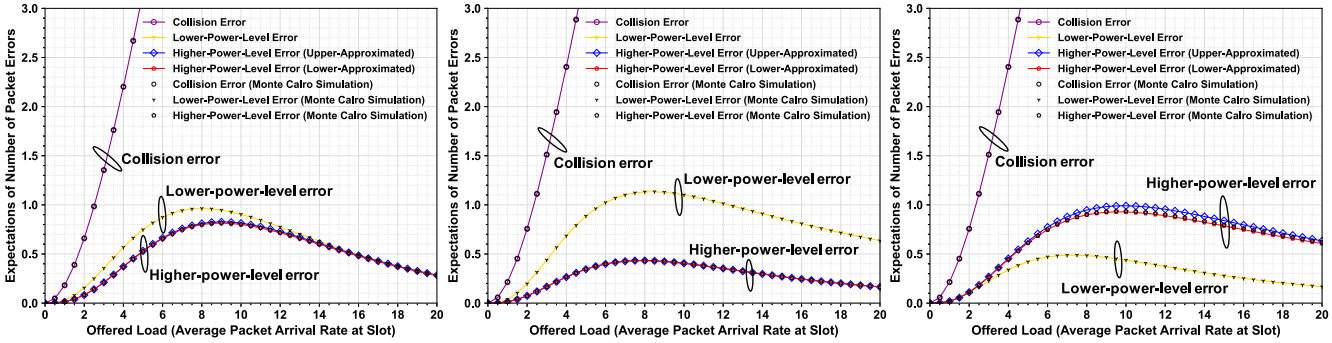

Fig. 6. Expectations of the number of packet errors in the proposed model with increasing the offered load, i.e., λ, at $L = 5$ in the uniform (left), sloping (middle), and rising (right) selection schemes, respectively.

than the other schemes, but the proposed model showed enough accuracy around the peak throughput. In the proposed model, $T^{-}(\boldsymbol{\lambda})$ and $T^{+}(\boldsymbol{\lambda})$ were ϵ of only 2% and 0.07% for the exhaustive search at $\lambda = L = 5$, respectively. Then, $\hat{T}^{-}(\boldsymbol{\lambda})$ and $\hat{T}^{+}(\boldsymbol{\lambda})$ in the existing model showed ϵ of 37% and 72%, respectively. At $L = 15$, the proposed lower-approximated throughput provided 0.1% lower throughput than the Monte Carlo simulation at $\lambda = 5$, providing the peak throughput. Then, the existing lower-approximated one provided 34% less throughput than the Monte Carlo simulation. Even in the schemes, these results emphasized that the proposed model achieved enough accurate throughput to express the exact throughput around the peak one.

C. Throughput and Packet Error Analysis

This section analyses the normalized throughput and packet errors. First, Fig. 7 shows the normalized throughput for the offered load in the three selection schemes at $L = 5, 15$. Here, the normalized throughput is the upper-approximated throughput because the upper-approximated one was more accurate than the lower-approximated one in Fig. 5. The horizontal axis is the offered load, i.e., λ, and the vertical axis is the normalized throughput. Fig. 8 shows the expectations of the number of packet errors in the proposed model at each power level at $L = 5$ at the offered load $\lambda = L/2 = 2.5$ around the peak throughput, in these selection schemes, respectively. The horizontal axis are the power level, i.e., ℓ, and the vertical axis are the throughput and the expectations of the number of packet errors.

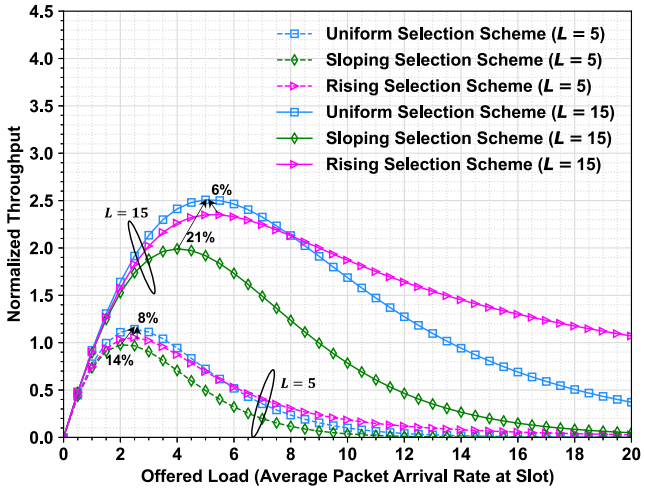

Fig. 7. Normalized throughput with increasing the offered load, i.e., λ, at $L = 5, 15$, in the uniform, sloping, and rising selection schemes.

In Fig. 7, we compared the characteristics of the normalized throughput with the per-level offered loads obtained by the selection schemes. At $L = 5$, the uniform selection scheme increased the normalized throughput over the sloping and rising ones by 14% and 8%, respectively. At $L = 15$, the uniform selection scheme increased the normalized throughput over the sloping and rising ones by 21% and 6%, respectively. The following subsections break down packet errors and results per level in each selection scheme.

1) Uniform Selection Scheme: The left graph in Fig. 6 highlights that the lower-power-level errors were more

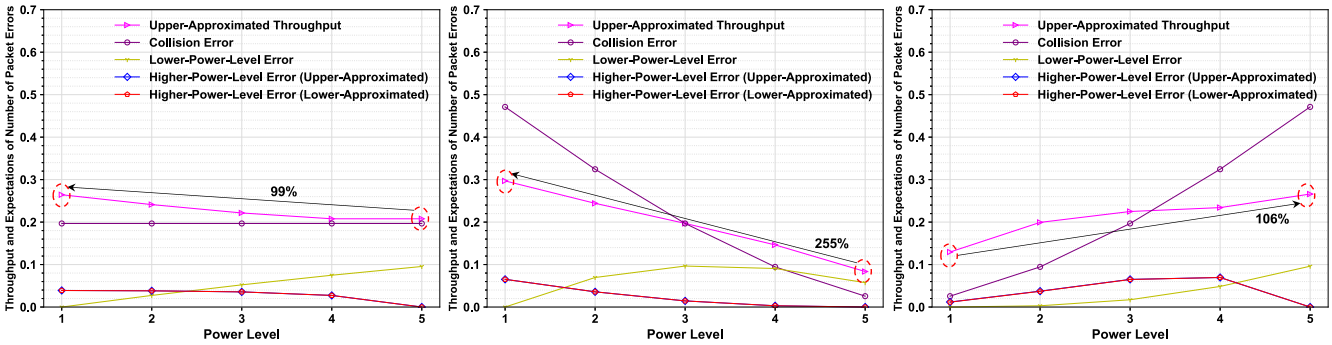

Fig. 8. Expectations of the number of packet errors in the proposed model at each power level at $L = 5$ at the offered load $\lambda = (L/2) = 2.5$ around the peak throughput, in the uniform (left), sloping (middle), and rising (right) selection schemes, respectively.

dominant than the higher-power-level errors, as discussed in Section IV-E. At $\lambda = L = 5$, the lower-power-level errors occurred 40% more frequently than the higher-power-level errors. Also, at a larger λ, collision errors were more dominant to the throughput, and the other errors occurred less frequently. The left graph in Fig. 8 shows that a lower-power level provided lower throughput, selecting the first level experienced 99% higher throughput than selecting the fifth level.

2) Sloping Selection Scheme: Also, the middle graph in Fig. 6 highlights that the sloping selection scheme increased the impacts of the lower-power-level errors on the throughput over the uniform selection scheme, as discussed in Section IV-E. At $\lambda = L = 5$, the lower-power-level errors occurred 157% more frequently than the higher-power-level errors. This characteristic increased the difference of the throughput over the power levels, as shown in the middle graph in Fig. 8, selecting the first level experienced 255% higher throughput than selecting the fifth level.

3) Rising Selection Scheme: Also, this scheme showed the inverse characteristics of the packet errors to the other two schemes. The right graph in Fig. 6 presents that the higher-power-level errors occurred more frequently than the lower-power-level errors, as discussed in Section IV-E, the difference of expectations was 33% at $\lambda = L = 5$. Also, the right graph in Fig. 8 shows that a higher power level experienced lower throughput, specifically selecting the fifth level showed 106% higher throughput than selecting the first level.

VI. PSO-Based Optimization Framework for Per-Level Offered Load

This section discusses a PSO-based optimization framework for a per-level offered load as applicable examples of the analytical model. This section discusses the following two typical optimization problems using the proposed analytical model: 1) the throughput maximization problem and 2) the energy minimization problem under a throughput condition. In the following subsections, we describe these optimization problems and how to solve them by PSO, which is a basic optimization technique and has a few design parameters.

A. Throughput Maximization Problem

This optimization problem aims to maximize the throughput of GF-NOMA based on the analytical model. To guarantee the exact throughput, we use the lower-approximated throughput. This problem is formulated as follows:

$$\max_{\boldsymbol{\lambda}} \quad T^{-}(\boldsymbol{\lambda}) \tag{50a}$$

$$\text{s.t.} \quad \lambda_\ell > 0 \quad {}^{\forall}\ell. \tag{50b}$$

Considering the uniform selection scheme of the power levels, the above problem has the following constraint:

$$\max_{\boldsymbol{\lambda}} \quad T^{-}(\boldsymbol{\lambda}) \tag{51a}$$

$$\text{s.t.} \quad \lambda_\ell = \lambda_{\hat{\ell}} > 0 \quad {}^{\forall}\ell\, {}^{\forall}\hat{\ell}. \tag{51b}$$

To simply solve this problem, we transform this problem for λ as follows:

$$\max_{\lambda} \quad T_u^{-}\left(\frac{\lambda}{L}\right) \tag{52a}$$

$$\text{s.t.} \quad \lambda > 0. \tag{52b}$$

In this optimization framework with the existing analytical model, named the existing framework, $T^{-}(\boldsymbol{\lambda})$ is replaced to $\hat{T}^{-}(\boldsymbol{\lambda})$, and $T_u^{-}\lambda/L$ is replaced to $\hat{T}_u^{-}\lambda/L$.

B. Energy Minimization Problem

This problem minimizes the energy consumption, given $\boldsymbol{\lambda}$ provides equal or higher throughput than the optimally uniform per-level offered load, denoted as $\boldsymbol{\lambda}_u^*$, i.e., $T(\boldsymbol{\lambda}) \geq T(\boldsymbol{\lambda}_u^*)$. The energy consumed by each user depends on the selection scheme of the power levels and its channel from its transmission power by (1). The channel depends on the user distribution model, and thus, this article focuses on the expectation of the transmission power of an user k as follows:

$$\mathbb{E}[p_k] = \sum_{\ell_k=1}^{L} \frac{P_{\ell_k}}{|h_k|^2} \psi_{\ell_k} = \sum_{\ell_k=1}^{L} \frac{P_{\ell_k}}{|h_k|^2} \frac{\lambda_{\ell_k}}{\lambda}. \tag{53}$$

The objective function to minimize this energy is simply formulated as the following expectations of target-received power values based on the selection scheme based on $\boldsymbol{\lambda}$:

$$\mathbb{E}[P_\ell \mid \boldsymbol{\lambda}] = \sum_{\ell=1}^{L} P_\ell \frac{\lambda_\ell}{\lambda}. \tag{54}$$

To guarantee $T(\boldsymbol{\lambda}) \geq T(\boldsymbol{\lambda}_u^*)$, we use a constraint where the lower-approximated throughput for $\boldsymbol{\lambda}^*$ is equal to the

upper-approximated throughput for $\boldsymbol{\lambda}_u^*$ or larger. Then, the optimization problem is presented as follows:

$$\max_{\boldsymbol{\lambda}} \ \mathbb{E}[P_\ell \mid \boldsymbol{\lambda}] \tag{55a}$$

$$\text{s.t} \ \ \lambda_\ell > 0 \quad {}^\forall\ell \tag{55b}$$

$$T^-(\boldsymbol{\lambda}) \geq T^+\left(\boldsymbol{\lambda}_u^*\right). \tag{55c}$$

Note that the uniform per-level offered load shows the following energy: $\mathbb{E}[P_\ell \mid \boldsymbol{\lambda}_u^*] = 1/L \sum_{\ell=1}^{L} P_\ell$. As well as the above throughput maximization problem, in the existing framework, $T^-(\boldsymbol{\lambda})$ is replaced to $\hat{T}^-(\boldsymbol{\lambda})$, and $T^+(\boldsymbol{\lambda}_u)$ is replaced to $\hat{T}^+(\boldsymbol{\lambda}_u)$.

C. PSO-Based Optimization Framework

This article uses a PSO technique to solve the above nonlinear and complex problems directly. Let us consider that $\mathfrak{Z}$ is the set of particles. In PSO, each particle $z \in \mathfrak{Z}$ on a position at time t, denoted as $\boldsymbol{x}_z^{(t)}$, explores a better position for a fitness function than the past positions of itself and other particles. In our framework, $\boldsymbol{x}_z^{(t)} = \boldsymbol{\lambda} = \left(\lambda_1 \ \lambda_2 \cdots \lambda_L\right)$. Its position is updated by using its velocity, denoted as $\boldsymbol{v}_z^{(t)}$, as follows:

$$\boldsymbol{x}_z^{(t+1)} = \boldsymbol{x}_z^{(t)} + \boldsymbol{v}_z^{(t+1)}. \tag{56}$$

The velocity is defined as follows:

$$\begin{aligned}\boldsymbol{v}_z^{(t+1)} = w_m \boldsymbol{v}_z^{(t)} &+ \boldsymbol{r}_p^{(t)} \odot \left(\boldsymbol{x}_{z,p}^{(t)} - \boldsymbol{x}_z^{(t)}\right) \\ &+ \boldsymbol{r}_g^{(t)} \odot \left(\boldsymbol{x}_g^{(t)} - \boldsymbol{x}_z^{(t)}\right)\end{aligned} \tag{57}$$

where w_m is a constant value, and $\boldsymbol{r}_p \in [0, w_p]^L$ and $\boldsymbol{r}_g \in [0, w_g]^L$ are the vectors with the random values. w_p and w_g are the constant values to decide the ranges of the random values, respectively. Also, $\boldsymbol{x}_{z,p}^{(t)}$ is the personally best position in a particle z until t, and $\boldsymbol{x}_g^{(t)}$ is the globally best position in all the particles until t. To solve the throughput optimization problem, these positions are designed as follows:

$$\begin{aligned}\boldsymbol{x}_{z,p}^{(t)} &= \operatorname{argmax}_{\{\boldsymbol{x}_z^{(\hat{t})} | \hat{t} \leq t\}} T^-\left(\boldsymbol{x}_z^{(\hat{t})}\right) \\ \boldsymbol{x}_g^{(t)} &= \operatorname{argmax}_{\{\boldsymbol{x}_{z,p}^{(t)} | z \in \mathfrak{Z}\}} T^-\left(\boldsymbol{x}_{z,p}^{(t)}\right)\end{aligned} \tag{58}$$

and the positions for the energy minimization problem are presented as follows:

$$\begin{aligned}\boldsymbol{x}_{z,p}^{(t)} &= \operatorname{argmax}_{\{\boldsymbol{x}_z^{(\hat{t})} | \hat{t} \leq t\}} E\left(\boldsymbol{x}_z^{(\hat{t})}\right) \\ \boldsymbol{x}_g^{(t)} &= \operatorname{argmax}_{\{\boldsymbol{x}_{z,p}^{(t)} | z \in \mathfrak{Z}\}} E\left(\boldsymbol{x}_{z,p}^{(t)}\right)\end{aligned} \tag{59}$$

where $E(\cdot)$ is an energy function based on (55a) with the penalty for the constraint of the throughput as follows:

$$E(\boldsymbol{\lambda}) = \begin{cases} \mathbb{E}\left[P_\ell \mid \boldsymbol{\lambda}_u^*\right] + T^+\left(\boldsymbol{\lambda}_u^*\right) - T^-(\boldsymbol{\lambda}) & T^-(\boldsymbol{\lambda}) < T^+\left(\boldsymbol{\lambda}_u^*\right) \\ \mathbb{E}[P_\ell \mid \boldsymbol{\lambda}] & T^-(\boldsymbol{\lambda}) \geq T^+\left(\boldsymbol{\lambda}_u^*\right). \end{cases} \tag{60}$$

In the penalty term, i.e., the function at $T^-(\boldsymbol{\lambda}) < T^+(\boldsymbol{\lambda}_u^*)$, the second term for unsatisfying the throughput condition aims for each particle to move positions achieving higher throughput than $T^+(\boldsymbol{\lambda}_u^*)$ based on (55c). The first term, i.e., $\mathbb{E}[P_\ell \mid \boldsymbol{\lambda}_u^*]$, avoids that the value in $T^-(\boldsymbol{\lambda}) < T^+(\boldsymbol{\lambda}_u^*)$ is smaller than $\mathbb{E}[P_\ell \mid \boldsymbol{\lambda}]$ in $T^-(\boldsymbol{\lambda}) \geq T^+(\boldsymbol{\lambda}_u^*)$. Also, the initial position and velocity of each particle are randomly selected as follows:

$$\boldsymbol{x}_i^{(0)} \in [0, 1]^L, \quad \boldsymbol{v}_i^{(0)} = \boldsymbol{0}. \tag{61}$$

For the rapid convergence, $\mathfrak{Z}$ contains a particle z with $\boldsymbol{x}_z^{(0)} = \boldsymbol{\lambda}_u^*$. Note that each function used in the above explanations is replaced along with the optimization problems described in Sections VI-A and VI-B.

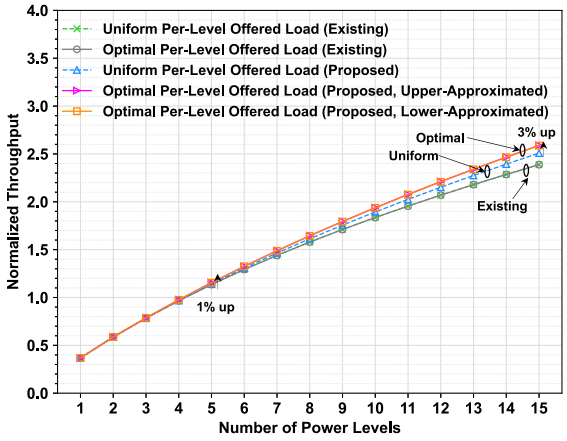

Fig. 9. Achievable normalized throughput in the optimal per-level offered load with increasing the number of power levels.

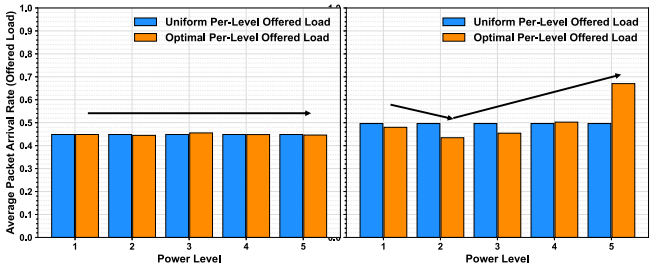

Fig. 10. Optimal per-level offered load in the PSO-based optimization framework with the existing model (left) and proposed model (right) at $L = 5$, compared with the optimally uniform per-level offered load in each framework.

D. Characteristics of Optimal Per-Level Offered Load

First, this section qualitatively discusses the characteristics of the optimal per-level offered load by the proposed framework. For the throughput maximization problem, the optimal $\boldsymbol{\lambda}$ depends on balancing the event probabilities of packet error patterns, as discussed in Section IV-E and analytical results shown in Section V. A key characteristic is that the lower-power-level errors occur more frequently than the higher-power-level errors. Based on this characteristic, the per-level offered load with the rising selection scheme is expected to be near-optimal. Another key characteristic is that the first power level experiences no lower-power-level errors, and the Lth power level experiences no higher-power-level errors. Suitably balancing these characteristics yields the maximum throughput. For the energy minimization problem, the expected energy depends on the optimal per-level offered load in addition to balancing the characteristics. Selecting a lower power level at a larger probability suppresses the

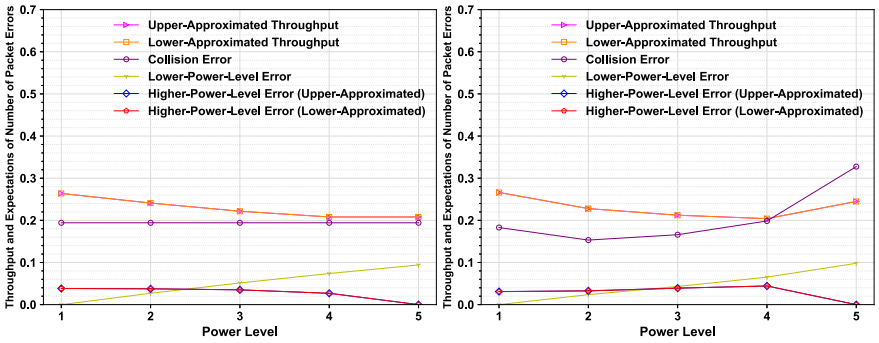

Fig. 11. Normalized throughput and the expectations of packet errors at each level in $\boldsymbol{\lambda}_u^*$ (left) and $\boldsymbol{\lambda}^*$ (right) in the proposed framework.

TABLE III
ANALYSIS PARAMETERS

Parameter	Value	Parameter	Value
Γ	$4 \approx 6$ dB	N_0	-110 dBm
$\|\mathcal{I}\|$	1000	Number of steps	100
w_m	0.8	w_p and w_g	0.2

expected energy, and as a result, a rising selection scheme is expected to be near-optimal.

Next, this section qualitatively compares the optimal per-level offered load of the proposed framework with the existing framework that uses the existing analytical model. The proposed framework provides a better per-level offered load than the existing framework for the accuracy of the analytical model. The existing model shows that higher-power-level errors are more dominant to the throughput than the proposed model. This approximation error is expected to cause the existing framework to provide a near-uniform selection scheme as the optimal per-level offered load. For the energy minimization problem, the existing framework provides the throughput less accurately, and thus prevents satisfying (55c) more remarkably than the proposed framework. As a result, $\boldsymbol{\lambda}$ is less likely to reach $\boldsymbol{\lambda}^*$.

VII. Optimizing Per-Level Offered Load

This section describes the optimal per-level offered load in the two optimization problems discussed in Section VI as application examples of the proposed analytical model. First, this section describes the optimization results in the throughput maximization problem. Second, this section describes the per-level offered load minimizing the energy under satisfying the required throughput.

A. Parameters

This section used typical wireless parameters. $\Gamma = 4$, the same value in Section V. $N_0 = -110$ dBm, used in many related works [20]. The PSO parameters were typical values, summarized in Table III. We compared the optimal $\boldsymbol{\lambda}$ obtained from the proposed framework with the existing framework, which is the PSO-based framework using the existing analytical model denoted in the related work [18].

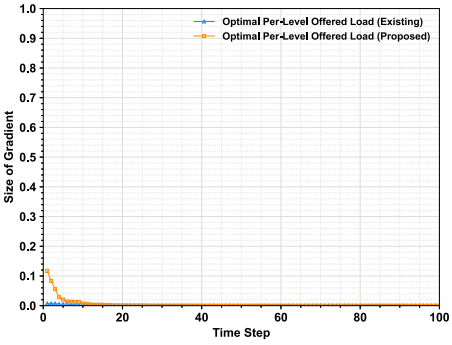

Fig. 12. Gradient of the throughput over time steps in PSO at $L = 5$.

B. Throughput Maximization

This section shows the results of the throughput maximization. Fig. 9 shows the achievable throughput in the per-level offered load optimized by the proposed optimization framework, i.e., the PSO-based framework with the proposed analytical throughput model, compared with the existing one. The horizontal axis represents the number of power levels, i.e., L, and the vertical axis shows the proposed analytical throughput for the optimal per-level offered load by each framework. This graph showed that optimal per-level offered loads formulated by (52a), named uniform per-level offered load, and (50a), named optimal per-level offered load, by the proposed and existing frameworks.

From Fig. 9, our framework showed more throughput than the existing framework by optimizing $\boldsymbol{\lambda}$. In the proposed framework, $\boldsymbol{\lambda}^*$ provided 1% more throughput than $\boldsymbol{\lambda}_u^*$ at $L = 5$. Then, the proposed framework achieved at most 2% more throughput than the existing framework in the optimal per-level offered load. Also, at $L = 15$, $\boldsymbol{\lambda}^*$ optimized by the proposed framework provided at most 3% and 8% more throughput than $\boldsymbol{\lambda}_u^*$ and the optimized $\boldsymbol{\lambda}$ by the existing framework, respectively. Our results highlighted that the nonuniform selection scheme had small impacts on the throughput. Note that, the upper-approximated throughput provided only 0.04% higher than the lower-approximated throughput, but this difference in the existing analytical model was 18% at $\boldsymbol{\lambda}$ optimized by the existing framework at $L = 5$. Our results highlighted that the proposed analysis was accurate enough even during the optimization process.

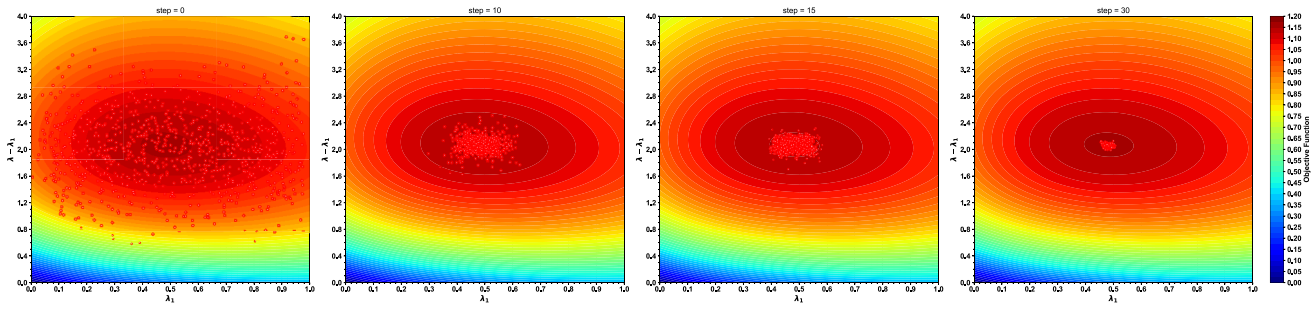

Fig. 13. Positions of particles at the PSO time steps $t = 0, 10, 15, 30$ at $L = 5$ in the proposed PSO-based throughput optimization framework.

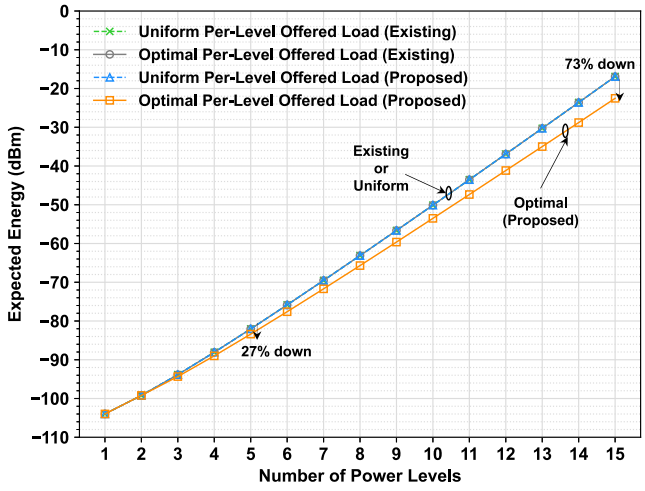

Fig. 14. Expected energy in the optimized per-level offered load with increasing the number of power levels.

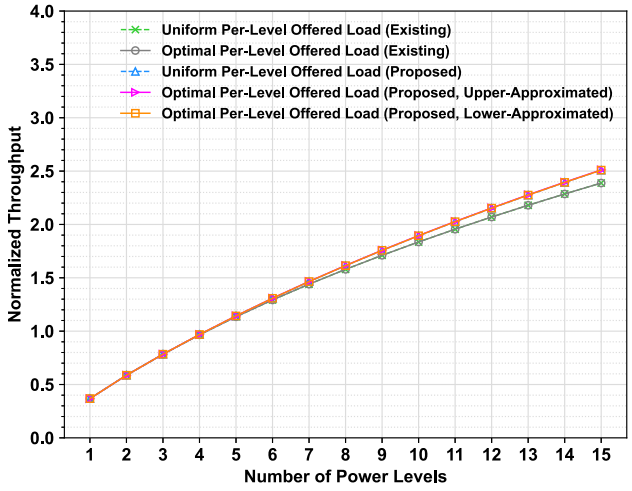

Fig. 15. Normalized throughput in the optimized per-level offered load with increasing the number of power levels.

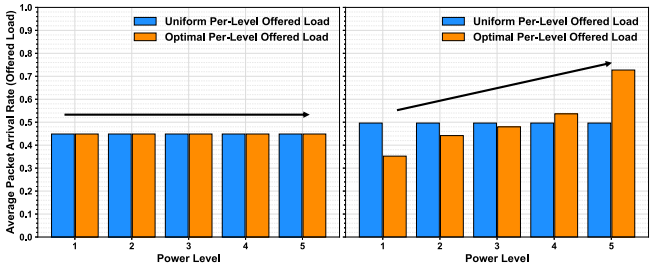

Fig. 16. Optimal per-level offered loads in the PSO-based optimization framework with the existing model (left) and the proposed model (right) at $L = 5$.

From Fig. 10 and 11, we analyse the per-level offered load. Fig. 10 shows the per-level offered load optimized by these frameworks, the left graph shows $\boldsymbol{\lambda}$ by the existing framework, and the right graph shows $\boldsymbol{\lambda}$ by the proposed framework. The horizontal axis are the power level, and the vertical axis are the offered load. Also, Fig. 11 shows the normalized throughput and expectations of the number of packet errors in these optimal per-level offered loads. The horizontal axis and vertical axis are the same as Fig. 8, respectively. The right figure in Fig. 10 emphasizes that the proposed framework optimally balanced the lower and higher-power-level errors; the per-level offered load was mainly based on the rising selection scheme except for the first level. The optimal per-level offered load showed an unimodal per-level offered load; λ_1, impacting no lower-power-level errors, was higher than λ_2, and monotonically increasing the offered load at the other levels involved suppressing the lower-power-level errors, as discussed in Section VI-D, while suppressing the higher-power-level errors. From the left graph in 11, $\boldsymbol{\lambda}^*$ decreased the lower-power-level errors below $\boldsymbol{\lambda}_u^*$, although $\boldsymbol{\lambda}^*$ increased the higher-power-level errors above $\boldsymbol{\lambda}_u^*$. $\boldsymbol{\lambda}^*$ highlighted the throughput at $\ell = 5$ was higher than $\ell = 2, 3, 4$. In contrast, the existing framework discovered an uniform-like per-level offered load as the optimal $\boldsymbol{\lambda}$.

Figs. 12 and 13 show the optimization process. The horizontal axis of Fig. 12 is the time step, and the vertical axis is the size of the numerical gradient of the throughput. Fig. 13 shows the positions of particles, shown in circles, in the proposed framework at $L = 5$ at $t = 0, 10, 15, 30$. The horizontal axis is the offered load at the first level, i.e., λ_1, and the vertical axis is the totally-remained offered load, i.e., $\lambda - \lambda_1$. The background color shows $T(\boldsymbol{\lambda})$. The value was converged around the PSO time step $t = 30$ in Fig. 13, and then the numerical gradient was small enough for zero in Fig. 12.

C. Energy Minimization

This section shows the results of the energy minimization. Fig. 14 shows the expected energy achieved by the proposed

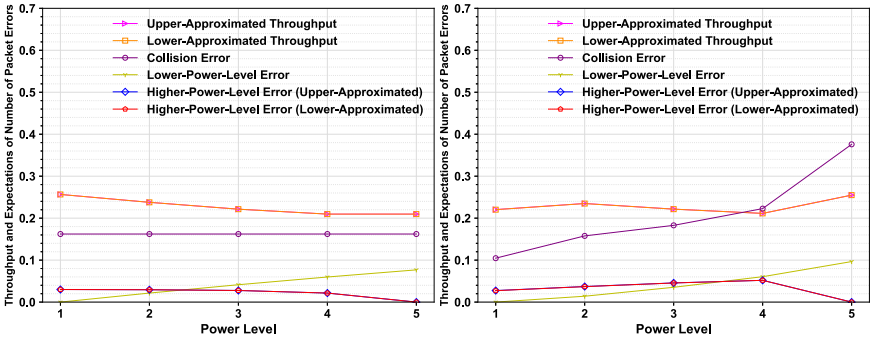

Fig. 17. Normalized throughput and the expectations of packet errors at each level in $\boldsymbol{\lambda}$ optimized by the existing framework (left) and the proposed framework (right).

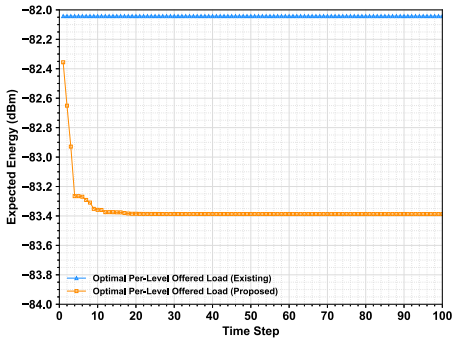

Fig. 18. Expected energy over time steps in PSO at $L = 5$.

and existing framework in the dBm domain with increasing the number of the power levels. The horizontal axis is the number of power levels, i.e., L, and the vertical axis is the expected energy. Fig. 15 shows the normalized throughput with increasing the number of power levels. The horizontal and vertical axis are the same as Fig. 9. From Fig. 14, at $L = 5$, $\boldsymbol{\lambda}^*$ achieved 27% (i.e., 1.3 dB) less energy than $\boldsymbol{\lambda}_u^*$ and the optimal $\boldsymbol{\lambda}$ optimized by the existing one. Then, in Fig. 15, $\boldsymbol{\lambda}^*$ provided 0.01% higher throughput than $\boldsymbol{\lambda}_u^*$ with an enough small approximation error. Also, the proposed framework suppressed the energy consumption more significantly at a larger L. At $L = 15$, the optimal per-level offered load consumed 73% (i.e., 5.6 dB) less energy than the existing optimal per-level offered load and $\boldsymbol{\lambda}_u^*$.

Fig. 16 shows the optimal per-level offered loads at $L = 5$ to analyse the above results more deeply. The left graph shows $\boldsymbol{\lambda}$ optimized by the existing framework, and the right graph shows $\boldsymbol{\lambda}_u^*$ and $\boldsymbol{\lambda}^*$ by the proposed framework described in Section VI. The horizontal and vertical axis are the same as Fig. 10. Fig. 16 highlights that the existing framework did not find per-level offered loads satisfying the condition, and then, the optimal per-level offered load was the same uniform per-level offered load. The proposed method discovered an optimally nonuniform $\boldsymbol{\lambda}$, as shown in Fig. 16. The optimal per-level offered load presented a larger offered load at a higher power level, i.e., a rising selection scheme. The selection scheme has a small selection probability of the first level enough to reduce the energy consumed in the first level, as discussed in Section VI-D. From the characteristic of GF-NOMA analysed in Section VII-B, the throughput was less sensitive for the offered load, and then, GF-NOMA suppressed the energy while guaranteeing similar throughput to the optimally uniform per-level offered load. Fig. 17 shows the throughput and the expectations of the number of packet errors in $\boldsymbol{\lambda}$ optimized by each framework. The horizontal and vertical axis are the same as Fig. 8. As a result of the optimization, the optimal per-level offered load showed lower throughput at the first level than the optimally uniform power selection scheme.

As well as the throughput maximization, we confirmed the optimization process in the energy minimization in Figs. 18 and 19. Fig. 18 shows the expected energy over time steps in PSO at $L = 5$. The horizontal axis is the time step, and the vertical axis is the expected energy, calculating no numerical gradients of the value because of no smooth function. Fig. 19 shows the position of each particle at time step $t = 0, 10, 15, 30$ in the solution space defined by $E(\cdot)$ in (60). The horizontal and vertical axis are the same as Fig. 13. The color in the solution space shows $E(\cdot)$ in (60). Each point shows the position of each particle, and its color shows the throughput condition of (55c), specifically, the particles satisfying the condition are painted red, and the others are painted blue. The expected energy was converged around $t = 30$ in Fig. 18, and then, the particles also converged to the optimal per-level offered load at $t = 30$ in Fig. 19.

VIII. Conclusion

This article proposed a scalable analytical framework for the throughput of GF-NOMA in mMTC and analysed the achievable performance for the per-level offered load optimized by the proposed analytical framework with our model. The proposed model focuses on the power level design of GF-NOMA to reflect the higher-power-level errors more accurately, even in the large-scale GF-NOMA systems, than the existing analytical models. Our analytical model enables scalably optimizing the per-level offered load for the typical optimization problems of GF-NOMA: the throughput maximization and the energy minimization. Our analytical results showed that the proposed analytical model provided an approximation error percentage of 0.4% for the exact throughput at $L = 5$, although the existing one showed that of 25%. These results highlighted that the proposed analytical model

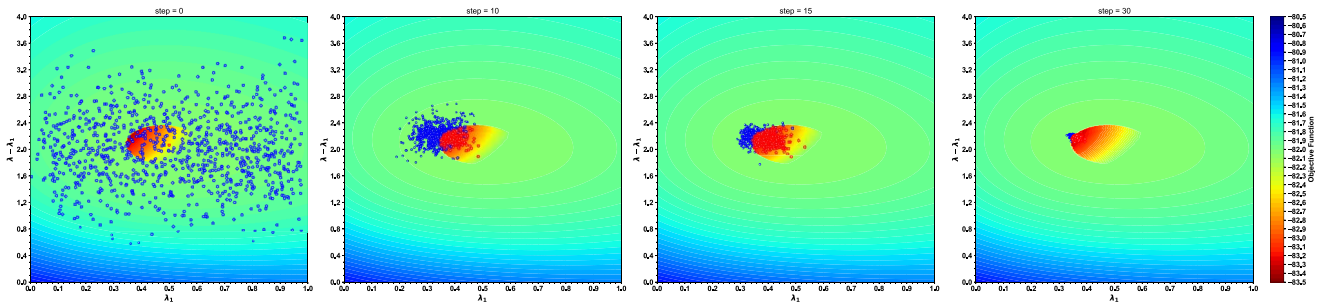

Fig. 19. Positions of particles in time steps $t = 0, 10, 15, 30$ in $L = 5$ in the proposed PSO-based energy minimization framework.

achieved enough accurate throughput to optimally design new access protocols for the large-scale GF-NOMA systems. As one of the examples, our optimization results emphasized that our proposed PSO-based optimization framework discovered a 27% more energy-efficient per-level offered load at $L = 5$ than the optimally uniform per-level offered load and the optimal per-level offered load on the existing analytical model. Then, the proposed one guarantees higher throughput than the optimally uniform one.

REFERENCES

[1] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[2] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, "Towards enabling critical mMTC: A review of URLLC within mMTC," *IEEE Access*, vol. 8, pp. 131796–131813, 2020.

[3] M. U. A. Siddiqui, H. Abumarshoud, L. Bariah, S. Muhaidat, M. A. Imran, and L. Mohjazi, "URLLC in beyond 5G and 6G networks: An interference management perspective," *IEEE Access*, vol. 11, pp. 54639–54663, 2023.

[4] J.-B. Seo, B. C. Jung, and H. Jin, "Online backoff control for NOMA-enabled random access procedure for cellular networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1158–1162, Jun. 2021.

[5] J.-B. Seo, S. Pack, and H. Jin, "Uplink NOMA random access for UAV-assisted communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8289–8293, Aug. 2019.

[6] J.-B. Seo, H. Jin, and B. C. Jung, "Multichannel uplink NOMA random access: Selection diversity and bistability," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1515–1519, Sep. 2019.

[7] Z. Chen, Y. Liu, L. X. Cai, Y. Cheng, R. Zhang, and M. Han, "Performance study of random access NOMA with truncated channel inversion power control," in *Proc. IEEE ICC*, 2021, pp. 1–6.

[8] Z. Chen, Y. Liu, S. Khairy, L. X. Cai, Y. Cheng, and R. Zhang, "Optimizing non-orthogonal multiple access in random access networks," in *Proc. IEEE 91st VTC*, 2020, pp. 1–5.

[9] J. Choi, "Multichannel NOMA-ALOHA game with fading," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4997–5007, Oct. 2018.

[10] J. Choi and J.-B. Seo, "Evolutionary game for hybrid uplink NOMA with truncated channel inversion power control," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8655–8665, Dec. 2019.

[11] J. Choi and Y. Ko, "On asymmetric game for NOMA-ALOHA under fading," in *Proc. IEEE 95th VTC*, 2022, pp. 1–5.

[12] J.-B. Seo, S. De, and H. Jin, "Real-time transmission control for multichannel NOMA random access systems," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 8984–8995, May 2023.

[13] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.

[14] J. Choi, "Layered non-orthogonal random access with SIC and transmit diversity for reliable transmissions," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1262–1272, Mar. 2018.

[15] Y. Jin and T.-J. Lee, "Throughput analysis of NOMA-ALOHA," *IEEE Trans. Mobile Comput.*, vol. 21, no. 4, pp. 1463–1475, Apr. 2022.

[16] W. Yu, C. H. Foh, A. U. Quddus, Y. Liu, and R. Tafazolli, "Throughput analysis and user barring design for uplink NOMA-enabled random access," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6298–6314, Oct. 2021.

[17] J. Choi, "On throughput bounds of NOMA-ALOHA," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 165–168, Jan. 2022.

[18] J.-B. Seo, B. C. Jung, and H. Jin, "Performance analysis of NOMA random access," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2242–2245, Nov. 2018.

[19] T. Hirai, R. Oda, and N. Wakamiya, "Power level design-aware throughput analysis of grant-free power-domain NOMA in mMTC," in *Proc. IEEE GLOBECOM*, 2022, pp. 105–110.

[20] H. Jiang, Q. Cui, Y. Gu, X. Qin, X. Zhang, and X. Tao, "Distributed layered grant-free non-orthogonal multiple access for massive MTC," in *Proc. IEEE 29th PIMRC*, 2018, pp. 1–7.

[21] Z. Shi, W. Gao, J. Liu, N. Kato, and Y. Zhang, "Distributed Q-learning-assisted grant-free NORA for massive machine-type communications," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–5.

[22] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "Transmit power pool design for grant-free NOMA-IoT networks via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7626–7641, Nov. 2021.

[23] S. A. Tegos, P. D. Diamantoulakis, A. S. Lioumpas, P. G. Sarigiannidis, and G. K. Karagiannidis, "Slotted ALOHA with NOMA for the next generation IoT," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6289–6301, Oct. 2020.

[24] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A novel analytical framework for massive grant-free NOMA," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, Mar. 2019.

[25] T. Hirai and P. Spasojevic, "Link-level performance evaluations of sparse code multiple access for PC5-based cellular-V2X with heterogeneous channel estimation errors," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC)*, 2021, pp. 1–5.

Takeshi Hirai (Member, IEEE) was born in Toyama, Japan, in 1994. He received the B.S. degree in engineering, the M.S. degree in information science, and the Ph.D. degree in informatics from Nagoya University, Nagoya, Japan, in 2016, 2018, and 2021, respectively.

He has been an Assistant Professor with Osaka University, Osaka, Japan, since 2021. His research interests include wireless communications, wireless LAN systems, cellular networks, V2X, UAV communications, nonorthogonal multiple access, IoT, QoS control, congestion control, broadcast communications, MAC layer protocol, driver assistance systems, and autonomous driving.

Dr. Hirai received the IEICE Young Researcher's Award in 2019 and the IEICE Tokai Section Student Award in 2017. He received the Research Fellowship for Young Scientists (DC2) from the Japan Society for the Promotion of Science, the Kurata Grant from the Hitachi Global Foundation, a Travel Grant from the Telecommunications Advancement Foundation, and publication grants from the Research Foundation for the Electrotechnology of Chubu in 2018 and 2020. He is a member of IEICE.

Rei Oda received the B.S. degree in engineering from Osaka University, Suita, Japan, in 2022, where he is currently pursuing the M.E. degree.

His research interests include wireless communications, wireless LAN systems, cellular networks, and nonorthogonal multiple access.

Naoki Wakamiya (Member, IEEE) received the M.E. and Ph.D. degrees from Osaka University, Suita, Japan, in 1994 and 1996, respectively.

He has been a Research Associate with the Graduate School of Engineering Science and the Educational Center for Information Processing, an Assistant Professor with the Graduate School of Engineering Science, and an Associate Professor with the Graduate School of Information Science and Technology, Osaka University, where he became a Professor in 2011. His research interests include biologically and brain inspired information and communication technology and self-organizing network control.