

Performance evaluation of a power allocation algorithm based on dynamic blocklength estimation for URLLC in the multicarrier downlink NOMA systems

Won Jae RYU¹, Soo Young SHIN*¹

Department of IT Convergence Engineering, Faculty of Engineering, Kumoh National Institute of Technology, Gumi, Republic of Korea

Received: 20.04.2020

Accepted/Published Online: 20.08.2020

Final Version: 27.01.2021

Abstract: This study investigates a power allocation algorithm using blocklength estimation by the finite blocklength (FBL) regime in a multicarrier downlink nonorthogonal multiple access (NOMA) system for ultrareliable low latency communication (URLLC) that is one of the services in 5G networks, requiring exceedingly high reliability and low latency. As NOMA systems can boost the capacity and increase the spectrum efficiency, it can be considered as a solution for URLLC. A multicarrier downlink NOMA system using blocklength estimation based on the FBL regime is proposed for effective resource allocation in this study. The FBL is used to derive the equation for dynamic blocklength estimation for an efficient power allocation algorithm in multicarrier downlink NOMA systems. The proposed power allocation algorithm is compared with other power allocation methods in terms of blocklength, missed deadlines, and energy efficiency; it performed the best compared with the other methods.

Key words: URLLC, finite blocklength regime, NOMA, PSO, power allocation

1. Introduction

Ultrareliable low latency communication (URLLC) is a service category in the 5G New Radio (NR) standard characterized by short data and stringent latency and reliability constraints, such as industrial communications, smart grids, robotic telesurgery, and autonomous driving, and is regarded as the most demanding service because having stringent latency and reliability simultaneously is difficult [1]. The target reliability of a general URLLC is at least 99.999% and the latency is at most 1 ms [2]. Although URLLC packets are shorts, they must be transmitted with exceedingly high reliability and low latency. For that reason, conventional approaches, such as the Shannon capacity assuming infinite blocklength [3], will be unsuitable for this case [4]. Therefore, the finite blocklength (FBL) capacity model, where the blocklength is comparable to the length of the actual data, has been proposed [5–7].

To satisfy the requirements of URLLC, studies have been conducted regarding the maximal channel coding rate achievable at a given blocklength and error probability in the FBL regime. Destounis et al. studied the delay performance of queuing systems and scheduling policies with URLLC in the FBL regime [8]. Pan et al. studied the joint optimization of blocklength allocation and unmanned aerial vehicle (UAV) location to minimize decoding error probability subject to latency [9]. Ren et al. introduced the resource allocation problem in a factory automation scenario in orthogonal multiple access (OMA) and relay-assisted transmission and aimed to

*Correspondence: wdragon@kumoh.ac.kr

jointly optimize blocklength and power allocation to minimize the error probability of the actuator [10]. Anand et al. studied the design of a wireless system supporting downlink URLLC traffic using a queuing network-based model for a wireless system with OMA and hybrid automatic repeat request (HARQ) [11]. Sun et al. aimed to obtain the global optimal resource allocation for URLLC in OMA [12]. However, the abovementioned studies focused on OMA systems that have lower spectrum efficiencies than nonorthogonal multiple access (NOMA), which allows all users to share the entire time and frequency resource [13, 14]. NOMA offers potential benefits while integrating with the short packets of URLLC. Kassab et al. studied OMA and NOMA for the multiplexing of enhanced mobile broadband (eMBB) and URLLC users in the uplink of a multicell cloud/fog radio access architecture [15, 16]. Wang et al. adopted outage probability as an important URLLC metric and deduced two closed form approximate bounds of system OP for finite and infinite battery capacities to characterize the relationship between reliability and latency [17]. Kotaba et al. introduced the concept of NOMA-HARQ to solve bandwidth limitations during retransmission [18]. Dosti et al. presented an FBL comparison between the OMA and NOMA schemes for an uplink channel in a Gaussian channel and a Rayleigh fading channel [19]. However, these studies were performed based on the FBL capacity with a specific and fixed blocklength for URLLC, not the dynamic blocklength.

In this study, the multicarrier downlink NOMA system was adopted to comply with URLLC requirements in terms of the dynamic blocklength derived from the FBL capacity equation. Furthermore, based on the dynamic blocklength equation, a power allocation algorithm was proposed for two users in a user pair.

The main contributions of this work can be summarized as follows:

1. An equation based on the FBL regime is proposed to estimate blocklength with the specific error rate of URLLC and channel state.
2. The power allocation algorithm in a multicarrier downlink NOMA system to reduce blocklength is suggested based on an equation to estimate the blocklength and particle swarm optimization (PSO) to satisfy URLLC requirements. Most of the works for power allocation in NOMA have focused on capacity maximization. Instead, the proposed power allocation focused on blocklength minimization, not capacity maximization.
3. The proposed power allocation algorithm is compared with other power allocation methods in terms of the blocklength, missed deadlines, and energy efficiency.

The rest of this paper is structured as follows. Section 2 describes the proposed system model. Section 3 describes the proposed power allocation algorithm in the multicarrier downlink NOMA. Section 4 provides a discussion of simulation results. Finally, conclusions and future studies are presented in Section 5.

2. System model

The multicarrier downlink NOMA system was adopted with the dynamic blocklength estimation derived from the FBL regime for the resource allocation of URLLC packets, which is illustrated in Figure 1. The system comprises one base station and several near and far users using NOMA. Two users comprising a near and a far user are paired by the user pair algorithm, referred to as the next largest difference-based user pairing (UP) algorithm [20, 21]. This means that the signals from the paired users are superposed by power differentiation. As shown in Figure 1, to satisfy the specific error rate and latency of URLLC, users in the worse channel have longer blocklengths than those in the better channel. Herein, the blocklength refers to the number of

subcarriers to be used for transmission. Depending on the power allocation algorithm and channel states, the power differentiation between near and far users varies as shown in Figure 1. As all the users in the system are regarded as URLLC users, all the transmissions in this system are with the specific error rate required by URLLC. To satisfy the error rate of URLLC, the users in the worse channel should transmit data with a longer blocklength than that in the better channel. In this system, as the multicarrier system is adopted, the blocklength is regarded as the number of subcarriers. The worse channel requires a longer blocklength, regarded as the number of subcarriers because it is not easy to achieve high reliability with a shorter blocklength. In other words, the better channel requires a shorter blocklength than the worse channel. The correlation between the channel state and blocklength is shown in Figure 2, which explains the FBL regime. In terms of the FBL regime, the channel state alters the data rate and blocklength to satisfy the specific error rate that can be the reliability requirement of URLLC. The equation of the FBL regime is as follows [19]:

$$\begin{aligned}
 R_1 &= C(x_1) - \sqrt{\frac{V(x_1)}{n}} Q^{-1}(\epsilon) \\
 R_2 &= C(x_2) - \sqrt{\frac{V(x_2)}{n}} Q^{-1}(\epsilon)
 \end{aligned}
 \tag{1}$$

R represents the data rate based on the FBL regime, R_1 is data rate for the near user, R_2 is data rate for the far user, C is the conventional capacity, $C(x) = \log_2(1+x)$, V is the channel dispersion, $V(x) = 1 - \frac{1}{(1+x)^2}$ in [22], x_1 is the signal-to-interference-plus-noise ratio (SINR) of the near user, x_2 is the SINR of the far user, n is the blocklength, Q is the Q-function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$, and ϵ is the error rate of URLLC. Eq. (1) expresses that the longer blocklength, the closer is the Shannon limit, as shown in Figure 2a. A longer blocklength implies more resources required. Therefore, the transmission will fail if the number of resource blocks is insufficient until the latency constraints of URLLC.

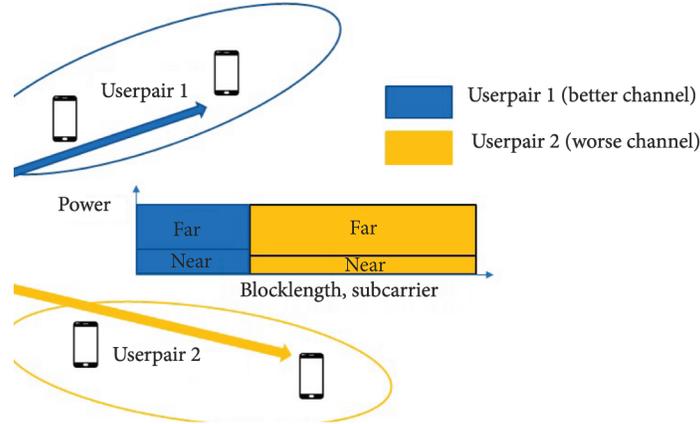


Figure 1. System description.

Instead, Figure 2b shows blocklength over SINR increasing. It means that the better channel has a shorter blocklength. Accordingly, based on the channel state, blocklength can be estimated [23].

As (1) means that the data rate is in one block, the data rate of the total blocks T can be described as

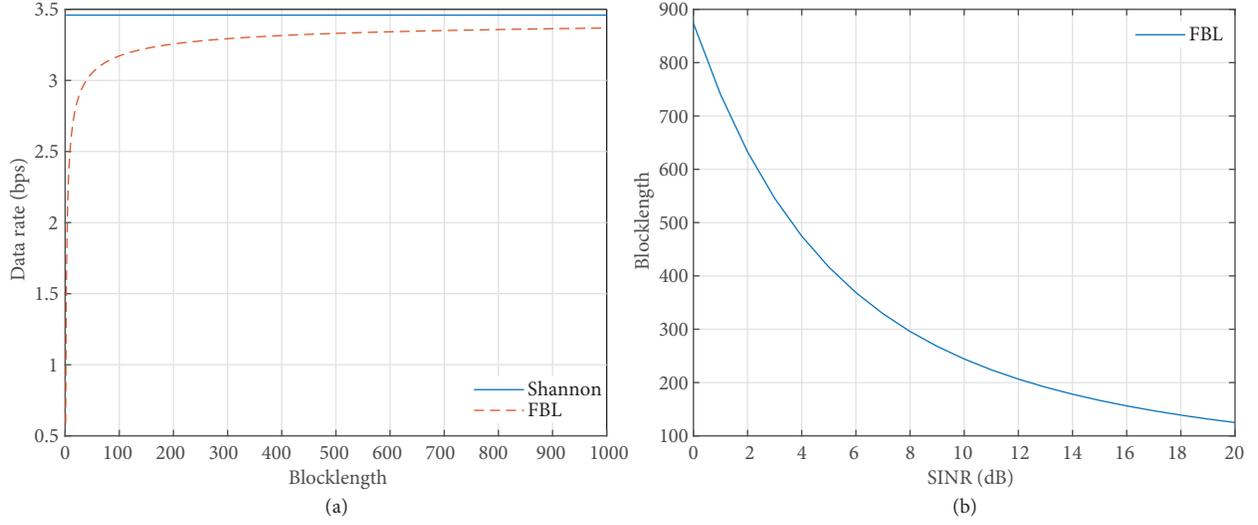


Figure 2. (a) Data rate along blocklength at SINR 10, and (b) blocklength with 100 bytes packet at error rate 0.00001 in a Gaussian channel.

follows:

$$T = nR = n(C(x) - \sqrt{\frac{V(x)}{n}}Q^{-1}(\epsilon)) \quad (2)$$

To estimate blocklength n ,

$$n = \left(\frac{\sqrt{4TC(x)Q(\epsilon)^2 + V(x)} + \sqrt{V(x)}}{2C(x)Q(\epsilon)} \right)^2 \quad (3)$$

Therefore, n will be used in the form of below

$$n(x, \epsilon, T) = \left(\frac{\sqrt{4TC(x)Q(\epsilon)^2 + V(x)} + \sqrt{V(x)}}{2C(x)Q(\epsilon)} \right)^2 \quad (4)$$

Derivation of (4) is described in Appendix A.

Based on (4), the blocklength of data can be calculated and applied to the evaluations. The longer blocklength is required, the more resource blocks are required. Therefore, the power allocation algorithm to minimize the number of resource blocks in the next section is described.

3. Power allocation algorithm based on dynamic blocklength estimation

In this section, the power allocation algorithm based on (4) and PSO in the multicarrier downlink NOMA system is proposed.

Using the sum of blocklength between the powers of the near user p_1 and far user p_2 , the point that enables an efficient transmission can be obtained.

$$\begin{aligned} 0 < \alpha < 1 \\ p_1 &= \alpha \\ p_2 &= 1 - \alpha \end{aligned} \quad (5)$$

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \left[n(x_1, \epsilon, T_1) + n(x_2, \epsilon, T_2) \right] = \underset{\alpha}{\operatorname{argmin}} \left[n\left(\frac{\alpha|h_1|^2}{N_{0,1}}, \epsilon, T_1\right) + n\left(\frac{(1-\alpha)|h_2|^2}{(\alpha|h_2| + N_{0,2})}, \epsilon, T_2\right) \right] \quad (6)$$

x_1 and x_2 , $\frac{p_1|h_1|^2}{N_{0,1}}$ and $\frac{p_2|h_2|^2}{(p_1|h_2|^2 + N_{0,2})}$, are the SINRs for the near and the far users, respectively; $N_{0,1}$ and $N_{0,2}$ are the noise variances for the near and far users, respectively; h_1 and h_2 are the channel states of the near and far users, respectively. In the case of the near user, only $N_{0,1}$ without any interference is considered because the signal for the far user is canceled by the successive interference cancellation (SIC). Instead, the far user experiences interference which is the signal for the near user, because the far user has a larger power ratio than of the near user. Therefore the far user can decode the signal without the SIC. As the far user does not do the SIC, $\alpha|h_2|$ is added as an interference to the far user. T_1 and T_2 are the packet sizes for the near and far users, respectively; α is the power allocation ratio for the near user.

Based on (4), the blocklengths for both the near and the far users are combined in (6). The point α that is used for minimizing the sum blocklength between the near and far users the point is considered as the efficient power allocation ratio in terms of blocklength. Therefore, as stated by (6), the power allocation α^* was obtained to minimize the required resource blocks.

To obtain the point α^* , i.e. the power allocation factor, PSO was adopted [24]. Algorithm 1 describes the mechanism of PSO for minimizing the required blocklength. The reason why PSO was chosen is that PSO is a metaheuristic optimization algorithm that is simple to implement, has short computational time, and able to run parallel computation [25]. It means that PSO can be suitable and easily used for finding a near optimal power allocation factor in real-time services rather than other optimization technics. Therefore to find a near optimal power allocation factor for user pairs in the NOMA system, PSO can be regarded as a good method.

The initial positions of particles were set in α . Each particle j was updated until the maximum iteration max . Velocity v was computed and position α , which is the power allocation factor, was updated. In particular, the *getFitness* function was used to evaluate each particle using (6) to obtain the blocklength. After the procedures in Algorithm 1 have been deployed, the minimum point α^* will be decided and used to transmit data.

4. Simulation

4.1. Simulation description

We compared five algorithms to evaluate the proposed algorithm: the power allocation scheme based on the blocklength estimation proposed herein; fixed power allocation; fractional transmit power control (FTPC) [13, 21, 26]; gradient descent algorithm (GDA) [27], which superposes all users in the entire bandwidth; and GDA for two users, which superposes two users in a user pair. In the case of the fixed power allocation, the power ratio for the near user is 0.2 and the power ratio for the far user is 0.8. For FTPC, the equation is below,

$$p_k = \frac{1}{\sum_{j \in S} \left(\frac{h_j}{N_{0,j}}\right)^{-\alpha_{ftpc}}} \left(\frac{h_k}{N_{0,k}}\right)^{-\alpha_{ftpc}} \quad (7)$$

p_k is the power ratio for the k_{th} user; h_k is the channel state for the k_{th} user; $N_{0,k}$ is the noise variance for the k_{th} user; S is the set of the users in a user-pair; $-\alpha_{ftpc}$ is the decay factor, $0 \leq \alpha_{ftpc} \leq 1$. When α_{ftpc} is 0, equal power allocation is achieved. α_{ftpc} increasing, the user with the worse α_{ftpc} channel state will get more power.

Algorithm 1: PSO-based power allocation.

Input: Channel of far user h_f , channel of near user h_n , noise N_0 , maximum iterations max
Output: Power for near user p_1 , power for far user p_2
Data: Positions of particles (power) for near user α , velocity v , acceleration coefficients c_1 and c_2 , random value r_1 and r_2 , global best α^* , the inertial coefficient w , individual best of each particles α^{ib}

```

1  $\alpha = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$ 
2  $N = length(\alpha)$ 
3  $\alpha.score = getFitness(\alpha, h_f, h_n, N_0)$  // Based on eq (6)
4  $\alpha^{ib} = \alpha$ 
5  $\alpha^* = findBestFitness(\alpha)$ 
6 while  $i \leq max$  do
7   while  $j \leq N$  do
8      $v(j) = w * v(j) + c_1 * r_1 * (\alpha^{ib} - \alpha(j)) + c_2 * r_2 * (\alpha^* - \alpha(j))$ 
9      $\alpha(j) = \alpha(j) + v(j)$ 
10     $\alpha(j).score = getFitness(\alpha(j), h_f, h_n, N_0)$ 
11    if  $\alpha(j).fitness \leq \alpha^{ib}(j).fitness$  then
12       $\alpha^{ib}(j) = \alpha(j)$ 
13    if  $\alpha(j).fitness \leq \alpha^*.fitness$  then
14       $\alpha^* = \alpha(j)$ 
15     $w = w * 0.99$ 
16  $p_1 = \alpha^*$ 
17  $p_2 = 1 - \alpha^*$ 

```

GDA is designed to focus on the maximization of the capacity of R_{sum} . Therefore, the maximization problem for power allocation using GDA is

$$P = (p_1, p_2, \dots, p_K)$$

$$R_{sum} = \sum_{n=1}^N \sum_{i=1}^K \log_2 \left(1 + \frac{p_i |h_i|^2}{\sum_{j=1}^{i-1} p_j |h_i|^2 + N_{0,i}} \right) \tag{8}$$

$$= \sum_{n=1}^N \sum_{i=1}^{K-1} \log_2 \left(\frac{\sum_{j=1}^i p_j + \frac{N_{0,i}}{|h_i|^2}}{\sum_{j=1}^i p_j + \frac{N_{0,i+1}}{|h_{i+1}|^2}} \right) + \sum_{n=1}^N \log_2 \left(\frac{\sum_{i=1}^K p_i + \frac{N_{0,K}}{|h_K|^2}}{\frac{N_{0,1}}{|h_1|^2}} \right).$$

Eq. (8) is a concave function of power allocation factors for each user superposed, P . N is the total number of subcarriers; K is the total number of users. To solve the problem of (8), the projected gradient descent method [28] is chosen.

The simulation parameters are shown in Table . The subcarrier spacing, transmission time interval (TTI), and TTI duration correspond to the minislot in the defined NR terminology [29]. And all the nodes are assumed as URLLC users. The URLLC packet for each user comes out in accordance with the packet arrival rate.

As the channel state can differ between a near and a far user, the blocklength between them can differ as well. This means that one user between them might have finished earlier during the TTI. In this case, the power allocated to the finished user will be assigned to the other user during the TTI. In other words, the blocklength

Table . Simulation parameters.

Normalized distance between the users and the basestation	0.0–1.0
Basestation maximum transmit power	2 W
Bandwidth	2.5 MHz
Subband bandwidth	15 kHz
The number of subcarriers	166
Transmission time interval (TTI)	0.143 ms
TTI duration	2 symbols
Channel model	Rayleigh fading
Path loss exponent	4
Target error rate	10^{-5}
Latency constraint	0.576 ms
URLLC packet size	32 bytes
Simulation time	1430 ms
The number of basestation	1
The number of nodes	20
Iterations of PSO	10

of a user pair can be expressed as follows:

$$n_{superposed} = \min\left(n\left(\frac{\alpha|h_1|^2}{N_{0,1}}, \epsilon, T_1\right), n\left(\frac{(1-\alpha)|h_2|^2}{(\alpha|h_2| + N_{0,2})}, \epsilon, T_2\right)\right) \quad (9)$$

In the case where the blocklength of the near user is $n_{superposed}$,

$$\begin{aligned} T_2 &= T_2 - (R_2 \times n_{superposed}) \\ n_{solo} &= n\left(\frac{|h_2|^2}{N_{0,2}}, \epsilon, T_2\right). \end{aligned} \quad (10)$$

Therefore, the final blocklength of the user pair $n_{userpair}$ is

$$n_{userpair} = n_{superposed} + n_{solo}. \quad (11)$$

Eq. (11) is described in Figure 3a, whereas the opposite case is described in Figure 3b. Using (11), the blocklength that is assigned to the subcarriers can be calculated. Therefore, the missed deadlines can be decided based on the blocklength assigned to the limited bandwidth. If the number of subcarriers is not enough to be assigned compared to the blocklength until the deadline, the data transmission with the blocklength will be counted as a missed deadline.

As the estimated blocklength is calculated using (4) with the error rate of URLLC, the reliability of URLLC is satisfied. However, the latency of URLLC is not guaranteed in (4). If the number of resource blocks is infinite, then the latency constraint is satisfied without exception. In other words, as the number of resource blocks is limited in practical situations, the shorter blocklength can be considered advantageous in terms of the latency constraint. Therefore, blocklength comparisons and missed deadlines among the power allocation

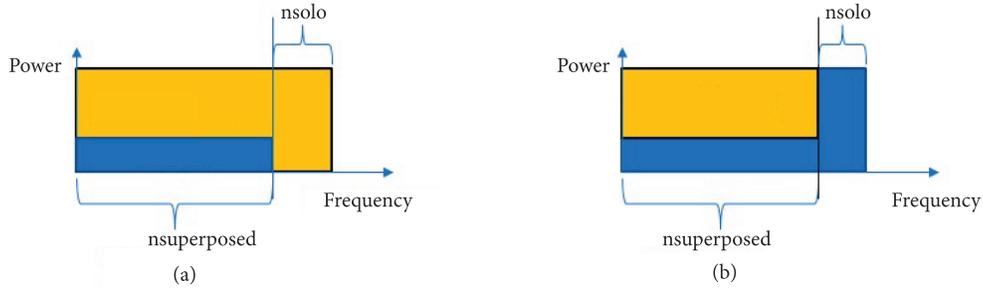


Figure 3. (a) Power allocation in the case where near user finished earlier during transmission, and (b) power allocation in the case where far user finished earlier during transmission.

algorithms were evaluated. One-half of 1 ms, which is generally considered as the latency constraint in URLLC, was used as the latency constraint in the simulation as the focus of the study was on the downlink system. To demonstrate additional advantages of the proposed algorithm, a simulation was performed to reflect the energy efficiency (EE).

$$EE = \frac{R_{tot} \times t}{P_{tot} \times t} \quad (12)$$

EE was calculated using (12). P_{tot} is the total transmitted power, and R_{tot} is the total data rate.

4.2. Simulation results

Figure 4 compares the blocklength of each user pair. The FBL is the proposed algorithm, whereas the fixed power allocation was set to 0.2 to the near user and 0.8 to the far user. The GDA was not included because it superposed all users in the entire bandwidth without a user pair. Instead, the GDA for two users, which is modified to consider only two users in a user pair, was compared. Its FBL tends to be shorter around all user pairs than those of the other power allocation methods.

Figure 5 compares the average sum of blocklength of the algorithms when 10 user-pairs transmit signals simultaneously. The FBL requires the shortest blocklength over all SINRs. This means that the proposed algorithm requires fewer resource blocks and more appropriate to satisfy URLLC latency requirements than the other methods.

Figure 6 compares the missed deadlines. It is noteworthy that 1 ms is generally considered as the latency of URLLC including uplink and downlink. As the focus of this study is on the downlink system, the latency constraint was set to 0.576 ms, which is approximately one-half of 1 ms. In all transmissions during the simulation, all cases where the algorithms missed the deadlines of transmissions were considered. By increasing the packet arrival rate, the number of cases with missed deadlines increased owing to the same reason contributing to the increase in the number of transmissions. The proposed algorithm indicated the least missed deadlines. This means that the proposed algorithm performed the best in terms of deadline compliance. By contrast, the GDA performed the worst because it was designed to focus on the maximization of the sum capacity. This means that the GDA may not be optimal in terms of the blocklength.

Figure 7 compares EE. The results indicate the EE consumption of the proposed algorithm. When the SINR was low, EE was not vastly different among the algorithms. However, over an increasing SINR, the gap in EE among them enlarged. The proposed algorithm demonstrated the best EE, whereas the GDA, the worst. Therefore, using the proposed algorithm, the energy required to transfer signals can be reduced. Compared

with other algorithms, the proposed algorithm is more applicable to cases such as those involving small IoT devices and military operations with limited battery life.

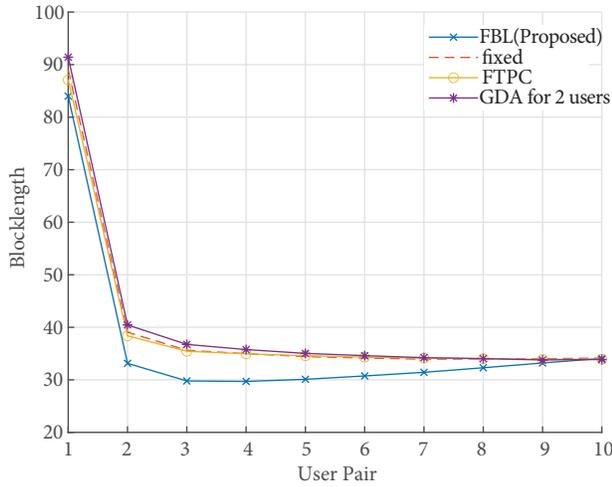


Figure 4. Energy efficiency.

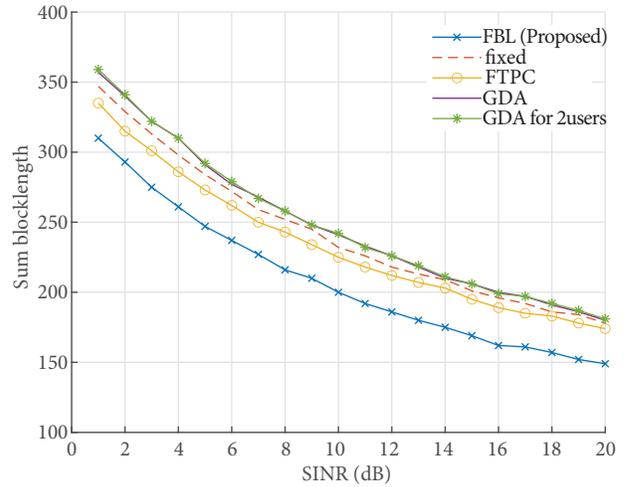


Figure 5. Average sum blocklength

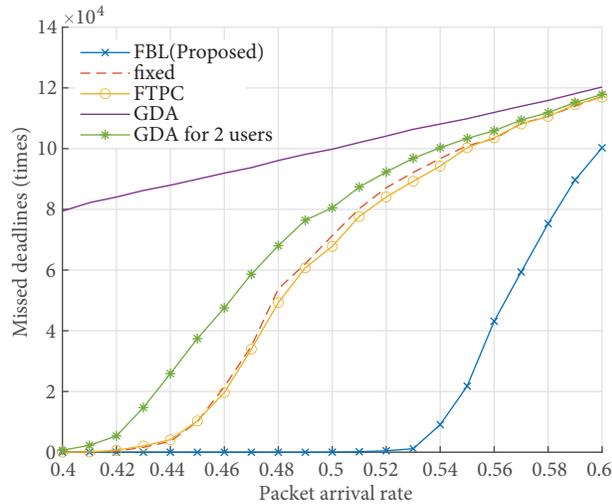


Figure 6. Missed deadlines

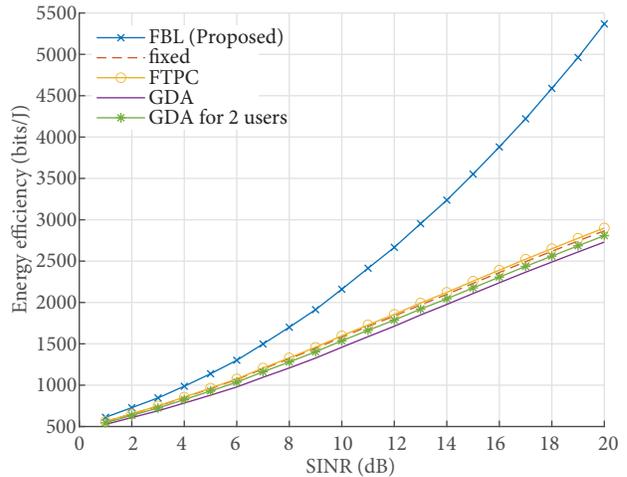


Figure 7. Energy efficiency

5. Conclusion

In this study, the multicarrier downlink NOMA system for URLLC was investigated. A resource allocation methodology was proposed in terms of blocklength to satisfy the requirements of URLLC. To estimate the blocklength prior to resource allocation, an equation at a fixed error rate based on the FBL regime was derived. Subsequently, based on the equation and PSO, the power allocation algorithm for two users in a NOMA system was proposed to satisfy the URLLC requirements. From the perspective of blocklength, missed deadlines, and energy efficiency, the proposed algorithm was compared with fixed power allocation, FTPC, and GDA in the multicarrier downlink NOMA system. Simulation results indicated that the proposed algorithm performed

better and is more appropriate for URLLCs than the other algorithms. Furthermore, in terms of energy efficiency, the proposed algorithm performed better than the others for systems with limited battery life, such as small IoT devices and military operations. Future studies regarding UP optimization and the superposition of more than two users will be conducted based on this study.

Acknowledgment

This research was supported by the Grand Information Technology Research Center Program through the Institute of Information and Communications Technology, and Planning and Evaluation (IITP) funded by the Ministry of Science and ICT (MSIT), Korea (IITP-2020-2020-0-01612).

References

- [1] Bennis M, Debbah M, Poor HV. Ultrareliable and low-latency wireless communication: tail, risk, and scale. *Proceedings of the IEEE* 2018; 106 (10): 1834-1853
- [2] Strinati EC, Mueck M, Clemente A, Kim JH, Noh GS et al. 5GCHAMPION disruptive 5G technologies for rollout. *ETRI Journal* 2018; 40 (1): 10-25.
- [3] Shannon CE. A mathematical theory of communication. *Bell System Technical Journal* 1948; 27 (3): 379-423.
- [4] Popovski P, Nielsen JJ, Stefanovic C, Carvalho Ed, Strom E et al. Wireless access for ultra-reliable low-latency communication: principles and building blocks. *IEEE Network* 2018; 32 (2): 16-23.
- [5] Polyanskiy Y, Poor HV, Verdu S. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory* 2010; 56 (5): 2307-2359.
- [6] Hu Y, Schmink A, Gross J. Blocklength-limited performance of relaying under quasi-static Rayleigh channels. *IEEE Transactions on Wireless Communications* 2016; 15 (7): 4548-4558.
- [7] Ozger M, Vondra M, Cavdar C. Towards beyond visual line of sight piloting of UAVs with ultra reliable low latency communication. In: *IEEE Global Communications Conference*; Abu Dhabi, United Arab Emirates; 2018. pp. 1-6.
- [8] Destounis A, Paschos GS, Arnau J, Kountouris M. Scheduling URLLC users with reliable latency guarantees. In: *16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*; Shanghai; 2018. pp. 1-8.
- [9] Pan C, Ren H, Deng Y, Elkashlan M, Nallanathan A. Joint blocklength and location optimization for URLLC-enabled UAV relay systems. *IEEE Communications Letters* 2019; 23 (3): 498-501.
- [10] Ren H, Pan C, Deng Y, Elkashlan M, Nallanathan A. Resource allocation for URLLC in 5G mission-critical IoT networks. In: *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*; Shanghai, China; 2019. pp. 1-6.
- [11] Anand A, De Veciana G. Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks. *IEEE Journal on Selected Areas in Communications* 2018; 36 (11): 2411-2421.
- [12] Sun C, She C, Yang C, Quek TQS, Li Y et al. Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications. *IEEE Transactions on Wireless Communications* 2019; 18 (1): 402-415.
- [13] Saito Y, Kishiyama Y, Benjebbour A, Nakamura T, Li A et al. Non-orthogonal multiple access (NOMA) for cellular future radio access. In: *IEEE 77th Vehicular Technology Conference (VTC Spring)*; Dresden, Germany; 2013. pp. 1-5.
- [14] Dai L, Wang B, Yuan Y, Han S, Chih-lin I et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. *IEEE Communications Magazine* 2015; 53 (9): 74-81.

- [15] Kassab R, Simeone O, Popovski P. Coexistence of URLLC and eMBB services in the C-RAN uplink: an information-theoretic study. In: IEEE Global Communications Conference; Abu Dhabi, United Arab Emirates; 2018. pp. 1-6.
- [16] Kassab R, Simeone O, Popovski P, Islam T. Non-orthogonal multiplexing of ultra-reliable and broadband services in fog-radio architectures. IEEE Access 2019; 7: 13035-13049.
- [17] Wang Z, Lv T, Lin Z, Zeng J, Mathiopoulos PT. Outage performance of URLLC NOMA systems with wireless power transfer. IEEE Wireless Communications Letters 2020; 9 (3): 380-384.
- [18] Kotaba R, Manchon CN, Pratas NMK, Balercia T, Popovski P. Improving spectral efficiency in URLLC via NOMA-based retransmissions. In: IEEE International Conference on Communications; Shanghai, China; 2019. pp. 1-7.
- [19] Dosti E, Shehab M, Alves H, Latva-aho M. On the performance of non-orthogonal multiple access in the finite blocklength regime. Ad Hoc Networks 2019; 84: 148-157.
- [20] Ding Z, Fan P, Poor HV. Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. IEEE Transactions on Vehicular Technology 2016; 65 (8): 6010-6023.
- [21] Islam SMR, Zeng M, Dobre OA, Kwak K. Resource allocation for downlink NOMA systems: key techniques and open issues. IEEE Wireless Communications 2018; 25 (2): 40-47.
- [22] Berardinelli G. Reliability analysis of uplink grant-free transmission over shared resources. IEEE Access 2018; 6: 23602-23611.
- [23] Ryu WJ, Shin SY. Power allocation for URLLC using finite blocklength regime in downlink NOMA systems. In: 2019 International Conference on Information and Communication Technology Convergence (ICTC); Jeju Island, Korea (South); 2019. pp. 770-773.
- [24] Eiben AE, Smith JE. Introduction to Evolutionary Computing. USA: Springer, 2003.
- [25] Abdmouleh Z, Gastli A, Ben-Brahim L, Haouari M, Al-Emadi NA. Review of optimization techniques applied for the integration of distributed generation from renewable energy sources. Renewable Energy 2017; 113: 266-280.
- [26] Otao N, Kishiyama Y, Higuchi K. Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation. In: International Symposium on Wireless Communication Systems (ISWCS); Paris, France; 2012. pp. 476-480.
- [27] Fu Y, Salaün L, Sung CW, Chen CS. Subcarrier and power allocation for the downlink of multicarrier NOMA systems. IEEE Transactions on Vehicular Technology 2018; 67 (12): 11833-11847.
- [28] Boyd S, Vandenberghe L. Convex Optimization. USA: Cambridge University Press, 2004.
- [29] Study on New Radio Access Technology Physical Layer Aspects, document 38.802 v14.2.0. 3rd Generation Partnership Project. Valbonne, France: 3GPP Mobile Competence Centre, 2017. Generation Partnership Project, 2017.

Appendix A
Derivation of the equation for blocklength n

Derivation of (4) is described as follows:

$$T = nR = n(C(x) - \sqrt{\frac{V(x)}{n}}Q^{-1}(\epsilon)) \quad (\text{D1})$$

$$T = nC(x) - \frac{\sqrt{nV(x)}}{Q(\epsilon)} \quad (\text{D2})$$

$$T = (\sqrt{nC(x)} - \frac{\sqrt{V(x)}}{2Q(x)\sqrt{C(x)}})^2 - \frac{V(x)}{4Q(\epsilon)^2C(x)} \quad (\text{D3})$$

$$(\sqrt{nC(x)} - \frac{\sqrt{V(x)}}{2Q(\epsilon)\sqrt{C(x)}})^2 = T + \frac{V(x)}{4Q(\epsilon)^2\sqrt{C(x)}} \quad (\text{D4})$$

$$\sqrt{nC(x)} = \sqrt{T + \frac{V(x)}{4Q(\epsilon)^2\sqrt{C(x)}}} + \frac{\sqrt{V(x)}}{2Q(\epsilon)\sqrt{C(x)}} \quad (\text{D5})$$

$$\sqrt{n} = \sqrt{\frac{4TQ(\epsilon)^2C(x) + V(x)}{4Q(\epsilon)^2C(x)^2}} + \frac{\sqrt{V(x)}}{2Q(\epsilon)C(x)} \quad (\text{D6})$$

$$n = (\frac{\sqrt{4TC(x)Q(\epsilon)^2 + V(x)} + \sqrt{V(x)}}{2C(x)Q(\epsilon)})^2 \quad (\text{D7})$$