Joint Block Length and Pilot Length Optimization for URLLC in the Finite Block Length Regime

Jie Cao[†], Xu Zhu[*], Yufei Jiang[†], Yujie Liu[*], and Fu-Chun Zheng[†]

† School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China
* Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, UK
Emails: 18b952056@stu.hit.edu.cn, xuzhu@liverpool.ac.uk, jiangyufei@hit.edu.cn

Abstract—In this paper, we maximize the system throughput of a point-to-point ultra-reliable low-latency communications (URLLC) system by jointly optimizing its block length and pilot length under the constraints of latency and block error probability. A finite block length (FBL) is adopted to enable low transmission latency. We prove that the throughput is approximately concave with respect to pilot length, given a block length, and that there exists a unique optimal block length in terms of throughput, with a given pilot length. Closed-form expressions are derived for the near-optimal pilot length with a given block length, as well as the asymptotic block error probability with respect to both block length and pilot length. A low-complexity iterative algorithm is proposed for joint optimization of block length and pilot length, which converges within only 1-3 iterations. Simulation results show that the proposed joint optimization scheme achieves a near-optimal throughput performance of an FBL URLLC system, with a much lower complexity than exhaustive search. It also significantly outperforms the previous approaches that considered either block length optimization or pilot length optimization only.

Index Terms—Ultra-reliable low-latency communications, finite block length, pilot length, joint optimization

I. INTRODUCTION

One of the features of the fifth generation (5G) mobile networks is the support for Ultra-reliable low-latency communications (URLLC) [1], [2]. URLLC can be used for mission critical services such as autonomous driving, factory automation and remote piloting of drones [3]–[6], requiring end-to-end latency of less than 1 ms and very high reliability of 99.999%. How to achieve a high system throughput under the requirements of URLLC is a particularly challenging task and the focus of this paper.

Many techniques have been proposed to improve reliability, but often have to sacrifice latency. On the other hand, methods that reduce latency could cause decreasing reliability. Adopting a short block structure is a straightforward way to reduce the transmission latency. However, in the finite block length (FBL) regime, the decoding error probability should be taken into account and conventional metrics based on Shannon's theorem become inaccurate, *i.e.*, channel capacity and outage probability. In [7], the authors derived an accurate approximation of the achievable rate with respect to block length, error probability and signal-to-noise ratio (SNR) over additive white Gaussian noise (AWGN) channel. Subsequently, the results have been applied in relaying networks [8], non-orthogonal multiple access (NOMA) systems [9], [10] and

quality of service (QoS)-constrained networks [11], [12]. Power and bandwidth optimization have been considered in the above references to maximize the throughput or achievable rate. However, most of the previous studies assumed fixed block length, which may not be optimal. The work in [13] discussed the trade-off between the block length and the number of retransmissions. However, the specific block length optimization for one shot transmission was not analyzed in [13].

In addition, channel estimation accuracy has a great impact on the latency and reliability of URLLC [2], [3]. Furthermore, in the FBL regime, the pilot overhead for channel estimation is not negligible [14]. In the aforementioned work [10], [11], [13], however, no pilot length design was considered and perfect channel state information (CSI) was assumed. In [14], only the pilot overhead optimization was addressed without considering the block length. In addition, the pilot-based channel estimation also introduces estimation errors, which are not negligible. Therefore, it is essential to consider both pilot length and block length when designing a short-block structure for URLLC.

In this paper, a joint block length and pilot length optimization scheme is proposed to maximize the throughput of a point-to-point system under the constraints of high reliability and low latency. The main contributions of this paper are summerized as follows:

- To the best of our knowledge, this is the first work to investigate joint optimization of block length and pilot length of an FBL URLLC system. While the previous work has considered either block length [13] or pilot length [14] in block structure design. A closed-form expression of the asymptotic block error probability with respect to both block length and pilot length in the FBL regime is derived, while in the previous work only the block error probability with respect to block length was investigated [11], [13].

- We consider the impact of both block length and pilot length on system throughput of an FBL URLLC system. It is proven that the throughput is approximately concave with respect to pilot length, given a block length, and that there exists a unique optimal block length in terms of throughput, with a given pilot length. We also derive a closed-form expression of the near-optimal pilot length with a given block length. In addition, we show that the

impact of pilot length on throughput is more significant in low SNR region, while block length plays a dominant role in medium to high SNR region. The proposed joint optimization scheme achieves much higher throughput than the approaches in [13] and [14], where only block length [13] or pilot length [14] was optimized.

- A low-complexity joint block length and pilot length optimization (JBLPLO) algorithm is proposed, which operates in an iterative manner. Thanks to the closed-form expression of the near-optimal pilot length derived, only a search for the optimal block length is needed, and the algorithm converges within 1-3 iterations. Numerical results show that the proposed scheme achieves near-optimal performance in terms of throughput, with a tremendous complexity reduction over exhaustive search.

The rest of this paper is organized as follows. Section II presents the system model and problem formulation. In Section III, we derive the expression of the asymptotic block error probability in terms of both block length and pilot length. In Section IV, the impact of block length and pilot length on system throughput is analyzed and the corresponding JBLPLO algorithm is presented. Numerical results are given in Section V. Section VI draws conclusion.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We consider a point-to-point URLLC system with FBL. A pragmatic block structure is illustrated in Fig. 1, where a block of N symbols of M-ary modulation consists of $L/\log_2 M$ information symbols with L denoting the number of information bits, n_t pilot symbols for channel estimation, and redundant symbols for error correction code (ECC). In this paper, the specific ECC is not studied.
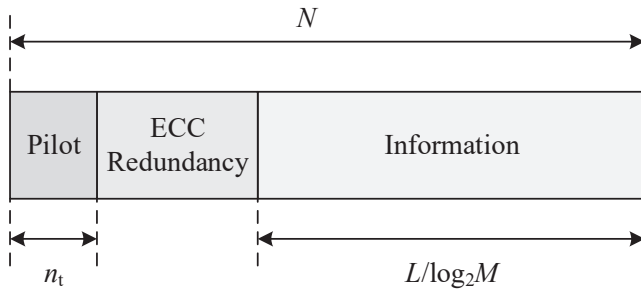

Fig. 1. Block Structure

A quasi-static Rayleigh fading channel is considered in this system, where the channel gain remains constant for the duration of one block. Denote x_i as the i-th ($i = 1, \ldots, N$) transmitted symbol of average symbol energy E_s, the i-th received signal is given by

$$y_i = hx_i + w_i, \tag{1}$$

where w_i is AWGN with zero-mean and variance N_0, and h is the Rayleigh fading channel gain within a block. The received instantaneous SNR is denoted by

$$\gamma = \frac{|h|^2 E_s}{N_0}. \tag{2}$$

B. Problem Formulation

In this subsection, we formulate an optimization problem to maximize the system throughput with respect to the block length and the pilot length under the constraints of both the latency and the reliability. Based on [9] and [10], the effective throughput is adopted as the system performance metric. The throughput (bits per channel use or bpcu for short) is obtained by

$$\eta = \frac{L}{\tau} (1 - \epsilon), \tag{3}$$

where ϵ is the decoding block error probability, and τ is the total number of channel uses in the block transmission, which is utilized to indicate latency.

In this paper, we focus on the transmission latency. As a consequence, we have $\tau = N$. The reliability is characterized by block error probability in terms of block length N and pilot length n_t, which is derived in Section III. Then the block length and pilot length are determined by solving the following optimization problem

$$\max_{N, n_t} \quad \eta \tag{4a}$$

$$\text{s.t.} \quad N - N_{\text{thr}} \leq 0, \tag{4b}$$

$$\epsilon - \epsilon_{\text{thr}} \leq 0, \tag{4c}$$

$$N - n_t \geq L/\log_2 M, \tag{4d}$$

$$N, n_t \in \mathbb{N}, \tag{4e}$$

where (4b) and (4c) limit the maximum allowable latency N_{thr} and maximum error probability ϵ_{thr}, respectively. Date must be transmitted with the reliability $1 - \epsilon_{\text{thr}}$ within the latency constraint N_{thr}. (4d) shows the ranges of N and n_t that mainly depend on the amount of information that needs to be transmitted.

III. BLOCK ERROR PROBABILITY ANALYSIS

A. Impact of Block Length on Block Error Probability

As disscussed in [7], for given block length N, block error probability ϵ and instantaneous SNR γ with perfect CSI, the achievable rate of the Rayleigh fading channel can be tightly approximated as

$$R = \frac{L}{N} \approx C(\gamma) - \sqrt{\frac{V(\gamma)}{N}} Q^{-1}(\epsilon), \tag{5}$$

where $C(\gamma) = \log_2(1 + \gamma)$ and $V(\gamma) = \left(1 - \frac{1}{(1+\gamma)^2}\right) (\log_2 e)^2$. In addition, $Q^{-1}(\cdot)$ is the inverse of the Q-function given by $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.

Method	Analytical Computational Complexity
Proposed JBLPLO algorithm	$O\left(I \log_2 N_{\mathrm{A}}\right)$
Optimal block length + fixed pilot length [13]	$O\left(\log_2 N_{\mathrm{A}}\right)$
Optimal pilot length + fixed block length [14]	$O\left(\log_2 N_{\mathrm{B}}\right)$
Exhaustive search	$O\left(\frac{N_{\mathrm{A}}^2}{2}\right)$

Subsequently, for a direct transmission with block length N, achievable rate R and received SNR γ, the block error probability is given by [11]

$$\epsilon \approx Q\left(\sqrt{\frac{N}{V(\gamma)}}\left(C(\gamma) - \frac{L}{N}\right)\right), \qquad (6)$$

Since Q-function is monotonically decreasing, the block error probability decreases as the block length increases in the FBL regime.

B. Impact of Pilot Length on Block Error Probability

In this section, we study the impact of pilot length on the block error probability. First, the length of pilot affects the number of data symbols and hence the block error probability can be expressed as

$$\epsilon \approx Q\left(\sqrt{\frac{N - n_{\mathrm{t}}}{V(\gamma)}}\left(C(\gamma) - \frac{L}{N - n_{\mathrm{t}}}\right)\right), \qquad (7)$$

which shows that an increase in n_{t} leads to an increase in the block error probability.

Then we evaluate the impact of n_{t} on channel estimation. n_{t} pilot symbols are employed to estimate channel gain that is assumed to remain constant within a block in the quasi-static channel. In this paper, minimum mean square error (MMSE) estimator is used in block fading model and we can obtain MMSE as $\delta^2 = \frac{1}{1 + n_{\mathrm{t}}\gamma}$ [14]. The effective SNR can be formulated as

$$\gamma_{n_{\mathrm{t}}} = \frac{\gamma(1 - \delta^2)}{1 + \gamma\delta^2} = \frac{n_{\mathrm{t}}\gamma^2}{n_{\mathrm{t}}\gamma + \gamma + 1}, \qquad (8)$$

which shows that the effective SNR $\gamma_{n_{\mathrm{t}}}$ approaches the SNR γ as n_{t} increases.

Hence, based on the above derivations, the asymptotic block error probability in terms of both block length and pilot length can be obtained as

$$\epsilon \approx Q\left(\sqrt{\frac{N - n_{\mathrm{t}}}{V(\gamma_{n_{\mathrm{t}}})}}\left(C(\gamma_{n_{\mathrm{t}}}) - \frac{L}{N - n_{\mathrm{t}}}\right)\right). \qquad (9)$$

IV. BLOCK STRUCTURE OPTIMIZATION

In this section, we propose a low-complexity JBLPLO algorithm to solve Problem (4) based on decomposition methods

[15]. The key idea of our algorithm is to decompose the original problem into two solvable subproblems and conduct an iterative search. In the proposed algorithm, the optimal values of the block length N and the pilot length n_{t} are determined under the fixed latency constraint τ_{thr} and reliability constraint ϵ_{thr}.

A. Optimal Block Length With Given Pilot Length

Proposition 1. With a given pilot length n_{t} and a block error probability of $\epsilon < 0.5$, there exists a unique optimal N_{opt} that maximizes the FBL system throughput η under the constraints of (4b), (4c) and (4d).

Proof: See Appendix A.

According to Proposition 1, optimal block length N_{opt} with a given pilot length can be obtained by using bisection search.

B. Optimal Pilot Length With Given Block Length

Proposition 2. *With a given block length N, the FBL system throughput is approximately concave with respect to pilot length n_{t}.*

Proof: See Appendix B.

Based on Proposition 2, the optimal pilot length n_{topt} can be derived by letting (20) be zero, which is equivalent to solving the first derivative of block error probability (17) with respect to n_{t}. The closed-form expression of the near-optimal pilot length is given by

$$n_{\mathrm{topt}} \approx \frac{\sqrt{\ln(2)}B_1 + B_2}{(2\ln(2)L - 2)\gamma} \qquad (10)$$

where

$B_1 = \sqrt{(4LN^2 + 4LN + \ln(2)L^2)\gamma^2 + (8LN + 4L)\gamma + 4L}$.

$B_2 = (-2N - \ln(2)L)\gamma - 2\ln(2)L$.

C. Joint Optimization of Block Length and Pilot Length

The overall algorithm for solving Problem (4) is provided in Algorithm 1. In particular, we start with random initialized value of n_{t}^0 and calculate the feasible range $[N_{\mathrm{min}}, N_{\mathrm{thr}}]$ according to (4b), (4c) and (4d). Then the optimal solution N^1 can be obtained by using bisection search in $[N_{\mathrm{min}}, N_{\mathrm{thr}}]$. Therefore, the optimal solution n_{t}^1 can be determined by (10) based on N^1. Then, we repeat these two steps till the solution converges. Thanks to the closed-form expression of the near-optimal pilot length given in (10), only a search for the optimal block length is needed, which implies low complexity.

Algorithm 1: The Iterative JBLPLO Algorithm with Block Length N and Pilot Length n_t

1 initialize: $n_t{}^i$, N^i, iterative index $i = 1$, maximum iterative times i_{max}, error tolerance θ;

2 repeat

3 For a given $n_t{}^{i-1}$, solve (13) to obtain the optimal N^i by using bisection search in feasible region $[N_{min}, N_{max}]$;

4 With a fixed N^i, the solution $n_t{}^i$ is obtained by (10).

5 $i = i + 1$;

6 until $i \geq i_{max}$ or $n_t{}^i - n_t{}^{i-1} < \theta$ & $N^i - N^{i-1} < \theta$

Fig. 2. Impact of SNR on throughput

D. Complexity Analysis

Then we analyze the complexity of the proposed algorithm. We denote $N_A = N_{thr} - N_{min}$ and $N_B = N - L/\log_2 M$. Generally, the complexity of exhaustive search is $O\left(\frac{N_A{}^2}{2}\right)$. The complexity of separately optimizing the block length and the pilot length are $O(\log_2 N_A)$ and $O(\log_2 N_B)$, respectively. The complexity of the proposed joint optimization algorithm is $O(I \log_2 N_A)$, where I is the number of iterations. The complexity of the above methods are shown in Table I. With $N_A = 100$, $N_B = 50$ and $I = 3$, The proposed algorithm is found to be very computationally efficient, with complexity reduction of around 200-fold than the exhaustive search. It also achieves a complexity comparable to the approaches that optimize either block length [13] or pilot length [14] only.

V. NUMERICAL RESULTS

Simulations have been carried out to demonstrate the effectiveness of the proposed JBLPLO algorithm for URLLC communication systems. The system parameters are specified as follows: the constraints for the maximum latency and the block error probability are $N_{thr} = 150$ and $\epsilon_{thr} = 1e - 5$, respectively; the transmitted information bit length L is set to 100; the modulation order is $M = 4$ and the number of iterations of JBLPLO is set to $I = 3$, except for Fig. 5.

A. Throughput

In Fig. 2, we compare the proposed joint optimization algorithm (labeled as 'Proposed JBLPLO algorithm') with the following methods: 1) exhaustive search algorithm (labeled as 'Exhaustive search'), 2) optimal block length with fixed pilot length $n_t = 1$ [13] (labeled as 'Optimal block length + fixed pilot length'), 3) fixed mean block length $N = \frac{N_\epsilon + N_{max}}{2}$ that satisfies the demand of URLLC with the optimal pilot length [14] (labeled as 'Fixed block length + optimal pilot length'). The performace of different methods are shown in Fig. 2.

As shown in Fig. 2, the throughput of all methods increases as SNR increases from 2 dB to 15 dB and then achieve a convergence gradually. It is shown that the results of our proposed JBLPLO algorithm are consistent with the exhaustive search and significantly outperforms the other methods, which demonstrates the effectiveness of joint optimization. At SNR=12 dB, the throughpu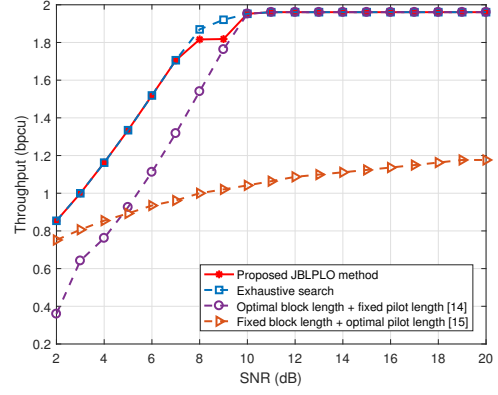t of the approaches that optimize only either block length or pilot length are enhanced approximately 0.6 npcu by the proposed iterative JBLPLO algorithm. In addition, the method with optimizing the pilot length is better than the method whithout optimization in the low SNR region, and optimizing the block length is more effective in the high SNR region.

B. Optimal Block Structure

Fig. 3 shows the impact of block length N and pilot length n_t on throughput at different SNRs. In general, the system performance is remarkably affected by the block length, which indicates that we can optimize block length N to maximize the throughput. It is shown that the impact of pilot length on throughput is significant in the low SNR region, which indicates that optimizing the pilot length in low SNR is more effective. As shown in Fig. 3(a) and 3(c), for the short block length region, the throughput increases rapidly as the block length increases. However, as the block size increases, the throughput decreases for the long block length region. This is due to the fact that with the increase of block length, the channel use becomes the major limit of the throughput instead of block error probability. Similarly, the throughput increases with an increase of the pilot length first and then starts to decrease slowly. Compared with Fig. 3(a), Fig. 3(c) shows that the optimal block length decreases as the SNR increases. By comparing Fig. 3(b) and Fig. 3(d), we observe that when the length of transmitted bits is fixed, the optimal pilot length and the optimal block length both decrease as the SNR increases.

Fig. 4 shows the relationship between the optimal block structure and SNR. It is shown that the optimal block length (excluding pilot length) and the optimal pilot length generally decreases as SNR increases.The optimal block struture remains almost constant at high SNR. In addition, the optimal pilot overhead that refers to the ratio of the optimal pilot length to the optimal block length is low in the high SNR region. Furthermore, it is shown that optimal pilot length and optimal block length obtained by the proposed JBLPLO algorithm are nearly the same to those by exhaustive search.
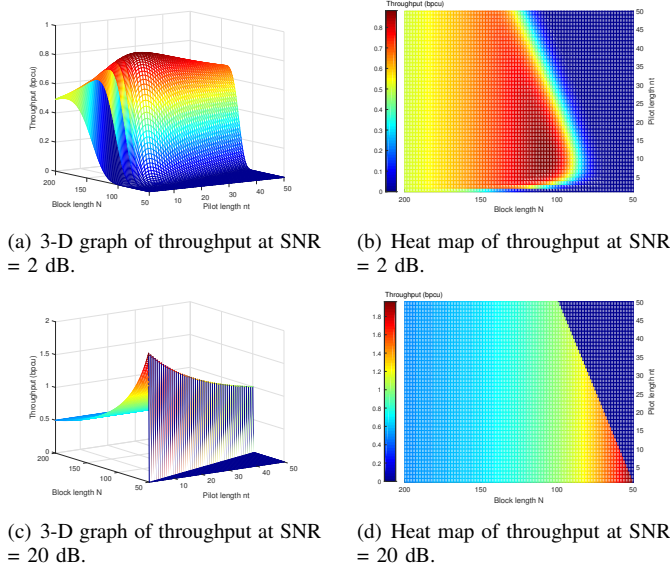
(a) 3-D graph of throughput at SNR = 2 dB.

(b) Heat map of throughput at SNR = 2 dB.

(c) 3-D graph of throughput at SNR = 20 dB.

(d) Heat map of throughput at SNR = 20 dB.

Fig. 3. Joint impact of block length and pilot length on the throughput for different SNRs
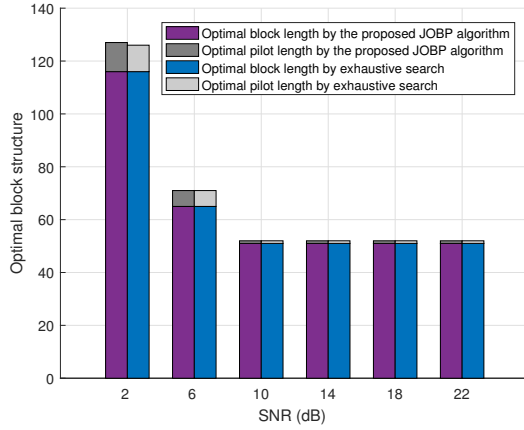

Fig. 4. Optimal Block structure for different SNRs

C. Convergence Behavior

Fig. 5 demonstrates the system throughput against the number of iterations of the proposed JBLPLO algorithm, from SNR = 2 dB to SNR = 18 dB with a step size of 4 dB. It is easily seen that the proposed iterative JBLPLO algorithm converges fast within 3 iterations, regardless of SNR.

VI. CONCLUSION

In this paper, we have jointly optimized pilot length and block length to maximize the throughput in a point-to-point FBL system under the constraints of high reliability and low latency, and derived the closed-form expression of the near-optimal pilot length with a given block length. A low-complexity iterative JBLPLO algorithm has been proposed, which requires only 1-3 iterations for convergence. Numerical results confirm that the proposed joint optimization algorithm achieve the near-optimal performance with a 200-fold
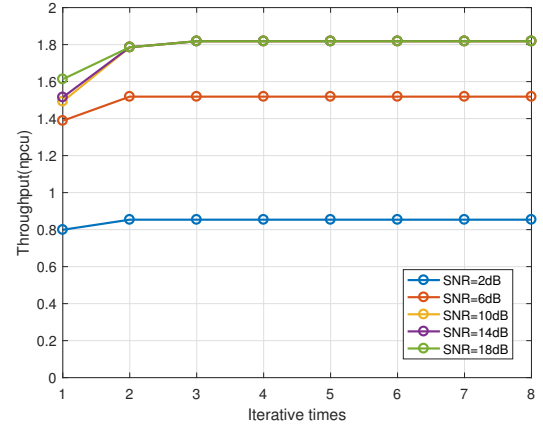

Fig. 5. Convergence performance of the proposed algorithm for various SNR

complexity reduction over exhaustive search. The throughput of the approaches that optimize only either block length or pilot length are enhanced approximately 0.6-0.8 npcu by the proposed JBLPLO algorithm. In addition, the optimal pilot overhead decreases with the increase of SNR and saturates at only 2% at high SNR.

APPENDIX A
PROOF OF PROPOSITION 1

For ease of analysis, we denote $\tilde{\eta} = -\eta$, and thus the maximization problem in (4) is transformed into a minimization problem. Then the objective fuction is given by

$$\min_{N,n_t} \quad \tilde{\eta} \tag{11}$$

First, according to [13], block length N can be approximated as a function of the block error probability ϵ as follows:

$$N_\epsilon \approx \frac{L}{C(\gamma_{n_t})} + W_1\left(1 + W_2\right), \tag{12}$$

where

$$W_1 = \frac{\left(Q^{-1}(\epsilon)\right)^2 V(\gamma_{n_t})}{2\left(C(\gamma_{n_t})\right)^2}.$$

$$W_2 = \sqrt{1 + \frac{4LC(\gamma_{n_t})}{V(\gamma_{n_t})\left(Q^{-1}(\epsilon)\right)^2}}.$$

Since ϵ is monotonically decreasing with respect to N, in order to satisfy $\epsilon \leq \epsilon_{thr}$, the block length needs to meet $N \geq N_\epsilon$. Therefore, the feasible range of N is $[N_{min}, N_{max}]$, where $N_{min} = \max\{N_\epsilon, L/\log_2 M\}$, $N_{max} = N_{thr}$.

Then we prove that there exists only one optimal N_{opt} that minimizes $\tilde{\eta}$. To simplify the notation, we denote $W_{N,n_t} = \sqrt{(N - n_t)/V(\gamma_{n_t})} * \left(C(\gamma_{n_t}) - \frac{L}{N - n_t}\right)$. The first order of $\tilde{\eta}$ with respect to N is given by

$$\frac{\partial \tilde{\eta}}{\partial N} = \frac{L}{N^2}\left(\frac{\partial Q\left(W_{N,n_t}\right)}{\partial N}N + 1 - Q\left(W_{N,n_t}\right)\right). \tag{13}$$

Let $R = \frac{\partial Q\left(W_{N,n_t}\right)}{\partial N}N + 1 - Q\left(W_{N,n_t}\right)$. $\frac{\partial \tilde{\eta}}{\partial N} > 0$ holds if $R > 0$, vice versa. The first derivative of R is given by

$$\frac{\partial R}{\partial N} = \left(\frac{\partial^2 Q\left(W_{N,n_t}\right)}{\partial N^2}N\right), \tag{14}$$

where

$$\frac{\partial^2 Q\left(W_{N,n_t}\right)}{\partial N^2} = G\left(W_{N,n_t}\left(\frac{\partial W_{N,n_t}}{\partial N}\right)^2 - \frac{\partial^2 W_{N,n_t}}{\partial N^2}\right), \tag{15}$$

where $G = \frac{1}{\sqrt{2\pi}}e^{-\frac{W_{N,n_t}^2}{2}}$ and

$$\frac{\partial^2 W_{N,n_t}}{\partial N^2} = -\frac{(N - n_t)C(\gamma_{n_t}) + 3L}{4\left(N - n_t\right)V\left(\gamma_{n_t}\right)^2\left(\frac{N-n_t}{V\left(\gamma_{n_t}\right)}\right)^{\frac{3}{2}}} < 0. \tag{16}$$

Note that $\frac{\partial R}{\partial N} > 0$ holds if $W_{N,n_t} > 0$, and $W_{N,n_t} > 0$ holds if $\epsilon < 0.5$. $\frac{\partial R}{\partial N} > 0$ suggests that R is monotonically increasing with respect to N. As N increases, there are three cases. 1) $R < 0$, the objective function $\tilde{\eta}$ is monotonically decreasing with respect to N in the feasible range $[N_{\min}, N_{\text{thr}}]$. The optimal block length is $N_{\text{opt}} = N_{\max}$. 2) R gradually changes from $R < 0$ to $R > 0$. There exists a unique optimal N_{opt} in $(N_{\min}, N_{\text{thr}})$. Then we apply bisection search to obtain the optimal N_{opt} by solving (13). 3) $R > 0$, the objective function $\tilde{\eta}$ is monotonically increasing with respect to N in $[N_{\min}, N_{\max}]$. The optimal block length is $N_{\text{opt}} = N_{\min}$.

APPENDIX B
PROOF OF PROPOSITION 2

Given N, minimizing $\tilde{\eta}$ is equivalent to minimize ϵ. The first and second order of ϵ with respect to n_t are given by

$$\frac{\partial \epsilon}{\partial n_t} = -G\frac{\partial W_{N,n_t}}{\partial n_t}. \tag{17}$$

$$\frac{\partial^2 \epsilon}{\partial n_t^2} = G\left(W_{N,n_t}\left(\frac{\partial W_{N,n_t}}{\partial n_t}\right)^2 - \frac{\partial^2 W_{N,n_t}}{\partial n_t^2}\right). \tag{18}$$

It is clear that $\frac{\partial^2 \epsilon}{\partial n_t^2} \geq 0$ holds if $\frac{\partial^2 W_{N,n_t}}{\partial n_t^2} \leq 0$. In the following, we mainly analyze $\frac{\partial^2 W_{N,n_t}}{\partial n_t^2}$ in detail. First, $V\left(\gamma_{n_t}\right)$ can be approximated as $V\left(\gamma_{n_t}\right) \approx \frac{1}{2}\left(\log_2 e\right)^2$ when the received SNR is higher than 5 dB [12]. Since n_t is much smaller than N, we simplify W_{N,n_t} as

$$Ws_{N,n_t} = \ln 2\sqrt{2N}\left(C(\gamma_{n_t}) - \frac{L}{N - n_t}\right). \tag{19}$$

Similarly, the convexity of ϵ can be proved by proving $\frac{\partial^2 Ws_{N,n_t}}{\partial n_t^2} \leq 0$. The first derivative of Ws_{N,n_t} is given by

$$\frac{\partial Ws_{N,n_t}}{\partial n_t} = \ln 2\sqrt{2N}\left(A_1 - \frac{L}{(N - n_t)^2}\right), \tag{20}$$

where

$$A_1 = \frac{\gamma^2/\ln(2)}{2\left(n_t\gamma + 1\right)^2\left(\frac{\gamma}{n_t\gamma + 1} + 1\right)}. \tag{21}$$

The second order of Ws_{N,n_t} is given by

$$\frac{\partial^2 Ws_{N,n_t}}{\partial n_t^2} = \ln(2)\sqrt{2N}\left(A_2 + A_3 + A_4\right), \tag{22}$$

where

$$A_2 = -\frac{\gamma^3}{\ln(2)\left(\gamma n_t + 1\right)^3\left(\frac{\gamma}{\gamma n_t + 1} + 1\right)},$$

$$A_3 = \frac{\gamma^4}{2\ln(2)\left(\gamma n_t + 1\right)^4\left(\frac{\gamma}{\gamma n_t + 1} + 1\right)^2},$$

$$A_4 = -\frac{2L}{(N - n_t)^3},$$

Then we have $\sum_{i=2}^4 A_i < 0$. Therefore $\frac{\partial^2 Ws_{N,n_t}}{\partial n_t^2} < 0$ and ϵ is convex with respect to n_t. It also demonstrate that the throughput η is approximately concave with respect to n_t.

REFERENCES

[1] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, pp. 16–23, Mar. 2018.

[2] Y. Liu, X. Zhu, E. G. Lim, Y. Jiang, and Y. Huang, "Fast iterative semi-blind receiver for URLLC in short-frame full-duplex systems with CFO," *IEEE J. Sel. Areas Commun.*, vol. 37, pp. 839–853, Apr. 2019.

[3] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, pp. 124–130, Jun. 2018.

[4] M. Ozger, M. Vondra, and C. Cavdar, "Towards beyond visual line of sight piloting of uavs with ultra reliable low latency communication," in *Proc. IEEE Globecom 2018*, Abu Dhabi, UAE, pp. 1–6, Dec. 2018.

[5] M. Gapeyenko, V. Petrov, D. Moltchanov, S. Andreev, N. Himayat, and Y. Koucheryavy, "Flexible and reliable uav-assisted backhaul operation in 5G mmwave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 2486–2496, Nov. 2018.

[6] L. Wang, Y. L. Che, J. Long, L. Duan, and K. Wu, "Multiple access mmwave design for uav-aided 5G communications," *IEEE Wireless Commun.*, vol. 26, pp. 64–71, Feb. 2019.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.

[8] Y. Hu, A. Schmeink, and J. Gross, "Blocklength-limited performance of relaying under quasi-static rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 4548–4558, Jul. 2016.

[9] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Downlink NOMA transmission for low-latency short-packet communications," in *Proc. IEEE ICC Workshops 2018*, Kansas, USA, pp. 1–6, May 2018.

[10] E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, "On the performance of non-orthogonal multiple access in the finite blocklength regime," *Ad Hoc Networks*, vol. 84, pp. 148 – 157, 2019.

[11] Y. Hu, M. Ozmen, M. C. Gursoy, and A. Schmeink, "Optimal power allocation for QoS-constrained downlink multi-user networks in the finite blocklength regime," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 5827–5840, Sep. 2018.

[12] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, pp. 402–415, Jan. 2019.

[13] A. Anand and G. de Veciana, "Resource allocation and harq optimization for urllc traffic in 5g wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, pp. 2411–2421, Nov. 2018.

[14] M. Mousaei and B. Smida, "Optimizing pilot overhead for ultra-reliable short-packet transmission," in *Proc. IEEE ICC 2017*, Paris, France, pp. 1–5, May 2017.

[15] D. P. Palomar and Mung Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1439–1451, Aug. 2006.