Independent Encoding vs. Joint Encoding: Short Frame Structure Optimization for Heterogeneous URLLC Systems

Xiayue Liu, *Student Member, IEEE,* Xu Zhu, *Senior Member, IEEE,* Yufei Jiang, *Member, IEEE,* Jie Cao, *Member, IEEE*, Sumei Sun, *Fellow, IEEE*, and Vincent K. N. Lau, *Fellow, IEEE*

***Abstract*—Short frame structure and its optimization plays an important role in ultra reliable and low latency communication (URLLC). We investigate and compare the latency and throughput performances of the independent encoding (IE) and joint encoding (JE) frame structures for heterogeneous multi-device URLLC in the finite block length regime. There is a counter-intuitive finding that, despite a longer frame, IE enables a much lower average latency and higher reliability than JE, thanks to lower queuing latency, while JE achieves higher throughput with lower traffic heterogeneity, thanks to less channel dispersion. It is also shown that traffic heterogeneity has less adverse effects on the performance of the IE frame structure, and can even help reduce its average latency with the shortest block length first (SBF) scheduling rule proposed. We also provide an intensive analysis of the trade-off between pilot power and pilot overhead, with near-optimal pilot power and block length derived in closed form. Low-complexity joint pilot power, pilot length and block length optimization algorithms are proposed for IE and JE frame structures. Numerical results verify the effectiveness of the proposed algorithms, and also show that pilot power optimization plays a significant role in enhancing throughput at low to medium SNR.**

***Index Terms*—ultra reliable and low latency communication, joint encoding, joint optimization, pilot power boosting, heterogeneous traffic**

I. Introduction

Ultra-reliable and low-latency communication (URLLC) serves as a cornerstone for beyond-fifth-generation (B5G) and 6G wireless communications, enabling mission-critical applications such as the industrial metaverses and vehicle-to-everything (V2X) interfaces [2]–[6]. These real-time Ethernet applications require high reliability (*i.e.*, 99.999% $\sim$ 99.999999%) and low latency (*i.e.*, 500 μs$\sim$10 ms)[3], [4].

This work was supported in part by the National Natural Science Foundation of China under Grant 62171161 and Grant 11901209; in part by the Guangdong Provincial Science and Technology Program under Grant 2022A0505050022; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022B1515120018; in part by the Shenzhen Science and Technology Program under Grant ZDSYS20210623091808025, Grant KQTD20190929172545139, and Grant GXWD20220817133854003; in part by Grant BCIC-23-K10 from Guangxi Key Laboratory of Brain-inspired Computing and Intelligent Chips; and in part by Grant AoE/HA474. *(Corresponding author: Xu Zhu.)*

Xiayue Liu, Xu Zhu, Yufei Jiang and Jie Cao are with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China. (e-mail:liuxiayue@stu.hit.edu.cn; xuzhu@ieee.org; jiangyufei@hit.edu.cn; caojhitsz@ieee.org). Part of this work has been presented in 2021 Globecom [1].

Sumei Sun is with Agency for Science Technology and Research, Institute for Infocomm Research, Singapore.

Vincent K. N. Lau is with Hong Kong University of Science and Technology, Hong Kong, China.

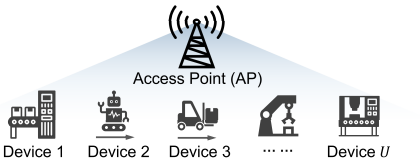


Fig. 1. Heterogeneous Downlink System

Within industrial Internet of Things (IIoT) systems, the message sizes range from 10 to 250 bytes, and the devices' speeds can reach up to 75 km/h [5]. The use of diverse devices inherently leads to heterogeneous traffic [7], [8]. For example, the controllers assign tasks to assembly robots and mobile robots [5], as illustrated in Fig. 1 . To meet the requirement of small message size and low latency, finite block length (FBL) transmission is a feasible solution, reducing transmission and processing latency [3]. However, FBL brings several new challenges, such as the non-negligible block error probability (BLEP), metadata and pilot overhead [9]. The short block length directly influences data rate, latency, and BLEP [9]–[11], thereby establishing a trade-off between bandwidth, rate, reliability, energy, and latency in the FBL regime [10]. Independent encoding (IE) [8] and joint encoding (JE) [9] are commonly used techniques for short-frame frame structures. Therefore, it is significant to investigate and compare the IE and JE frame structures under multi-device heterogeneous URLLC requirements.

A. Related Work

Theory of FBL: The channel capacity derived from Shannon's theorem in the infinite block length (IBL) regime proves imprecise in short packet communications due to non-negligible packet errors and channel dispersion caused by the short block length [11]. In [11], the relationship between achievable data rate and block length in the FBL regime is revealed over the additive white Gaussian noise (AWGN) channel and subsequently extended to multiple-antenna quasi-static fading [12], Rayleigh-fading [13], and coherent block-fading channels [14]. In [13], the diversity-multiplexing trade-off and the outage capacity are investigated. A large number of antennas at the receiver instead of at the transmitter help in reducing channel use [14]. Further, the throughput analysis in the FBL regime has been conducted across diverse scenarios, such as quality of service (QoS)-constrained networks [15], wireless-powered IoT networks [16] and massive multiple-input multiple-output (MIMO) IIoT networks [17].

IE and JE Frame Structures: How to improve the throughput of a multi-device URLLC system is an important

issue. Prior research mainly focuses on the IE frame structure, which encodes the data for each device into separate blocks, as depicted in Fig. 2(b). Power allocation is one of the efficient ways to improve system throughput, such as joint pilot and payload power optimization [17], and joint user power and block length optimization [18]. Recently, studies have demonstrated that the frame length can be condensed to improve throughput by refining the frame structure, such as the pilot sharing scheme explored in [19]. In [20], an efficient JE frame structure is proposed, where messages from different devices are combined and jointly encoded into a single packet, as illustrated in Fig. 3(b). The achievable data rate in the FBL regime [11] indicates that there is a contradiction between low latency and high data rate. Given the signal-to-noise ratio (SNR) and BLEP, the data rate decreases with the reduced number of transmitted symbols due to the channel dispersion. There is a critical trade-off between block length and data rate. The JE frame structure leverages the reduced channel dispersion to encode a large number of messages, thereby reducing transmission latency within the shortened frame length [9], [20].

The study in [9] suggests new trade-offs in the JE frame structure, such as the trade-off between encoding efficiency and message processing delays. In [20], the trade-off between the average frame length and power consumption in the JE frame structure is investigated. In [21], the trade-off between the age of information (AoI) and the number of packets encoded jointly is studied. [22] derives the AoI expressions of the IE and JE frame structures in homogeneous scenarios, respectively. Recently, in [23], the AoI performance is compared between the broadcast (JE) and unicast (IE) first-come-first-served (FCFS) systems, and found that the JE frame structure is better suited for the systems that serve a limited number of devices, transmitting a substantial amount of common information, and covering a small area. [24] explores the trade-off between the queuing delay and reliability in the JE frame structure. In [25], a novel JE structure scheme is designed to provide inter-user orthogonal channel utilization, offering a 20% improvement over the IE frame structure under the assumption of perfect channel state information (CSI). However, the existing works [20]–[25] do not conduct a comparative performance analysis between the IE and JE frame structures in terms of latency and throughput. Moreover, the impact of imperfect CSI on the JE frame structure remains unexplored. Overall, there has remained several open issues in the literature. Firstly, regarding pilot-based channel estimation, the superior performance of either the IE or JE frame structures under specific conditions is not known. Secondly, the influence of inherent heterogeneity on latency and throughput performance in multi-user URLLC systems remains unexamined.

Block Length and Pilot Length Optimization for URLLC: Pilot symbols play an important role in practical wireless communication systems, such as channel estimation [26], synchronization [27], and radar sensing [28]. It is worth noting that, contrary to conventional IBL systems, FBL systems require a pilot overhead exceeding 25% to ensure high reliability [29]. Therefore, pilot overhead needs to be minimized, while balancing the trade-off between low latency, high reliability, and high throughput. In the context of IBL systems, an optimal pilot overhead is derived by unifying block fading and continuous fading channels [30]. However, the optimization in the IBL regime is not applicable in the FBL regime, due to the non-negligible BLEP caused by the FBL [29]. In [31], the block length was optimized to minimize the average Age of Information (AoI) for a single-user FCFS system under non-preemption, preemption, and retransmission schemes. The investigation of joint optimization of pilot length and block length in the FBL regime [32] illustrates a better throughput performance compared to algorithms that only optimize either block length [33] or pilot length [29]. Furthermore, existing works of block structure optimization in the FBL regime [17], [32], [33] are based on the assumption of block fading, which is applicable to static or low-speed moving devices. Nonetheless, analyses based on the block fading model cannot meet the requirement of fast-moving devices [1], such as mobile robots in IIoT systems [5] and vehicles in V2X systems [6]. Therefore, it is important to investigate the joint optimization of pilot power and pilot overhead under continuous fading. Notably, however, the transcendental equations with respect to efficient SNR under continuous fading are particularly complex to solve and unexplored in the previous literature.

Pilot Power Boosting: Pilot power boosting is an effective way to reduce pilot overhead while improving throughput [30], [34]. Boosting pilot power can enhance the channel estimation accuracy to obtain a better performance in the limited transmission power [17]. Pilot symbols with a high SNR also help improve synchronization performance, thus reducing BLEP in practical systems [27]. In previous works, pilot power boosting has been investigated in the IBL regime [27], [30], [34], but not in the FBL regime. In [17], the trade-off between pilot power and payload power is studied in the FBL regime, but the pilot length has not been optimized. In the conference version of this paper [1], the joint impact of pilot overhead and pilot power on throughput is investigated for single-block-transmission point-to-point FBL systems. The joint impact on latency and throughout has not been investigated for multi-device systems.

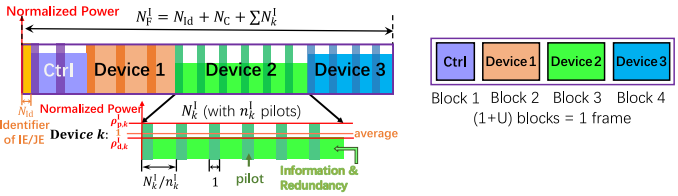
B. Contributions

Motivated by the above open issues, we investigate the IE and JE frame structures for a heterogeneous multi-device URLLC downlink in the FBL regime over continuous fading channels, respectively. Distinct from our prior work focused on a single-device system model [1], this study not only extends to a multi-device framework but also delivers a deeper comparative analysis on latency and throughput performances, introducing novel algorithms and offering more comparative insights with a richer array of simulation results. We aim to answer open questions as follows:

1) Which of the IE and JE frame structures achieves a better throughput and latency performance in heterogeneous URLLC systems over continuous fading channels?
2) How to utilize traffic heterogeneity to achieve throughput enhancement and latency reduction?
3) How to strike the balance between pilot power and pilot overhead in the FBL regime?

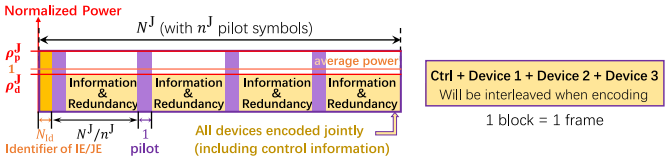
The above questions reflect closely related aspects of the performance optimization problem and are addressed in this paper. Our main contributions are as follows:

- To the best of our knowledge, this is the first work to investigate and compare the latency and throughput performances of the IE and JE frame structures in FBL heterogeneous URLLC systems. We reveal a counter-intuitive finding that, despite a longer frame, the IE structure features a much lower average latency and higher reliability than the JE frame structure, thanks to the lower queuing latency for each user within the frame when transmitting the same information. This is different from the suggestion in [9] that considered transmission latency only. On the other hand, the JE frame structure achieves higher throughput with lower traffic heterogeneity, thanks to the lower channel dispersion with a higher block length. These findings are supported by the closed-form expressions derived for the average latency gain of the IE frame structure over the JE frame structure and the throughput gain of the JE frame structure over the IE frame structure. In addition, the impacts of channel capacity, the number of devices, and the number of information bits per device on the throughput gain of the JE frame structure is also investigated.
- We investigate the effect of traffic heterogeneity on the average latency and throughput performances of the IE and JE frame structures, respectively, while in the previous work [9], [21], [22], [24], the impact of heterogeneity was not considered. It is shown that higher traffic heterogeneity has a less adverse effect on the performance of the IE frame structure than on that of the JE frame structure. It is further proved that the traffic heterogeneity can be utilized in scheduling to help reduce the average latency of the IE frame structure. A shortest block length first (SBF) scheduling rule is proposed accordingly, which allows the devices with less block lengths to experience a lower queuing latency, thereby minimizing the average latency of the IE frame structure. In contrast, the first-come-first-serve round-robin scheduling rule in [22] results in a higher average latency with higher traffic heterogeneity. The JE frame structure is shown to be preferable for the low heterogeneity case in terms of throughput, however, under high heterogeneity conditions, the JE frame structure's performance is less favorable as the users with higher transmission requirements play a more significant role in the overall system performance.
- We provide an intensive analysis of the trade-off between pilot power and pilot overhead through a throughput maximization problem in the FBL regime. Closed-form expressions are derived for the near optimal pilot power and block length for each device over continuous fading, by solving complex transcendental equations. We propose two low-complexity algorithms to jointly optimize pilot power, pilot length and block length, namely the joint optimization for the multi-device (JOMD) IE frame structure (JOMD-IE) algorithm and the JOMD JE frame structure (JOMD-JE) algorithm. These algorithms are built upon the single-device joint pilot power, pilot length, and block length optimization (JPLLO) algorithm in our previous work [1]. It is revealed that boosting pilot power helps reduce pilot overhead and improve throughput especially at low to medium SNR, but over-boosting can be counterproductive. Conversely, at high SNR, the impact of pilot overhead on throughput becomes more significant. Numerical results also show that the proposed JOMD-IE and JOMD-JE algorithms achieve a near-optimal throughput performance with a dramatic complexity reduction over exhaustive search, and also outperform the previous work with no pilot power boosting [19], [32].



(a) Frame structure (b) Arrangement of information bits

Fig. 2. Frame structure of the independent-encoding (IE) scheme. In the frame structure, for each data block, the light color part indicates its data symbols (information and redundancy), and the dark shade indicates its pilots.



(a) Frame structure (b) Arrangement of information bits

Fig. 3. Frame structure of the joint-encoding (JE) scheme.

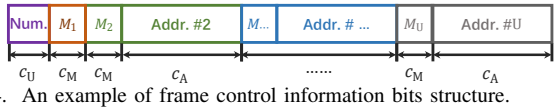


Fig. 4. An example of frame control information bits structure.

The rest of this paper is organized as follows. Section II introduces the frame structures and channel model. Section III presents the analysis of the impact of pilot power on pilot overhead and system throughput in the FBL regime under continuous fading. In Section IV, joint optimization of frame length, pilot lengths and pilot power is investigated to maximize throughput for the IE and JE frame structures, respectively. The average latency analysis and performance comparison between the two frame structures are presented in Section V. Simulation results are given in Section VI. Finally, Section VII draws the conclusion. Table I summarizes key notation in this paper.

II. System Model

We consider a heterogeneous downlink URLLC system with a single-antenna access point (AP) and U single-antenna devices, as illustrated in Fig. 1. The heterogeneity among these devices is manifested in their distinct information size, modulation order, SNR, Doppler frequency, BLEP limit, pilot power limit, block length and pilot length. These devices operate under a time-division access (TDMA) framework.

TABLE I
KEY NOTATION GLOSSARY

Parameters	Description	Parameters	Description
Ω	Superscript indicating the type of frame structure, $\Omega = \mathrm{I}, \mathrm{J}$ indicates the IE and JE structure, respectively	$N^{\Omega}_{\epsilon_{\mathrm{thr},u}}$	Minimum block length subject to $\epsilon_{\mathrm{thr},u}$
u, k	Device and block indexes, $k = u(\Omega = \mathrm{I}), \mathrm{k} = 1(\Omega = \mathrm{J})$	$N^{\Omega}_{\mathrm{opt},u}, P_{\mathrm{P_{opt}},k}$	Optimal block length and optimal total pilot power
$\alpha^{\Omega}_{\mathrm{min,C}}, \alpha^{\Omega}_{\mathrm{min},k}$	Minimum pilot overhead of the control and the k-th data block	$N^{\Omega}_{\epsilon_{\mathrm{thr},u}}$	Minimum block length subject to $\epsilon_{\mathrm{thr},u}$
β	Frame length ratio of the JE to the IE frame structure $\rho^{\mathrm{I}}_{\mathrm{pthr},k}$	$n^{\Omega}_k, \alpha^{\Omega}_k$	Pilot length and pilot overhead
β^{D}	Threshold of β that the average latency of the JE frame structure is lower than that of the IE frame structure	$\rho^{\Omega}_{\mathrm{d,k}}, \rho^{\Omega}_{\mathrm{p,k}}$	Normalized power of data and pilot symbols
$\epsilon^{\Omega}_{\mathrm{C},u}, \epsilon^{\Omega}_u$	BLEP of control and data information of the u-th device	$\rho^{\mathrm{I}}_{\mathrm{pthr},k}$	Maximum limit of normalized power of pilot symbols
$\epsilon_{\mathrm{thr,C}}, \epsilon_{\mathrm{thr},u}$	Maximum limit of BLEP of control and data information	$P^{\Omega}_k, P^{\Omega}_{\mathrm{p,k}}$	Average power and total pilot power
$\eta^{\Omega}_{\mathrm{F}}, \eta^{\Omega}_k$	Throughput of the frame and the k-th block	$T^{\Omega}_u, \overline{T^{\Omega}}$	Overall latency of device u and average latency of U devices
$f_{\mathrm{D},u}, f_{m,u}$	Normalized Doppler frequency and Doppler frequency	U	Number of devices
$L_u, L_k,$ $L_{\mathrm{C}}, L_{\mathrm{sum}}$	Number of information bits of device data, block, frame control information and frame	$\overline{T^{\Omega}_{\mathrm{Het}}}, \overline{T^{\Omega}_{\mathrm{Homo}}}$	Average latency under heterogeneous and homogeneous scenarios
M_u, M_k	Modulation order of the u-th device and the k-th block	$\gamma_u, \gamma^{\Omega}_{\mathrm{e},u}$	Average SNR and effective SNR
$N_{\mathrm{Id}}, N_{\mathrm{C}}$	Block length of frame structure type indicator and control block	$\Delta\overline{T}$	Average latency reduction ratio of the IE frame structure to the JE frame structure
$N^{\Omega}_{\mathrm{F}}, N^{\Omega}_k$	Frame length and the data block length	$\Delta\eta_{\mathrm{F}}$	Throughput improvement of the JE frame structure over the IE frame structure
$N_{\mathrm{Fthr}}, N^{\Omega}_{\mathrm{thr},k}$	Maximum limit of frame length and block length	$\sigma^*_\gamma, \sigma^*_N$	Normalized standard deviation of SNR and block length

A. Frame Structure

We consider two frame structures in this work, *i.e.*, the IE frame structure and the JE frame structure [9], as illustrated in Figs. 2 and 3 , respectively. Let N^{Ω}_{F} $(\Omega = \{\mathrm{I}, \mathrm{J}\})$ denote the frame length, where $N^{\mathrm{I}}_{\mathrm{F}}$ indicates the IE frame length and $N^{\mathrm{J}}_{\mathrm{F}}$ indicates the JE frame length. N_{Id} symbols are placed in the front of the frame, allowing the device to know the type of frame structure to decode correctly, either the IE or JE frame structure.

The IE frame structure allows the data information of U devices independently to be encoded into U blocks together with a control information block of N_{C} symbols containing L_{C} information bits in the front, as shown in Fig. 2 . Let L_u denote the information bits of the u-th $(u = 1, 2, \ldots, U)$ device. Let N^{Ω}_k denote the number of symbols in the k-th $(k = 1, 2, \ldots, U)$ block, each with $L_k/\log_2(M_k)$ information symbols, where L_k is the number of information bits in the k-th block, and M_k is the M_k-ary modulation. For the IE frame structure, the u-th device is mapped into the k-th block. The control information is used at each device to identify which block the device should decode. For example, Fig. 4 demonstrates the control information structure. The first c_{U} bits are used to recognize the value of the number of devices in the frame, followed by c_{M} and c_{A} bits to represent modulation order of M_u and the starting address of the first symbol for each device, respectively. The values of c_{U}, c_{M} and c_{A} are designed properly to reduce the control information overhead. Please note that the starting address of the first device is not recorded, as the address can be calculated via the number of devices and control information structure. The address of the first device equals the bits of control information, given as $L_{\mathrm{C}} = c_{\mathrm{U}} + c_{\mathrm{M}} \cdot U + c_{\mathrm{A}} \cdot (U - 1)$. This can reduce the number of bits in the control information block of the short packet transmission, while the other methods in the literature [19]–[21], [24], [25] do not investigate the important property.

For the JE frame structure, devices' data and control information are encoded jointly into a single block that consists of $N^{\mathrm{J}}_{\mathrm{F}}$ symbols as $k = 1$ when $\Omega = \mathrm{J}$, as shown in Fig. 3 . Thus, a single block is the frame for the JE frame structure. For convenience, let N^{J} denote the number of symbols for the JE frame structure. The starting address in the control information in the IE frame structure corresponds to the first symbol of each device. However, this is not applicable to the JE frame structure, as it is difficult to extract the starting address of the first symbol for two reasons. First, the redundancy information is shared by all devices. Second, the information bits of all devices are usually interleaved encoding for high reliability. The first symbol of each device shuffles in the JE frame. Therefore, the starting address of each device in the JE frame structure is extracted in the decoded bits rather than in the received symbol in the IE frame structure.

Let $\alpha^{\Omega}_k = n^{\Omega}_k/N^{\Omega}_k$ denote the pilot overhead ratio, with n^{Ω}_k denoting pilots in the k-th block. In the case of continuous fading, both the IE and JE frame structures follow the regular periodic placement (RPP)-1 as the optimal pattern [26], [35], where a pilot symbol is inserted every $N^{\Omega}_k/n^{\Omega}_k - 1$ data symbols, as shown in Figs. 2(a) and 3(a) . The pilot symbols are not shared between blocks. Let P^{Ω}_k denote the average power of all transmitted symbols in the k-th block, which includes both pilot and data symbols. Furthermore, let $\rho^{\Omega}_{\mathrm{p},k}$ and $\rho^{\Omega}_{\mathrm{d},k}$ denote the power of the pilot and data symbols normalized to P^{Ω}_k, respectively. Thus, the power of each pilot symbol is given by $\rho^{\Omega}_{\mathrm{p},k}P^{\Omega}_k$ and that of each data symbol by $\rho^{\Omega}_{\mathrm{d},k}P^{\Omega}_k$. Given $\rho^{\Omega}_{\mathrm{p},k}$ and α^{Ω}_k, we can determine the normalized transmission power of the data symbols by:

$$\rho^{\Omega}_{\mathrm{d},k} = \frac{1 - \rho^{\Omega}_{\mathrm{p},k}\alpha^{\Omega}_k}{1 - \alpha^{\Omega}_k}. \tag{1}$$

For simplicity, let $n^{\mathrm{J}}, \alpha^{\mathrm{J}}, P^{\mathrm{J}}, \rho^{\mathrm{J}}_{\mathrm{p}}$ and $\rho^{\mathrm{J}}_{\mathrm{d}}$ denote the pilot length, pilot overhead ratio, average transmission power, normalized transmission power of pilot symbols, and normalized transmission power of data symbols for the JE frame structure with a single block, respectively.

B. Channel Model

We consider the Rayleigh fading channel with continuous amplitude variations (termed as continuous fading [30]), where all U devices are in mobility. Let $h_{k,l,u}$ denote the continuous

channel in the l-th ($l = 1, 2, \ldots, N_k^{\Omega}$) symbol of the k-th block between the AP and the u-th device. In the continuous fading, $h_{k,l,u}$ is a discrete-time complex Gaussian stationary random process, with the spectral distribution function satisfying the Doppler spectrum. Let $f_{\mathrm{D},u} = T_{\mathrm{s}} f_{m,u}$ denote the normalized Doppler frequency of the u-th device, where T_{s} is the symbol period, $f_{m,u} = \upsilon_u/\lambda$ is the Doppler frequency, υ_u is the relative velocity between the u-th device and AP in meters per second (m/s), and λ is the carrier wavelength in meter (m). The coherence time is $T_{c,u} \approx 0.423/f_{m,u} = 0.423\, T_{\mathrm{s}}/f_{\mathrm{D},u}$. There are two important Doppler spectra [26]. One is the Clarke-Jakes spectrum:

$$S_{\mathrm{H}}(\nu) = 1 \Big/ \left(\pi \sqrt{f_{\mathrm{D},u}^2 - (\nu T_{\mathrm{s}})^2}\right). \tag{2}$$

The other one is the rectangular spectrum, given by:

$$S_{\mathrm{H}}(\nu) = 1/(2 f_{\mathrm{D},u}). \tag{3}$$

where ν denotes the frequency shift relative to the carrier frequency, the definition range of the spectra is $|\nu| \leq f_{m,u}$. According to NB-IoT standard in 5G NR at 3.5 GHz [36] with sub-carrier space (SCS) of 15 kHz, there are $T_{c,u} \approx$1.09 ms and $f_{\mathrm{D},u} \approx 0.026$, when υ_u=120 km/h. The block fading can be treated as a special case of the continuous fading, when $h_{k,l,u}$ remains constant in the duration of one block.

Let $x_{k,l}$ denote the transmit signal in the l-th symbol of the k-th block, with independent identically distributed (i.i.d.) complex Gaussian random variables of zero mean and unit variance, *i.e.*, $x_{k,l} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ [34]. The l-th symbol of the k-th block received by the u-th device in the FBL system is:

$$y_{k,l,u} = \begin{cases} h_{k,l,u} \rho_{\mathrm{p},k}^{\Omega} P_k^{\Omega} x_{k,l} + w_{k,l,u}, & (l \bmod [N_k^{\Omega}/n_k^{\Omega}]) = 1 \\ h_{k,l,u} \rho_{\mathrm{d},k}^{\Omega} P_k^{\Omega} x_{k,l} + w_{k,l,u}, & \text{others}, \end{cases} \tag{4}$$

where $\bmod$ is the modulo operation that returns the remainder of a division, $[\cdot]$ denotes rounding to the nearest integer, and $w_{k,l,u}$ denotes the additive white Gaussian noise (AWGN) in the l-th symbol of the k-th block between the AP and the u-th device, with zero mean and variance of ϖ_0, *i.e.*, $w_{k,l,u} \sim \mathcal{N}_{\mathbb{C}}(0, \varpi_0)$. The average SNR of the u-th device receiving the k-th block is defined as:

$$\gamma_{u,k} = \frac{P_k^{\Omega} \sum_{l=1}^{N_k^{\Omega}} |h_{k,l,u}|^2}{\varpi_0 N_k^{\Omega}}. \tag{5}$$

For simplicity, $\gamma_{u,k}$ is noted as γ_u, since $k = u$ with $\Omega = \mathrm{I}$ and $k = 1$ with $\Omega = \mathrm{J}$.

To ensure that the channel observed through the pilot transmissions has an unaliased spectrum, the minimum pilot overhead of the k-th block is given by [30]:

$$\alpha_{\mathsf{min},k}^{\Omega} = 2 \max_u \{f_{\mathrm{D},u}\}. \tag{6}$$

The minimum pilot overhead of the control block of the IE frame structure $\alpha_{\mathsf{min},\mathrm{C}}^{\Omega}$ is also given by (6).

III. Throughput Analysis under Continuous Fading FBL Systems

In this section, we derive the throughput of the IE and JE frame structure with closed-form expressions of BLEP of data and control information. The impact of SNR, mobility and pilot power on pilot overhead and throughput are analyzed.

A. Throughput Analysis

Considering U devices in a single frame, devices can decode the received data correctly only if both the data and control information are received correctly. Note that since the control and data messages are jointly encoded in the JE frame structure, ϵ_u^{J} is actually the probability that errors happen for either the data or control messages. Since the system is concerned with the data information of devices, the control information is not counted in the throughput.

The throughput of the frame for U devices in bits per channel use (bpcu) is given by:

$$\eta_{\mathrm{F}}^{\Omega} = \begin{cases} \frac{\sum_{u=1}^{U} L_u \left(1-\epsilon_u^{\mathrm{I}}\right)\left(1-\epsilon_{\mathrm{C},u}^{\mathrm{I}}\right)}{N_{\mathrm{F}}^{\mathrm{I}}}, & \Omega = \mathrm{I} \\ \frac{\sum_{u=1}^{U} L_u \left(1-\epsilon_u^{\mathrm{J}}\right)}{N_{\mathrm{F}}^{\mathrm{J}}}, & \Omega = \mathrm{J}, \end{cases} \tag{7}$$

where ϵ_u^{Ω} denotes the BLEP of the data information of the u-th device, and $\epsilon_{\mathrm{C},u}^{\Omega}$ denotes the BLEP of the control information received by the u-th device.

Based on the system model, the data of devices and control information are arranged into different blocks in the IE frame structure. Hence, the frame length of the IE frame structure $N_{\mathrm{F}}^{\mathrm{I}}$ is given by:

$$N_{\mathrm{F}}^{\mathrm{I}} = N_{\mathrm{Id}} + N_{\mathrm{C}} + \sum_{k=1}^{K} N_k^{\mathrm{I}}. \tag{8}$$

For the JE frame structure, the data of devices and control information is encoded jointly in a single block as shown in Fig. 3 . The frame length of the JE frame structure is as:

$$N_{\mathrm{F}}^{\mathrm{J}} = N_{\mathrm{Id}} + N^{\mathrm{J}}. \tag{9}$$

According to the FBL theory, the achievable rate of short packets is given by $R \approx C(\gamma) - \sqrt{V(\gamma)/N}\, Q^{-1}(\epsilon)$ [11], where $C(\gamma)$ denotes the channel capacity, $V(\gamma)$ denotes the channel dispersion, ϵ denotes the BLEP, N denotes block length, and $Q^{-1}(\cdot)$ is the inverse of the Q-function given by $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. Specifically, channel capacity $C(\gamma) = \log_2(1+\gamma)$ with $\gamma > 0$, and channel dispersion $V(\gamma) = (1 - \frac{1}{(1+\gamma)^2})(\log_2(e))^2$, with e denoting the Euler's number. Considering pilots for channel estimation, the BLEP of data information of the u-th device is derived as [1], [32]:

$$\epsilon_u^{\Omega} = \begin{cases} Q\left(\sqrt{\frac{N_k^{\mathrm{I}} - n_k^{\mathrm{I}}}{V\left(\gamma_{\mathrm{e},u}^{\mathrm{I}}\right)}}\left(C\left(\gamma_{\mathrm{e},u}^{\mathrm{I}}\right) - \frac{L_u}{N_k^{\mathrm{I}} - n_k^{\mathrm{I}}} + \frac{\log_2\left(N_k^{\mathrm{I}} - n_k^{\mathrm{I}}\right)}{2\left(N_k^{\mathrm{I}} - n_k^{\mathrm{I}}\right)}\right)\right), & \Omega = \mathrm{I} \\ Q\left(\sqrt{\frac{N^{\mathrm{J}} - n^{\mathrm{J}}}{V\left(\gamma_{\mathrm{e},u}^{\mathrm{J}}\right)}}\left(C\left(\gamma_{\mathrm{e},u}^{\mathrm{J}}\right) - \frac{L_{\mathrm{sum}}}{N^{\mathrm{J}} - n^{\mathrm{J}}} + \frac{\log_2\left(N^{\mathrm{J}} - n^{\mathrm{J}}\right)}{2\left(N^{\mathrm{J}} - n^{\mathrm{J}}\right)}\right)\right), & \Omega = \mathrm{J}. \end{cases} \tag{10}$$

where $\gamma_{\mathrm{e},u}^{\Omega}$ denotes the effective SNR of the u-th device after channel estimation, and $L_{\mathrm{sum}} = L_{\mathrm{C}} + \sum_{u=1}^{U} L_u$ denotes the total information bits in the frame. The specific formula for effective SNR is defined in the subsequent subsection. Different from the throughput, the BLEP calculates both data and control information in the block rather than just the devices' valid information in the JE frame structure.

Due to different channels of devices, the same control information transmitted to different devices has different BLEP. In the IE frame structure, the control information is in an independent block. Thus, $\epsilon_{\mathrm{C},u}^{\mathrm{I}}$ relates to the control block structure and the channel state of the u-th device as:

$$\epsilon_{\mathrm{C},u}^{\mathrm{I}}=Q\left(\sqrt{\frac{N_{\mathrm{C}}^{\mathrm{I}}-n_{\mathrm{C}}^{\mathrm{I}}}{V\left(\gamma_{\mathrm{e},u}^{\mathrm{I}}\right)}}\left(C\left(\gamma_{\mathrm{e},u}^{\mathrm{I}}\right)-\frac{L_{\mathrm{C}}}{N_{\mathrm{C}}^{\mathrm{I}}-n_{\mathrm{C}}^{\mathrm{I}}}+\frac{\log_2\left(N_{\mathrm{C}}^{\mathrm{I}}-n_{\mathrm{C}}^{\mathrm{I}}\right)}{2\left(N_{\mathrm{C}}^{\mathrm{I}}-n_{\mathrm{C}}^{\mathrm{I}}\right)}\right)\right). \tag{11}$$

In the JE frame structure, the control information mixes up with devices' data in a shared block. The BLEP of the control information is equal to that of the device as:

$$\epsilon_{\mathrm{C},u}^{\mathrm{J}}=\epsilon_u^{\mathrm{J}}. \tag{12}$$

B. Impact of Pilot Power with Effective SNR Analysis

The effective SNR refers to the SNR after channel estimation, related to SNR γ_u, mean square error MSE_u^{Ω} and the normalized power of data symbols $\rho_{\mathrm{d},k}^{\Omega}$ [30], as shown as:

$$\gamma_{\mathrm{e},u}^{\Omega}=\frac{\gamma_u\left(1-\mathsf{MSE}_u^{\Omega}\right)}{1/\rho_{\mathrm{d},k}^{\Omega}+\gamma_u\mathsf{MSE}_u^{\Omega}}. \tag{13}$$

As in (13), $\gamma_{\mathrm{e},u}^{\Omega}$ is a monotonically decreasing function with respect to MSE_u^{Ω}. This is proved by the partial derivative of $\gamma_{\mathrm{e},u}^{\Omega}$ with respect to MSE_u^{Ω} as $\partial\gamma_{\mathrm{e},u}^{\Omega}/\partial\mathsf{MSE}_u^{\Omega}=-\gamma_u(\gamma_u+1/\rho_{\mathrm{d},k}^{\Omega})/(\gamma_u\mathsf{MSE}_u^{\Omega}+1/\rho_{\mathrm{d},k}^{\Omega})^2<0$.

We assume that the minimum mean square error (MMSE)-based channel estimator is used. In the continuous fading model, the rectangular spectrum results in the worst-case estimation error [26]. The mean square error MSE_u^{Ω} of channel estimation for the u-th device is given by [30]:

$$\mathsf{MSE}_u^{\Omega}=\frac{1}{1+\frac{\alpha_k^{\Omega}}{2f_{\mathrm{D},u}}\rho_{\mathrm{p},k}^{\Omega}\gamma_u}=\frac{1}{1+\frac{n_k^{\Omega}}{2f_{\mathrm{D},u}N_k^{\Omega}}\rho_{\mathrm{p},k}^{\Omega}\gamma_u}. \tag{14}$$

It indicates that MSE_u^{Ω} increases under high mobility or low SNR. The increased MSE_u^{Ω} can be reduced by increasing pilot power instead of a large increment in pilot overhead. Besides, the block fading model provides the effect of mean square error and effective SNR the same as the continuous fading model. Thus, $f_{\mathrm{D},u}$ in the continuous fading is replaced by $f_{\mathrm{D},k}=1/(2N_k^{\Omega})$ in block fading [30].

We derive a closed-form expression for the optimal block length in (22) in Subsection IV-A . This expression can be used to analyze how pilot power $\rho_{\mathrm{p},k}^{\Omega}$ that impacts pilot overhead α_k^{Ω} and throughput $\eta_{\mathrm{F}}^{\Omega}$.

Proposition 1: The effective SNR $\gamma_{\mathrm{e},u}^{\Omega}$ is concave with respect to normalized pilot power $\rho_{\mathrm{p},k}^{\Omega}$. With a given pilot overhead α_k^{Ω}, the optimal $\rho_{\mathrm{p},k}^{\Omega}$ for the u-th device to maximize $\gamma_{\mathrm{e},u}^{\Omega}$, denoted as $\widehat{\rho_{\mathrm{p},k,u}^{\Omega}}(\alpha_k^{\Omega})$, is derived as:

$$\widehat{\rho_{\mathrm{p},k,u}^{\Omega}}(\alpha_k^{\Omega})=\left(-\sqrt{2f_{\mathrm{D},u}\left(2f_{\mathrm{D},u}+\gamma_{\mathrm{e},u}^{\Omega}\right)\left(\alpha_k^{\Omega}-1\right)\left(\alpha_k^{\Omega}-\gamma_{\mathrm{e},u}^{\Omega}-1\right)}+2f_{\mathrm{D},u}\left(\alpha_k^{\Omega}-\gamma_{\mathrm{e},u}^{\Omega}-1\right)\right)\Big/\left(\alpha_k^{\Omega}\gamma_{\mathrm{e},u}^{\Omega}\left(2f_{\mathrm{D},u}+\alpha_k^{\Omega}-1\right)\right) \tag{15}$$

Proof: See Appendix A . ■

Proposition 2: The minimum data and redundant symbol length satisfying the BLEP threshold, defined as $\mathring{N}_u^{\Omega}=(N_{\epsilon_{\mathrm{thr},u}}-n_k^{\Omega})$, is monotonically decreasing with respect to $\gamma_{\mathrm{e},u}^{\Omega}$.

Proof: See Appendix B . ■

Remark 1: Based on **Propositions 1** and **2**, we deduce that boosting pilot power $\rho_{\mathrm{p},k}^{\Omega}$ can reduce pilot overhead α_k^{Ω} and increase throughput $\eta_{\mathrm{F}}^{\Omega}$, but over-boosting adversely affect $\eta_{\mathrm{F}}^{\Omega}$. These propositions show that $\rho_{\mathrm{p},k}^{\Omega}$ influences the effective SNR $\gamma_{\mathrm{e},u}^{\Omega}$, which in turn affects $\mathring{N}_u^{\Omega}$. From (7), when block length drops, $\eta_{\mathrm{F}}^{\Omega}$ rises (considering negligible BLEP in the URLLC system parameters). Thus, as $\rho_{\mathrm{p},k}^{\Omega}$ grows, there's an optimal point for maximizing throughput. However, when $\rho_{\mathrm{p},k}^{\Omega}$ becomes excessively high, $\eta_{\mathrm{F}}^{\Omega}$ would decrease. An intuitive explanation is that an excessive $\rho_{\mathrm{p},k}^{\Omega}$ tends to diminish the power of data symbols, resulting in a reduced $\gamma_{\mathrm{e},u}^{\Omega}$ as outlined in (13). Besides, to achieve high throughput, we can reduce the pilot overhead, as long as we maintain $\gamma_{\mathrm{e},u}^{\Omega}$ unchanged by increasing pilot power as indicated in (14). Notice that the pilot overhead cannot be reduced below the minimum value defined by (6). In summary, in weak channel conditions, an appropriate increase in $\rho_{\mathrm{p},k}^{\Omega}$ can optimize α_k^{Ω} and $\eta_{\mathrm{F}}^{\Omega}$, but over-boosting $\rho_{\mathrm{p},k}^{\Omega}$ results in a point where α_k^{Ω} cannot be minimized further, and $\eta_{\mathrm{F}}^{\Omega}$ goes low. This analysis is verified by Figs. 10 and 11 .

IV. Frame Structure Optimization of IE and JE

To compare IE and JE frame structures, throughput maximization problems are formulated. An in-depth analysis of the throughput maximization problem concerning the IE frame structure is undertaken in Subsection IV-A, wherein the closed-form optimal solutions in resolving the problem are also derived. The proposed JOMD-IE and JOMD-JE algorithms are detailed in Subsections IV-B and IV-C , respectively.

A. Throughput Maximization of the IE Frame Structure: Problem Formulation and Analysis

Based on the analysis in Subsection III-A , we formulate an optimization problem P1 to maximize the throughput of the IE frame structure by jointly optimizing block lengths $\mathbf{N}^{\mathrm{I}}$, pilot lengths $\mathbf{n}^{\mathrm{I}}$ and the normalized pilot powers $\boldsymbol{\rho}_{\mathbf{p}}^{\mathrm{I}}$ for each block, as:

$$\begin{aligned}
\mathbf{P1:}\quad & \max_{\mathbf{N}^{\mathrm{I}},\mathbf{n}^{\mathrm{I}},\boldsymbol{\rho}_{\mathbf{p}}^{\mathrm{I}}} \quad \eta_{\mathrm{F}}^{\mathrm{I}} && (16)\\
\text{s.t.}\quad & (C1): \quad N_{\mathrm{F}}^{\mathrm{I}}-N_{\mathrm{Fthr}}\leqslant 0,\\
& (C2): \quad \epsilon_{\mathrm{C},u}^{\mathrm{I}}-\epsilon_{\mathrm{thr,C}}\leqslant 0,\ \forall u,\\
& (C3): \quad \epsilon_u^{\mathrm{I}}-\epsilon_{\mathrm{thr},u}\leqslant 0,\ \forall u,\\
& (C4): \quad 1\leqslant\rho_{\mathrm{p}_k}^{\mathrm{I}}\leqslant\rho_{\mathrm{pthr},k}^{\mathrm{I}},\ \forall k,\\
& (C5): \quad N_k^{\mathrm{I}}-n_k^{\mathrm{I}}\geqslant L_k^{\mathrm{I}}/\log_2\left(M_k^{\mathrm{I}}\right),\ \forall k,\\
& (C6): \quad n_k^{\mathrm{I}}/N_k^{\mathrm{I}}\geqslant\alpha_{\mathsf{min},k}^{\mathrm{I}},\ \forall k,\\
& (C7): \quad N_k^{\mathrm{I}},n_k^{\mathrm{I}}\in\mathbb{N},\ \forall k,\\
& \qquad u=1,...,U,\\
& \qquad k=1,...,U,\mathrm{C}.
\end{aligned}$$

where $\mathbf{N}^{\mathrm{I}}=[N_1^{\mathrm{I}},...,N_U^{\mathrm{I}},N_{\mathrm{C}}^{\mathrm{I}}]$ is the block length vector, $\mathbf{n}^{\mathrm{I}}=[n_1^{\mathrm{I}},...,n_U^{\mathrm{I}},n_{\mathrm{C}}^{\mathrm{I}}]$ is the pilot length vector, $\boldsymbol{\rho}_{\mathbf{p}}^{\mathrm{I}}=[\rho_{\mathrm{p},1}^{\mathrm{I}},...,\rho_{\mathrm{p},U}^{\mathrm{I}},\rho_{\mathrm{p,C}}^{\mathrm{I}}]$ is the vector of normalized power of pilot symbols, with $n_{\mathrm{C}}^{\mathrm{I}}$ and $\rho_{\mathrm{p,C}}^{\mathrm{I}}$ denoting the pilot length and the normalized pilot power in the control information block, respectively. For the IE frame structure, the u-th device is mapped into the k-th block. Thus, in Constraints $(C4)\sim(C7)$, $k=u$ when $k=1,...,U$.

Constraint $(C1)$ limits the maximum frame length N_{Fthr}. Constraint $(C2)$ specifies the control information BLEP threshold $\epsilon_{\mathrm{thr,C}}$. Constraint $(C3)$ specifies the data information BLEP

threshold of the u-th device $\epsilon_{\text{thr},u}$. Constraint $(C4)$ limits the normalized pilot power of the k-th block by $\rho^{\text{I}}_{\text{pthr},k}$. Constraint $(C5)$ ensures there are enough symbols to include information bits in each block. Constraint $(C6)$ meets the minimum pilot ratio requirement in (6).

Problem P1 is a mixed-integer optimization problem with $3(U+1)$ variables and $1+2U+4(U+1)$ constraints. The complexity of the exhaustive search for the optimal solution is high. Notice that the control block and devices' data blocks can be regarded as being independent of each other as in (7), and the change of parameters of one block does not affect the parameters of the other blocks when N_{Fthr} is sufficient. Therefore, we decompose the problem P1 (16) into $(U+1)$ individual sub-blocks' block structure optimization problems with 3 variables each. The sub-problem of the device data block is formulated with 6 constraints as follows:

$$\textbf{P1.1:} \quad \max_{N_k^\Omega, n_k^\Omega, \rho_{\text{p}k}^\Omega} \quad \eta_k^\Omega = L_u(1-\epsilon_u^\Omega)/N_k^\Omega \qquad (17)$$
$$\text{s.t.} \quad (C3) \sim (C7),$$
$$(C8): \quad N_k^\Omega - N_{\text{thr},k}^\Omega \leqslant 0.$$

where $N_{\text{thr},k}^\Omega$ is the block length limit for the k-th block. Constraint $(C8)$ is an alternative version of constraint $(C1)$ in the case of block structure optimization. There are U $(C2)$ constraints in the control block structure optimization sub-problem, since the control block's structure is jointly affected by U devices' channels. This sub-problem is similar to problem P1.1 and can be solved with the help of its results.

The block structure optimization problem P1.1 is a mixed integer optimization problem, which can be further decomposed into the total pilot power optimization sub-problem P1.1.1 and the block length optimization sub-problem P1.1.2, as follows:

$$\textbf{P1.1.1:} \quad \max_{P_{\text{p},k}^\Omega} \quad \eta_k^\Omega \quad \text{s.t.} \quad (C3) \sim (C8), \qquad (18)$$

$$\textbf{P1.1.2:} \quad \max_{N_k^\Omega} \quad \eta_k^\Omega \quad \text{s.t.} \quad (C3) \sim (C8). \qquad (19)$$

where $P_{\text{p},k}^\Omega = n_k^\Omega \rho_{\text{p},k}^\Omega$ is the total pilot power of the k-th block. The optimal solution of P1.1 can be obtained by the iterative way.

Proposition 3: With a given block length N_k^Ω, the throughput is concave with respect to the total pilot power $P_{\text{p},k}^\Omega$.

Proof: See Appendix C. ∎

Therefore, the total pilot power optimization sub-problem P1.1.2 can be solved by convex optimization methods.

Proposition 4: With a given pilot length n_k^Ω and normalized pilot power $\rho_{\text{p},k}^\Omega$, the throughput η_k^Ω is mono-decreasing with respect to block length N_k^Ω.

Proof: For a practical FBL URLLC system, BLEP is $\epsilon_u^\Omega < 10^{-5}$ and N_k^Ω is less than 1000 [4], [5]. As long as ϵ_u^Ω satisfies the constraint $\epsilon_{\text{thr},u}$, ϵ_u^Ω has little effect on the throughput η_k^Ω. Thus, $\eta_k^\Omega \approx L_u/N_k^\Omega$. It is obvious that η_k^Ω is mono-decreasing with respect to N_k^Ω. ∎

Therefore, the optimal solution of the block length optimization sub-problem P1.1.1 only requires the minimum value of block length that satisfies constraints $(C3), (C5) \sim (C7)$. However, transcendental equations are needed to be solved because $C(\gamma_{\text{e},u}^\Omega)$ and $V(\gamma_{\text{e},u}^\Omega)$ relate to N_k^Ω by (14) under the continuous fading model. It is hard to derive the precise analytical solution [1]. The closed-form expression of the approximate solution can be obtained. **Lemma 1** and **Lemma 2** show the approximate solutions for the two sub-problems, respectively.

Based on **Proposition 3**, $P_{\text{p}_{\text{opt},k}}$ can be derived by forcing $\partial W_{\text{S},u}^\Omega / \partial P_{\text{p},k}^\Omega = 0$. It is a transcendental equation too. In practical scenarios, for the sake of efficiency, the power of pilot symbols is typically much smaller than the overall transmission power, allowing us to assume $N_k^\Omega \gg P_{\text{p},k}^\Omega$. Additionally, since the value of function $V(x)$ exhibits little variation with respect to x, it is reasonable to simplify the expression $\sqrt{(N_k^\Omega - P_{\text{p},k}^\Omega)/V(\gamma_{\text{e},u}^\Omega)}$ in (35) to $\sqrt{N_k^\Omega / V(\gamma_u)}$. With the above simplification, $\partial W_{\text{S},u}^\Omega / \partial P_{\text{p},k}^\Omega = 0$ is a quadratic equation, which can be easily solved.

Lemma 1: With a given block length N_k^Ω and device's parameters, and under the condition of $N_k^\Omega \gg P_{\text{p},k}^\Omega$, the closed-form expression of the near-optimal total pilot power $P_{\text{p}_{\text{opt}},k}$ is given by:

$$P_{\text{p}_{\text{opt}},k} = \frac{N_k^\Omega (B_1 - B_2)}{\left(L_u \ln(2) - 2 N_k^\Omega f_{\text{D},u}\right)\gamma_u}, \qquad (20)$$

where $B_1 = \sqrt{\ln(2) L_u f_{\text{D},u}} \sqrt{L_u \ln(2) f_{\text{D},u} \gamma_u^2 + 8 B_3 N_k^\Omega}$,
$B_2 = \left(L_u(\gamma_u+2)\ln(2) + 2N_k^\Omega \gamma_u\right) f_{\text{D},u}$,
$B_3 = (f_{\text{D},u}(\gamma_u+1) + \gamma_u/2)(\gamma_u/2 + f_{\text{D},u})$.

We derive the optimal N_k^Ω for problem P1.1.1 as follows. As **Proposition 4**, the transcendental inequality $(C3)$ is needed to be solved. Since the value of $C(\gamma_{\text{e},u}^\Omega)$ is much larger than the value of $\log_2\left(N_k^\Omega - n_k^\Omega\right)/\left(2\left(N_k^\Omega - n_k^\Omega\right)\right)$, we simplify $(C3)$ as $Q\left(\sqrt{(N_k^\Omega - n_k^\Omega)/V(\gamma_{\text{e},u}^\Omega)}\left(C(\gamma_{\text{e},u}^\Omega) - L_u/(N_k^\Omega - n_k^\Omega)\right)\right) \leq \epsilon_{\text{thr},u}$. Equations (10), (13) and (14) reveal that the affect of $(N_k^\Omega - n_k^\Omega)$ on ϵ_u^Ω is greater than that on $\gamma_{\text{e},u}^\Omega$ when N_k^Ω changes a little. Thus, we treat $\gamma_{\text{e},u}^\Omega$ as a constant. The origin transcendental inequality becomes a linear inequality. The minimum block length subject to $\epsilon_{\text{thr},u}$ is derived as:

$$N_{\epsilon_{\text{thr},u}}^\Omega = \frac{A_{q,u}^\Omega + A_u^\Omega}{2\left(C\left(\gamma_{\text{e},u}^\Omega\right)\right)^2} + \frac{L_u}{C\left(\gamma_{\text{e},u}^\Omega\right)} + n_k^\Omega, \qquad (21)$$

where $A_u^\Omega = \sqrt{A_{qu}^\Omega + 4C(\gamma_{\text{e},u}^\Omega) L_u} \sqrt{A_{qu}^\Omega}$, $A_{qu}^\Omega = V\left(\gamma_{\text{e},u}^\Omega\right)\left(Q^{-1}(\epsilon_{\text{thr},u})\right)^2$. After combining the constraints $(C5) \sim (C7)$, we obtain:

$$N_{\text{opt},u}^\Omega = \left\lceil \min\left\{ \max\left\{ \frac{L_u}{\log_2(M_u)} + n_k^\Omega, N_{\epsilon_{\text{thr},u}}^\Omega \right\}, \min\left\{ N_{\text{thr},k}, \frac{n_k^\Omega}{2 f_{\text{D},u}} \right\} \right\} \right\rceil. \qquad (22)$$

where $\lceil \cdot \rceil$ means rounding a number up to an integer. Note that the solution of $N_{\epsilon_{\text{thr},u}}$ may not be precisely accurate since treating $\gamma_{\text{e},u}^\Omega$ as a constant. In order to get an accurate $N_{\text{opt},u}$, an iterative method is used, as shown in **Algorithm 1**. First, an approximate $N_{\epsilon_{\text{thr}}}^{(0)}$ is computed as in (21) from a random or empirical initial value. Then we use $N_{\epsilon_{\text{thr}}}^{(0)}$ to update $\gamma_{\text{e},u}^\Omega$ as in (21) to obtain an accurate $N_{\epsilon\text{thr}}^{(1)}$.

Lemma 2: With a given pilot length n_k^Ω and normalized pilot power $\rho_{\text{p},k}^\Omega$, the closed-form expression of the near-optimal block length for the u-th device, *i.e.*, $N_{\text{opt},u}^\Omega$, is given by **Algorithm 1**.

Algorithm 1 A proposed iterative algorithm to obtain the optimal block length

Input: $n_k^{\Omega}, \rho_{\mathrm{p},k}^{\Omega}, N_{\mathrm{thr},k}, \epsilon_{\mathrm{thr},u}, L_u, M_u, \gamma_u, f_{\mathrm{D},u}$;
Execute:
1: Let $\alpha = \max((0.1 \sim 0.2), 3 \times 2 f_{\mathrm{D},u}), N^{(0)} = n_k^{\Omega}/\alpha$;
2: With $N^{(0)}$, $N_{\epsilon_{\mathrm{thr}}}^{(0)}$ is obtained by (14) (13) (21);
3: With $N_{\epsilon_{\mathrm{thr}}}^{(0)}$, $N_{\epsilon_{\mathrm{thr}}}^{(1)}$ is obtained by(14) (13) (21);
4: **return** $N_{\mathrm{opt},u}$ by (22) with $N_{\epsilon_{\mathrm{thr},u}} = N_{\epsilon_{\mathrm{thr}}}^{(1)}$.

Algorithm 2 The Proposed JPLLO Algorithm

Input: $N_{\mathrm{thr},k}, \epsilon_{\mathrm{thr},u}, \rho_{\mathrm{pthr},k}^{\Omega}, L_u, M_u, \gamma_u, f_{\mathrm{D},u}, I_{\max}$;
Execute:
1: Initialize $N^{(0)} = 2L_u / \log_2(M_u)$;
2: **for** $i = 1$ **to** $I_{\max}$ **do**
3: Given $N^{(i-1)}$, $P_{\mathrm{p}}^{(i)}$ is obtained by (20);
4: $n^{(i)} = \lceil \max\{P_{\mathrm{p}}^{(i)}/\rho_{\mathrm{pthr},k}^{\Omega}, 2N^{(i-1)} f_{\mathrm{D},u}\} \rceil$;
5: $\rho_{\mathrm{p}}^{(i)} = [P_{\mathrm{p}}^{(i)}/\rho_{\mathrm{pthr},k}^{\Omega} < 2N^{(i-1)} f_{\mathrm{D},u}]$ **?** $[P_{\mathrm{p}}^{(i)}/n^{(i)}]$ **:** $\rho_{\mathrm{pthr},k}^{\Omega}$;
6: Given $n^{(i)}$ as n_k^{Ω} and $\rho_{\mathrm{p}}^{(i)}$ as $\rho_{\mathrm{p},k}^{\Omega}$, $N^{(i)}$ is obtained by **Algorithm 1**;
7: **end for**
8: **return** $N^{(i)}$, $n^{(i)}$, $\rho_{\mathrm{p}}^{(i)}$ as optimal result.

B. Throughput Maximization of the IE Frame Structure: Algorithm Design

Based on the above discussion, the JPLLO algorithm for solving single block structure optimization problem P1.1 is proposed, as shown in Algorithm 2 . At first, a proper empirical initial value of block length N_k^{Ω} helps the algorithm converge faster. Given N_k^{Ω}, the optimal total pilot power $P_{\mathrm{p},k}^{\Omega}$ is obtained as in (20) by **Lemma 1**. To reduce pilot length n_k^{Ω}, a straightforward approach is to let pilot power reach its limit. However, it is important to ensure that this choice of n_k^{Ω} satisfies the constraint $(C6)$. This condition can be verified using step 4, which determines the final value of n. To avoid limiting the pilot power excessively, the algorithm restricts it further only when the pilot length falls below the lower limit. This restriction is implemented in step 5 using the trinomial operator $\mathsf{X} = \mathsf{A}\,?\,\mathsf{B} : \mathsf{C}$, which assigns X to B if the condition expression A is true, and C otherwise. This step also prevents performance loss caused by high $\rho_{\mathrm{pthr},k}^{\Omega}$, as analyzed in Subsection III-B . Subsequently, in step 6, N_k^{Ω} is obtained by **Algorithm 1** by **Lemma 2** . We repeat these steps until the number of iterations reaches $I_{\max}$. Thanks to the closed-form expressions of (20) and (21), the proposed JPLLO algorithm enjoys a low complexity.

The JOMD-IE algorithm for multi-device system is proposed as **Algorithm 3** . This algorithm is based on the results of the single-device block structure optimization problem P1.1 (17), which was previously solved using the JPLLO algorithm (see **Algorithm 2**). The JOMD-IE algorithm functions in two stages. Initially, the control block structure is optimized (steps 2-7) to select the block structure consuming the most transmission resources, thus catering to the highest device demands. Subsequently, the data block structures of the devices are optimized (steps 8-11). Devices scheduled later in the process may encounter a shortage of block lengths to meet their BLEP

Algorithm 3 The Proposed JOMD-IE Algorithm

Input: $N_{\mathrm{Fthr}}, \boldsymbol{\epsilon}_{\mathrm{thr}}, \boldsymbol{\rho}_{\mathrm{pthr}}, \mathbf{L}, \mathbf{M}, \boldsymbol{\gamma}, \mathbf{f}_{\mathrm{D}}, I_{\max}, U, \boldsymbol{\lambda}$;
Execute:
1: Reorder devices by descending $\boldsymbol{\lambda}$ values.
2: $N_{\mathrm{thr,C}} = N_{\mathrm{Fthr}} - \sum_{k=1}^{U_{\mathrm{F}}} L_k / \log_2(M_k) - L_{\mathrm{C}} / \log_2(M_{\mathrm{C}})$;
3: **for** $u = 1$ **to** U **do**
4: Given $N_{\mathrm{thr,C}}, \epsilon_{\mathrm{thr,C}}, \rho_{\mathrm{pthr,C}}, L_{\mathrm{C}}, M_{\mathrm{C}}, \gamma_u, f_{\mathrm{D},u}, I_{\max}$, the optimal parameters of the control block structure for the u-th device are obtained by **Algorithm 2** , *i.e.*, $N_{\mathrm{C},u}$, $n_{\mathrm{C},u}$ and $\rho_{\mathrm{p}_{\mathrm{C},u}}$;
5: **end for**
6: Let $u^* = \arg\max_u \{N_{\mathrm{C},u}\}$;
7: $N_{\mathrm{C}} = N_{\mathrm{C},u^*}, n_{\mathrm{C}} = n_{\mathrm{C},u^*}, \rho_{\mathrm{p_C}} = \rho_{\mathrm{p}_{\mathrm{C},u^*}}$;
8: **for** $k = 1$ **to** U **do**
9: $N_{\mathrm{thr},k} = N_{\mathrm{Fthr}} - N_{\mathrm{thr,C}} - \sum_{i=1}^{k} N_i - \sum_{i=k}^{U_{\mathrm{F}}} L_i / \log_2(M_i)$;
10: Given $N_{\mathrm{thr},k}, \epsilon_{\mathrm{thr},k}, \rho_{\mathrm{pthr},k}, L_k, M_k, \gamma_k, f_{\mathrm{D},k}, I_{\max}$, the optimal parameters of the data block structure for the u-th device are obtained by **Algorithm 2** , *i.e.*, N_k, n_k, ρ_{p_k} ;
11: **end for**
12: **return** $\boldsymbol{N}, \boldsymbol{n}, \boldsymbol{\rho}_{\mathrm{p}}$

constraints. This raises a fairness concern, as these devices are at risk of being disregarded. To address this issue, we introduce a weight variable, denoted as $\boldsymbol{\lambda} = \{\lambda_u | u = 1, \ldots, U\}$, in step 1 to determine the scheduling order of devices. Initially, λ_u is assigned a value within the range of 1 to U, based on the outcomes of existing scheduling algorithms. If device u is previously disregarded due to insufficient block lengths, its new weight for the current schedule will be updated to $\lambda_u' = \lambda_u + U$. This adjustment is designed to prioritize devices that were previously unable to get necessary resources, thereby improving fairness in the allocation process in subsequent rounds of scheduling.

C. Throughput Maximization of the JE Frame Structure

Based on the analysis in Subsection III-A , the optimization problem is formulated by maximizing the throughput of the JE frame structure $\eta_{\mathrm{F}}^{\mathrm{J}}$ with respect to block length $N_{\mathrm{F}}^{\mathrm{J}}$, pilot length n^{J} and the normalized pilot power $\rho_{\mathrm{p}}^{\mathrm{J}}$, as follows:

$$\mathbf{P2:} \quad \max_{N^{\mathrm{J}}, n^{\mathrm{J}}, \rho_{\mathrm{p}}^{\mathrm{J}}} \quad \eta_{\mathrm{F}}^{\mathrm{J}}, \quad \text{s.t.} \quad (C1) \sim (C7) \qquad (23)$$
$$\text{with } \Omega = \mathrm{J}, \quad u = 1, ..., U, \quad k = 1.$$

where $L^{\mathrm{J}} = L_{\mathrm{C}} + \sum_{u=1}^{U} L_u$. The constraints are basically the same as problem P1. Problem P2 has a total number of $5 + 2U$ constraints since there is only a single block in the JE frame structure. Note that the JE frame structure should satisfy the requirement of the most demanding device, *i.e.*, $\rho_{\mathrm{pthr}}^{\mathrm{J}} = \min\{\rho_{\mathrm{pthr},u}\}, M^{\mathrm{J}} = \min\{M_u\}$, since devices' data and control information are encoded together in a single block.

The JE frame structure can be treated as a singular data block, as both devices' data and control information are encoded within it. Therefore, the issue of optimizing the JE frame structure can be efficiently tackled with the help of the JPLLO algorithm. The proposed JOMD-JE algorithm is presented in Algorithm 4. The data block in the JE frame must meet the strictest transmission requirements from all devices

Algorithm 4 The Proposed JOMD-JE Algorithm

Input: $N_{\text{Fthr}}, \boldsymbol{\epsilon}_{\text{thr}}, \boldsymbol{\rho}_{\text{pthr}}, \mathbf{L}, \mathbf{M}, \boldsymbol{\gamma}, \mathbf{f}_{\text{D}}, I_{\max}, U$;

Execute:

1: **for** $u = 1$ **to** U **do**

2: $\quad \epsilon_{\text{thr},u} = \min\{\epsilon_{\text{thr},u}, \epsilon_{\text{thr,C}}\}$;

3: $\quad$ Given $N_{\text{Fthr}}, \epsilon_{\text{thr},u}, \rho^{\text{J}}_{\text{pthr}}, L^{\text{J}}, M^{\text{J}}, \gamma_u, f_{\text{D},u}, I_{\max}$, the optimal JE frame structure parameters for the u-th device are obtained by **Algorithm 2**, *i.e.*, N_u, n_u, ρ_u;

4: **end for**

5: Let $u^* = \arg\max_u\{N_u\}$

6: **return** N_{u^*}, n_{u^*}, ρ_{p,u^*}

and control information. Thus, the constraint of a lower BLEP between the devices' and control information is used for further calculation, as shown in step 2. The proposed JPLLO algorithm is then deployed to independently determine optimal frame structures that satisfy the constraints of each device. The structure with the longest frame length is determined as the final result, thus ensuring the fulfillment of transmission requirements for all devices, as described in step 5.

V. Latency and Throughput Comparison between the IE and JE Frame Structures

The throughput analysis between the IE and JE frame structures in homogeneous and heterogeneous scenarios is presented in Subsection V-A. In Subsection V-B, we analyze the latency of the IE frame structure in general heterogeneous scenarios. Subsection V-C compares the average latency between the two frame structures. The complexity analysis of the proposed algorithms is shown in V-D.

A. Throughput Comparison

Let β denote the frame length ratio between the JE and IE frame structures while transmitting an equivalent amount of information, *i.e.*, $N^{\text{J}} = \beta N^{\text{I}}_{\text{F}}$.

We use $\mathcal{G}$ to denote the throughput gain of the JE frame structure over the IE frame structure during transmission of an equivalent amount of information. We have $\mathcal{G} = 1/\beta$ as given in equation (7), when the negligible influence of BLEP is disregarded. As β decreases, the relative frame length of the JE frame structure to the IE frame structure becomes smaller, and correspondingly, the throughput gain $\mathcal{G}$ increases. For facilitating the derivation, we examine the monotonicity of β. The examination enables us to discover how to improve the throughput gain of the JE frame structure.

1) Homogeneous Scenarios: Homogeneous scenarios are the extreme case of low heterogeneous scenarios. In homogeneous scenarios, we have $L_u = L_1$ and $N^{\text{I}}_u = N^{\text{I}}_1$ ($u = 1, \ldots, U$). Thus, $N^{\text{I}}_{\text{F}} = N^{\text{I}}_{\text{C}} + UN^{\text{I}}_1$. And we denote $L^{\text{J}} = L_1 U + L_{\text{C}}$ for the sake of brevity.

Proposition 5: In homogeneous scenarios, the ratio of the JE to IE frame length β can be approximately derived in (24). β is mono-increasing with respect to channel capacity, and the amount of information bits L_u, respectively, and is mono-decreasing with respect to the number of devices U.

$$\beta \approx \widetilde{N}^{\text{J}} / \left(\widetilde{N}^{\text{I}}_{\text{C}} + UN^{\text{I}}_1\right), \tag{24}$$

where $\widetilde{N}^{\text{I}}_{\text{C}} = \frac{1}{\left(1-\alpha^{\text{I}}_1\right)}\left(\frac{A^{\text{I}}_{q,1}+A^{\text{I}}_{\text{C}}}{2\left(C\left(\gamma^{\text{I}}_{\text{e},1}\right)\right)^2} + \frac{L_{\text{C}}}{C\left(\gamma^{\text{I}}_{\text{e},1}\right)}\right)$, $\widetilde{N}^{\text{J}} = \frac{1}{\left(1-\alpha^{\text{I}}_1\right)}\left(\frac{A^{\text{I}}_{q,1}+A^{\text{J}}}{2\left(C\left(\gamma^{\text{I}}_{\text{e},1}\right)\right)^2} + \frac{L^{\text{J}}}{C\left(\gamma^{\text{I}}_{\text{e},1}\right)}\right)$, $A^{\text{J}} = \sqrt{A^{\text{I}}_{q,1} + 4C\left(\gamma^{\text{I}}_{\text{e},1}\right)L^{\text{J}}}\sqrt{A^{\text{I}}_{q,1}}$, $A^{\text{I}}_{\text{C}} = \sqrt{A^{\text{I}}_{q,1} + 4C\left(\gamma^{\text{I}}_{\text{e},1}\right)L_{\text{C}}}\sqrt{A^{\text{I}}_{q,1}}$ at low to moderate SNRs; $\widetilde{N}^{\text{I}}_{\text{C}} = \frac{1}{\left(1-\alpha^{\text{I}}_1\right)}\frac{L_{\text{C}}}{\log_2(M_{\text{C}})}$, $\widetilde{N}^{\text{J}} = \frac{1}{\left(1-\alpha^{\text{J}}_{\min}\right)}\frac{L^{\text{J}}}{\log_2(M_1)}$ at high SNR. $N^{\text{I}}_1, \alpha^{\text{I}}_1, \gamma^{\text{I}}_{\text{e},1}, A^{\text{I}}_{q,1}$ are obtained from the optimal block structure of the first device, by the proposed JPLLO algorithm with equations (20) and (22).

Proof: See Appendix D. ∎

The analysis demonstrates that the throughput of the JE frame structure outperforms that of the IE frame structure. In addition, the throughput gain of the JE frame structure $\mathcal{G}$ increases when the channel capacity decreases, the number of information bits in a block decreases, or the number of homogeneous devices increases. According to the FBL theory [11], a smaller L results in a wider channel dispersion, which yields a lower achievable rate. The JE frame structure accommodates a larger number of information bits in a block compared to the IE frame structure, resulting in higher throughput. Moreover, the JE frame structure allows devices to share redundant and pilot symbols, whereas the IE frame structure requires independent redundant and pilot symbols for each device. Consequently, the JE frame structure decreases the quantity of redundant and pilot symbols, enjoying a shorter frame length and higher throughput than the IE frame structure, particularly at low average SNR. However, as the average SNR increases, the need for redundant symbols decreases, leaving more space for data and pilot symbols in the frame. In such case, the performance gain from the JE frame structure, through shared redundancy, is not as pronounced. The validity of this analysis is verified in Figs. 5, 9 and 12.

2) Heterogeneous Scenarios: In heterogeneous scenarios, accurate estimation of $\widetilde{N}^{\text{I}}_{\text{C}}$ and $\widetilde{N}^{\text{J}}$ based on a particular device block becomes challenging. This is due to the significant variation in the effective SNR $\gamma^{\Omega}_{\text{e},u}$ across different devices. For the JE frame structure, devices with high $\gamma^{\Omega}_{\text{e},u}$ must use a large number of resources to be compatible with devices sustaining low $\gamma^{\Omega}_{\text{e},u}$. Under such circumstances, the JE frame structure requires a higher number of pilot and redundant symbols than that in homogeneous scenarios, given that the total information in a frame remains constant. This may result in a reduction in the throughput gain of the JE frame structure, and in some instances, it may even lead to inferior performance compared to the IE frame structure, *i.e.*, $\mathcal{G} < 1$. For the IE frame structure, the data of each device is encoded into blocks independently. The heterogeneity only affects the length of the control block, with minimal impact on the total throughput. As a result, the throughput in heterogeneous scenarios remains virtually identical to that in homogeneous scenarios, given that the total information in a frame remains constant. The analysis is verified in Fig. 5.

B. Latency Analysis for the IE Frame Structure

For the IE frame structure, the transmission latency of the k-th block is N^{I}_k, and the queuing latency of the k-th block is $N^{\text{I}}_{\text{C}} + \sum_{m=1}^{k-1} N_m$. Thus, Let $T^{\text{I}}_u = N^{\text{I}}_{\text{C}} + \sum_{k=1}^{u} N^{\text{I}}_k$ denote

the overall latency of the u-th device in the IE frame. Let $\overline{T^{\Omega}}$ denote the average latency of U devices. The average latency of the IE frame structure under heterogeneous scenarios is derived as:

$$\overline{T^{\mathrm{I}}_{\mathrm{Het}}}=\frac{1}{U}T^{\mathrm{I}}_{u}=N^{\mathrm{I}}_{\mathrm{C}}+\frac{1}{U}\sum_{k=1}^{U}(U-k+1)\,N^{\mathrm{I}}_{k}=N^{\mathrm{I}}_{\mathrm{F}}-\frac{1}{U}\sum_{k=1}^{U}(k-1)\,N^{\mathrm{I}}_{k}. \tag{25}$$

Proposition 6: The SBF scheduling rule achieves the minimum average latency for the IE frame structure.

Proof: We prove the **Proposition 6** by contradiction. Suppose a non-SBF schedule $\mathcal{S}$ is optimal. Suppose two adjacent blocks i, j with $i = j-1$ and $N^{\mathrm{I}}_i > N^{\mathrm{I}}_j$ in $\mathcal{S}$. Perform an adjacent pairwise interchange on the two blocks. All other blocks remain in their original positions. Call the new schedule $\mathcal{S}'$. Let $\overline{T^{\mathrm{I}}_{\mathcal{S}}}$ and $\overline{T^{\mathrm{I}}_{\mathcal{S}'}}$ denote the average latencies of $\mathcal{S}$ and $\mathcal{S}'$, respectively. We have $\overline{T^{\mathrm{I}}_{\mathcal{S}}} = \mathring{T}^{\mathrm{I}} - (i-1)N^{\mathrm{I}}_i - (j-1)N^{\mathrm{I}}_j$ and $\overline{T^{\mathrm{I}}_{\mathcal{S}'}} = \mathring{T}^{\mathrm{I}} - (i-1)N^{\mathrm{I}}_j - (j-1)N^{\mathrm{I}}_i$ by (25), where $\mathring{T}^{\mathrm{I}} = N^{\mathrm{I}}_{\mathrm{F}} - \frac{1}{U}\sum_{k=1,k\neq i,j}^{U}(k-1)\,N^{\mathrm{I}}_k$. Thus, $\overline{T^{\mathrm{I}}_{\mathcal{S}}} - \overline{T^{\mathrm{I}}_{\mathcal{S}'}} = N^{\mathrm{I}}_i - N^{\mathrm{I}}_j > 0$. That is, the average latency of $\mathcal{S}$ is higher than that of $\mathcal{S}'$. This contradicts the optimality of $\mathcal{S}$ and completes the proof. ■

Proposition 7: In the IE frame structure, the average latency decreases as the device heterogeneity increases under the SBF scheduling, when the total frame length remains unchanged.

Proof: The homogeneous scenario is the extreme case of heterogeneous scenario. With the same frame length $N^{\mathrm{I}}_{\mathrm{F}}$ as the heterogeneous scenario, the average latency in the homogeneous scenario is derived as:

$$\overline{T^{\mathrm{I}}_{\mathrm{Homo}}} = N^{\mathrm{I}}_{\mathrm{F}} - \frac{1}{U}\sum_{k=1}^{U}(k-1)\,\frac{N^{\mathrm{I}}_{\mathrm{F}}-N^{\mathrm{I}}_{\mathrm{C}}}{U} = N^{\mathrm{I}}_{\mathrm{F}} - \frac{1+U}{2U}\sum_{k=1}^{U}N^{\mathrm{I}}_{k}. \tag{26}$$

Thus, $\overline{T^{\mathrm{I}}_{\mathrm{Het}}} - \overline{T^{\mathrm{I}}_{\mathrm{Homo}}} = \frac{1}{U}\left(\frac{1+U}{2}\sum_{k=1}^{U}N^{\mathrm{I}}_{k} - \sum_{k=1}^{U}kN^{\mathrm{I}}_{k}\right)$.

When U is odd,

$$\overline{T^{\mathrm{I}}_{\mathrm{Het}}} - \overline{T^{\mathrm{I}}_{\mathrm{Homo}}} = \frac{1}{U}\left(\frac{1+U}{2}\sum_{k=1}^{U}N^{\mathrm{I}}_{k} - \sum_{k=1}^{U}kN^{\mathrm{I}}_{k}\right)$$

$$= \frac{1}{U}\left[\frac{1+U}{2}\sum_{k=1}^{U}N^{\mathrm{I}}_{k} - \left(\sum_{k=1}^{(U-1)/2}kN^{\mathrm{I}}_{k} + \frac{1+U}{2}N^{\mathrm{I}}_{\frac{1+U}{2}} + \sum_{k=(U+3)/2}^{U}kN^{\mathrm{I}}_{k}\right)\right]$$

$$\overset{k'=U-k+1}{=} \frac{1}{U}\left[\sum_{k=1}^{(U-1)/2}\left(\frac{1+U}{2}-k\right)N^{\mathrm{I}}_{k} - \sum_{k'=1}^{(U-1)/2}\left(\frac{1+U}{2}-k'\right)N^{\mathrm{I}}_{U-k'+1}\right]$$

$$= \frac{1}{U}\left[\sum_{k=1}^{(U-1)/2}\left(\frac{1+U}{2}-k\right)\left(N^{\mathrm{I}}_{k} - N^{\mathrm{I}}_{U-k+1}\right)\right]. \tag{27}$$

Similarly, when U is even,

$$\overline{T^{\mathrm{I}}_{\mathrm{Het}}} - \overline{T^{\mathrm{I}}_{\mathrm{Homo}}} = \frac{1}{U}\left[\frac{1+U}{2}\sum_{k=1}^{U}N^{\mathrm{I}}_{k} - \left(\sum_{k=1}^{U/2}kN^{\mathrm{I}}_{k} + \sum_{k=U/2+1}^{U}kN^{\mathrm{I}}_{k}\right)\right]$$

$$= \frac{1}{U}\left[\sum_{k=1}^{U/2}\left(\frac{1+U}{2}-k\right)\left(N^{\mathrm{I}}_{k} - N^{\mathrm{I}}_{U-k+1}\right)\right]. \tag{28}$$

Under the SBF scheduling, $N^{\mathrm{I}}_k - N^{\mathrm{I}}_{U-k+1} \leqslant 0$ since $N^{\mathrm{I}}_1 \leq N^{\mathrm{I}}_2 \leqslant \cdots \leqslant N^{\mathrm{I}}_U$. Thus, $\overline{T^{\mathrm{I}}_{\mathrm{Het}}} \leqslant \overline{T^{\mathrm{I}}_{\mathrm{Homo}}}$. The $N^{\mathrm{I}}_k - N^{\mathrm{I}}_{U-k+1}$ is greater when there is strong heterogeneity, therefore the average latency is lower. ■

Proposition 6 and **7** suggest that, in heterogeneous scenarios, using SBF scheduling for the IE frame structure can help the data transmitted to devices in a timely manner.

C. Average Latency Comparison

For the JE frame structure, the overall latency of each device is the transmission latency of the frame, *i.e.*, $T^{\mathrm{J}}_u = N^{\mathrm{J}}$. Thus, the average latency of the JE frame structure is given as:

$$\overline{T^{\mathrm{J}}} = N^{\mathrm{J}} = \beta N^{\mathrm{I}}_{\mathrm{F}}. \tag{29}$$

Proposition 8: The average latency of the JE frame structure is lower than that of the IE frame structure only when β is less than a certain threshold, denoted as β^{D} (as shown in (30)), *i.e.*, $\beta < \beta^{\mathrm{D}}$. Moreover, $\beta^{\mathrm{D}} \leq 1$ and is mono-decreasing with respect to the number of devices U in homogeneous scenarios.

$$\beta^{\mathrm{D}} = \begin{cases} \beta^{\mathrm{D}}_{\mathrm{Het}} := 1 - \frac{1}{UN^{\mathrm{I}}_{\mathrm{F}}}\sum_{k=1}^{U}(k-1)\,N^{\mathrm{I}}_{k}, & \text{Heterogeneous,} \\ \beta^{\mathrm{D}}_{\mathrm{Homo}} := \frac{N^{\mathrm{I}}_{\mathrm{C}}+(1+U)N^{\mathrm{I}}_{1}/2}{N^{\mathrm{I}}_{\mathrm{C}}+UN^{\mathrm{I}}_{1}}, & \text{Homogeneous.} \end{cases} \tag{30}$$

Proof: From (25) (26) and (29), β^{D} can be derived by solving $\overline{T^{\mathrm{J}}} \leqslant \overline{T^{\mathrm{I}}}$. For heterogeneous scenarios, it is obvious that $\beta^{\mathrm{D}} \leq 1$. For homogeneous scenarios, $\beta^{\mathrm{D}} = 1$ with $U = 1$, and $\partial\beta^{\mathrm{D}}_{\mathrm{Homo}}/\partial U = -N^{\mathrm{I}}_1(N^{\mathrm{I}}_{\mathrm{C}} + N^{\mathrm{I}}_1)/(2(N^{\mathrm{I}}_{\mathrm{C}} + UN^{\mathrm{I}}_1)^2) < 0$. Thus, $\beta^{\mathrm{D}} \leq 1$ and is mono-decreasing with respect to U. ■

Remark: Essentially, β^{D} is the point at which the average latencies of the JE and IE frame structures are equal. We define the corresponding throughput gain of the JE structure at this point as $\mathcal{G}^{\mathrm{D}}$, where $\mathcal{G}^{\mathrm{D}} = 1/\beta^{\mathrm{D}}$. This means if $\beta < \beta^{\mathrm{D}}$, and correspondingly, $\mathcal{G} > \mathcal{G}^{\mathrm{D}}$, the JE frame structure has a lower average latency than the IE frame structure, and vice versa.

Considering the intra-frame queuing latency from the perspective of devices, the average latency of each device within the JE frame structure is not always lower than that of the IE frame structure. This differs from the discussion of [9], which just considers the transmission latency. From the perspective of the frame's transmission latency, as long as the frame length of the JE frame structure $N^{\mathrm{J}}_{\mathrm{F}}$ is shorter than that of the IE frame structure $N^{\mathrm{I}}_{\mathrm{F}}$, *i.e.*, $\beta < 1$ ($\mathcal{G} > 1$), the transmission latency of the JE frame structure is lower than the IE frame structure. However, from the perspective of the average latency of each device within the frame, the average latency of the JE frame structure is lower only when $\beta < \beta^{\mathrm{D}}$. In such case, *i.e.*, $\mathcal{G} > \mathcal{G}^{\mathrm{D}} > 1$, corresponding to scenarios with low average SNR or/and a small L, the JE frame structure demonstrates superior performance in both throughput and average latency compared to the IE frame structure. Generally, the average latency of the JE frame structure tends to be higher than that of the IE frame structure, especially in heterogeneous scenarios, *i.e.*, $\beta > \beta^{\mathrm{D}}$ ($\mathcal{G} < \mathcal{G}^{\mathrm{D}}$). As heterogeneity increases, β rises due to a larger increment in $N^{\mathrm{J}}_{\mathrm{F}}$ compared to $N^{\mathrm{I}}_{\mathrm{F}}$, as explained in Subsection V-A . Concurrently, β^{D} decreases due to the beneficial effects of using the proposed SBF scheduling in reducing $\overline{T^{\mathrm{I}}}$, and due to the increased $\overline{T^{\mathrm{J}}}$ caused by an expanding total frame length, as discussed in Subsection V-B . This makes the condition $\beta < \beta^{\mathrm{D}}$ ($\mathcal{G} > \mathcal{G}^{\mathrm{D}}$) less likely to hold. In general, compared to the IE frame structure, the JE frame structure provides higher throughput but higher average latency. The illustration is shown in Fig. 9 .

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON OF FRAME STRUCTURE OPTIMIZATION ALGORITHMS

Algorithm	Analytical Computational Complexity	Normalized Numerical Computational Complexity	
		$U=2$	$U=3$
JOMD-IE	$2UI_{\mathrm{A}}$	2	3
JOMD-JE	UI_{A}	1	1.5
ES with IE	$\prod_{u=1}^{U} m_{\rho,u}^{\mathrm{I}}(m_{N,u}^{\mathrm{I}})^2/2$	4.2×10^{8}	1.4×10^{13}
ES with JE	$m_{\rho}^{\mathrm{J}}(m_{N}^{\mathrm{J}})^2/2$	3.3×10^{4}	5×10^{5}
D-IPBO [19]	$I_{\text{D-I}}I_{\mathrm{D}}U^4$	10	36

Notations: U: number of devices; I_{A}: the number of iterations to achieve a steady status of the proposed JPLLO algorithm; $m_{\rho,u}^{\mathrm{I}}$: the number of searching points of normalized pilot power for the u-th device within the IE frame structure; m_{ρ}^{J}: the number of searching points of normalized pilot power of the JE frame structure; $m_{N,u}^{\mathrm{I}}$: the feasible points of block lengths for the u-th device within the IE frame structure; m_{N}^{J}: feasible points of the JE frame length; $I_{\text{D-I}}$: the number of iterations of the D-IPBO algorithm; I_{D}: the number of iterations of Dinkelbach's algorithm. ES stands for exhaustive search method.

D. Complexity Analysis

The computational complexity of the proposed JOMD-IE and JOMD-JE algorithms are shown in Table II, in terms of the number of calculations, in comparison to the dynamic joint independent-pilot length and block length optimization (D-IPBO) algorithm [19]. The proposed JPLLO algorithm is performed by the derived closed-form expressions for the near-optimal block length (22) and total pilot power (20), at complexity of I_{A}, where I_{A} is the number of iterations to achieve a steady status of the proposed JPLLO. Thus, the complexities of the JOMD-IE and JOMD-JE algorithm are $2UI_{\mathrm{A}}$ and UI_{A}, respectively. The exhaustive search (ES) is used as a benchmark. The complexities for the IE and JE frame structures with the exhaustive search are $\prod_{u=1}^{U} m_{\rho,u}^{\mathrm{I}}(m_{N,u}^{\mathrm{I}})^2/2$ and $m_{\rho}^{\mathrm{J}}(m_{N}^{\mathrm{J}})^2/2$ respectively, where $m_{\rho,u}^{\mathrm{I}}$ and m_{ρ}^{J} are the number of searching points of $\rho_{\mathrm{p},u}^{\mathrm{I}}$ and $\rho_{\mathrm{p}}^{\mathrm{J}}$ respectively, $m_{N,u}^{\mathrm{I}}$ and m_{N}^{J} are the feasible points of block lengths N_k^{I} and N^{J} respectively. The complexity of the D-IPBO algorithm [19] designed for the IE frame structure is $I_{\text{D-I}}I_{\mathrm{D}}U^4$ [19], where $I_{\text{D-I}}$ is the number of iterations of the D-IPBO algorithm [19], and I_{D} is the number of iterations of Dinkelbach's algorithm. Also, the D-IPBO algorithm [19] does not take pilot power into consideration.

With $I_{\mathrm{A}}=3$, $m_{\rho,u}^{\mathrm{I}}=m_{\rho}^{\mathrm{J}}=10$, $m_{N,u}^{\mathrm{I}}=100$, $m_{N}^{\mathrm{J}}=m_{N,u}^{\mathrm{I}}\cdot U$, $I_{\text{D-I}}=2$ and $I_{\mathrm{D}}=2$, the normalized numerical complexities are shown in Table II. The complexities of the proposed algorithms grow linearly with the number of devices U, but the complexities of the exhaustive search methods grow exponentially. Considering one more group of optimization variables, *i.e.*, pilot power, our proposed algorithms still achieve greater than 10-fold complexity reduction over the D-IPBO algorithm [19]. Therefore, the proposed JOMD-IE and JOMD-JE algorithms reduce complexity significantly.

VI. SIMULATION RESULTS

A. Simulation Setup

In this section, we conduct numerical simulations to verify the theoretical analysis and the proposed algorithms. We set the symbol period $T_{\mathrm{s}}=1/B=10\,\mu$s with the bandwidth $B=100$ kHz [36]. The frame structure parameters are $N_{\mathrm{Id}}=1, c_{\mathrm{U}}=3, c_{\mathrm{M}}=2, c_{\mathrm{A}}=8$. The maximum frame length is set to $N_{\mathrm{Fthr}}=1000$. The homogeneous scenario settings in Figs. 5 and 9 are as follows: The information bits of each device is $L_u=100$ bits. The modulation order is $M_u=4$. The BLEP threshold of device data is $\epsilon_{\mathrm{thr},u}=10^{-7}$. The BLEP threshold of control information is $\epsilon_{\mathrm{thr,C}}=10^{-7}$. The normalized pilot power threshold is $\rho_{\mathrm{pthr},u}=5$ dB. The normalized Doppler frequency is $f_{\mathrm{D},u}=0.02$. The number of device number is $U=3$ in Fig. 5. The heterogeneous scenario settings in Figs. 5, 6, 7 and 12 are as follows: $U=3, \mathbf{L}=[100,50,150], \mathbf{M}=[4,4,4], \boldsymbol{\epsilon}_{\mathrm{thr}}=[10^{-7},10^{-6},10^{-8}], \epsilon_{\mathrm{thr,C}}=10^{-8}$, $\boldsymbol{\rho}_{\mathrm{pthr}}=[3,1,5]$ dB, and $\mathbf{f}_{\mathrm{D}}=[0.02,0.01,0.02]$. The simulation parameters in Figs. 10 and 11 are set as: $U=1, L=200, M=4, f_{\mathrm{D}}=0.02$, and $\epsilon_{\mathrm{thr}}=10^{-6}$.

The results displayed in Fig. 7 and Fig. 8 are a culmination of an analysis conducted on one million samples. The samples' parameters vary within the following ranges: U spans from 2 to 6. L_u is from 50 to 350 bits. M remains constant at 4. $\epsilon_{\mathrm{thr},u}$ varies from 10^{-8} to 10^{-5}. $\rho_{\mathrm{thr},u}$ ranges from 0 to 7 dB. $f_{\mathrm{D},u}$ extends from 0.001 to 0.03. γ_u varies between 2 to 12 dB. Let $\sigma_X^*=\sigma_X/\overline{X}$ denote the normalized standard deviation of a random variable X, where $\overline{X}$ denotes the mean of X, σ_X denotes the standard deviation of X. In Fig. 7, we set the normalized standard deviation of SNR, denoted as σ_γ^*, to 0.2 for different devices within each sample frame. In Fig. 8, the frame length N_{F} is set to 500.

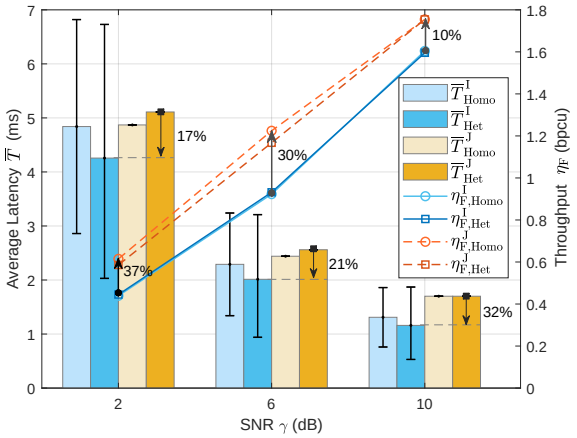


Fig. 5. Average latency and throughput between IE and JE frame structures in homogeneous and heterogeneous scenarios with the same amount of transmitted information $\sum_{u=1}^{U}L_u=300$ bits.

B. Latency and Throughput Performance Comparison

We define the relative average latency reduction of the IE frame structure over the JE frame structure as $\Delta\overline{T}=\frac{\overline{T}^{\mathrm{J}}-\overline{T}^{\mathrm{I}}}{\overline{T}^{\mathrm{J}}}$, and define the throughput improvement of the JE frame structure over the IE frame structure as $\Delta\eta_{\mathrm{F}}=\frac{\eta_{\mathrm{F}}^{\mathrm{J}}-\eta_{\mathrm{F}}^{\mathrm{I}}}{\eta_{\mathrm{F}}^{\mathrm{I}}}=\mathcal{G}-1$.

Fig. 5 presents a comparison of latency and throughput between IE and JE frame structures, with the same amount of transmitted information but within different scenarios. Compared to the IE frame structure with the proposed SBF scheduling, the JE frame structure provides a higher throughput but also a higher average latency, which aligns with the analyses conducted in Subsection V-A and Subsection

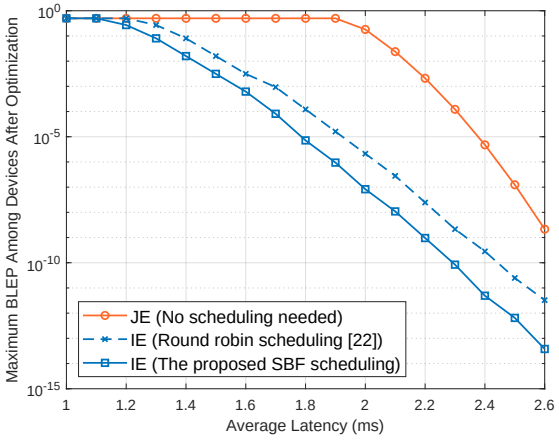


Fig. 6. Maximum BLEP among devices through frame structure optimizing, given average latencies, in a heterogeneous scenario transmitting identical information at average SNR $= 6$ dB.

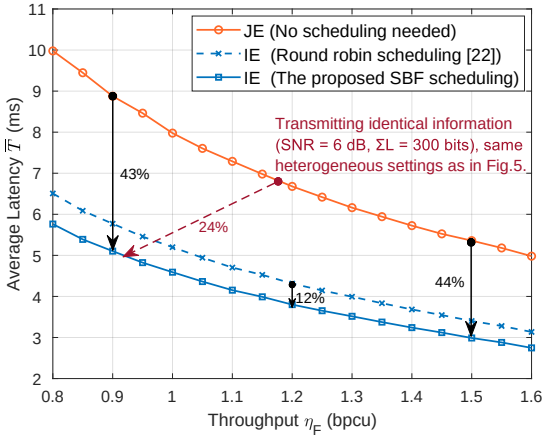


Fig. 7. Average latency at different throughputs derived from 1 million samples, with the normalized standard deviation of SNR $\sigma_\gamma^* = 0.2$.

V-C. Specifically, the JE frame structure achieves superior throughput ($\Delta\eta_F$ is up to more than 35%) due to reduced channel dispersion with a shorter frame, while the IE frame structure offers significantly lower average latency ($\Delta\overline{T}$ is up to more than 30%) because of decreased queuing latency. Notably, $\Delta\eta_F$ declines gradually as the SNR increases, and $\Delta\overline{T}$ rises with increasing SNR. This behavior is attributed to the JE frame structure's ability to reduce redundant and pilot symbols, leading to a shorter frame length and consequently enhanced throughput, especially at lower average SNRs, as analyzed in **Proposition 5**. Correspondingly, $\Delta\overline{T}$ is less pronounced at low SNRs.

With regard to the IE frame structure, the average latency is lower in heterogeneous scenarios than in homogeneous scenarios, thanks to SBF scheduling, which allows devices with shorter block lengths to experience lower queuing latency, in agreement with **Proposition 7**. In heterogeneous scenarios, due to the JE frame structure being designed for the device with the poorest channel (often termed the "cask effect"), the frame length increases. As a result, throughput decreases and average latency rises, corroborating the analysis in Subsection V-A. The scheduling fairness of the IE and JE frame structures is shown by error-bars. The JE frame structure maintains fairness across all devices but incurs a higher average latency. In contrast, the IE frame structure offers lower latency to devices with shorter block lengths, while it causes a higher latency on devices with larger block lengths.

Fig. 6 illustrates the variations in BLEP under different transmission schemes given average latencies. The worst BLEP among devices is used for evaluation. The IE frame structure excels over the JE frame structure by providing superior BLEP performance at the same average latency. This is attributed to the longer frame length of the IE frame structure under equivalent average latency, indicating that the IE frame structure trades off throughput for reliability and low latency. Furthermore, when employing the IE frame structure, the proposed SBF scheduling offers a lower average latency than round robin scheduling [22] at the same BLEP. This can be explained by the fact that, for a given frame length,

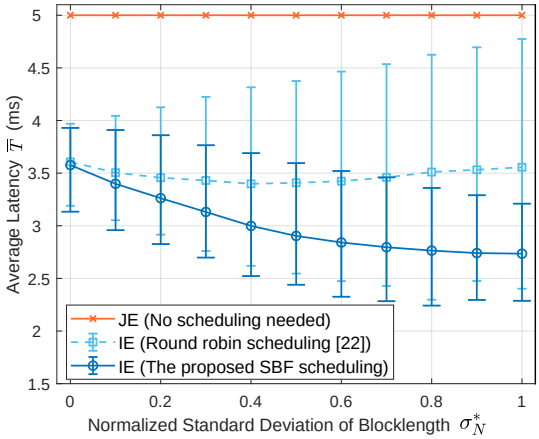


Fig. 8. Impact of heterogeneity of the proposed SBF scheduling on average latency with the fixed frame length $N_F = 500$ symbols, derived from 1 million samples.

the proposed SBF scheduling achieves the minimum average latency for the IE frame structure, proved as **Proposition 6** in Subsection V-B.

Fig. 7 presents the average latency at different throughputs, with data points representing mean values derived from corresponding throughput samples. Compared to the JE frame structure, the IE frame structure significantly reduces average latency by over 40% at equivalent throughputs. This can be attributed to IE having a lower queuing latency for each user within the frame. Furthermore, thanks to the SBF allowing devices with shorter block lengths to experience lower queuing latency, the proposed SBF scheduling rule decreases average latency about 12% more than the round robin scheduling [22], verifying **Proposition 6** in Subsection V-B.

Fig. 8 shows the impact of heterogeneity of the proposed SBF scheduling on average latency at the fixed frame length. As heterogeneity increases, the normalized standard deviation of block length σ_N^* also increases. Compared to round-robin scheduling [22], the proposed SBF scheduling lowers the statistical average, minimum, and maximum values of average latency, as indicated by the error bars. Moreover, as heterogeneity increases, the average latency of the proposed

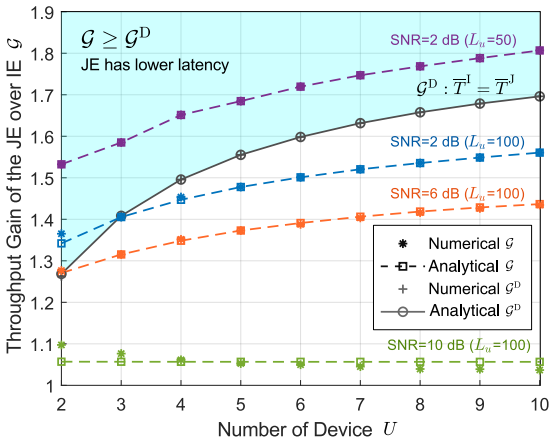


Fig. 9. Impact of various parameters (SNR, device information quantity L_u, number of users U) on the throughput gain of the JE over the IE frame structure in homogeneous scenarios.

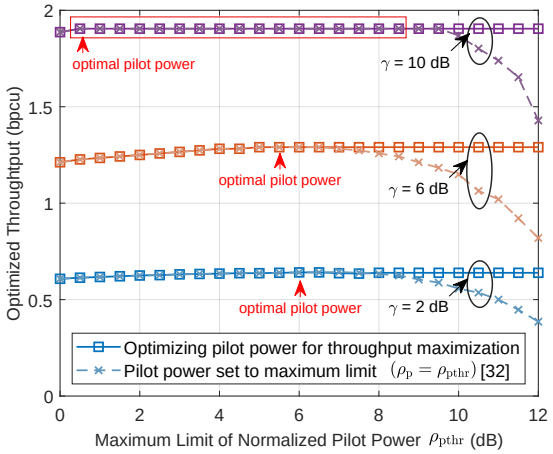


Fig. 10. Impact of pilot power on optimal throughput under two pilot power allocation schemes in a single block, with information quantity $L = 200$ bits and BLEP limit $\epsilon_{\mathrm{thr}} = 10^{-6}$.

SBF scheduling decreases and the scheduling gain increases, verifying **Proposition 7**. Notice that the difference in error-bar lengths between the round-robin and the proposed SBF schedules widens as heterogeneity increases, indicating that the proposed SBF scheduling provides a stable average latency. This can be attributed to the fact that SBF scheduling consistently schedules users with the shortest block lengths at the beginning of the frame, while the round-robin scheduling [22] does not consider block length in its order.

Fig. 9 illustrates the variation of throughput gain of the JE frame structure $\mathcal{G}$ with different parameters of SNR, user information quantity, and the number of users. The $\mathcal{G}$ values under different parameters are presented by colored dashed lines and markers. And $\mathcal{G}^{\mathrm{D}}$, the throughput gain of JE when the average latency of IE frame structure equals to that of the JE frame structures, is presented by a gray solid line and markers. For clarity, only one line representing $\mathcal{G}^{\mathrm{D}}$ is drawn, as the values of $\mathcal{G}^{\mathrm{D}}$ nearly overlap at different SNRs. The analytical values of $\mathcal{G}$ and $\mathcal{G}^{\mathrm{D}}$, derived from **Propositions 5** and **8**, respectively, correspond closely with their numerical results. The area where $\mathcal{G} \geq \mathcal{G}^{\mathrm{D}}$ is filled in blue. When $\mathcal{G}$ falls in this region, the JE frame structure outperforms the IE frame structure in both throughput and average latency. This advantage is particularly apparent when the SNR and L_u values are small. This is because, in these settings, the JE frame structure mitigates the adverse effects of short packet channel dispersion on throughput. In most cases, $\mathcal{G} > 1$ but $\mathcal{G} < \mathcal{G}^{\mathrm{D}}$, suggesting that the JE frame structure yields higher throughput but also higher average latency compared to the IE frame structure. Additionally, $\mathcal{G}$ increases with U in scenarios of low to moderate SNRs. This suggests that employing the JE frame structure, leveraging the high throughput advantage of long packets, is more beneficial in scenarios with a large number of devices. The results in Fig. 9 align with the analyses presented in Subsections V-A and V-C.

C. Frame Structure Optimization

1) Impact of Pilot Power: The impacts of pilot power on throughput and pilot overhead are depicted in Figs. 10 and 11.

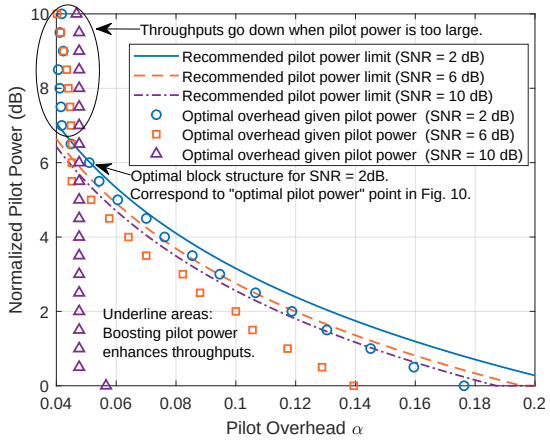


Fig. 11. Trade-off between pilot overhead and pilot power for maximum throughput in a single block at different SNRs γ, with information quantity $L = 200$ bits and BLEP limit $\epsilon_{\mathrm{thr}} = 10^{-6}$.

Fig. 10 illustrates the impact of pilot power on optimal throughput. When using the proposed JPLLO algorithm, the throughput maximizes regardless of the pilot power limit, as shown in solid lines. However, unoptimized pilot power, as in [32], leads to an initial increase in throughput followed by a sharp decline. Fig. 10 also highlights the need for higher pilot power at lower SNRs to achieve peak throughput. Therefore, it is significant to optimize pilot power in order to maximize the system throughput.

Fig. 11 presents the trade-off between pilot overhead and power. The curves labeled "recommended pilot power limit" provide suggested pilot power values for certain pilot overheads, as derived from **Proposition 1**. Overshooting these values risks throughput reduction, while staying below can enhance throughput. The markers, labeled as "optimal overhead given pilot power", correspond to data from Fig. 10 labeled "Pilot power set to a maximum limit". While Fig. 10 focuses on throughput as a function of given pilot power, Fig. 11 provides the optimal pilot overhead associated with those pilot power values. As pilot power increases, the overhead decreases towards $\alpha_{\min}$ as (6). In summary, boosting pilot power can reduce the overhead and improve throughput, but over-boosting may degrade the throughput. The intersection

TABLE III
CONCLUSION OF PROPOSITIONS

Propositions	Verification	Main Findings
1 and 2	Figs. 10 and 11	Boosting pilot power properly can reduce pilot overhead and increase throughput.
3 and 4	Fig. 12	Throughput is concave *w.r.t.* total pilot power and mono-decreasing *w.r.t.* block length.
5 and 8	Figs. 5, 9 and 12	The JE frame structure outperforms the IE frame structure in terms of throughput and average latency in low average SNRs or/and small information sizes. Otherwise, the average latency is lower with the IE frame structure.
6 and 7	Figs. 6, 7 and 8	The SBF scheduling rule minimizes average latency in the IE frame structure, and under this rule, the average latency decreases with increased device heterogeneity.

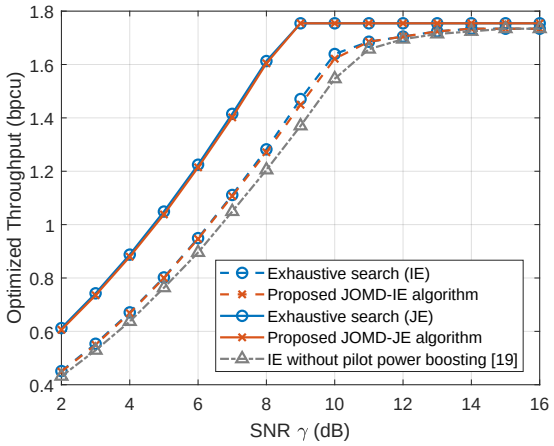


Fig. 12. Throughput performance of the proposed JOMD-IE and JOMD-JE algorithms in the heterogeneous scenario as Fig. 5, comparing with other algorithms.

points between the curves and markers in Fig. 11, labeled as "Optimal block structure", align with the "optimal pilot power" markers in Fig. 10, indicating a balance between pilot power and pilot overhead.

In summary, optimizing pilot power is crucial for maximizing throughput, especially in low to medium SNR scenarios. Our results, as supported by the figures, confirm the effectiveness of the proposed JPLLO algorithm and the validity of our analysis in Subsection III-B.

2) Multi-Device Frame Optimization: In Fig. 12, we use the global optimal results obtained by the exhaustive search as our benchmark. The throughputs yielded by our proposed JOMD-IE and JOMD-JE algorithms closely match the optimal results of the IE and JE frame structures, respectively. As analyzed in the previous Subsection V-A, $\Delta\eta_{\mathrm{F}}$ decreases as SNR increases. By employing pilot power boosting, our proposed JOMD-IE and JOMD-JE algorithms outperform the D-IPBO algorithm [19] with constant pilot power, in throughput performance.

VII. CONCLUSION

In this paper, we intensively investigated the IE and JE frame structures for heterogeneous URLLC systems in the FBL regime over continuous fading channels. Table III summarizes the main findings of propositions 1-7 in this paper. We derive the closed-form expressions for the average latency gain of the IE frame structure over the JE frame structure, and the throughput gain of the JE frame structure over the IE frame structure. It becomes evident that when transmitting the same information, the IE frame structure offers significantly lower average latency and BLEP compared to the JE frame structure. In particular, the IE frame structure provides a substantial reduction in average latency exceeding 40% relative to the JE frame structure when maintaining the same throughput. Conversely, the JE frame structure is favored in scenarios with low heterogeneity in terms of throughput, demonstrating an improvement of over 30% in low SNRs compared to the IE frame structure when transmitting identical information. Furthermore, we show that traffic heterogeneity has less adverse effects on the performance of the IE frame structure than on the JE frame structure. Moreover, traffic heterogeneity can be leveraged to reduce the average latency and improve reliability of the IE frame structure by the proposed shortest block length first (SBF) scheduling. To solve the throughput maximization problems with low complexity, we derive the closed-form expressions of near-optimal total pilot power and near-optimal block length by solving very complicated transcendental equations due to device mobility. The proposed JOMD-IE and JOMD-JE algorithms achieve near-optimal performance with a tremendous complexity reduction over the exhaustive search. Finally, we highlight the trade-off between pilot power and pilot overhead, showing that proper pilot power boosting can reduce pilot overhead and substantially improve throughput, especially in low and moderate SNRs. At high SNRs, however, the pilot overhead emerges as a significant factor in enhancing throughput.

APPENDIX A
PROOF OF PROPOSITION 1

For the sake of brevity, in this appendix, Let z denote $\gamma_{\mathrm{e},u}^{\Omega}$, let p denote $\rho_{\mathrm{p},k}^{\Omega}$, let a denote α_k^{Ω}, let f denote $f_{\mathrm{D},u}$, respectively. To establish that z is concave with respect to p, we need to demonstrate that $\frac{\partial^2 z}{\partial p^2} \leq 0$. First, we have:

$$\frac{\partial z}{\partial p} = \frac{az^2\left(\left(a^2z(2f+a-1)\right)p^2+(4af(a-z-1))\,p-2f(a-z-1)\right)}{\left((a^2z+(2f-1)\,za)\,p+2f\,(a-z-1)\right)^2}. \tag{31}$$

Let $\mathcal{G}_1(p)$ and $\mathcal{G}_2(p)$ denote the numerator and denominator of $\frac{\partial z}{\partial p}$, respectively. The second-order partial derivative of z with respect to p can be expressed as:

$$\frac{\partial^2 z}{\partial p^2} = \frac{\mathcal{G}_2(p)\cdot\frac{d}{dp}(\mathcal{G}_1(p)) - \mathcal{G}_1(p)\cdot\frac{d}{dp}(\mathcal{G}_2(p))}{(\mathcal{G}_2(p))^2}. \tag{32}$$

Let $\mathcal{G}_3(p)$ represent the numerator of $\frac{\partial^2 z}{\partial p^2}$. The expression for $\mathcal{G}_3(p)$ is given by:$\mathcal{G}_3(p) = a^2fz^2\,(a-z-1)\,(a-1)\left(f+\frac{z}{2}\right)((2f+a-1)\,azp-2fz+2f\,(a-1))$. Considering the properties of the system: the pilot overhead is less than 1, thus $a<1$. Given that the pilot power and SNR are non-negative, so $p \geq 0$

and $z \geq 0$. If the condition $(2f+a-1) < 0$ holds, *i.e.*, $a < 1-2f$, then it follows that $\mathcal{G}_3(p) \leq 0$. This condition is plausible in practice as the value of f is unlikely to exceed 0.1 and a would typically not be greater than 0.8. For instance, $f \approx 0.1$ when $\upsilon_u = 300$ km/h at 6 GHz with the SCS of 15 kHz. Therefore, we deduce that $\frac{\partial^2 z}{\partial p^2} \leq 0$, confirming that z is concave with respect to p.

The optimal p to maximize z with given a, denoted as p^*, can be obtained by solving $\mathcal{G}_1(p) = 0$. That is, $p^* = \frac{2f(a-z-1)-\sqrt{2f(2f+z)(a-1)(a-z-1)}}{az(2f+a-1)}$. When $p < p^*$, boosting pilot power enhances the effective SNR. However, when $p > p^*$, the effective SNR tends to decrease.

Appendix B
Proof of Proposition 2

For the sake of brevity, in this appendix, let z, q, $\mathcal{L}$ denote $\gamma^{\Omega}_{\mathrm{e},u}$, $(Q^{-1}(\epsilon_{\mathrm{thr},u})\log_2 e)^2$ and L_u, respectively. We derive the first-order partial derivative of $\mathring{N}^{\Omega}_u$ with respect to $\gamma^{\Omega}_{\mathrm{e},u}$ as:

$$\frac{\partial \mathring{N}^{\Omega}_u}{\partial z} = -\frac{\sqrt{\ln(2)}}{\mathcal{A}_1 \left(\ln(1+z)\right)^3 (1+z)^5}\left(\mathcal{A}_2 + \mathcal{A}_4 q \ln(2)\right), \quad (33)$$

where $\mathcal{A}_1 = (1+z)^{-2}\sqrt{z(z+2)\,q\ln(2)+4\mathcal{L}\ln(1+z)(1+z)^2}\cdot\sqrt{z(z+2)\,q}$, $\mathcal{A}_2 = (1+z)^2\mathcal{A}_1\mathcal{A}_3$, $\mathcal{A}_3 = q(\ln 2)^{3/2} z(z+2) + \left(-q(\ln 2)^{\frac{3}{2}} + \mathcal{L}(1+z)^2\sqrt{\ln 2}\right)\ln(1+z)$, $\mathcal{A}_4 = z^2 q\ln(2)(z+2)^2 - 2\mathcal{L}(1+z)^2(\ln(1+z))^2 - z(z+2)\left(q\ln(2) - 3\mathcal{L}(1+z)^2\right)\ln(1+z) + z^2 q\ln(2)(z+2)^2$. We prove that $\mathcal{A}_2$ and $\mathcal{A}_4$ is larger than 0 in the domain of z ($z > 0$), and therefore $\frac{\partial \mathring{N}^{\Omega}_u}{\partial z} < 0$, indicating $\mathring{N}^{\Omega}_u$ is monotonically decreasing with respect to $\gamma^{\Omega}_{\mathrm{e},u}$.

Proof of $\mathcal{A}_2 > 0$: It is obvious that $\mathcal{A}_1 > 0$. As long as $\mathcal{A}_3 > 0$, then $\mathcal{A}_2 > 0$. We have $\partial\mathcal{A}_3/\partial z = l(1+z)(2\ln(1+z)+1) + q\left(2z^2+4z+1\right)\ln(2)/(1+z) > 0$, which means $\mathcal{A}_3$ is monotonically increasing with respect to z. And since $\mathcal{A}_3|_{z=0} = 0$, $\mathcal{A}_3 > 0$. Therefore, $\mathcal{A}_2 > 0$ for $z > 0$.

Proof of $\mathcal{A}_4 > 0$: We have $\mathcal{A}_4|_{z=0} = 0$ and $\partial\mathcal{A}_4/\partial z = \mathcal{A}_5/(1+z)$, where $\mathcal{A}_5 = -4\mathcal{L}(1+z)^2(\ln(1+z))^2 - 2(1+z)^2\left(q\ln(2) - 6\left(z^2+2z+1/6\right)\mathcal{L}\right)\ln(1+z) + 4(z+2)\left((z+3/2)(z+1/2)\,q\ln(2) + 3/4\mathcal{L}(1+z)^2\right)z$. As long as $\mathcal{A}_5 > 0$, then $\mathcal{A}_4 > 0$. For $\mathcal{A}_5$, we have $\mathcal{A}_5|_{z=0} = 0$ and $\partial\mathcal{A}_5/\partial z = 16\left(\mathcal{A}_6 + 3/2\left(z^2+2z+1/3\right)\mathcal{L}\right)(1+z)$, where $\mathcal{A}_6 = -\mathcal{L}(\ln(1+z))^2/2 + qz(z+2)\ln(2) + 3\left(z^2+2z+5/12\right)\mathcal{L}$. As long as $\mathcal{A}_6 > 0$, then $\mathcal{A}_5 > 0$. For $\mathcal{A}_6$, we have $\mathcal{A}_6|_{z=0} = 5\mathcal{L}/4$ and $\partial\mathcal{A}_6/\partial z = 2q\ln(2) + \mathcal{L}\left(6z^2+\ln(1+z)+12z+5\right)/(1+z)^2 > 0$. Thus, $\mathcal{A}_4 > 0$ for $z > 0$.

Appendix C
Proof of Proposition 3

Letting $\rho^{\Omega}_{\mathrm{p},k} = 1$, n^{Ω}_k can be regarded as $P^{\Omega}_{\mathrm{p},k}$ in (10). By the definition of throughput (17), given N^{Ω}_k, maximize η^{Ω}_k is equivalent to minimize ϵ^{Ω}_u. So we can proof the throughput $\eta^{\Omega}_{\mathrm{F}}$ is concave with respect to total pilot power $P^{\Omega}_{\mathrm{p},k}$ by proofing ϵ^{Ω}_u is convex with respect to $P^{\Omega}_{\mathrm{p},k}$. For the sake of concise writing, in this appendix, let $\mathcal{P}, \mathcal{L}, \mathcal{N}$ denote $P^{\Omega}_{\mathrm{p},k}$, L_u, N^{Ω}_k, respectively. By defining $W^{\Omega}_u = Q^{-1}(\epsilon^{\Omega}_u)$, we have:

$$\frac{\partial^2 \epsilon^{\Omega}_u}{\partial \mathcal{P}^2} = \frac{1}{\sqrt{2\pi}} e^{-{W^{\Omega}_u}^2/2}\left(W^{\Omega}_u\left(\frac{\partial W^{\Omega}_u}{\partial \mathcal{P}}\right)^2 - \frac{\partial^2 W^{\Omega}_u}{\partial \mathcal{P}^2}\right). \quad (34)$$

It is clear that $\frac{\partial^2 \epsilon}{\partial \mathcal{P}^2} > 0$ holds if $\frac{\partial^2 W^{\Omega}_u}{\partial \mathcal{P}^2} < 0$. For simplicity, the W^{Ω}_u is appropriately simplified. Firstly, $V\left(\gamma^{\Omega}_{\mathrm{e},u}\right)$ changes little with respect to $\mathcal{P}$ and can be approximated as $\frac{1}{2}(\log_2 e)^2$ [32], so we simplify $\sqrt{(\mathcal{N}-\mathcal{P})/V(\gamma^{\Omega}_{\mathrm{e},u})}$ as $\sqrt{\mathcal{N}/V(\gamma_u)}$. Secondly, since $\mathcal{L} \gg \log_2(\mathcal{N}-\mathcal{P})$, we remove the term $\frac{\log_2(\mathcal{N}-\mathcal{P})}{2(\mathcal{N}-\mathcal{P})}$. The proposition still can be proofed without removing this term with a tedious process. Finally we simplify W^{Ω}_u as:

$$W^{\Omega}_{\mathrm{S},u} = \left(\sqrt{\frac{\mathcal{N}-\mathcal{P}}{V(\gamma_u)}}\left(C\left(\gamma^{\Omega}_{\mathrm{e},u}\right) - \frac{\mathcal{L}}{\mathcal{N}-\mathcal{P}}\right)\right). \quad (35)$$

It's second derivative with respect to n^{Ω}_k is as:

$$\frac{\partial^2 W^{\Omega}_{\mathrm{S},u}}{\partial \mathcal{P}^2} = \sqrt{\frac{1}{V(\gamma_u)}}\Big(\mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3 + \sqrt{\mathcal{N}-\mathcal{P}}\left(\mathcal{B}_4 + \mathcal{B}_5\right)\Big), \quad (36)$$

where $\mathcal{B}_1 = -\Big(C\left(\gamma^{\Omega}_{\mathrm{e},u}\right) - \frac{\mathcal{L}}{\mathcal{N}-\mathcal{P}}\Big)/4(\mathcal{N}-\mathcal{P})^{3/2}$, $\mathcal{B}_2 = -\mathcal{B}_6(\mathcal{N}-\mathcal{P})^{-\frac{1}{2}}/\left(\mathcal{B}_7(\mathcal{B}_8)^2\right)$, $\mathcal{B}_3 = -\mathcal{L}(\mathcal{N}-\mathcal{P})^{-5/2}$, $\mathcal{B}_4 = -\gamma_u\mathcal{B}_6/\left(\mathcal{B}_7(\mathcal{B}_8)^3\right)$, $\mathcal{B}_5 = -\gamma_u\mathcal{B}_6^2/\left(\mathcal{B}_7^2(\mathcal{B}_8)^4/\ln 2\right)$, $\mathcal{B}_6 = {\gamma_u}^2(\gamma_u+1)\mathcal{N}f_{\mathrm{D}}$, $\mathcal{B}_7 = 2\ln(2)\left(\gamma^{\Omega}_{\mathrm{e},u}+1\right)$, $\mathcal{B}_8 = (\mathcal{N}f_{\mathrm{D}} + \mathcal{P}/2)\gamma_u + \mathcal{N}f_{\mathrm{D}}$. Since $\epsilon < 0.5$ and $\mathcal{N} > \mathcal{P}$ in practice, $C\left(\gamma^{\Omega}_{\mathrm{e},u}\right) - \frac{\mathcal{L}}{\mathcal{N}-\mathcal{P}}$ in $\mathcal{B}_1$ is greater than zero. So obviously, $\sum_{i=1}^{5}\mathcal{B}_i < 0$. $\frac{\partial^2 W^{\Omega}_{\mathrm{S},u}}{\partial \mathcal{P}^2} < 0$ is proven. Therefore ϵ is convex with respect to $\mathcal{P}$, *i.e.*, η is concave with respect to $\mathcal{P}$.

Appendix D
Proof of Proposition 5

In homogeneous scenarios, the SNR and effective SNR are consistent across all users. The optimal block structure of the data block of a single device is first obtained by the JPLLO algorithm. The control block of the IE frame structure has a small number of information bits L_{C}. It has little impact on some optimal control block parameters, *i.e.*, $\alpha^{\mathrm{I}}_{\mathrm{C}} \approx \alpha^{\mathrm{I}}_1, \gamma^{\mathrm{I}}_{\mathrm{e,C}} \approx \gamma^{\mathrm{I}}_{\mathrm{e},1}, A^{\mathrm{I}}_{q,\mathrm{C}} \approx A^{\mathrm{I}}_{q,1}$. Similarly, although the number of information bits of the single block of the JE frame structure is larger, the parameters of the optimal frame structure can also be approximated by a single device data block, *i.e.*, $\alpha^{\mathrm{J}} \approx \alpha^{\mathrm{I}}_1, \gamma^{\mathrm{J}}_{\mathrm{e}} \approx \gamma^{\mathrm{I}}_{\mathrm{e},1}, A^{\mathrm{J}}_{q,u} \approx A^{\mathrm{I}}_{q,1}$. With $\alpha^{\Omega}_k = n^{\Omega}_k/N^{\Omega}_k$, and $N^{\Omega}_k = N^{\Omega}_{k,\mathrm{Data}} + n^{\Omega}_k$, N^{Ω}_k can be rewritten as:

$$N^{\Omega}_k = N^{\Omega}_{k,\mathrm{Data}} \cdot \frac{1}{1-\alpha^{\Omega}_k} \quad (37)$$

where $N^{\Omega}_{k,\mathrm{Data}}$ is the length of data symbols in a block. In low to moderate SNRs, $N^{\Omega}_{\mathrm{opt},u} = N^{\Omega}_{\epsilon_{\mathrm{thr},u}}$. At high SNR, $N^{\Omega}_{\mathrm{opt},u} = \frac{L_u}{\log_2(M_u)} + n^{\Omega}_k$ as in (22). Thus, the ratio of JE to IE frame length β can be approximated as in (24).

Proof of β is mono-increasing with respect to channel capacity:

$$\frac{\partial \beta}{\partial C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)} = \frac{-2\left(2\mathcal{D}_1\mathcal{D}_4 C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)L^{\mathrm{J}}\sqrt{A^{\mathrm{I}}_{q,1}} + \mathcal{D}_2\mathcal{D}_5 + \mathcal{D}_8\right)}{\mathcal{D}_1\mathcal{D}_2\mathcal{D}_3\left(\sqrt{A^{\mathrm{I}}_{q,1}}(\mathcal{D}_1 + \mathcal{D}_2 U) + 2C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)L^{\mathrm{J}} + A^{\mathrm{I}}_{q,1}(U+1)\right)}, \quad (38)$$

where $\mathcal{D}_1 = \sqrt{A^{\mathrm{I}}_{q,1} + 4C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)L_{\mathrm{C}}}$, $\mathcal{D}_2 = \sqrt{A^{\mathrm{I}}_{q,1} + 4C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)L_1}$, $\mathcal{D}_3 = \sqrt{A^{\mathrm{I}}_{q,1} + 4C\left(\gamma^{\mathrm{J}}_{\mathrm{e}}\right)L^{\mathrm{J}}}$, $\mathcal{D}_4 = L^{\mathrm{J}}\mathcal{D}_2 - UL_1\mathcal{D}_3$, $\mathcal{D}_5 = \mathcal{D}_1\mathcal{D}_6 + \mathcal{D}_3\mathcal{D}_7 + \left(A^{\mathrm{I}}_{q,1}\right)^2\left(L_{\mathrm{C}} - L^{\mathrm{J}}\right)$, $\mathcal{D}_6 = \left(\left(L_1 - L^{\mathrm{J}}\right)U + L_{\mathrm{C}} - L^{\mathrm{J}}\right)\left(A^{\mathrm{I}}_{q,1}\mathcal{D}_3 + \left(A^{\mathrm{I}}_{q,1}\right)^{\frac{3}{2}}\right)$, $\mathcal{D}_7 = \left(L_{\mathrm{C}} - L^{\mathrm{J}}\right)\left(A^{\mathrm{I}}_{q,1}\right)^{\frac{3}{2}} - 2L_{\mathrm{C}}L^{\mathrm{J}}C\left(\gamma^{\mathrm{I}}_{\mathrm{e},1}\right)\sqrt{A^{\mathrm{I}}_{q,1}}$, $\mathcal{D}_8 = -U\mathcal{D}_1\left(L^{\mathrm{J}} - L_1\right)\left(\mathcal{D}_3\left(A^{\mathrm{I}}_{q,1}\right)^{\frac{3}{2}} + \left(A^{\mathrm{I}}_{q,1}\right)^2\right)$. Obviously,

$\mathcal{D}_1 > 0$, $\mathcal{D}_2 > 0$ and $\mathcal{D}_3 > 0$. We have $L_{\rm C} < L^{\rm J}$ and $L_1 < L^{\rm J}$ with $L^{\rm J} = L_{\rm C} + \sum_{u=1}^{U} L_u$. Thus, $\mathcal{D}_5 < 0$ with $\mathcal{D}_6 < 0$ and $\mathcal{D}_7 < 0$. And with $\mathcal{D}_4 < 0$ and $\mathcal{D}_8 < 0$, $\frac{\partial \beta}{\partial C\left(\gamma_{\rm e,1}^{\rm I}\right)} > 0$.

Proof of β is mono-increasing with respect to the amount of information bits L_1:

$$\frac{\partial \beta}{\partial L_1} = \frac{4UC\left(\gamma_{\rm e,1}^{\rm I}\right)\left(\mathcal{D}_9 + \mathcal{D}_{10}\right)}{\mathcal{D}_2 \mathcal{D}_3 \left(\mathcal{D}_{13}\right)^2 \left(A_{q,1}^{\rm I}\right)^2} \tag{39}$$

where $\mathcal{D}_9 = C\left(\gamma_{\rm e,1}^{\rm I}\right)\sqrt{A_{q,1}^{\rm I}}\left(\left(UL_1 - L_{\rm C}\right)\mathcal{D}_3 - \left(UL_1 + L_{\rm C}\right)\mathcal{D}_2\right)$, $\mathcal{D}_{10} = 2\mathcal{D}_3\mathcal{D}_{11} + 2\mathcal{D}_2\left(\mathcal{D}_1 A_{q,1}^{\rm I} + \left(A_{q,1}^{\rm I}\right)^{\frac{3}{2}} U\right) + \mathcal{D}_{12}$, $\mathcal{D}_{11} = \left(\mathcal{D}_1\sqrt{A_{q,1}^{\rm I}} + A_{q,1}^{\rm I} U\right)\mathcal{D}_2 + \left(U-1\right)\left(A_{q,1}^{\rm I}\right)^{\frac{3}{2}}$, $\mathcal{D}_{12} = 2A_{q,1}^{\rm I}\left(A_{q,1}^{\rm I} U/4 - A_{q,1}^{\rm I}/4 - L_{\rm C} C\left(\gamma_{\rm e,1}^{\rm I}\right)\right)$, $\mathcal{D}_{13} = 2C\left(\gamma_{\rm e,1}^{\rm I}\right) L_1 U + \sqrt{A_{q,1}^{\rm I}}\mathcal{D}_2 + \sqrt{A_{q,1}^{\rm I}}\mathcal{D}_1 + + A_{q,1}^{\rm I}\left(U+1\right)$. In general, $L_{\rm C}$ can be neglected since $UL_1 \gg L_{\rm C}$. Thus, $\mathcal{D}_9 > 0$ as $\mathcal{D}_3 > \mathcal{D}_2$. $\mathcal{D}_{10} > 0$ because $\mathcal{D}_{11} > 0$ and $\mathcal{D}_{12} > 0$. And with $\mathcal{D}_{13} > 0$, $\frac{\partial \beta}{\partial L_1} > 0$.

Proof of β is mono-decreasing with respect to the number of devices U: Generally , $L_{\rm C}$ is a mono-increasing linear function with respect to U, as $L_{\rm C} = m_1 U + m_2$. We have:

$$\frac{\partial \frac{\widetilde{N}^{\rm J}}{U\widetilde{N}_1^{\rm I}}}{\partial U} = \frac{\left(m_2 C\left(\gamma_{\rm e,1}^{\rm I}\right) + A_{q,1}^{\rm I}/2\right)\mathcal{D}_3 + \left(A_{q,1}^{\rm I}\right)^{\frac{3}{2}}/2 + L^{\rm J} C\left(\gamma_{\rm e,1}^{\rm I}\right)\sqrt{A_{q,1}^{\rm I}}}{-\mathcal{D}_3\left(A_{q,1}^{\rm I}\right)^2\left(2C\left(\gamma_{\rm e,1}^{\rm I}\right)L_1 + \sqrt{A_{q,1}^{\rm I}}\mathcal{D}_2 + A_{q,1}^{\rm I}\right)/2}. \tag{40}$$

$\frac{\partial}{\partial U}\left(\frac{\widetilde{N}^{\rm J}}{U\widetilde{N}_1^{\rm I}}\right) < 0$ since $\mathcal{D}_2 > 0$ and $\mathcal{D}_3 > 0$. That is, $\frac{\widetilde{N}^{\rm J}}{U\widetilde{N}_1^{\rm I}}$ is mono-decreasing with respect to the number of devices U. And because $\frac{\widetilde{N}^{\rm J}}{\widetilde{N}_{\rm C}^{\rm I} + U\widetilde{N}_1^{\rm I}} < \frac{\widetilde{N}^{\rm J}}{U\widetilde{N}_1^{\rm I}}$, β is mono-decreasing with respect to the number of devices U.

REFERENCES

[1] Liu *et al.*, "Pilot overhead vs. pilot power: Short packet structure optimization for URLLC over continuous fading," in *Proc. 2021 IEEE Global Communications Conference (Globecom2021)*, Madrid, Spain, Dec. 2021.

[2] You *et al.*, "Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, pp. 1–74, Jan. 2021.

[3] Cao *et al.*, "Toward industrial metaverse: Age of information, latency and reliability of short-packet transmission in 6g," *IEEE Wireless Communications*, vol. 30, no. 2, pp. 40–47, Apr. 2023.

[4] Ma *et al.*, "High-reliability and low-latency wireless communication for internet of things: Challenges, fundamentals, and enabling technologies," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7946–7970, May 2019.

[5] 3GPP, "Service requirements for cyber-physical control applications in vertical domains," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.104, Dec. 2021, version 18.3.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3528

[6] 3GPP, "Service requirements for enhanced V2X scenarios," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.186, Apr. 2022, version 17.0.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3180

[7] Qiu *et al.*, "How can heterogeneous internet of things build our future: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2011–2027, Mar. 2018.

[8] Cao *et al.*, "Short frame structure optimization for industrial IoT with heterogeneous traffic and shared pilot," in *Proc. 2020 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2020, pp. 1–6.

[9] Popovski *et al.*, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Feb. 2018.

[10] Popovski *et al.*, "Wireless access in ultra-reliable low-latency communication (urllc)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[11] Polyanskiy *et al.*, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[12] Yang *et al.*, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[13] Durisi *et al.*, "Short-packet communications over multiple-antenna rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.

[14] Collins *et al.*, "Coherent multiple-antenna block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 380–405, Jan. 2019.

[15] Sun *et al.*, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 402–415, Jan. 2019.

[16] Chen *et al.*, "Resource allocation for wireless-powered IoT networks with short packet communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1447–1461, Feb. 2019.

[17] Ren *et al.*, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[18] Ren *et al.*, "Resource allocation for secure URLLC in mission-critical IoT scenarios," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5793–5807, Sep. 2020.

[19] Cao *et al.*, "Independent pilots versus shared pilots: Short frame structure optimization for heterogeneous-traffic urllc networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 5755–5769, Aug. 2022.

[20] Trillingsgaard *et al.*, "Downlink transmission of short packets: Framing and control information revisited," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2048–2061, May 2017.

[21] Chan *et al.*, "Age of information with joint packet coding in industrial IoT," *IEEE Commun. Lett.*, vol. 10, no. 11, pp. 2499–2503, Nov. 2021.

[22] Tang *et al.*, "The age of information of short-packet communications: Joint or distributed encoding?" in *Proc. 2022 IEEE International Conference on Communications (ICC)*, May 2022, pp. 2175–2180.

[23] Tang *et al.*, "Age of information in downlink systems: Broadcast or unicast transmission?" *IEEE Journal on Selected Areas in Communications*, Jul. 2023.

[24] Zhao *et al.*, "Joint framing and finite-blocklength coding for URLLC in multi-user downlinks," in *Proc. 2020 IEEE International Conference on Communications (ICC)*, Jun. 2020, pp. 1–6.

[25] Zhao *et al.*, "Achieving extremely low-latency in industrial internet of things: Joint finite blocklength coding, resource block matching, and performance analysis," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6529–6544, Oct. 2021.

[26] F.Hlawatsch *et al.*, *Wireless communications over rapidly time-varying channels*. Academic press, May 2011.

[27] Spalvieri *et al.*, "Pilot-aided carrier recovery in the presence of phase noise," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1966–1974, Jul. 2011.

[28] Liu *et al.*, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[29] Mousaei *et al.*, "Optimizing pilot overhead for ultra-reliable short-packet transmission," in *Proc. 2017 IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–5.

[30] Jindal *et al.*, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[31] Wang *et al.*, "On the age of information of short-packet communications with packet management," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, Dec. 2019, pp. 1–6.

[32] Cao *et al.*, "Proc. joint block length and pilot length optimization for URLLC in the finite block length regime," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[33] Chen *et al.*, "Resource allocation for wireless-powered IoT networks with short packet communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1447–1461, Feb. 2019.

[34] Ohno *et al.*, "Average-rate optimal psam transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Apr. 2002.

[35] Dong *et al.*, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[36] 3GPP, "Coexistence between NB-IoT and NR," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 37.824, Jul. 2020, version 16.0.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3663