Effective Joint Scheduling and Power Allocation for URLLC-Oriented V2I Communications

Jing Li, Yong Niu, *Senior Member, IEEE*, Hao Wu, *Member, IEEE*, Bo Ai, *Fellow, IEEE*, Tony Q. S. Quek, *Fellow, IEEE*, Ning Wang, *Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

***Abstract*—Future vehicular applications, such as safety guarantee and autonomous driving, rely on vehicular-to-infrastructure (V2I) ultra-reliable low-latency communication (URLLC). This paper investigates the flow scheduling and power allocation mechanism to improve the transmission capacity of the downlink V2I orthogonal frequency division multiplexing (OFDM) URLLC network. Given the stringent latency requirements, short package transmission is adopted and the approximation of the finite blocklength codes capacity is introduced for the algorithm design. Also in the system design, we fully consider the effect of Doppler spread caused by high vehicular mobility. We formulate the problem of maximizing the number of flows that satisfy delay and reliability requirements while meeting the constrained radio and power resources. To solve this challenging non-convex problem, we propose a joint optimization framework for flow scheduling and power allocation. In the scheduling phase, we propose a deferred acceptance based flow scheduling algorithm by leveraging matching game. In the power allocation phase, we design a collection-reallocation algorithm for local power optimization while fully considering the dynamic characteristics of V2I scenarios. Numerical results show that the proposed scheme effectively enhances the system performance compared to other state-of-art mechanisms.**

***Index Terms*—Flow scheduling, matching theory, OFDM, power allocation, URLLC.**

I. Introduction

INTELLIGENT transportation system (ITS) is emerging as one of the most fundamental scenarios in the future communication era, which promises to improve network availability and reliability, reduce latency, and conserve energy [1], [2]. Among the technologies proposed for ITS, enhanced vehicle communication provides a more efficient and safer traffic experience for our daily life. On the one hand, vehicle-to-vehicle (V2V) and vehicle-to-pedestrian (V2P) communications adopt the device-to-device (D2D) mode, which supports extended sensing and enhanced mobile broadband (eMBB) services [3]. On the other hand, vehicle-to-infrastructure (V2I) communication achieves point-to-multipoint transmission, extends transmission range, and provides high data rates for nodes in coverage, which promises to enable ultra-reliable low-latency communication (URLLC) for mission-critical applications, such as traffic management, autonomous driving, collision prevention, etc. [4], [5]. Typically, URLLC in V2I imposes strict requirements on latency (e.g., $3-10$ ms) and reliability (e.g., 10^{-5}), with the payload size limited to around 300 bytes [6]. To meet these stringent metrics, research and development for V2I URLLC are rapidly advancing.

From the standards aspect, the third generation partnership project (3GPP) introduces the fifth generation (5G) new radio (NR) to vehicular systems, which proposes advanced physical (PHY) layer design and medium access control (MAC) layer operation for V2I communications [7]. 5G NR achieves flexible transmission durations and minimum control signaling overhead, where mini-slots, containing 2, 4 or 7 symbols, as short and agile transmission units can be applied to facilitate low-latency applications [8]. Orthogonal frequency-division multiplexing (OFDM) forms the baseline waveform in NR mobile systems, which can easily be combined with multiple antenna configurations for spectral efficiency and reliability enhancement. Evaluated by [9], 5G NR outperforms other standards, in terms of reliability, latency and data rates, which plays an essential role in motivating URLLC cases of vehicular scenarios. In the transmission phase, short-packet with finite blocklength codes (FBC) is introduced to satisfy latency requirements, where the achievable data rate can typically be modeled as a function of decoding error probability (DEP) and blocklength [10]. The authors of [11] have further extended the above result to quasi-static fading channels, and the work [12] has demonstrated the feasibility of achieving URLLC requirements in the 5.9 GHz vehicular system, providing a holistic study for promoting future mobile URLLC applications.

However, due to the limited available resources in mobile vehicular networks, scheduling and resource allocation becomes

Manuscript received 10 August 2023; revised 6 February 2024; accepted 10 March 2024. Date of publication 26 March 2024; date of current version 15 August 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB2900301, in part by the National Natural Science Foundation of China under Grant 62221001, Grant 62231009 and Grant U21A20445, in part by the Beijing Natural Science Foundation under Grant L232042, and in part by the Fundamental Research Funds for the Central Universities under Grant 2023JBMC030, Grant 2022JBXT001 and Grant 2022JBQY004. The review of this article was coordinated by Prof. Joao Catalao. *(Corresponding author: Yong Niu.)*

Jing Li, Yong Niu, Hao Wu, and Bo Ai are with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: jinglee@bjtu.edu.cn; niuy11@163.com; hwu@bjtu.edu.cn; boai@bjtu.edu.cn).

Tony Q. S. Quek is with the Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

Ning Wang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: ienwang@zzu.edu.cn).

Sheng Chen is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: sqc@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TVT.2024.3381924

a significant part of practical URLLC system design. On the one side, considering the constrained radio resources, it is necessary to design the scheduling policies that satisfy the latency objectives of all users while addressing the uncertainty of the vehicle channel [13]. Hence, several recent efforts have been made to achieve this goal. The studies [14] and [15] reviewed the packet scheduling in the existing vehicular URLLC systems and presented potential challenges, thereby providing deep insights for further research. The work [13] innovatively modeled the scheduling problem in the mobile network as a Markov decision process, and proposed a dynamic programming based strategy to solve it. In addition, some researchers concentrated on the joint URLLC and eMBB traffic scheduler, attempting to maximize the utility of eMBB traffic while satisfying URLLC demands, for which puncturing-based transmission schemes [16], [17], [18] have been proposed. Nevertheless, these mechanisms mainly focus on static or low-mobility scenarios and thus cannot be directly applied to V2I networks.

On another side, power allocation is also crucial for the V2I URLLC system, especially in multi-user cases. Typically, this is a challenging task due to the time-varying channel state and the complex interactions among transmission signal power, block-length and DEP. To tackle this challenging problem, several recent literature have conducted the related research [19], [20], [21], [22], [23], [24]. Specifically, the study [19] proposed a minorization-maximization-based algorithm for power allocation and beamforming design, thereby increasing the sum FBL rate while reducing complexity. The work [20] adopted a reinforcement learning approach to maximize the overall capacity while guaranteeing the reliable and delay requirements of V2V links. Also, the authors of [21] investigated a cooperative multi-agent deep reinforcement learning framework for real-time power allocation to ensure the reliability of multi-connectivity URLLC vehicular networks. By characterizing the distribution of extreme events through extreme value theory, the study [22] proposed a federated learning framework to minimize the power consumption in V2V networks with URLLC constraints, which also yielded significant gains in reducing queue length. By noticing the coexistence of URLLC and eMBB services in vehicular networks, the studies [23] and [24] exploited the resource management schemes based on network slicing and puncturing technique to improve performance. These works devised feasible solutions for vehicular resource allocation, whereas the prioritized link scheduling as an important part of quality of service (QoS) guarantee has not been fully considered. In addition, efforts on satisfying URLLC requirements for V2I links are still sparse.

Motivated by the above discussion, we investigate the joint framework for flow scheduling and power allocation in a multi-user downlink OFDM V2I network, aiming to maximize the number of successful flows while guaranteeing their specific delay and reliability constraints. The proposed scheme fits well with the dynamic channel and enables efficient power utilization, serving as a lightweight candidate technique to support vehicular URLLC networks. The main contributions are summarized as follows.

- Based on the 5G NR standard, we design a system model for the downlink V2I URLLC network, where the effect of vehicle mobility, i.e., the interference caused by Doppler shift is fully considered. To guarantee the delivery of safety-critical messages, we formulate an optimization problem with the objective to maximize the number of successful flows by jointly optimizing radio resources and transmission power allocation subject to stringent QoS requirement constraints.
- Since the proposed optimization problem is non-convex and difficult to tackle directly, we decouple it into the sub-problems of flow scheduling and power allocation. The flow scheduling sub-problem can be viewed as a matching process between unscheduled flows and available frequency bins (FBs), which can be solved by the deferred acceptance based flow scheduling algorithm (DAFS). By fully leveraging the quasi-static characteristic of the vehicular channel, we realize the local power optimization by the collection-reallocation algorithm (CRA), which enhances the successful transmission probability effectively. Then the joint scheduling and power allocation algorithm (JSPA) is designed to solve the original problem.
- Extensive simulations are conducted to evaluate the performance of our proposed scheme under diverse network environments, including different numbers of users, power constraints, delay requirements and mobility speeds. The results show that the proposed algorithm always outperforms the existing works, in terms of the completed flows and system throughput.

The remainder of this paper is organized as follows. Section II introduces the system model of the multi-user V2I URLLC system and formulates the optimization problem to maximize the number of successful flows. To solve this challenging problem, the JSPA is proposed in Section III. Then Section IV provides numerical results and Section V concludes this paper. Note that the notations and descriptions in this paper are listed in Table I.

II. System Model and Problem Formulation

A. Network Model

The downlink V2I network is illustrated in Fig. 1, where the power-constrained Road Side Unit (RSU) attempts to send safety-critical information to K vehicles indexed by $k \in \mathcal{K} = \{1, 2, \cdots, K\}$ via short packages. For ease of exposition, we assume that vehicles are equipped with single antenna and are traveling in the same direction at a constant speed of v. Denote the sizes of contents to transmit by $\mathcal{N} = \{N_1, N_2, \ldots, N_K\}$, where N_k represents the payload size of the k-th flow. Herein, to satisfy the users' latency and reliability requirements, each URLLC flow should be successfully transmitted within D_k mini-slots at a maximum block error rate (BLER) of ϵ_k. Thus, we summarize the QoS requirements of each flow as $\{N_k, D_k, \epsilon_k\}$.

In the transmission design, to avoid strong co-channel interference among different flows as well as to achieve the low-latency objective, OFDM modulation is employed with a subcarrier spacing of 15 kHz and supporting mini-slot transmission, which complies with the 5G NR standard [25], [26]. In this case, a

TABLE I
SUMMARY OF NOTATION AND DESCRIPTION

Notation	Description
K, I, J	Number of vehicles/flows, number of mini-slots, number of FBs
$\mathcal{K}$, $\mathcal{I}$, $\mathcal{J}$, $\mathcal{N}$	Set of vehicles/flows, set of mini-slots, set of FBs, set of payload sizes
N_k, D_k, ϵ_k	Payload size of flow k, delay requirement of flow k, BLER of flow k
(i, j)	RB in the i-th mini-slot and j-th FB
G_0	Transmit antenna gain
$A(\theta, \psi)$, $A_H(\theta)$, $A_V(\psi)$	Total receive antenna gain, antenna gain in the horizontal plane, antenna gain the vertical plane
$\mathcal{T}_c, T_c$	Channel coherent time, number of mini-slots $\mathcal{T}_c$ occupies, named coherent interval
τ	Duration of each mini-slot
f_d, v	Maximum Doppler frequency shift, speed of vehicles
$x_{i,j,k}$	Binary variable indicating the allocation result of RB (i, j) ($x_{i,j,k} = 1$ represents RB (i, j) is allocated to vehicle k while $x_{i,j,k} = 0$ denotes not)
$P_{i,j,k}$	Transmit power of flow k in RB (i, j)
Λ, T_o, L, B	Number of OFDM symbols per RB, duration of each OFDM symbol, number of subcarriers per RB, bandwidth of each RB
$S_{i,j,k}$, $S^e_{i,j,k;i,j',k'}$, $S^a_{i,j,k}$	Total interference power received by flow k in RB (i, j), inter-RB interference from flow k' of RB (i, j') to flow k transmitted at RB (i, j), intra RB interference when transmitting flow k in RB (i, j)
$R_{i,j,k}$	Achievable data rate of flow k in RB (i, j)
H	Blocklength
C_k	Total number of data bits received at vehicle k
δ_k	Binary variable representing transmission status of flow k ($\delta_k = 1$ indicates that flow k has been transmitted successfully while $\delta_k = 0$ indicates that the transmission failed)
δ	Total number of completed flows, i.e., $\sum_k \delta_k$

mini-slot in the time domain and 12 sub-carriers in the frequency domain form the basic scheduling unit called the resource block (RB). There are a total of $|\mathcal{I}| \cdot |\mathcal{J}|$ RBs, where $\mathcal{I} = \{1, 2, \ldots, I\}$ and $\mathcal{J} = \{1, 2, \ldots, J\}$ represent respectively the sets of mini-slots and FBs. For clarity, we denote the RB in the i-th mini-slot and j-th FB by (i, j).

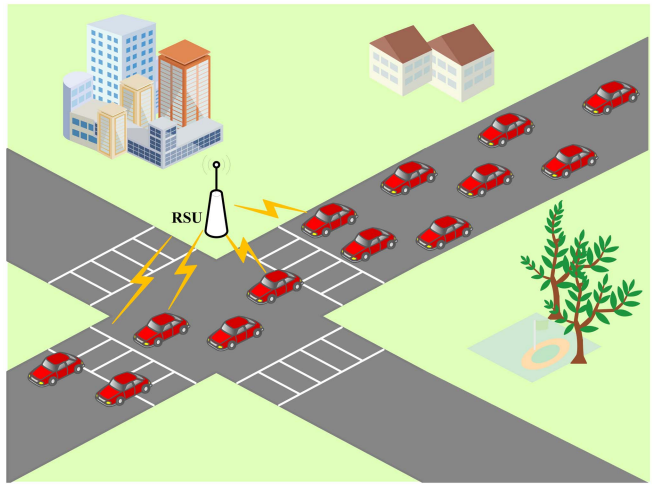


Fig. 1. Vehicular network.

Since both radio and power resources are limited in this V2I network, it is challenging to fulfill the QoS requirements of all flows. Therefore, the RSU performs both radio resource scheduling and transmit power allocation, based on the principles: 1) each RB can only be occupied by one flow at most, and 2) the total power consumption should be no more than the constrained power. In practice, resource management should also take into account the Doppler shift caused by the mobility of vehicles, and this is considered in this paper.

B. Antenna Model

Conventional studies assume the antenna radiation pattern at the RSU as omnidirectional since vehicles can be in any direction relative to it, which is however not throughput-effective [27], [28]. To achieve performance improvement, we deploy four 90° sector antennas on the RSU, while neglecting the interference effects of antenna co-location [29]. Such an antenna configuration has the same transmission characteristics as the omnidirectional ones and can realize the maximum gain $G_0 = 6$ dBi [30].

For vehicles, we consider the popular 3D antenna model in 3GPP TR 37.885, the power pattern of which is described as a function of horizontal angle θ and vertical angle ψ, with the total gain expressed as:

$$A(\theta, \psi) = -\min\{-(A_H(\theta) + A_V(\psi)), 3\} \text{ [dBi]}, \quad (1)$$

where $A_H(\theta)$ and $A_V(\psi)$ represent the antenna gains in the horizontal plane and the vertical plane, respectively.

C. Channel Model

As multiple learning techniques [31] have been developed to perform mobile channel estimation, it is reasonable to assume that the perfect channel state information (CSI) is available at the RSU, which is consistent with the studies [19], [22], [32], [33]. We consider a quasi-static flat fading channel in which the channel gain remains constant in the coherent time $\mathcal{T}_c$ but varies independently across consecutive coherent-time durations. The

most common definition of coherence time is given by [34]

$$\mathcal{T}_c = \sqrt{\frac{9}{16\pi f_d^2}} = \frac{0.423}{f_d}, \tag{2}$$

where $f_d = \frac{v}{c} f_c$ is the maximum Doppler frequency shift, with c being the speed of light and f_c the carrier frequency. Let τ denote the mini-slot duration, and $\mathcal{T}_c \geq \tau$ holds. We define $T_c = \left\lceil \frac{\mathcal{T}_c}{\tau} \right\rceil$ as the coherent interval. Besides, the received power at the k-th vehicle in RB (i, j) can be written as

$$P^r_{i,j,k} = \beta P_{i,j,k} G_0 A\left(\theta_{i,j,k}, \psi_{i,j,k}\right) g_{i,j,k} d^{-\alpha}_{i,j,k}, \tag{3}$$

where β is a constant proportional to $(\frac{\lambda}{4\pi})^2$ with λ being the wavelength, G_0 is the antenna gain at the RSU, α is the path loss (PL) exponent, $P_{i,j,k}$ and $A(\theta_{i,j,k}, \psi_{i,j,k})$ are the transmit power and the antenna gain at the k-th vehicle in RB (i, j), while $d_{i,j,k}$ is the distance between the RSU and the k-th vehicle. Moreover, based on the channel measurement and analysis results [35], [36], [37], it is reasonable to assume that the channel gain $g_{i,j,k}$ follows the Rician fading, with the amplitude probability density function (PDF) [38]:

$$f(g) = \frac{2(\mathrm{K}_f+1)g}{\Omega} \mathrm{e}^{-\mathrm{K}_f} \mathrm{e}^{-\frac{(\mathrm{K}_f+1)}{\Omega} g^2} I_0\left(2\sqrt{\frac{\mathrm{K}_f(\mathrm{K}_f+1)}{\Omega}} g\right), \tag{4}$$

where K_f denotes the Rician K-factor, Ω is the average envelope, and $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind.

D. Data Rate

Although RBs in OFDM systems are designed to be orthogonal, the high-speed motion of vehicles causes the Doppler shift, which destroys the orthogonality between subcarriers and introduces inter-RB interference. Hence, the interference power received by vehicle k in RB (i, j), denoted by $S_{i,j,k}$, includes both inter-RB and intra-RB interferences, and can be expressed as

$$S_{i,j,k} = x_{i,j,k} S^e_{i,j,k;i,j',k'} + S^a_{i,j,k}. \tag{5}$$

In (5), $x_{i,j,k}$ is a binary variable with $x_{i,j,k} = 1$ only when RB (i, j) is allocated to vehicle k. $S^e_{i,j,k;i,j',k'}$ is the inter-RB interference from vehicle k' of RB (i, j'), given by [39]

$$S^e_{i,j,k;i,j',k'} = \frac{P^r_{i,j,k;i,j',k'} \Lambda (T_o f_d)^2}{2\ L} \times \left(\sum_{m=1}^{L} \sum_{n=1, n\neq m}^{L} \frac{1}{((Lj+n)-(Lj'+m))^2} \right), \tag{6}$$

in which $P^r_{i,j,k;i,j',k'}$ and L stand for the received power at RB (i, j) from RB (i, j') and the number of subcarriers per RB, while Λ and T_o are the number of OFDM symbols and the duration of each symbol. In addition, the intra RB (i, j) interference can be calculated as

$$S^a_{i,j,k} = \frac{P^r_{i,j,k} \Lambda (T_o f_d)^2}{2\ L} \left(\sum_{m=1}^{L} \sum_{n=1, n\neq m}^{L} \frac{1}{(m-n)^2} \right). \tag{7}$$

Therefore, the signal-to-interference-plus-noise ratio (SINR) of the k-th flow in RB (i, j) is given by

$$\gamma_{i,j,k} = \frac{P^r_{i,j,k}}{S_{i,j,k} + N_0 B}, \tag{8}$$

where N_0 and B denote the noise power density and the bandwidth of each RB, respectively.

As aforementioned, URLLC systems involve small payloads and employ short packet transmission, rendering the classic Shannon capacity no longer appropriate to describe the maximum achievable data rate. To this end, the FBC capacity formula given in [10] is adopted to characterize the relationship among the achievable rate, latency and BLER, with the achievable rate of flow k in RB (i, j) given by

$$R_{i,j,k} = x_{i,j,k} B \left(\log(1+\gamma_{i,j,k}) - \frac{Q^{-1}(\epsilon_k)}{\ln 2} \sqrt{\frac{V(\gamma_{i,j,k})}{H}} \right), \tag{9}$$

where $Q^{-1}(\epsilon_k)$ stands for the inverse function of the Gaussian tail function $Q(\epsilon_k) = \frac{1}{\sqrt{2\pi}} \int_{\epsilon_k}^{\infty} e^{-t^2/2} dt$, $V(\gamma_{i,j,k}) = 1 - (1 + \gamma_{i,j,k})^{-2}$ defines the channel dispersion, and H denotes the blocklength.

Then based on the joint channel coding theory [40], we encode packets of each flow over all the scheduled RBs, and the maximum number of the received data bits for the k-th vehicle can be approximated as

$$C_k = \sum_i \sum_j \tau R_{i,j,k}. \tag{10}$$

E. Problem Formulation

In this V2I network, the RSU receives the transmission requests with QoS requirements from vehicles at mini-slot 0. Then the transmission of the desired data flows begins when the scheduling is completed. Note that the RSU has the complete knowledge of vehicles' location and the system resource utilization. Our objective is to fully exploit the transmission ability of the V2I network, i.e., to accomplish the traffic demand of URLLC services as much as possible, by jointly optimizing the flow scheduling and power allocation, given the constrained radio resource and limited transmission power. Mathematically, this joint flow scheduling and power allocation optimization problem can be described as

$$\mathcal{P}: \max \sum_k \delta_k, \tag{11}$$

$$\text{s.t. } x_{i,j,k} \in \{0, 1\}, \forall\, i, j, k, \tag{12}$$

$$\sum_k x_{i,j,k} \leq 1, \forall\, i, j, \tag{13}$$

$$x_{i,j,k} = 0, \forall\, i \geq D_k, \tag{14}$$

$$x_{i,j,k}P_{i,j,k} \geq 0, \forall i,j,k, \tag{15}$$

$$\sum_i \sum_j \sum_k x_{i,j,k}P_{i,j,k} \leq P_{\text{sum}}, \forall i,j,k, \tag{16}$$

$$\delta_k = \begin{cases} 1, & C_k \geq N_k \text{ within } D_k, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

In $\mathcal{P}$, δ_k is a binary variable indicating whether flow k is successfully transmitted, so the objective value $\sum_k \delta_k$ denotes the number of completed flows. Constraints (12) and (13) indicate that each RB is assigned to at most one flow, to eliminate mutual interference. To meet the delay requirement, constraint (14) specifies that vehicle k is served within the first D_k mini-slots. Constraint (15) guarantees the non-negative power allocation, and constraint (16) restricts the total available power of the RSU to no more than P_{sum}. Constraint (17) defines the value of δ_k, i.e., within the allowable delay, only if the destination vehicle of flow k receives data that satisfies its target payload demand, i.e., $C_k \geq N_k$ within D_k, $\delta_k = 1$; otherwise $\delta_k = 0$. Therefore, the joint flow scheduling and power allocation optimization is to maximize the number of successful flows, while guaranteeing their specific rate, delay and reliability requirements as well as meeting the RSU's radio resource and power constraints.

III. Proposed Algorithm for URLLC Service Scheduling and Power Allocation

The joint flow scheduling and power allocation optimization formulated in Subsection II-E involves both binary and continuous variables, and it is a mixed-integer nonlinear programming problem (MINLP), which is NP-hard and difficult to solve directly. Inspired by the concurrent scheduling mechanism in time division multiple access systems, we propose a heuristic algorithm to solve it. Specifically, we separately address the flow scheduling and power allocation sub-problems while considering content sizes, latency, and reliability requirements. Subsequently, an efficient joint optimization framework is designed to solve this MINLP.

A. Sub-Problem 1: Flow Scheduling

Given that each flow can use multiple RBs but each RB can only be assigned to one flow, the previous studies [33], [41] modeled the flow scheduling as a many-to-one matching game (MG) and solve it accordingly. However, we notice that each flow tends to transmit in consecutive RBs with stable CSI, i.e., not change its transmission state within a coherence-time duration unless completed. Therefore, we reorganize the entire radio resource and formulate a one-to-one MG as follows.

Definition 1 (Matching Concept): Let $\widehat{\mathcal{J}}(i)$ be the set of available radio resource in the i-th mini-slot, each element of which represents an idle FB. Let $\widehat{\mathcal{K}}$ denote the set of remaining flows that load URLLC services. Initially, we have $\widehat{\mathcal{J}}(1) = \mathcal{J}$ and $\widehat{\mathcal{K}} = \{1, 2, \ldots, K\}$. A two-side matching is defined as the mapping result between flows in $\widehat{\mathcal{K}}$ and resource in $\widehat{\mathcal{J}}(i)$, with each available FB exclusively assigned to one flow for collision avoidance, where

1) $\forall j \in \widehat{\mathcal{J}}$, $\mathcal{M}(j) \in \widehat{\mathcal{K}}$ and $|\mathcal{M}(j)| = 1$,

2) $\forall k \in \widehat{\mathcal{K}}$, $\mathcal{M}(k) \in \widehat{\mathcal{J}}(i) \cup \emptyset$ and $|\mathcal{M}(k)| = \{0, 1\}$.

Herein, $\mathcal{M}(\cdot)$ represents the partner of player j or k under matching $\mathcal{M}$, and $|\mathcal{M}(\cdot)|$ denotes the cardinality of the matched partners. To avoid resource waste and possible extra latency, FBs should be fully allocated, such that $|\mathcal{M}(j)| = 1$ holds. However, due to the limited radio resource, not all flows can be scheduled such that $|\mathcal{M}(k)| = \{0, 1\}$.

Definition 2 (Preference List): The matching $\mathcal{M}$ contains two groups of players whose preference relations are utilized to describe the decision process. On the one hand, each flow aims to maximize its achievable rate, and its preference list can be built by calculating (9) over all available FBs. Denote the preference relation by $\succ_k$, we have $j \succ_k j' \Leftrightarrow R_{i,j,k} > R_{i,j',k}$. On the other hand, each FB tends to accept the flow most likely to be successfully transmitted, with its preference relation denoted by $\succ_j$. Herein, $k \succ_j k' \Leftrightarrow \frac{1}{N_k \cdot D_k} > \frac{1}{N_{k'} \cdot D_{k'}}$ indicates that priority ones are flows with strict latency requirements and light loads, in that a) flows with strict latency constraints should be prioritized to increase the probability of successful transmission; b) the content size N_k should be considered to realize the optimization goal of completing URLLC services as much as possible. Since content-loaded flows take up a large amount of resources, they are less preferred than lightly loaded ones.

Definition 3 (Deferred Acceptance Based Algorithm): Since both the FBs and flows are selfish and rational, the above matching process can be formulated as a game in which the preference list of each player is strict and complete [42]. According to [43], there exist stable states for such a game and the classical deferred acceptance (DA) algorithm can find the stable matching that maximizes the sum-utility.

Therefore, we propose the DA-based solution to obtain the optimal flow scheduling results, as shown in Algorithm 1. The matching procedure is performed at the RSU side, which receives the overall URLLC requests from vehicles beforehand and captures real-time system parameters such as vehicle location, flow scheduling, FB allocation and preference lists. Since the complexity of pairing relation establishment and scheduling result determination is $\mathcal{O}(|\mathcal{J}| + |\mathcal{K}|)$ and $\mathcal{O}(|\mathcal{J}|)$, respectively, the complexity of Algorithm 1 is $\mathcal{O}(|\mathcal{K}|)$.

B. Sub-Problem 2: Local Power Allocation

We adjust the power of the scheduled flows to optimize the transmission. In view of the quasi-static channel, it is reasonable to locally adjust the power in each $\mathcal{T}_c$, with the basic idea of: 1) collecting available power from flows that have more power than needed to complete their payloads before the tolerable delay; 2) and reallocating it to flows that cannot meet the latency requirements, thereby increasing the number of completed flows. Algorithm 2 shows the two-phase pseudo-code of the proposed algorithm, where $a = \lceil \frac{i}{T_c} \rceil$ denotes the index of the current coherent interval.

1) Phase I: First we calculate the data rate of each flow at the average power $\overline{P}$, and then check whether its transmission deadline D_{k_μ} lies in the current coherent-time duration. If so, we further examine whether and how much available power this flow can provide, as indicated by lines 4–14. Specifically,

Algorithm 1: Deferred Acceptance Based Flow Scheduling Algorithm (DAFS).

Input : Set of flows that have not been scheduled, $\widehat{\mathcal{K}}$, and their corresponding content sizes, $\widehat{\mathcal{N}}$; Set of available radio resource in time slot i, $\widehat{\mathcal{J}}(i)$;
Output: Flow scheduling results; Index of each scheduled flow, k_μ.

1 Each flow in $\widehat{\mathcal{K}}$ builds its preference list over $\widehat{\mathcal{J}}(i)$ by calculating (9);
2 Each flow proposes to its top-ranked FB j that has not rejected it before;
3 **if** *j receives one request* **then**
4 accept it;
5 **if** *j receives two or more requests* **then**
6 accept the top-ranked one by calculating $\frac{1}{N_k \cdot D_k}$;
7 Repeat steps 2-6 until no flow requests or each j has been assigned.

this algorithm obtains the local achievable data of flow k_μ by multiplexing its data rate R_{i,j,k_μ} and transmission duration $(D_{k_\mu} - (a-1)\tau T_c)$, after which the following two cases exist. a) When the achievable data size exceeds its remaining content, decrease the power to P_{k_u} that just can complete its transmission. Then increase the number of completed flows by 1 and store the available power in P_a, as in lines 9-10. b) When flow k_μ fails to accomplish transmission within the required delay, record it in $\mathcal{K}_\mu$ and store its unfinished data size $\widehat{N}_{k_\mu}$ in $\widehat{\mathcal{N}}_{k_\mu}$, as described in lines 12-14. As for the flow that does not end in this interval, just update its remaining content size $\widehat{N}_{k_\mu}$ as given in line 16. In this phase, the total available power in the a-th T_c, denoted by ΔP_a, is collected.

2) Phase II: Considering that the flows with less content remaining tend to be successful and deserve priority for power reallocation, Phase II ranks the content sizes in $\widehat{\mathcal{N}}_{k_\mu}$ in line 19. Then ΔP_a is assigned to the flows most likely to complete until all the flows are considered or the available power runs out, as shown in lines 20-26. In this process, once a flow's payload is accomplished, increase δ by 1 and subtract the power consumed in ΔP_a. Since the Phase I with complexity $\mathcal{O}(|\mathcal{J}|)$ and the Phase II with complexity $\mathcal{O}(|\mathcal{J}|-1)$ are executed sequentially, the complexity of Algorithm 2 is $\mathcal{O}(|\mathcal{J}|)$.

Algorithm 2: Collection-Reallocation Algorithm (CRA).

Input : Index of each scheduled flow at time slot i, k_μ; Delay and BLER requirements, and remaining content size of each scheduled flow, D_{k_μ}, ϵ_{k_μ}, and $\widehat{N}_{k_\mu}$; Initial transmit power of each flow, $\overline{P}$;
Output: Completed number of flows δ;

1 **Phase I: Available power collection**
2 **foreach** *scheduled flow k_μ* **do**
3 Calculate corresponding R_{i,j,k_μ} by (9);
4 **if** $D_{k_\mu} \le a \cdot \tau \cdot T_c$ **then**
5 **if** $R_{i,j,k_\mu} \cdot (D_{k_\mu} - (a-1)\tau T_c) \ge \widehat{N}_{k_\mu}$ **then**
6 Decrease power allocated to k_μ from $\overline{P}$ to
7 P_{k_μ} so that:
8 $R'_{i,j,k_\mu} \cdot (D_{k_\mu} - (a-1)\tau T_c) = \widehat{N}_{k_\mu}$;
9 $\delta = \delta + 1$;
10 $\Delta P_a = \Delta P_a + (\overline{P} - P_{k_\mu}) \cdot (D_{k_\mu} - (a-1)\tau T_c)$;
11 **else**
12 $\mathcal{K}_\mu = \mathcal{K}_\mu \cup k_\mu$;
13 $\widehat{N}_{k_\mu} = \widehat{N}_{k_\mu} - R_{i,j,k_\mu} \cdot (D_{k_\mu} - (a-1)\tau T_c)$;
14 $\widehat{\mathcal{N}}_{k_\mu} = \widehat{\mathcal{N}}_{k_\mu} \cup \widehat{N}_{k_\mu}$;
15 **else**
16 $\widehat{N}_{k_\mu} = \widehat{N}_{k_\mu} - R_{i,j,k_\mu} \cdot \tau \cdot T_c$;
17 **Phase II: Power reallocation**
18 **if** $|\mathcal{K}_\mu| \ge 1$ **then**
19 Sort elements in $\widehat{\mathcal{N}}_{k_\mu}$ in ascending order;
20 **for** $l = 1 : |\mathcal{K}_\mu|$ **do**
21 **if** $\Delta P_a > 0$ **then**
22 Calculate power required by flow l to achieve $R'_{i,j,l} \cdot (D_l - (a-1)\tau T_c) = \widehat{N}_l$;
23 **if** $P_l - \overline{P} \le \Delta P_a$ **then**
24 $\delta = \delta + 1$;
25 $\Delta P_a = \Delta P_a$
26 $-(P_l - \overline{P}) \cdot (D_l - (a-1)\tau T_c)$.

C. Joint Flow Scheduling and Power Allocation

Based on the solutions to the flow scheduling and power allocation sub-problems, we propose the JSPA to solve $\mathcal{P}$. The basic idea is to schedule flows for idle FBs and adjust local power at the first mini-slot of each coherent-time interval or when a flow transmission completes. To illustrate the principle clearly, an example of 5 FBs is presented in Fig. 2, with the specific operations of JSPA at the following time points:

1) $aT_c + 1$*:* No idle FB, run CRA for all ongoing flows;
2) i_1*:* Run DAFS for the idle FB 4 and CRA for the newly scheduled flow;
3) $(a+1)T_c + 1$*:* Run DAFS for FBs 5, 2 and then conduct CRA for all ongoing flows;
4) i_2*:* For FB 1, execute the same procedure as in i_1;
5) i_3*:* For FB 3, execute the same procedure as in i_1.

The corresponding pseudo-code is summarized in Algorithm 3. In the system, I mini-slots are divided by coherent interval T_c into A segments, where $I = \mathrm{A} \cdot T_c$. After the initialization, the algorithm sequentially executes the joint optimization algorithm for A intervals. In each T_c, we capture the CSI, update the vehicles' location, and initialize the available power as 0. Then in lines 4-15, we make scheduling decisions and conduct power allocation mini-slot by mini-slot. Specifically, it is reasonable to adjust the power of the scheduled flows at the first mini-slot of T_c according to the updated system parameters, before which whether idle FB exists is checked. If so, run Algorithm 1 to assign

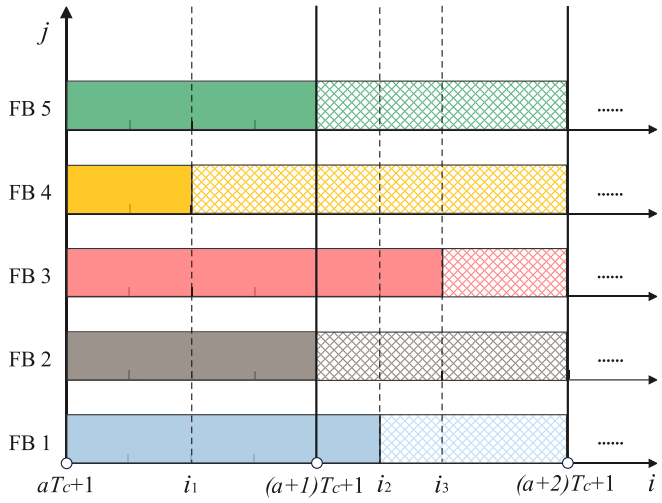


Fig. 2. Illustration of JSPA principle.

Algorithm 3: Joint Scheduling and Power Allocation Algorithm (JSPA).

Input : Set of vehicles $\mathcal{K}$; Set of radio resource $\mathcal{J}$; Initial transmit power of each flow $\overline{P}$; Content sizes of overall flows $\mathcal{N}$;
Initialization: Remaining flows $\widehat{\mathcal{K}} = \mathcal{K}$; Remaining content sizes $\widehat{\mathcal{N}} = \mathcal{N}$; Completed flows $\delta = 0$;

1 **for** $a = 1 : \mathrm{A}$ **do**
2 Update system parameters;
3 Reallocated power $\Delta P_a = 0$;
4 **if** *slot* $i = ((a-1)T_c + 1)$ **then**
5 **if** *idle FB exists* **then**
6 **foreach** *idle FB* **do**
7 Assign a flow k' from $\widehat{\mathcal{K}}$ by using Algorithm 1;
8 Update $\widehat{\mathcal{K}} = \widehat{\mathcal{K}} - k'$;
9 Execute CRA using Algorithm 2 for all scheduled flows;
10 $i = i + 1$;
11 **for** *slot* $i = ((a-1)T_c + 2) : aT_c$ **do**
12 **if** *idle FB exists* **then**
13 Execute steps 6-8;
14 Conduct CRA using Algorithm 2 for newly scheduled flows;
15 $i = i + 1$;
16 $a = a + 1$;
17 **return** δ.

the flows from $\widehat{\mathcal{K}}$ to idle FBs, and then conduct the CRA in Algorithm 2 for the transmitting flows. In the following mini-slots, once a flow reaches its content requirement, the corresponding FB turns idle. At this time, execute the same procedure as lines 6-8 and conduct the CRA for the newly scheduled flows. In this way, we fully utilize the overall radio resource in each interval and maximize the number of completed flows.

1) Complexity: In Algorithm 3, the outer loop executes A times, while in the inner loop, the part (lines 11-15) repeats at most $|\mathcal{J}|T_c$ times, which is generally more than that in lines 4-9 and dominates the inner complexity. Thus, the worst-case computational complexity of Algorithm 3 is $\mathcal{O}(A|\mathcal{J}|T_c)$, which can be implemented in URLLC systems.

2) Convergence: On the one hand, the number of completed flows in Algorithm 3 is non-decreasing with mini-slots because Algorithm 2 reallocates the available power to flows needing them, trying to complete more flows while ensuring the successful transmission of basic ones. On the other hand, the number of accomplished flows is upper-bounded due to the limited number of flows and constrained resources. Therefore, Algorithm 3 is guaranteed to converge.

IV. Performance Evaluation

In this section, we provide numerical results to validate the performance of our proposed scheme for V2I OFDM URLLC, and investigate the impacts of the network parameters, such as number of users, transmit power, delay requirement and mobility speed, on the achievable system performance.

A. Simulation Setup

In the simulation, we consider a wireless vehicular network in the 5.9 GHz band, where an RSU is located at the center of the coverage area with a radius of 500 m. Vehicles requiring URLLC services are randomly distributed within the cell and are moving in the same direction at a constant speed of 60 km/h. The antenna heights of the RSU and vehicles are respectively 5 m and 1.5 m [30]. According to the 5G NR specifications [44], we set the duration of one mini-slot to 0.28 ms, consisting of 4 OFDM symbols. Then each RB is composed of 12 subcarriers with 15 kHz spacing in the frequency domain and 4 OFDM symbols in the time domain. Thus, the bandwidth of each RB is 180 kHz. Note that we consider URLLC services with different QoS requirements in the evaluation, which is more practical than the mean payload size assumption in most of the previous works. The complete list of the simulation parameters is given in Table II. Unless otherwise specifically stated, these default parameters are used in the simulations.

Based on the analysis of [1], [33], [46], we choose the number of completed flows and system throughput as the key evaluation metrics. Their specific definitions are as follows.

1) *Number of completed flows:* The number of flows that accomplish payload transmission under specific delay and reliability constraints, denoted by δ.
2) *System throughput:* After counting the number of data bits reached by successfully transmitted flows, divide it by the actual total time consumption/completion time of the last successful flow to obtain the system throughput.

Next, we consider the four benchmark schemes for performance comparison, which are

TABLE II
DEFAULT SYSTEM PARAMETERS

Parameter	Symbol	Value
Carrier frequency	f	5.9 GHz
System bandwidth	W	20 MHz
Number of mini-slots	I	17
Number of FBs	J	20
Symbol duration	T_o	66.67μs
Blocklength	H	256 [45]
Total transmit power	P_{sum}	30 dBm
Noise power density	N_0	-169 dBm/Hz [40]
Rician K-factor	K_f	7
PL exponent	α	2
Speed of vehicles	v	60 km/h
Number of vehicles	K	34
BLER	ϵ_k	1×10^{-5}
Latency requirement	D_k	$3-5$ ms [6]
Package size	Q_k	$2400-3000$ bits [6], [33]

1) *MG only (MGO) [47]:* This scheme utilizes MG to solve the flow scheduling sub-problem while allocating equal power to all ongoing flows. Its complexity is $\mathcal{O}(|\mathcal{K}|)$.
2) *Local power allocation only (LPAO):* This scheme runs the proposed scheme without matching. Specifically, it randomly selects an unscheduled flow for each idle FB and conducts the CRA for local power allocation. The computational complexity is $\mathcal{O}(AT_c)$.
3) *Water-filling power allocation only (WPAO) [48]:* This scheme utilizes random flow scheduling while conducting water-filling power allocation for all scheduled flows. Notice that its complexity is $\mathcal{O}(A|\mathcal{J}|T_c)$, the same as the proposed JSPA.
4) *Random transmission scheme (RTS):* This scheme runs random flow scheduling for idle FBs while maintaining the same average power during transmission, and the complexity is $\mathcal{O}(A)$.

All the simulation results are averaged over 500 realizations.

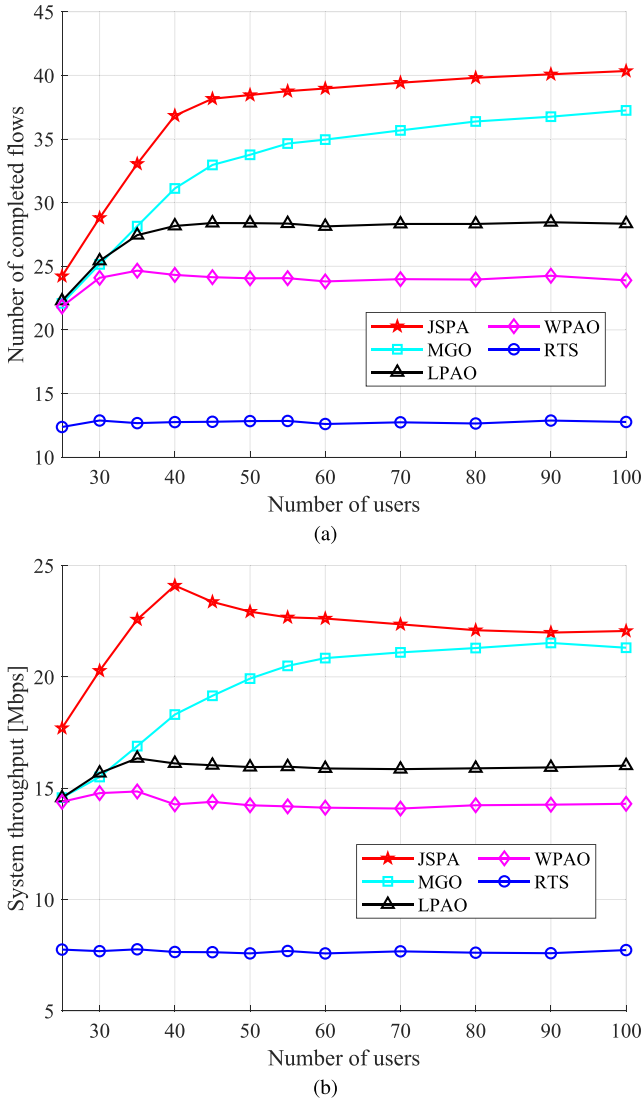


Fig. 3. Comparison of performance as the functions of number of users for five schemes: (a) number of completed flows and (b) system throughput.

B. Performance Under Different Numbers of Users

Fig. 3 compares the performance versus the number of users for the proposed JSPA and four benchmark schemes.

From Fig. 3(a), it can be observed that as the number of request users increases, the completed flows increase for all schemes except RTS. In addition, LPAO and WPAO reach the flow saturation values of 27 and 24 respectively, while that of JSPA and MGO have not been reached but are definitely higher. Evidently, the proposed JSPA attains the best performance, and the proposed local power allocation (LPAO) outperforms the traditional water-filling mechanism (WPAO). Therefore, the simulation results demonstrate that both our flow scheduling and power allocation are effective in improving the transmission capacity.

Fig. 3(b) plots the achievable system throughput of the five schemes as the functions of user numbers. Again we observe that our JSPA achieves the best performance, MGO is better than LPAO and WPAO, while RTS has the worst performance. It can also be seen that the achievable system throughput of JSPA first increases as K increases and starts to decrease after reaching a saturation value of 24.5 Mbps at $K = 40$. The reason is that 1) When $K \leq 40$, most of flows can be successfully transmitted, contributing to the throughput improvement. 2) However, when K exceeds 40, JSPA selects flows with stricter latency requirements and lighter payloads, which may not be throughput-optimal. The result is reasonable because our goal is to increase the number of completed flows rather than maximizing the overall throughput.

C. Performance Under Different Maximum Power Constraints

As shown in Fig. 4, both the number of completed flows and system throughput increase with the available total power P_{sum}. This is because larger available power enables higher data rate and lower transmission delay, whereby more flows

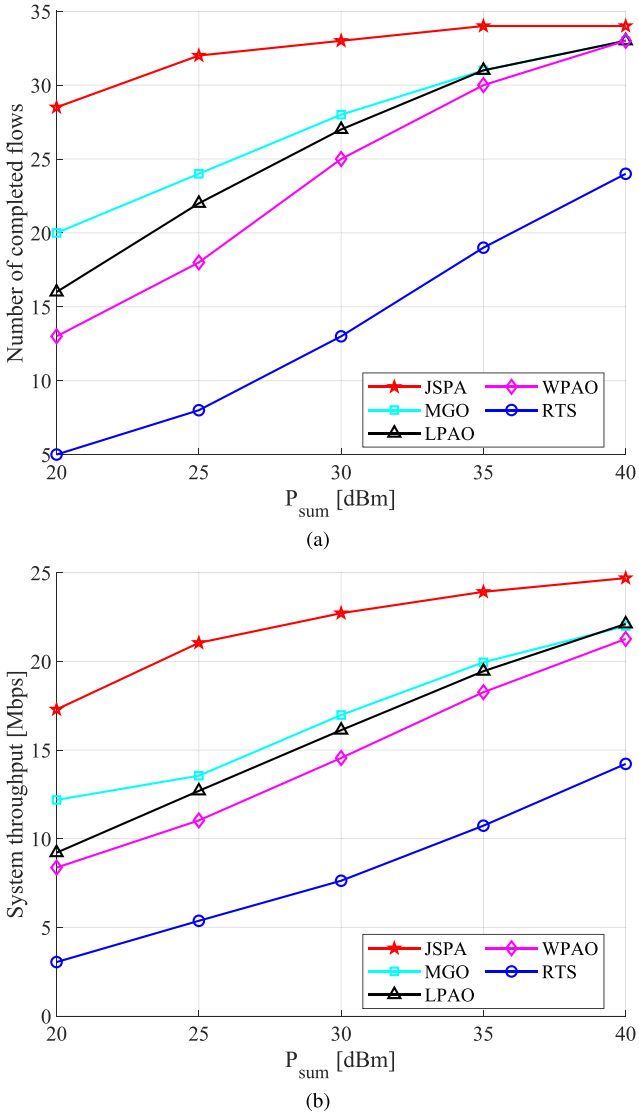


Fig. 4. Comparison of performance as the functions of total power constraint for five schemes: (a) number of completed flows and (b) system throughput.

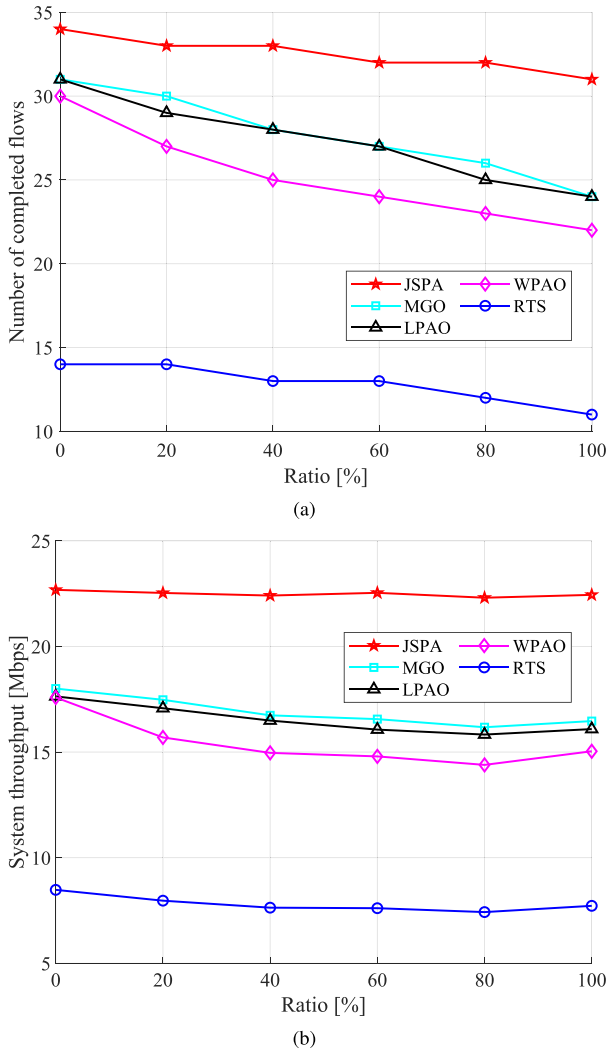


Fig. 5. Comparison of performance as the functions of delay requirement for five schemes: (a) number of completed flows and (b) system throughput.

can be transmitted successfully in the limited time. It can be seen that our proposed JSPA outperforms the other schemes, in terms of both number of completed flows and achievable system throughput. We also notice that the performance gap between the best JSPA and the second-best MGO narrows as $P_{\rm sum}$ increases, because almost all flows can be accomplished with sufficient power. However, since most practical wireless communication systems are power-constrained, efficient and effective resource allocation is required.

D. Performance Under Different Delay Requirements

To investigate the impact of delay requirements on system performance, in the simulation, we define a stricter delay model of $3-4$ ms and a looser delay model of $4-5$ ms. Fig. 5 compares the performance of the five schemes, in terms of completed flows and system throughput, as the functions of the ratio (%) of stricter-delay flows to overall flows.

It can be seen from Fig. 5(a) that our JSPA completes the most flows, which is expected. Also as the percentage of stricter-delay flows increases, fewer flows are completed. This is because when the delay requirements are looser, precious radio resource can be allocated first to those flows with tighter requirements. Then, once a flow is accomplished, one of the remaining flows with looser delay demand can be scheduled to the idle FB and transmitted successfully. However, in the case with only stricter-delay flows, all the flows require to be considered for transmission immediately or they easily fail to meet the delay requirements.

Fig. 5(b) shows that our JSPA attains the best throughput performance and its achievable throughput remains constant as the percentage of stricter-delay flows increases, demonstrating its effectiveness. However, the system throughput of MGO, LPAO, WPAO and RTS first decrease with the increase of the percentage of stricter-delay flows, and then improve slightly when the percentage exceeds 80%. The reason is that when most flows impose stricter latency requirements, the total time

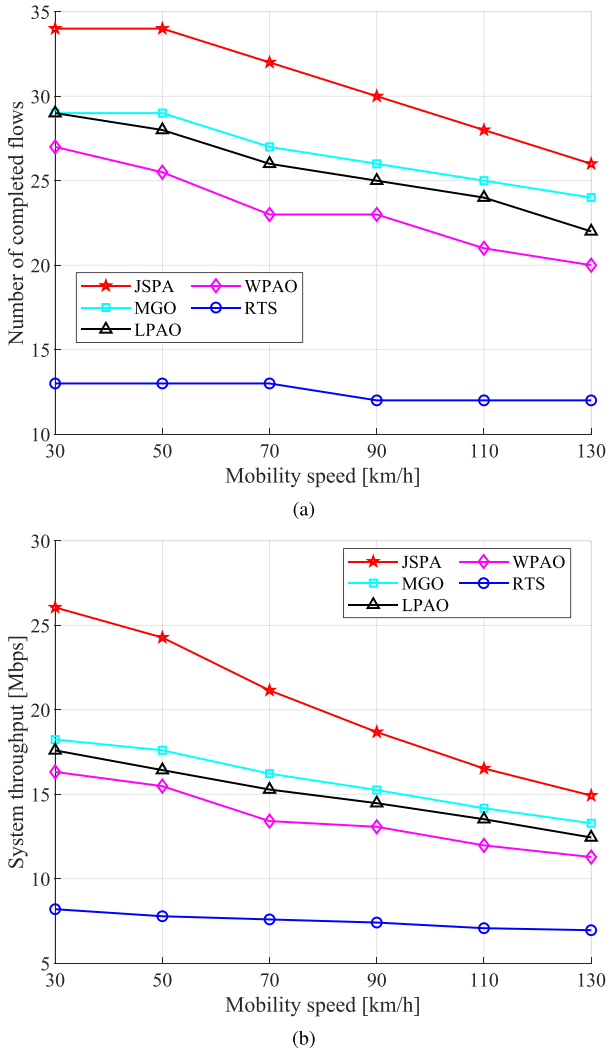


Fig. 6. Comparison of performance as the functions of mobility speed for five schemes: (a) number of completed flows, and (b) system throughput.

consumption drops with less transmitted data, which is slightly beneficial to throughput.

E. Impact of Mobility

Fig. 6 depicts the achievable performance of the five schemes in terms of number of completed flows and system throughput, given different mobility speeds. As expected, our JSPA achieves the best performance in both the number of completed flows and system throughput, indicating its ability to adapt to mobile scenarios. It can also be seen that both the numbers of completed flows and achievable system throughput for JSPA, MGO, LPAO and WFPO degrade significantly with the increase of mobile speed. The reason is that when the Doppler shift enlarges as the speed increases, the RB interference increases, which is detrimental to the network performance. Besides, the performance of the worst RTS scheme degrades slightly as the mobility speed increases, since it can only transmit a small amount of flows at low data rates, thereby less affected by RB interference.

V. Conclusion

This paper has proposed an effective JSPA scheme to maximize the number of successful flows in the downlink V2I OFDM URLLC network. Since this problem is NP-hard, we have derived a tractable framework to decouple it into the subproblems of flow scheduling and power allocation. As the flow scheduling process can be viewed as a MG between unscheduled flows and available FBs, we have proposed the low-complexity DAFS solution. We have also designed the CRA for local power allocation, which fully leverages the quasi-static nature of the vehicular channel and further improves the system performance. Subsequently, the overall optimization algorithm has been summarized as JSPA. Numerical results provided have demonstrated the superiority of the proposed JSPA solution over the existing solutions, in terms of the number of completed flows and system throughput. Particularly, it improves the system performance without increasing the algorithm complexity compared to the classic water-filling power allocation. In future work, we will extend this work along the following directions. 1) Evaluate the impact of imperfect CSI on the proposed scheme and study effective resource allocation for URLLC systems with partial CSI. 2) Design retransmission schemes to improve the reliability of URLLC in mobile systems.

References

[1] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink MISO OFDMA-URLLC systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7184–7200, Nov. 2020.

[2] B. Wang, R. Shi, F. Shi, and J. Hu, "mmWave-NOMA-based low-latency and high-reliable communications for enhancement of V2X services," *IEEE Access*, vol. 8, pp. 57049–57062, 2020.

[3] Z. Dong et al., "Dynamic manager selection assisted resource allocation in URLLC with finite block length for 5G-V2X platoons," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11336–11350, Nov. 2022.

[4] M. Klapez, C. A. Grazia, and M. Casoni, "Application-level performance of IEEE 802.11p in safety-related V2X field trials," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3850–3860, May 2020.

[5] H. Yang, K. Zheng, L. Zhao, and L. Hanzo, "Twin-timescale radio resource management for ultra-reliable and low-latency vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1023–1036, Jan. 2020.

[6] H.-S. Park, Y. Lee, T.-J. Kim, B.-C. Kim, and J.-Y. Lee, "Handover mechanism in NR for ultra-reliable low-latency communications," *IEEE Netw.*, vol. 32, no. 2, pp. 41–47, Mar./Apr. 2018.

[7] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.

[8] S. A. Ashraf, R. Blasco, H. Do, G. Fodor, C. Zhang, and W. Sun, "Supporting vehicle-to-everything services by 5G new radio release-16 systems," *IEEE Commun. Standards Mag.*, vol. 4, no. 1, pp. 26–32, Mar. 2020.

[9] W. Anwar, N. Franchi, and G. Fettweis, "Physical layer evaluation of V2X communications technologies: 5G NR-V2X, LTE-V2X, IEEE 802.11bd, and IEEE 802.11p," in *Proc. IEEE 90th Veh. Technol. Conf.*, Honolulu, HI, USA, 2019, pp. 1–7.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[11] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[12] J. Yan and J. Härri, "On the feasibility of URLLC for 5G-NR V2X sidelink communication at 5.9 GHz," in *Proc. IEEE Glob. Commun. Conf.*, Rio de Janeiro, Brazil, 2022, pp. 3599–3604.

[13] A. Destounis, G. S. Paschos, J. Arnau, and M. Kountouris, "Scheduling URLLC users with reliable latency guarantees," in *Proc. 16th Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw.*, Shanghai, China, 2018, pp. 1–8.

[14] G. J. Sutton et al., "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Surv. Tuts.*, vol. 21, no. 3, pp. 2488–2524, Thirdquarter 2019.
[15] M. E. Haque, F. Tariq, M. R. A. Khandaker, K.-K. Wong, and Y. Zhang, "A survey of scheduling in 5G URLLC and outlook for emerging 6G systems," *IEEE Access*, vol. 11, pp. 34372–34396, 2023.
[16] M. Almekhlafi, M. Chraiti, M. A. Arfaoui, C. Assi, A. Ghrayeb, and A. Alloum, "A downlink puncturing scheme for simultaneous transmission of URLLC and eMBB traffic by exploiting data similarity," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13087–13100, Dec. 2021.
[17] A. K. Bairagi et al., "Coexistence mechanism between eMBB and uRLLC in 5G wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1736–1749, Mar. 2021.
[18] M. Darabi, V. Jamali, L. Lampe, and R. Schober, "Hybrid puncturing and superposition scheme for joint scheduling of URLLC and eMBB traffic," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1081–1085, May 2022.
[19] Z. Li, H. Shen, W. Xu, P. Zhu, and C. Zhao, "Resource allocation for IRS-assisted uplink URLLC systems," *IEEE Commun. Lett.*, vol. 27, no. 6, pp. 1540–1544, Jun. 2023.
[20] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, May 2019.
[21] J. Xue, K. Yu, T. Zhang, H. Zhou, L. Zhao, and X. Shen, "Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks," *IEEE Trans. Mobile Comput.*, early access, Jan. 03, 2024, doi: 10.1109/TMC.2023.3347580.
[22] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1146–1159, Feb. 2020.
[23] R. M. Sohaib, O. Onireti, Y. Sambo, R. Swash, S. Ansari, and M. A. Imran, "Intelligent resource management for eMBB and URLLC in 5G and beyond wireless networks," *IEEE Access*, vol. 11, pp. 65205–65221, 2023.
[24] Q. Chen, H. Jiang, and G. Yu, "Service oriented resource management in spatial reuse-based C-V2X networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 91–94, Jan. 2020.
[25] A. A. Zaidi et al., "OFDM numerology design for 5G new radio to support IoT, eMBB, and MBSFN," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 78–83, Jun. 2018.
[26] W. Kim and B. Shim, "Ultra-mini slot transmission for 5G and 6G URLLC network," in *Proc. IEEE 92nd Veh. Technol. Conf.*, Victoria, BC, Canada, 2020, pp. 1–5.
[27] E. Abdul-Rahman and D. N. Aloi, "Design of a 5G sub-6GHz vehicular cellular antenna element with consistent radiation pattern using characteristic mode analysis," *Sensors*, vol. 22, Nov. 2022, Art. no. 8862.
[28] T. T. d. Almeida, L. d. C. Gomes, F. M. Ortiz, J. G. R. Júnior, and L. H. M. K. Costa, "Comparative analysis of a vehicular safety application in NS-3 and veins," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 620–629, Jan. 2022.
[29] M. Noor-A-Rahim, G. G. M. N. Ali, H. Nguyen, and Y. L. Guan, "Performance analysis of IEEE 802.11p safety message broadcast with and without relaying at road intersection," *IEEE Access*, vol. 6, pp. 23786–23799, 2018.
[30] 3rd Generation Partnership Project, "Study on evaluation methodology of new vehicle-to-everything (V2X) use cases for LTE and NR," 3rd Generation Partnership Project, Sophia Antipolis, France, Tech. Report 37.885 V15.3.0, Jun. 2019.
[31] M. M. Awad, K. G. Seddik, and A. Elezabi, "Low-complexity semi-blind channel estimation algorithms for vehicular communications using the IEEE 802.11p standard," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1739–1748, May 2019.
[32] X. Li, L. Ma, R. Shankaran, Y. Xu, and M. A. Orgun, "Joint power control and resource allocation mode selection for safety-related V2X communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7970–7986, Aug. 2019.
[33] C.-H. Fang, K.-T. Feng, and L.-L. Yang, "Resource allocation for URLLC service in in-band full-duplex-based V2I networks," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3266–3281, May 2022.
[34] Y. S. Cho, J. Kim, W. Y. Yang, and C.-G. Kang, *MIMO-OFDM Wireless Communications With MATLAB*. Hoboken, NJ, USA: Wiley, 2010.
[35] S. Yan, R. Malaney, I. Nevat, and G. W. Peters, "Location verification systems for VANETs in Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5652–5664, Jul. 2016.
[36] J. Gozalvez, M. Sepulcre, and R. Bauza, "IEEE 802.11p vehicle to infrastructure communications in urban environments," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 176–183, May 2012.
[37] J. Wang et al., "Network-ELAA beamforming and coverage analysis for eMBB/URLLC in spatially non-stationary Rician channels," in *Proc. IEEE Int. Conf. Commun.*, Seoul, Korea, 2022, pp. 3508–3513.
[38] M. A. Al-Jarrah, K.-H. Park, A. Al-Dweik, and M.-S. Alouini, "Error rate analysis of amplitude-coherent detection over Rician fading channels with receiver diversity," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 134–147, Jan. 2020.
[39] T. Wang, J. G. Proakis, E. Masry, and J. R. Zeidler, "Performance degradation of OFDM systems due to Doppler spreading," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1422–1432, Jun. 2006.
[40] J. Cheng, C. Shen, and S. Xia, "Robust URLLC packet scheduling of OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Seoul, Korea, 2020, pp. 1–6.
[41] T. Hößler, P. Schulz, E. A. Jorswieck, M. Simsek, and G. P. Fettweis, "Stable matching for wireless URLLC in multi-cellular, multi-user systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5228–5241, Aug. 2020.
[42] Y. Xiao, D. Niyato, K.-C. Chen, and Z. Han, "Enhance device-to-device communication with social awareness: A belief-based stable marriage game framework," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 36–44, Aug. 2016.
[43] E. A. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *Proc. 17th Int. Conf. Digit. Signal Process.*, Corfu, Greece, 2011, pp. 1–8.
[44] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G radio network design for ultra-reliable low-latency communication," *IEEE Netw.*, vol. 32, no. 2, pp. 24–31, Mar./Apr. 2018.
[45] Y. Xu, C. Shen, T. -H. Chang, S. -C. Lin, Y. Zhao, and G. Zhu, "Transmission energy minimization for heterogeneous low-latency NOMA downlink," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1054–1069, Feb. 2020.
[46] L. Li et al., "Resource allocation and computation offloading in a millimeter-wave train-ground network," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10615–10630, Oct. 2022.
[47] S. Sekander, H. Tabassum, and E. Hossain, "Decoupled uplink-downlink user association in multi-tier full-duplex cellular networks: A two-sided matching game," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2778–2791, Oct. 2017.
[48] S. Zhang, B. Di, L. Song, and Y. Li, "Sub-channel and power allocation for non-orthogonal multiple access relay networks with amplify-and-forward protocol," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2249–2261, Apr. 2017.

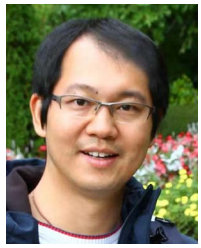
Jing Li received the M.E. degree from Beijing Jiaotong University, Beijing, China, in 2020. She is currently working toward the Ph.D. degree with Beijing Jiaotong University. Her research interests include millimeter wave communications and medium access control.

Yong Niu (IEEE Senior Member) received the B.E. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2011, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2016. He is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. From 2014 to 2015, he visited University of Florida, FL, USA, as a Visiting Scholar. His research interests include networking and communications, including millimeter wave communications, device-to-device communication, medium access control, and software-defined networks. He received the Ph.D. National Scholarship of China in 2015, Outstanding Ph.D. Graduates and Outstanding Doctoral thesis of Tsinghua University in 2016, and Outstanding Ph.D. Graduates of Beijing in 2016. He was Technical Program Committee (TPC) member for CHINACOM 2015 and IWCMC 2017, and also Session Chair for IWCMC 2017 and CHINACOM 2017.

Hao Wu (Member, IEEE) received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2000. She is currently a Full Professor with the State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University (BJTU), Beijing, China. She has authored or coauthored more than 100 papers in international journals and conferences. Her research interests include Intelligent Transportation Systems (ITS), security and QoS issues in wireless networks (VANETs, MANETs and WSNs), wireless communications, and Internet of Things (IoT). She is a member of a Reviewer of its major conferences and journals in wireless networks and security.

Bo Ai (Fellow, IEEE) received the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2002 and 2004, respectively. He is currently a Full Professor and Ph.D. degree Candidate Advisor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China. He is the Deputy Director of the State Key Laboratory of Rail Traffic Control and Safety. He has authored or co-authored six books and published more than 230 academic research papers. He holds 21 invention patents. He is an Institution of Engineering and Technology fellow. He is an Associate Editor for IEEE TRANSACTIONS ON CONSUMER ELECTRONICS and an Editorial Committee Member of *Wireless Personal Communications*.

Tony Q. S. Quek (Fellow, IEEE) received the B.E. and M.E. degrees in electrical and electronics engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008. He is currently the Cheng Tsang Man Chair Professor with Singapore University of Technology and Design (SUTD), Singapore, and ST Engineering Distinguished Professor. He is also the Director of the Future Communications R&D Programme, the Head of ISTD Pillar, and the Deputy Director of the SUTD-ZJU IDEA. His current research interests include wireless communications and networking, network intelligence, non-terrestrial networks, open radio access network, and 6G. Dr. Quek has been actively involved in organizing and chairing sessions, and was a member of the Technical Program Committee as well as symposium chairs in a number of international conferences. He is currently an Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Dr. Quek was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards - Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, and the 2022 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the Academy of Engineering Singapore.



Ning Wang (Member, IEEE) received the B.E. degree in communication engineering from Tianjin University, Tianjin, China, in 2004, the M.A.Sc. degree in electrical engineering from The University of British Columbia, Vancouver, BC, Canada, in 2010, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2013. From 2004 to 2008, he was with the China Information Technology Design and Consulting Institute, as a Mobile Communication System Engineer, specializing in planning and design of commercial mobile communication networks, network traffic analysis, and radio network optimization. From 2013 to 2015, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia. Since 2015, he has been with the School of Information Engineering, Zhengzhou University, Zhengzhou, China, where he is currently an Associate Professor. He also holds adjunct appointments with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, and the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. His research interests include resource allocation and security designs of future cellular networks, channel modeling for wireless communications, statistical signal processing, and cooperative wireless communications. He has served on the technical program committees of international conferences, including IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and CyberC. He was on the Finalist of the Governor Generals Gold Medal for Outstanding Graduating Doctoral Student with the University of Victoria in 2013.

Sheng Chen (Fellow, IEEE) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, U.K., in 1986, and the higher Doctoral (D.Sc.) degree from the University of Southampton, Southampton, U.K., in 2005. He held research and academic appointments with the University of Sheffield, Sheffield, U.K., the University of Edinburgh, Edinburgh, U.K., and the University of Portsmouth, Portsmouth, U.K., from 1986 to 1999. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he is currently a Professor of intelligent systems and signal processing. He has authored more than 700 research papers. He have more than 19 000 Web of Science citations with h-index 60, and more than 37 000 Google Scholar citations with h-index 82. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization. Dr Chen was elected to a Fellow of the United Kingdom Royal Academy of Engineering in 2014. He is a Fellow of Asia-Pacific Artificial Intelligence Association (FAAIA), a Fellow of IET, and an original ISI Highly Cited Researcher in engineering in March 2004.