Energy Efficiency Optimization of Ultra-Reliable Low-Latency Communication for High-speed Rail

Yuanyuan Qiao, Yong Niu, *Senior Member, IEEE*, Sheng Chen, *Life Fellow, IEEE*, Zhangdui Zhong, *Fellow, IEEE*, Changming Zhang, *Member, IEEE*, Ning Wang, *Member, IEEE*, Bo Ai, *Fellow, IEEE*

***Abstract*—With the rapid development of new communication technology, notably ultra reliable low latency communication (URLLC) in the 5th generation cellular network, and the wide deployment of high-speed rail (HSR), it becomes critically important to enhance the energy efficiency of the HSR communication system by optimally allocating the bandwidth and transmit power for users, while ensuring the quality of service (QoS) requirements of URLLC. In this paper, we tackle this challenging resource allocation problem of the URLLC system for HSR. Specifically, we establish the train-ground URLLC model for the HSR communication system with mobile relays (MRs), based on which we formulate the optimization problem with the QoS requirements of URLLC as constraints for maximizing the system's energy efficiency. By decomposing this challenging optimization problem into two subproblems, a heuristic algorithm is adopted to optimally allocate the transmit power and bandwidth of users with the block coordinate descent (BCD) approach. The simulation results show that compared with the existing algorithms, the proposed algorithm achieves better performance in resource allocation optimization of user bandwidth and transmit power.**

***Index Terms*—Ultra-reliable low-latency communication, high-speed railway, energy efficiency optimization, resource allocation.**

I. Introduction

With the technological progress and rapid deployment, high-speed rail (HSR) has been widely spread to connect cities of all sizes across some countries, such as in China. With its features of fast speed, comfortable experience, safety and economy, HSR plays an important role in people's life [1]. In the post COVID-19 pandemic era, people's demand for travel has increased dramatically, and they expect to get the same excellent communication experience on HSR as on the ground. In the Internet era, a variety of applications, including mobile games, telecommuting and telemedicine, demand ultra reliable and low delay services from the communication network. People hope that they can experience these applications without obstacles on HSR. However, communication interruption and high power consumption of mobile devices currently experienced in high-speed scenarios [2] mean that this urgent requirement could not be met yet, and this fuels the current research effort on HSR communications.

Ultra reliable low latency communication (URLLC) is one of the three application scenarios or usage cases of the 5th generation (5G) cellular network technology [3]. URLLC is a key feature that makes 5G qualitatively different from the previous generations of mobile communication technology. High reliability and low latency requirements of URLLC become a potential promoter of a large number of new emerging applications [4]. For example, URLLC plays a key role in enabling mission-critical applications, such as factory intelligence, vehicle-to-vehicle communications, and remote surgery [5, 6]. These applications require ultra low communication delay and high reliability [7]. Overall expectations of 3GPP [7] for URLLC requirements are: 1) Overall reliability requirement of 99.9999% and radio delay of 1 ms for the user plane are met; 2) The average transmission delay of the user plane of the uplink and downlink are less than 0.5 ms.

In order to ensure strict end-to-end (E2E) delay, transmission delays, queuing delays, encoding and processing delays, and delays in backhaul and routing need to be taken into account in the uplink and downlink [8]. To meet URLLC requirement of low delay, short packet communication becomes necessary, as the main way to reduce delay is to adopt packets with short frame structure [7]. In short packet communication, the relationship among the achievable communication rate, decoding error probability, and transmission delay becomes different from that in long packet communication, and the familiar Shannon formula is no longer applicable [9]. This is because using Shannon formula to analyze the system performance in short packet communication will underestimate the error of reliability and delay, leading to the failure of meeting the quality of service (QoS) requirements of URLLC [10]. In a nutshell, in short packet communication, 1) the classical information theory is no longer applicable because the law of large number cannot be applied; and 2) the size of the control information (metadata) in data packets is similar to the size of the payload, and the inefficiency of metadata encoding will significantly affect the overall efficiency of transmission [9]. Therefore, it is a huge challenge to efficiently allocate short packet communication resources for URLLC

Yuanyuan Qiao is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China, and also with Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications, Beijing Jiaotong University, Beijing 100044, China (email: qiaoyuanyuan@bjtu.edu.cn).

Yong Niu is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China (email: niuy11@163.com).

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

Zhangdui Zhong is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: zhdzhong@bjtu.edu.cn).

Changming Zhang is with the Institute of Intelligent Network, Zhejiang Lab, Hangzhou 311121, China (e-mail: zhangcm@zhejianglab.com).

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

Bo Ai is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China, and also with Henan Joint International Research Laboratory of Intelligent Networking and Data Analysis, Zhengzhou University, Zhengzhou 450001, China (email: boai@bjtu.edu.cn).

applications of users, particularly in high-speed scenarios.

Against this background, in this paper, we address the challenging task of the optimization of URLLC resource allocation for the application to the HSR communication assisted by mobile relays (MRs), where an MR is deployed in the middle of the roof top of each train carriage to relay the communications between the remote radio heads (RRHs) and the users in the carriage, in order to mitigate penetration loss. Specifically, our goal is to maximize the system's achievable energy efficiency. Since this optimization problem is NP-hard, we decompose the optimization problem into two sub-problems, which can then be solved by the block coordinate descent (BCD) approach based on a heuristic algorithm that is widely used in solving communication resource allocation problems [11]. The main contributions of this paper are summarized as follows.

- We establish the train-ground URLLC model for the HSR with MRs by introducing the directional antenna model, millimeter wave (mmwave) communication model in 3GPP protocol [12], and the traffic model of users as well as the QoS requirements of URLLC. This is different from previous work related to HSR train-ground communication and URLLC resource allocation, and to the best of the authors' knowledge, this paper is the first to consider them together.
- Based on the proposed train-ground URLLC model, we establish the system's energy efficiency model and formulate the optimization problem for maximizing the system's energy efficiency. A heuristic algorithm with BCD is proposed to solve the problem of optimal bandwidth and transmit power allocation of MR to users. First, the optimization problem is decomposed into two subproblems to address the coupling between them, which are solved by the heuristic algorithm with BCD, respectively. Then, the bandwidth and transmit power are updated alternately, to obtain the final solution. The heuristic algorithm adopted in this paper can effectively solve search problems with high complexity. In addition, its better robustness can adapt to the changing driving environment of HSR.
- By configuring the network with different parameters, we compare the proposed algorithm with other existing schemes from multiple perspectives. Simulation results show that our algorithm has better resource allocation ability, and it is capable of improving the energy efficiency of the system significantly by allocating the communication bandwidth and transmit power to the personalized service users.

The rest of this paper is organized as follows. Section II reviews the related work. Section III introduces the train-ground communication model with MRs for URLLC, and formulates the optimization problem for maximizing the system's energy efficiency. Section IV presents the heuristic algorithm with BCD to solve the problem of optimal bandwidth and transmit power allocation to users. In Section V, the performance comparison of the proposed algorithm and other existing algorithms is conducted under different system parameters. Finally, Section VI draws the conclusions of this paper.

II. Related Work

Currently, there exist rich literature on energy efficiency (EE) optimization of HSR communication [13–16]. For the mmwave based HSR communication system with multiple MRs on top of train, Wang et al. [13] proposed a dynamic power control scheme for train-ground communication that minimizes energy consumption under the constraints of transmit data and transmit power budgets. By embedding power adjustment into the existing communication switching process, Lu et al. [14] studied the influence of power adjustment on the switching performance in the HSR communication system and showed that the performance can be improved without additional energy consumption by appropriate power adjustment. Jiang et al. [15] studied the non-orthogonal multiple access (NOMA) based HSR communication system and optimized the uplink EE of the NOMA system by considering the QoS and maximum transmission power constraints. Hu et al. [16] investigated a joint transmission mode selection and power optimization scheme to maximize the EE of the distributed antenna based HSR communication system and showed that on the one hand, the system's EE can be maximized while meeting the rate requirements of the users on train by power optimization, and on the other hand, the transmission mode selection strategy can be used to further improve the system's achievable EE. However, the aforementioned works do not consider the application of URLLC to HSR, which is an important usage case for the vertical business development of 5G.

The resource allocation in URLLC is vital for the effective use of communication resources and the improvement of EE [17–20]. Deng et al. [17] proposed a hybrid resource allocation method based on NOMA technology to meet the ultra reliability and low delay requirements of emerging URLLC services. Specifically, the system adopts NOMA technology to share URLLC user resources to improve the utilization of frequency resources. When the shared resources cannot support the transmission requirements of URLLC users, private resources are used for transmission. Ghanem et al. [18] considered the resource allocation for the downlink multi-input single-output orthogonal frequency division multiple access URLLC system. A low complexity sub-optimal resource allocation algorithm was designed based on successive convex approximation and difference of convex programming to maximize the weighted system throughput, constrained by the QoS requirements of the number of transmitted bits, packet error probability, and latency for URLLC users. Chen et al. [19] studied a unmanned aerial vehicle (UAV)-assisted URLLC service system. The average uplink transmit power is minimized by jointly optimizing the device scheduling and association, power control and resource allocation as well as UAV deployment, using an iterative algorithm based on BCD with Lagrange dual decomposition. Simulation results presented in [19] show that a performance gain of 15%-20% can be achieved in the average transmit power of URLLC. Based on uplink transmission, Chang et al. [20] investigated the resource allocation problem of URLLC in real-time wireless control system. The problem is solved by optimizing bandwidth and transmission power

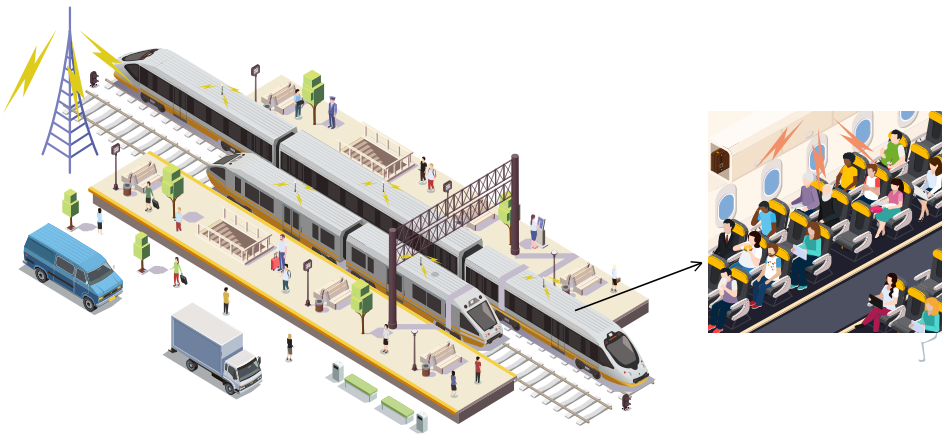

Fig. 1: Train-ground URLLC system in HSR scenario with MRs.

allocation in URLLC, and controlling the convergence rate subject to communication and control constraints. Compared with traditional URLLC methods that satisfy fixed QoS, the method of [20] can adjust the optimal spectrum allocation to maximize the efficiency of the communication spectrum, leading to significant performance improvement in spectrum efficiency and control cost. However, the aforementioned works do not consider the resource allocation of URLLC in the HSR scenario. By contrast, our work specifically considers the EE optimization for URLLC in the application to HSR with MR, and we design a resource allocation scheme of bandwidth and power for HSR users under the URLLC QoS constraints on delay and reliability.

Table I shows the difference between this paper and related studies, in which the literature [13, 15, 16, 21, 22] studied the resource allocation of HSR train-ground communication, but none of them considered network conditions of URLLC, and the vast majority of them did not consider the Doppler effect caused by high speed. Literature [17-20] studied resource allocation with URLLC, but all of them were at the static condition. In this paper, we develop a model for HSR train-ground communication with URLLC, taking into account the impact of the Doppler effect due to high speed and considering the user's personalized network services, which is the first time according to the authors.

III. System Overview And Problem Formulation

In this section, we first introduce the train-ground URLLC system for HSR with MRs. Next we describe the mmwave communication model, URLLC rate, traffic model, and QoS requirements. Then the EE model is derived, which forms the optimization objective. Finally, we formulate the resource allocation problem for maximizing the system's EE.

A. System Description

URLLC is one of the three usage cases in 5G, where the classical information theory is inadequate to describe the relationship of communication rate, error probability and transmission delay [9]. As mentioned previously, current URLLC studies mostly focus on static scenarios. In this paper, we study the train-ground URLLC system for HSR assisted by MRs. Our goal is to maximize the system's EE through the bandwidth and transmit power resource allocation to users.

Fig. 1 illustrate the conceived train-ground URLLC system operated at the mmwave band for HSR, where an MR is deployed at the middle of the top of each train carriage to relay/serve the users in the carriage. The left picture of Fig. 1 depicts the HSR communication system in which RRH communicates with MR. The right picture of Fig. 1 sketches the scenario of URLLC communication between the MR and the users inside the carriage. The system adopts the single frequency network (SFN), in which multiple RRHs send and receive the signals at the same frequency in a cell [23]. SFN has the advantages of small interference from adjacent cells, large coverage area of cells and low signal switching frequency, and it is suitable for HSR applications. Since RRH communicates with MR at the mmwave band, omnidirectional antenna is unsuitable, as it has high energy consumption and is incapable of compensating for large path loss. Therefore, directional antenna with unidirectional coverage as defined in 3GPP 38.913 [7] is deployed at RRH to communicate with MR. This unidirectional antenna based HSR SFN communication system is shown in Fig. 2, which is the only model considered for HSR 30 GHz deployment in 3GPP [7]. In order to overcome the severe penetration loss of the carriage, MR is deployed to act as the intermediate medium for the communication between the RRH and the users [24].

In the model of [7] depicted in Fig. 2, v is the speed of the train, $D_{\mathrm{RRH\_rail}}$ is the horizontal distance between RRH and the rail, and D_{RRH} is the horizontal distance between adjacent RRHs, while h_T is the height of MR at the top of

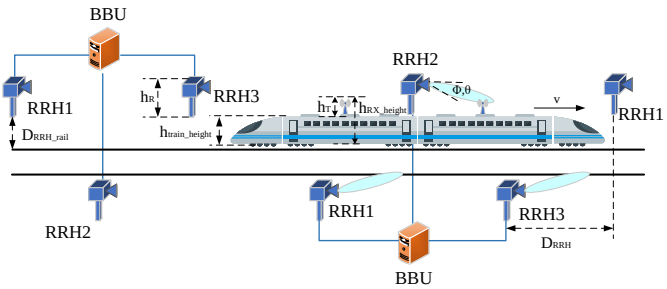

Fig. 2: HSR SFN communication system with unidirectional antenna.

TABLE I
DIFFERENCES BETWEEN THIS PAPER AND RELATED WORKS

	Research scene	Object	Algorithm	Motion state	QoS of URLLC	Doppler	User-Defined Services
this paper	A two-stage HSR train-to-ground communication MR-assisted model for user-defined services	Max(EE)	Heuristic Optimization Algorithm Based on BCD for Bandwidth and Transmit Power	High speed	Maximum latency: 1 ms Maximum system error: 3e-7	Yes	Yes
[13]	A control/user-plane splitting network for HSR communications with MRs	Min(power)	The Multiplier Punitive Function-Based Power Control Algorithm	High speed	No	Yes	No
[15]	A high-speed railway transmission model of two-hop NOMA uplink	Max(EE)	The optimal EE search iterative algorithm based on Dinkelbach method	High speed	No	No	No
[16]	HSR in a single-cell with distributed antenna transmission system model	Max(EE)	A energy-efficient power optimization with transmission mode selection scheme	High speed	No	No	No
[21]	An mm-wave train-ground communication system using FD MRs	Max(capacity)	A sequential quadratic programming algorithm based on Lagrangian function	High speed	No	No	No
[22]	An mm-wave train-ground communication system using FD MR	Max(flows)	Coalition Game Based Algorithm for User Association and Transmission Scheduling	High speed	No	No	No
[17]	An uplink URLLC user transport system with N users	Max(rate)	A hybrid resource allocation method based on nonorthogonal multiple access	Static	Reliability: 99.999 %	No	No
[18]	A system model with N-antenna BS and K single-antenna users	Max(throughput)	A sub-optimal resource allocation algorithm based on successive convex approximation.	Static	Error tolerance: 0.01 Packet error probability: 1e-6	No	No
[19]	A UAV-assisted IoT network in a URLLC service scenario	Min(power)	A block coordinate descent optimization algorithm (BCDOA)	Static	E2E delay threshold: 1ms Decoding error threshold: 1e-5	No	No
[20]	A typical centralized real-time wireless communication-control system	Max(SE)	An iterative algorithm for optimal resource allocation	Static	Maximum time delay: 0.5 ms Packet error probability: 1e-5	No	No

the train, h_R is the height of RRH, and $h_{\text{train\_height}}$ is the height of train. Hence the height of receive antenna at top of the train is $h_{\text{RX\_height}} = h_T + h_{\text{train\_height}}$. Furthermore, Φ and θ denote the beam direction of RRH. As illustrated in Fig. 2, every three RRHs, two located at one side of the rail and the other at the other side, are connected to the same baseband unit (BBU) through optical fibers, and this pattern is alternatively deployed along the entire length of the track. The orientation of the RRH panel is towards the adjacent rail track, and therefore the RRH's beam can always be aligned with the MR's beam [23, 25].

B. Mmwave Communication Model

According to Fig. 2 and the link budget, the receive power of MR, denoted by P_r, can be expressed as [23]

$$\begin{aligned} P_r(x,\theta_E,\phi_E) =& P_t - PL(x) + A_B^t(\theta_E,\phi_E) + A_B^r(\theta_E,\phi_E) \\ &+ A_E^t(\theta_E,\phi_E) + A_E^r(\theta_E,\phi_E), \end{aligned} \tag{1}$$

where P_t is the transmit power of the RRH, and $PL(x)$ is the path loss (PL) with x being the distance between the RRH and the MR. The beamforming gain in mmwave is composed of antenna element gain A_E and composite array radiation gain A_B. The antenna element gain in (1) includes the transmit antenna (i.e., RRH antenna) gain A_E^t and the receive antenna (i.e., MR antenna) gain A_E^r. The composite array radiation gain in (1) includes the RRH's composite array radiation gain A_B^t and the MR's composite array radiation gain A_B^r. The values of the four antenna gains, A_E^t, A_E^r, A_B^t and A_B^r, depend on the down-tilting angle θ_E and azimuth angle ϕ_E of the vector linking the antenna elements of the MR and RRH.

Since the HSR track is mainly over flat terrain and there is no blockage between RRH and train, the line of sight transmission is dominated and the PL model can be considered to be [12]:

$$PL(d_{2D}) = \begin{cases} PL_1(d_{2D}), & 10\,\text{m} \le d_{2D} \le d_{BP}, \\ PL_2(d_{2D}), & d_{BP} \le d_{2D} \le 10\,\text{km}, \end{cases} \tag{2}$$

where

$$\begin{aligned} PL_1(d_{2D}) =& 20\log_{10}\left(\frac{40\pi \cdot d_{3D} \cdot f_c}{3}\right) \\ &- \min\left\{0.044h^{1.72}, 14.77\right\} \\ &+ \min\left\{0.03h^{1.72}, 10\right\} \cdot \log_{10}(d_{3D}) \\ &+ 0.002\log_{10}(h) \cdot d_{3D}, \end{aligned} \tag{3}$$

$$PL_2(d_{2D}) = PL_1(d_{BP}) + 40\log_{10}\left(\frac{d_{3D}}{d_{BP}}\right). \tag{4}$$

The definitions of d_{2D} and d_{3D} in (2) to (4) are given in Fig. 3, while d_{BP} in (2) and (4) is the breakpoint distance, given by

$$d_{BP} = \frac{2\pi \cdot h_R \cdot h_{\text{RX\_height}} \cdot f_c}{c}, \tag{5}$$

in which f_c is the central frequency in Hz and c is the velocity of light (3×10^8 m/s). Furthermore, h in (3) represents the average height of buildings in HSR communication environment and hence $h \in [5,\ 50]$ m.

In this paper, the calculation of antenna element gain A_E

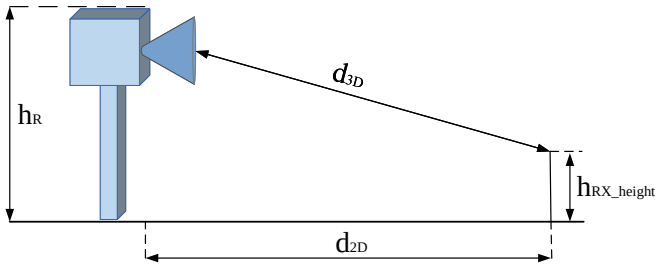

Fig. 3: Definition of d_{2D} and d_{3D}.

can be expressed as (6), (7) and (8) according to [12]:

$$A_E\left(\theta'_E, \phi'_E = 0^\circ\right) = -\min\left\{12\left(\frac{\theta'_E - 90^\circ}{\theta_{3dB}}\right)^2, SLA_V\right\}, \tag{6}$$

$$A_E\left(\theta'_E = 90^\circ, \phi'_E\right) = -\min\left\{12\left(\frac{\phi'_E}{\phi_{3dB}}\right)^2, A_{\max}\right\}, \tag{7}$$

$$\begin{aligned} A_E\left(\theta'_E, \phi'_E\right) =& G_{\max} - \min\Big\{-\Big(A_E\left(\theta'_E, \phi'_E = 0^\circ\right) \\ &+ A_E\left(\theta'_E = 90^\circ, \phi'_E\right)\Big), A_{\max}\Big\}, \end{aligned} \tag{8}$$

where θ'_E and ϕ'_E are the results in the global coordinate system that are converted from θ_E and ϕ_E in the local coordinate system [12], θ_{3dB} is the vertical 3 dB beamwidth and ϕ_{3dB} is the horizontal 3 dB beamwidth, while SLA_V is the vertical side-lobe attenuation, $A_{\max}$ is the maximum attenuation and $G_{\max}$ is the maximum directional again of an antenna element. Furthermore, $A_E\left(\theta'_E, \phi'_E = 0^\circ\right)$ in (6) and $A_E\left(\theta'_E = 90^\circ, \phi'_E\right)$ in (7) represent the values of the vertical and horizontal cuts of the radiant power pattern, respectively, while $-\min\Big\{-\Big(A_E\left(\theta'_E, \phi'_E = 0^\circ\right) + A_E\left(\theta'_E = 90^\circ, \phi'_E\right)\Big), A_{\max}\Big\}$ in (8) is the attenuation gain of an antenna element.

The composite array radiation gain A_B can be expressed as

$$A_B(\theta_E, \phi_E) = 10\log_{10}\left(\left|\sum_{m=1}^{N_H}\sum_{n=1}^{N_V} w_{n,m} \cdot v_{n,m}\right|^2\right), \tag{9}$$

where N_H is the number of antenna elements on the panel and N_V is the number of antenna elements with the same polarization in each column, while $w_{n,m}$ and $v_{n,m}$ are the weight and super position vectors, expressed respectively as

$$\begin{aligned} w_{n,m} =& \frac{1}{\sqrt{N_H N_V}} \exp\Bigg(\mathrm{j}\cdot 2\pi\Bigg(\frac{(n-1)\cdot d_V \cdot \sin(\theta_B)}{\lambda} \\ &+ \frac{(m-1)\cdot d_H \cdot \cos(\theta_B)\cdot \sin(\phi_B)}{\lambda}\Bigg)\Bigg), \end{aligned} \tag{10}$$

$$\begin{aligned} v_{n,m} =& \exp\Bigg(\mathrm{j}\cdot 2\pi\Bigg(\frac{(n-1)\cdot d_V \cdot \cos(\theta_E)}{\lambda} \\ &+ \frac{(m-1)\cdot d_H \cdot \sin(\theta_E)\cdot \sin(\phi_E)}{\lambda}\Bigg)\Bigg), \end{aligned} \tag{11}$$

in which $\mathrm{j} = \sqrt{-1}$ denotes the imaginary axis, λ is the wavelength, d_V and d_H are the distances between the neighbouring antenna elements in the vertical and horizontal directions, respectively. In this paper, we set $d_V = d_H = \frac{\lambda}{2}$. Furthermore, θ_B and ϕ_B in (10) are the beam direction, which includes the beam direction θ_B^t and ϕ_B^t for the RRH, and beam direction θ_B^r and ϕ_B^r for the MR.

The main focus of this paper is to design an allocation scheme of bandwidth and power that maximizes the energy efficiency for users with different network requirements, while meeting the constraints of URLLC on delay and reliability. We assume that an MR is in the middle of the top of the train, and set up an RRH to communicate with the MR, to analyze the resource allocation scheme of bandwidth and power for different users. There have been many studies on the beam alignment of high speed rail, e.g., [23, 24]. Therefore, we assume that the beams between the RRH and MR are always aligned, i.e., θ_B and ϕ_B are known and they are not the design variables, to simplify the model.

In order to quantify the ICI power P_{ICI} due to Doppler expansion caused by the high speed of the train, we adopt the widely used ICI approximation model [24], as shown in the following equation

$$P_{ICI} = \int_{-1}^{1} (1 - |\tau|)\, \mathrm{J}_0\left(2\pi f_{d,\max} T_s \tau\right) d\tau, \tag{12}$$

where T_s denotes the symbol duration, and $\mathrm{J}_0(\cdot)$ is the first type of zero-order Bessel function. $f_{d,\max} = v \cdot f_c / c$ represents the maximum Doppler expansion, where v is the speed of the train. f_c and c represent the carrier frequency and the speed of light, respectively.

C. URLLC Communication Rate

The communication capacity described by Shannon formula is widely used to express the traditional communication service rate. However, the impact of decoding error in URLLC is more prominent and cannot be ignored, because of the short packet structure [7]. In short packet communication, the mathematical relationship among the achievable communication rate, transmission delay and decoding error probability is different from that in traditional long packet communication [9]. In addition, MR is relatively stationary with respect to the user in the carriage. Therefore, even when the train is running at high speed, the communication channel between MR and user is an interference-free single antenna system affected by the quasi-static flat fading channel. With short packet length, the maximum achievable rate of user k can be accurately approximated as [26]

$$r_k \approx \frac{\tau W_k}{u \ln 2}\left(\ln\left(1 + \frac{\alpha_k g_k P_k}{N_0 W_k + P_{ICI}}\right) - \sqrt{\frac{V_k}{\tau W_k}} Q_G^{-1}(\varepsilon_k^c)\right), \tag{13}$$

where W_k and P_k are the bandwidth and transmit power allocated by the MR for downlink transmission to user k. α_k is the large scale channel gain including the log distance path loss with exponent n and the log-normal shadow fading with mean 0 and standard deviation σ [27], that includes the gain of

URLLC user G_{ue} and the gain of URLLC antenna $G_{antenna}$. The log distance path loss model is defined as $L(d) = L(d_0)+10n\lg\left(\frac{d}{d_0}\right)$ and $L(d_0)$ is $32.45+20\lg(f)+20\lg(d_0)$ that f is the signal transmission frequency, d_0 is the reference distance, and d is the communication transmission distance [28]. g_k is the small scale channel gain with Nakagami-m fading model [29], and the probability density function is given by $f_{g_k}(z) = \frac{m^m z^{m-1}}{\Gamma(m)}\exp(-mz)$ [30]. u is the number of bytes in a packet. N_0 is the single side noise power spectral density (PSD), $\tau \in (0,\ T_f)$ is the time duration that can be used for downlink transmission in a frame of the length T_f, ε_k^c is the transmission error probability of user k, $Q_G^{-1}(x)$ is the inverse of the Gaussian Q function, and V_k is the channel dispersion given by [26].

$$V_k = 1 - \frac{1}{\left(1+\frac{\alpha_k g_k P_k}{N_0 W_k + P_{ICI}}\right)^2} \approx 1. \tag{14}$$

When the signal-to-noise ratio (SNR) at the receiver is higher than 5 dB, V_k in (14) is approximately equal to 1 [31], which is satisfied in URLLC communication most of the time. On the other hand, under the condition of SNR lower than 5 dB, $V < 1$. But if we substitute $V = 1$ into the URLLC rate expression (13), we obtain the lower limit of the rate that URLLC can achieve. If this lower limit is applied to the subsequent analysis of resource allocation, the reliability and latency requirements can be met [32].

D. Traffic Model

In the system traffic model, the time is divided into frames [33]. The duration of each frame is denoted as T_f. It is assumed that user k has A_k apps connected through the network, that is, applications consuming network traffic, and the user activates apps with probability κ in a frame, wherein the activation of each app is independent and identically distributed (i.i.d.). Therefore, the packet arrival process of each user is modeled as a Poisson process with an average arrival rate of $\lambda_k = A_k\kappa$ packets/frame [34]. Hence, there exists a set that contains a series of samples, each of which is the number of apps used by a user, and these samples conform to the Gaussian distribution.

E. QoS Requirements

As illustrated in Fig. 4, the system delay $D_{\max}$ consists of the UL transmission delay D_{UL}, the mmwave delay of RRH-MR D_{mmwave}, the queue delay D_{queue} and the DL transmission delay D_{DL}, which can be expressed as

$$D_{\max} = D_{UL} + D_{DL} + D_{mmwave} + D_{queue}. \tag{15}$$

Since the packet size is very small, e.g., 20 bytes [7], we assume that a packet transmission of UL and DL can be completed in one frame with a given probability of error without retransmission [35]. Therefore, the queue delay D_{queue} that ensures the system delay, denoted as $D^q_{\max}$, is given by

$$D^q_{\max} = D_{\max} - T_f - D_{mmwave}. \tag{16}$$

For simplicity, we assume that the mmwave delay satisfies $D_{mmwave} = 0.5$ ms [36].

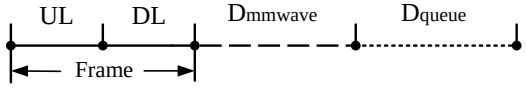

Fig. 4: System delay model.

If the queuing delay of a packet is greater than the delay limit $D^q_{\max}$, the packet is discarded, and the queuing delay violation probability is denoted as ε^q_k. In order to meet the queuing delay requirement of limited transmission power, the proactive packet dropping mechanism [37] can be applied, and the proactive packet dropping probabilities is expressed as ε^h_k. These two probabilities together with the transmission error probability ε^c_k of user k should satisfy the condition

$$\varepsilon^c_k + \varepsilon^q_k + \varepsilon^h_k \le \varepsilon_D, \tag{17}$$

to ensure the overall reliability of URLLC, where ε_D is the maximum tolerable error rate required to ensure the overall reliability of URLLC [35].

To ensure that the queue delay and queue delay violation probability, $D^q_{\max}$ and ε^q_k, as well as the DL transmission error probability, ε^c_k, meet the QoS of URLLC, we introduce the concept of effective bandwidth [38]. The work [37] indicates that the effective bandwidth can be used to analyze the queuing delay of Poisson process at the transmitter if the queuing delay violation probability is small. Since the queuing delay in URLLC is usually shorter than the channel coherence time, the service rate is constant [37]. For Poisson arrival process λ_k, the effective bandwidth can be expressed as [37, 38]

$$E^B_k = \frac{T_f \ln\left(\frac{1}{\varepsilon^q_k}\right)}{D^q_{\max}\ln\left(1+\frac{T_f\ln\left(\frac{1}{\varepsilon^q_k}\right)}{\lambda_k D^q_{\max}}\right)}. \tag{18}$$

To meet the requirement of $(D^q_{\max}, \varepsilon^q_k)$, the constant packet service rate should not be less than the effective bandwidth [38], i.e., $r_k \ge E^B_k$. Substituting (13) into (18) leads to

$$\gamma_k \ge \exp\left(\frac{E^B_k u \ln 2}{\tau W_k} + \frac{Q_G^{-1}(\varepsilon^c_k)}{\sqrt{\tau W_k}}\right) - 1, \tag{19}$$

where the SNR γ_k is defined as $\gamma_k = \frac{\alpha_k g_k P_k}{N_0 W_k}$.

In order to ensure that the proactive packet dropping probabilities ε^h_k meets the requirement of URLLC, according to the derivation in [35, 37], ε^h_k can be approximated as

$$\varepsilon^h_k \approx B_{N_t}\left(g^{th}_k\right), \tag{20}$$

where g^{th}_k and $B_{N_t}\left(g^{th}_k\right)$ are given respectively by [35]

$$g^{th}_k = \frac{(N_0 W_k + P_{ICI})\gamma_k}{\alpha_k P^{th}_k}, \tag{21}$$

$$B_{N_t}\left(g^{th}_k\right) = \int_0^{g^{th}_k}\left(1 - \frac{\ln\left(1+\frac{g\gamma_k}{g^{th}_k}\right)}{\ln(1+\gamma_k)}\right) f_{N_t}(g)\,dg, \tag{22}$$

with $f_{N_t}(g) = \frac{g^{N_t-1}}{(N_t-1)!}\exp(-g)$.

An upper limit of the approximate proactive packet dropping probability $B_{N_t}\left(g_k^{th}\right)$ is given by [35]

$$F_{N_t}\left(g_k^{th}\right) = (N_t-1)\int_0^{g_k^{th}}\left(\frac{1}{g}-\frac{1}{g_k^{th}}\right)f_{N_t}(g)dg = \left(1-\frac{N_t-1}{g_k^{th}}\right)\sum_{n=0}^{N_t-2} f_{n+1}\left(g_k^{th}\right)+f_{N_t-1}\left(g_k^{th}\right). \quad (23)$$

Hence $F_{N_t}\left(g_k^{th}\right)$ is taken as the QoS constraint on the proactive packet dropping probability ε_k^h. It is assumed that there are enough antennas to meet the requirements of URLLC. It is shown in [37] that $\varepsilon_k^c = \varepsilon_k^q = \varepsilon_k^h = \frac{\varepsilon_D}{3}$ is an approximate optimal combination to ensure (17).

F. Energy Efficiency Model

The total power consumption P_{tot} of the aforementioned HSR URLLC communication model is composed of the transmit power of RRH P_{BS} and the transmit power of MR P_{MR} as well as a fixed circuit power consumption P_0^C, which can be expressed as

$$P_{tot} = \frac{1}{\rho}\left(P_{BS}+P_{MR}\right)+P_0^C, \quad (24)$$

where $\rho \in (0,\ 1]$ is the power amplifier efficiency [35]. P_{MR} includes the transmit power allocated to all the users, namely, $P_{MR} = \sum_{k=1}^{K} P_k$, and P_{BS} satisfies $R = \left(1-P_{l_{BS,MR}}^{out}\right) W\log(1+\frac{\alpha g P_{BS}}{W N_0})$, in which W is the communication bandwidth between RRH and MR. α is the large scale channel gain, as shown in (2). g is the small scale channel gain with rayleigh channel model [39]. N_0 is single-side noise PSD. $P_{l_{BS,MR}}^{out}$ represents the interruption probability between the transmitter and the receiver during the mmwave transmission, and it can be denoted as $P_{a,b}^{out} = 1-\exp\left(-\beta l_{ab}\right)$, where l_{ab} is the distance between the transmitter a and the receiver b, and β is a parameter that reflects the density and size of obstacles.

Therefore, the energy efficiency model can be expressed as

$$\eta = \frac{(1-\varepsilon_D)\left(\frac{u\sum_{k=1}^{K}\lambda_k}{T_f}\right)\cdot t}{\int_0^t P_{\text{tot}}dt}. \quad (25)$$

The energy efficiency η of (25) is defined as the ratio of the successful communication byte size of all the users to the energy consumed between two adjacent location bins. Let the distance between two adjacent location bins be σ_D. Then the traveling time t between two adjacent location bins in (25) is $t = \frac{\sigma_D}{v}$. In order to simplify the analysis, when we study resource allocation, the bandwidth and transmit power of users are optimized every σ_D for the entire travel distance.

G. Model Analysis

In this paper, we optimize resource allocation to maximize the energy efficiency η while meeting the QoS constraints of URLLC. Note that the communication rate of user k is determined according to the Poisson arrival process of packets obtained from the activation of apps, which is almost unaffected by resource allocation. In addition, due to the requirements of URLLC for high reliability and low latency, $1-\varepsilon_D \approx 1$. Therefore, 'maximum energy efficiency' is equivalent to 'minimum power consumption' [35], i.e.,

$$\max\{\eta\} = \min\{P_{tot}\}. \quad (26)$$

Therefore, the conceived energy efficiency optimization problem can be formulated as

$$\min_{\{P_k,W_k\}_{k=1}^K}\left\{\frac{1}{\rho}\left(P_{BS}+P_{MR}\right)+P_0^C\right\}, \quad (27)$$

$$\text{s.t.} \quad \gamma_k \geq \exp\left(\frac{E_k^B u\ln 2}{\tau W_k}+\frac{Q_G^{-1}(\varepsilon_k^c)}{\sqrt{\tau W_k}}\right)-1, \forall k, \quad (27a)$$

$$F_{N_t}\left(g_k^{th}\right) \leq \varepsilon_k^h = \frac{\varepsilon_D}{3}, \forall k, \quad (27b)$$

$$P_k \leq P_k^{th}, \forall k, \sum_{k=1}^{K} P_k \leq P_{\max}, \quad (27c)$$

$$W_k \leq W_k^{th}, \forall k, \sum_{k=1}^{K} W_k \leq W_{\max}. \quad (27d)$$

The constraints (27a) ensure the queuing delay violation probability ε_k^q and DL decoding error probability ε_k^c, and (27b) are the constraints on the proactive packet dropping probabilities ε_k^h. Essentially, (27a) and (27b) ensure high reliability and low latency of URLLC. We express the maximum transmit power of the MR as $P_{\max}$. Then, the transmit power assigned by the MR to all the users should meet $\sum_{k=1}^{K} P_k \leq P_{\max}$. Under this total power constraint, the transmit power allocated to each user depends on the channels of other users, and therefore it is difficult to obtain the average transmit power per user in closed form [35]. In order to facilitate optimization, the maximum transmit power constraint, $P_k \leq P_k^{th}$, is introduced for each user in (27c). The constraints (27d) make sure that the sum of the bandwidth allocated to all the users should be less than the total bandwidth of the MR, i.e., $\sum_{k=1}^{K} W_k \leq W_{\max}$, and the bandwidth of each user should be less than the bandwidth threshold W_k^{th}.

IV. Proposed Resource Allocation Algorithm

Analyzing optimization problem (27), it can be observed that within variables, both P_k and W_k for each user are continuous. Therefore, for all K users, the complexity of the problem is $\mathrm{O}\left(K\cdot R^2\right)$. This makes it challenging to find an algorithm in polynomial time to jointly optimizing the continuous variables of transmit power and bandwidth allocation for K users in (27). Additionally, the constraints are related to SINR, which is an non-convex optimization [40], and there is strong coupling between variables P_k and W_k, making problem difficult to solve directly and challenging. Therefore, optimization problem (27) is an NP-hard problem [41, 42].

Since the energy efficiency optimization problem (27) is NP-hard, directly solving it is challenging. We decompose the problem (27) into two subproblems, and use the BCD approach to optimize $\{W_k\}$ and $\{P_k\}$, respectively. For each subproblem, a heuristic algorithm is used to optimize the individual optimization variables one by one while fixing the

other variables in an iterative procedure. The proposed resource allocation algorithm for optimizing the system's energy efficiency consists of the BCD procedure involving iteratively the heuristic algorithm for optimizing user bandwidth and the heuristic algorithm for optimizing user transmit power.

A. Problem Decomposition

In order to effectively solve the system energy efficiency optimization problem proposed in Section III, we decompose the optimization problem (27) into two sub-problems, namely, the heuristic optimization of user bandwidth and the heuristic optimization of user transmitted power.

1) Optimization of User Bandwidth: Given the transmit power allocated to users by the MR, this subproblem deals with how the MR should allocate the bandwidth to K users in order to minimize the total power consumption of the system. When all the users' transmit power $\{P_k\}_{k=1}^K$ are fixed which satisfy the constraints (27c), the optimization problem (27) can be written as

$$\min_{\{W_k\}_{k=1}^K} \left\{ \frac{1}{\rho}\left(P_{BS}+P_{MR}\right)+P_0^C \right\}, \tag{28}$$

$$\text{s.t.} \quad \gamma_k \geq \exp\left(\frac{E_k^B u \ln 2}{\tau W_k}+\frac{Q_G^{-1}(\varepsilon_k^c)}{\sqrt{\tau W_k}}\right)-1, \forall k, \tag{28a}$$

$$F_{N_t}\left(g_k^{th}\right) \leq \varepsilon_k^h = \frac{\varepsilon_D}{3}, \forall k, \tag{28b}$$

$$W_k \leq W_k^{th}, \forall k, \sum_{k=1}^K W_k \leq W_{\max}. \tag{28c}$$

2) Optimization of User transmit Power: Given the bandwidth allocated to users by the MR, this subproblem deals with how the MR should allocate the transmit power to K users in order to minimize the total power consumption of the system. When all the users' bandwidth $\{W_k\}_{k=1}^K$ are fixed which satisfy the constraints (27d), the optimization problem (27) can be written as

$$\min_{\{P_k\}_{k=1}^K} \left\{ \frac{1}{\rho}\left(P_{BS}+P_{MR}\right)+P_0^C \right\}, \tag{29}$$

$$\text{s.t.} \quad \gamma_k \geq \exp\left(\frac{E_k^B u \ln 2}{\tau W_k}+\frac{Q_G^{-1}(\varepsilon_k^c)}{\sqrt{\tau W_k}}\right)-1, \forall k, \tag{29a}$$

$$F_{N_t}\left(g_k^{th}\right) \leq \varepsilon_k^h = \frac{\varepsilon_D}{3}, \forall k, \tag{29b}$$

$$P_k \leq P_k^{th}, \forall k, \sum_{k=1}^K P_k \leq P_{\max}. \tag{29c}$$

B. Heuristic Algorithm

Heuristic algorithms are often used to solve problems that are mathematically hard to solve or have high computational complexity in a reasonable amount of time, such as combinatorial optimization problems [43]. It explores possible solutions through limited computational resources rather than trying to exhaust all possibilities, and is therefore useful and often robust when dealing with high-dimensional or complex search spaces. Also heuristic algorithms are less dependent on initial conditions, this means they perform well on different problem instances without the need for large-scale problem-specific tuning. While heuristic algorithms are not necessarily guaranteed to find a globally optimal solution, they are usually able to find locally optimal solutions that are close to the optimal solution. In each iteration step, the BCD selects a block of variables, and optimizes the selected block while keeping the other blocks fixed. This single-direction update significantly reduces the possibility of getting trapped in saddle points when moving in multiple directions simultaneously [44]. This is often sufficient in practical applications, especially when the global optimal solution is difficult to obtain [45].

The variables of the optimization problem in this paper are the transmit power P_k and the transmission bandwidth W_k allocated by MR to the users in the carriage, because the number of user in the carriage is usually large. And the network requirements between different user services are significant differences, which makes it necessary for the algorithm to be optimized for each user's communication resource, with the high reliability and low latency required of URLLC. In addition, the total transmit power and the total bandwidth allocated by MR to users are limited.

The above factors lead to a complex search space for the optimization model in this paper. At the same time, the traveling route of HSR is characterized by large distance span and terrain changes. The total transmit power and total bandwidth of MR change significantly with the different types of accessed BSs. The heuristic algorithm can effectively solve the search problem with high complexity such as communication resource allocation. In addition, the better robustness of the heuristic algorithm can adapt to the changing driving environment of the HSR. It has been applied in the related literature of communication resource [11, 46, 47].

C. Heuristic Algorithm for User Bandwidth

We now present a heuristic algorithm for solving the optimization problem (28). The strategy adopted is to optimize the individual bandwidth variables W_k one by one while fixing the other bandwidth variables in an iterative procedure. Observe that (28b) does not directly restrict the optimization variables W_k, while (28a) imposes constraint on W_k. The constraint (28a) can be mathematically transformed to:

$$\frac{\alpha_k g_k P_k}{N_0} \geq \left(\exp\left(\frac{E_k^B u \ln 2}{\tau W_k}+\frac{Q_G^{-1}\left(\varepsilon_k^c\right)}{\sqrt{\tau W_k}}\right)-1\right) W_k, \tag{30}$$

where E_k^B is given in (18). Note that since the Doppler effect P_{ICI} can be considered to be unchanged when the train speed and the carrier frequency are fixed, here we omit the expression P_{ICI} in (30) to simplify the analysis of the inequality relation. Let the right hand part of inequality (30) be equal to $y_k(W_k)$, namely,

$$y_k(W_k) = \left(\exp\left(\frac{E_k^B u \ln 2}{\tau W_k}+\frac{Q_G^{-1}\left(\varepsilon_k^c\right)}{\sqrt{\tau W_k}}\right)-1\right) W_k. \tag{31}$$

We have the following two properties of $y_k(W_k)$ [32, 35].

Property 1: $y_k(W_k)$ first strictly decreases and then strictly increases as W_k increases.

Property 2: $y_k(W_k)$ is strictly convex in W_k when $W_k \in \left(0,\ W_k^{th}\right)$.

Algorithm 1 Heuristic Algorithm for User Bandwidth

Initialization: Give transmit power $\{P_k\}_{k=1}^K$ of K users; Randomly generate initial bandwidth $\left\{W_k^{origin}\right\}_{k=1}^K$ of K users, calculate inflection points $\left\{W_k^{knee}\right\}_{k=1}^K$ of $y_k(W_k)$ for $1 \le k \le K$ and critical bandwidth $\left\{W_k^{thr}\right\}_{k=1}^K$ that ensure constraints (27a), set accuracy $\varepsilon = 0.05$, and iterative index $\gamma = 0$;

1: Calculate η^γ according to (25);
2: **repeat**
3: $\gamma = \gamma + 1$;
4: **for** $(k = 1; k \le K; k = k + 1)$ **do**
5: Bandwidth of users $\{1, 2, ..., k-1, k+1, ..., K\}$ are fixed;
6: **if** $W_k^{origin} < W_k^{knee} < \min\{W_k^{thr}, W_k^{th}\}$ **then**
7: **for** $W_k^\gamma = W_k^{origin} : W_{gap} : \min\{W_k^{thr}, W_k^{th}\}$ **do**
8: Calculate P_{tot} with W_k^γ according to (24) and store them;
9: **end for**
10: **else if** $W_k^{knee} < W_k^{origin} < \min\{W_k^{thr}, W_k^{th}\}$ **then**
11: **for** $W_k^\gamma = W_k^{knee} : W_{gap} : \min\{W_k^{thr}, W_k^{th}\}$ **do**
12: Calculate P_{tot} with W_k^γ according to (24) and store them;
13: **end for**
14: **end if**
15: Find bandwidth W_k^γ of user k that minimizes P_{tot}, and set $W_k^{origin} = W_k^\gamma$;
16: **end for**
17: Calculate η^γ according to (25);
18: **until** $\frac{|\eta^\gamma - \eta^{\gamma-1}|}{\eta^\gamma} < \varepsilon$;
19: **Output** η^γ and $\left\{W_k^\gamma\right\}_{k=1}^K$.

In the optimization of the user bandwidth (28), the user transmit power $\{P_k\}_{k=1}^K$ allocated by the MR to the K users are fixed. The heuristic algorithm for user bandwidth allocation is summarized in **Algorithm 1**.

In the initialization, the bandwidth of the K users are randomly initialized as $\left\{W_k^{origin}\right\}_{k=1}^K$. According to *Property 1*, there exists a unique inflection point W_k^{knee} that minimizes $y_k(W_k)$ of (31), where $1 \le k \le K$. In addition, the critical bandwidth W_k^{thr}, $1 \le k \le K$, for which the constraints (28a) hold, are also calculated.

The main iterative procedure (lines 2 to 18) involves optimizing individual bandwidth variables one by one while fixing the other bandwidth variables (lines 4 to 16).

(a) Specifically, for user k, if its initial bandwidth W_k^{origin} satisfies $W_k^{origin} < W_k^{knee} < \min\{W_k^{thr}, W_k^{th}\}$, according to *Property 1*, the function $\frac{\alpha_k g_k P_k}{N_0} - y_k(W_k)$ in the interval $w_k \in \left[W_k^{origin}, \ \min\{W_k^{thr}, W_k^{th}\}\right]$ will first increase and then decrease as the user bandwidth W_k increases. Not that using $\min\{W_k^{thr}, W_k^{th}\}$ is to ensure the constraints (28a) as well as meet the user bandwidth threshold. At the inflection point W_k^{knee}, the function value is maximized. According to (13) and the characteristics of MR-RRH communication, by adjusting the k-th user bandwidth W_k according to the increment $W_{gap} = \frac{|\min\{W_k^{thr}, W_k^{th}\} - W_k^{origin}|}{M}$, where M is the value chosen to control the size of the increment, the mmwave transmit power P_{BS} of the RRH-MR communication and user transmit power P_k will vary accordingly. Therefore, the bandwidth W_k of user k that minimizes the system power P_{tot} is found under the constraints (28a) and (28c), in the range of $\left[W_k^{origin}, \ \min\{W_k^{thr}, W_k^{th}\}\right]$ (lines 6-9).

(b) If $W_k^{knee} < W_k^{origin} < \min\{W_k^{thr}, W_k^{th}\}$, according to *Properties 1* and *2*, the function $\frac{\alpha_k g_k P_k}{N_0} - y_k(W_k)$ in the interval $w_k \in \left[W_k^{knee}, \ \min\{W_k^{thr}, W_k^{th}\}\right]$ decreases as the user bandwidth W_k increases. With the increment $W_{gap} = \frac{|\min\{W_k^{thr}, W_k^{th}\} - W_k^{knee}|}{M}$, the bandwidth W_k of user k that minimizes the system power P_{tot} is found in the interval $\left[W_k^{knee}, \ \min\{W_k^{thr}, W_k^{th}\}\right]$, subject to the constraints (28a) and (28c) (lines 10 to 13).

D. Heuristic Algorithm for Transmit Power

In the optimization of the user transmit power (29), the user bandwidth $\{W_k\}_{k=1}^K$ allocated by the MR to the K users are fixed. Similar to Subsection IV-B, the strategy adopted to solve the problem (29) is to optimize the individual transmit power variables P_k one by one while fixing the other transmit power variables in an iterative procedure. Examining the optimization problem (29) reveals that the constraints (29b) restrict P_k, given the fixed user bandwidth W_k. Specifically, the function $z_k(P_k) = F_{N_t}\left(g_k^{th}\right) - \frac{\varepsilon_D}{3}$ monotonically decreases as P_k increases. Define the lower bound of P_k that satisfies the constraint (29b) as $P_k^{th(\min)}$. Then P_k cannot be smaller than $P_k^{th(\min)}$. The heuristic algorithm for the user transmit power allocation is summarized in **Algorithm 2**.

The initialization involves randomly generating the initial transmit power $\left\{P_k^{origin}\right\}_{k=1}^K$ for the K users. The main iterative procedure (lines 2 to 12) involves optimizing individual

Algorithm 2 Heuristic Algorithm for Transmit Power

Initialization: Give bandwidth $\{W_k\}_{k=1}^K$ of K users; Randomly generate initial transmit power $\left\{P_k^{origin}\right\}_{k=1}^K$ for K users, set accuracy $\xi = 0.05$ and iterative index $\gamma = 0$;

1: Calculate η^γ according to (25);
2: **repeat**
3: $\gamma = \gamma + 1$;
4: **for** $(k = 1; k \le K; k = k + 1)$ **do**
5: Transmit power of users $\{1, 2, \cdots, k-1, k+1, \cdots, K\}$ are fixed;
6: **for** $P_k^\gamma = P_k^{origin} : -P_{gap} : P_k^{th(\min)}$ **do**
7: Calculate P_{tot} with P_k^γ according to (24) and store them;
8: **end for**
9: Find transmit power P_k^γ of user k that minimizes P_{tot}, and set $P_k^{origin} = P_k^\gamma$;
10: **end for**
11: Calculate η^γ according to (25);
12: **until** $\frac{|\eta^\gamma - \eta^{\gamma-1}|}{\eta^\gamma} < \xi$;
13: **Output** η^γ and $\left\{P_k^\gamma\right\}_{k=1}^K$.

Algorithm 3 Resource Allocation Optimization of User Bandwidth and Transmit Power Based on BCD

Initialization: Randomly generate initial transmit power $\left\{P_k^{origin}\right\}_{k=1}^{K}$ and initial bandwidth $\left\{W_k^{origin}\right\}_{k=1}^{K}$ for K users, set iterative index $\gamma = 0$ and accuracy $\zeta = 0.05$;

1: Calculate η^{γ} according to (25);
2: **repeat**
3: With fixed transmit power $\left\{P_k^{\gamma}\right\}_{k=1}^{K}$, update bandwidth $\left\{W_k^{\gamma+1}\right\}_{k=1}^{K}$ for K users using **Algorithm 1**;
4: With fixed user bandwidth $\left\{W_k^{\gamma}\right\}_{k=1}^{K}$, update transmit power $\left\{P_k^{\gamma+1}\right\}_{k=1}^{K}$ for K users using **Algorithm 2**;
5: Calculate η^{γ} according to (25);
6: $\gamma = \gamma + 1$;
7: **until** $\frac{\left|\eta^{\gamma}-\eta^{\gamma-1}\right|}{\eta^{\gamma}} < \zeta$;
8: **Output** η^{γ}, $\{P_k^{\gamma}\}_{k=1}^{K}$ and $\{W_1^{\gamma}\}_{k=1}^{K}$.

transmit power variables one by one while fixing the other transmit power variables (lines 4 to 10). Specifically, for user k, by adjusting P_k in the interval $\left[P_k^{th(\min)},\ P_k^{origin}\right]$ with the increment $P_{gap} = \frac{|P_k^{origin}-P_k^{th(\min)}|}{M}$, the transmit power P_k of user k that minimizes P_{tot} is found under the constraints of (29b) and (29c).

E. Proposed Algorithm

1) Algorithm Summary: The proposed resource allocation algorithm decomposes the joint user bandwidth and transmit power optimization (27) into the two subproblems of optimizing user bandwidth given user transmit power and optimizing user transmit power given user bandwidth, separately, and solves these two subproblems iteratively based on BCD. The idea of the BCD approach is that during each iteration, only one variable is optimized, while the remaining variables are fixed [48]. **Algorithm 3** summarizes the proposed resource allocation optimization of user bandwidth and transmit power using the BCD based on **Algorithm 1** and **Algorithm 2**.

2) Block Coordinate Descent: BCD is a more generalization of coordinate descent, which decomposes the original problem into multiple sub-problems by simultaneously optimizing a subset of variables. The order of updates during the descent can be deterministic or random [48]. The solution idea of BCD is to optimize the solution for only one variable in each iteration, keeping the remaining variables constant, and then solving alternately.

Consider an optimization task as follows

$$\min\ F(x_1, ..., x_s) \equiv f(x_1, ..., x_s) + \sum\nolimits_{i=1}^{s} r_i(x_i). \quad (32)$$

A generic framework for BCD is shown in **Algorithm 4**. In the general framework of BCD, the most commonly used update scheme is block minimization, i.e., $x_i^k = \arg\min_{x_i} F\left(x_{<i}^k, x_i, x_{>i}^{k-1}\right)$. For (32), we can use coordinate descent to seek a minimum value, and we start with an initial $x^{(0)}$ that loops over k, as described in the following unfolding.

Algorithm 4 Block coordinate descent

Initialization: choose $(x_1^0, ..., x_s^0)$

1: **for** k=1,2,... **do**
2: **for** i=1,2,...,s **do**
3: update x_i^k with all other blocks fixed
4: **end for**
5: **if** stopping criterion is satisified **then**
6: return $(x_1^k, ..., x_s^k)$.
7: **end if**
8: **end for**

$$\begin{aligned} x_1^{(k)} &= \arg\min_{x_1} F(x_1, x_2^{(k-1)}, x_3^{(k-1)}, ..., x_n^{(k-1)}), \\ x_2^{(k)} &= \arg\min_{x_2} F(x_1^{(k)}, x_2, x_3^{(k-1)}, ..., x_n^{(k-1)}), \\ &... \\ x_n^{(k)} &= \arg\min_{x_n} F(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, ..., x_n). \end{aligned} \quad (33)$$

Next, we will discuss the BCD applied in Algorithm 3. We first perform the initialization at the moment $\gamma = 0$, including the allocation scheme (W_k, P_k) for user communication resources. In the next step, the two sub-problems of the section III-G decomposition are computed by BCD, by Algorithm 1 and Algorithm 2. The results are compared with the values calculated in the previous calculation, as shown in step 7 of Algorithm 3. The first sub-problem is to fix the transmit power P_k assigned by MR and optimize the communication bandwidth W_k allocated to users, to obtain the local optimal solution; The second is to fix the bandwidth W_k and optimize transmit power P_k, to obtain the local optimal solution. Finally, we determine whether the end-of-iteration condition is satisfied. If the iteration does not end, the next alternate iteration of optimization is performed by BCD. Finally, when the algorithm converges, we will get the optimal solution.

3) Complexity Analysis: The complexity of **Algorithm 3** depends on the number of users N_{UE} in the carriage as well as the complexity of the heuristic optimization algorithms for the user bandwidth allocation subproblem and the user transmit power allocation subproblem, namely, **Algorithm 1** and **Algorithm 2**.

For the heuristic optimization algorithm of the user bandwidth allocation, let user k perform N_k^{band} bandwidth optimization calculations to find the locally optimal bandwidth. Assume that **Algorithm 1** needs to perform N_1 iterations to achieve the convergence condition $\left|\eta^{\gamma} - \eta^{\gamma-1}\right| / \eta^{\gamma} < \varepsilon$. Then the complexity of **Algorithm 1** is on the order of $\mathrm{O}\left(N_1 \cdot \sum_{k=1}^{N_{UE}} N_k^{band}\right)$.

For the heuristic optimization algorithm of the user transmit power allocation, assume that user k performs N_k^{power} power optimization calculations to find the locally optimal transmitted power and **Algorithm 2** needs N_2 iterations to achieve the convergence condition $\left|\eta^{\gamma} - \eta^{\gamma-1}\right| / \eta^{\gamma} < \varepsilon$. Then the complexity of **Algorithm 2** is $\mathrm{O}\left(N_2 \cdot \sum_{k=1}^{N_{UE}} N_k^{power}\right)$.

Let the number of iterations for **Algorithm 3** to achieve the convergence condition $\left|\eta^{\gamma}-\eta^{\gamma-1}\right|/\eta^{\gamma}<\zeta$ be N_3. The complexity of the proposed algorithm is on the order of $\mathrm{O}\left(N_3\cdot\left(N_1\cdot\sum_{k=1}^{N_{UE}}N_k^{band}+N_2\cdot\sum_{k=1}^{N_{UE}}N_k^{power}\right)\right)$.

4) Convergence Analysis: The Algorithm 3 presented in this paper, that is Resource Allocation Optimization of User Bandwidth and Transmit Power Based on BCD, involves the iterative solution of two subproblems by BCD.

A note on the convergence of Algorithm 1. When the transmit power P_k allocated by MR to the users is fixed, Algorithm 1 is used to solve the optimal solution of the user bandwidth W_k. In Algorithm 1, for the objective function $P_{tot}^{\gamma}(W^{\gamma},P^{*})$, the communication bandwidths of the remaining $K-1$ users are fixed in the γ_{th} iteration using the heuristic algorithm (i.e., line 2 - line 18) to find the optimal bandwidth for user k, to minimize the total power consumption. Note that in this paper, because of the non-monotonic constraints of (28a) (i.e., property 1 and 2) and the continuity of the variable, the complexity of the optimization algorithm is reduced by discretizing the variable W_k by exploiting W_{gap} that depends on the initial variable W_k^{origin} and the inflection point W_k^{knee} in (28a). Finally, user k can obtain the bandwidth W_k that minimizes the total system power consumption P_{tot} and satisfies both constraints (28a) and (28c). Next, the optimal bandwidth W_{k+1} for the user $k+1$ is solved. Thus in the γ_{th} iteration of Algorithm 1,the heuristic algorithm is utilized to obtain the optimal solution $\{W_k^{\gamma}\}_{k=1}^{K}$ for the bandwidth of all the users. Because it is searching for the bandwidth W_k^{γ} for which the user k minimizes the total power consumption P_{tot}, $P_{tot}^{\gamma+1}(W^{\gamma+1},P^{*})\le P_{tot}^{\gamma}(W^{\gamma},P^{*})$, i.e., $P_{tot}^{\gamma}(W^{\gamma},P^{*})$ is nonincreasing. Moreover, since W_{gap} discretizes the continuous variable, this makes the system power consumption $P_{tot}^{\gamma}(W^{\gamma},P^{*})$ bounded. Therefore, Algorithm 1 is convergent.

A note on the convergence of Algorithm 2. When the user bandwidth W_k allocated by the MR is fixed, Algorithm 2 is used to solve for the transmit power P_k allocated by the MR to the user. Due to the monotonicity of the constraint (29b) and the continuity of the variable, the discretization of the variable by P_{gap} depending on the initial variable P^{origin} and the threshold $P^{th(\min)}$ of the constraint (29b) is utilized to reduce the computational complexity. The proof of convergence is similar to that of Algorithm 1. In the γ_{th} iteration, the optimal solution P^{γ} of the transmit power of all users is obtained by the heuristic algorithm (i.e., line 2 - line 12). Each iteration of Algorithm 2 is computed with the aim of minimizing the system power consumption P_{tot} and eventually obtaining P_k^{γ} corresponding to the system power consumption P_{tot}^{*}, and hence $P_{tot}^{\gamma+1}(W^{*},P^{\gamma+1})\le P_{tot}^{\gamma}(W^{*},P^{\gamma})$. Also the system power consumption P_k^{γ} is bounded due to the discretization of the variables. Therefore, Algorithm 2 is convergent.

A note on the convergence of Algorithm 3. Algorithm 3 utilizes the BCD algorithm to iteratively optimize Algorithm 1 and Algorithm 2. The solution idea of BCD is to optimize only one variable per iteration, keeping the rest of the variables constant, and then solving alternately, which has been described in detail in the previous description of the BCD algorithm. In line 3 of Algorithm 3, the transmit power P_k is fixed and the user bandwidth allocation scheme is optimized to obtain a locally optimal solution through Algorithm 1. In line 4, the user bandwidth W_k is fixed and the transmit power allocation scheme is optimized to obtain a locally optimal solution through Algorithm 2. The convergence of Algorithm 1 and Algorithm 2 has been proved earlier. Meanwhile, in the γ_{th} iteration in Algorithm 3, the corresponding converged local optimum values can be obtained for both line 3 and line 4, and line 4 is relatively non-increasing with respect to the value obtained for line 3. So in Algorithm 3, there exists $P_{tot}^{\gamma+1}(W^{\gamma+1},P^{\gamma+1})\le P_{tot}^{\gamma+1}(W^{\gamma},P^{\gamma})$, which indicates that the objective function of the original problem is non-increasing. Also due to the discretization of variables W_k and P_k, this makes the problem $P_{tot}^{\gamma}(W^{\gamma},P^{\gamma})$ of minimizing the power consumption of the system bounded. Thus Algorithm 3, Resource Allocation Optimization of User Bandwidth and Transmit Power Based on BCD, proposed in this paper, is convergent. And the convergence performance of the Algorithm 3 is showed in Fig. 11.

V. Performance Evaluation

In this section, we evaluate the performance of the proposed algorithm in simulation using the HSR model with MR for vehicle-ground URLLC communication.

A. Simulation Setup

The deployment model of RRH and MR in the HSR scenario is 3GPP 38.913 [7], which is detailed in Subsection III-A. The mmwave communication model between RRH and MR is 3GPP 38.901 [12], which is discussed in Subsection III-B. The expression of URLLC [26] is shown in Subsection III-C. The traffic model based on Poisson arrival process of packet [34] is given in Subsection III-D. The QoS of URLLC system [35] is presented in Subsection III-E. The energy efficiency of the system is derived in Subsection III-F. In the simulation, the train runs at a speed of $v=250$ km/h from the position of 100 m horizontal distance to the RRH. The train runs through eight location bins, and the length of a location bin is 50 m. The number of bytes u in a packet is 20. The carrier frequency of mmwave communication is $f_c=30$ GHz. The RRH communicates with the user in the carriage via MR. There are N_{UE} users in a carriage, and the communication devices carried by users have N_{App} apps that consume traffic on average, which conform to the Gaussian distribution. These apps are activated with probability κ. The total bandwidth of MR is $W_{\max}$, and the total transmit power of MR is $P_{\max}$. ε_D is the maximum tolerable error rate required to ensure the overall reliability of URLLC. Table II summarizes the default parameters of the simulation system. Unless otherwise specified, these default parameters are used.

Two existing resource allocation algorithms are chosen as the benchmarks to compare with the proposed algorithm in the system performance evaluation simulation under different train locations and parameter configurations. For the convenience, the proposed algorithm that optimizes the resource allocation of both users' bandwidth and transmit power based on BCD is

TABLE II
PARAMETERS OF HSR SIMULATION CONFIGURATION

Parameter	Value
Distance between RRH and rail $D_{\text{RRH\_rail}}$	30 m
Height of RRH h_R	15 m
Height of MR h_T	1.5 m
Height of train $h_{\text{train\_height}}$	4.5 m
Length of train $h_{\text{train\_length}}$	30 m
Height of receiver antenna $h_{\text{RX\_height}}$	6 m
Radius of location bin σ_D	50 m
Distance between adjacent RRH D_{RRH}	600 m
Average height of building h	5 m
Vertical 3 dB beamwidth at RRH θ_{RRH_3dB}	65°
Vertical 3 dB beamwidth at MR θ_{MR_3dB}	90°
Horizontal 3 dB beamwidth at RRH ϕ_{RRH_3dB}	65°
Horizontal 3 dB beamwidth at MR ϕ_{MR_3dB}	90°
RRH vertical sidelobe attenuation $SLA_{V_{RRH}}$	30 dB
MR vertical sidelobe attenuation $SLA_{V_{MR}}$	25 dB
RRH maximum attenuation $A_{\max\_RRH}$	30 dB
MR maximum attenuation $A_{\max\_MR}$	25 dB
RRH maximum directional gain $G_{\max\_RRH}$	8 dBi
MR maximum directional gain $G_{\max\_MR}$	5 dBi
Mmwave carrier frequency f_c	30 GHz
Density and size of obstacles β	0.001
Noise power spectrum density N_0	-174 dBm/Hz
Mmwave fixed rate at RRH R_{mmwave}	4×10^8 bit/s
URLLC communication carrier frequency f_{urllc}	2.5 GHz
Standard deviation of shadow fading σ	5
Path loss exponent n	3.8
Total MR bandwidth $W_{\max}$	20 MHz
URLLC user gain G_{ue}	3 dBi
URLLC antenna gain $G_{antenna}$	25 dBi
High-speed train speed v	250 km/h
URLLC maximum latency tolerance $D_{\max}$	1 ms
Duration of frame T_f	0.1 ms
Duration of DL/UL transmission τ	0.05 ms
Duration of RRH-MR transmission D_{mmwave}	0.5 ms
Packet size u	160 bits
Queuing delay requirement $D^q_{\max}$	0.85 ms
Power amplifier efficiency ρ	0.5
Fixed circuit power consumption P_0^C	50 mw
Average number of Apps for users N_{App}	15
Activation probability of Apps κ	0.5
Maximum total system error rate ε_D	3×10^{-7}
Number of users in a carriage N_{UE}	20

denoted PA. The other two resource allocation algorithms include the random algorithm (RA), and the resource allocation algorithm based on two-stage dynamic K-means clustering (KA) [49], which are further elaborated as follows.

- **Random Algorithm (RA)**. This wireless resource scheduling method configures the bandwidth and transmit power allocated by MR to users in a random method. Therefore, the difference between RA and PA is that our PA optimizes the parameter configuration with BCD.
- **Resource allocation algorithm based on two-stage dynamic K-means clustering (KA) [49]**. First, the unsupervised K-means algorithm is used to divide the users into K groups based on the number of activated apps and their relative positions to MR, and the bandwidth and power available from MR are divided according to the relative relationships of the cluster centers of the K groups. Then the bandwidth and power in each group are optimized by the K-means clustering to minimize the energy consumption of the system under the system and URLLC constraints.

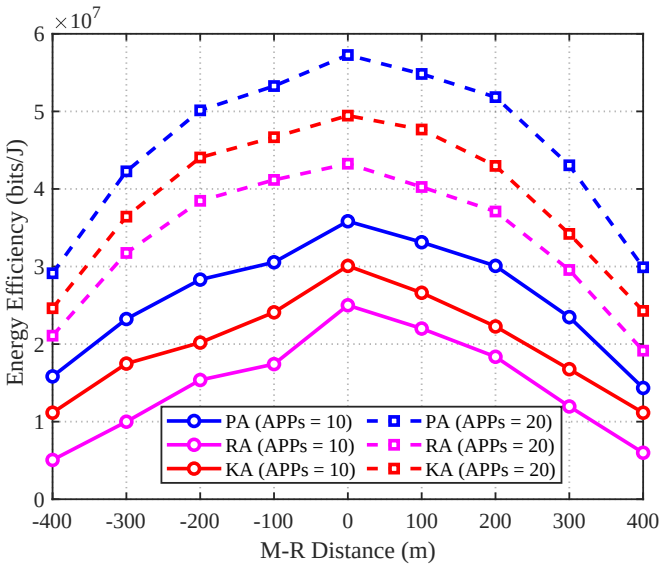

Fig. 5: System energy efficiency comparison of three resource allocation algorithms at different M-R distance for two different numbers of Apps.

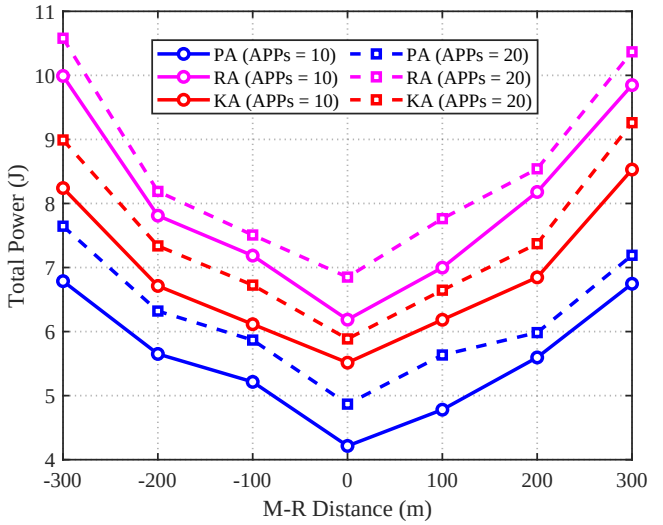

Fig. 6: Total system power consumption comparison of three resource allocation algorithms at different M-R distance for two different numbers of Apps.

B. Performance Comparison

In the simulation, we change the configuration of the system parameters and evaluate the achievable performance of the proposed algorithm and the two benchmark algorithms. Note that the relative horizontal distance between the MR and RRH (M-R distance) is considered here, and the positive and negative distances indicate that the train is on the left and right sides of the RRH, respectively.

1) Impact of Average Number of Apps N_{Apps}: Fig. 5 and Fig. 6 investigate the impact of the average number of Apps N_{Apps} on the achievable energy efficiency and total system power consumption for the three resource allocation algorithms, PA, RA and KA, respectively. The simulation results of Fig. 5 clearly show that the system energy efficiency increases with N_{Apps}. On one hand, as the average number of Apps on

each user device increases, the bandwidth threshold allocated to each user remains the same since the same total MR bandwidth is given. However, users with more Apps would need to be allocated with more bandwidth otherwise fewer available low-power solutions would meet the reliability and latency constraints. Therefore, the total power consumption has to increase, as shown in Fig. 6. On the other hand, as N_{Apps} increases, packet arrival rates grow faster, which more than compensates for the negative effect of higher total power consumption, leading to an improvement in the energy efficiency. The results again demonstrate that our PA is the best, and KA is the second best, while RA is the worst, in terms of system energy efficiency. When the distance $M-R=200\,\text{m}$ and the average number of apps $N_{Apps}=20$, for example, the energy efficiency achieved by our PA is 21% higher than that of KA, 39% higher than that of RA.

The simulation results clearly show that our PA attains the best performance, and KA achieves the second best performance, while RA performs the worst. For example, when the horizontal distance of the train to the RRH is $M-R=100\,\text{m}$ and the number of users is $N_{Apps}=20$, the energy efficiency of our PA is around 15% higher than that of KA, and more than 35% higher than that of RA. Note that the performance of RA is achieved with the initial bandwidth and power allocation for PA. The significant performance gain of PA over RA therefore demonstrates the effectiveness of the proposed optimization approach for alternatively optimizing the bandwidth and power allocated to the users, separately. The KA scheme classifies the users with an approximate number of apps and relative location relationships into a group by clustering. Then it performs a second clustering to optimize the bandwidth and power of each group. The advantage of clustering is that it is low complexity. But it suffers from the disadvantage of not optimizing for each user, leading to an inferior performance than PA. When the distance reaches $M-R=400\,\text{m}$, the differences between PA, KA and RA are not as large as when the absolute value of the M-R distance is less than 400 m, because the large path loss reduces the room that can be exploited by resource allocation optimization.

2) Influence of Total MR Bandwidth $W_{\max}$: Fig. 7 and Fig. 8 illustrate the influence of the total MR bandwidth available on the achievable energy efficiency and total system power consumption with different train speed, for three comparison algorithms, PA, RA and KA. Note that in the simulation about different train speeds, distance is $M-R=250$ m. It can be observed that the system energy efficiency increases with the total MR bandwidth. By increasing $W_{\max}$, the bandwidth available per user increases. Under the condition that the reliability and delay constraints of URLLC are satisfied, the available low-power solutions will increase, and the total power consumed decreases accordingly, as shown in Fig. 8. This leads to an increase in the energy efficiency. Furthermore, we observe that the system energy efficiency of all three resource allocation algorithms decreases as the train speed increases, corresponding to an increase in the total system power consumption. This is because in this paper, we consider the Doppler effect P_{ICI} due to high speed, and as shown in (12), P_{ICI} increases as the train speed increases, leading to

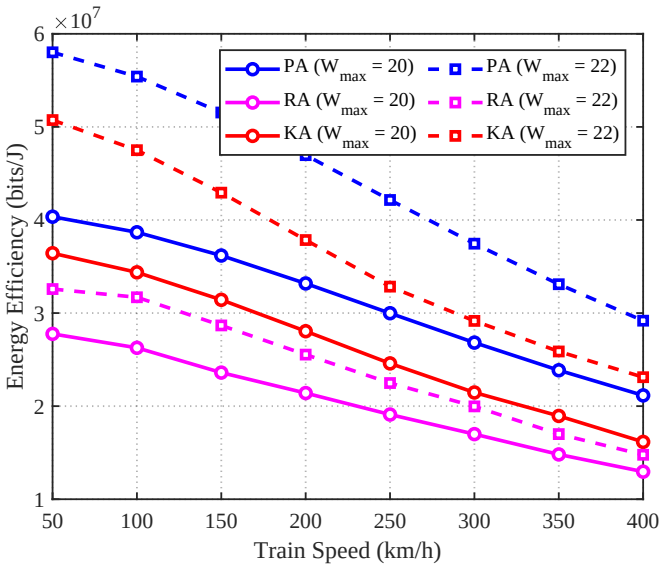

Fig. 7: System energy efficiency comparison of three resource allocation algorithms at different train speed for two different MR bandwidth.

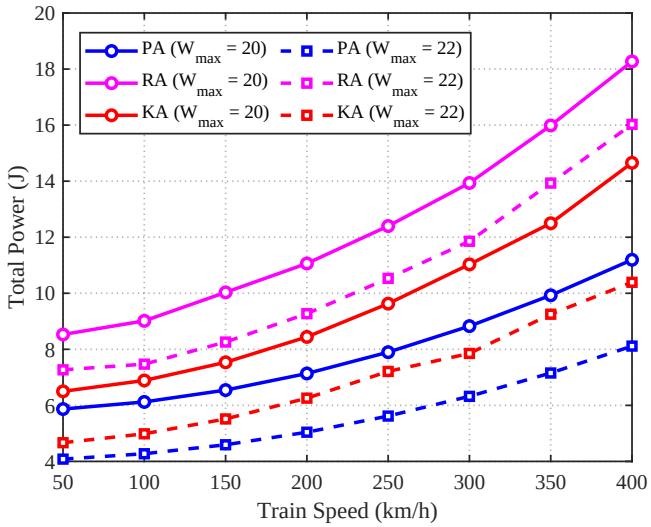

Fig. 8: Total system power consumption of three resource allocation algorithms at different train speed for two different MR bandwidth.

a decrease in the achievable rate in (13), which can drive the system to require more power to satisfy the QoS of URLLC. According to the simulation results of Fig. 7, PA is superior to the other two resource allocation algorithms. At the train speed $v=250\,\text{km/h}$ and with the MR bandwidth $W_{\max}=22\,\text{MHz}$, for instance, the energy efficiency of PA is 28% higher than that of KA, 87% higher than that of RA.

3) Impact of Total MR transmit Power $P_{\max}$: Fig. 9 depicts the effect of the maximum total MR transmit power $P_{\max}$ on the achievable energy efficiency for three algorithms, PA, KA, and RA. According to Fig. 9, increasing $P_{\max}$ decreases the system energy efficiency. This is because as the maximum MR transmit power increases, the power allocated to each user can increase accordingly to reduce the total bandwidth and the RRH transmit power consumed. This causes an increase

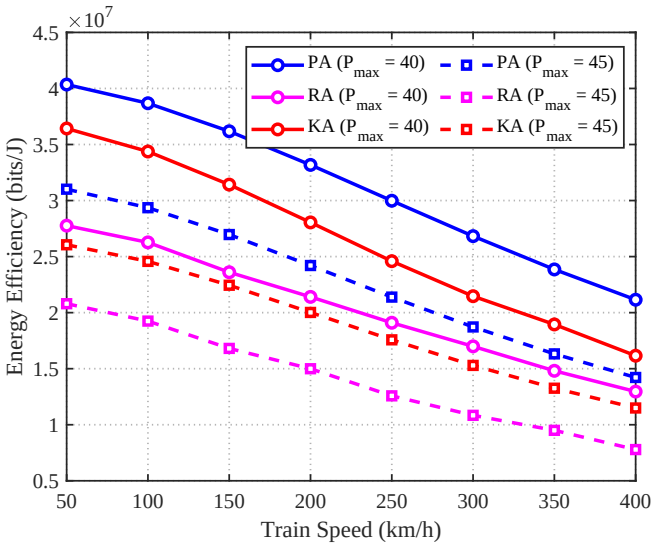

Fig. 9: System energy efficiency comparison of three resource allocation algorithms at different train speed for two different maximum total MR power.

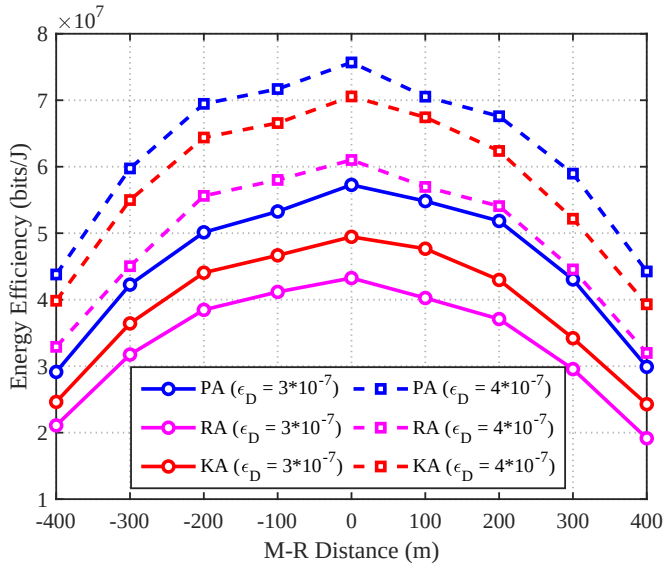

Fig. 10: System energy efficiency comparison of three resource allocation algorithms at different M-R distance for two different maximum tolerable total error probability.

in the total power consumption, leading to a reduction in system energy efficiency. Due to the Doppler effect P_{ICI} caused by high speed, the energy efficiency of the system correspondingly decreases as the speed increases. Again, PA is superior to the other two resource allocation algorithms. In particular, at the train speed $v = 300$ km/h and with the maximum total MR transmit power $P_{\max} = 40$ dBm, the energy efficiency achieved by PA is 25% higher than that of KA, and 58% higher than that of RA.

4) Influence of Maximum Tolerable Error Rate ε_D: Fig. 10 investigates the influence of the maximum tolerable total error probability ε_D on the system energy efficiency for PA, KA and RA, which shows that increases ε_D leads to better achievable system energy efficiency. This is because by increasing the maximum tolerable error probability, the power consumption of each user can be decreased while transmitting the same

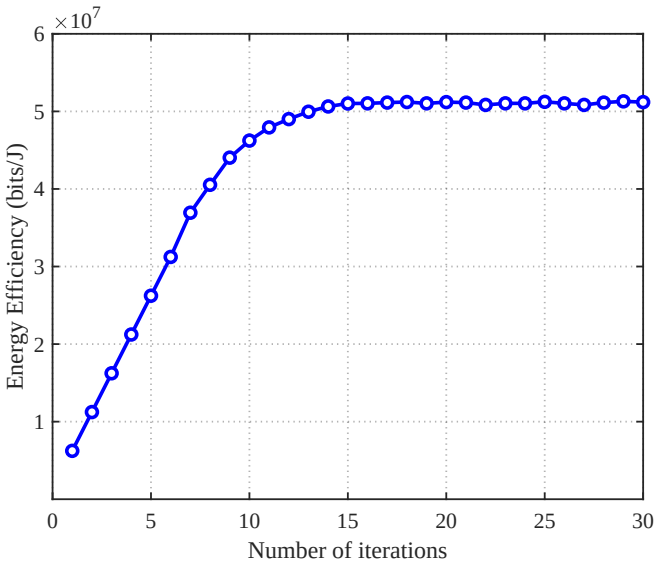

Fig. 11: Convergence performance analysis of Algorithm 3.

amount of data. At the same time, due to $(1 - \varepsilon_D) \approx 1$ in the energy efficiency formula (25), the package arrival rate changes little. Consequently, the total energy efficiency is increased. Again the simulation results confirm the superior performance of PA over the other two resource allocation algorithms. For example, at the distance $M - R = 200$ m and with $\varepsilon_D = 3 \times 10^{-7}$, the energy efficiency achieved by PA is 21% higher than that of KA, 40% higher than that of RA.

VI. Conclusions

In this paper, we have addressed optimal allocation of bandwidth and transmit power to users in the high-speed rail URLLC-based communication system assisted by mobile relays. Our focus has been on maximizing the system energy efficiency while meeting the QoS requirements of URLLC. More specifically, we have developed a mmwave train-to-ground URLLC model incorporating MR, which aims to reduce the penetration loss of communication between RRH and users in the carriage. Based on this model, we have formulated the system energy efficiency maximization problem subject to the URLLC QoS requirements by jointly optimizing the users' bandwidth and transmit power allocation. In order to efficiently solve this NP-hard optimization, we have alternatively optimized the user bandwidth allocation and the user transmit power allocation using a heuristic BCD-based approach. An extensive simulation study under various system parameter configurations has demonstrated the effectiveness of the proposed algorithm in improving system energy efficiency. In particular, the simulation results have shown that our proposed algorithm consistently outperforms a very latest resource allocation optimization algorithm.

References

[1] B. Ai, A. F. Molisch, M. Rupp, and Z.-D. Zhong, "5G key technologies for smart railways," *Proceedings of the IEEE*, vol. 108, no. 6, pp. 856–893, June 2020.

[2] D. He, B. Ai, K. Guan, Z. Zhong, B. Hui, J. Kim, H. Chung, and I. Kim, "Channel measurement, simulation, and analysis

for high-speed railway communications in $5G$ millimeter-wave band," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3144–3158, Oct. 2018.

[3] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[4] M. E. Porter and J. E. Heppelmann, "How smart, connected products are transforming competition," *Harvard business review*, vol. 92, no. 11, pp. 64–88, Nov. 2014.

[5] O. N. C. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency $5G$ communication for a factory automation use case," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, London, UK, June 2015, pp. 1190–1195.

[6] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "$5G$-enabled tactile internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[7] 3GPP, "Study on scenarios and requirements for next generation access technologies: $TR38.913$ $V17.0.0$," *Technical Specification Group Radio Access Network, Technical Report 38.913*, 2022.

[8] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency critical iot applications in $5G$: Perspective on the design of radio interface and network architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[9] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, Sept. 2016.

[10] M. Serror, C. Dombrowski, K. Wehrle, and J. Gross, "Channel coding versus cooperative arq: Reducing outage probability in ultra-low latency wireless communications," in *2015 IEEE Globecom Workshops (GC Wkshps)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[11] Y. Chen, B. Ai, Y. Niu, R. He, Z. Zhong, and Z. Han, "Resource allocation for device-to-device communications in multi-cell multi-band heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4760–4773, May 2019.

[12] 3GPP, "Study on channel model for frequencies from 0.5 to $100GHz$: $TR38.901$," 2017.

[13] L. Wang, B. Ai, Y. Niu, Z. Zhong, S. Mao, N. Wang, and Z. Han, "Energy efficient train-ground mmwave mobile relay system for high speed railways," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, July 2022.

[14] Y. Lu, K. Xiong, P. Fan, Z. Zhong, and B. Ai, "The effect of power adjustment on handover in high-speed railway communication networks," *IEEE Access*, vol. 5, pp. 26 237–26 250, Nov. 2017.

[15] T. Jiang, X. Liu, Y. Wang, and W. Wang, "Research on optimal energy efficient power allocation for noma system in high-speed railway scenarios," in *2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE)*, Shenzhen, China, May 2022, pp. 529–532.

[16] J. Hu, X. Wang, and Y. Xu, "Energy-efficient power optimization and transmission mode selection for distributed antenna system in hsr communications," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, Kuala Lumpur, Malaysia, Apr. 2019, pp. 1–5.

[17] H. Deng, F. Luo, and Q. Li, "A hybrid resource allocation method for urllc based on noma," in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, Tianjin, China, Oct. 2021, pp. 902–905.

[18] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink $MISO$ ofdma-urllc systems," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7184–7200, 2020.

[19] K. Chen, Y. Wang, J. Zhao, X. Wang, and Z. Fei, "Urllc-oriented joint power control and resource allocation in uav-assisted networks," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 10 103–10 116, Jan. 2021.

[20] B. Chang, L. Zhang, L. Li, G. Zhao, and Z. Chen, "Optimizing resource allocation in urllc for real-time wireless control systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8916–8927, Sept. 2019.

[21] X. Zhang, Y. Niu, S. Mao, Y. Cai, R. He, B. Ai, Z. Zhong, and Y. Liu, "Resource allocation for millimeter-wave train-ground communications in high-speed railway scenarios," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4823–4838, May 2021.

[22] X. Zhang, Y. Niu, T. Yang, X. Xiao, J. Ding, S. Chen, Z. Zhong, N. Wang, and B. Ai, "Qos-aware user association and transmission scheduling for millimeter-wave train-ground communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 9, pp. 9532–9545, Sept. 2023.

[23] X. Zhou, X. Zhang, C. Chen, Y. Niu, Z. Han, H. Wang, C. Sun, B. Ai, and N. Wang, "Deep reinforcement learning coordinated receiver beamforming for millimeter-wave train-ground communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5156–5171, May 2022.

[24] M. Gao, B. Ai, Y. Niu, Z. Zhong, Y. Liu, G. Ma, Z. Zhang, and D. Li, "Dynamic mmwave beam tracking for high speed railway communications," in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Barcelona, Spain, Apr. 2018, pp. 278–283.

[25] 3GPP, "Study on performance enhancements for high speed scenario in lte: $TR36.878$," 2015.

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

[27] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for $D2D$-enabled vehicular communications," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3186–3197, July 2017.

[28] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[29] X. Zhang, Q. Zhu, and H. V. Poor, "Minimum-energy and error-rate for urllc networks over nakagami-m channels: A finite-blocklength analysis," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1–6.

[30] C. She, C. Liu, T. Q. S. Quek, C. Yang, and Y. Li, "Ultra-reliable and low-latency communications in unmanned aerial vehicle communication systems," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3768–3781, May 2019.

[31] S. Schiessl, J. Gross, and H. Al-Zubaidy, "Delay analysis for wireless fading channels with finite blocklength channel coding," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Nov. 2015, pp. 13–22.

[32] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 402–415, Jan. 2019.

[33] P. Kela, M. Costa, J. Salmi, K. Leppanen, J. Turkka, T. Hiltunen, and M. Hronec, "A novel radio frame structure for $5G$ dense outdoor radio access networks," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, UK, July 2015, pp. 1–6.

[34] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance modeling of safety messages broadcast in vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 380–387, Mar. 2013.

[35] C. Sun, C. She, and C. Yang, "Energy-efficient resource allocation for ultra-reliable and low-latency communications,"

in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, Jan. 2017, pp. 1–6.

[36] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for $5G$ cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[37] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 127–141, Jan. 2018.

[38] C.-S. Chang and J. Thomas, "Effective bandwidth in high-speed digital networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1091–1100, Aug. 1995.

[39] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, Y. Zhu, and D. O. Wu, "Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2674–2687, Mar. 2017.

[40] Y. Chen, B. Ai, H. Zhang, Y. Niu, L. Song, Z. Han, and H. Vincent Poor, "Reconfigurable intelligent surface assisted device-to-device communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 2792–2804, May 2021.

[41] M. Gao, B. Ai, Y. Niu, W. Wu, P. Yang, F. Lyu, and X. Shen, "Efficient hybrid beamforming with anti-blockage design for high-speed railway communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9643–9655, Sept. 2020.

[42] C. Kai, H. Li, L. Xu, Y. Li, and T. Jiang, "Joint subcarrier assignment with power allocation for sum rate maximization of $D2D$ communications in wireless cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4748–4759, May 2019.

[43] N. Kokash, "An introduction to heuristic algorithms," *Department of Informatics and Telecommunications*, pp. 1–8, 2005.

[44] E. Song, Z. Shen, and Q. Shi, "Block coordinate descent only converge to minimizers," *arXiv preprint arXiv:1710.09047*, 2017.

[45] S. Desale, A. Rasool, S. Andhale, and P. Rane, "Heuristic and meta-heuristic algorithms and their relevance to the real world: a survey," *Int. J. Comput. Eng. Res. Trends*, vol. 351, no. 5, pp. 2349–7084, 2015.

[46] X. Hu, Y. Wang, Z. Liu, X. Du, W. Wang, and F. M. Ghannouchi, "Dynamic power allocation in high throughput satellite communications: A two-stage advanced heuristic learning approach," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 3502–3516, Mar. 2023.

[47] A. A. Z. Ibrahim, F. Hashim, N. K. Noordin, A. Sali, K. Navaie, and S. M. E. Fadul, "Heuristic resource allocation algorithm for controller placement in multi-control $5G$ based on sdn/nfv architecture," *IEEE Access*, vol. 9, pp. 2602–2617, Dec. 2021.

[48] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[49] M. Katwe, K. Singh, P. K. Sharma, C.-P. Li, and Z. Ding, "Dynamic user clustering and optimal power allocation in uav-assisted full-duplex hybrid noma system," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2573–2590, Apr. 2022.