A Low-Latency Collaborative HARQ Scheme for Control/User-Plane Decoupled Railway Wireless Networks

Li Yan, *Student Member, IEEE*, Xuming Fang, *Member, IEEE*, Geyong Min, *Member, IEEE*, and Yuguang Fang, *Fellow, IEEE*

***Abstract*—The control/user (C/U) plane decoupled railway wireless network is an innovative architecture recently proposed to meet the communication demands of both train control systems and onboard passengers. The core idea is to completely separate the C-plane and the U-plane into different network nodes operating at different frequency bands. Although the system capacity of this network architecture can be highly increased, the forwarding latency of X3 interfaces to link the C-plane and the U-plane becomes a serious problem, particularly for hybrid automatic repeat request (HARQ) protocols that demand frequent interactions between the C-plane and the U-plane. To address this challenging problem, we propose a low-latency collaborative HARQ scheme in this paper. Specifically, we develop a new collaborative transmission framework where the possible spare resources on lower frequency bands of macrocells excluding those used by C-plane transmissions can be utilized to help small cells relay erroneously received data. Compared with the conventional HARQ scheme, the proposed scheme requires fewer retransmissions to reach the same transmission reliability, thereby mitigating the latency problem caused by HARQ retransmissions. Furthermore, channel mapping is also redesigned to conform to the proposed collaborative transmission framework. Through theoretical analysis, we derive the expression of the average number of retransmissions related to the sum of independent Gamma variables. Finally, the results of simulation experiments show that the proposed scheme can largely decrease the retransmission latency for railway wireless networks.**

***Index Terms*—C-plane and U-plane decoupling, collaborative transmission, low latency, HARQ, railway communications.**

I. Introduction

RECENT rapid developments of railway technologies and mobile Internet have stimulated increasingly pressing demands on high-speed wireless access. However, no mobile network operators would provide full and dependable wireless coverage for sparsely-populated railway scenarios with low revenue return. Furthermore, due to the limited signal processing capability of mobile equipments, it is difficult to overcome the severe challenges in high-speed railway scenarios, such as large penetration loss, high mobility and fast group handovers [1]. As a result, a novel unified railway wireless network is urgently needed to meet the communication demands of both train control systems and onboard passengers. To this end, industrial participants have reached the consensus that the current narrowband Global System for Mobile Communication for Railway (GSM-R) will evolve to Long Term Evolution (LTE) for railway communication systems in the near future [2], [3]. In addition, a C/U-plane decoupled railway wireless network was proposed [4], [5]. To address the problems caused by the direct connection between onboard passengers and wayside base stations, an access point (AP) is deployed inside the train to collect passengers' services in this network. These services are then relayed to wayside base stations by a mobile relay (MR) deployed on the roofs outside the train [1], [6]. More importantly, a concept of decoupling C-plane and U-plane is applied to this network. To guarantee the transmission reliability and coverage, the more important C-plane is kept in the macro cells with relatively high-quality lower frequency bands, i.e., conventional 800 MHz $\sim$ 2 GHz frequency bands. In contrast, as the main data carrier, the U-plane needs more transmission capacity, thereby being moved to small cells operating at broadband higher frequency bands, including frequency bands higher than 5 GHz, up to 300 GHz. In [4], to provide a better understanding of this network architecture, we studied the bandwidth matching between macro cells and small cells, where 96 passengers' C-plane can be supported in a subframe with a bandwidth of 5 MHz in macro cells, and the corresponding required bandwidth in small cells is 21.2 MHz with a date rate of 500 Kbps per passenger. With more available bandwidth in macro cells, more passengers can be served, and therefore more bandwidth is required in small cells. Besides, as the data rate of passengers increases, the required bandwidth in small cells will also be enlarged. Considering more stringent requirements for transmission reliability, train control information is entirely kept at macro cells without decoupling. Through a newly introduced interface, namely, X3, macro cells and small cells can interact and synchronize with each other. From a purely technical point of view, X3 interfaces employ the same protocols of traditional X2 interfaces in LTE networks [7].

Manuscript received September 30, 2015; revised November 13, 2015; accepted January 11, 2016. Date of publication February 15, 2016; date of current version July 29, 2016. This work was supported in part by the 973 Program of China under Grant 2012CB316100, by the National Natural Science Foundation of China under Grant 61471303, and by the EU FP7 QUICK project under Grant PIRSES-GA-2013-612652. The Associate Editor for this paper was F.-Y. Wang.

L. Yan and X. Fang are with the Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: liyan12047001@my.swjtu.edu.cn; xmfang@swjtu.edu.cn).

G. Min is with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, U.K. (e-mail: g.min@exeter.ac.uk).

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: fang@ece.ufl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2518189

Through retransmissions to ensure transmission reliability, the hybrid automatic repeat request (HARQ) protocol plays an important role in LTE networks [8]–[11]. However, in C/U-plane decoupled railway wireless networks, the C-plane signaling, including HARQ acknowledgments, is transmitted on relatively high-quality lower frequency bands, while the U-plane data are carried by higher frequency bands. In other words, the ACK/NACK message receiving and data transmitting of an HARQ process are in a macro cell and a small cell, respectively, not in the same network node. Therefore, macro cells need to forward the received HARQ acknowledgments to small cells via X3 to instruct small cells what to do next, and handle retransmissions if the acknowledgement is an NACK or send new data if it is an ACK. Obviously, the decoupling C-plane and U-plane cause more overhead in X3. Since in this paper we focus on the HARQ retransmission latency, the overhead is taken into account in the form of forwarding latency of X3 interfaces. According to the field test results in [12], the forwarding latency of X2 interfaces is about 1 ms on average and 6 ms at maximum. Because X3 interfaces adopt the same protocols as X2 interfaces, the above results also apply to X3 interfaces. Consequently, the time interval between two adjacent (re)transmissions of an HARQ process in C/U-plane decoupled railway wireless networks is enlarged. How to reduce the aggravated latency of HARQ retransmissions in this network becomes a significant challenge.

In the conventional network architecture, C-plane and U-plane share the same frequency resources in macro cells. The available resources for the C-plane in LTE networks are very limited because only the first three OFDM symbols in a subframe are used for control channels [11]. In contrast, in the C/U-plane decoupled railway wireless network, the resources of lower frequency bands in macro cells are all used for the C-plane. As a consequence, some relatively high-quality and spare spectrum resources may become available in macro cells, which could be further exploited to assist small cells to handle retransmissions, aiming at reducing the retransmission latency. To sum up, the main contributions and novelties of this paper include:

1) The timing of the conventional HARQ scheme under C/U-plane decoupled railway wireless networks is analyzed to investigate how decoupling C-plane and U-plane can affect the retransmission latency of HARQ, demonstrating the critical issue of aggravating the latency problem of HARQ in C/U-plane decoupled railway wireless networks.
2) To reduce the HARQ retransmission latency in C/U-plane decoupled railway wireless networks, we propose a collaborative HARQ scheme in which the potential spare lower frequency band resources in macro cells excluding those used for C-plane transmissions are exploited to help small cells handle retransmissions. Furthermore, double copies of data are concurrently transmitted in macro cells and small cells in the process of retransmission in order to reduce the total number of required retransmissions and the retransmission latency. Accordingly, the timing of the proposed collaborative HARQ scheme is designed. Considering the significant difference of signal propagation between the lower and higher frequency bands, the process of data combining is studied from the hardware point of view.
3) To determine how to realize collaborative retransmissions between different network nodes in macro cells and small cells, a collaborative transmission framework between macro cells and small cells is proposed to reduce the aggravated latency and enhance the flexibility of bandwidth extension for C/U-plane decoupled railway wireless networks. This framework can also be generalized to common U-plane data collaborative transmissions beyond railway wireless networks. To this end, the channel mapping is redesigned to conform to this framework.
4) For both the conventional and proposed schemes, we conduct theoretical analysis of the average number of retransmissions related to the sum of independent Gamma variables. Based on the analytical results, the average number of retransmissions, the average retransmission latency and the average transmission rate of the conventional and proposed schemes are compared through numerical experiments.

The remainder of this paper is organized as follows. In Section II, we describe the timing of the conventional HARQ scheme applied to C/U-plane decoupled railway wireless networks and present the aggravated retransmission latency problem. In Section III, the collaborative HARQ scheme to mitigate this problem is proposed. In Section IV, to enhance the flexibility of bandwidth extension of the C/U-plane decoupled railway wireless networks, a general collaborative transmission framework is proposed. In Section V, analytical results for the average number of retransmissions related to the sum of independent Gamma variables are derived. In Section VI, simulation results are illustrated and analyzed in details. Finally, conclusions are drawn in Section VII.

II. Conventional HARQ Scheme in C/U-Plane Decoupled Railway Wireless Networks

This section presents the C/U-plane decoupled railway wireless network architecture and then discusses the problems of timing and aggravated retransmission latency of the conventional HARQ scheme working in this network architecture.

A. *C/U-Plane Decoupled Railway Wireless Networks*

In order to increase the system capacity to meet the wireless access demands of train passengers, a C/U-plane decoupled railway wireless network was proposed in [4], [5]. As shown in Fig. 1, an AP and an MR are deployed on the roofs inside and outside trains, respectively, so as to provide a dependable connection for train passengers. The inside AP firstly collects the services of train passengers, and then these services are forwarded to the wayside base stations via the MR. In this way, the problems caused by the direct connection between onboard passengers and wayside base stations, such as large penetration loss, high mobility and group handovers, can be avoided. According to [1], backhaul links between MRs and wayside

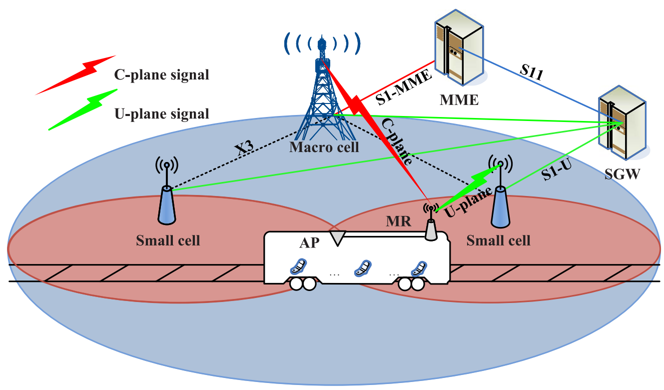


Fig. 1. C/U-plane decoupled railway wireless networks.

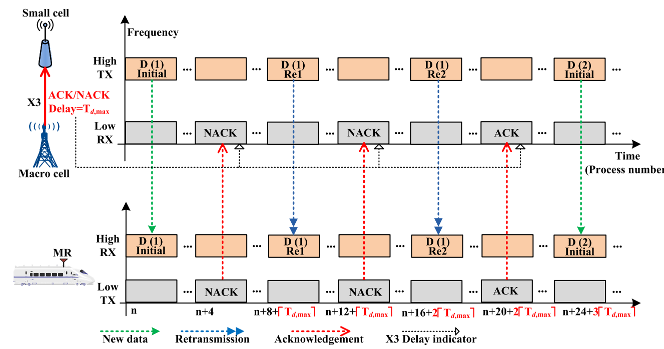


Fig. 2. Timing of the conventional HARQ scheme in C/U-plane decoupled railway wireless networks.

base stations are the key capacity bottlenecks. Therefore, the C/U-plane decoupled architecture is applied to these links. To enable efficient mobility support, the critical C-plane is kept in macro cells operating at relatively high-quality lower frequency bands. The corresponding U-plane, which is the main data carrier, is moved to broadband but relatively poor-quality higher frequency bands to expand the system capacity. Specifically, some low-rate services with absolute demands on transmission reliability, e.g., train control information, can be entirely distributed to macro cells without decoupling. Since small cells are only responsible for the U-plane without handling any control functions, they are solely connected to the SGW, but not to the MME as shown in Fig. 1. Via X3 interfaces, macro cells and small cells can exchange control signaling and data as well as synchronize with each other. In order to provide some low-rate services that have stringent requirements for transmission reliability, in addition to the MME, macro cells can also be connected to the SGW.

B. Timing of the Conventional HARQ in C/U-Plane Decoupled Railway Wireless Networks

Stop and Wait (SaW) is a typical HARQ retransmission mechanism where the next transmission is performed only when the acknowledgment to the previous transmission is received [11]. To fully utilize the resources while waiting for acknowledgments, multiple HARQ processes are carried out. In the conventional network architecture, the number of synchronous HARQ processes in frequency division duplex (FDD) systems is eight, i.e., the maximum latency caused by a retransmission is 8 ms. While in the C/U-plane decoupled railway wireless networks, the maximum latency of HARQ is $8 + \lceil T_{d,\max} \rceil$, where $\lceil T_{d,\max} \rceil$ denotes the maximum forwarding latency of X3 interfaces. According to the field test results in [12], the forwarding latency of X2 interfaces is about 6 ms at maximum. Then, the time interval between two adjacent (re)transmissions of an HARQ process in this network is 14 ms, which is much larger than that of HARQ in the conventional networks. For the asynchronous HARQ and time division duplex (TDD) systems, the time interval is even larger. For clarity, in this paper we take the downlink synchronous HARQ of FDD systems as a case study. Nevertheless, the same analysis can be generalized to other cases as well. Moreover, to achieve a compromise between the system complexity and performance, chase combining in which retransmissions contain the same set of coded bits as the initial transmission [13] is adopted to combine the retransmitted data and the initially transmitted data [14]–[16]. After each retransmission, the receiver uses maximum ratio combination (MRC) to combine each received signals and finally feeds them to the decoder.

Fig. 2 illustrates the timing of the conventional HARQ scheme applied to C/U-plane decoupled railway wireless networks. For clarity, we focus on the timing of one HARQ process. In downlink, the small cell transmits the initial data (1) (denoted by D(1)) on subframe n. After decoding the received data, the MR generates the corresponding acknowledgments. To guarantee the transmission reliability, the HARQ acknowledgments, as a kind of C-plane signaling, are kept at relatively high-quality lower frequency bands in macro cells. In the situation shown in Fig. 2, the data in the initial transmission are not correctly received. Thus, the MR generates an NACK and sends it to the macro cell on subframe $(n + 4)$. Then, the macro cell forwards this NACK to the small cell via X3. In consideration of the forwarding latency of X3 interfaces, the subsequent retransmission for these erroneously received data is performed on subframe $(n + 8 + \lceil T_{d,\max} \rceil)$. As shown in Fig. 2, two retransmissions are performed before the data are successfully decoded. For D(1), the total latency caused by retransmissions is $2 \times (8 + \lceil T_{d,\max} \rceil) = (16 + 2\lceil T_{d,\max} \rceil)$ ms. Compared to the conventional networks in which the latency caused by two retransmissions is 16 ms, $2\lceil T_{d,\max} \rceil$ ms extra latency is brought in, aggravating the latency problem in HARQ especially for latency-sensitive services. For clarity, the wireless transmission delay is neglected in Fig. 2. Therefore, the subframes, on which the same data are transmitted and received, are aligned in the time domain.

III. Proposed Collaborative HARQ Scheme

In the above section, the aggravated retransmission latency problem of the conventional HARQ scheme working in C/U-plane decoupled railway wireless networks is discussed. To mitigate the problem, the principle and timing of the proposed collaborative HARQ scheme are presented in this section.

A. Principle of the Collaborative HARQ Scheme

Due to the change in C/U-plane decoupled railway wireless networks where the whole lower frequency bands of macro

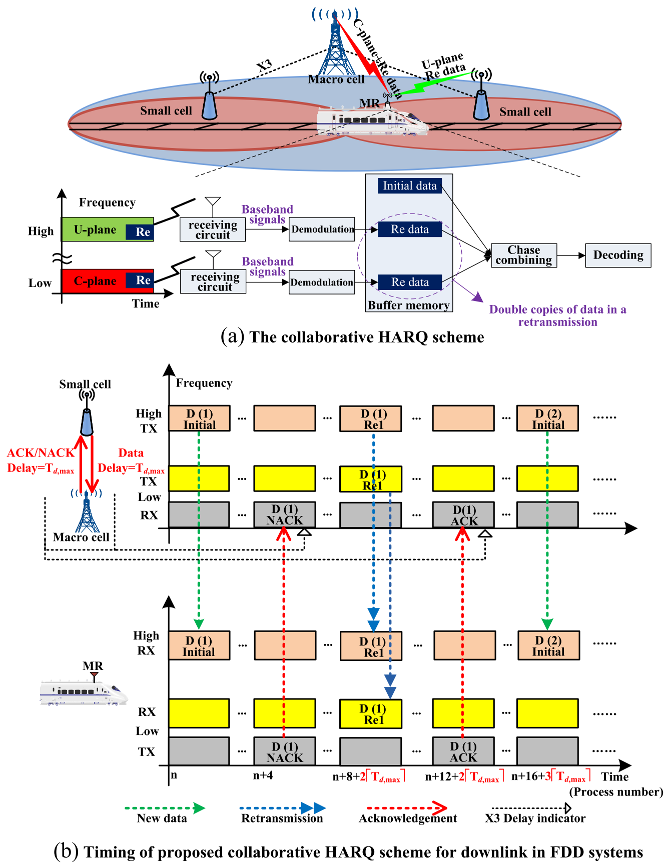


Fig. 3. (a) Principle of the proposed collaborative HARQ scheme for C/U-plane decoupled railway wireless networks. (b) Timing of the proposed scheme.

cells are used for C-plane transmissions, there may be spare spectrum resources on these high-quality bands. Suppose that the bandwidth of a macro cell is 10 MHz corresponding to 8400 available resource elements (REs) in 1 ms, and 30 percent of 1000 users in a train are active and scheduled in a frame of 10 ms. To make a compromise, two control channel elements (CCEs), consisting of 72REs are assigned to the physical downlink control channel (PDCCH) of a user, and every eight users share a physical HARQ indicator channel (PHICH) as in conventional LTE networks. Then, the total consumed resources by control channels and reference signals in 1 ms are 3008REs and there are still 5392REs unused.

Based on this observation, to mitigate the aforementioned retransmission latency problem, a low-latency collaborative HARQ scheme that exploits the possible spare resources in macro cells to help small cells handle retransmissions is proposed, as shown in Fig. 3(a). Considering the large signal propagation difference between discontinuous higher and lower frequency bands, if necessary, two dedicated antennas and receiving circuits are integrated at the MR side to individually handle the radio frequency (RF) signals received from macro cells and small cells. Obviously, compared to the conventional network architecture, although the C/U-plane decoupled network architecture can greatly increase the system capacity, it comes at the expense of higher hardware complexity. After converting RF signals to baseband signals, the receiving circuits output the two parallel baseband signals of a retransmission to the demodulators. Then, through chase combining, the two demodulated bit streams are combined with the initial transmission stored in the buffer memory. Finally, the combined signals are fed to the decoder for a new decoding attempt.

B. Timing of the Collaborative HARQ Scheme

Correspondingly, the downlink timing of the proposed collaborative HARQ scheme is depicted in Fig. 3(b). After receiving an NACK on subframe $(n+4)$, which is the HARQ acknowledgment to the initially transmitted D(1) on subframe n, the macro cell forwards it to the small cell via X3, resulting in a maximum $T_{d,\max}$ latency. If there are spare spectrum resources in the macro cell, the small cell will forward the retransmitted data to the macro cell via X3. Similarly, a maximum $T_{d,\max}$ latency is induced again. As a case study, under the same situation as depicted in Fig. 2, in the proposed scheme the data are correctly decoded after one retransmission which actually includes double copies of the retransmitted data. The latency of a retransmission with double copies of data concurrently transmitted in the proposed scheme is $(8+2\lceil T_{d,\max}\rceil)$ ms. Compared to the conventional scheme shown in Fig. 2, under the same situation, the total latency is reduced by 8 ms in the proposed scheme. When D(1) is correctly decoded, the MR feeds back an ACK to the macro cell on the subsequent subframe after 4 ms. Then, with a maximum $T_{d,\max}$ latency, the acknowledgment is further forwarded to the small cell. As it is an ACK, the small cell directly performs another new data transmission on subframe $(n+16+3\lceil T_{d,\max}\rceil)$ without the need to forward data to the macro cell any more. Based on the above analysis, we can figure out the timing between two adjacent (re)transmissions in the proposed collaborative HARQ scheme. With respect to a retransmission, the time interval between this retransmission and the previous (re)transmission is $(8+2\lceil T_{d,\max}\rceil)$ ms. The time interval between a new data transmission and the previous (re)transmission is $(8+\lceil T_{d,\max}\rceil)$ ms. According to these timing, the receiver can receive all data in sequence accurately.

IV. Proposed Collaborative Transmission Framework

Different from the conventional network architecture, C-plane and U-plane in the C/U-plane decoupled wireless network architecture are completely separated into different network nodes. Therefore, channel mappings need to be redesigned. In this paper, we take LTE protocols as a design benchmark. As shown in Fig. 4, in the newly redesigned channel mapping method, all control channels are kept at macro cells, and all traffic channels are moved to small cells. In macro cells, a new MAC uplink (or downlink) control channel (i.e., U(D)L-CCH) is defined to carry some of up-layer RLC control channels, which is finally mapped to a converged physical uplink (or downlink) control channel (CPU(D)CCH). In small cells, a new MAC traffic channel, uplink (or downlink) traffic channel (U(D)L-TCH), is defined and finally mapped to a physical uplink (or downlink) traffic channel (PU(D)TCH). In terms of physical resources, all C-plane channels are eventually mapped to lower frequency bands, and all U-plane channels

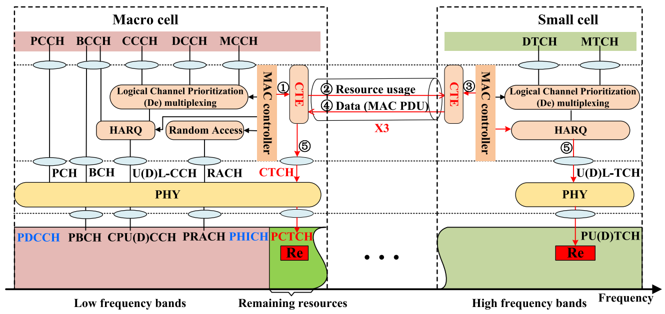


Fig. 4. Collaborative transmission framework.

are mapped to higher frequency bands. Therefore, the physical control format indicator channel (PCFICH) in the conventional coupled LTE networks, which is used to discriminate the boundary of the control and traffic channels sharing the same frequency bands, is saved in the C/U-plane decoupled network architecture.

To realize collaborative transmissions between two different network nodes, i.e., macro cells and small cells, new peer entities, namely, collaborative transmission entities (CTEs) are proposed to be deployed in the MAC layers of macro cells and small cells. Through an X3 interface, the CTEs in different base stations can communicate with each other. Note that only the potential spare resources in collaborative transmissions macro cells are exploited to help transmit U-plane data of small cells. The real control of these U-plane data transmissions, including HARQ decisions, is still under the control of small cells. Therefore, as shown Fig. 4, the forwarded U-plane MAC PDUs from small cells are directly mapped to physical resources of macro cells without going through the HARQ module of macro cells. After receiving the data, MRs will send the corresponding HARQ acknowledgements contained in PHICH on more dependable lower frequency bands, and then they are forwarded by macro cells to small cells. Finally, based on this information, small cells make decisions for the forthcoming transmissions. For clarity, the whole collaborative transmission process consists of the following five steps:

- Step 1: In the macro cell, its MAC controller periodically informs its CTE of the resource usage of lower frequency bands.
- Step 2: Via X3, the resource usage of the macro cell is forwarded by its CTE to the CTE of the small cell.
- Step 3: If there are spare spectrum resources in the macro cell, the MAC controller of the small cell will transfer some data to its own CTE.
- Step 4: The CTE of the small cell forwards these data to the CTE of the macro cell. Then, based on the current system states, the MAC controller of the macro cell makes a decision on the resources and transmission formats used for collaborative transmissions.
- Step 5: With the help of the macro cell, the U-plane data can be transmitted on both lower and higher frequency bands, enhancing the spectrum utilization of lower frequency bands. After receiving the data, MRs will feed back HARQ acknowledgements contained in PHICH on more dependable lower frequency bands. Then, macro cells forward these acknowledgements to small cells via X3, based on which small cells make decisions for the following transmissions.

Moreover, as shown in Fig. 4, in the macro cell, a new MAC channel, namely, collaborative transmission channel (CTCH) and a new PHY channel, namely, physical collaborative transmission channel (PCTCH) are brought in to accommodate collaborative transmissions. To instruct MRs to precisely receive data in the framework, the message indicating whether a collaborative transmission is active or not and the resources as well as the transmission formats used for transmissions in macro cells and small cells is carried by PDCCHs. From the physical resource point of view, the proposed framework equivalently establishes a virtual bridge between completely discontinuous lower and higher frequency bands, and thus enhancing the flexibility of bandwidth extension for C/U-plane decoupled railway wireless networks. Considering the general applicability, the framework is described from a more general perspective above. Although the system capacity can be intuitively increased if the possible spare resources are used to transmit the general U-plane data, for broadband C/U-plane decoupled railway wireless networks the capacity is not a key problem temporarily. Nevertheless, as what will be shown in Section VI, if they are used to help retransmissions, the average retransmission latency can be highly reduced. Therefore, to mitigate the latency problem, in this paper we use this collaborative transmission framework to support collaborative retransmissions between macro cells and small cells. That is, the forwarded data by small cells are retransmitted data, and they are mapped on both higher frequency bands and lower frequency bands in duplication. It is also notable that if there is no spare spectrum resource in macro cells, the system will degenerate to the conventional scheme as shown in Fig. 2.

V. Analytical Models

In this section, we first describe the wireless channel of high-speed railway scenarios. Then, theoretical analysis of the average number of retransmissions related to the sum of independent Gamma variables is conducted for the conventional and proposed schemes.

A. Wireless Channel Modeling

Doppler effect in the high-speed movement scenarios is a severe hindrance to high performance transmissions. Fortunately, in the special railway scenarios, the characteristics of determined running directions, regular running tracks and repetitive movements of trains along fixed running tracks lead to a regular, repetitive and predictable Doppler shift curve, thereby making it easier to trace and compensate the Doppler effect under this scenario [17]. Moreover, there have been many research studies focusing on the Doppler effect estimation and compensation for high-speed railway scenarios, such as in [18], and [19]. Based on this observation, in this paper we assume that the Doppler effect can be perfectly compensated and has almost no influence on the final performance for both the conventional scheme and the proposed scheme, which also ensures the fairness when

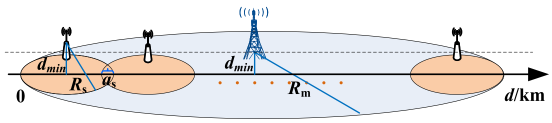


Fig. 5. Geometric sketch for the theoretical analysis.

comparing the performance of two schemes in both theoretical analysis and simulation experiments.

For high-speed railway scenarios, the most typical terrain is viaduct, in which wireless channels can be approximated as line of sight (LOS) following a Rician distribution [20], [21]. To study the combining performance of (re)transmissions, we need to derive the joint probability density function (PDF) and cumulative distribution function (CDF) of multiple Rician variables. However, according to [22], it is almost impossible to obtain an exact analytical joint PDF or CDF for more than three Rician variables. Fortunately, it is widely reported that Nakagami statistics can closely approximate Rician statistics with a relationship between the Rician parameter K and Nakagami parameter m as [23], [24]

$$m = \frac{(K+1)^2}{(2K+1)}. \tag{1}$$

Under the Nakagami model, the received signal to noise ratio (SNR) follows a Gamma distribution, i.e., $\gamma \sim \text{Gamma}(m, m/\overline{\gamma})$, of which the PDF can be expressed as

$$f_\gamma(\gamma) = \left(\frac{m}{\overline{\gamma}}\right)^m \frac{\gamma^{m-1}}{\Gamma(m)} e^{-m\gamma/\overline{\gamma}} \cdot S(\gamma) \tag{2}$$

where $S(\gamma)$ is a unit step function and is defined as

$$S(\gamma) = \begin{cases} 1, & \gamma > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and $\overline{\gamma}$ denotes the average received large-scale SNR, i.e., $\overline{\gamma} = E[\gamma]$, where $E[\cdot]$ is the expectation operator. Fig. 5 describes the geometric sketch to calculate the average large-scale SNR of received signals from macro cells and small cells in C/U-plane decoupled railway wireless networks. In consideration of the linear topology of wireless communication networks in railway scenarios, the macro cells and small cells in Fig. 5 are deployed on a straight line with vertical distance of $d_{\min}$ to the rail. R_s, R_m and a_s denote the radius of a macro cell, the radius of a small cell and the overlapping distance of small cells, respectively. To clarify the analysis, only one macro cell is considered in this paper, i.e., the analysis scope of d is from a_s to $2R_m - a_s$. Since macro cells and small cells operate at different frequency bands with different characteristics such as different coverage radiuses, for clarity we use the subscript m and s to represent the parameters of macro cells and small cells, respectively.

As shown in Fig. 5, suppose that the train starts from the original point and travels through distance of d along the abscissa-axis direction, then the signal propagation distance, $x_m(d)$, from the macro cell to the train is

$$x_m(d) = \sqrt{(d - R_m)^2 + d_{\min}^2}. \tag{4}$$

TABLE I
DESCRIPTIONS OF MCS MODES

	Mode 1	Mode 2	Mode 3
Modulation	QPSK 3/4	16QAM 3/4	64QAM 3/4
R_i(bit/sym.)	1.5	3	4.5
a_i	3.8e5	2.5e5	4.4e4
g_i	5.5	1.3	0.5
γ_i^{th}	2.3	9.2	20.7
AMC$_{th}$(dB)	0.4	8.5	15.6

Similarly, the signal propagation distance, $x_s(d)$, from the current serving small cell to the train is

$$x_s(d) = \sqrt{\left(d - Di_s \cdot \left\lfloor \frac{d}{Di_s} \right\rfloor - \frac{Di_s}{2}\right)^2 + d_{\min}^2} \tag{5}$$

where $Di_s = 2R_s - a_s$ is defined to simplify the expression of Eq. (5).

Generally, the average received large-scale SNR can be calculated as

$$\overline{\gamma}(x) = \frac{P_t \cdot PL(x)}{N_0} \tag{6}$$

where P_t is the transmit power, $PL(x)$ is the large-scale path loss, and N_0 is the noise power. Then, the average large-scale SNR of received signals from the macro cell and the current serving small cell can be separately expressed as

$$\overline{\gamma}_m(x_m) = \frac{P_{m,t} \cdot PL_m(x_m)}{N_0} \tag{7}$$

$$\overline{\gamma}_s(x_s) = \frac{P_{s,t} \cdot PL_s(x_s)}{N_0}. \tag{8}$$

Based on [25], for a given modulation and coding scheme (MCS) i, the packet error rate (PER) related to the instantaneous SNR can be approximated as

$$F_i(\gamma) \approx \begin{cases} 1, & \text{if } \gamma < \gamma_i^{th} \\ a_i \exp(-g_i \gamma), & \text{if } \gamma > \gamma_i^{th} \end{cases} \tag{9}$$

where a_i, g_i, and γ_i^{th} are three parameters determined by MCS i as listed in Table I [25], [26].

B. Theoretical Analysis of the Conventional HARQ Scheme

To make the analysis more understandable, the effects of incorrect receiving or losing of HARQ acknowledgments on the overall performance are not considered. Let $N_{\max}^{\text{con}}$ denote the maximum number of permissible retransmissions for the conventional HARQ scheme in C/U-plane decoupled railway wireless networks. Since in the conventional HARQ scheme small cells perform all initial transmissions and retransmissions, for MCS i_s where $i_s \in \{1, 2, 3\}$, the average number of required retransmissions of the conventional scheme, excluding the initial transmission, can be derived as

$$\overline{L}_{i_s,\text{con}} = E\left[\sum_{n=1}^{N_{\max}^{\text{con}}} \prod_{\ell=0}^{n-1} F_{i_s}\left(\sum_{j=0}^{\ell} \gamma_{s,j}\right)\right] \tag{10}$$

where $\gamma_{s,0}$ is the received SNR of the initial transmission in the small cell. The detailed derivation of Eq. (10) is given in Appendix A.

Suppose that the velocity of the train is $v = 360$ km/h and the center frequency of small cells is $f_s = 5$ GHz. Then, we can get the maximum Doppler Shift in small cells as $f_{d,s} = vf_s/c = 1.67$ kHz, where c is the light speed. Correspondingly, the coherence time of wireless channels in small cells is $T_{c,s} = 0.423/f_{d,s} = 0.25$ ms [27], which is much shorter than the time interval between two adjacent (re)transmissions of an HARQ process. Thus, it is reasonable to assume that the (re)transmissions of an HARQ process experience independent wireless channels. Additionally, taking LTE networks as the analysis baseline, the value of $N_{\max}^{\text{con}}$ is 6 [28], and then the time interval between the initial transmission and the sixth retransmission is $(48 + 6\lceil T_{d,\max}\rceil)$ ms. Based on the field test results in [12], let $T_{d,\max} = 6$ ms. And then $48 + 6\lceil T_{d,\max}\rceil = 84$ ms. During the time interval of 84 ms, the running distance of the train is 8.4 m, which is relatively much shorter than the distance between base stations and the rail. Therefore, all (re)transmissions in an HARQ process can be assumed to have the same average large-scale SNR. As a result, $\gamma_{s,j}$ with $j = 0, 1, 2\ldots$ in Eq. (10) follows an independent and identical distribution. According to [29], we can get $\Upsilon_{s,\ell} = \sum_{j=0}^{\ell}\gamma_{s,j} \sim \text{Gamma}((\ell+1)m_s, m_s/\overline{\gamma}_s)$. Therefore, the average number of retransmissions of the conventional HARQ scheme in C/U-plane decoupled railway wireless networks can be rewritten as

$$\overline{L}_{i_s,\text{con}} = \sum_{n=1}^{N_{\max}^{\text{con}}}\prod_{\ell=0}^{n-1}\left(\frac{\Gamma_{\text{low}}\left(\frac{m_s\gamma_{i_s}^{th}}{\overline{\gamma}_s}, (\ell+1)m_s\right)}{\Gamma\left((\ell+1)m_s\right)} + a_{i_s}\left(1+g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-(\ell+1)m_s} \times\left(1-\frac{\Gamma_{\text{low}}\left(\left(\frac{m_s}{\overline{\gamma}_s}+g_{i_s}\right)\gamma_{i_s}^{th}, (\ell+1)m_s\right)}{\Gamma\left((\ell+1)m_s\right)}\right)\right) \tag{11}$$

where $\Gamma_{\text{low}}(y, m)$ is the lower incomplete Gamma function, defined as $\Gamma_{\text{low}}(y, m) = \int_0^y x^{m-1}e^{-x}dx$. The detailed derivation of Eq. (11) is given in Appendix A.

With the obtained average number of retransmissions, the average latency caused by retransmissions of the conventional HARQ scheme in C/U-plane decoupled railway wireless networks can be expressed as

$$\overline{D}_{\text{con}} = (8 + \lceil T_{d,\max}\rceil)\,\overline{L}_{i_s,\text{con}}. \tag{12}$$

In the conventional HARQ scheme, only a copy of the data that are unsuccessfully decoded in the initial transmission is carried in a retransmission. Accordingly, only a copy of resource is consumed in a retransmission. For MCS i_s, after $\overline{L}_{i_s,\text{con}}$ retransmissions, the average system transmission rate is

$$\overline{R}_{i_s,\text{con}} = \frac{R_{i_s}}{\overline{L}_{i_s,\text{con}} + 1}. \tag{13}$$

C. Theoretical Analysis of the Proposed Collaborative HARQ Scheme

In the proposed collaborative HARQ scheme, initial transmissions are carried by small cells, and the following retransmissions will be collaboratively accomplished by small cells and macro cells. Due to the differences of signal propagation characteristics on lower and higher frequency bands, for the train at the same geographic location as shown in Fig. 5, the average large-scale SNR of received signals from macro cells and small cells may be different. Hence, the macro cell and small cell probably choose different MCS for the two copies of data in a single retransmission based on the assumption that the adaptive modulation and coding (AMC) technique is used. In this case, we cannot obtain the exact analytical expression of the average number of retransmissions for the proposed scheme. Nevertheless, based on the MCS with a higher order among the two MCSs selected by the macro cell and small cell, the upper limit of the average number of required retransmissions can be obtained for the proposed scheme. Similarly, based on the MCS with a lower order, the lower limit of the average number of required retransmissions can be derived. Let i_m denote the selected MCS by the macro cell, where $i_m \in \{1, 2, 3\}$. As listed in Table I, the lower the values of i_s and i_m are, the lower the order of MCSs is. With the definition of $x = \max(i_m, i_s)$, the upper limit of the average number of required retransmissions in the proposed collaborative HARQ scheme is obtained as

$$\overline{L}_{\text{pro}} \le E\left[F_{i_s}(\gamma_{s,0}) + \sum_{n=2}^{N_{\max}^{\text{pro}}} F_{i_s}(\gamma_{s,0}) \times\prod_{\ell=1}^{n-1} F_x\left(\sum_{j=0}^{\ell}\gamma_{s,j} + \sum_{j=1}^{\ell}\gamma_{m,j}\right)\right]. \tag{14}$$

The detailed derivation of Eq. (14) is given in Appendix B. For clarity, we separately conduct the derivation for each item in Eq. (14). Based on the above analysis, we can obtain

$$\begin{aligned} E\left[F_{i_s}(\gamma_{s,0})\right] &= P\left(\gamma_{s,0} < \gamma_{i_s}^{th}\right) + a_{i_s}e^{-g_{i_s}\gamma_{s,0}}P\left(\gamma_{s,0} > \gamma_{i_s}^{th}\right) \\ &= \frac{\Gamma_{\text{low}}\left(\frac{m_s\gamma_{i_s}^{th}}{\overline{\gamma}_s}, m_s\right)}{\Gamma(m_s)} + a_{i_s}\left(1+g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-m_s} \\ &\quad\times\left(1-\frac{\Gamma_{\text{low}}\left((m_s/\overline{\gamma}_s + g_{i_s})\gamma_{i_s}^{th}, m_s\right)}{\Gamma(m_s)}\right). \end{aligned} \tag{15}$$

Similarly, suppose that the frequency center of macro cells is $f_m = 2$ GHz. Then, the coherence time of wireless channels in macro cells is $T_{c,m} = 0.423/(vf_m/c) = 0.63$ ms, which is much shorter than the time interval between two adjacent (re)transmissions in the proposed scheme. Therefore, we can assume that the wireless channels experienced by (re)transmissions of an HARQ process in the macro cell are independent on each other. Let $T_{d,\max} = 6$ ms. Considering that double copies of data are concurrently transmitted in a retransmission in the proposed scheme, the maximum number of permissible retransmissions of the proposed scheme is set to $N_{\max}^{\text{pro}} = 3$. Then, the time interval between the initial transmission and the third retransmission is $3 \times (8 + 2\lceil T_{d,\max} \rceil) =$ 60 ms. During this period, the running distance of the train is 6 m, which is relatively much smaller than the distance between base stations and the rail. Hence, the average large-scale SNR of all (re)transmissions in an HARQ process is almost equivalent. As a consequence, it can be assumed that wireless channels experienced by all (re)transmissions of an HARQ process in the macro cell follow an independent and identical distribution. Then,

$$\Upsilon_{s,\ell} = \sum_{j=0}^{\ell} \gamma_{s,j} \sim \text{Gamma}\left((\ell+1)m_s, \frac{m_s}{\overline{\gamma}_s}\right)$$
$$\Upsilon_{m,\ell} = \sum_{j=1}^{\ell} \gamma_{m,j} \sim \text{Gamma}\left(\ell m_m, \frac{m_m}{\overline{\gamma}_m}\right). \tag{16}$$

For clarity, we define

$$\alpha_{s,\ell} = (\ell+1)m_s, \frac{1}{\beta_s} = \frac{m_s}{\overline{\gamma}_s}$$
$$\alpha_{m,\ell} = \ell m_m, \frac{1}{\beta_m} = \frac{m_m}{\overline{\gamma}_m}. \tag{17}$$

As macro cells and small cells operate at completely different frequency bands, we can assume that transmissions in macro cells and small cells are diverse and independent. That is, $\Upsilon_{s,\ell}$ and $\Upsilon_{m,\ell}$ are two independent Gamma variables with different parameters. According to [30], [31], the PDF of the sum of the two variables, denoted by $Y = \Upsilon_{s,\ell} + \Upsilon_{m,\ell}$, can be derived as

$$g(y) = \left(C \sum_{\upsilon=0}^{\infty} \frac{\delta_\upsilon y^{\rho+\upsilon-1} e^{-y/\beta_q}}{\Gamma(\rho+\upsilon)\beta_q^{\rho+\upsilon}}\right) \cdot S(y) \tag{18}$$

where q, C, ρ and δ are interim coefficients to simplify the expression of Eq. (18), and

$$q = \min_{q \in \{s,m\}} (\beta_s, \beta_m)$$
$$C = \left(\frac{\beta_q}{\beta_{\{s,m\}-q}}\right)^{\alpha_{\{s,m\}-q,\ell}}$$
$$\rho = \alpha_{s,\ell} + \alpha_{m,\ell}$$
$$\begin{cases} \delta_{\upsilon+1} = \frac{1}{\upsilon+1} \sum\limits_{\kappa=1}^{\upsilon+1} \alpha_{\{s,m\}-q,\ell} \left(1 - \frac{\beta_q}{\beta_{\{s,m\}-q}}\right)^{\kappa} \delta_{\upsilon+1-\kappa} \\ \delta_0 = 1. \end{cases} \tag{19}$$

Then, we can obtain the expectation of the last item in Eq. (14) as

$$E\left[F_x\left(\sum_{j=0}^{\ell} \gamma_{s,j} + \sum_{j=1}^{\ell} \gamma_{m,j}\right)\right]$$
$$= C \sum_{\upsilon=0}^{\infty} \delta_\upsilon \left(\frac{\Gamma_{\text{low}}\left(\gamma_x^{th}/\beta_q, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)} + a_x(1+g_x\beta_q)^{-(\rho+\upsilon)} \times \left(1 - \frac{\Gamma_{\text{low}}\left((1/\beta_q + g_x)\gamma_x^{th}, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)}\right)\right). \tag{20}$$

Considering the practical application, we take the first Ω items of the infinite series in Eq. (20) as an approximation, that is,

$$E\left[F_x\left(\sum_{j=0}^{\ell} \gamma_{s,j} + \sum_{j=1}^{\ell} \gamma_{m,j}\right)\right]$$
$$\approx C \sum_{\upsilon=0}^{\Omega} \delta_\upsilon \left(\frac{\Gamma_{\text{low}}\left(\frac{\gamma_x^{th}}{\beta_q}, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)} + a_x(1+g_x\beta_q)^{-(\rho+\upsilon)} \times \left(1 - \frac{\Gamma_{\text{low}}\left((1/\beta_q + g_x)\gamma_x^{th}, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)}\right)\right). \tag{21}$$

To substitute the above results into Eq. (14), the upper limit of the average number of retransmissions in the proposed collaborative HARQ scheme, denoted by $\overline{L}_{\text{pro}}^{\text{upper}}$, can be rewritten as Eq. (22), shown at the bottom of the page. Similarly, if we define $z = \min(i_m, i_s)$, the lower limit of the average number of required retransmissions in the proposed scheme, denoted by $\overline{L}_{\text{pro}}^{\text{lower}}$, can be obtained as Eq. (23), shown at the bottom of the next page.

With the above results, the scope of the latency caused by retransmissions in the proposed scheme is

$$(8 + 2\lceil T_{d,\max} \rceil)\, \overline{L}_{\text{pro}}^{\text{lower}} \leq \overline{D}_{\text{pro}} \leq (8 + 2\lceil T_{d,\max} \rceil)\, \overline{L}_{\text{pro}}^{\text{upper}}. \tag{24}$$

In the proposed collaborative HARQ scheme, double copies of data are concurrently transmitted in a retransmission, leading to the consumption of double copies of resources. Therefore,

$$\overline{L}_{\text{pro}} \leq \overline{L}_{\text{pro}}^{\text{upper}} = \frac{\Gamma_{\text{low}}\left(m_s \gamma_{i_s}^{th}/\overline{\gamma}_s, m_s\right)}{\Gamma(m_s)} + a_{i_s}\left(1 + g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-m_s}\left(1 - \frac{\Gamma_{\text{low}}\left((m_s/\overline{\gamma}_s + g_{i_s})\gamma_{i_s}^{th}, m_s\right)}{\Gamma(m_s)}\right)$$
$$+ \sum_{n=2}^{N_{\max}^{\text{pro}}}\left(\left(\frac{\Gamma_{\text{low}}\left(m_s \gamma_{i_s}^{th}/\overline{\gamma}_s, m_s\right)}{\Gamma(m_s)} + a_{i_s}\left(1 + g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-m_s}\left(1 - \frac{\Gamma_{\text{low}}\left((m_s/\overline{\gamma}_s + g_{i_s})\gamma_{i_s}^{th}, m_s\right)}{\Gamma(m_s)}\right)\right)\right.$$
$$\left.\times \prod_{\ell=1}^{n-1}\left(C\sum_{\upsilon=0}^{\Omega}\delta_\upsilon\left(\frac{\Gamma_{\text{low}}\left(\gamma_x^{th}/\beta_q, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)} + a_x(1+g_x\beta_q)^{-(\rho+\upsilon)}\left(1 - \frac{\Gamma_{\text{low}}\left((1/\beta_q + g_x)\gamma_x^{th}, \rho+\upsilon\right)}{\Gamma(\rho+\upsilon)}\right)\right)\right)\right) \tag{22}$$

the scope of the average system transmission rate in the proposed scheme is

$$\frac{R_{i_s}}{2\overline{L}_{\rm pro}^{\rm upper}+1} \leq \overline{R}_{\rm pro} \leq \frac{R_{i_s}}{2\overline{L}_{\rm pro}^{\rm lower}+1}. \tag{25}$$

VI. Performance Analysis and Comparison

Based on the theoretical analyses in Section V, numerical results are provided in this section to compare the performance of the conventional and proposed schemes. Simulation experiments are performed under two conditions, i.e., under fixed MCS and under AMC. Due to the fact that AMC can enhance the transmission reliability in some degree, the retransmission latency performance improvement of the proposed scheme under AMC is less remarkable than that under fixed MCS.

A. Performance Comparisons Under Fixed MCS

In this section, to conduct fair and comprehensive performance comparisons, simulation experiments are performed under the condition that the same MCS is used in small cells and macro cells and is not adaptively changed when the train runs through the whole wireless coverage. In this way, we can investigate the pure performance improvements of the proposed scheme without the influence of AMC techniques on the transmission reliability. Performance comparisons of the average number of retransmissions, the average retransmission latency and the average system transmission rate are illustrated in Figs. 6–8, respectively. Detailed simulation parameter settings are listed in Table II [4], [5]. For clarity, here two small cells symmetrically located in the coverage of a macro cell are considered. Therefore, the following simulation results are symmetrical to the position of the macro cell. For the purpose of performance analysis, we focus on the results of the left-hand side as case studies.

As shown in Fig. 6, overall the proposed scheme requires much fewer retransmissions compared to the conventional scheme. Moreover, the higher the MCS order is, the more remarkable the performance improvement of the proposed scheme is. For instance, under the MCS of 64QAM, when the train is at the edges of the macro cell which also corresponds to the edges of small cells, e.g., at $d = 0.2$ km, due to the low SNR in these regions both schemes arrive at their maximum permissible retransmissions. As the train moves towards the small cell, with double copies of data concurrently transmitted

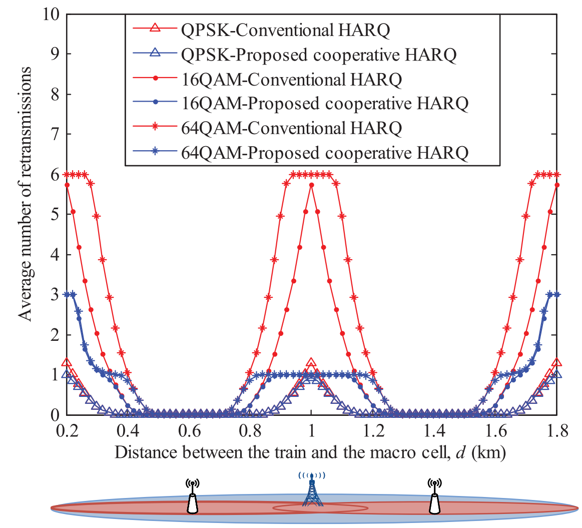


Fig. 6. Performance comparisons of the average number of retransmissions for different MCS modes.

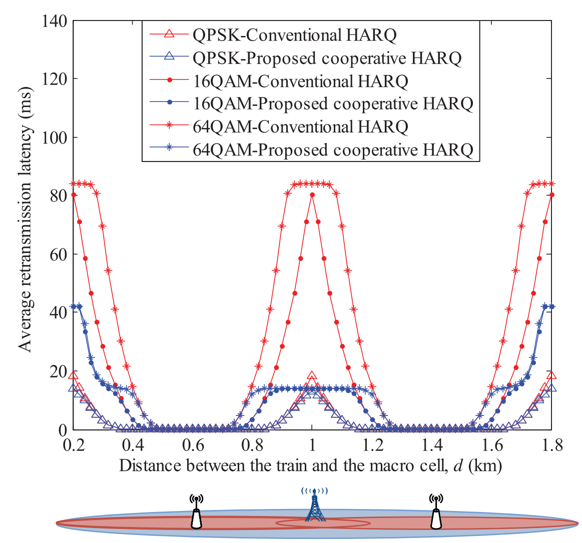


Fig. 7. Performance comparisons of the average retransmission latency for different MCS modes.

in a retransmission, the number of required retransmissions of the proposed scheme is greatly reduced. For instance, at $d = 0.3$ km, the average numbers of retransmissions of the conventional scheme and the proposed scheme are 2 and 1, respectively. At the center areas of small cells in which the

$$\begin{aligned}\overline{L}_{\rm pro} \geq \overline{L}_{\rm pro}^{\rm lower} =& \frac{\Gamma_{\rm low}\left(m_s\gamma_{i_s}^{th}/\overline{\gamma}_s, m_s\right)}{\Gamma(m_s)} + a_{i_s}\left(1+g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-m_s}\left(1-\frac{\Gamma_{\rm low}\left(\left(m_s/\overline{\gamma}_s+g_{i_s}\right)\gamma_{i_s}^{th}, m_s\right)}{\Gamma(m_s)}\right)\\ &+\sum_{n=2}^{N_{\max}^{\rm pro}}\left(\frac{\Gamma_{\rm low}\left(m_s\gamma_{i_s}^{th}/\overline{\gamma}_s, m_s\right)}{\Gamma(m_s)} + a_{i_s}\left(1+g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-m_s}\left(1-\frac{\Gamma_{\rm low}\left(\left(m_s/\overline{\gamma}_s+g_{i_s}\right)\gamma_{i_s}^{th}, m_s\right)}{\Gamma(m_s)}\right)\right)\\ &\times\prod_{\ell=1}^{n-1}\left(C\sum_{v=0}^{\Omega}\delta_v\left(\frac{\Gamma_{\rm low}\left(\gamma_z^{th}/\beta_q, \rho+v\right)}{\Gamma(\rho+v)}+a_z(1+g_z\beta_q)^{-(\rho+v)}\left(1-\frac{\Gamma_{\rm low}\left((1/\beta_q+g_z)\gamma_z^{th}, \rho+v\right)}{\Gamma(\rho+v)}\right)\right)\right)\end{aligned} \tag{23}$$

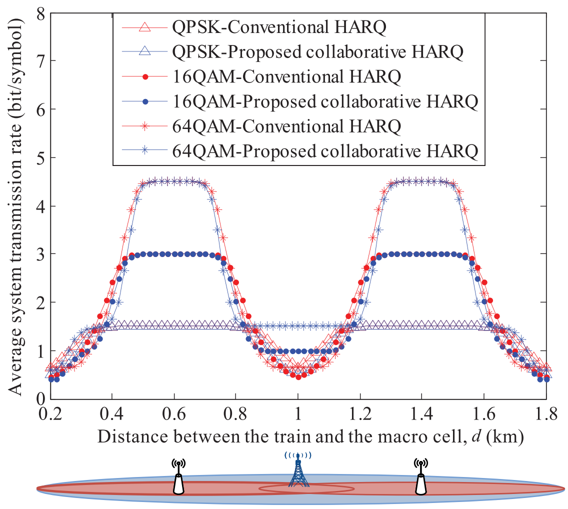


Fig. 8. Performance comparisons of the average system transmission rate for different MCS modes.

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Frequency of the macro cell	2GHz
Frequency of the small cell	5GHz
Transmit power of the macro cell	43dBm
Transmit power of the small cell	30dBm
Path loss model of the macro cell	Hata
Path loss model of the small cell	M.2135
Radius of the macro cell	1km
Radius of the small cell	0.6km
K factor of the macro cell	6dB
K factor of the small cell	8dB
Distance between base station and rail	0.1km
Overlapping area distance	0.2km
$T_{d,max}$	6ms
$N_{\max}^{con}$	6
$N_{\max}^{pro}$	3
Ω	100

SNR of received signals from small cells is high, e.g., at $d \in (0.5, 0.7)$ km, the numbers of required retransmissions of both schemes are close to zero. When the train is at the center area of the macro cell which corresponds to the edges of these two small cells, thanks to the high SNR collaborative retransmissions from the macro cell, the performance improvement of the proposed scheme in this region is the most remarkable. At $d = 1$ km, the number of required retransmissions of the proposed scheme is about 1, while for the conventional scheme it reaches up to 6. For other MCSs, the whole trend is the same as 64QAM. Nevertheless, because lower order MCSs can achieve higher retransmission reliability, as shown in Fig. 6, the lower the MCS order is, the less remarkable the performance improvement is.

Based on the above results, as shown in Fig. 7, with double copies of data concurrently transmitted in a retransmission in the proposed scheme, to achieve the same transmission reliability, the proposed scheme requires fewer retransmissions, thereby mitigating the latency problem caused by HARQ retransmissions. For example, at $d = 1$ km, the average retransmission latency of 64QAM, 16QAM and QPSK is reduced by the proposed scheme from about 84 ms to 14 ms, from about 80 ms to 14 ms and from about 18 ms to 13 ms, respectively. Also double copies of resources are consumed by a retransmission in the proposed scheme. As depicted in Fig. 8, when the train is at the edges of the macro cell, the conventional scheme outperforms the proposed scheme in terms of the average system transmission rate. Nevertheless, with reduced retransmissions in the proposed scheme, the performance gap between two schemes is very small. On the contrary, for 16QAM and 64QAM, at the center area of the macro cell, thanks to the high SNR collaborative retransmissions from the macro cell, much fewer retransmissions are needed in the proposed scheme. As aforementioned, at $d = 1$ km in Fig. 6, the proposed scheme reduces the average number of required retransmissions from about 5.8 to 1 and from 6 to 1 for 16QAM and 64QAM, respectively. Therefore, in this region, the proposed scheme obtains a higher average system transmission rate as shown in Fig. 8. Correspondingly, at $d = 1$ km, the performance improvement of the proposed scheme under 64QAM and 16QAM is about 0.9 bit/symbols and 0.6 bit/symbols, respectively. However, as discussed above, for QPSK which has a higher ability in enhancing the transmission reliability, the performance improvement of the average number of required retransmissions is less remarkable. At $d = 1$ km of Fig. 6, the average number of required retransmissions are reduced from about 1.3 to 0.9 by the proposed scheme. Therefore, in Fig. 8, even at the center area of the macro cell, the conventional scheme still obtains a higher average system transmission rate than the proposed scheme.

B. Performance Comparisons Under AMC

In practical wireless communication systems, AMC techniques are usually employed to adapt to various wireless channels. If a wireless channel is of high quality, higher order MCSs with higher transmission rate will be used to improve the spectrum efficiency. Otherwise, lower order MCSs will be implemented to enhance the transmission reliability. That is to say, AMC can enhance the transmission reliability in some degree. Hence, under AMC, the performance improvement of the proposed scheme will not be very remarkable. To make this more understandable, the MCS adaption results of the macro cell and small cell during the train running through the coverage of the macro cell are illustrated in Fig. 9. The higher the received signal quality is, the higher the MCS order is selected. For the center areas of the macro cell and small cells, the selected MCS mode is 64QAM. In contrast, QPSK is used for cell edges. And 16QAM is used for other areas with intermediate signal quality. For simplicity, in this paper, we only consider the three MCSs listed in Table I, in which the SNR thresholds of MCS switching are given in the last row. Nevertheless, by substituting the corresponding MCS related parameters, the same analysis method also holds for other MCSs.

Fig. 10 depicts the average number of retransmissions of both schemes under AMC, confirming the fact that the proposed scheme still outperforms the conventional scheme. Based on the theoretical analysis, in the regions of $d \in (0.2, 0.34)$ km and $d \in (0.72, 0.74)$ km where both the macro cell and small cell adopt the same MCS as shown in Fig. 9, the obtained number of retransmissions of the proposed scheme in Fig. 10

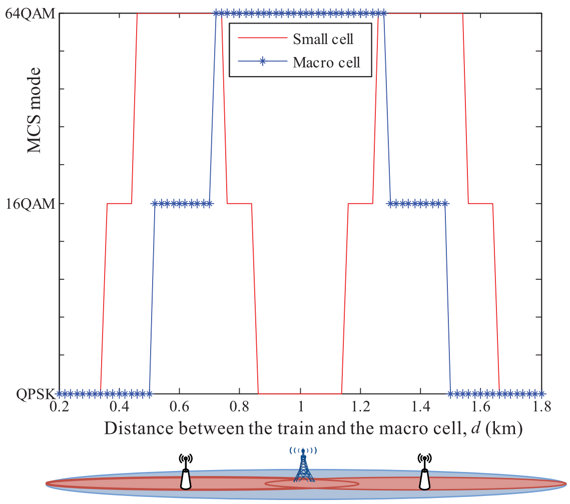


Fig. 9. MCS adaptation results.

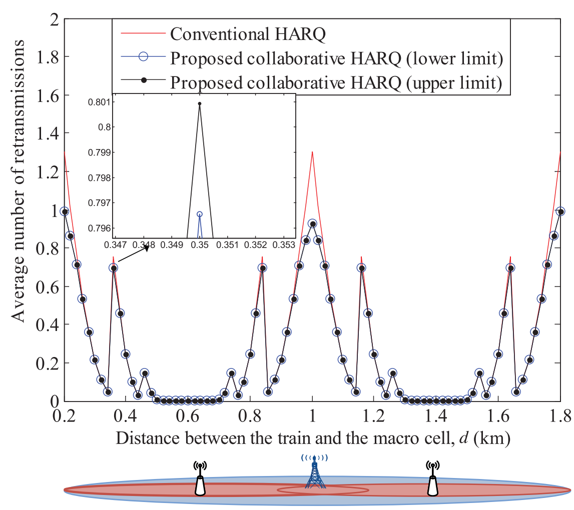


Fig. 10. Performance comparison of the average number of retransmissions under AMC.

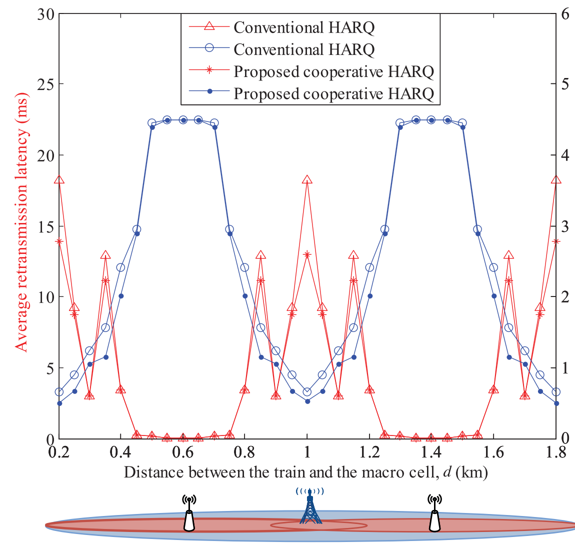


Fig. 11. Performance comparisons of the average retransmission latency and average system transmission rate under AMC.

are exact values, not the upper or lower limit. For other regions, as shown in Fig. 9, the macro cell and small cell always use different MCSs. And we can only obtain the upper limit and lower limit of the average number of retransmissions in those regions for the proposed scheme. Besides, as shown in Fig. 9, when the train is in those regions, either the macro cell or the small cell can provide a high SNR retransmission. Therefore, in the proposed scheme which employs both the macro cell and small cell to handle retransmissions, for all MCSs, the numbers of required retransmissions are very low and the performance difference among them is small. As a consequence, as shown in Fig. 10, the two curves of the upper limit and lower limit of the average number of retransmissions in the proposed scheme are almost overlapped. Then, in the following analysis, we can use one of them as an approximate curve of the real average number of retransmissions for the proposed scheme. Due to the fact that AMC techniques can enhance the transmission reliability in some degree, as shown in Fig. 10 the total number of required retransmissions of both schemes are less than 1.5 and the performance improvement of the proposed scheme is not very remarkable.

Fig. 11 compares the average retransmission latency and the average system transmission rate comparisons between two schemes under AMC. In the proposed scheme, with double copies of data concurrently transmitted in a retransmission, the HARQ retransmission latency is reduced. Nevertheless, as aforementioned, due to the fact that AMC techniques can enhance the transmission reliability in some degree, the performance improvement is not very remarkable. As a kind of sacrifice, also with double copies of resources consumed in a retransmission, the proposed scheme has lower average system transmission rate compared to the conventional scheme. However, the performance gap is very small. In addition, thanks to the concept of C-plane and U-plane decoupling which dramatically expands the bandwidth, the system capacity in this network is not a critical problem. On the contrary, it is very important and necessary to mitigate the latency problem when linking the completely separated C-plane and U-plane in different network nodes, which is also the main goal of this paper.

Alternatively, excluding the spectrum resources used by C-plane transmissions, if there are still spare resources in macro cells, they can also be exploited to transmit general U-plane data via the proposed collaborative transmission framework. Based on this observation, in order to present a more comprehensive performance study, in Fig. 12, the effective transmission rates of spare resources of macro cells used in two different cases, i.e., collaborative HARQ retransmissions and general transmissions, are compared. As shown in Fig. 12, for the case that spare resources are used for HARQ retransmissions, in the regions of $d \in (0.2, 0.5)$ km and $d \in (0.7, 1)$ km these resources are invested to repeatedly retransmit the received erroneous packets, leading to a lower effective transmission rate in these regions compared to the case that these resources are used for general transmissions. Nevertheless, as demonstrated above, with the help of these resources to handle retransmissions, the retransmission latency is highly reduced. As a

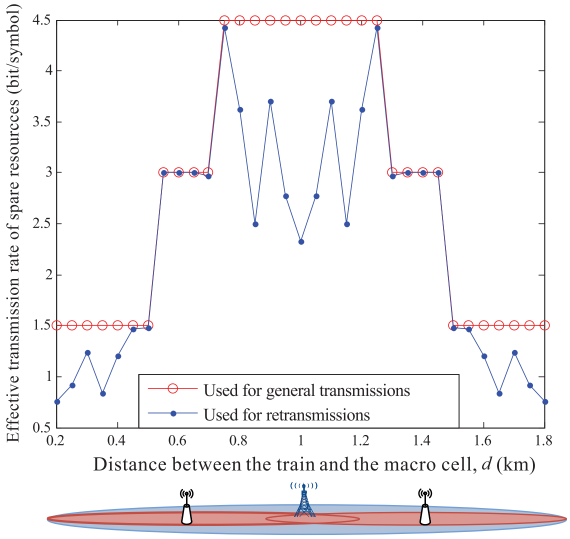


Fig. 12. Performance comparison of the effective transmission rates of spare resources used in different situations.

matter of fact, this is very common for the field of wireless communications where some performance is sacrificed for the improvements of other performance. Fortunately, for broadband C/U-plane decoupled railway wireless networks, the capacity is not a key problem temporarily. On the contrary, the aggravated retransmission latency problem in this network is more pronounced. Hence, it is more beneficial here to utilize the possible spare high-quality resources of macro cells to handle retransmissions.

VII. Conclusion

The recently proposed C/U-plane decoupled railway wireless networks aim to meet dramatically growing wireless access demands of train passengers as well as the reliable transmission requirements of train control systems. Although the whole system capacity is highly increased in this network with the U-plane moved to broadband higher frequency bands, how to mitigate the latency problem when linking the completely separated C-plane and U-plane in different physical nodes becomes important, especially for HARQ protocols which require frequent interactions between the C-plane and U-plane. To address this challenge, in this paper we have proposed a low-latency collaborative HARQ scheme, in which if there are spare spectrum resources in macro cells excluding those used by C-plane transmissions, they will be exploited to help small cells handle retransmissions. To realize collaborative transmissions between two different network nodes, i.e., macro cells and small cells, a novel collaborative transmission framework is proposed. Although the framework in this paper is used for HARQ retransmissions, it can also be developed for general collaborative transmissions to enhance the flexibility of bandwidth extension for C/U-plane decoupled railway wireless networks. Accordingly, the channel mapping is also redesigned to conform to this framework.

Both the theoretical analysis and simulation experiments are conducted under two different conditions, i.e., with and without AMC. No matter under which condition, the proposed scheme always outperforms the conventional scheme in terms of the average retransmission latency. Due to the fact that AMC techniques can enhance the transmission reliability in some degree, under the condition with AMC, the average system transmission rate of the proposed scheme is slightly lower than that of the conventional scheme. Nevertheless, for C/U-plane decoupled railway wireless networks in which broadband higher frequency bands are integrated, the system capacity is not a key point temporarily. In contrast, in this network, how to mitigate the aggravated latency during HARQ retransmissions becomes important, which is exactly the research focus of this paper. As a matter of fact, in the field of wireless communications, sacrificing some spectrum resources to gain the transmission reliability is a common means. For our future work, we will consider the differences of signal propagation characteristics on higher and lower frequency bands, such as different Doppler shifts and frame structures when combining the data received from macro cells and small cells, so as to make the collaborative transmission scheme more practical and further increase the data combining performance.

Appendix A
Derivation of Equations (10) and (11)

$$
\begin{aligned}
\overline{L}_{i_s,\mathrm{con}} &= \Pr\left(Su_{s,1}|Fa_{s,0}\right)+2\Pr(Su_{s,2}|Fa_{s,0},Fa_{s,1})+,\ldots,\\
&\quad+\left(N_{\max}^{\mathrm{con}}-1\right)\Pr\left(Su_{s,N_{\max}^{\mathrm{con}}-1}\left|Fa_{s,0},Fa_{s,1},\ldots,\right.\right.\\
&\qquad\qquad\left.Fa_{s,N_{\max}^{\mathrm{con}}-2}\right)+N_{\max}^{\mathrm{con}}\\
&\quad\times\Pr\left(Fa_{s,0},Fa_{s,1},\ldots,Fa_{s,N_{\max}^{\mathrm{con}}-1}\right)\\
&=E\left[\sum_{n=1}^{N_{\max}^{\mathrm{con}}-1} n\left(1-F_{i_s}\left(\sum_{j=0}^{n}\gamma_{s,j}\right)\right)\right.\\
&\qquad\left.\times\prod_{\ell=0}^{n-1}F_{i_s}\left(\sum_{j=0}^{\ell}\gamma_{s,j}\right)+N_{\max}^{\mathrm{con}}\prod_{\ell=0}^{N_{\max}^{\mathrm{con}}-1}F_{i_s}\left(\sum_{j=0}^{\ell}\gamma_{s,j}\right)\right]\\
&=E\left[\sum_{n=1}^{N_{\max}^{\mathrm{con}}}\prod_{\ell=0}^{n-1}F_{i_s}\left(\sum_{j=0}^{\ell}\gamma_{s,j}\right)\right]\\
&=\sum_{n=1}^{N_{\max}^{\mathrm{con}}}\prod_{\ell=0}^{n-1}\left(P\left(\Upsilon_{s,\ell}<\gamma_{i_s}^{th}\right)+a_{i_s}e^{-g_{i_s}\Upsilon_{s,\ell}}P\left(\Upsilon_{s,\ell}>\gamma_{i_s}^{th}\right)\right)\\
&=\sum_{n=1}^{N_{\max}^{\mathrm{con}}}\prod_{\ell=0}^{n-1}\left(\frac{\Gamma_{\mathrm{low}}\left(\frac{m_s\gamma_{i_s}^{th}}{\overline{\gamma}_s},(\ell+1)m_s\right)}{\Gamma\left((\ell+1)m_s\right)}+a_{i_s}\right.\\
&\qquad\times\left(1+g_{i_s}\frac{\overline{\gamma}_s}{m_s}\right)^{-(\ell+1)m_s}\\
&\qquad\left.\times\left(1-\frac{\Gamma_{\mathrm{low}}\left(\left(\frac{m_s}{\overline{\gamma}_s}+g_{i_s}\right)\gamma_{i_s}^{th},(\ell+1)m_s\right)}{\Gamma\left((\ell+1)m_s\right)}\right)\right)
\end{aligned}
\tag{26}
$$

where Su indicates the event that data in this transmission are successfully decoded, and Fa denotes the otherwise case. The first three equations are the derivation steps of Eq. (10), and the last two equations are the derivation steps of Eq. (11).

APPENDIX B
DERIVATION OF EQUATION (14)

$$\begin{aligned}
\overline{L}_{\rm pro} &= \Pr(Su_{m,s,1}|Fa_{s,0}) + 2\Pr(Su_{m,s,2}|Fa_{s,0}, Fa_{m,s,1}) \\
&\quad +,\ldots,+ (N_{\max}^{\rm pro}-1)\Pr\left(Su_{m,s,N_{\max}^{\rm pro}-1}|Fa_{s,0},\ldots,\right. \\
&\qquad\qquad \left. Fa_{m,s,N_{\max}^{\rm pro}-2}\right) \\
&\quad + N_{\max}^{\rm pro}\Pr\left(Fa_{s,0}, Fa_{m,s,1},\ldots,Fa_{m,s,N_{\max}^{\rm pro}-1}\right) \\
&\le E\left[\sum_{n=2}^{N_{\max}^{\rm pro}-1} n\left(1-F_x\left(\sum_{j=0}^{n}\gamma_{s,j}+\sum_{j=1}^{n}\gamma_{m,j}\right)\right)\right. \\
&\quad \times F_{i_s}(\gamma_{s,0})\prod_{\ell=1}^{n-1}F_x\left(\sum_{j=0}^{\ell}\gamma_{s,j}+\sum_{j=1}^{\ell}\gamma_{m,j}\right) \\
&\quad + \left(1-F_x\left(\sum_{j=0}^{1}\gamma_{s,j}+\gamma_{m,1}\right)\right)\cdot F_{i_s}(\gamma_{s,0}) \\
&\quad \left. + N_{\max}^{\rm pro}F_{i_s}(\gamma_{s,0})\prod_{\ell=1}^{N_{\max}^{\rm pro}-1}F_x\left(\sum_{j=0}^{\ell}\gamma_{s,j}+\sum_{j=1}^{\ell}\gamma_{m,j}\right)\right] \\
&= E\left[F_{i_s}(\gamma_{s,0}) + \sum_{n=2}^{N_{\max}^{\rm pro}}F_{i_s}(\gamma_{s,0})\right. \\
&\quad \left.\times\prod_{\ell=1}^{n-1}F_x\left(\sum_{j=0}^{\ell}\gamma_{s,j}+\sum_{j=1}^{\ell}\gamma_{m,j}\right)\right]. \qquad (27)
\end{aligned}$$

REFERENCES

[1] X. Zhu, S. Chen, H. Hu, X. Su, and Y. Shi, “TDD-based mobile communication solutions for high-speed railway scenarios,” *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 22–29, Dec. 2013.

[2] B. Ai *et al.*, “Challenges toward wireless communications for high-speed railway,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2143–2158, Oct. 2014.

[3] J. Calle-Sanchez, M. Molina-Garcia, J. I. Alonso, and A. Fernandez-Duran, “Long term evolution in high speed railway environments: Feasibility and challenges,” *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 237–253, Sep. 2013.

[4] L. Yan, X. Fang, and Y. Fang, “Control and data signaling decoupled architecture for railway wireless networks,” *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 103–111, Feb. 2015.

[5] L. Yan and X. Fang, “Reliability evaluation of 5G C/U-plane decoupled architecture for high-speed railway,” *EURASIP J. Wireless Commun. Netw.*, vol. 2014, pp. 1–11, 2014.

[6] C.-X. Wang *et al.*, “Cellular architecture and key technologies for 5G wireless communication networks,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[7] S. Ahmad and D. Datla, “Distributed power allocations in heterogeneous networks with dual connectivity using backhaul state information,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4574–4581, Aug. 2015.

[8] L. Lu, M. Xiao, and L. Rasmussen, “Relay-aided broadcasting with instantaneously decodable binary network codes,” in *Proc. IEEE ICCCN*, Maui, HI, USA, Jul. 2011, pp. 1–5.

[9] M. Xiao and M. Skoglund, “Multiple-user cooperative communications based on linear network coding,” *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3345–3351, Dec. 2010.

[10] L. Lu, M. Xiao, L. Rasmussen, and M. Skoglund, “Efficient scheduling for relay-aided broadcasting with random network codes,” in *Proc. IEEE PIMRC*, Toronto, ON, Canada, Sep. 2011, pp. 1815–1819.

[11] E. Dahlman, S. Parkvall, and J. Sköld,*LTE/LTE-Advanced for Mobile Broadband*. Amsterdam, The Netherlands: Elsevier, 2011.

[12] L. Zhang, Y. Nagai, T. Okamawari, and T. Fujii, “Field experiment of network control architecture for CoMP JT in LTE-advanced over asynchronous X2 interface,” in *Proc. IEEE 77th VTC—Spring*, Dresden, Germany, Jun. 2013, pp. 1–5.

[13] D. Chase, “Code combining—A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets,” *IEEE Trans. Commun.*, vol. 33, no. 5, pp. 385–393, May 1985.

[14] P. Frenger, S. Parkvall, and E. Dahlman, “Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA,” in *Proc. IEEE 54th Veh. Technol. Conf.*, Atlantic City, NJ, USA, 2001, vol. 3, pp. 1829–1833.

[15] “Performance comparison of Hybrid-ARQ schemes,” Motorola, Stockholm, Sweden, TSG R1-17-00-1396, 3GPP Input Paper, 2000.

[16] K. C. Beh, A. Doufexi, and S. Armour, “Performance evaluation of hybrid ARQ schemes of 3GPP LTE OFDMA system,” in *Proc. IEEE 18th Int. Symp. PIMRC*, Athens, Greece, Sep. 2007, pp. 1–5.

[17] Y. Zhao, C. Yin, and J. Li, “Learning- and optimization-based channel estimation for cognitive high-speed rail broadband wireless communications,” in *Proc. IEEE ISCIT*, Gold Coast, Qld., Australia, Oct. 2012, pp. 562–567.

[18] A. Kalakech, M. Berbineau, I. Dayoub, and E. Simon, “Time domain LMMSE channel estimator based on sliding window for OFDM systems in high mobility situations,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5728–5740, Dec. 2015.

[19] Q. Du, G. Wu, Q. Yu, and S. Li, “ICI mitigation by Doppler frequency shift estimation and pre-compensation in LTE-R systems,” in *Proc. IEEE ICCC*, Beijing, China, Aug. 2012, pp. 469–474.

[20] L. Liu *et al.*, “Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 834–845, May 2012.

[21] W. Luo, X. Fang, M. Cheng, and Y. Zhao, “Efficient Multiple-Group Multiple-Antenna (MGMA) scheme for high-speed railway viaducts,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2558–2569, Jul. 2013.

[22] P. Dharmawansa, N. Rajatheva, and C. Tellambura, “On the trivariate Rician distribution,” *IEEE Trans. Commun.*, vol. 56, no. 12, pp. 1993–1997, Dec. 2008.

[23] M. Yacoub, M. Barbin, M. de Castro, and J. B. Vargas, “Level crossing rate of Nakagami-m fading signal: Field trials and validation,” *Electron. Lett.*, vol. 36, no. 4, pp. 355–357, Feb. 2000.

[24] J. Proakis,*Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2000.

[25] H. Jin, C. Cho, N.-O. Song, and D. K. Sung, “Optimal rate selection for persistent scheduling with HARQ in time-correlated Nakagami-m fading channels,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 637–647, Feb. 2011.

[26] J. Francis and N. Mehta, “EESM-based link adaptation in OFDM: Modeling and analysis,” in *Proc. IEEE GLOBECOM*, Atlanta, GA, USA, Dec. 2013, pp. 3703–3708.

[27] T. S. Rappaport,*Wireless Communications: Principles and Practice*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[28] W. Luo, X. Fang, S. Li, and Y. Xia, “Position assisted coordinate HARQ in LTE systems for high speed railway,” in *Proc. IEEE 77th VTC—Spring*, Dresden, Germany, Jun. 2013, pp. 1–5.

[29] S. Kotz and J. W. Adams, “Distribution of sum of identically distributed exponentially correlated gamma-variables,” *Ann. Math. Statist.*, vol. 35, no. 1, pp. 277–283, Mar. 1964.

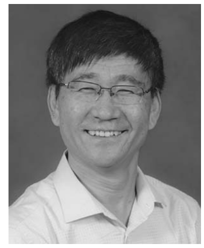
[30] P. G. Moschopoulos, “The distribution of the sum of independent gamma random variables,” *Ann. Inst. Statist. Math., A*, vol. 37, pp. 541–544, Mar. 1985.

[31] M.-S. Alouini, A. Abdi, and M. Kaveh, “Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels,” *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1471–1480, Nov. 2001.

Li Yan (S'14) received the B.E. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2012, where she is currently working toward the Ph.D. degree with the Key Laboratory of Information Coding and Transmission, School of Information Science and Technology. Her research interests include handover, network architecture, and reliable wireless communication for high-speed railways.

Xuming Fang (M'00) received the B.E. degree in electrical engineering, the M.E. degree in computer engineering, and the Ph.D. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 1984, 1989, and 1999, respectively. In September 1984, he was a Faculty Member with the Department of Electrical Engineering, Tongji University, Shanghai, China. He then joined the School of Information Science and Technology, Southwest Jiaotong University, where he has been a Professor since 2001 and the Chair of the Department of Communication Engineering since 2006. He held visiting positions with the Institute of Railway Technology, Technical University at Berlin, Berlin, Germany, in 1998 and 1999 and with the Center for Advanced Telecommunication Systems and Services, University of Texas at Dallas, Richardson, TX, USA, in 2000 and 2001. He has, to his credit, around 200 high-quality research papers in journals and conference publications. He has authored or coauthored five books or textbooks. His research interests include wireless broadband access control, radio resource management, multihop relay networks, and broadband wireless access for high-speed railway. Dr. Fang is the Chair of the IEEE Vehicular Technology Society of the Chengdu Chapter and an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Geyong Min (M'01) received the B.Sc. degree in computer science from Huazhong University of Science and Technology, Wuhan, China, in 1995 and the Ph.D. degree in computing science from the University of Glasgow, Glasgow, U.K., in 2003. He is a Professor of high-performance computing and networking with the Department of Mathematics and Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, U.K. His research interests include future Internet, computer networks, wireless communications, multimedia systems, high-performance computing, modeling, and performance engineering.



Yuguang "Michael" Fang (F'08) received the B.S./M.S. degree from Qufu Normal University, Shandong, China, in 1987; the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 1994; and the Ph.D. degree from Boston University, Boston, MA, USA, in 1997. In 2000, he joined the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, where he has been a Full Professor since 2005. He held a University of Florida Research Foundation Professorship from 2006 to 2009; a Changjiang Scholar Chair Professorship with Xidian University, Xi'an, China, from 2008 to 2011; and a Guest Chair Professorship with Tsinghua University, Beijing, China, from 2009 to 2012.

Dr. Fang received the U.S. National Science Foundation Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002 and was a recipient of the Best Paper Award from IEEE ICNP (2006). He has also received the 2010–2011 UF Doctoral Dissertation Advisor/Mentoring Award, the 2011 Florida Blue Key/UF Homecoming Distinguished Faculty Award, and the 2009 UF College of Engineering Faculty Mentoring Award. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, was the Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2009–2012), and serves/served on the Editorial Board of several journals, including the IEEE TRANSACTIONS ON MOBILE COMPUTING (2003–2008, 2011–present), the IEEE TRANSACTIONS ON COMMUNICATIONS (2000–2011), and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2002–2009). He has been actively participating in conference organizations such as serving as the Technical Program Cochair for IEEE INOFOCOM 2014 and the Technical Program Vice Chair for IEEE INFOCOM 2005.