Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2233503, 12 pages
https://doi.org/10.1155/2022/2233503

Research Article

Network-Assisted Full-Duplex Enabled Ultrareliable and Low-Latency Communications

Yue Zhu,[1] **Jiamin Li**,[1,2] **Pengcheng Zhu**,[1] **Dongming Wang**,[1,2] **and Xiaohu You**[1,2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*
[2]*Purple Mountain Laboratories, Nanjing 211111, China*

Correspondence should be addressed to Jiamin Li; jiaminli@seu.edu.cn

Received 24 June 2022; Accepted 3 October 2022; Published 14 October 2022

Academic Editor: Yejun He

The ultrareliable and low-latency communication (URLLC) is one of the key scenarios of the current 5G new radio (NR). The performance of the duplex technologies in URLLC to assist in meeting the needs of low-latency services is of great significance. However, time division duplex (TDD) has poor delay performance due to the extra data waiting delay caused by the frequent uplink/downlink switching. Besides, TDD systems cannot dynamically adapt to the real-time changes of user requirements due to the problem of frame structure configuration, while frequency division duplex (FDD) has more pilot overhead due to channel estimation and worse spectrum utilization which are great drawbacks in 5G NR enabled URLLC. In this paper, we further proposed the strengths of a novel duplex technology named network-assisted full-duplex (NAFD) in terms of URLLC latency performance compared with TDD and FDD. NAFD scheme exploits the resources of the time and spatial domains to form a special duplex mechanism where the central processing unit (CPU) can dynamically assign the modes of different access points (APs) as downlink transmission or uplink receiving in the same coherent time block. Some improvement of the existing NAFD scheme for better fitting for URLLC is also proposed in this paper. The simulation results and analysis showed that the NAFD scheme outperforms other duplex technologies under URLLC scenarios.

1. Introduction

Ultra-reliable and low-latency communication (URLLC) is one of the three major use cases defined by 3GPP. Duplex technologies are widely used because of the double throughput contributing to the systems [1]. The duplex technologies have been widely employed into URLLC scenarios, such as robotics or industrial automation application and traffic condition monitoring [2]. Further, dynamic time division duplexing (TDD) is the major duplexing technology for 5G new radio (NR) because of the wide spectrum availability of unpaired bands [3]. Additionally, the frequency division duplexing (FD-D) also supports 5G NR and is considered especially relevant for deployments at bands below 6 GHz [4]. However, the two existing duplex technologies also have their own disadvantages in supporting ultrareliable and low-latency services.

The delay of the TDD systems mainly comes from the extra data waiting delay caused by the duplex feature and the uplink/downlink conversion of the frame structure, which cannot be compensated by the minislot technology in URLLC [5]. Dynamic TDD systems are also difficult to meet the stringent URLLC latency and reliability targets which is attributed to the nonconcurrent availability of the downlink (DL) and uplink (UL) transmission opportunities, and the additional cross-link interference (CLI) between base stations (BSs) and user equipment (UE) [6]. While FDD systems are more manageable, both BSs and UE always have simultaneous UL and DL transmission opportunities. From this perspective, FDD systems can better fulfill the URLLC requirements because of the absence of the UL-DL switching delay. However, FDD systems exhibit more pilot overhead, higher complexity, and lower reliability for channel estimation due to the channel nonreciprocity [5]. On the other hand, FDD systems cannot support asymmetric spectrum services as TDD systems, which is a great drawback for real-time service scenarios with asymmetric user requirements.

Since the TDD and FDD systems both have drawbacks in meeting the stringent requirements of URLLC for high-reliable and low-latency services in some specific aspects, new duplex scheme should be investigated to better fit the URLLC requirements. A new duplex technology named network-assisted full-duplex (NAFD) was proposed in [7]. In NAFD scheme, each access point (AP) is connected to the central processing unit (CPU) through fronthaul link. The baseband processing is jointly carried out at the CPU. Each AP has a half-duplex transceiver device performing either transmitting or receiving which are decided by CPU according to the traffic load of the whole network. Any UE with a specific data demand can find some nearby APs operating in the corresponding mode in the same slot. In-band full-duplex could be achieved under cell-free massive multiple-input multiple-output (MIMO) with existing half-duplex hardware devices. Further, the joint processing of the signals at the CPU can mitigate the CLI [8]. NAFD scheme flexibly allocates resources in the two dimensions of space and time to achieve duplex at the network level without sacrificing bandwidth. Compared with the traditional TDD systems, NAFD scheme can provide low-latency services. Compared with traditional FDD systems, NAFD scheme can support asymmetric services without reducing spectrum utilization. The superiority of NAFD scheme and the analysis of the duplex operation mechanism behind why NAFD has better performance in URLLC scenarios are highlighted in Section 3.

Kant Chaudhary et al. [9] investigated the fronthaul latency in the uplink of a C-RAN system with massive MIMO-based RRUs and 3GPP functional Split 7. The latency gain of an interactive URLLC device when using flexible TDD and a decoupled UL/DL access was investigated in [10]. Reddy et al. [11] worked out on latency analysis of IMT-2020 radio interface technology for both FDD and TDD modes while [5] presented a system level analysis of the URLLC outage performance within the 5G new radio flexible TDD systems. Specifically, [5] also studied the feasibility of the URLLC outage targets compared to the case with the 5G FDD and with numerous 5G design variants

Our three main contributions in this article are outlined as follows:

(i) We model the latency of control plane (CP) and user plane (UP) in URLLC scenario for three different duplex technologies. Different from the previous studies, we do not simply list the delay components of TDD and FDD systems. Instead, we mainly focus on how the duplex mechanism affects communications under URLLC scenarios based on the latency analysis for both control plane and user plane

(ii) We analyze the superiority of the novel NAFD scheme compared with TDD/FDD systems in terms of reliability and latency tradeoff. The analysis suggests that the proposed NAFD scheme can make up for some of the deficiencies of TDD and FDD systems under URLLC scenarios

(iii) We further propose the open-loop communications and load-aware mode selection for NAFD scheme to further reduce the latency. NAFD scheme employs some specific duplex mechanisms to expand the envelope of URLLC latency-reliability curves. The proposed NAFD scheme is based on the cell-free distributed MIMO-based architecture, which introduces cooperation between APs. It broadens the dimension of system performance optimization

2. URLLC Radio Latency Analysis for Different Duplex Mode

Latency on the ultra-reliable and low-latency communications is defined as the average time between the transmission of packet and the reception of an acknowledgment. The latency performance of a communication system is analyzed for both control plane and user plane. The IMT-2020 proposal [12] defines the minimum latency support for CP and UP for URLLC is within 20 ms and 1 ms, respectively.

2.1. Control Plane Latency. According to 3GPP TR 38.913 [13], control plane latency is defined as “the time to move from a battery efficient state e.g. IDLE to the start of continuous data transfer e.g. ACTIVE.” With minislots, the transmit time intervals (TTIs) can have different lengths such as 2, 4, or 7 symbols. For simplicity, the processing delay is set to 1TTI in both BSs and UE. Data transmission happens in two ways from the UE, once when the UE has a dedicated radio resource control (RRC) which is the signal of physical uplink shared channel (PUSCH) resource available and once when the UE needs to access the network and then begin data transmission. When no dedicated connection is established, a scheduling request will be transmitted on random access control channel (RACH) called as random access scheduling request. The process of accessing the network when no dedicated RRC is established or when the UE transmits for the first time is called random access (RA), and the channel that plays the major role in this aspect is called RACH channel. The RRC processing delays are assumed to be of a fixed value of 3 ms [11]. A preamble is sent by UE to base station over physical random access channel (PRACH) to obtain the UL synchronization. Here, the transmission delay of RACH preamble and preamble detection and processing delay in BS are set to be 1TTI. Then, the base station sends RA response to all the UE which has sent RA preamble intimating them that resources have been reserved for them. Later, the UE will send RRC connection request. In the RRC messages, the UE sends its identifier to the base station to make it aware of its identity. This identifier is used to solve any contention resolution that can happen. The latency metrics of control plane for TDD/FDD/NAFD systems in TTIs are shown in Table 1, where an alternating UL-DL sequence is designed for the TDD slot sequence. The delay for NAFD scheme presented in Table 1 is within a coherent

Table 1: CP latency in TTIs in TDD/FDD/NAFD systems.

Latency composition	TDD latency	FDD latency	NAFD latency
Worst-case delay due to random access channel (RACH) scheduling period	2TTI	1TTI	2TTI
Transmission of RACH preamble	1TTI	1TTI	1TTI
Preamble detection and processing in BS	1TTI	1TTI	1TTI
DL slot alignment	1TTI	0TTI	0TTI
Transmission of random access (RA) response	1TTI	1TTI	1TTI
UE processing delay	1TTI	1TTI	1TTI
UL slot alignment	1TTI	0TTI	0TTI
Transmission of RRC connection resume request	1TTI	1TTI	1TTI
Processing delay in BS	3 ms	3 ms	3 ms
DL slot alignment	1TTI	0TTI	0TTI
Transmission of RRC connection resume	1TTI	1TTI	1TTI
Processing delay in the UE	3 ms	3 ms	3 ms
UL slot alignment	1TTI	0TTI	0TTI
Transmission of RRC connection resume complete	1TTI	1TTI	1TTI
Processing delay in BS	1TTI	1TTI	1TTI
Total delay	14TTI + 6 ms	9TTI + 6 ms	10TTI + 6 ms

time block, thus excluding additional alignment delays caused by the sequence of DL and UL slots. FDD systems are also free of UL-DL link-switching delay.

2.2. User Plane Latency. The user plane latency is analyzed as the radio interface latency from the time when transmitter packet data convergence protocol (PDCP) receives an Internet protocol (IP) packet to the time when receiver PDCP successfully receives the IP packet and delivers the packet to the upper layer [11]. The average one-way URLLC latency for TDD systems in the DL direction T_{dl} is given in Table 2 and can be split into Equation (1) [5].

$$T_{\mathrm{dl}} = t_{\mathrm{bsp}} + t_{\mathrm{tdd}} + t_{\mathrm{qd}} + t_{\mathrm{fa}} + t_{\mathrm{tti}} + \alpha t_{\mathrm{harq}} + t_{\mathrm{uep}}, \quad (1)$$

where t_{bsp}, t_{tdd}, t_{qd}, t_{fa}, t_{tti}, t_{harq}, and t_{uep} denote the BS processing, UL-DL link-switching delay, DL total queuing, DL frame alignment, DL packet transmission, DL hybrid automatic repeat request (HARQ) retransmission, and UE processing delays, respectively. α implies the target block error rate (BLER). When UL/DL data comes, it needs to wait for the next UL/DL time slot to send; there will be a long queuing delay. That corresponds to the UL-DL link-switching delay. The DL HARQ delay t_{harq} is expressed as

$$t_{\mathrm{harq}} = t_{\mathrm{uep}} + t_{\mathrm{fa}} + t_{\mathrm{nack}} + t_{\mathrm{bsp}} + t_{\mathrm{tdd}} + t_{\mathrm{qd}} + t_{\mathrm{fa}} + t_{\mathrm{tti}}, \quad (2)$$

where t_{nack} denotes the hybrid automatic repeat request (HARQ) negative acknowledgment (NACK) transmission time.

For FDD systems, there is no UL-DL link-switching delay. However, since FDD systems do not have channel reciprocity, in the downlink transmission, we need to send pilots to estimate the channel quality, which will take more time than TDD systems. Besides, it will take more time than TDD systems in the processing time of the BS for the more complicated channel estimations. For NAFD scheme, we suppose there is no UL-DL link-switching delay in a coherent time block.

Similarly, the one-way URLLC UL latency T_{ul} in TDD systems follows a similar behavior as T_{dl}. Here, the grant-free (GF) UL scheduling is considered.

$$T_{\mathrm{ul}} = t_{\mathrm{uep}} + t_{\mathrm{tdd}} + t_{\mathrm{qu}} + t_{\mathrm{fa}} + t_{\mathrm{tti}} + \alpha t_{\mathrm{harq}} + t_{\mathrm{bsp}}, \quad (3)$$

where t_{uep}, t_{tdd}, t_{qu}, t_{fa}, t_{tti}, t_{harq}, and t_{bsp} denote the UE processing delay, TDD UL-DL link-switching delay, UL total queuing, UL frame alignment delay, UL payload transmission delay, UL HARQ delay, and the BS processing delays, respectively. The UL HARQ delay t_{harq} is expressed as

$$t_{\mathrm{harq}} = t_{\mathrm{bsp}} + t_{\mathrm{fa}} + t_{\mathrm{nack}} + t_{\mathrm{uep}} + t_{\mathrm{tdd}} + t_{\mathrm{qu}} + t_{\mathrm{fa}} + t_{\mathrm{tti}}. \quad (4)$$

For FDD systems and NAFD scheme within a coherent time block, there is no additional alignment delays caused by the sequence of DL and UL slots. User plane latency in TTIs for uplink transmission is assumed in Table 3.

3. Latency Assessment for Different Duplex Modes

3.1. URLLC Latency versus the Length of Coherence Block for NAFD Scheme. As shown in Figure 1, in NAFD scheme, each AP has a transceiver device performing either transmitting or receiving which are decided by CPU according to the traffic load of the whole network. Through the working mode selection of each AP as uplink receiving or downlink transmission, the configuration ratio of the uplink AP and the downlink AP in the entire system remains unchanged within a coherence block. Therefore, in several coherence

TABLE 2: User plane latency in TTIs for downlink in TDD/FDD/NAFD systems.

Latency composition	TDD latency	FDD latency	NAFD latency
BS processing	1.5TTI	2TTI	1.5TTI
UL-DL link-switching delay	1TTI	0TTI	0TTI
DL total queuing	t_{qd}	t_{qd}	t_{qd}
DL frame alignment	1TTI	1TTI	1TTI
DL packet transmission	1TTI	1TTI	1TTI
DL hybrid automatic repeat request (HARQ) retransmission	αt_{harq}TDD	αt_{harq}FDD	αt_{harq}NAFD
UE processing delay	1.5TTI	1.5TTI	1.5TTI
Pilot sending for channel estimation delay	0TTI	3TTI	0TTI
Total delay	$6\text{TTI} + t_{qd} + \alpha(8\text{TTI} + t_{qd})$	$8.5\text{TTI} + t_{qd} + \alpha(7.5\text{TTI} + t_{qd})$	$5\text{TTI} + t_{qd} + \alpha(7\text{TTI} + t_{qd})$

TABLE 3: User plane latency in TTIs for uplink in TDD/FDD/NAFD systems.

Latency composition	TDD latency	FDD latency	NAFD latency
UE processing delay	1.5TTI	1.5TTI	1.5TTI
UL-DL link-switching delay	1TTI	0TTI	0TTI
UL total queuing	t_{qu}	t_{qu}	t_{qu}
UL frame alignment	1TTI	1TTI	1TTI
UL packet transmission	1TTI	1TTI	1TTI
UL HARQ retransmission	αt_{harq}TDD	αt_{harq}FDD	αt_{harq}NAFD
BS processing delay	1.5TTI	1.5TTI	1.5TTI
Total delay	$6\text{TTI} + t_{qu} + \alpha(8\text{TTI} + t_{qu})$	$5\text{TTI} + t_{qu} + \alpha(7\text{TTI} + t_{qu})$	$5\text{TTI} + t_{qu} + \alpha(7\text{TTI} + t_{qu})$

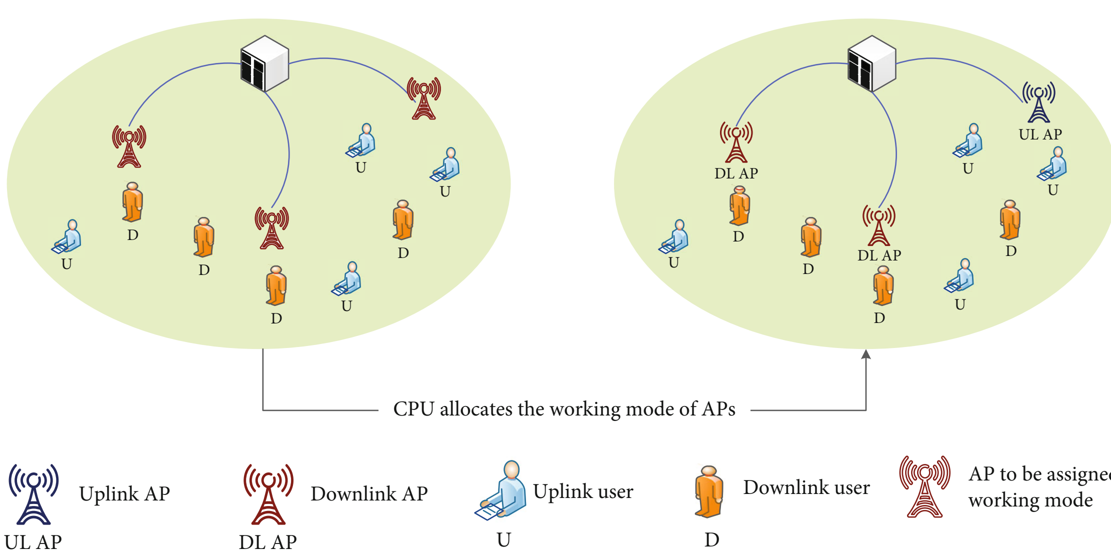

FIGURE 1: NAFD system model.

blocks, there is still the switching delay of uplink and downlink, and how much the delay is reduced compared to dynamic TDD depends on the length of the coherence block. The longer the coherence block, the less total TDD buffering delays. The number of transmission symbols per coherence block ranges from hundreds (with high mobility and high channel dispersion) to hundreds of thousands of samples (with low mobility and low channel dispersion) [14]. τ is defined as the transmission symbols per coherence block contains. We can safely assume that $\tau \geq 200$.

Taking CP latency for example, subcarrier spacing (SCS) is set to be 30 kHz, and TTI is set to be 2 symbols. We further analyzed the delay in NAFD scheme with the length of coherence block. As shown in Figure 2, for CP delay, FDD systems achieve the best performance. NAFD scheme performs better than TDD systems. The longer the length of the coherence block, the lower delay of NAFD scheme can be achieved. Since NAFD scheme still has the UL-DL switching delay from the perspective of several coherence blocks, it can not exceed FDD systems in terms of the CP latency performance. More factors should be considered as regards to the UP latency, such as pilot sending delay for channel estimation in FDD systems. Therefore, it is impossible to clearly distinguish the advantages of FDD and TDD systems in terms of delay performance in some specific scenarios. By contrast, for NAFD scheme, the UP latency still decreases as coherence block grows.

3.2. UP Latency versus the Traffic Loads. An equivalent FDD bandwidth allocation is adopted to maintain fairness when compared with TDD/NAFD. Because FDD systems have half the bandwidth of TDD/NAFD systems, when the amount of data is relatively larger, there will be more queuing and buffering delays. As shown in Figure 3, with the slight traffic loads, FDD systems outperform TDD/NAFD systems for completely free of UL-DL switching delay. With the traffic loads increasing, UP latency of FDD systems grows larger due to the rapid growth of queuing and buffering data. NAFD scheme always has better performance in terms of UP latency than TDD systems because of the absence of UL-DL switching delay within a coherence block in NAFD scheme. Hence, we can conclude that when the traffic loads are heavy, NAFD scheme outperforms other two duplex modes in terms of UP latency.

The FDD systems sacrifice frequency band resources in exchange for simultaneous uplink and downlink transmission opportunities, which directly results in half the bandwidth and half the rate. NAFD scheme sacrifices spatial freedom in exchange for cotime cofrequency duplex at the network level. In NAFD scheme, we could virtually regard the cotime cofrequency full-duplex (CCFD) APs as two APs: one for uplink reception and one for downlink transmission. For equality, the number of antennas in NAFD scheme is equivalent to half that of TDD systems, but the achievable rate is higher than half of TDD systems [7]. This is a very important reason why NAFD is better than FDD. In [15], it is proved that NAFD scheme with half-duplex APs outperforms the TDD-based cell-free massive MIMO systems with the same antenna density in terms of the achievable rate. Hence, in Figure 3, with the traffic loads increasing, the advantage for higher achievable rate of NAFD scheme is highlighted. The gap between the three duplex modes is growing.

3.3. CP/UP Latency versus SCS Size. Figure 4 shows CP latency versus SCS size. Unlike the 4G standards, 5G NR adopts different SCSs for its diverse service classes. Here, TTI size is set to 2 symbols. FDD performs best in terms of CP latency, followed by NAFD and TDD, respectively. With the increase of SCS, the delay difference between the three duplex technologies becomes smaller. The larger SCS size offers reduced processing time due to the shorter OFDM symbols in time.

Figure 5 shows UP latency with SCS size. Here, TTI size is also set to 2 symbols, and we suppose that the traffic loads are light without additional queuing and buffering delay for FDD systems. The load pattern is set to be UL-DL pattern for UP latency. The NAFD scheme exhibits best, while the performance of FDD is worst since FDD systems spend a lot in channel estimation. Also, with the increase of SCS, the delay difference between the three becomes smaller.

3.4. CP/UP Latency versus TTI Size. Figures 6 and 7 show CP and UP latency versus TTI size, respectively. The load pattern is set to be UL-DL pattern for UP latency. The TTI length determines the packet transmission periodicity. The larger the TTI size, the larger the time delay of which the incoming packets shall exhibit in the scheduling buffers. Also, with the increase of TTI size, the delay difference between the three becomes larger.

3.5. Latency versus Inter-BS CLI. The inter-BS CLI is considered as the most critical challenge against the 5G NR TDD systems [5]. Here, we do not consider the extra delay caused by the low rate ratio and large channel estimation overhead due to the FDD spectrum halving mentioned above but mainly focus on comparing the impact of CLI interference cancellation on the URLLC duplex systems. As shown in Figure 8, the FDD systems provide the best latency performance, mainly due to the absolute absence of the inter-BS CLI. TDD systems have the worst performance mainly because of the packets getting retransmitted several times prior to a successful decoding, due to the severe BS-BS CLI, leading to significantly large retransmission delay. In NAFD scheme, the cross-link interference can be eliminated through joint processing in CPU. Besides, by decoupling uplink and downlink, NAFD is able to allocate system resources more flexibly, and CLI will be reduced after balancing the loads. However, it can not reach the totally absence of inter-BS CLI state as FDD systems do.

3.6. Superiority of NAFD Scheme Implemented in URLLC Scenarios. The superiority of NAFD scheme implemented in URLLC scenarios in the future 6G systems is highlighted as follows:

(i) Due to the influence of the fixed frame structure configuration of TDD systems, it is difficult to improve the delay performance of the TDD systems. For example, when DL data comes, it needs to wait for the next DL time slot to send; there will be a long queuing delay. However, in NAFD scheme, the data of UE is not only transmitted in the time slot of one specific AP; UE can choose to transmit or receive data from the assigned DL/UL APs through mode selection. Hence, it does not need to wait for the UL/DL time slots to be transmitted. The mode pattern of uplink and downlink APs will not change within a coherent period in NAFD scheme. Like FDD systems, there is no UL/DL switching delay

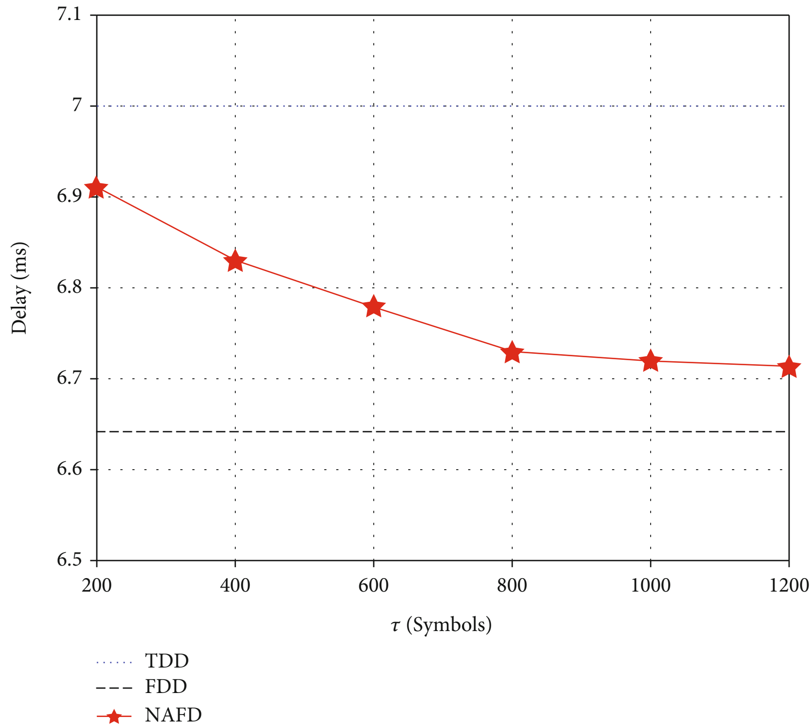

Figure 2: CP latency with the length of coherence block.

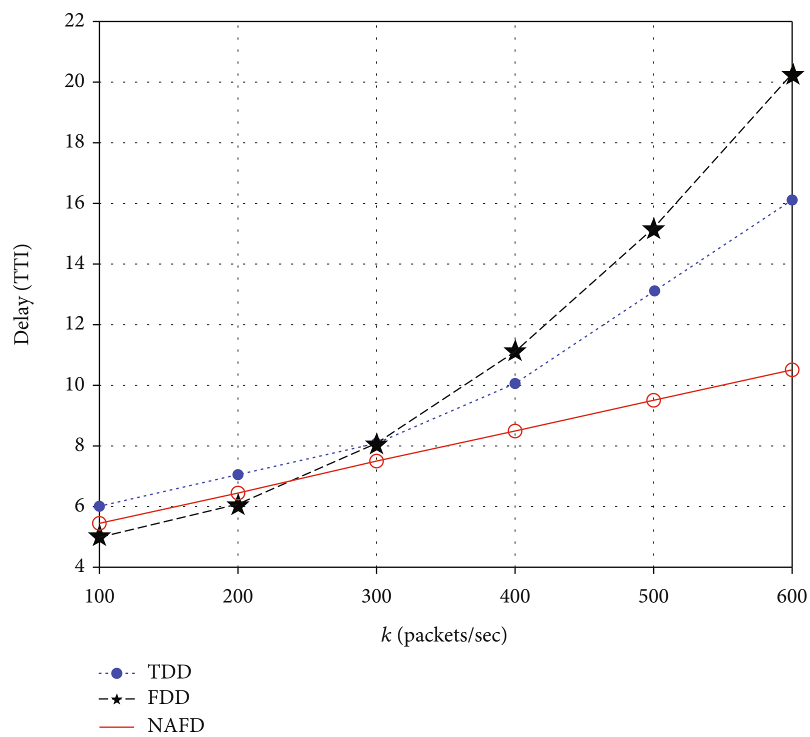

Figure 3: UP latency with the traffic load.

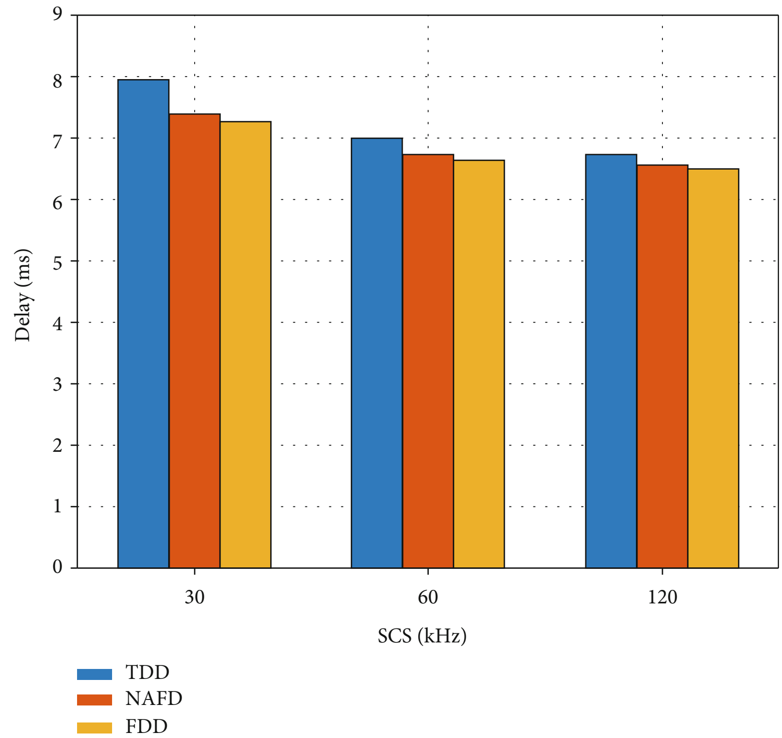

FIGURE 4: CP latency with SCS size.

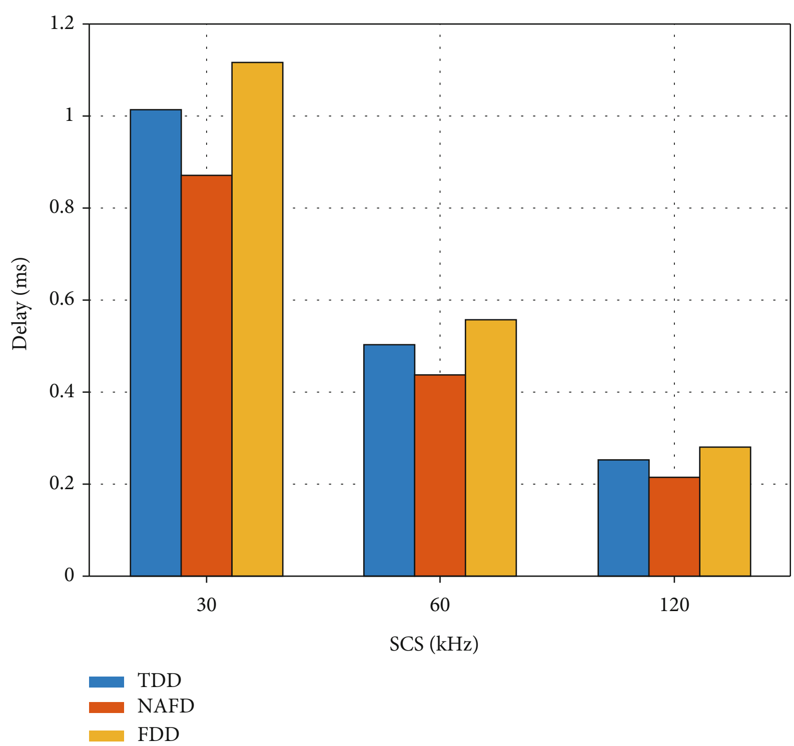

FIGURE 5: UP latency with SCS size.

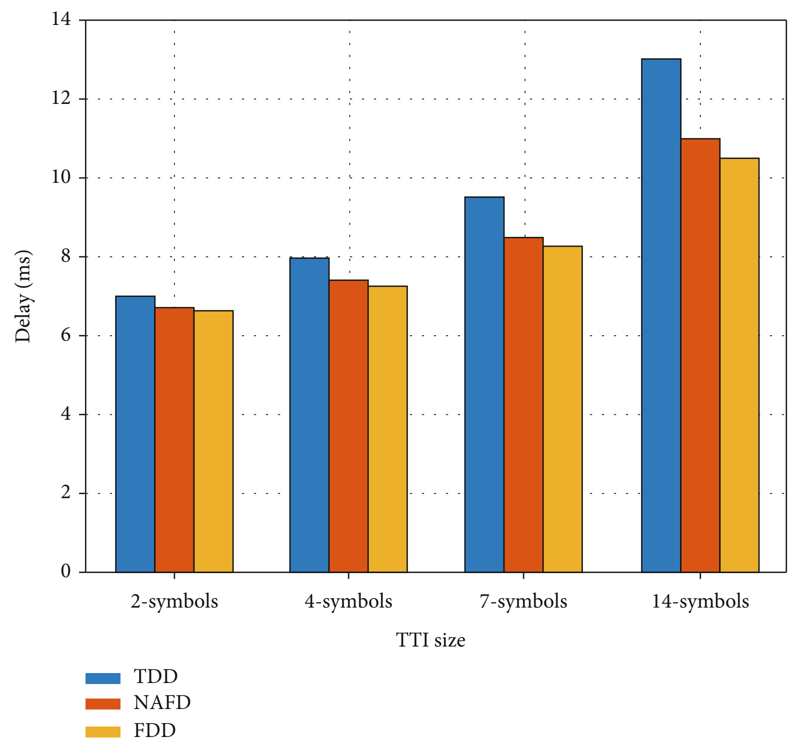

Figure 6: CP latency with TTI size.

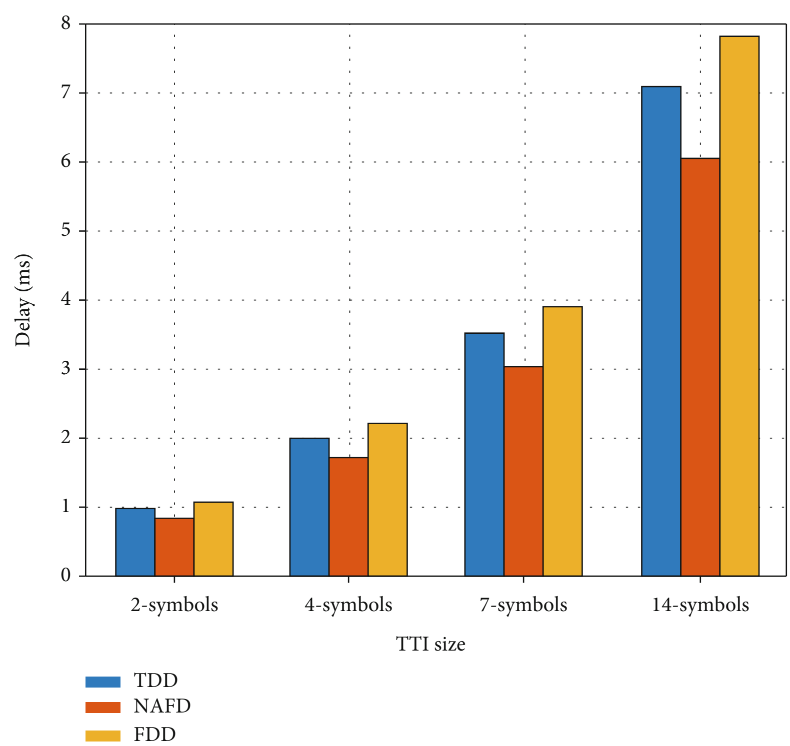

Figure 7: UP latency with TTI size.

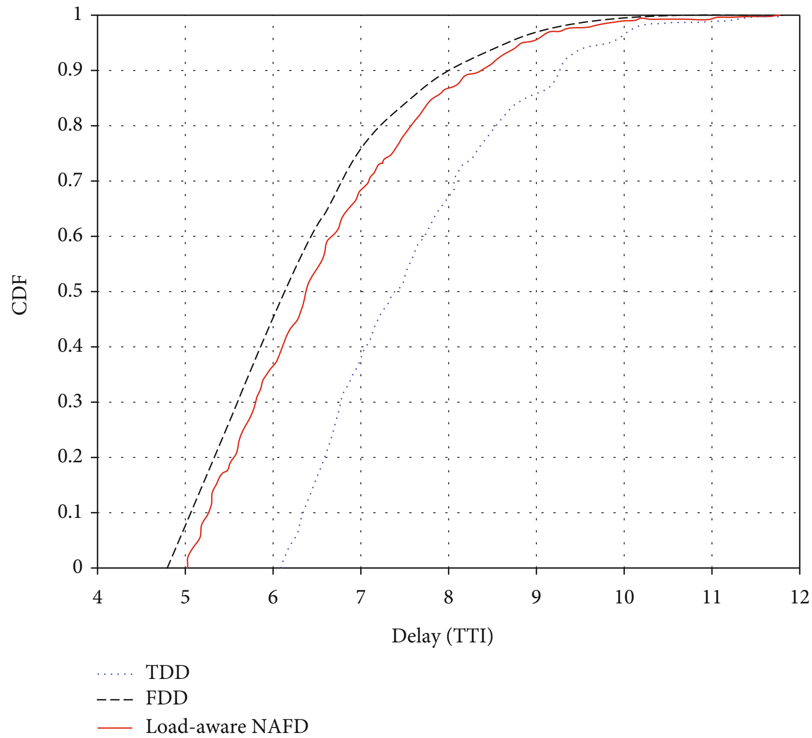

Figure 8: Latency versus inter-BS CLI.

in a coherent block. NAFD scheme exploits spatial freedom to reduce the delay of waiting for uplink and downlink handover for TDD

(ii) In dynamic TDD, an adapted frame structure is generally selected for a specific scene such as UL-heavy or DL-heavy. After the frame structure is configured, it cannot be flexibly changed immediately after the UE changes in the scene, such as load changes or position changes. However, NAFD can dynamically select the uplink or downlink working modes of APs within a coherent block according to the load change and position change of the UE in the scenario, realizing flexible duplexing in the true sense

(iii) From the perspective of a single AP, NAFD scheme also fills the uplink and downlink working modes in the time domain. Hence, similar to TDD, NAFD also supports asymmetric services, has higher spectrum resource utilization, and doubles the bandwidth of FDD. FDD has more pilot overhead due to channel estimation and worse spectrum utilization thus not widely used in actual scenes. NAFD performs channel estimation based on the reciprocity of the channel, which reduces the pilot overhead for channel estimation and improves the transmission reliability

(iv) FDD systems sacrifice frequency band resources in exchange for simultaneous uplink and downlink transmission opportunities, which directly results in half the bandwidth and half the rate. NAFD scheme sacrifices spatial freedom in exchange for cotime cofrequency duplex at the network level. In NAFD scheme, we could virtually regard the cotime cofrequency full-duplex (CCFD) APs as two APs, one for uplink reception and one for downlink transmission. For equality, the number of antennas in NAFD scheme is equivalent to half that of TDD systems, but the achievable rate is higher than half of TDD systems [8]. This is a very important reason why NAFD is better than FDD

(v) By decoupling the uplink and downlink in the spatial domain, NAFD is able to allocate system resources more flexibly. By balancing the load, NAFD can improve the successful access probability, thus avoiding the extra delay caused by repeated retransmissions. Another advantage of decoupling uplink and downlink is that after balancing the load, the interference between UE and APs will be reduced. Compared with FDD/TDD, NAFD is more critical in terms of cross-link interference elimination through joint processing, which can reduce local processing delay and increase access reliability

4. Improvement of NAFD Scheme for URLLC

The simulation results and analysis have showed that the proposed NAFD scheme outperforms other duplex technologies

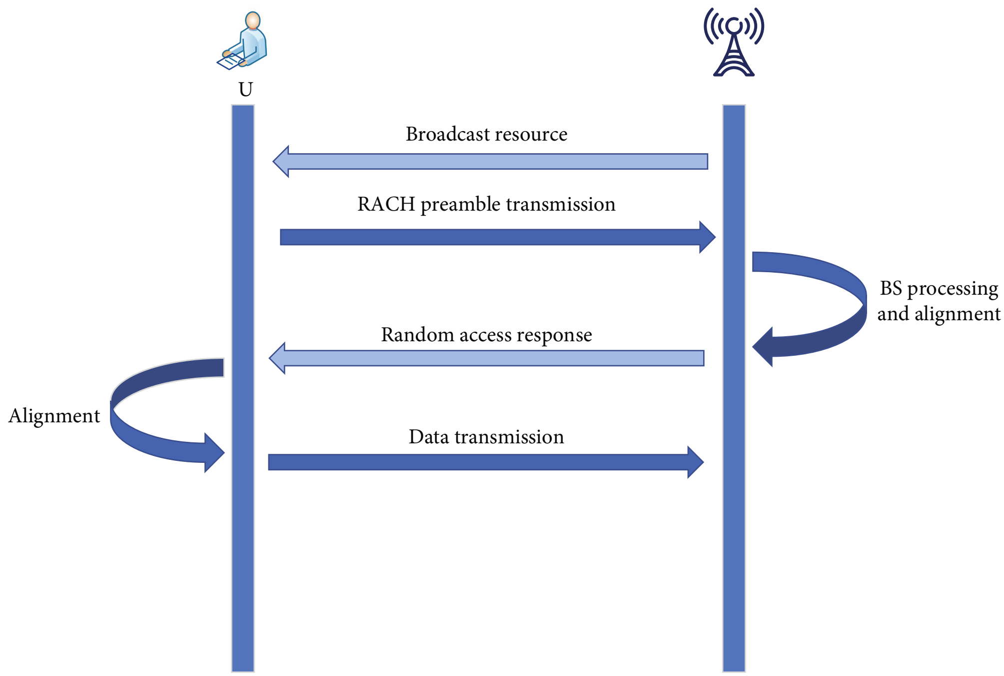

Figure 9: CP delay in OLC mode for NAFD scheme.

under URLLC scenarios. NAFD retains the channel reciprocity of TDD, while utilizing the freedom of time and space. Without sacrificing the spectrum, it avoids the complex channel estimation overhead of FDD and the problem of low spectrum utilization. The extra delay overhead caused by frequent uplink and downlink switching of TDD is also avoided within the coherence block. Nowadays, the dynamic TDD scheme is proposed where each cell base station can adaptively adjust the ratio of uplink and downlink time slots of each transmission frame according to the actual flow of the cell. From the perspective of a certain moment, NAFD scheme and existing dynamic TDD technology both regulate the uplink and downlink transmission modes of the APs or cell base stations, so that the system can flexibly adapt to the business needs of asymmetric upstream and downstream. However, from a certain coherent time block, the uplink and downlink working modes of the AP in the flexible duplex system are flexibly scheduled and selected by the CPU and remain unchanged within a coherent time block. NAFD scheme greatly reduces the switching delay of uplink and downlink compared with dynamic TDD systems. In addition, the proposed NAFD scheme is based on the cell-free distributed MIMO-based architecture, which introduces cooperation between APs. The working modes of all APs are uniformly scheduled by the CPU which can achieve better performance in terms of CLI elimination and resource allocation compared with dynamic TDD systems that only adjust the transmission mode of the base station according to the traffic volume in its own cell. The proposed NAFD scheme broadens the dimension of system performance optimization.

4.1. Open-Loop Communication Mode for NAFD Scheme. Short packet traffic in mission-critical services is usually bursty and sporadic, such as collision prewarning alert in autonomous driving. When such short packet traffic arrives, the random and proactive radio resource allocation strategy can realize the shortest possible scheduling time and hence improve transmission latency. For CP delay, the traditional close-loop mode may not satisfy low-latency requirements of mission-critical services due to feedback and retransmissions. As shown in Figure 9, to reduce the latency and alleviate signaling overhead, the open-loop communication (OLC) mode [16] can be adopted in NAFD scheme, which is feedback free. The CP delay components are listed as follows:

$$t_{cp} = t_{broadcast} + t_{RACH} + t_{BS} + t_{alignment} + t_{RAres} + t_{tti}, \quad (5)$$

where $t_{broadcast}$ denotes the delay for BSs to broadcast available resources. t_{RACH}, t_{BS}, $t_{alignment}$, t_{RAres}, and t_{tti} represent the RACH preamble transmission delay, BS processing delay, slot alignment delay, RA response transmission delay, and data transmission delay, respectively. It is obvious that the CP delay is reduced in the OLC mode. OLC mode is an effective strategy to reduce transmission delay, but due to low-level resource regulation and lack of feedback, OLC mode will lead to reduced reliability.

Considering in NAFD scheme, all APs jointly serve for UE. UE associates with multiple nearby base stations at the same time, thereby receiving the same information over multiple links. In the NAFD scheme, the multiconnection feature in which a user associates with multiple base stations can naturally improve the reliability of transmission due to its own macrodiversity gain and solve the problems brought by the OLC mode. Latency for the control plane will be greatly reduced while still guarantee the access reliability.

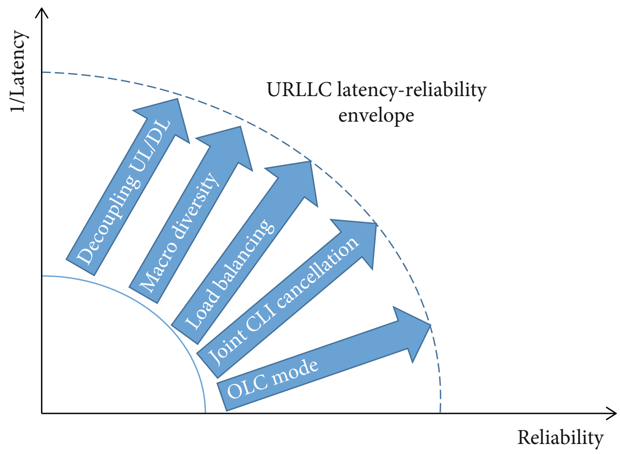

Figure 10: URLLC latency-reliability curve for NAFD scheme.

4.2. Load-Aware Dynamic Mode Selection for NAFD Scheme. We have studied load-aware dynamic mode selection for NAFD scheme [17]. Intelligent reinforcement learning algorithms are proposed which proved that NAFD scheme enabled cell-free massive MIMO system with appropriately scheduled APs is a promising solution to meet the heterogeneous traffic loads in next generation wireless systems. By balancing the load, NAFD can improve the successful access probability and reduce the inter-UE and inter-BS interference, thus avoiding the extra delay caused by repeated retransmissions.

Due to the density difference and nonuniformity of the dynamic distribution of base stations and users in actual duplex scenarios, the traditional mode of binding the uplink and downlink channels to the same base station cannot ensure optimal performance. In the duplex scheme of NAFD, the uplink and downlink channels are flexibly decoupled according to the relative distance between the users and the base stations, transmission power, and traffic loads; that is, the uplink and downlink channels are associated with different APs. Using spatial freedom to realize elastic access in the scenario of massive terminal-enabled URLLC is more conducive to dynamic resource allocation and has a great impact on the network interference structure. Through intelligent cooperation between APs and dynamic mode selection of APs as downlink transmission or uplink receiving, we can establish the trade-off relationship between air interface latency and reliability with the elastic access mode. NAFD scheme expands URLLC latency-reliability compromise envelope with some unique duplex mechanisms as depicted in Figure 10.

5. Conclusion

URLLC requires new duplex technologies to improve latency performance under some low-latency scenarios. NAFD scheme is a free duplex method based on a cellular-free architecture. We investigated TDD/FDD/NAFD three duplex technologies in terms of latency performance under URLLC scenarios including control-plane latency and user-plane latency. Compared with the traditional TDD, NAFD can provide low-latency services. Compared with the traditional FDD, NAFD can support asymmetric services without reducing spectrum utilization. NAFD scheme employs some specific duplex mechanisms to expand the envelope of URLLC latency-reliability curves, such as decoupling the uplink and downlink, macrodiversity, load balancing, and joint CLI cancellation. Open-loop communication and load-aware mode selection for NAFD scheme can further reduce the latency. Hence, NAFD scheme can support URLLC communication for a large number of terminals and has great application prospects under URLLC scenarios.

Data Availability

We do not use external data; the data are all obtained from our own simulation experiment platform.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2021YFB2900300 and by the National Natural Science Foundation of China (NSFC) under Grants 61971127, 61871465, and 61871122.

References

[1] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, “Full-duplex MIMO relaying: achievable rates under limited dynamic range,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, p. 1541C1553, 2012.

[2] K. Singh, S. Biswas, M. -L. Ku, and M. F. Flanagan, “Transceiver design and power control for full-duplex ultra-reliable low-latency communication systems,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1392–1406, 2022.

[3] J. Lee, E. Tejedor, K. Ranta-aho et al., “Spectrum for 5G: global status, challenges, and enabling technologies,” *IEEE Communications Magazine*, vol. 56, no. 3, pp. 12–18, 2018.

[4] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, “A flexible 5G frame structure design for frequency-division duplex cases,” *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, 2016.

[5] A. A. Esswie and K. I. Pedersen, “On the ultra-reliable and low-latency communications in flexible TDD/FDD 5G networks,” in *Proceedings IEEE CCNC*, Las Vegas, NV, USA, Jan. 2020.

[6] A. A. Esswie and K. I. Pedersen, “Analysis of outage latency and throughput performance in industrial factory 5G TDD deployments,” in *Proceedings IEEE VTC-Spring*, pp. 1–6, Piscataway, NJ, USA, 2021.

[7] D. Wang, M. Wang, P. Zhu, J. Li, J. Wang, and X. You, “Performance of network-assisted full-duplex for cell-free massive MIMO,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1464–1478, 2020.

[8] J. Li, Q. Lv, P. Zhu, D. Wang, J. Wang, and X. You, “Network-assisted full-duplex distributed massive MIMO systems with beamforming training based CSI estimation,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2190–2204, 2021.

[9] J. Kant Chaudhary, J. Francis, A. Noll Barreto, and G. Fettweis, "Latency in the uplink of massive MIMO CRAN with packetized fronthaul: modeling and analysis," in *Proceedings IEEE WCNC*, pp. 1–7, Piscataway, NJ, USA, 2019.

[10] B. Soret, P. Popovski, and K. Stern, "A queueing approach to the latency of decoupled UL/DL with flexible TDD and asymmetric services," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1704–1708, 2019.

[11] A. P. K. Reddy, N. Kumar, S. S. A. Tirumalasetty, S. Srinivasan, and J. V. B. James, "Latency analysis for IMT-2020 radio interface technology evaluation," in *2020 IEEE 3rd 5G World Forum (5GWF)*, pp. 613–618, Piscataway, NJ, USA, 2020.

[12] ITU-R, "IMT for 2020 and beyond," International Telecommunication Union Radiocommunication sector (ITU-R), Article, 07 2020. [Online]. Available: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx.

[13] 3GPP, "Study on scenarios and requirements for next generation access technologies," 3rd Generation Partnership Project (3GP-P), Technical Report (TR) 38.913, 07 2020, version 16.0.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/Specifi-cationDetails.aspx?specificationId=2996.

[14] Z. T. Demir, E. Bjrnson, and L. Sanguinetti, "Foundations of user-centric cell- free massive MIMO," *Foundations and Trends in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2020.

[15] A. Chowdhury, R. Chopra, and C. R. Murthy, "Can dynamic TDD enabled half-duplex cell-free massive MIMO outperform full-duplex cellular massive MIMO?," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4867–4883, 2022.

[16] C. Zheng, F.-C. Zheng, J. Luo, and D. Feng, "Open-loop communications for up-link URLLC under clustered user distribution," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11509–11522, 2021.

[17] Y. Zhu, J. Li, P. Zhu, D. Wang, H. Ye, and X. You, "Load-aware dynamic mode selection for network-assisted full-duplex cell-free large-scale distributed MIMO systems," *IEEE Access*, vol. 10, pp. 22301–22310, 2022.