In-Band Full-Duplex Relaying for 5G Cellular Networks with Wireless Virtualization

Gang Liu, F. Richard Yu, Hong Ji, Victor C. M. Leung, and Xi Li

Abstract

Recent advances in self-interference cancellation techniques enable in-band FDR systems, which transmit and receive simultaneously in the same frequency band with high spectrum efficiency. Meanwhile, wireless virtualization enables abstraction and sharing of infrastructure and radio spectrum resources in wireless networks. Both in-band FDR and wireless virtualization are promising technologies for 5G cellular networks. In this article, we first give a brief survey of them to present their basic ideas, enabling technologies, characteristics, and so on. Then we propose a virtual resource management architecture for in-band FDR networks, where multiple service providers and multiple mobile network operators coexist. Simulation results are presented to show the effectiveness of the proposed scheme.

With the development of a global economy and advances in information and communications technology (ICT), there has been a tremendous growth in the subscription of mobile cellular and mobile broadband services. It is predicted that the worldwide mobile traffic by 2020 will reach a 33 times increase compared to the figures in 2010. Meanwhile, Internet access will become dominated by wireless devices such as smartphones, tablets, machines, and sensors. This unprecedented growth in wireless devices and mobile traffic has motivated the research and development of next generation cellular networks (i.e., fifth generation, 5G) with higher data rate, spectrum efficiency, and energy efficiency, as well as lower latency [1].

Among the emerging technologies for 5G cellular networks, in-band full-duplex wireless [2] has become a hot research topic. In the past, most researchers thought it was generally not possible for radios (e.g., base stations, relays, or mobiles) to receive and transmit on the same frequency band at the same time due to the strong self-interference from the transmitter to the receiver. As a result, a long-held assumption in wireless system design is that radios can only operate in half-duplex or out-of-band full-duplex mode, meaning that they transmit and receive either at different times or over different frequency bands. With recent advances in self-interference cancellation technologies, research works [2–5] have shown the feasibility of in-band full-duplex wireless, which allows radios to receive and transmit on the same frequency band simultaneously. In fact, in-band full-duplex wireless not only has the potential to double the spectrum efficiency in the physical layer, but also can help to solve some important problems in existing wireless networks, such as hidden terminals, loss of throughput due to congestion, and large end-to-end delays [2].

Gang Liu, Hong Ji, and Xi Li are with Beijing University of Posts and Telecommunications.

F. Richard Yu is with Carleton University.

Victor C. M. Leung is with the University of British Columbia.

Another important technology, *relaying*, which can improve the throughput and coverage significantly [6], has been regarded as a key component of 5G networks. Nevertheless, most existing works on relaying only focus on *half-duplex relaying* (HDR) due to its implementation simplicity, which suffers serious loss in spectrum efficiency. To overcome this problem, several effective schemes have been proposed, including successive relaying [7], two-way relaying [7], buffer-aided relaying [8], frame-level virtual full-duplex relaying (FDR) [9], and out-of-band FDR [10]. However, most of them are still restrained by the basic limitation of half-duplex or out-of-band full-duplex operation. To combat the spectrum efficiency loss more fundamentally, it is natural to make relays operate in the in-band full-duplex mode. As a typical application of in-band full-duplex wireless, *in-band FDR* is expected to integrate the merits of in-band full-duplex wireless and relaying technology.

Meanwhile, *wireless virtualization* [11, 12] is another emerging trend for next generation wireless networks. Wireless virtualization enables abstraction and sharing of infrastructure and radio spectrum resources. Consequently, the overall expenses of wireless network deployment and operation can be reduced significantly. Moreover, wireless virtualization can provide easier migration to newer products or technologies by isolating part of the network. In addition, the emerging heterogeneous wireless networks need a convergent and powerful network management mechanism, which can be provided by wireless virtualization.

Although some excellent work has been done on in-band FDR and wireless virtualization, these two important technologies are separately studied in existing works. However, they are in fact closely correlated in next generation wireless networks. For example, on one hand, the full-duplex capability of in-band FDR can be leveraged to facilitate the virtual resource slicing process in wireless virtualization; on the other hand, wireless virtualization can provide powerful and flexible resource management mechanisms for in-band FDR networks to improve the performance.

In this article, we jointly consider in-band FDR and wireless

 0890-8044/15/$25.00 © 2015 IEEE

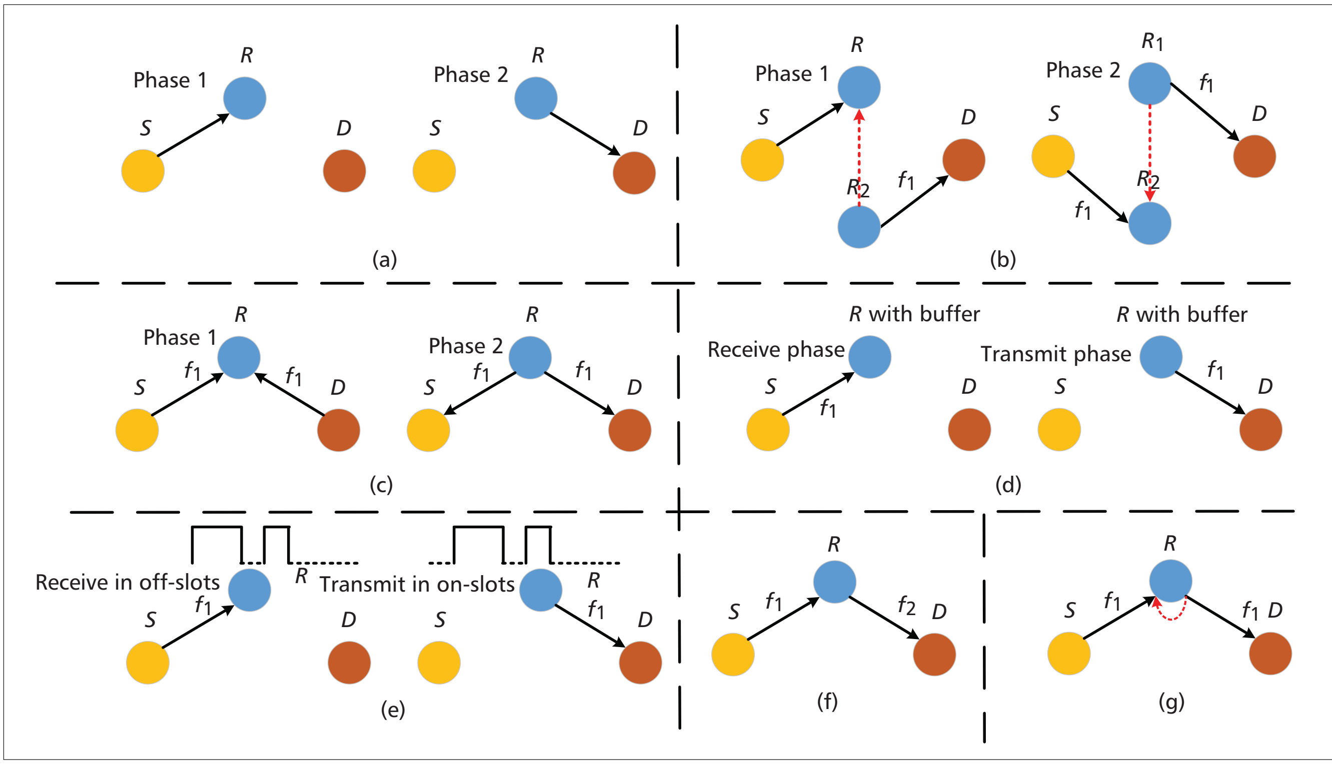

Figure 1. The illustration of different relaying protocols (The direct link from source to destination is omitted for simplicity): a) half-duplex relaying; b) successive relaying; c) two-way relaying; d) buffer-aided relaying; e) frame-level virtual FDR; f) out-of-band full-duplex relaying; g) in-band full-duplex relaying.

virtualization in 5G cellular networks. We first give a brief survey of these two technologies. For in-band FDR, we discuss the historical perspective, self-interference cancellation technologies, and merits. For wireless virtualization, the basic idea, a multi-dimensional perspective, and merits are presented. Then we propose a virtual resource management architecture for in-band FDR networks, where multiple service providers (SPs) and multiple mobile network operators (MNOs) coexist. In the proposed resource management scheme, in addition to radio spectrum, both base stations (BSs) and full-duplex ralay stations (RSs) are virtualized as virtual resources, which can be dynamically allocated to different users from different SPs. Simulation results are presented to show the effectiveness of the proposed scheme.

In-Band Full-Duplex Relaying

In-Band FDR: A Historic Perspective

In traditional relaying networks, *half-duplex relaying* [6] is commonly used due to its implementation simplicity. As shown in Fig. 1a, the transmission from a source node S to a destination node D consists of two phases, where node S transmits its data to a relay node R in the first phase, and then the relay forwards it to node D in the second phase. Clearly, HDR leads to inefficient use of system resources as extra dedicated time slot or frequency is usually required for relay transmissions. To overcome this problem, several schemes have been proposed, which may fall into one of the following five categories.

Successive Relaying — As shown in Fig. 1b, successive relaying [7] tries to mimic in-band FDR via two HDRs. In each phase, one of the relays (in the receive mode) receives new data from the source while the other relay (in the transmit mode) forwards the processed data (obtained in the previous time interval) to the destination. The role of the relays is swapped at a proper time interval. In this way, the source can send a new message to the destination in each phase as if in-band FDR was employed.

However, inter-relay interference, as shown by the red dashed arrows in Fig. 1b, is a crucial problem.

Two-Way Relaying: As shown in Fig. 1c, two-way relaying [7] is another promising spectrum-efficient transmission technique to enable bidirectional information exchange between two terminals with the help of an HDR node. In the first phase, nodes S and D transmit their data to the relay simultaneously. Then the relay broadcasts the processed superposition signal to nodes S and D in the second phase. Since parts of the superposition signal transmitted by the relay are known at nodes S and D, each node will first subtract its own message from the received signal and then detect the information from its partner at the end of the second phase. As a result, only two phases are required to accomplish the bidirectional information exchange between nodes S and D.

Buffer-Aided Relaying: As shown in Fig. 1d, buffer-aided relaying [8] is realized via traditional HDR with a buffer. The key idea is that relay R dynamically chooses its reception and transmission time based on the quality of the source-relay and relay-destination links. Thus, the relay transmits when the relay-destination channel is comparatively stronger than the source-relay channel, and otherwise receives and stores the received information in its buffer.

Frame-Level Virtual FDR: As shown in Fig. 1e, according to basic idea of rapid on-off-division duplex (RODD) [9], the frame-level virtual FDR works as follows. Unlike conventional HDR, where the relay's transmission frames are scheduled away from its reception, relay node R transmits its signal through the on-slots of a randomly generated on-off duplex mask (or signature) over every frame interval, and receives a signal through each of its own off-slots. Over the period of a

Domain	Approaches	Channel awareness	Active/ passive	Target SI	Features
Propagation domain	Path loss	Channel-unaware	Passive	Direct paths	While avoiding saturation of the receiver circuitry, the desired (incoming or outgoing) signal might be accidentally suppressed.
	Directional antenna	Channel-unaware	Passive	Direct paths	
	Antenna placement	Channel-unaware	Passive	Direct paths	
	Cross-polarization	Channel-unaware	Passive	Direct paths	
	Duplexer	Channel-unaware	Passive	Direct paths	
	Transmit beamforming	Channel-aware	Active	Direct and reflected paths	
Analog-circuit domain	Adaptive SI canceller	Channel-aware	Active	Direct and reflected paths	There is a trade-off between circuit non-ideality and signal processing complexity.
	Non-adaptive SI canceller	Channel-unaware	Active	Direct paths	
Digital domain	Digital SI canceller	Channel-aware	Active	Direct and reflected paths	Limited by the dynamic-range of ADC; a discrete-time model is needed.
	Receive beamforming	Channel-aware	Active	Direct and reflected paths	

Table 1. Different self-interference mitigation technologies.

single frame, relay node *R* can transmit a message to destination node *D*, and may simultaneously receive a message from source node *S*. Thus, RODD achieves frame-level virtual FDR using HDR and can simplify the design of higher-layer network protocols significantly.

Out-of-Band FDR: Different from HDR, out-of-band FDR [10] transmits and receives simultaneously, as shown in Fig. 1f. However, the transmitter and receiver of the relay node operate at different frequency. To avoid self-interference, the main prerequisite for good support of out-of-band FDR is to have enough frequency separation between respective frequency resources for the transmitter and receiver of the relay.

In summary, many excellent approaches have been proposed to tackle the problem of spectrum efficiency loss in HDR systems. As shown in [7–9], they can alleviate this problem to some extent. However, they are still restrained by the basic limitation of half-duplex or out-of-band full-duplex operation at relays, meaning that the relay itself can only transmit and receive either at different times or over different frequency bands. To combat the spectrum efficiency loss more fundamentally, *in-band FDR* [2, 5] transmits and receives simultaneously in the same frequency band with high spectrum efficiency, as shown in Fig. 1g. Ideal in-band FDR is promised to double the spectrum efficiency of HDR. However, the receiver of in-band FDR may suffer serious self-interference from its transmitter. To enable in-band FDR, the self-interference needs to be well suppressed using different self-interference mitigation technologies. In the rest of this article, we use "FDR" and "in-band FDR" equivalently for simplicity.

Enabling Technologies for FDR

Generally, different self-interference mitigation technologies [2, 5] can be realized in three domains: *propagation domain*, *analog-circuit domain* and *digital domain*, which could be considered as three lines of defense against self-interference. Moreover, according to whether or not they adapt to environmental effects like multipath propagation, they can operate in either a *channel-aware* or *channel-unaware* manner. In addition, existing self-interference mitigation technologies would also be divided into the following two categories: *passive suppression*, which electromagnetically isolates the transmit and receive antennas, and *active cancellation*, which exploits a node's knowledge of its own transmit signal to cancel the self-interference.

The characteristics of different self-interference (SI) mitigation technologies are summarized in Table 1. In fact, different mitigation technologies in different domains can be adopted to suppress the self-interference in FDR systems; these are discussed below.

Propagation-Domain Self-Interference Suppression — To avoid receiver saturation, it is indispensable for FDR to use propagation-domain suppression technologies as the first line of defense. For FDR systems with separate antenna deployment [5], *path loss* can be leveraged to suppress the self-interference by increasing the physical distance between transmit and receive antennas, or exploiting the surrounding obstacles (e.g., buildings, tunnels, and shielding plates) to block the self-interference propagating directly from the transmit chain to the receive chain (i.e., direct paths). Besides, *directional antennas* can be equipped such that the gain of the transmit antennas is low in the direction of the receive antennas and vice versa. Moreover, one can create a null position where the receive antenna hears much weaker self-interference from the transmit antenna via careful *antenna placement*. In addition, this goal can be achieved by placing two transmit antennas asymmetrically at ℓ and $\ell + (\lambda/2)$ distance from a receive antenna, which allows the transmit signals to add π out of phase and hence cancel each other at the receive antenna. Furthermore, *cross-polarization* serves as another approach that electromagnetically increases the isolation between the transmit and receive antennas. For instance, the transmit antenna of an FDR can be horizontally polarized, while its receive antenna only receives vertically polarized signals with the goal of avoiding interference between them. For FDR systems with shared antenna deployment [4], propagation-domain isolation is generally accomplished using a *duplexer*. For example, one can adopt a circulator, which routes the transmit signal from the transmitter to the shared antenna and routes the received signal on the antenna to the receiver to provide certain level of isolation between the transmitter and receiver front-end.

Although these techniques are effective to passively combat the direct paths of self-interference, they are sensitive to device size, antenna placement, and environmental effects.

To further deal with the reflected paths of self-interference caused by nearby scatterers, channel-aware technologies are desirable. Take MIMO FDR systems as an example: *transmit beamforming* has been widely used to spatially suppress the self-interference (both direct paths and reflected paths) [5]. The basic idea of *transmit beamforming* is to electronically steer the transmit weight of different antenna elements in an attempt to zero the radiation pattern at receive antennas.

While the propagation-domain self-interference suppression techniques above adjust the transmit/receive patterns to mitigate self-interference, the desired (incoming or outgoing) signals might be accidentally suppressed as well [2].

Analog-Circuit-Domain Self-Interference Cancellation: As the second line of defense, analog-circuit-domain self-interference cancellation technologies can be adopted to further cancel the self-interference in the analog receive-chain circuitry by subtracting a predicted copy of self-interference from the received signal before it is digitized. According to whether they can respond to the changing environmental effects, they can be either *non-adaptive* or *adaptive* [4]. The non-adaptive ones are unaware of the changes in environment and use fixed parameters (e.g., gain, phase, and delay) to form the predicted self-interference when an FDR is designed or calibrated. Hence, they might be sensitive to the reflected paths of self-interference. In contrast, the adaptive ones dynamically adjust the parameters according to the reflection channel, and hence can effectively mitigate both the direct and reflected paths of self-interference.

Generally, there are several ways to form a copy of the predicted self-interference [2]. One can tap the transmit signal at the transmit antenna feed and electronically process it to form the predicted self-interference in the analog circuit domain. In this way, the non-idealities like oscillator phase noise and HPA distortion can be better captured. However, doing so requires analog-domain signal processing, which becomes difficult in the case of wideband signal. Also, one can also generate the predicted self-interference by tapping the transmit signal in the digital domain, adjusting the gain/phase/delay digitally and then converting it to analog signal for self-interference cancellation. In this way, we can take advantage of digital signal processing techniques to cancel the reflected paths of self-interference. Nevertheless, the effectiveness of self-interference cancellation is affected by the downstream analog circuit non-idealities.

Digital-Domain Self-Interference Mitigation: With the self-interference mitigation technologies above, the self-interference can be partially suppressed, which might be enough for amplify-and-forward FDR. However, such suppression is not sufficient in some cases, especially for decode-and-forward FDR. Hence, digital-domain self-interference mitigation technologies can be adopted after the analog-to-digital conversion (ADC) as the last line of defense against self-interference. Generally, digital-domain self-interference mitigation technologies include a *digital self-interference canceller* and *receive beamforming* [2]. A digital self-interference canceller first estimates the residual self-interference after propagation and analog-circuit-domain suppression, and then this prediction is subtracted from the received baseband samples in the digital domain. Moreover, receive beamforming can also be adopted in MIMO FDR systems, in which the per-antenna received signals are weighted by separate adaptive complex-valued gains before being summed together. Although receive beamforming could in principle be implemented in the analog domain, it is far more common in practice to implement it in the digital domain for reasons of circuit complexity and power consumption.

In general, there are transmitter distortion and receiver distortion due to the non-ideality of amplifiers, oscillators, ADCs, and digital-to-analog converters (DACs). To implement digital-domain self-interference mitigation technologies in FDR systems, it is essential to build an equivalent discrete-time baseband model that captures everything between the DAC and ADC of the FDR. An accurate model needs to capture not only the transmitter and receiver distortion, but also the propagation and analog-circuit domain self-interference suppression, and even the inter-antenna propagation channels. Such a model would be matrices of z-domain polynomials in the case of a frequency-selective channel with multiple transmit and receive antennas. Furthermore, it is simplified to matrices of complex numbers in the non-selective case, or scalar values in the case of single transmit and receive antenna. When an accurate self-interference model is used, the residual self-interference can be effectively suppressed to the noise floor [4].

Feasibility and Merits of FDR

Although FDR was considered impractical in the past, recent research has demonstrated its feasibility using a combination of different self-interference mitigation techniques in different domains described above [2–5]. Indeed, FDR has received a lot of attention from both academia and industry. For example, several proposals on FDR have been released by standardization groups (e.g., IEEE and 3GPP) for sufficient isolation between the transmit and receive antennas at the relay. As a typical application of in-band full-duplex wireless, FDR is expected to integrate the merits of in-band full-duplex wireless (high spectrum efficiency) and relaying technology (improved throughput, coverage, and reliability). Therefore, we believe that FDR will be a promising technology for 5G cellular networks.

Wireless Virtualization

The Basic Idea of Wireless Virtualization

In the ICT sector, *virtualization* has become a popular concept in different areas, for example, virtual memory, virtual machine, virtual storage, and virtual data center. Virtualization involves abstraction and sharing of resources among different parties. With virtualization, the overall cost of equipment and management can be significantly reduced due to the increased hardware utilization, decoupled functionalities from infrastructure, easier migration to newer services and products, and flexible management[11].

In wired networks, virtualization has occurred for decades, for example, virtual private networks (VPNs) over wide area networks and virtual local area networks (VLANs) in enterprise networks. Recently, network virtualization has been actively used in Internet research testbeds (e.g., G-Lab and 4WARD) and also applied in the cloud computing environment. It aims to overcome the resistance of the current Internet to fundamental architecture changes. Network virtualization has been considered as one of the most promising technologies for the future Internet [13].

With the tremendous growth in wireless traffic and services, it is natural to extend virtualization to wireless networks. Similar to wired network virtualization, in which the physical infrastructure owned by one or more providers can be shared among multiple service providers, wireless virtualization [11] needs the physical wireless infrastructure and radio resources to be abstracted and isolated to a number of virtual slices, which then can be offered to different service providers. That is, virtualization, regardless of networks being wired or wireless, can be considered as a process splitting the entire network. However, the distinctive properties of the wireless environment, in terms of time-varying channels, attenuation,

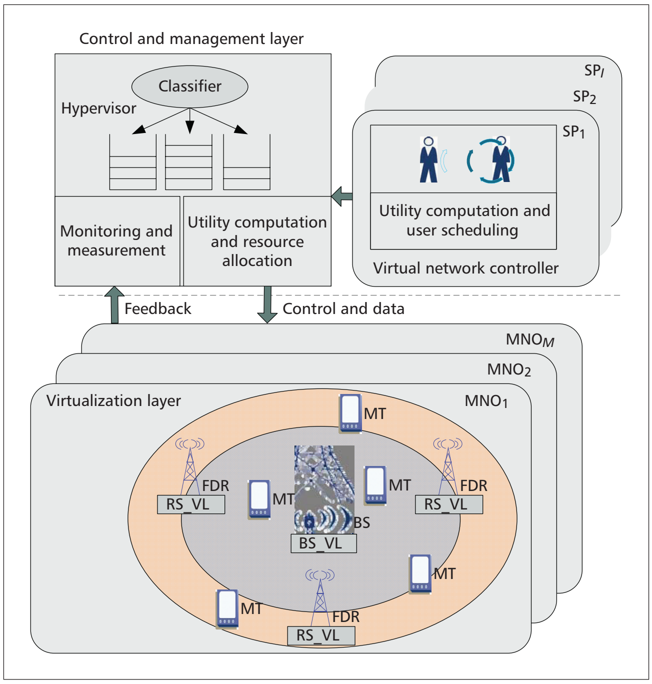

Figure 2. The resource management architecture for virtualized FDR networks with multiple MNOs and multiple SPs.

mobility, broadcast, and so on, make the problem more complicated. Furthermore, wireless networks contain many more access technologies compared to wired networks, and each access technology has its particular characteristics, making convergence, sharing, and abstraction difficult to achieve. In short, wireless virtualization can be considered as technologies in which *physical wireless network infrastructure* and *physical radio resources* can be *abstracted* and *sliced* into *virtual wireless resources*, and *shared* by multiple *parties* with a certain degree of *isolation* between them. In other words, virtualizing a wireless network is to realize the process of abstracting, slicing, isolating, and sharing wireless resources.

A Multi-Dimensional Perspective Toward Wireless Virtualization

Generally, wireless virtualization is a broad area of research and has different aspects. To understand it from an overall viewpoint, we present a multi-dimensional perspective toward wireless virtualization.

Technology-Dependent Wireless Virtualization: Obviously, virtualization can be applied to a large variety of wireless networks, including cellular networks, wireless local area networks, personal area networks, sensor networks, and so on. Nevertheless, different wireless networks may have different characteristics, such as different management architecture (centralized, distributed, etc.), different multi-access technologies — space-division multiple access (SDMA), time-division multiple access (TDMA), frequency-division multiple access (FDMA), code-division multiple access (CDMA), orthogonal FDMA (OFDMA), carrier sense multiple access with collision avoidance (CSMA/CA), and so on — different duplexing modes (time-division duplexing [TDD], frequency-division duplexing [FDD], etc.), and different operation frequency bands (e.g., unlicensed, licensed). Those differences may lead to different ways to abstract and share the virtual resources. Therefore, wireless virtualization is generally technology-dependent.

Scope of Wireless Virtualization: In fact, the scope of wireless virtualization can also vary from one network to another. Specifically, one may focus on the virtualization of an individual wireless device (e.g., base station, mobile terminal). Also, one can virtualize a whole wireless access network by abstracting and sharing all the physical resources (wireless devices, frequency bands, transmission power, etc.). Ultimately, one or more entire mobile networks, including both access networks and core networks, can also be fully virtualized.

Depth of Wireless Virtualization: According to the extent of penetration of slicing and partitioning on the wireless resources, the depth of virtualization can also be different [11]. For instance, a *flow-based wireless virtualization* scheme focuses on providing isolation, scheduling, management, and service differentiation between both uplink and downlink traffic flows from different slices. With flow-based virtualization, commercial hardware can be treated as a black box component, and all virtual slices share the same wireless protocol stack. However, it has limitations in terms of the granularity of control and resource allocation. With deeper virtualization, *protocol-based wireless virtualization* focuses on the isolation, customization, and management of multiple wireless protocol instances. These implementations can simultaneously support different wireless protocol stacks on the same radio front-end by decoupling the wireless protocols from the physical hardware. In addition, *RF front-end and spectrum-based wireless virtualization* are the deepest form of slicing currently possible. It involves the abstraction and dynamic allocation of the frequency spectrum into slices through spectrum reshaping and radio slicing techniques. It allows a given wireless protocol stack to use arbitrary and potentially non-contiguous frequency bands. At last, we would like to note that the coexistence of these three virtualization levels is also possible in the same virtualized network.

Merits of Wireless Virtualization

With wireless virtualization, network infrastructure can be decoupled from the services it provides. Hence, differentiated services can coexist on the same infrastructure while maximizing its utilization. Consequently, multiple wireless virtual networks operated by different SPs can dynamically share the physical substrate wireless networks operated by MNOs. Since wireless virtualization enables the sharing of infrastructure and radio spectrum resources, the capital expenses and operation expenses of wireless (radio) access networks, as well as core networks, can be reduced significantly. Moreover, the SPs, who may provide different telecom services (e.g., VoIP, video, and over-the-top services), can help MNOs attract more users, while MNOs can produce more revenue by leasing the isolated virtualized networks to them. Meanwhile, wireless virtualization provides easier migration to newer products or technologies while supporting legacy products by isolating parts of the networks. In addition, the emerging heterogeneous wireless networks need a convergent and powerful network management mechanism, which can be provided by wireless virtualization. Due to those benefits, wireless virtualization can play an important role in the research and development of 5G networks.

Virtual Resource Management for FDR 5G Cellular Networks

Virtual Resource Management Architecture

Based on the descriptions above, we can see that both FDR and wireless virtualization are promising technologies in 5G networks. Although these two technologies are separately studied in existing works, they are in fact closely correlated. On one hand, recent advances in self-interference cancellation technologies enable in-band full-duplex terminals to simultaneously transmit and receive on separate and arbitrary spectrum fragments using a single RF front-end [1, 14]. Therefore, one can leverage this full-duplex capability to build an abstraction, which allows an FDR to be flexibly sliced into separate independent virtual FDRs operating on different spectrum fragments. In this way, we can even realize a deep level (e.g., the RF front-end level) of virtualization for FDR networks. Of course, other levels (e.g., flow-based and protocol-based levels) of virtualization can also be implemented. On the other hand, while FDR can improve performance significantly, resource management is a complicated problem in FDR networks since the residual self-interference couples the transmission and reception of FDR. Wireless virtualization can provide powerful and flexible resource management mechanisms for FDR networks.

Motivated by the observations above, we consider wireless virtualization and FDR jointly in cellular networks. We then propose a resource management architecture for virtualized FDR networks as shown in Fig. 2, where multiple MNOs and multiple SPs coexist. We assume that M MNOs offer wireless access services simultaneously in a certain geographical area, which is common in practice, especially in urban areas. Each MNO creates and manages a cellular network with one BS, several full-duplex RSs, and a certain amount of wireless spectrum (e.g., subchannels). There are I SPs, which provide various services to their subscribers through the same substrate network consisting of the physical networks of multiple MNOs. In order to efficiently share the physical resources from multiple MNOs among multiple SPs, the resource management architecture is described as follows. According to the general frameworks of wireless virtualization in [11, 12], the considered resource management architecture is divided into two separate layers: the *virtualization layer* (VL) and the *control and management layer* (CML). On one hand, the VL, either BS_VL or RS_VL, is responsible for the virtualization and abstraction of physical resources in BSs and RSs. Both the infrastructure nodes (e.g., BSs and RSs) and the spectrums from different MNOs are virtualized and shared by the SPs. Moreover, the VL also provides the CML with the interfaces needed to control virtualized resources (e.g., BSs, RSs, subchannels, transmission power). On the other hand, the CML is responsible for the management of the virtualized resources (e.g., resource allocation and scheduling). The resource management functions in the CML are realized by several *virtual network controllers* and a *hypervisor*. Each SP has a virtual network controller, which is used to schedule users, determine their quality of service (QoS) requirements, and report them to the hypervisor. The whole virtualized network has one hypervisor, which is used to dynamically allocate the virtual resources from multiple MNOs to different SPs based on diverse QoS requirements and feedback information (e.g., available subchannels and transmission power). With the virtual resource management architecture above, each user of a certain SP could access its service via different access points (either BSs or RSs) and different spectrum resources from different MNOs. Also, the SPs do not need to know the channel state information (CSI) of their subscribers, and they can only focus on the service status of their subscribers.

Virtual Resource Management Optimization

To facilitate virtual resource management optimization, the utility functions of different parties are defined as follows. For a given user u, the utility function can be defined as its total service rate

$$R(u) = R_u^1 + R_u^2 \quad (1)$$

where R_u^1 is the service rate of user u when it is served by any BS directly (i.e., one-hop) and R_u^2 is the service rate of user u when it is served by any RS indirectly (i.e., two-hop). In fact, $R(u)$ is a function of the CSI information, access point selection indicators, subchannel allocation indicators, and the transmission power on each subchannel. For a given SP i with user set U_i, the utility function is defined as the weighted sum service rate of its served users. It can be expressed as

$$F_{SP}(U_i) = \sum_{u \in U_i} \omega_u R(u) \quad (2)$$

where ω_u is a positive weight for user u specified by each SP based on different user scheduling criteria. For the MNOs, the total utility considers not only the revenue earned by serving users but also the cost of energy consumption. It is defined as the difference between the weighted sum utility of all SPs and the total energy consumption cost for all MNOs, that is,

$$F_{MNOs}(U) = \sum_{i=1}^{I} \sum_{u \in U_i} \beta_i \omega_u R(u) - cP_{total} \quad (3)$$

where β_i is the price for SP i charged by the hypervisor, the first term is the total amount of money that the SPs pay to the MNOs, the second term is the total cost for energy consumption of all the MNOs, c is the price of per unit energy consumption, and P_{total} is the sum of the transmission power and circuit energy consumption at all the BSs and RSs. Moreover, the circuit energy consumption includes the power dissipations in the transmit filters, mixers, frequency synthesizers, and DACs. For simplicity, we omit the explicit expressions for $R(u)$ and P_{total}.

In order to utilize the virtual resources efficiently, the resource management problem can be formulated as an optimization problem at the hypervisor, which maximizes the total utility (Eq. 3) of all MNOs under the constraints of individual power consumption at each BS/RS, the minimum rate constraint for each user, and the constraints of subchannel allocation. The formulated problem jointly optimizes the access point selection (BS or RS), subchannel allocation, and transmission power allocation across all the MNOs and SPs. However, this joint optimization problem is a mixed combinatorial and non-convex optimization problem, which involves enormous overhead to solve. Therefore, how to solve the resource optimization problem efficiently is crucial. To reduce the computational complexity, the considered problem can be converted into a suboptimal convex problem at first by relaxing the resource allocation indicators for access point selection and subchannel allocation to a real value instead of Boolean. Then, to reduce the signaling overheads involved in CSI exchange, we develop an efficient distributed algorithm to solve the transformed problem based on the emerging alternating direction method of multipliers (ADMM) [15], which is a simple but powerful algorithm for distributed convex optimization. The proposed ADMM-based distributed algorithm takes the form of a decomposition-coordination procedure, in which the solutions to small local subproblems are coordinated to find a solution to a large global problem. More specifically, in different MNOs each BS only determines its local variables based on the local information, and the hypervisor is responsible for achieving consensus between the local variables and the global variables according to the global constraints. In this

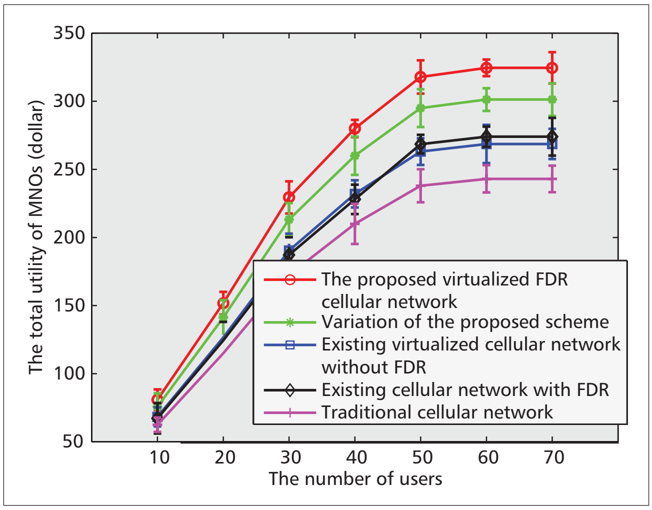

Figure 3. The total utility of MNOs for different schemes with two MNOs, two SPs, and uniform traffic distribution.

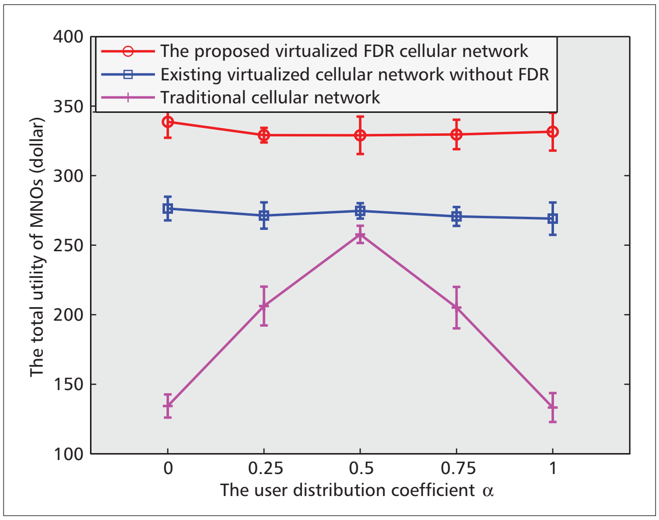

Figure 4. The effect of user distribution. (There are two MNOs and two SPs, and the number of users equals the total number of subchannels.)

way, there is no need to exchange CSI information between BSs and the hypervisor, which will reduce the signaling overhead significantly.

Simulation Results and Discussions

In this part, the effectiveness of the proposed virtualized FDR network and its resource management architechture is demonstrated by simulation. We first compare the performance against several other schemes, and then investigate the effects of some important parameters (e.g., user distribution) on its performance. The simulation parameters are given as follows.

We consider a geographical area in an urban environment, which is covered by several MNOs simultaneously. Each MNO has one macrocell in this area. The number of cells per MNO are varied in different simulations. The basic configuration for each cell is based on the 3GPP case 1 relay scenario in TR36.814. Each cell has one BS, three RSs, and 10 MHz spectrum resource. The whole spectrum band consists of 24 subchannels with bandwidth equal to 375 kHz. Each BS, RS, and user terminal is equipped with two antennas. The maximum transmission power at each BS and RS is set to 46 dBm and 30 dBm, respectively. The minimum transmission rate for each user terminal is 0.5 Mb/s. Each user can only be served by one distinct subchannel; hence, each cell can only serve at most 24 users simultaneously if virtualization is not used. The results below are averaged from 20 simulation runs, and the 95 percent confidence interval is also shown in each figure.

Performance Comparison — To illustrate the performance of the proposed virtualized FDR network, we compare the following schemes:

1. A traditional cellular network without FDR and virtualization
2. A cellular network with FDR but without virtualization
3. A virtualized cellular network without FDR
4. A variation of the proposed virtualized FDR network with only spectrum virtualization;
5. The proposed virtualized FDR network with spectrum and infrastructure virtualization

In this part, we assume there are two MNOs and two SPs. All the weights ω_u are set to one. In addition, β_1 = US\$25/Mb/s, β_2 = US\$30/Mb/s, and c = 5 US\$/W. Also, we define an user distribution coefficient α, which is the ratio of the access requests cominge from SP_1 to the total number of access requests in the whole area. In this subsection, we only consider a uniform traffic distribution scenario, in which the access requests equally come from SP_1 and SP_2 (i.e., $\alpha = 0.5$).

The total utility of MNOs is illustrated in Fig. 3. It can be observed that the proposed scheme and its variation (i.e., scheme 4) always outperform the existing virtualized network without FDR. This is because the proposed scheme with FDR is able to further enhance the service rate of cell edge users with higher spectrum efficiency and lower power consumption, which results in larger utility for MNOs, SPs, and users. Also, it can be found that the variation of the proposed scheme with only spectrum virtualization is superior to the existing cellular network with FDR but without virtualization. Moreover, there is another appreciable performance gain of our proposed scheme compared to its variation. The reason is that with both spectrum and infrastructure virtualization, there is a higher degree of freedom for resource allocation, and hence a user is able to connect to a better access point via a better subchannel with better channel conditions. Therefore, better utility can be obtained from our proposed virtualized FDR networks. That is, we can benefit from both FDR and wireless virtualization simultaneously in 5G cellular networks.

Meanwhile, it can also be found from Fig. 3 that when the number of users is larger than the number of available subchannels, there is only a small performance gain for each scheme. This is because the maximum number of served users is equal to the number of available subchannels. More users can only lead to limited multi-user diversity gain. In addition, it can be observed that the performance of the existing virtualized network without FDR is not necessarily always better than that of the existing FDR network without virtualization. This is because the performance of those two schemes is closely related to the number of users, user distribution, channel conditions, and the ratio of one-hop and two-hop users. Different parameters may result in different performance.

The Effects of Different Parameters: In this part, the effects of some important parameters (e.g., user distribution and the network scale) is demonstrated. As shown in Fig. 4, the proposed virtualized FDR network is robust to user distribution. Moreover, its performance is better than the traditional cellular network, especially when the user distribution is not balanced. This is because no matter which SP the users belong to, they can share the physical resources from different MNOs in our proposed virtualized FDR networks. By contrast, they can only access the physical resources from their corresponding MNO in traditional cellular networks, and hence some of them may not be served due to limited resources when unbalanced user

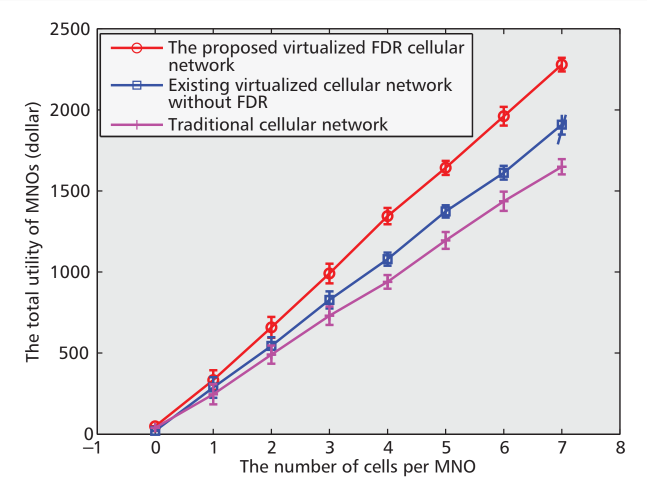

Figure 5. The effect of network scale. (There are two MNOs and two SPs. Every cell is fully loaded.)

distribution occurs. Finally, we expand the network scale by increasing the number of cells in each MNO. Then the effect of network scale is shown in Fig. 5, when every cell is fully loaded. We can observe that the total utility of MNOs increases almost linearly with the network scale. In other words, the proposed scheme can be extended to large-scale networks.

Conclusions and Future Work

In-band FDR and wireless virtualization are two promising technologies for 5G cellular networks. In this article, we first give a brief survey of them. For in-band FDR, a historic perspective, the self-interference cancellation technologies, and the merits are discussed. For wireless virtualization, we present the basic idea and a multi-dimensional perspective. Then we propose a virtual resource management architecture for in-band FDR networks. It is demonstrated that the proposed scheme can substantially improve the performance of virtualized FDR networks, where SPs, MNOs, and users can benefit from these two emerging technologies in 5G cellular networks. Future work is in progress to consider content caching in the proposed framework.

Acknowledgment

This work is jointly supported by the National 863 Project of China (Grant No. 2014AA01A705) and the National Natural Science Foundation of China (Grant No. 61271182). We would like to thank the anonymous reviewers for their constructive suggestions and comments, which have greatly helped us to improve the content, quality, and presentation of this article. We also express our gratitude to the Editor-in-Chief, Prof. Xuemin (Sherman) Shen, for his encouragement during the revision of this article.

References

[1] S. Hong *et al.*, "Applications of Self-Interference Cancellation in 5G and Beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 114–21.
[2] A. Sabharwal *et al.*, "In-Band Full-Duplex Wireless: Challenges and Opportunities," arXiv:1311.0456v1 [cs.IT], submitted 3 Nov. 2013.
[3] A. K. Khandani, "Two-Way (True Full-Duplex) Wireless," *Proc. 13th Canadian Wksp. Info. Theory*, Toronto, Canada, June 2013.
[4] D. Bharadia, E. McMilin, and S. Katti, "Full Duplex Radios," *Proc. ACM SIGCOMM '13*, Hong Kong, China, Aug. 2013.
[5] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of Loopback Self-Interference in Full-Duplex MIMO Relays," *IEEE Trans. Signal Processing*, vol. 59, no. 12, Dec. 2011, pp. 5983–93.
[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Info. Theory*, vol. 50, no. 12, Dec. 2004, pp. 3062–80.
[7] Z. Ding *et al.*, "On Combating the Half-Duplex Constraint in Modern Cooperative Networks: Protocols and Techniques," *IEEE Wireless Commun.*, vol. 19, no. 6, Dec. 2012, pp. 20–27.
[8] N. Zlatanov and R. Schober, "Buffer-Aided Half-Duplex Relaying Can Outperform Ideal Full-Duplex Relaying," *IEEE Commun. Letters.*, vol. 17, no. 3, Mar. 2013, pp. 479–82.
[9] L. Zhang and D. Guo, "Virtual Full Duplex Wireless Broadcasting via Compressed Sensing," *IEEE/ACM Trans. Networking*, Oct. 2013.
[10] "Full Duplex Configuration of Un and Uu Subframes for Type I Relay," Samsung, tech. rep. 3GPP TSG RAN WG1 R1-100139, Jan. 2010.
[11] H. Wen, P. K. Tiwary, and T. Le-Ngoc, *Wireless Virtualization*, Springer-Briefs in Computer Science, Sept. 2013.
[12] F. Fangwen and U. C. Kozat, "Stochastic Game for Wireless Network Virtualization," *IEEE/ACM Trans. Networking*, vol. 21, no. 1, Apr. 2013, pp. 84–97.
[13] N. M. K. Chowdhury and R. Boutaba, "A Survey of Network Virtualization," *Comp. Net.*, vol. 54, no. 5, Apr. 2010, pp. 862–76.
[14] S. Hong, J. Mehlman, and S. Katti, "Picasso: Flexible RF and Spectrum Slicing," *Proc. ACM SIGCOMM '12*, Helsinki, Finland, Aug. 2012.
[15] S. Boyd *et al.*, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, Jan. 2011, pp. 1–122.

Biographies

GANG LIU received his B.S. degree in electronics and information engineering from Sichuan University, China, in 2010. He is currently working toward his Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), China. He has also been with the University of British Columbia as a visiting scholar since November 2013. His current research interests include 5G cellular networks, full-duplex wireless, resource management, cross-layer design, and TCP/IP in satellite communication systems. He won the second prize in the 2009 National Undergraduate Electronic Design Contest and Best Paper Award at IEEE ICC '14. He has served as a reviewer for the *International Journal of Communication Systems*, *China Communications*, IEEE ICC '12, IEEE GLOBECOM '13, and so on.

F. RICHARD YU [SM] is an associate professor at Carleton University, Canada. He received the IEEE Outstanding Leadership Award in 2013, Carleton Research Achievement Award in 2012, Ontario Early Researcher Award (formerly Premier's Research Excellence Award) in 2011, Excellent Contribution Award at IEEE/IFIP TrustCom 2010, Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009, andthe Best Paper Awards at IEEE ICC 2014, IEEE GLOBECOM 2012, IEEE/IFIP TrustCom 2009, and the International Conference on Networking 2005. His research interests include cross-layer design, security, green IT, and QoS provisioning in wireless networks. He serves on the Editorial Boards of several journals, including *IEEE Transactions on Vehicular Technology* and *IEEE Communications Surveys & Tutorials*. He has served as a member of the Technical Program Committees (TPCs) and TPC Co-Chair of several major conferences.

HONG JI [SM] received her B.S. degree in communications engineering and M.S. and Ph.D degrees in information and communications engineering from BUPT in 1989, 1992, and 2002, respectively. From June to December 2006, she was a visiting Sscholar with the University of British Columbia (UBC), Vancouver, Canada. She is currently a professor at BUPT. She also works on national science research projects, including the Hi-Tech Research and Development Program of China, the National Natural Science Foundation of China, and so on. Her research interests include heterogeneous networks, peer-to-peer protocols, and cognitive radio.

VICTOR C.M. LEUNG [F] received his B.A.Sc. (Hons.) degree in electrical engineering from UBC in 1977, and was awarded the APEBC Gold Medal as head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Natural Sciences and Engineering Research Council Postgraduate Scholarship and completed his Ph.D. degree in electrical engineering in 1981. From 1981 to 1987, he was a senior member of technical staff in the satellite systems group at MPR Teltech Ltd. He started his academic career in the Department of Electronics at the Chinese University of Hong Kong in 1988. He returned to UBC as a faculty member in 1989, where he currently holds the positions of professor and TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering.

XI LI is an associate professor in the School of Information and Communication Engineering, BUPT. She received her B.E. and Ph.D. degrees in communication and information systems from BUPT in 2005 and 2010, respectively. She has served as Chair of the Special Track on Cognitive Testbeds in Chinacom 2011, a TPC member of the MAC Layer Track of IEEE WCNC 2012/2013, and the MAC and Cross-Layer Design Track of PIMRC 2012. Her current research interests include broadband wireless communications and heterogeneous networks.