Full-Duplex Non-Orthogonal Multiple Access for Next Generation Wireless Systems

Full-Duplex Non-Orthogonal Multiple Access for Next Generation Wireless Systems

Mohammadali Mohammadi, *Member, IEEE,* Xiaoyan Shi, *Student Member, IEEE,* Batu K. Chalise, *Senior Member, IEEE,* Zhiguo Ding, *Senior Member, IEEE,* Himal A. Suraweera, *Senior Member, IEEE,* Caijun Zhong, *Senior Member, IEEE,* and John S. Thompson, *Fellow, IEEE*

***Abstract*—In this article, we study the combination of non-orthogonal multiple access (NOMA) and full-duplex operation as a promising solution to improve the capacity of next-generation wireless systems. We study the application of full-duplex NOMA transmission in wireless cellular, relay and cognitive radio networks, and demonstrate achievable performance gains. It is shown that the effects of self-interference and inter-user interference due to full-duplex operation can be effectively mitigated by optimizing/enhancing the beamforming, power control, and link scheduling techniques. We also discuss research challenges and future directions so that full-duplex NOMA can be made practical in the near future.**

I. Introduction

The proliferation of communication devices and high-speed data services has led to a demand for increased spectral efficiency in emerging wireless systems such as 5G. To meet this demand, full-duplex communications, i.e., simultaneously transmitting and receiving radio signals on the same frequency band, has become a promising solution [1]. Although, the practical throughput gain of full-duplex operation is limited by the self-interference (SI) [2], recent advances in antenna and transceiver design show promise to cancel the SI up to the receiver noise floor [3], thus making it a promising solution for implementing future wireless systems.

Non-orthogonal multiple access (NOMA) is a another technology recognized for implementing next generation wireless networks [4]. In NOMA, although user multiplexing through power allocation reduces the allocated power to each user, users with higher channel gains (NOMA-strong user) and users with smaller channel gains (NOMA-weak user) benefit from being scheduled more frequently with more assigned bandwidth. Therefore, the coverage and average throughput can be increased significantly with a moderate increase in the system complexity [4], [5]. Moreover, NOMA is compatible with orthogonal frequency division multiple access in the downlink (DL), thus many users can be served simultaneously, making it ideal for massive connectivity and low latency transmission in 5G networks. NOMA can also be applied to multi-antenna networks for performance improvement [6].

Since NOMA and full-duplex are complementary in principle, their integration is worthwhile to study in detail [7]. A typical objective of full-duplex transceivers in NOMA systems is to enable reliable and simultaneous transmission in uplink (UL) and DL channels. More specifically, data from UL users, and data to be sent to DL users are received and transmitted simultaneously at the same frequency. Furthermore, by incorporating the full-duplex operation, a NOMA user close to the base station (BS) can simultaneously receive and forward data to a distant NOMA user. Recent studies have shown that integration of NOMA and full-duplex operation can not be accomplished trivially. For example, SI and inter-user interference due to full-duplex operation can significantly degrade the NOMA performance. As such, additional design constraints should be considered for user pairing and beamforming design.

This article is focused on full-duplex NOMA systems. Specifically, we quantify the performance gains of wireless systems with UL/DL, cooperative and cognitive radio transmission due to the use of NOMA and full-duplex operation. We also discuss resource management for a full-duplex NOMA system. Finally, research challenges and future directions are summarized.

II. Full-duplex NOMA Transmissions

A. NOMA Basics

Different from orthogonal multiple access (OMA), NOMA uses the power domain to realize multiple access [4]. Unlike the "water-filling" principle, NOMA allocates lower power to users with good channel realizations and higher power to users with poor channel realizations, thus, imposing an additional fairness constraint.

Let us consider single-cell DL NOMA transmission with a single BS and two users, denoted as DL1 and DL2, respectively. Assume that DL2 is located close to the cell-edge. The BS simply transmits a waveform which is a sum of both users' signals. According to the NOMA principle, transmit power of the information signal for DL2 must be larger than that of DL1. Since DL2 experiences a poor channel condition, interference from DL1's superimposed signal will not significantly

M. Mohammadi is with the Faculty of Engineering, Shahrekord University, Shahrekord 115, Iran (email: m.a.mohammadi@sku.ac.ir).

B. K. Chalise is with the Department of Electrical and Computer Engineering, New York Institute of Technology, Old Westbury, New York, USA (email: batu.k.chalise@ieee.org).

Z. Ding is with School of Electrical and Electronic Engineering, the University of Manchester, Manchester, the UK (email: zhiguo.ding@manchester.ac.uk).

H. A. Suraweera is with the Department of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya 20400, Sri Lanka (email: himal@ee.pdn.ac.lk).

C. Zhong is with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (email: caijunzhong@zju.edu.cn).

X. Shi and J. S. Thompson are with the Institute for Digital Communications, School of Engineering, University of Edinburgh, United Kingdom (email: Xiaoyan.Shi, John.Thompson@ed.ac.uk).

affect DL2. Accordingly, DL2 directly decodes its information by assuming the interference due to DL1's signal as noise. In contrast, DL1 can decode its own information after removing DL2's signal using a successive interference cancellation (SIC) receiver. Therefore, the system throughput can be significantly improved compared to OMA schemes.

B. Review of Full-duplex NOMA Systems

Two cases of the NOMA concept benefiting from full-duplex operation are i) when a full-duplex BS serves UL and DL users at the same time in the same frequency [7], and ii) cooperative NOMA systems where full-duplex NOMA-strong users [8], [9] or dedicated full-duplex relays [10] assist the transmissions between the source and NOMA-weak users. In [7], a full-duplex resource allocation method is proposed for multicarrier transmissions using NOMA, where the BS simultaneously serves multiple DL and UL users. The corresponding approach for maximizing the weighted sum throughput is formulated as a non-convex optimization: the resulting optimal subcarrier and power allocation policies have been found in [7].

In the literature, cooperative schemes that use relays have been extensively studied as an effective method for utilizing spatial diversity and increasing coverage. Hence, cooperative NOMA is a natural extension of traditional relaying systems that can take advantage of the reduced attenuation between a relay node and NOMA-weak users [8]–[11]. In [8], a full-duplex device-to-device cooperative NOMA scheme, where the NOMA-strong user operates in a full-duplex mode, has been proposed. Moreover, an adaptive multiple access scheme, which adaptively switches among cooperative NOMA, conventional NOMA, and OMA schemes, according to the residual SI level and link quality are described in [8]. It has been shown in [9], that if available the direct link can be exploited to increase the diversity gain and thereby to improve the performance of the weak user in cooperative full-duplex NOMA systems. The work in [10], has considered a dual-user NOMA system with a full-duplex relay which assists information transmission to the weak user with an unfavorable channel. Methods to determine possible data rates of a full-duplex cooperative NOMA system has been reported in [11].

Current work on full-duplex NOMA shows that the achievable gains are sensitive to the residual SI strength [8]–[10]. In practice, residual SI brings new challenges in making a SIC receiver feasible. Each full-duplex receiver needs to cancel its own SI first, and then proceed to decode the inter-user interference signals and finally recover its own signal. Therefore, the accuracy of the adopted SI canceler will play a pivotal role in SIC. The elevated interference at the user terminals in full-duplex scenarios will make it difficult for a receiver to cancel the strong signal before decoding its own information. Hence, to fully exploit full-duplex NOMA, advanced interference cancellation techniques, such as coordinated multi-point or interference alignment coupled with sophisticated error correction coding schemes are needed.

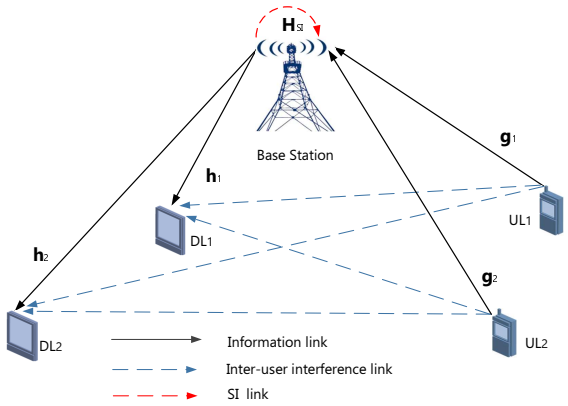


Fig. 1: Full-duplex NOMA system with two DL users (DL1 and DL2) and two UL users (UL1 and UL2).

III. Uplink and Downlink Full-duplex NOMA

In legacy cellular networks, number of users supported simultaneously is limited due to orthogonal resource allocation. NOMA can accommodate multiple users simultaneously via non-orthogonal resource allocation. Initial implementations of NOMA in cellular networks demonstrate that it achieves a superior spectral efficiency as compared to OMA. One way to achieve higher spectral efficiencies in NOMA-based cellular networks is to use full-duplex transmission at the BS. However, the benefits of full-duplex communication does not come without a cost. As an illustrative example, Fig. 1 shows a simple dual-user system where DL1 (UL1) and DL2 (UL2) represent a NOMA-strong DL (UL) user and a NOMA-weak DL (UL) user, respectively. We see that due to full-duplex operation, UL transmissions suffer from SI at the BS while DL transmissions are degraded by the inter-user interference.

The application of multiple antennas in NOMA provides additional performance improvement. An UL/DL multi-antenna NOMA set up which can deliver performance gains in terms of the reception reliability was proposed in [6]. A full-duplex BS could utilize its multiple antennas for spatial SI cancellation and consequently for improving the signal-to-interference-plus-noise ratio (SINR) of both DL and UL transmissions. In Fig. 2 we present the ergodic sum capacity of a UL/DL full-duplex system as applicable in single-cell and multi-cell scenarios. We assume that each cell on the hexagonal grid has six interfering cells in the surrounding tier. Moreover, each cell was populated with two DL NOMA users and two UL NOMA users uniformly distributed in the cell. Simulation parameters are based on 3GPP LTE-A small cell scenarios. The transmit beamformer is designed using the maximum ratio transmission (MRT) principle, while the receive beamformer at the BS uses zero-forcing (ZF). In the single-cell environment, half-duplex system outperforms the full-duplex counterpart with the maximum ratio combining (MRC)/MRT scheme at high signal-to-noise ratio (SNR). In the multi-cell scenario, full-duplex operation has a superior performance over the entire SNR regime. Full-duplex NOMA system achieves 75% higher average sum rate than its half-duplex counterpart at a SNR of 20 dB. In a multi-cell environment, co-channel

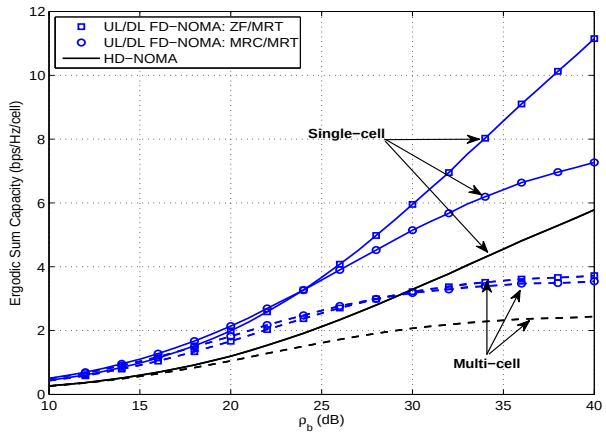
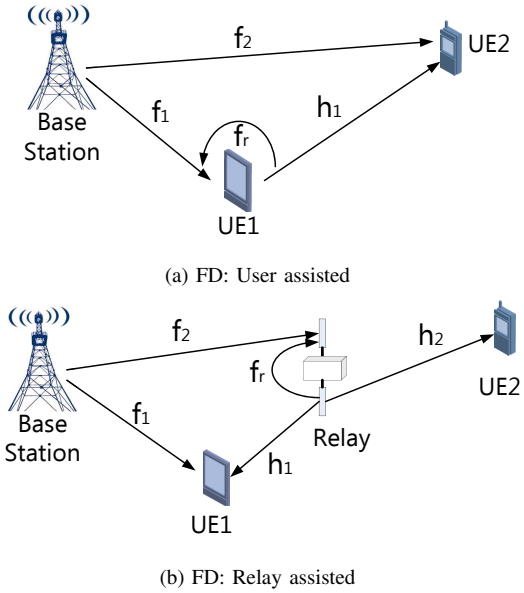


Fig. 2: Ergodic capacity of UL/DL full-duplex and half-duplex NOMA systems for a three transmit/two receive antenna configuration at the BS and for $\sigma^2_{\mathsf{SI}} = -10$ dBW.



(a) FD: User assisted

(b) FD: Relay assisted

Fig. 3: System model: Dual-user cooperative NOMA.

interference will determine the performance gap between the full-duplex and half-duplex modes. This interference can be reduced significantly by exploiting interference coordination.

IV. Cooperative Full-duplex NOMA

Cooperative communications can be used to enlarge the radio coverage and improve the communication reliability, using low-complexity signal processing techniques. In the literature, there exists two different kinds of cooperative NOMA systems, namely, user-assisted cooperative NOMA (UC-NOMA) [12] and relay-assisted cooperative NOMA (RC-NOMA) [13]. In UC-NOMA, the NOMA-strong user assists the NOMA-weak user since it can decode the both users information, while in RC-NOMA, a dedicated relay is deployed to help the NOMA-weak user.

While cooperation improves the reliability of NOMA systems, implementation of cooperation among users or relays incurs additional bandwidth. Since one major motivation of NOMA is to enhance the spectral efficiency, reduction of spectral efficiency due to cooperation is undesirable. To tackle this issue, we propose to empower the cooperative node with full-duplex operation as illustrated in Fig. 3(a) and 3(b), where the NOMA-strong user or the relay adopts full-duplex transmission. In half-duplex relays, two time slots are required to forward data to a relay and then to the destination. However, full-duplex cooperative systems eliminate the extra time slot required for cooperation. Despite the promise, adoption of full-duplex may significantly elevate the interference level. For example, consider the full-duplex RC-NOMA. In addition to the SI at the full-duplex relay, the NOMA-strong user will also receive harmful interference.

To illustrate the practical gain of cooperative NOMA due to full-duplex mechanism, let us focus on the model depicted in Fig. 3(d), where the BS transmits the superimposed signal to the user equipment UE1 and the relay, while the relay not only overhears the superimposed signal, but also simultaneously transmits the previously received signal intended for UE2. Since both UE1 and the relay are aware of the signal intended for UE2, known interference cancellation techniques can be applied. However, due to imperfect cancellation, residual interference is likely to exist. Let k_1 and k_2 denote the residual interference power level at UE1 and the relay, respectively. According to Fig. 4, compared to the half-duplex RC-NOMA system, the ergodic sum capacity of full-duplex RC-NOMA system is higher in the low and moderate SNR region. Interestingly, the difference in the strength of residual SI does not have a significant impact on the sum capacity gain. However, when the SNR increases, the ergodic sum capacity of the full-duplex RC-NOMA system becomes worse, mainly due to high interference level, which indicates the critical importance of SI mitigation.

V. NOMA in Full-duplex Cognitive Radio Networks

NOMA can be considered as a particular case of cognitive radio networks (CRNs), where a user with a strong channel condition exploits the spectrum occupied by a user with a poor channel condition. The strong-user and weak-user can be, respectively, viewed as a primary user and cognitive user in the CRN. Therefore, according to the principle of CRNs, the transmit power of the strong-user is constrained by the weak-user's SINR. Following this concept, a variation of NOMA termed as cognitive radio inspired NOMA has been proposed in [5].

The NOMA concept is applicable for underlay CRNs, where a cognitive source (CS) communicates with two or more cognitive users (CUs). Cognitive radio terminals are permitted to access the frequency bands of the primary network, provided they satisfy the predefined threshold of interference received at primary users. Given this interference constraint, the performance of the cognitive network is strictly limited. Cooperative techniques help to reduce the transmission power as the communication distances in cognitive networks are decreased by using a cognitive relay (CR). Hence, the mutual

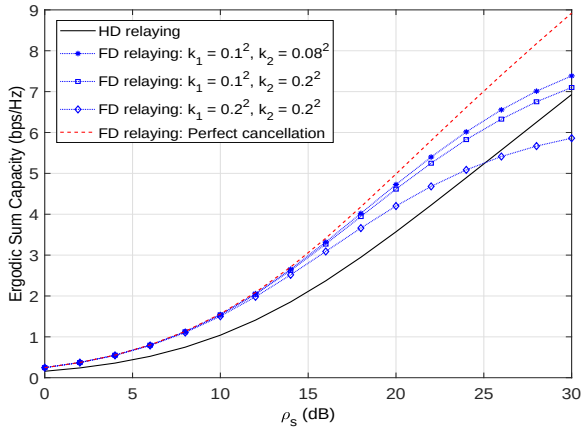


Fig. 4: Ergodic capacity of RC-NOMA systems with relay transmit power $\rho_r = \rho_b/2$, power allocation coefficients $a_1 = 0.05$ and $a_2 = 0.95$, channel gains for f_1 being $\lambda_{b1} = 1$, for f_2, h_1 and h_2 being $\lambda_{br} = \lambda_{r1} = \lambda_{r2} = 0.5$, while for f_r being $\lambda_{rr} = 0.3$.

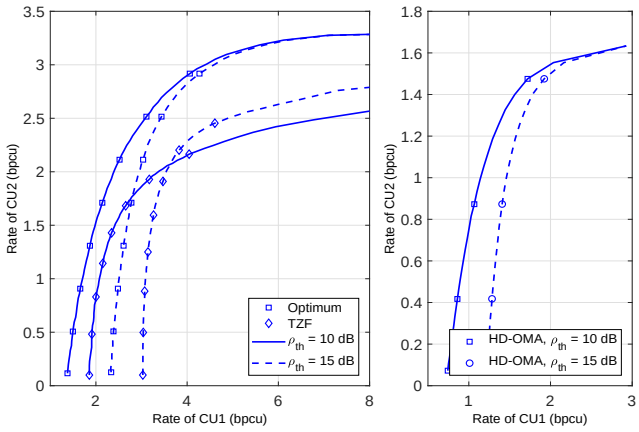


Fig. 5: Rate region achieved by the optimum and the sub-optimum schemes for full-duplex and half-duplex CR-NOMA network.

interference between primary and cognitive networks can be managed effectively. If a half-duplex CR is adopted, a loss of spectral efficiency should be expected and it can be mitigated through utilizing of full-duplex radio at the CR. On the other hand, however, adopting the full-duplex relay into the CRNs can result in several problems. In particular, the primary receiver will simultaneously receive interference from the CS and CR. Therefore, the transmission powers at the CS and CR must be lower than those in the half-duplex case to comply with the interference constraint.

Moreover, signal reception at the CR and NOMA-strong user must deal with SI and co-channel interference, respectively. To deal with these challenges, we propose to employ attractive techniques, such as joint power allocation between the CS and CR and/or CR beamforming design (in case of MIMO-CR). Fig. 5 shows the achievable rate region of a full-duplex relay-assisted CRN with one primary transmitter-receiver pair for different levels of the predetermined maximum tolerable interference level, ρ_{th}, at the primary receiver, where the CR is equipped with $\mathsf{N_T} = 3$ transmit antennas and $\mathsf{N_R} = 2$ receive antennas, while the CS, NOMA-strong user (CU1), NOMA-weak user (CU2) and primary transmitter-receiver are single antenna terminals. NOMA power allocation coefficients at the CS are $a_1 = 0.1$, $a_2 = 0.9$. With the proposed optimum scheme, CR's receive and transmit beamformers and the transmit power of CS and CR are jointly optimized so that the CU1's capacity is maximized by guaranteeing that the CU2's capacity is above a certain value. With suboptimum design, transmit/receive beamformers at the CR are designed using the MRC and ZF principles (TZF), respectively, and the transmit power levels at CS and CR are optimized. As can be observed, the achievable rate of the CU1 and the CU2 increase significantly compared to the half-duplex CRN, particularly when the optimum scheme is used.

Full-duplex operation also offers the potential to achieve simultaneous sensing and transmission in cognitive radio NOMA networks. Specifically, a full-duplex cognitive transmitter (CS or CR) is able to dynamically sense the spectrum band and determine if the primary users are busy or idle, and at the same time decide to send data or keep silent. One possible approach to realize online spectrum sensing in CRNs is to use multiple antennas at the full-duplex cognitive transmitter, where some ($\mathsf{N_S}$) antennas are allocated for sensing, some ($\mathsf{N_T}$) antennas for transmitting data, and some ($\mathsf{N_R}$) antennas for receiving data. This approach, however, requires design of new signal processing techniques, resource allocation algorithms, and antenna selection rules. To this end, new algorithms require to jointly take into account different quality-of-service levels for NOMA users.

VI. Resource Management in Full-duplex NOMA Networks

Resource management is important for improving the system performance and at the same time to guarantee the quality-of-service (QoS) requirements. With full-duplex functionality applied to the BS, resource management strategies for DL and UL transmissions are closely related in a full-duplex system. NOMA further complicates the situation by enabling multiplexing in the power domain. In this section, we provide several insights into how the resource management problem in full-duplex NOMA networks can be addressed.

First, instead of merely considering time, frequency, and space as the basic elements for resource optimization, for NOMA transmission, it is also important to investigate whether it is beneficial for multiple users to share one 'resource block'. The two extra challenges related to FD-operations (as already mentioned in Section III) are the SI at the BS and the co-channel interference (CCI) between UL and DL users. From the perspective of resource management, the user scheduling policy must consider the UL-DL CCI so that the DL user is not significantly interfered with the UL signals and a good balance between DL and UL performance can be achieved. Fortunately, since the BSs usually transmit DL signals with stronger power than the UL users, UL-DL interference can be commonly treated as extra noise by the DL receivers

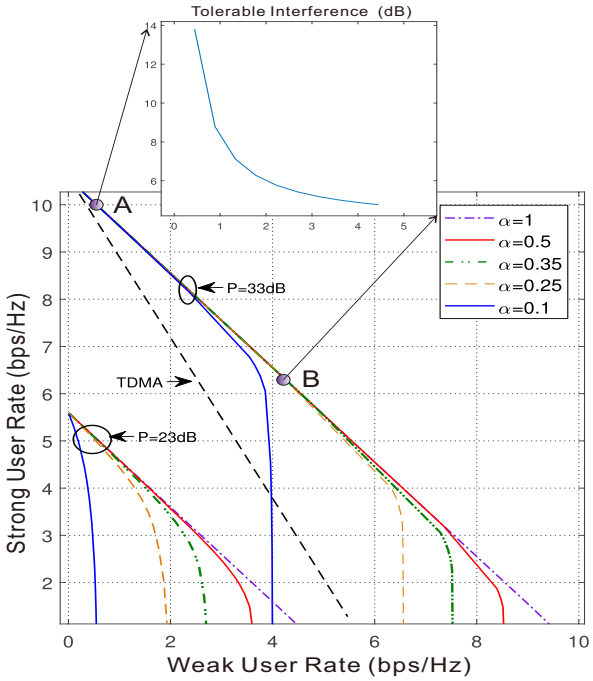


Fig. 6: The Two-user rate region for superposition coding with beamforming.

[7]. It is unlikely to be beneficial if we eliminate the UL-DL interference at the DL receiver end through SIC. As for residual SI at the BS, it causes degradation of the UL signal SNR and is typically modeled as extra Gaussian noise. The reduction of residual SI is achieved through various technologies such as massive MIMO [14] that suppress signal leakage from the BS transmitter to its receiver end. After absorbing the residual SI into the receiver noise, dynamic multiple access schemes that have already been studied in half-duplex systems can be applied. Usually these types of methods adapt to the current levels of receiver noise and link conditions by switching between NOMA and OMA modes [8]. It then leaves us only the UL-DL CCI, which is more complicated, to consider.

To decouple the challenges brought by NOMA and full-duplex operation for resource management, let us first consider the impact of NOMA on DL transmission. It is temporarily assumed that user selection for DL and UL transmissions are independent of each other and the UL-DL CCI is treated as extra noise at the DL users. We further consider that the BS, equipped with multiple antennas, is communicating with two DL single-antenna users in a NOMA fashion. In [15], such a beamforming transmission is denoted as SCBF (beamforming with superposition coding). Specifically, in [15], a novel algorithm has been proposed to compute the two-user rate region for the SCBF method. A rate region characterizes how the transmission rates can be traded off between the two users given a fixed total transmission power P. As can be seen in Fig. 6, for lower rates of the weak user, the two user rate curve is similar to that for TDMA operating in symmetric channels. As the rate of the weaker user increases, there is then a rapid deterioration of the strong user's rate.

Fig. 6 also shows how the two-user rate region changes according to the transmit power level (i.e. different values of P) and to the degree of channel asymmetry (i.e. different values of α, which is the attenuation factor of the weaker channel with respect to the stronger one). The comparison between SCBF and time-division multiplexing (TDM) becomes straightforward, as the rate region achieved by TDMA is simply the straight line from the y-axis intercept to the x-axis intercept for the SCBF rate region curve.

We then discuss how the user scheduling for the UL users operating in the same band as the DL users would affect the DL NOMA transmission just characterized by the rate region in Fig. 6. We highlight that the UL-DL CCI can actually be translated into an additional source of channel asymmetry. Suppose that the UL interferer is close to the weaker user which is usually further from the BS than the stronger user, and therefore the stronger user receives little interference from the UL interferer. The extra interference at the weaker user means smaller α value or a higher degree of channel asymmetry. Furthermore from Fig. 6, we see that for fixed total transmission power P, rate points belonging to a range of channel asymmetry would overlap at the low rate part of the weak user, such as points between A and B in Fig. 6. For example, when the operational point lies in Point B, as shown in Fig. 6, increasing the channel asymmetry, or equivalently introducing more UL-DL CCI to the weaker user, will not degrade the performance as long as α is above 0.25. The sub-figure in 6 plots the maximum interference that can be tolerated by the weak user at the rate points on line segment AB. This interesting property suggests that NOMA provides great robustness to the UL-DL CCI when the UL users are close to the weak DL user. The above results can be quite useful for lowering the complexity for the design of a user scheduling protocol. Notice that the complexity of the link scheduling problem usually grows rapidly with the number of users since the problem is of a combinatorial nature.

Our analysis has so far assumed only two DL users. For more general cases, user grouping strategy combined with orthogonal multiplexing methods, such as ZF and TDM can be used to strike a good balance between performance and computational complexity.

VII. Future Research Challenges

We now outline some interesting research challenges and future directions for full-duplex NOMA systems.

Interference management for full-duplex NOMA systems: Full-duplex operation creates a number of simultaneous transmissions in multi-cell settings which can cause elevated interference levels in wireless systems. Such interference will adversely effect a NOMA user's ability to successfully decode messages. Therefore, when full-duplex NOMA systems are deployed, interference must be carefully studied so that effective countermeasures can be designed. To this end, power control schemes, scheduling, and error control coding schemes should be investigated in detail. Moreover, employment of inter/intra-cell interference suppression techniques and approaches such as interference alignment with low complexity are promising as worthwhile directions to pursue.

Low complexity multi-antenna schemes: There has been significant work on low complexity MIMO systems. To this end, antenna selection is a popular technique, but it may not be directly applied because full-duplex NOMA systems must account for both SI as well as user link conditions to satisfy the NOMA constraints. Hence designing new antenna selection schemes for full-duplex NOMA systems that balance the performance against implementation complexity is a promising research direction.

User pairing in full-duplex settings: The performance of NOMA is dependent on which users are selected to pair. Most existing user pairing methods in NOMA systems have been proposed to support two users. To exploit the benefit of the full-duplex operation in NOMA systems, however, general user pairing algorithms which go beyond the dual-user pairing case should be developed.

Massive machine-to-machine (M2M) communication access and green communications: NOMA is a promising approach to implement massive access in emerging M2M communication networks and green communications. Devices in such networks suffer from energy constraint issues. To enable perpetual operation of M2M devices, energy harvesting techniques and wireless power transfer offer attractive solutions. Adoption of such techniques in full-duplex NOMA systems require sophisticated power and time-split optimization solutions as compared to OMA transmissions. Further, in addition to transmit power, SIC will increase the node level power consumption while additional interference would be created with full-duplex transmissions. Therefore, energy efficient node designs possible with multiple antenna cases and new architectures such as cloud RAN are essential to reap the benefits of full-duplex NOMA systems.

Security and ultra-reliable low latency communications: NOMA suffers from the several security risks; however, it can facilitate the implementation of physical layer security techniques. Use of NOMA ensures that the users messages are mixed; hence, certain common messages, together with full-duplex operation, can be used as jamming signals against eavesdropping. Additionally, the combination of NOMA and full-duplex show promise for ultra-reliable low latency communications. It is worthwhile to study and apply various new tools such machine learning, matching theory and deep learning which are useful to learn unknown channel conditions so that automatic encoding/decoding schemes can be deployed with low latency.

VIII. Conclusion

In this article, the concept of NOMA with full-duplex operation was discussed. We first reviewed the state-of-the-art and discussed UL and DL transmission, cooperative relay and cognitive radio with full-duplex NOMA operation. Further, the design of non-orthogonal beamforming with superposition coding was discussed as a natural way of extending NOMA to MIMO communications. Key research challenges and potential solutions that would hasten the practical deployment of full-duplex NOMA systems were also identified.

References

[1] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1637–1652, Sept. 2014.

[2] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, pp. 5983–5993, Dec. 2011.

[3] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 4296–4307, Dec. 2012.

[4] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC'13)*, Dresden, Germany, June 2013, pp. 1–5.

[5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, pp. 2181–2195, Oct. 2017.

[6] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 4438–4454, June 2016.

[7] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, pp. 1077–1091, Mar. 2017.

[8] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device aided cooperative non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, pp. 4467–4471, May 2017.

[9] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Exploiting full/half-duplex user relaying in NOMA systems," *IEEE Trans. Commun.*, vol. 66, pp. 560–575, Feb. 2018.

[10] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Commun. Lett.*, vol. 20, pp. 2478 –2481, Dec. 2016.

[11] J. So and Y. Sung, "Improving non-orthogonal multiple access by forming relaying broadcast channels," *IEEE Commun. Lett.*, vol. 20, pp. 1816–1819, Sept. 2016.

[12] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, pp. 1462–1465, Aug. 2015.

[13] J. Kim and I. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Commun. Lett.*, vol. 19, pp. 2037–2040, Nov. 2015.

[14] A. Yadav and O. A. Dobre, "All technologies work together for good: A glance at future mobile networks," *IEEE Wireless Commun. Mag.*, vol. 25, pp. 10–16, Aug. 2018.

[15] X. Shi, J. S. Thompson, M. Safari, and R. Liu, "Beamforming with superposition coding in multiple antenna satellite communications," in *Proc. IEEE Intl. Conf. Commun. (ICC'17), (Workshop on Satellite Communications)*, Paris, France, May 2017, pp. 705 –710.

Mohammadali Mohammadi [M] (m.a.mohammadi@sku.ac.ir) received the B.S. degree from Isfahan University of Technology in 2005, and the M.S. and Ph.D. degrees from K. N. Toosi University of Technology in 2007 and 2012. He was a visiting researcher at the Australian National University between 2010 and 2011. Currently, he is an assistant professor in the Faculty of Engineering, Shahrekord University. His current research interests include cooperative communications, energy harvesting communications, full-duplex communications and stochastic geometry.

Xiaoyan Shi [S] (Xiaoyan.Shi@ed.ac.uk) received the B.Sc. degree in electrical and information engineering from Beihang University in 2014. He is currently on the joint Ph.D. program between Beihang University the University of Edinburgh. His research directions mainly include beamforming algorithms for MIMO transmission, satellite communications and traffic scheduling algorithm designs for the radio access network.

Batu K. Chalise [SM] (batu.k.chalise@ieee.org) received the M.S. and Ph.D. degrees in electrical engineering from the University of Duisburg-Essen, Germany. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, New York Institute of Technology, NY. His research interests include signal processing for wireless and radar communications, wireless sensor networks, smart systems, and machine learning.

Zhiguo Ding [SM] (zhiguo.ding@manchester.ac.uk) received his B.Eng from BUPT in 2000, and his Ph.D degree from Imperial College in 2005. Currently, he is a professor at the University of Manchester. He has been serving as an Editor for IEEE TCOM and IEEE TVT, and served as an editor for IEEE WCL and CL. He received the best paper award in IET-CWMC-2009 and IEEE-WCSP-2015, the EU Marie-Curie Fellowship 2012-2014, IEEE TVT Top Editor 2017, IEEE Heinrich Hertz Award 2018, Jack Neubauer Memorial Award 2018, and Best Signal Processing Letter Award 2018.

Himal A. Suraweera [SM] (himal@ee.pdn.ac.lk) received his Ph.D. degree from Monash University, Australia, in 2007. Currently he is a Senior Lecturer at the University of Peradeniya, Sri Lanka. He has published more than 100 papers in the areas of wireless communications. His research interests include cooperative relay networks, massive MIMO, full-duplex radios and green communications. He serves as an Editor for several journals including IEEE Transactions on Wireless Communications.

Caijun Zhong [SM] (caijunzhong@zju.edu.cn) received the Ph.D. degree in Telecommunications in 2010 from University College London, United Kingdom. Since September 2011, he has been with Zhejiang University, China, where he is currently an associate professor. His research interests include massive MIMO systems, wireless power transfer, NOMA, and backscatter communication. Dr. Zhong is an Editor of the IEEE Transactions on Wireless Communications and IEEE Communications Letters.

John S. Thompson [F] (john.thompson@ed.ac.uk) is currently a Professor at the School of Engineering in the University of Edinburgh. He specializes in antenna array processing, cooperative communications systems and energy efficient wireless communications. He has published in excess of three hundred papers on these topics. In 2018 he was a technical programme co-chair for the IEEE Smartgridcomm conference in Aalborg, Denmark. In 2015-2018, he has been recognised by Thomson Reuters as a highly cited researcher.