DOI: 10.3969/j. issn. 1673-5188. 2017. S1. 003
http://kns.cnki.net/kcms/detail/34.1294.TN.20170606.1856.004.html, published online June 6, 2017

An Overview of Non-Orthogonal Multiple Access

Anass Benjebbour
(NTT DOCOMO, INC., Kanagawa 239-8536, Japan)

Abstract

In recent years, non-orthogonal multiple access (NOMA) has attracted a lot of attention as a novel and promising power-domain user multiplexing scheme for Long-Term Evolution (LTE) enhancement and 5G. NOMA is able to contribute to the improvement of the tradeoff between system capacity and user fairness (i.e., cell-edge user experience). This improvement becomes in particular emphasized in a cellular system where the channel conditions vary significantly among users due to the near-far effect. In this article, we provide an overview of the concept, design and performance of NOMA. In addition, we review the potential benefits and issues of NOMA over orthogonal multiple access (OMA) such as orthogonal frequency division multiple access (OFDMA) adopted by LTE, and the status of 3GPP standardization related to NOMA.

Keywords

multiple access; non-orthogonal multiple access (NOMA); power-domain; multi-user detection; MUST

1 Introduction

Significant gains in system capacity and quality of user experience (QoE) are required to respond to the anticipated exponential increase in the volume of mobile traffic in the next decade and the merge of enhanced mobile broadband (eMBB) services [1]. In cellular mobile communications, the design of the radio access technology (RAT) is one important aspect for improving system capacity in a cost-effective manner. Radio access technologies are typically characterized by the radio frame design, waveform design, multiple - input and multiple - output (MIMO) transmission scheme, and multiple access scheme. In particular, the design of the multiple access scheme is of great importance from a system perspective, since it provides the means for multiple users to access and share the system resources efficiently and simultaneously, e.g., frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency division multiple access (OFDMA). In the 3.9G and 4G mobile communication systems such as Long-Term Evolution (LTE) and LTE-Advanced [2], standardized by the 3rd Generation Partnership Project (3GPP), orthogonal multiple access (OMA) based on OFDMA for downlink and single carrier (SC)-FDMA for uplink are adopted. Orthogonal multiple access is a good choice for achieving good system-level throughput performance in packet-domain services with a simplified receiver design. However, non-orthogonal designs become of interest toward further enhancement of the system efficiency and QoE especially at the cell edge.

Recently, there have been several investigations on advanced schemes for non-orthogonal signal transmission within a user and non-orthogonal user multiplexing among multiple users. For example, Faster-than-Nyquist (FTN) signaling [3] is one approach for non-orthogonal signal transmission within a user by exploiting the excess bandwidth of the signal. Interleaved division multiple access (IDMA), where the channelization of respective user is achieved by the user-specific channel interleaver and multiuser detection at the receiver, is investigated to accommodate a large number of low-rate users [4], [5]. However, these schemes do not exploit the channel difference among users and generally require high complexity receivers for signal separation.

As a novel multiple access approach, a non-orthogonal multiple access (NOMA) scheme was proposed by NTT DOCOMO [6]–[18] including the author of this article. In the proposed NOMA, multiple users of different channel conditions are multiplexed in the power-domain on the transmitter side and multiuser signal separation on the receiver side is conducted. From an information-theoretic perspective, it is well-known that by using superposition coding at the transmitter and successive interference cancellation (SIC) at the receiver, non-orthogonal user multiplexing not only outperforms orthogonal multiplexing, but also can achieve the capacity region of the downlink broadcast channel [12], [19], [20]. NOMA can be also applied to the uplink (multiple access channel) [12], [15], [19]. For uplink, although both NOMA and OMA can achieve the capacity region, NOMA also provides improvements in the tradeoff of system capacity and user fairness (i.e., the cell-edge user experienced data rate) [12], [19].

Assuming a proportional fairness (PF) scheduler, the performance of NOMA has been heavily investigated for downlink and uplink from the system-level and link-level perspectives [6]–[18]. In addition, the transmitter and receiver designs for NOMA were considered for both closed-loop and open-loop MIMO and for both successive interference cancellation and non-SIC receivers [16], [17]. When applied to either downlink or uplink, NOMA is shown able to contribute to the improvement of the tradeoff between system capacity and user fairness. This improvement becomes in particular emphasized in a cellular system where the channel conditions vary significantly among users due to the near-far effect.

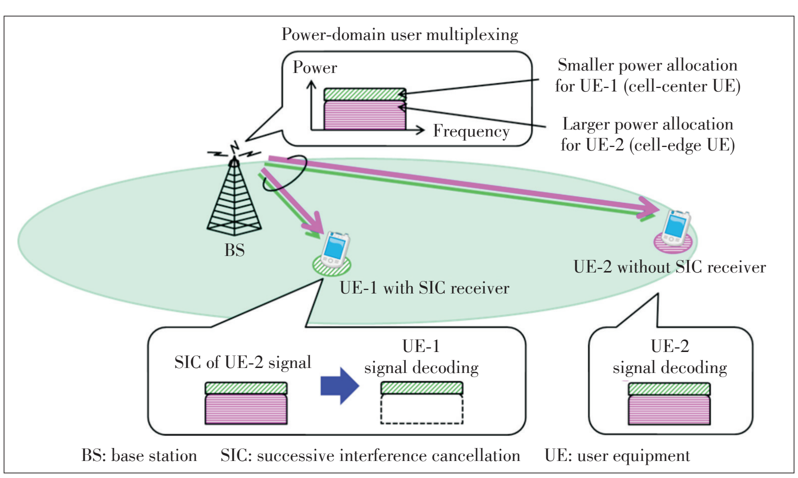
In this article, we introduce an overview of the concept, the design with a combination of MIMO and the performance of NOMA. We also review the potential benefits and issues of NOMA over orthogonal multiple access, and the status of standardization related to downlink NOMA, known as multi-user superposition transmission (MUST) in 3GPP.

The rest of this paper is organized as follows. Section 2 describes the concept. Section 3 discusses the expected benefits and issues of NOMA. Section 4 explains the combination of NOMA with MIMO. Section 5 reviews the performance of NOMA from link-level and system-level evaluations and trial results. In addition, the status of NOMA standardization in 3GPP LTE Release 14 is briefly summarized. Finally, Section 6 concludes the paper.

2 NOMA Concept

2.1 Downlink

Fig. 1 illustrates downlink NOMA with SIC for the case of one base station (BS) and two user equipments (UEs).



▲**Figure 1. Illustration of downlink NOMA with SIC.**

For the sake of simplicity, we assume in the following descriptions the case of single transmit and receive antennas. The overall system transmission bandwidth is assumed to be 1 Hz. The base station transmits a signal for UE-i (i = 1, 2), x_i, where $E[|x_i|^2] = 1$, with transmit power P_i and the sum of P_i is equal to P. In NOMA, x_1 and x_2 are superposed in the power-domain as follows:

$$x = \sqrt{P_1}x_1 + \sqrt{P_2}x_2 . \quad (1)$$

Thus, the received signal at UE-i is represented as

$$y_i = h_i x + w_i , \quad (2)$$

where h_i is the complex channel coefficient between UE-i and the BS. The variable w_i denotes additive white Gaussian noise (AWGN) including inter-cell interference. The power spectral density of w_i is $N_{0,i}$. In downlink NOMA, the SIC process is implemented at the UE receiver for the case where the decoding of the signal of desired UE and that of the superposed signals of other UEs are needed. The optimal order for SIC decoding is in the order of decreasing channel gain normalized by noise and inter-cell interference power, $|h_i|^2/N_{0,i}$ (called as simply channel gain in the following). Given this decoding order and assuming that any user can correctly decode the signals of other users whose decoding order comes before the corresponding user, each UE-i can remove the inter-user interference from the j-th user whose $|h_j|^2/N_{0,j}$ is lower than $|h_i|^2/N_{0,i}$. In a 2-UE case, assuming that $|h_1|^2/N_{0,1} > |h_2|^2/N_{0,2}$, UE-2 does not perform interference cancellation since it comes first in the decoding order. UE-1 first decodes x_2 and subtracts its component from the received signal y_1, then next, x_1 is decoded without interference from x_2. Assuming successful decoding and no error propagation, the throughput of UE-i, R_i, can be represented as

$$R_1 = \log_2\left(1 + \frac{P_1|h_1|^2}{N_{0,1}}\right), R_2 = \log_2\left(1 + \frac{P_2|h_2|^2}{P_1|h_2|^2 + N_{0,2}}\right). \quad (3)$$

From (3), it can be seen that power allocation for each UE greatly affects the user throughput performance and thus the modulation and coding scheme (MCS) used for data transmission of each UE. By adjusting the power allocation ratio, P_1/P_2, the BS can flexibly control the throughput of each UE and also optimize tradeoff between the system capacity and user fairness. By flexibly adjusting power allocation, the BS can control the throughput of each UE such that the signal designated to each UE is decodable at its corresponding receiver. Also, since the channel gain of the cell-center UE is higher than cell-edge UE, as long as the cell-edge UE signal is decodable at cell-edge UE receiver, its decoding at the cell-center UE receiver can be successful with high probability.

For OMA as orthogonal user multiplexing, the bandwidth of α Hz (0 < α < 1) is assigned to UE-1 and the remaining bandwidth, 1—α Hz, is assigned to UE-2. The throughput of UE-i, R_i, is represented as

$$R_1 = \alpha \log_2\left(1 + \frac{P_1|h_1|^2}{\alpha N_{0,1}}\right), R_2 = (1-\alpha)\log_2\left(1 + \frac{P_2|h_2|^2}{(1-\alpha)N_{0,2}}\right). \quad (4)$$

In NOMA, the performance gain compared to OMA increases when the difference in channel gains, e.g., path loss between UEs, is large. For example, as shown in **Fig. 2**, we assume a 2-UE case with a cell-interior UE and a cell-edge UE, where $|h_1|^2/N_{0,1}$ and $|h_2|^2/N_{0,2}$ are set to 20 dB and 0 dB, respectively. For OMA with equal bandwidth and equal transmission power are allocated to each UE (a = 0.5, $P_1 = P_2 = 1/2P$), the user rates are calculated according to (4) as R_1 = 3.33 bps and R_2 = 0.50 bps, respectively. On the other hand, for NOMA, when the power allocation is conducted as $P_1 = 1/5P$ and $P_2 = 4/5P$, the user rates are calculated according to (3) as R_1 = 4.39 bps and R_2 = 0.74 bps, respectively. The corresponding gains of NOMA over OMA are 32% and 48% for UE-1 and UE-2, respectively. According to this example where a 20 dB signal-to-noise ratio (SNR) difference between the 2 UEs is assumed, it can be seen that NOMA provides a higher sum rate than OMA.

2.2 Uplink

Fig. 3 illustrates uplink NOMA where two UEs transmit signals to the BS on the same frequency resource and at the same time, and SIC is conducted at BS for UE multi-user signal separation. Similar to downlink, we assume the case of single transmit and receive antennas, and the overall system transmission bandwidth is 1 Hz. The signal transmitted by UE-i (i = 1, 2) is denoted as x_i, where $E[|x_i|^2] = 1$, with transmit power P_i. In uplink NOMA, the received signal at BS is a superposed signal of x_1 and x_2 as follows:

$$y = h_1\sqrt{P_1}x_1 + h_2\sqrt{P_2}x_2 + w, \quad (5)$$

where h_i denotes the complex channel coefficient between UE-i and the BS. The variable w denotes inter-cell interference and noise observed at the BS with a power spectral density of N_0. We assume UE-1 is the cell-center user and UE-2 is the cell-edge user, i.e. $|h_1|^2/N_0 > |h_2|^2/N_0$, and the BS conducts SIC according to the descending order of channel gains. The throughput of UE-i, denoted as R_i, assuming no error propagation can be calculated as

$$R_1 = \log_2\left(1 + \frac{P_1|h_1|^2}{P_2|h_2|^2 + N_0}\right), R_2 = \log_2\left(1 + \frac{P_2|h_2|^2}{N_0}\right). \quad (6)$$

If the BS conducts SIC according to the ascending order of channel gains, the throughput of UE-i can be calculated as

$$R_1 = \log_2\left(1 + \frac{P_1|h_1|^2}{N_0}\right), R_2 = \log_2\left(1 + \frac{P_2|h_2|^2}{P_1|h_1|^2 + N_0}\right). \quad (7)$$

Interestingly, the total UE throughput is the same regardless of the SIC order of descending order or ascending order of channel gain, i.e.

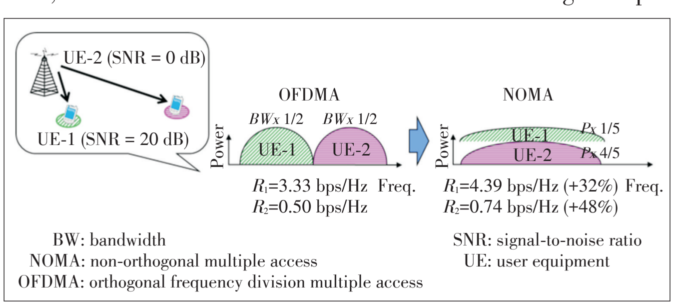
$$R_1 + R_2 = \log_2\left(1 + \frac{P_1|h_1|^2 + P_2|h_2|^2}{N_0}\right). \quad (8)$$

However, the conclusion that the total UE throughputs of different SIC orders are equal only holds under the assumption of no error propagation. In practical systems where we have error propagation, the best SIC order is in the decreasing order of channel gains.

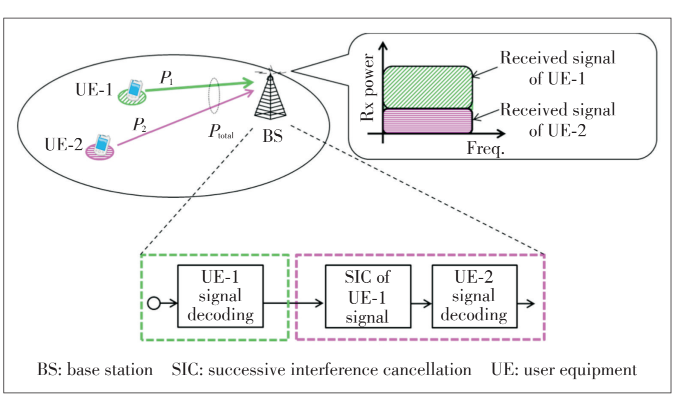
For OMA, we assume the bandwidth of α Hz (0 < α < 1) is assigned to UE-1 and the remaining bandwidth, 1—α Hz, is assigned to UE-2. The throughput of UE-i can be calculated as

$$R_1 = \alpha \log_2\left(1 + \frac{P_1|h_1|^2}{\alpha N_0}\right), R_2 = (1-\alpha)\log_2\left(1 + \frac{P_2|h_2|^2}{(1-\alpha)N_0}\right). \quad (9)$$

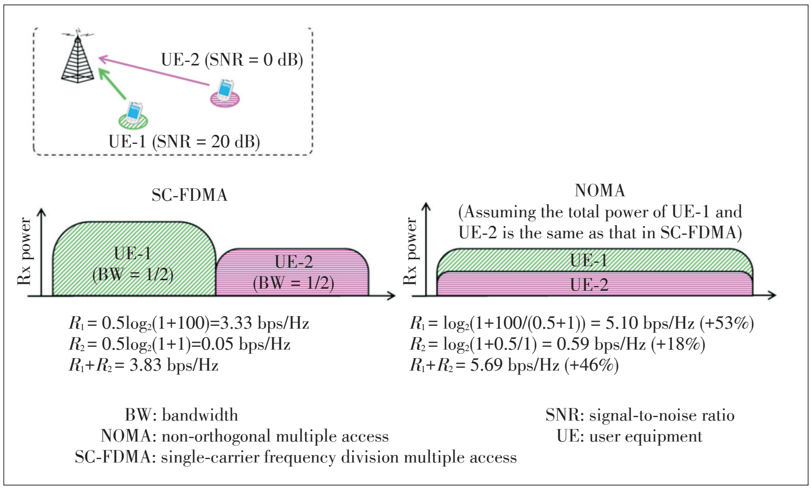
One comparison example of OMA and NOMA is shown in **Fig. 4** by assuming a 2-UE case with a cell-center UE and a cell-edge UE, where $|h_1|^2/N_0$ and $|h_2|^2/N_0$ are set to 20 dB and 0 dB, respectively. For OMA, we assume equal bandwidth is allocated to each UE (i.e., a = 0.5), the user rates are calculated according to (9) as R_1 = 3.33 bps and R_2 = 0.50 bps, respectively. On the other hand, in NOMA the total transmission power of each UE is assumed the same as that in OMA, the user rates are calculated according to (6) as R_1 = 5.10 bps and R_2 = 0.59



▲**Figure 2. Simple comparison example between NOMA and OFDMA for downlink.**



▲**Figure 3. Uplink NOMA with SIC applied at BS receiver.**



▲**Figure 4. Simple comparison example of NOMA and SC-FDMA for uplink.**

bps, respectively. The total UE throughput gain of NOMA over OMA is 46%. Therefore, for uplink NOMA, we can obtain similar performance gain as that for downlink NOMA.

3 Expected Benefits and Issues of NOMA

3.1 Benefits

NOMA is a promising multiple access scheme for the future owing to the following expected benefits.

1) Exploitation of channel gain difference among users

Unlike OMA (OFDMA) where the channel gain difference among users is translated into multi-user diversity gains via frequency-domain scheduling, in NOMA the channel gain difference is translated into multiplexing gains by superposing in the power-domain the transmit signals of multiple users of different channel gains. As shown in Figs. 1 and 2, by exploiting the channel gain difference in downlink NOMA, both UEs of high and low channel gains are in a win-win setup. Indeed, UEs with high channel gain (bandwidth-limited UEs) lose a little by being allocated less power, but can gain much more by being allocated more bandwidth, while UEs with low channel gain (power-limited UEs) also lose only a little by being allocated little less power and "effective" bandwidth (because of being interfered by the signal designated to the other UEs with high channel gain) but gain much more by being allocated more bandwidth. This win-win situation is also the main reason why NOMA gains over OMA increase when the difference in channel gains between NOMA paired UEs becomes larger [8].

2) Intentional non-orthogonality via power-domain user multiplexing and advanced receiver processing

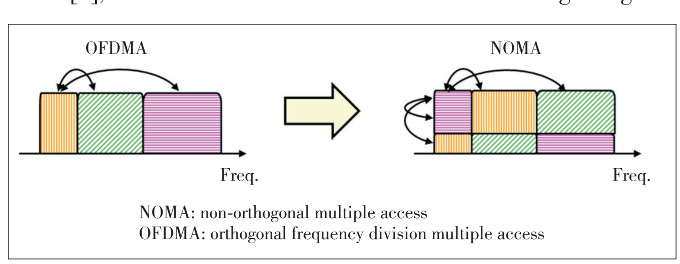
NOMA is a multiplexing scheme that utilizes an additional new domain, i.e., the power domain, which is not sufficiently utilized in previous systems. For downlink NOMA, non-orthogonality is intentionally introduced via power-domain user multiplexing as shown in **Fig. 5**; however, quasi-orthogonality can still be achieved. In fact, user demultiplexing is ensured via the allocation of large power difference between paired UEs and the application of SIC in power-domain. The UE with high channel gain (e.g., UE-1 in Figs. 1 and 2) is allocated less power and the UE with low channel gain (e.g., UE-2 in Figs. 1 and 2) is allocated more power. Such large power difference facilitates the successful decoding and the cancellation of the signal designated to UE-2 (being allocated high power) at UE-1 receiver and thus less complex receivers such as SIC can be used. In addition, at UE-2 receiver, the signal designated to UE-2 is decoded directly by treating the interference from the signal designated to UE-1 (being allocated low power) as noise.

On the other hand, NOMA captures well the evolution of device processing capabilities, generally following Moore's law, by relying on more advanced receiver processing such as SIC. In the same spirit, but for the purpose of inter-cell interference mitigation, network-assisted interference cancellation and suppression (NAICS), including SIC, is being discussed in LTE Release 12 [21]. Thus, NOMA is in fact a natural direction to extend the work in 3GPP on NAICS in LTE Release 13 and beyond, as it should be much easier to apply advanced receiver to deal with intra-cell interference than inter-cell interference. Moreover, the issue of the increased overhead is common to both intra-cell and inter-cell SIC since the signaling the information related to the demodulation and decoding of other UEs is needed. The signaling overhead issue is discussed later.

3) Robust performance gain in practical wide area deployments and high mobility scenarios

NOMA relies on power-domain instead of spatial domain for user multiplexing. Therefore, the knowledge of the transmitter of the instantaneous frequency-selective fading channels such as the frequency-selective channel quality indicator (CQI) or channel state information (CSI) is mainly used at the receiver for user pairing and multi-user power allocation. Thus, NOMA does require less fine CSI feedback compared to multi-user MIMO (MU-MIMO) and a robust performance gain in practical wide area deployments can be expected irrespective of UE mobility or CSI feedback latency.

In [9], downlink NOMA is shown to maintain good gains



▲**Figure 5. User multiplexing in power and frequency domains using NOMA.**

compared to OMA in particular with wideband scheduling. Thus, NOMA can be a promising multiple access to provide a good robustness to mobility by mainly relying on receiver side CSI and signal processing.

3.2 Issues

In the following, we discuss several issues regarding downlink NOMA, such as signaling overhead and receiver design.

3.2.1 Signaling Overhead

1) Multi-user scheduling

For OFDMA, both subband and wideband multi-user scheduling can be generally considered for frequency-domain scheduling. For the case of LTE which adopts OFDMA, irrespective of subband or wideband scheduling, the same channel coding rate (including rate matching) and data modulation scheme are assumed over all the subbands allocated to each single user. Thus, MCS selection is always wideband. However, when NOMA is applied over LTE and the user pairing and power allocation are conducted over each subband, a mismatch occurs between MCS selection granularity (i.e., wideband) and power allocation granularity (i.e., subband). Such a mismatch prevents the full exploitation of NOMA gains [14]. Thus, MCS selection over each subband, if introduced in 5G, could be beneficial for NOMA. On the other hand, when the NOMA user pairing and power allocation are conducted over each subband, the signaling overhead increases linearly with the number of subbands. Therefore, considerations on signaling overhead and performance tradeoffs need to be taken into account in the design of NOMA.

2) Multi-user power allocation

Because of the power-domain user multiplexing of NOMA, the transmit power allocation (TPA) to one user affects the achievable throughput of that user and also the throughput of other users. The best performance of downlink NOMA can obviously be achieved by exhaustive full search of user pairs and dynamic transmit power allocations. In case of full search power allocation (FSPA), all possible combinations of power allocations are considered for each candidate user set. However, FSPA remains computationally complex. Moreover, with such dynamic TPA, the signaling overhead associated with decoding order and power allocation ratio increases significantly. In order to reduce the signaling overhead associated with multi-user transmit power allocation of NOMA and to clarify the degree of impact of user pairing on the performance of NOMA, both exhaustive and simplified user pairing and power allocation schemes were explored [9]. In NOMA, users with large channel gain difference (e.g., large path-loss difference) are paired with high probability; thus, considering practical implementations, user pairing and TPA, could be simplified by using pre-defined user grouping and fixed per-group power allocation (FPA), where users are divided into multiple user groups according to the magnitude of their channel gains using pre-defined thresholds or according to their selected MCS level [15]. Pre-defined user grouping and fixed TPA can be promising in practical usage when the potential saving in signaling overhead is taken into account. For example, the order of SIC and information on power assignment do not need to be transmitted in every sub-frame but rather on a longer time scale.

For uplink NOMA, since the user separation process is implemented at the base station, we do not see a significant increase in the signaling overhead. In addition, the conventional control signaling assumed in LTE or LTE-Advanced may be reused in a straightforward manner.

3.2.2 Receiver Design and Resource Alignment

In practice, the impact of the receiver on NOMA performance remains as one concern. For the cell-edge UE, advanced receiver technologies may not necessarily be applied since the received signal power for this UE is greater than that for the cell-center UE, i.e., interfering UE. On the other hand, in order to decode the received signal for the cell-center UE, the application of interference cancellation is inevitable since the signal for the cell-center UE is significantly contaminated by that for the cell-edge UE in the same time and frequency resources. There are two types of interference cancellation receivers: symbol-level interference cancellation (SLIC) and codeword level interference cancellation (CWIC). For both receivers, the received data symbols for the cell-edge UE are first de-modulated by multiplying the received signal with the maximal ratio combining (MRC) weight or minimum mean squared error (MMSE) receiver, then the Log-likelihood ratio (LLR) corresponding to those de-modulated symbols are calculated.

For the CWIC, a sequence of LLRs which is called codeword is input to the Turbo decoder and a sequence of posteriori-LLRs is generated. After interleaving the sequence of posteriori-LLRs, the interleaved LLRs are used to calculate a soft symbol replica for the cell-edge UEs. On the other hand, for the SLIC, those LLRs are directly used to generate a symbol replica for the cell-edge UE.

The decoding performance of CWIC is basically better than that of SLIC. However, it is important to note that resource alignment and transmission power alignment highly impact the system performance and largely affect the receiver complexity and signaling overhead. For example, resource alignment among the paired UEs would be needed to facilitate the CWIC; however such a scheduling restriction may degrade the system-level performance due to reduction in scheduler flexibility and thus in the gains of frequency-domain scheduling. Also, some limitations on the UE pairing for the retransmissions need to be taking into account. On the other hand, when SLIC (e.g., reduced complexity ML (R-ML)) is applied, such restrictions related to resource allocation and retransmission can be relaxed and the frequency-domain scheduling gain can be obtained [25]. These tradeoffs need to be taken into account in the receiver choice.

4 Combination of NOMA and MIMO

MIMO is one of the key technologies to improve spectrum efficiency in LTE/LTE-Advanced. In general, MIMO techniques can be categorized into single-user MIMO (SU-MIMO), where only one UE is served in data transmission, and MU-MIMO, where more than one UE are served in data transmission. Because MIMO technology exploits spatial domain and NOMA exploits power domain, these two technologies can be combined to further boost the system performance. In single-input single-output (SISO) and single-input multiple-output (SIMO) downlink, the broadcast channel is degraded where superposition coding with SIC and dirty paper coding (DPC) are equivalent and optimal from the viewpoint of the achievable capacity region. However, for the downlink MIMO case, the broadcast channel is non-degraded and the superposition coding with SIC receiver becomes non-optimal, although DPC remains optimal [12], [19], [20]. These aspects need to be taken into account when NOMA is combined with MIMO.

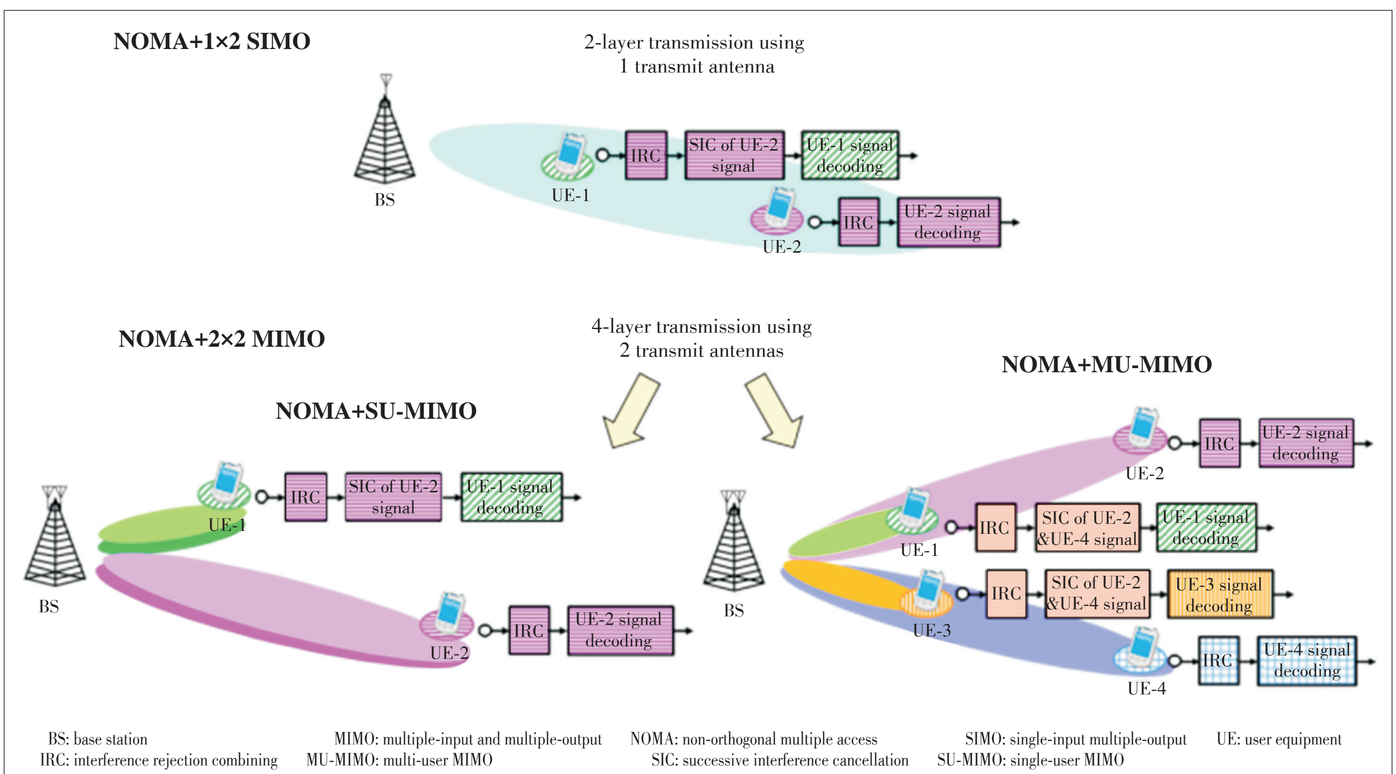
4.1 Downlink

There are two major approaches to combine downlink NOMA and MIMO technologies (**Fig. 6**).

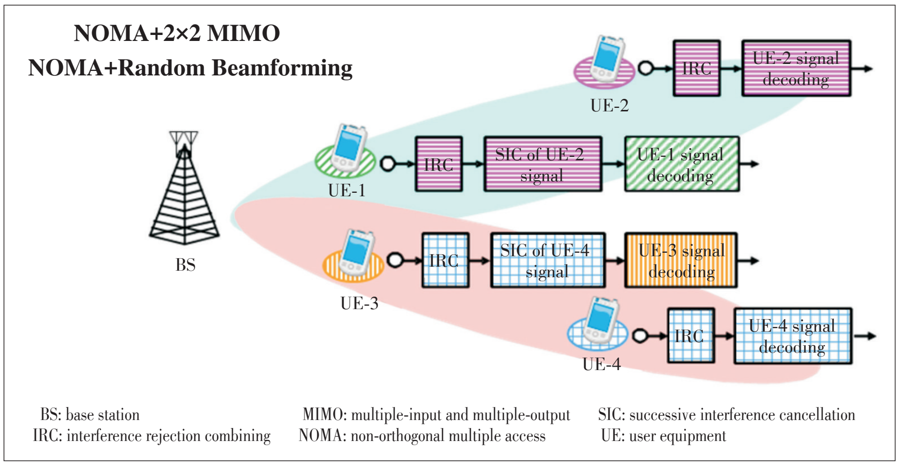
One approach is to use NOMA technique to create multiple power levels and apply SU-MIMO and/or MU-MIMO technique inside each power level. For example for NOMA with SU-MIMO (2x2), with up to 2 user multiplexing in the power-domain, non-orthogonal beam multiplexing enables up to 4 beam multiplexing using only 2 transmit antennas. In addition, the combination of NOMA with SU-MIMO can involve both open-loop MIMO (e.g., space frequency block coding (SFBC), large delay cyclic delay diversity (CDD)) and closed-loop MIMO (based on CSI such as the precoder indicator, channel quality indicator (CQI), rank indicator feedback by users)). Open-loop MIMO schemes when combined with NOMA are expected to provide robust performance in high mobility scenarios.

The other approach is to convert the non-degraded 2x2 MIMO channel into two degraded 1x2 SIMO channels, where NOMA is applied over each equivalent 1x2 SIMO channel separately, as shown in **Fig. 7** [10]. For this scheme, multiple transmit beams are created and superposition coding of signals designated to multiple users is applied within each transmit beam (i.e., intra-beam superposition coding). At the user terminal, the inter-beam interference is first suppressed by spatial filtering only by using multiple receive antennas, then multi-signal separation (e.g., SIC) is applied within each beam. This scheme can be considered as a combination of NOMA with MU-MIMO where fixed rank 1 transmission is applied to each user; thus, a large number of users would be required to obtain sufficient gains [11].

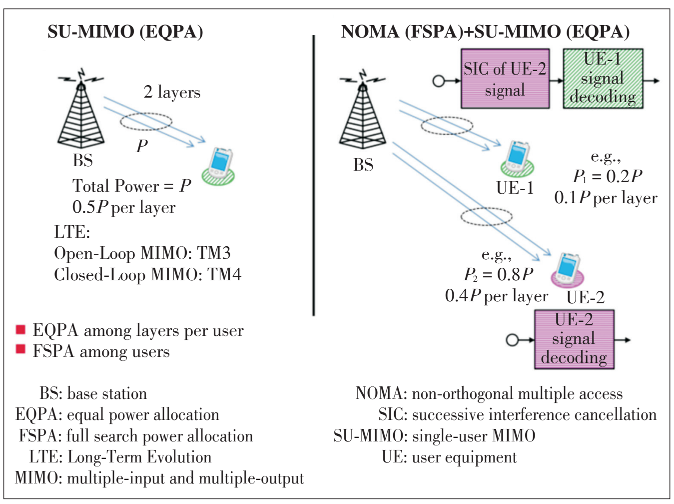
The combination of NOMA with SU-MIMO is illustrated in **Fig. 8** [20]. At the left side is the case of 2x2 SU-MIMO and at



▲**Figure 6. NOMA extension from 1x2 SIMO to 2x2 MIMO.**



▲**Figure 7. Downlink NOMA combined with 2x2 MIMO using random beamforming and applying IRC-SIC receivers.**



▲**Figure 8. Downlink NOMA with SIC combined with SU-MIMO (2x2 MIMO, 2-UE).**

the right side is the combination of NOMA with 2x2 SU-MIMO ($N_t = N_r = 2$), where the number of multiplexed UEs is 2. UE-1 and UE-2 are NOMA paired cell-center and cell-edge users, respectively.

By combining NOMA with SU-MIMO, up to 4-layer (4-beam) transmission is enabled using only 2 transmit antennas.

4.2 Uplink

Examples about uplink NOMA combined with MIMO assuming 2x2 antenna configuration are shown in **Fig. 9**. For the case of NOMA combined with SU-MIMO (left side), the UEs are separated in the power domain, and the spatial domain is used to multiplex multiple data streams of a single UE. For the case of NOMA combined with MU-MIMO (right side), UEs are separated in both power and spatial domains, i.e., within each user group of {UE-1, UE-2} and {UE-3, UE-4}, users are separated in the power domain. Among the {UE-1, UE-2} and {UE-3, UE-4} user groups, MU-MIMO transmission is applied to further separate the two user groups in the spatial domain. It can be seen that for the same MIMO antenna configuration, the same number of data streams are supported in uplink and downlink.

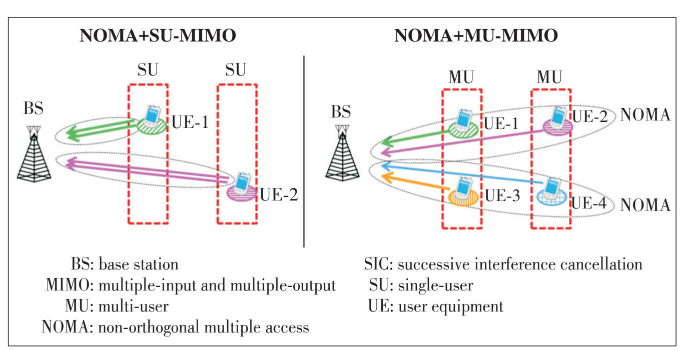
5 Performance of NOMA

5.1 NOMA Link-Level Performance

For the case of downlink NOMA with two UEs, we verified the effectiveness of CWIC for NOMA and the block error rate (BLER) performance of cell-center UE by evaluating link-level simulations. The number of transmit and receive antennas are both set to 2 and open-loop transmission mode 3 (TM3) is assumed as SU-MIMO transmission [16]. It is shown that almost the same BLER performance is obtained for NOMA with ideal SIC and CWIC. In addition, the power ratio (UE-1: UE-2 = P_1:P_2, $P_1 + P_2 = 1.0$, $P_1 < P_2$) for which the error propagation becomes dominant was investigated. The SIC error propagation increases for power sets with larger power ratio. In addition, the performance degrades with smaller values for the power ratio due to the increase of channel estimation error for the cell-center UE. Therefore, it would be important to limit the power sets to be used by the scheduler in order to maximize NOMA gains by limiting error propagation and the impact of channel estimation error, and ensuring that all chosen MCS combinations are decodable.

5.2 NOMA System-Level Performance

NOMA system-level performance has been investigated heavily with and without MIMO for both downlink and uplink. The multi-cell system-level simulation parameters are basically compliant with existing LTE specifications for an urban macro (UMa) scenario. The cell radius of the macro cells is set to 289 m (inter-site distance (ISD) = 500 m). 10 UEs are dropped randomly following a uniform distribution and full buffer traffic is assumed. Assuming proportional fairness scheduling, the performance gains of NOMA are measured in terms of cell



▲**Figure 9. Uplink NOMA combined with 2x2 MIMO.**

throughput (Mbps) and cell-edge user throughput (Mbps). The cell throughput is defined as the average aggregated throughput for users scheduled per a single cell, while the cell-edge user throughput is defined as the 5% value of the cumulative distribution function (CDF) of the user throughput.

The proportional fairness scheduler maximizes the geometric mean of the user throughputs, thus the tradeoff between user fairness and system throughput as shown in (10):

$$\sqrt[n]{\prod_{i\in U} R_i} = \frac{\sqrt[n]{\prod_{i\in U} R_i}}{\frac{1}{n}\sum_{i\in U} R_i} \times \frac{1}{n}\sum_{i\in U} R_i, \quad (10)$$

where U is the set of users scheduled. The first term on the right hand is the geometric mean of the user throughputs normalized with their arithmetic mean, representing a metric for user fairness, while the second term is the arithmetic mean of user throughputs, representing a metric for total system throughput.

5.2.1 Downlink

In [14], the user throughput of downlink NOMA is compared to that of OFDMA for both 1x2 SIMO and 2x2 MIMO. For MIMO, a comparison is made between the NOMA with SU-MIMO case and the OFDMA with SU-MIMO case for open-loop TM3 and closed-loop transmission mode 4 (TM4) MIMO. It is shown that NOMA with SU-MIMO provides gains over OMA with SU-MIMO by covering the entire user throughput region for both TM3 and TM4. The performance gains increase with the number of power sets. However, a hefty portion of the gains could be still achieved even with a few power sets.

5.2.2 Uplink

In [15], single-carrier frequency division multiple access (SC-FDMA) and NOMA are compared for uplink while taking uplink power control and resource contiguity constraint into account. A large performance gain in cell throughput is achieved for NOMA with very practical assumptions. This gain can be further increased by applying larger number of multiplexed users and/or enhanced schemes, e.g., advanced transmit power control (TPC). The large gain of NOMA mainly comes from the non-orthogonal multiplexing of users with large channel gain difference, which improves the resource utilization efficiency compared to SC-FDMA where only one UE exclusively occupies the radio resources.

When the user throughputs of SC-FDMA and NOMA are compared, it is observed that NOMA can achieve higher UE throughput than SC-FDMA for the most region of the CDF curve. However, for the cell-edge user throughput, i.e. 5% UE throughput, NOMA performance is worse than that of SC-FDMA. This is mainly due to two reasons. One reason is the increase of inter-cell interference in NOMA compared with SC-FDMA because more than one UE can be scheduled for simultaneous uplink transmission. The other reason is that the used TPC algorithm [8] is not fully optimized, where the total transmission power is controlled by a predefined parameter and the UEs in non-orthogonal transmission get less transmission power than what they get in SC-FDMA. Furthermore, the transmission power of the UEs is determined from large scale fading without considering instantaneous channel conditions.

To further balance between cell throughput and cell-edge throughput, three approaches are possible as listed below:

1) Introduction of weighted PF scheduling such that more resources are allocated to cell-edge user [12].

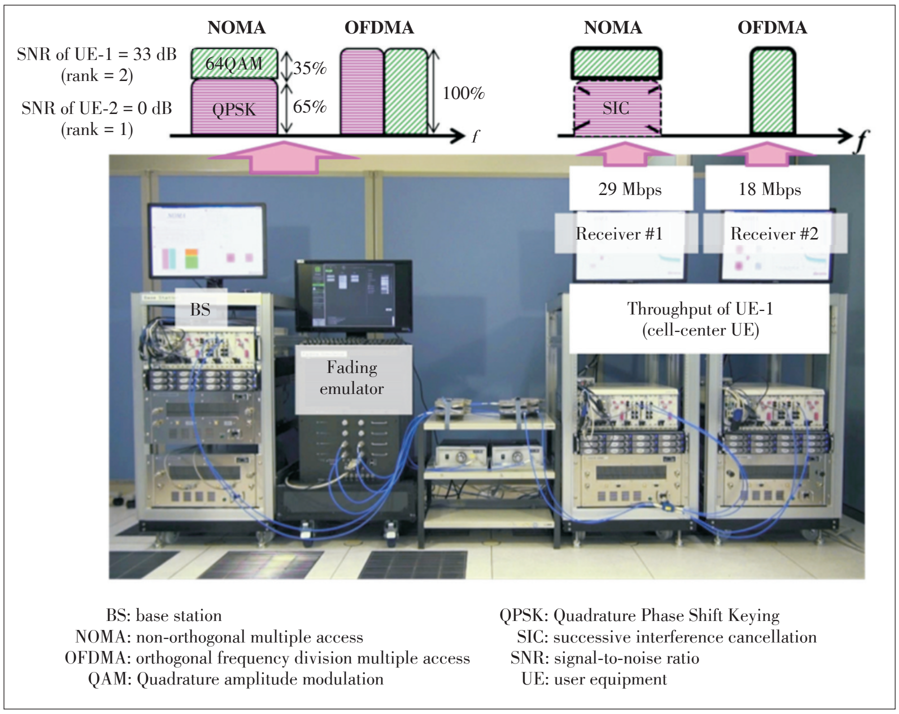
2) Combination of NOMA with other cell-edge performance enhancing technologies such as fractional frequency reuse (FFR) [15].

3) Introduction of sophisticated TPC algorithms designed for NOMA.

Taking the second approach above as an example, reference [15] shows that NOMA with FFR improves both the cell-edge throughput gain and the overall cell throughput gain. This possible improvement is due to the reduction in the inter-cell interference for both the cell-center UEs and cell-edge UEs.

5.2.3 NOMA Experimental Trial

A test-bed was developed to conduct experimental trials on NOMA, and to confirm NOMA performance with a real SIC receiver taking into account hardware (RF) impairments such as error vector magnitude (EVM) and the number of quantization bits of analog/digital (A/D) converter, etc. The test-bed assumed two UEs and used a carrier frequency of 3.9 GHz and bandwidth per user of 5.4 MHz for NOMA and of 2.7 MHz for OFDMA (a total bandwidth of 5.4 MHz for 2 users). LTE Release 8 frame structure is adopted and channel estimation is based on cell-specific reference signal (CRS). At the transmitter side, for each UE data, Turbo encoding, data modulation and multiplication by precoding vector are applied, then the precoded signal of the two UEs is superposed according to a predefined power ratio and goes through digital/analog (D/A) converter before up conversion to the carrier frequency of 3.9 GHz and transmission from two antennas. For MIMO transmission, LTE TM3 is utilized for open-loop 2-by-2 single user MIMO transmission. At the receiver side, two receive antennas are used to receive the RF signal, which is first down-converted and then goes through a 16-bit A/D converter. At the cell-center UE (UE-1), CWIC is applied. Using the fading emulator, for simplicity we set each link of the 2-by-2 MIMO channel to a 1-path channel with maximum Doppler frequency of 0.15 Hz. **Fig. 10** shows that the user throughputs of the cell-center UE, UE-1 (green color), with NOMA and SIC applied (29 Mbps) and with OFDMA only applied (18 Mbps). The user throughput of cell-edge UE, UE-2 (pink color) was adjusted for NOMA to be equal to the case of OFDMA. The measured gains of NOMA over OFDMA are the result of enabling three-layer transmission over a 2×2 MIMO channel while using twice the



▲**Figure 10. NOMA test-bed.**

bandwidth compared to OFDMA.

5.3 NOMA Standardization

NOMA was proposed to 3GPP LTE Release 13 [22] and a new study item (SI) under the name of "downlink multi-user superposition transmission (MUST)" was approved [23]. In 3GPP RAN1, the target scenarios, evaluation methodology, and the candidate non-orthogonal multiple access were discussed during the SI phase [24]–[26]. NOMA system-level performance with non-full buffer traffic and link-level performance for different receivers were evaluated [27], [28]. Based on Gray-mapped composite constellation with the same precoder but different transmit powers being applied to the superposed UEs, another NOMA multiplexing scheme is also considered in order to reduce signaling overhead and the receiver complexity compared to NOMA with SIC [26]. In such a scheme, coded bits for both the superposed UEs are jointly mapped onto the signal constellation based on Gray mapping, and then a reduced-maximum likelihood (R-ML) receiver is used for symbol-level interference cancellation [17]. The outcome of the SI in Release 13 was summarized under a technical report [29]. Later in Release 14, a work item (WI) was established to specify the necessary mechanisms to enable LTE to support downlink intra-cell multiuser superposition transmission for data channels with assistance information from serving BS to a UE regarding its experienced intra-cell interference [30]. In the WI, a MUST UE receiver is assumed to be capable to cancel or suppress intra-cell interference between co-scheduled MUST users for the following three cases:

- Case 1: Superposed data channels (i.e., Physical Downlink Shared Channels (PDSCHs)) are transmitted using the same transmission scheme and the same spatial precoding vector.
- Case 2: Superposed PDSCHs are transmitted using the same transmit diversity scheme.
- Case 3: Superposed PDSCHs are transmitted using the same transmission scheme, but their spatial precoding vectors are different.

During the WI phase, what was considered is up to 2 transmitter (Tx) CRS-based transmission schemes for cases 1 and 2, and up to 4 Tx CRS-based or up to 8 Tx DMRS-based transmission schemes for all three cases. The RAN1 agreements with RAN1 specification impacts made within the Release 14 WI are summarized in [31]. For example, for MUST Case 1 and Case 2, the higher layer and dynamic signaling mechanisms of MUST ON/OFF and of the power information of MUST users are specified.

6 Conclusions

This article presents an overview of the NOMA concept, design and its potential performance. Different from OFDMA, NOMA superposes multiple users in the power-domain, exploiting the channel gain difference between multiple UEs. NOMA contributes to the maximization of the tradeoff between system performance and user fairness. NOMA involves several aspects that need careful design, including the granularity in time and frequency of multi-user scheduling and multi-user power allocation, signaling overhead, receiver design, and combination with MIMO. NOMA can also be applied to uplink. For uplink, new issues arise including power control design to balance intra-cell and inter-cell interference and the design of the scheduling algorithm in case of single carrier transmission where consecutive resource allocation of non-orthogonally multiplexed UEs is taken into account.

From performance perspective, NOMA has shown promising gains for both downlink and uplink. These gains were investigated by link-level simulations, system-level simulations, and in experimental trials. Downlink NOMA was studied and specified in 3GPP RAN1 as MUST during LTE Release 13 and 14.

The design of sophisticated uplink power control schemes and of uplink reference signal for channel estimation to enable multiple user transmissions within the same frequency block is of interest to the future work.

Moreover, NOMA gains are expected to increase with more users, which correspond to the case of massive machine type communications (mMTC), i.e., massive sensors and devices with small packets being simultaneously transmitted over the cellular network. Further investigations and optimizations of NOMA for mMTC are also of interest.

References

[1] Y. Kishiyama, A. Benjebbour, H. Ishii, and T. Nakamura, "Evolution concept and candidate technologies for future steps of LTE - A," in *Proc. IEEE ICCS2012*, Singapore, Nov. 2012, pp. 473–477. doi: 10.1109/ICCS.2012.6406193.

[2] 3GPP, "Physical layer aspects for Evolved UTRA," TR 25.814,V7.1.0, Oct. 2006.

[3] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 764–775, Feb. 2009. doi: 10.1109/TIT.2008.2009832.

[4] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Transactions on Wireless Communications*, vol. 5, no. 4, pp. 938–947, Apr. 2006. doi: 10.1109/TWC.2006.1618943.

[5] L. Ping, Q. Guo, and J. Tong, "The OFDM-IDMA approach to wireless communication systems," *IEEE Wireless Communications*, vol. 14, no. 3, pp. 18–24, Jun. 2007. doi: 10.1109/MWC.2007.386608.

[6] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with successive interference cancellation for future radio access," in *Proc. APWCS 2012*, Kyoto, Japan, Aug. 2012.

[7] A. Benjebbour, Y. Saito, Y. Kishiyama, et al., "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. IEEE ISPACS*, Okinawa, Japan, Nov. 2013, pp. 770–774. doi: 10.1109/ISPACS.2013.6704653.

[8] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE PIMRC* 2013, London, UK, Sept. 2013, pp. 611–615. doi: 10.1109/PIMRC.2013.6666209

[9] A. Benjebbour, A. Li, Y. Saito, et al., "System-level performance of downlink NOMA for future LTE enhancements," in *Proc. IEEE Globecom*, Atlanta, USA, Dec. 2013, pp. 66–70. doi: 10.1109/GLOCOMW.2013.6824963.

[10] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE VTC2013-Fall*, Las Vegas, USA, Sept. 2013, pp. 1–5. doi: 10.1109/VTC Fall.2013.6692307.

[11] A. Li, A. Benjebbour, and A. Harada, "Performance evaluation of non-orthogonal multiple access combined with opportunistic beamforming," in *Proc. IEEE VTC Spring 2014*, Seoul, Korea, May 2014, pp. 1–5. doi: 10.1109/VTC-Spring.2014.7023050.

[12] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Transactions on Communications*, vol. E98-B, no. 3, pp. 403–414, Jan. 2015.

[13] Y. Lan, A. Benjebbour, X. Chen, A. Li, and H. Jiang, "Considerations on downlink non-orthogonal multiple access (NOMA) combined with closed-loop SU-MIMO," in *Proc. ICSPCS 2014*, Gold Coast, Australia, pp. 1–5. doi: 10.1109/ICSPCS.2014.7021086.

[14] A. Benjebbour, A. Li, Y. Kishiyama, H. Jiang, and T. Nakamura, "System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements," in *Proc. IEEE Globecom*, Austin, USA, Dec. 2014, pp. 706–710. doi: 10.1109/GLOCOMW.2014.7063515.

[15] A. Li, A. Benjebbour, X. Chen, H. Jiang, and H. Kayama, "Uplink non-orthogonal multiple access (NOMA) with single-carrier frequency division multiple access (SC-FDMA) for 5G Systems," *IEICE Transactions on Communications*, vol. E98-B, no. 8, pp. 1426–1435, Aug. 2015.

[16] K. Saito, A. Benjebbour, Y. Kishiyama, Y. Okumura, and T. Nakamura, "Performance and design of SIC receiver for downlink NOMA with open-loop SU-MIMO," in *Proc. IEEE ICC*, London, UK, Jun. 2015, pp. 1161–1165, doi: 10.1109/ICCW.2015.7247334.

[17] C. Yan, A. Harada, A. Benjebbour, et al., "Receiver design for downlink non-orthogonal multiple access (NOMA)," *Proc. IEEE VTC Spring 2015*, Glasgow, Scotland, May 2015, pp. 1–6. doi: 10.1109/VTCSpring.2015.7146043.

[18] A. Benjebbour, A. Li, K. Saito, Y. Kishiyama, and T. Nakamura, "Downlink non-orthogonal multiple access (NOMA) combined with single user MIMO (SU-MIMO)," *IEICE Transactions on Communications,* vol. E98-B, no. 8, pp. 1415-1425, Aug. 2015.

[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, England: Cambridge University Press, 2005.

[20] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1692–1706, Jul. 2003. doi: 10.1109/TIT.2003.813523.

[21] 3GPP, "Study on network-assisted interference cancellation and suppression for LTE," RP-130404, Feb. 2013.

[22] 3GPP, "Justification for NOMA in new study on enhanced multi-user transmission and network assisted interference cancellation for LTE," RP-141936, Dec. 2014.

[23] 3GPP, "New SI proposal: Study on downlink multiuser superposition transmission for LTE," RP-150496, Mar. 2015.

[24] 3GPP, "Deployment scenarios for downlink multiuser superposition transmissions," R1-152062, Apr. 2015.

[25] 3GPP, "Evaluation methodologies for downlink multiuser superposition transmissions," R1-153332, May 2015.

[26] 3GPP, "Candidate non-orthogonal multiplexing access scheme," R1-153335, May 2015.

[27] 3GPP, "System-level evaluation results for downlink multiuser superposition schemes," R1-154536, Aug. 2015.

[28] 3GPP, "Link-level evaluation results for downlink multiuser superposition schemes," R1-154537, Aug. 2015.

[29] 3GPP, "Study on downlink multiuser superposition transmission (MUST) for LTE (Release 13)," TR 36.859 V13.0.0, Dec. 2015.

[30] 3GPP, "New work item proposal: Downlink multiuser superposition transmission for LTE," RP-160680, Mar. 2016.

[31] 3GPP, "Summary of RAN1 agreements for Rel-14 DL MUST," R1-1613802, Nov. 2016.

Manuscript received: 2016-10-11

Biography

Anass Benjebbour (benjebbour@nttdocomo.com) obtained his Ph.D. and M.Sc. degrees in telecommunications in 2004 and 2001, respectively, and his B.Sc. diploma degree in electrical engineering in 1999, all from Kyoto University, Japan. In 2004, he joined NTT DOCOMO, INC. Since 2010, he has been a leading member of its 5G team. His research interests include novel system design concepts and radio access techniques for next generation mobile communication systems (5G), such as massive MIMO, NOMA, and waveform design. Dr. Benjebbour served as a 3GPP and ITU-R standardization delegate, a secretary of the IEICE RCS conference from 2012 to 2014, an associate editor for the *IEICE Communications Magazine* from 2010 to 2014, and an associate editor for the *IEICE Transactions on Communications* from 2014 to 2018. He is an author or a coauthor of 100+ technical publications, 4 book chapters and is an inventor of 50+ patent applications. He is a senior member of IEEE and IEICE.