Validation of NOMA System-Level Abstraction

Agnes Fastenbauer*†, Bashar Tahir*†, Stefan Schwarz*†, and Markus Rupp†
*Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion
†Institute of Telecommunications, TU Wien, Austria
agnes.fastenbauer@tuwien.ac.at

***Abstract*—Non-orthogonal multiple access (NOMA) transmission techniques promise reduced latency, massive user deployments, and higher data rates. To investigate how these performance enhancements translate to practical deployments in a network, a reliable system-level abstraction for downlink two-user power domain NOMA with successive interference cancellation at the receiver is needed. A simple and efficient abstraction is described and then validated through comparison with link-level simulation results. With the developed abstraction, the effects of the deployment of NOMA techniques in a wireless network with a large number of machine type communication users are evaluated in terms of cell throughput, number of users served simultaneously, and fairness.**

***Index Terms*—non-orthogonal multiple access, NOMA, multiple access, system-level, wireless network, simulations, mobile communications, cellular network, successive interference cancellation, SIC, multi-user superposition transmission, MUST**

I. Introduction

Analytical evaluation of non-orthogonal multiple access (NOMA) techniques has delivered promising results: NOMA allows to schedule more users [1], which leads to lower latency [2], [3], as well as to enlarge the achievable rate region [4] and thus to increase the throughput [5]. These improvements have led to NOMA techniques being adopted into standardization as multi-user superposition transmission (MUST) in the 3rd generation partnership project (3GPP) standardization body [6] and make NOMA a candidate for next generation multiple access for fifth generation (5G) and beyond [7].

However, the gains promised by NOMA can only be achieved under specific conditions when the channel quality of the multiplexed users shows large differences [5]. To evaluate whether the desired conditions are met in practical deployments, system-level simulations can be used. Simulations come with the additional benefits of being repeatable, cheap, and feasible even for heterogeneous networks [8]. Therefore, a number of system-level simulations have been performed to evaluate NOMA techniques, for example [9]–[12]. However, a validation of the employed NOMA system-level model is lacking as of yet. This lack is addressed in this paper.

The focus of this work lies on downlink two-user power domain NOMA with successive interference cancellation (SIC) at the receiver. SIC refers to iterative detection, where interfering signals are detected and subtracted from the received signal before detecting the desired signal. An example of a NOMA transmission is shown in Fig. 1, in which two users share the available resources in the power domain. A *near* user

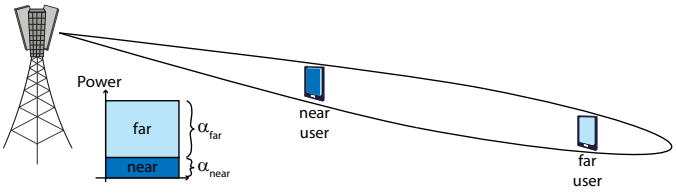


Fig. 1: A NOMA transmission.

with very good channel conditions, e.g., positioned close to the serving base station, is allocated a small share α_{near} of transmit power and a *far* user with poor channel conditions, e.g., at the cell edge, is allocated the larger remaining share of the available power $\alpha_{\text{far}} = 1 - \alpha_{\text{near}}$. The larger the difference in channel quality between the *near* and the *far* user is, the more effective the SIC and the NOMA transmission perform. Power allocation according to 3GPP MUST [6] is assumed.

In order to retrieve its data signal from the received signal, the *near* user has to perform SIC, where the signal of the *far* user is detected to then be subtracted from the received signal. Only then, the *near* user detects its own signal. Perfect interference cancellation at the *near* user is assumed in this work, thus no error propagation occurs during the detection of the *near* user signal. This means that the *near* user is not affected by interference from the *far* user in case of successful detection of the *far* user signal. The *far* user directly detects its signal treating the *near* user signal as additional noise.

The introduced NOMA system-level abstraction is not restricted to the used power allocation scheme and can be extended to an arbitrary number of superimposed users. A similar abstraction can be used for uplink NOMA, where the base station performs SIC for the received signals in an ascending order of channel quality. The trade-off struck for the enhanced performance achieved by NOMA is increased detection complexity due to SIC at the *near* user and the reduction in received signal quality.

To evaluate the gains achieved in a practical network deployment, an efficient system-level abstraction is described in detail in Section II and integrated in the Vienna Cellular Communications Simulators (VCCS). The VCCS is a suite of MATLAB based simulators, of which the Vienna 5G system-level simulator (SLS) [13] is extended for NOMA transmission and the Vienna 5G link-level simulator (LLS) [14] is used for validation of the NOMA system-level abstraction in Section III. The LLS implements the entire signal

978-1-6654-6975-3/22/$31.00 ©2022 IEEE

processing chain, including channel coding, multiple-input multiple-output (MIMO) processing, orthogonal frequency division multiplexing (OFDM), whereas the SLS works only with signal to interference and noise ratio (SINR) values on SINR to block error ratio (BLER) mapping tables. With the validated system-level abstraction, the performance of NOMA deployments with machine type communication (MTC) users is evaluated in Section IV.

II. System Model

The Vienna 5G SLS is used as a framework that is extended for two-user downlink power domain NOMA transmissions according to MUST with SIC at the receiver. The underlying system-level abstraction [15] is described in detail in [16, Section 3]. The system model is described for the case of a MIMO transmission with N_r antennas at the receiver, N_t antennas at the transmitter, that are used to transmit ν spatial beams that are called layers. The system model description for single-input single-output (SISO) can easily be obtained by the simplification $N_\mathrm{t} = N_\mathrm{r} = \nu = 1$.

In the system model, for each resource block scheduled to a user, the post-equalization SINR γ_l is calculated for each transmission layer $l = 1, \ldots, \nu$ according to the network geometry. Here, a resource block is a set of transmit symbols as defined in long term evolution-advanced (LTE-A) standardization for which the small scale fading will be assumed flat and constant for the duration of the resource block. The post-equalization SINR values γ_l are then mapped to a single value by means of a mutual information effective SINR mapping [17]. This single value is called effective SINR and is used to obtain the BLER and the throughput of the transmission.

A. Link Quality Model

The *link quality model* evaluates the quality of a link per resource block by calculating the post-equalization SINR γ_l for each layer l according to the experienced fading. This calculation is explained for only one resource block, but is performed for each resource block allocated to a user.

Assuming the use of a zero forcing (ZF) equalizer with perfect channel estimation, the estimated received symbol vector $\hat{\mathbf{x}} \in \mathbb{C}^{\nu_0 \times 1}$ after equalization with the ZF receive filter $\mathbf{F} \in \mathbb{C}^{\nu_0 \times N_\mathrm{r}}$ can be expressed as

$$\hat{\mathbf{x}} = \mathbf{F} \left(\underbrace{\mathbf{H}_0^{\mathrm{eff}} \mathbf{x}_0 + \sum_{i=1}^{N_\mathrm{int}} \mathbf{H}_i^{\mathrm{eff}} \mathbf{x}_i + \mathbf{n}}_{\text{receive signal}} \right), \tag{1}$$

in which N_int is the number of interfering transmitters, $\mathbf{x}_0 \in \mathbb{C}^{\nu_0 \times 1}$ is the transmit symbol vector of the desired transmitter, $\mathbf{x}_\mathrm{i} \in \mathbb{C}^{\nu_i \times 1}$ is the transmit symbol vector of the ith interfering transmitter using ν_i layers for transmission, $\mathbf{n} \in \mathbb{C}^{N_\mathrm{r} \times 1}$ represents the additive white Gaussian noise of noise power σ^2 at each receive antenna, and $\mathbf{H}_i^{\mathrm{eff}} \in \mathbb{C}^{N_\mathrm{r} \times \nu_i}$ represents the effective channel connecting the transmitter $i = 0, \ldots, N_\mathrm{int}$ to the receiver.

The effective channel can be separated into the macroscopic fading L_i, the normalized small scale fading matrix $\mathbf{H}_i \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$, the precoding matrix $\mathbf{W}_i \in \mathbb{C}^{N_\mathrm{t} \times \nu_i}$, and the transmit power allocated per layer $P_{\mathrm{tx},i}$:

$$\mathbf{H}_i^{\mathrm{eff}} = L_i \mathbf{H}_i \mathbf{W}_i P_{\mathrm{tx},i}. \tag{2}$$

For the small scale fading channel matrix $\mathbf{H}_i$, normalized refers to a matrix in which all entries have a mean power of one. The macroscopic fading L_i consists of the wall loss $L_{\mathrm{wall},i}$, the shadow fading $L_{\mathrm{shadow},i}$, the distance dependent path loss $L_{\mathrm{path},i}$, and the antenna gain G_i at the base station. The index i indicates that the loss affects the link to the ith transmitter. The complete macroscopic fading can be expressed as $L_i = L_{\mathrm{shadow},i} L_{\mathrm{wall},i} L_{\mathrm{path},i} G_i$. In the following, a homogeneous per-layer power distribution will be assumed, so that the average power received per layer $P_{\mathrm{rx},i}$ can be denoted as

$$P_{\mathrm{rx},i} = L_i P_{\mathrm{tx},i}. \tag{3}$$

When plugging Eq. (2) and Eq. (3) into Eq. (1), the estimated receive signal vector can be expressed as

$$\hat{\mathbf{x}} = \underbrace{P_{\mathrm{rx},0} \mathbf{F} \mathbf{H}_0 \mathbf{W}_0 \mathbf{x}_0}_{\text{desired signal}} + \underbrace{\sum_{i=1}^{N_\mathrm{int}} P_{\mathrm{rx},i} \mathbf{F} \mathbf{H}_i \mathbf{W}_i \mathbf{x}_i}_{\text{interference}} + \underbrace{\mathbf{F}\mathbf{n}}_{\text{noise}}. \tag{4}$$

With Eq. (4), the post-equalization SINR on layer l obtained with a ZF receive filter can be expressed as

$$\gamma_l = \frac{P_{\mathrm{rx},0}}{\underbrace{\sum_{i=1}^{N_\mathrm{int}} \sum_{l'=1}^{\nu_i} |\Theta_{i,ll'}|^2 P_{\mathrm{rx},i}}_{=I_l} + \underbrace{\sum_{r=1}^{N_\mathrm{r}} |F_{lr}|^2 \sigma^2}_{=N_l}} \tag{5}$$

$$= \frac{P_{\mathrm{rx},0}}{I_l + N_l}, \tag{6}$$

where A_{ij} represents the element in row i and column j of matrix $\mathbf{A}$ and the matrix $\mathbf{\Theta}_i = \mathbf{F}\mathbf{H}_i\mathbf{W}_i$ represents the interference enhancement with the receive filter $\mathbf{F} = ((\mathbf{H}_0\mathbf{W}_0)^\mathrm{H}\mathbf{H}_0\mathbf{W}_0)^{-1}(\mathbf{H}_0\mathbf{W}_0)^\mathrm{H}$. $\mathbf{A}^\mathrm{H}$ denotes the conjugate (or Hermitian) transpose of matrix $\mathbf{A}$.

In Eq. (5) the received power of the desired signal corresponds to the average received power due to the assumption of using a ZF equalizer. The terms I_l and N_l are introduced as the total interference power and the total noise power to allow for a shorter notation.

B. Extension for NOMA

The post-equalization SINR from Eq. (6) is now extended to include the effects of the NOMA power allocation and the interference of the *near* user at the *far* user and vice versa. At the *far* user, a reduction of the power allocated to the signal by the factor α_far and an increase of the interference by the *near* user signal is experienced. At the *near* user, SIC is more involved. The *far* user signal has to be detected successfully first and only then the signal of the *near* user can be extracted from the received signal. In the system-level abstraction this

means, that first the quality of the *far* user signal is evaluated at the *near* user through the post-equalization SINR $\gamma_{l,\text{near}}^{(\text{far})}$. If this post-equalization SINR is large enough to allow for a successful transmission of the *far* user signal to the *near* user, then the *near* user signal can be evaluated.

a) Far user: We first introduce an expression for the post-equalization SINR $\gamma_{l,\text{far}}$ of the signal received by the *far* user on layer l.

$$\gamma_{l,\text{far}} = \frac{\alpha_{\text{far}} P_{\text{rx,far},0}}{\alpha_{\text{near}} P_{\text{rx,far},0} + I_{l,\text{far}} + N_{l,\text{far}}}, \tag{7}$$

where $P_{\text{rx,far},0}$ represents the receive power of the combined transmit signal at the *far* user, $I_{l,\text{far}}$ is the interference power from other cells experienced by the *far* user, and $N_{l,\text{far}}$ is the noise power experienced at the *far* user. The additional term $\alpha_{\text{near}} P_{\text{rx,far},0}$ in the denominator represents the interference produced at the *far* user by the *near* user signal. From Eq. (7), it can be seen that the signal quality at the *far* user suffers from both the reduced power of the desired signal and the additional interference from the *near* user.

b) Near user: To model the SIC at the *near* user, first the post-equalization SINR of the far user signal at the *near* user is calculated as

$$\gamma_{l,\text{near}}^{(\text{far})} = \frac{\alpha_{\text{far}} P_{\text{rx,near},0}}{\alpha_{\text{near}} P_{\text{rx,near},0} + I_{l,\text{near}} + N_{l,\text{near}}}, \tag{8}$$

where $I_{l,\text{near}}$ and $N_{l,\text{near}}$ represent the interference and noise power experienced by the *near* user, and $P_{\text{rx,near},0}$ represents the receive power of the combined transmit signal at the *near* user. This post-equalization SINR $\gamma_{l,\text{near}}^{(\text{far})}$ is now used to determine whether the cancellation of the *far* user signal at the *near* user is successful. If the detection of the *far* user signal is unsuccessful, then the *near* user cannot remove the interference of the *far* user and cannot proceed with the detection of its own signal - the transmission of the *near* user signal fails. If the detection of the *far* user signal is successful, the post-equalization SINR for the transmission of the *near* user signal can be expressed as

$$\gamma_{l,\text{near}}^{(\text{near})} = \frac{\alpha_{\text{near}} P_{\text{rx,near},0}}{I_{l,\text{near}} + N_{l,\text{near}}}. \tag{9}$$

The post-equalization SINR in the final decoding iteration of the *near* user is only affected by the power share from the NOMA transmission, because perfect interference cancellation is assumed here. In order to account for interference from the *far* user after SIC, an additional term $\varepsilon \alpha_{\text{far}} P_{\text{rx,near},0}$ can be added to the denominator, in which ε represents the amount of interference that remains after imperfect interference cancellation. It can be seen in Eq. (9), that the quality of the *near* user signal suffers from the reduced signal power and the SINR is reduced by a factor of α_{near}. The probability for transmission failure increases, since two signals have to be detected successfully in order to retrieve the desired signal.

C. NOMA User Pairing

To decide which users share resources for a NOMA transmission, the user pairing scheme described in [5] is used. The user pairing scheme quantifies the quality of the channel condition of each user by evaluating the receive power $\nu_i L_i P_{\text{rx},i}$. The user with the strongest channel condition, i.e., SINR, in a cell is paired with the user with the weakest channel condition if the difference between their channel condition metrics is larger than the user pairing threshold. These users are then removed from the pairing candidates and the procedure is repeated until either a pair does not meet the condition of exceeding the user pairing metric, or fewer than two users are left as candidates. If two users have been paired, they share all resources allocated to either of them by the scheduling algorithm.

III. Validation of NOMA System-Level Abstraction

To validate the accuracy of the developed NOMA system-level abstraction, the simulation results using the system-level abstraction are compared to the results of the Vienna 5G LLS for SISO, multiple-input single-output (MISO), and MIMO transmission.

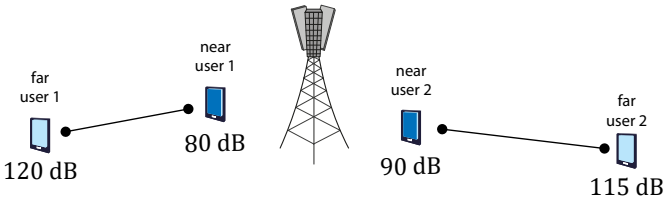
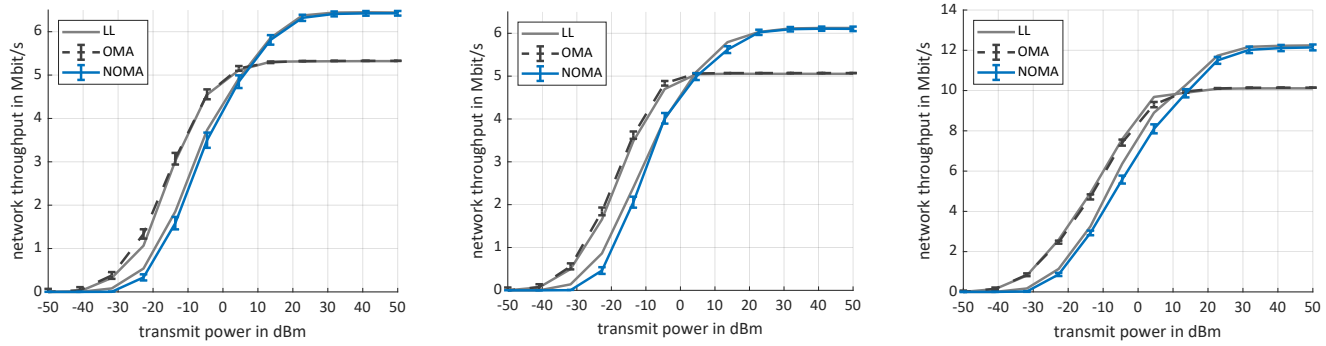


Fig. 2: Simulation scenario for comparison to the Vienna 5G LLS. The value under each user equipment indicates the path loss experienced by this user. The barbell connecting two users indicates that these users are paired for NOMA transmission.

A. Validation Simulation Scenario

TABLE I: Validation simulation parameters

Time and Frequency	
slot duration	1 ms
frequency	2.5 GHz
bandwidth	1.4 MHz
subcarrier spacing	15 kHz
Link Properties	
transmit mode	{SISO, MISO, MIMO}
$N_t \times N_r$	$\{1 \times 1, 2 \times 1, 2 \times 2\}$
small scale fading	Pedestrian A [18]
temporally correlated fading	true [19]
user speed	1 m/s
feedback delay	1 slot
precoder	3GPP TS 36.211 [20]
Coding and Decoding	
MUST index	01 [20]
coding	Turbo
decoding	Linear Log-MAP
decoding iterations	8



(a) $N_t \times N_r = 1 \times 1$ SISO (b) $N_t \times N_r = 2 \times 1$ MISO (c) $N_t \times N_r = 2 \times 2$ MIMO

Fig. 3: Simulation results for Vienna 5G LLS and SLS with 99% confidence intervals. The confidence intervals for link-level simulation are omitted for better readability, they are of similar magnitude as the confidence intervals of the system-level simulation results.

The simulation scenario used for validation is depicted in Fig. 2 and the simulation parameters are listed in Table I. A single cell with two users - *near* user 1 and *near* user 2 - is first simulated as a base line validation. Then, a *far* user is added to each *near* user for combined NOMA transmission. Each user pair is allocated half of the available resources. The general parameter settings are LTE-A compliant with MUST mode 1. The feedback is calculated according to [21] with a feedback delay of one slot. Each user is affected by a path loss as indicated in Fig. 2 and no other macroscopic fading mechanism is applied. The simulations are performed for SISO, MISO, and MIMO transmission modes with one or two transmit and receive antennas.

B. Validation Simulation Results

The cell throughput over transmit power is shown in Fig. 3 for three different transmission modes for an orthogonal multiple access (OMA) and a NOMA network for a system-level and a link-level simulation. It can be seen that the results from the system-level simulation closely match the results of the link-level simulation, which validates the system-level abstraction and proves its accuracy. The small differences are caused by the different receiver structures and the simplifications introduced with the abstraction. The system-level simulation uses SIC with a ZF equalizer and the link-level simulator performs maximum likelihood (ML) detection. Both detection schemes are optimal in the high signal to noise ratio (SNR) region and thus lead to comparable simulation results.

For low transmit powers, from -50 dBm to -40 dBm, all transmissions fail due to the lack of sufficiently high receive power and no network throughput is observed. The network throughput of the OMA network starts increasing at lower transmit powers because the combined NOMA transmission reduces the quality of the received signal and thus transmissions fail for transmit powers up to -30 dBm. At transmit powers of around 10 dBm, NOMA starts to outperform OMA and the network throughput exceeds the throughput in the OMA network. At high transmit powers, the network throughput saturates at the achievable throughput of the highest available modulation and coding scheme. This achievable throughput is higher in the NOMA network since two transmissions, the *near* and *far* user transmission, take place simultaneously. The transmission of the *near* user achieves the same throughput as the OMA user and the *far* user achieves the throughput of the modulation and coding scheme fixed in the MUST standardization, which corresponds to a channel quality indicator of 6.

It should be noted that the achieved throughput is almost doubled for MIMO transmission compared to SISO and MISO because two layers are available for transmission. The achieved throughput is slightly decreased for MISO transmission compared to SISO since more resources have to be used for signaling overhead.

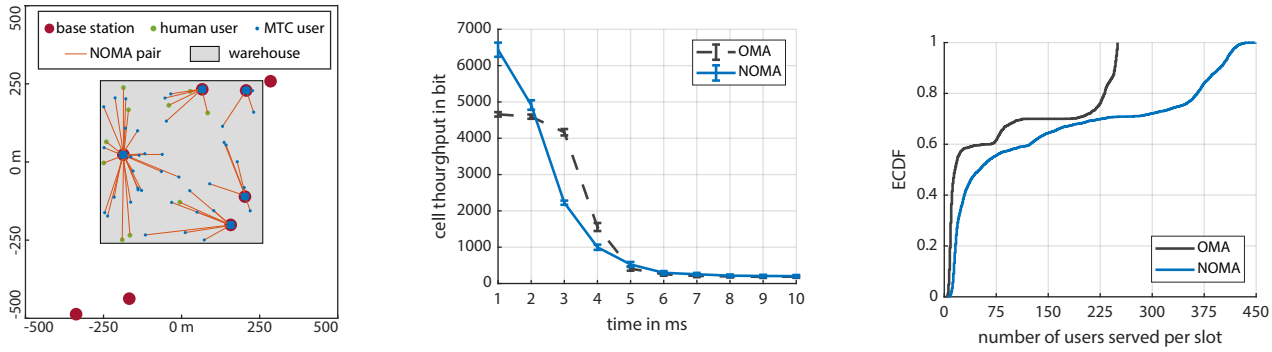
For NOMA to be beneficial, the receive power needs to be high enough. In case of poor link quality, the use of NOMA transmission will reduce the achieved throughput and increase the number of failed transmissions, as can be seen in the region of medium transmit power from -40 dBm to 10 dBm.

IV. MTC Performance Evaluation

A. Warehouse Scenario with Bursty MTC Traffic

To evaluate the performance improvement introduced by NOMA in a network, a warehouse with a large number of static MTC users and a small number of human and moving MTC users is considered. The MTC users are positioned on shelves that are equipped with small cells distributed uniformly in the warehouse, see Fig. 4a. Additionally, a smaller number of moving MTC and human users are uniformly distributed in the warehouse and move with a speed of 4 km/h. The warehouse is surrounded by 10 uniform randomly positioned macro base stations that introduce interference into the network. Simulation parameters are listed in Table II, parameters identical to those listed in Table I are omitted.

The MTC users have data occurring in traffic bursts: all MTC users generate a packet of 96 data bits at the same time. This corresponds to the amount of data transmitted in one



(a) Section of one network realization of the MTC simulation scenario.

(b) Average cell throughput with 95% confidence interval.

(c) ECDF of number of users served in a slot.

Fig. 4: MTC simulation scenario and results.

resource block using a channel quality indicator of 6. The human users generate traffic at a constant rate of 100 kbit/s in 100 bit-packets. The network behavior is analyzed for 10 ms after each MTC traffic burst. Each simulation consists of three traffic bursts and the simulation is repeated for 100 Monte Carlo runs with independent network and channel realizations.

A user pairing threshold of 0 dB is chosen to maximize the use of NOMA to allow analysis of the beneficial and detrimental effects of combined transmissions in the network. In a practical deployment, the pairing threshold should be increased to avert the negative effects of NOMA transmission for users with channel conditions that do not permit a shared use of the transmission resources.

TABLE II: MTC simulation parameters

Network Geometry	
scenario size	2000 m × 2000 m
warehouse size	500 m × 500 m
shelf size	15 m × 15 m
number of shelves	5
MTC users per shelf	150
human users	10
moving MTC users	40
moving user speed	4 km/h
Time and Frequency	
frequency	2 GHz
bandwidth	5 MHz
Link Properties	
$N_t \times N_r$	1×1
transmit power	20 W, 40 W (macro)
antenna pattern	omnidirectional
path loss	free space
shadow fading	space-correlated [22], $\mu = 0$, $\sigma = 1$
wall loss	20 dB
Resource Allocation	
user pairing threshold	0 dB
scheduler	round robin

B. Bursty MTC Network Performance

The simulation results show that the use of NOMA transmissions allows the network to recover more quickly from traffic bursts and to serve more users simultaneously, which increases the fairness in the network slightly. However, the number of failed transmissions is also increased, since the channel conditions are worsened by the combined transmission. This could be mitigated, but not prevented completely, with a more suited user pairing strategy.

The average cell throughput is shown in Fig. 4b. The time indicates the time since the bursty MTC user traffic has been generated. The cell throughput steadily decreases after the traffic burst, indicating that the data from the traffic burst is successfully transmitted to the MTC users and the network recovers and returns to serving a constant base traffic load from the human users. The cell throughput of the simulation with NOMA shows a steeper decrease, because more MTC users can be served in the initial two transmissions, whereas the the cell throughput of the OMA network needs more time to recover from the traffic burst. These results show that NOMA is a suitable technology to tackle the challenges brought by massive connectivity and can assist in fulfilling the requirements for lower latency in future networks.

The transmission resources are scheduled in units of resource blocks, where 50 resource blocks are available at each small cell. With five small cells, this results in 250 resource blocks that can be scheduled to users for transmission. In an OMA network this equates to 250 users that can be served simultaneously. This maximum number of users can be observed in Fig. 4c, where the empirical cumulative distribution function (ECDF) of served users reaches its maximum at 250 users. A user is considered as a served user if data has been successfully received in the current slot. In the NOMA network, at most 500 out of the 800 users in the simulation can be served simultaneously. This maximum number is not reached: the ECDF reaches its maximum at 450 users. Since all users are paired for NOMA transmission and, in the first transmission slot, all users have data, the lower number of served users indicates that a number of transmissions fail

in the first slot of the NOMA simulation. The ECDFs in Fig. 4c show multimodal distributions with three modes: the mode on the outmost right of the plot represents the initial stage where the maximum network capacity is used, the mode at the center represents the transition of the network from full load to base load, and the mode closest to the origin represents the distribution of users served at base load. Overall, NOMA consistently outperforms OMA in number of users served simultaneously. The increase in number of users served simultaneously also translates into increased fairness in the network: Jain's fairness index [23] is increased by 7.3%, from 0.406 to 0.436.

The increased connectivity comes with the downside of reduced channel quality for each individual user. This increases the number of users in outage, i.e., that are not served within the observed 10 ms of the traffic burst. On average 24.73 MTC users in the NOMA network and 14.98 in the OMA network remain with data to transmit at the end of the duration of the simulated 10 ms. This effect could be reduced, but not eliminated, by a more suited user pairing strategy or better selection of modulation and coding schemes.

V. Conclusion

The system-level abstraction for two-user power domain NOMA transmissions with SIC at the receiver has been validated through comparison with link-level simulation results and can thus be used to evaluate the performance of NOMA in large scale networks. This also validates results previously obtained with the same system-level abstraction. Simulations with MTC and human users show that NOMA techniques can reduce the duration of traffic peak loads and increase the number of users served simultaneously, making NOMA a promising candidate to tackle the challenges brought about by massive connectivity and the requirements for lower latency in future networks. However, the utilization of NOMA techniques reduces the channel quality of the individual user and care has to be taken not to push users already experiencing poor channel quality into outage.

Acknowledgment

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association is gratefully acknowledged.

References

[1] S. Mishra, L. Salaun, C. S. Chen, and K. Giridhar, "Analysis of downlink connectivity in NB-IoT networks employing NOMA with imperfect SIC," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Jun. 2021.

[2] C. Xiao, J. Zeng, W. Ni, X. Su, R. P. Liu, T. Lv, and J. Wang, "Downlink MIMO-NOMA for ultra-reliable low-latency communications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 780–794, Apr. 2019.

[3] M. Amjad and L. Musavian, "Performance analysis of NOMA for ultra-reliable and low-latency communications," in *IEEE Globecom Workshops*, Dec. 2018.

[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 *IEEE 77th Vehicular Technology Conference (VTC Spring)*, Jun. 2013.

[5] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[6] 3rd Generation Partnership Project (3GPP), "Technical specification group radio access network; study on downlink multiuser superposition transmission (MUST) for LTE," 3GPP, TR 36.859, Dec. 2015.

[7] Y. Liu, S. Zhang, X. Mu, Z. Ding, R. Schober, N. Al-Dhahir, E. Hossain, and X. Shen, "Evolution of NOMA toward next generation multiple access (NGMA) for 6G," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1037–1071, Apr. 2022.

[8] M. Rupp, S. Schwarz, and M. Taranetz, *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation*, ser. Signals and Communication Technology. Singapore: Springer, 2016.

[9] M. Kimura and K. Higuchi, "System-level throughput of NOMA with SIC in cellular downlink under FTP traffic model," in *International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2015.

[10] Y. Saito, A. Benjebbour, A. Li, K. Takeda, Y. Kishiyama, and T. Nakamura, "System-level evaluation of downlink non-orthogonal multiple access (NOMA) for non-full buffer traffic model," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, Oct. 2015.

[11] Z. Xie, W. Yi, X. Wu, Y. Liu, and A. Nallanathan, "STAR-RIS aided NOMA in multicell networks: A general analytical framework with gamma distributed channel modeling," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5629–5644, Aug. 2022.

[12] M. H. N. Shaikh, V. A. Bohara, A. Srivastava, and G. Ghatak, "On the performance of RIS-aided NOMA system with non-ideal transceiver over nakagami-m fading," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, Apr. 2022.

[13] M. K. Müller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. R. Elbal, A. Nabavi, L. Nagel, S. Schwarz, and M. Rupp, "Flexible multi-node simulation of cellular mobile communications: the Vienna 5G System Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 17, Sep. 2018.

[14] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz, and M. Rupp, "Versatile mobile communications simulation: the Vienna 5G Link Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 226, Sep. 2018.

[15] J. C. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *IEEE Vehicular Technology Conference (VTC Spring)*, Taipei, Taiwan, May 2010.

[16] J. C. Ikuno, "System level modeling and optimization of the LTE downlink," Ph.D. dissertation, E389, TU Wien, 2013. [Online]. Available: https://publik.tuwien.ac.at/files/PubDat_226618.pdf

[17] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2006.

[18] 3rd Generation Partnership Project (3GPP), "Technical specification group radio access network; high speed downlink packet access: UE radio transmission and reception," 3GPP, TR 25.890, May 2002.

[19] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, Jun. 2003.

[20] 3rd Generation Partnership Project (3GPP), "Evolved Universal Terrestrial Radio Access (E-UTRA) physical channels and modulation," 3GPP, TS 36.211, Jan. 2015.

[21] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *6th conference on Wireless advanced (WiAD)*. IEEE, 2010, pp. 1–6.

[22] T. Dittrich, M. Taranetz, and M. Rupp, "An efficient method for avoiding shadow fading maps in system level simulations," in *21st International ITG Workshop on Smart Antennas (WSA)*, Mar. 2017, pp. 1–8.

[23] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *Tech. Rep. TR-301, DEC*, Sep. 1984.