Improving NOMA Performance by Application of Autoencoders and Equidistant Power Allocation

Matthias Hummert, Niklas Bulk, Carsten Bockelmann, Dirk Wübben, Armin Dekorsy
Dept. of Communications Engineering,
University of Bremen, Bremen, Germany
Email: {hummert, bulk, bockelmann, wuebben, dekorsy}@ant.uni-bremen.de

***Abstract*—Non-orthogonal multiple access (NOMA) has been introduced as a promising scheme to allow for superposition of signals, such as the transmission of multiple services in the same resource block (time and frequency). In this paper, we propose the application of Deep Learning (DL) in an Autoencoder (AE) framework for simultaneous usage in multi-service NOMA transmission. While classical NOMA simultaneously incorporates locally separated user equipments (UEs), we focus on the simultaneous transmission of services within a single UE. Our scheme achieves improved performance compared to classical NOMA schemes and is capable of performing optimal estimation of the superimposed transmit signals. The scheme utilizes an equidistant power allocation scheme. The results show the potential of using DL to enhance the performance of NOMA systems and improve their adaptability and flexibility to different scenarios.**

***Index Terms*—Deep Learning, Autoencoder, NOMA, Resource Allocation, 6G, Neural Network**

I. Introduction

Today's industrial communication systems face the challenge of accommodating multiple service classes within a single radio system [1]. A promising solution is NOMA, designed to efficiently allocate resources by power domain superposition [2], [3]. The approach in [4] shifts from conventional NOMA to service-based NOMA, superimposing different services for a single UE. Service-based NOMA has to be proven optimal with equidistant power allocation [4]. This service-based NOMA approach is applicable to both uplink and downlink transmission. In this paper we focus on uplink.

DL and AEs have gained interest in wireless communication research [5]–[7]. Specifically, AEs jointly optimize transmitter and receiver structures, utilizing the power of neural networks (NNs). This optimization incorporates the channel to handle signal transmission and reception. Of particular interest is the work [5], as they were the first to propose the use of AE in the context of communication. They further applied simultaneous transmitter and receiver structures in a Multiple-Input-Multiple-Output (MIMO) environment to demonstrate the potential use cases of the idea.

Integrating these AE concepts into NOMA overcomes successive interference cancellation (SIC) complexities [6]. Techniques like weighted AE and models such as SICNet outperform SIC in cases of imperfect channel state information [7], [8].

As a main contribution, we present NOMA-AE, a DL-based AE, applied to a NOMA scenario. In our work two different services transmit in the same resource block and hence share the same channel. NOMA-AE employs NNs for learning a transmit signal and distinct power allocation for each service on the transmit side and provides different detection strategies via NN on the receiver side. Our approach offers complete flexibility and is suitable for modular implementation across numerous services. We provide multiple evaluations and simulation results, showing the potential of applying AE structures in the NOMA world by outperforming classical State-of-the-Art (SotA) SIC schemes.

II. System Overview

The goal of service-based NOMA is to simultaneously transmit S services from one UE in the same time-frequency band, hence leading to interference in the signals. For separation of the services, different power allocations are used. Typically, these services utilize different modulation schemes and have distinct Quality of Service (QoS) requirements, such as latency and block error rate (BLER). To distinguish the services on the receiver side SIC is used [9].

In this paper, we propose to replace the typical transmitter (modulation) and receiver structures (SIC) of service-based NOMA by applying NNs and machine learning (ML) in an AE structure, as shown in Fig. 1. We limit ourselves to two services in this paper, but in general this framework provides the flexibility to be extended to *more* than two services. Nevertheless, more services will introduce more interference and the performance will overall degrade.

The key idea is to jointly train transmitter and receiver (highlighted in orange) given a channel that disturbs the overall transmission. This overall framework allows to incorporate the superposition of services directly into the training. The overall system is *real* valued, as NNs cannot handle complex numbers straightforwardly. Hence, we stack real and imaginary part in two-dimensional vectors.

A. Transmitter

Specifically, the transmitter aims to transmit bitstreams of two services by constructing suitable transmit symbols $\mathbf{d}_s$

This work was funded the German Ministry of Education and Research (BMBF) under grant 16KIS1180K (FunKI) and 16KIS1012 (IRLG).

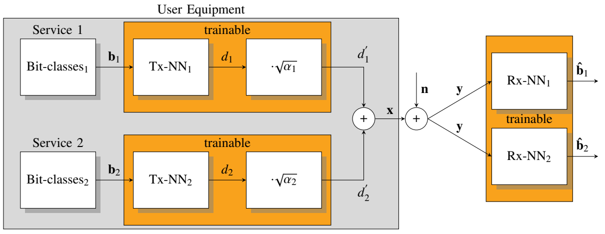

Fig. 1. Block diagram of NOMA-AE. Two independent sources create bits, ordered in vectors $\mathbf{b}_s$, which are fed to trainable Tx-NN$_s$ to be mapped to $\mathbf{d}_s$, weighted $\mathbf{d}_s'$, ensuring overall power normalization, and superimposed to $\mathbf{x}$. After an AWGN channel, the received signal $\mathbf{y}$ is fed into Rx-NN$_s$ to reconstruct $\hat{\mathbf{b}}_s$. Trainable parts of the proposed scheme are highlighted in orange.

via trainable NN for each service s, named Tx-NN$_s$ in the complex plane, i.e. a complex modulation scheme, where the index s runs from $s = 1, 2$. These learned modulation symbols are then scaled and superimposed to be sent over a channel. The complex modulations are stored in a real valued two-dimensional vector $\mathbf{d}_s \in \mathbb{R}^2$, containing real and imaginary part separately. The key idea is to jointly optimize the Tx-NN$_s$ for each service s so that a suited superimposed modulation is learned. The bitstream is represented, for each time instant, via one-hot encoded vectors $\mathbf{b}_s$. A one-hot coding is a vector with exactly one entry "1" and all others are "0". We use one-hot-vectors to have an easy represenatation for each modulation symbol and form a classification problem. M_s is the number of modulation symbols for each service s. E.g., for $M_1 = 4$ this results in four vectors $[1000, 0100, 0010, 0001]$. M_s directly represents the dimension of $\mathbf{b}_s$ as each modulation symbol is represented by one vector.

Tx-NNs are nonlinear functions with trainable parameters (weights), which can mathematically be written as

$$\mathbf{d}_s = f_s^t(\mathbf{b}_s, \mathbf{\Theta}_s^t), \tag{1}$$

where f_s^t is a nonlinear function with trainable weights $\mathbf{\Theta}_s^t$ for each service s, where t is used to identify a transmit NN. The Tx-NN$_s$ follow a feedforward structure with various hidden layers and Rectified Linear Unit (ReLu) activation functions. The output layer consists of a linear layer and each Tx-NN$_s$ for each service is power normalized to limit the total transmit power P. The details of Tx-NN$_s$ are listed in Table I. The structure of each Tx-NN$_s$ was selected experimentally after simulations, prioritizing the best performance while choosing the least computationally expensive NN. The output of the Tx-NN$_s$ is the learned modulation, contained in the two-dimensional vector $\mathbf{d}_s \in \mathbb{R}^2$. Each dimension of this vector represents real and imaginary part of the complex transmit symbols.

Afterwards, scaling via α_s is applied to ensure power normalization for the superimposed signal $\mathbf{x}$, fulfilling $\alpha_1 + \alpha_2 = 1$. The scaling via α_s can lead to destructive interference if chosen poorly, as the modulation symbols are directly scaled in amplitude. Please note that the α_s are adapted during the training process, while $\alpha_1 + \alpha_2 = 1$ is ensured. This constraint

TABLE I
TRANSMITTER AND RECEIVER NN CONFIGURATIONS

Name	Hidden Layer	width of layer	# of weights
Tx-NN$_s$	3	300, 200, 100	82002 & 85602
Rx-NN$_s$	3	500, 200, 100	122204 & 123416

is ensured by normalization. These α_s are not part of the NNs. Finally, the two services are superimposed via addition. Therefore, the transmit signal can be expressed as:

$$\mathbf{x} = \sum_{s=1}^{2} \left(\sqrt{\alpha_s \cdot P} \cdot \mathbf{d}_s \right) \tag{2}$$

B. Receiver

The goal of the receiver is to reconstruct the bit vectors/one-hot vectors $\mathbf{b}_s$. For the received signal, an additive white Gaussian noise (AWGN) channel is assumed. We choose AWGN as a first experiment to verify our approach. The received signal yields

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{3}$$

where $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2/2)$ is noise and $\mathbf{y}, \mathbf{x}, \mathbf{n} \in \mathbb{R}^2$. The noise power is noted per dimension of $\mathbf{n}$ of real and imaginary part. The received signal $\mathbf{y}$ is fed into trainbale NN for each service s, named Rx-NN$_s$. The key objective for the Rx-NN$_s$ is to reconstruct the one hot vectors $\hat{\mathbf{b}}_s$ given the received signal $\mathbf{y}$. The Tx-NN$_s$ and Rx-NN$_s$ are trained jointly, to cope with the channel. These two Rx-NN$_s$ use, again, multiple feedforward layers with ReLu activation. The output layer consists of a softmax activation to form probabilities for final classification. Mathematically, for the Rx-NN$_s$ we write

$$\hat{\mathbf{b}}_s = f_s^r(\mathbf{y}, \mathbf{\Theta}_s^r), \tag{4}$$

where f_s^r is again a nonlinear function with trainable parameters $\mathbf{\Theta}_s^r$. We use r to differ the Rx-NN from the Tx-NN. Again, the structure of these NN have been chosen experimentally and the details of the Rx-NN$_s$ are listed in Table I.

One advantage and distinguishing feature of NOMA-AE is its ability to run both Rx-NN$_s$ in parallel, resulting in no error propagation compared to SIC.

C. Supervised Learning and NN details

The general idea of supervised learning is to learn from labeled data, meaning, each training dataset contains ground truth label. The ground truth labels represent the optimal outcome for each training sample. Here we use NNs to create trainable functions with adaptable parameters $\mathbf{\Theta}$ to minimize a loss function $\mathcal{L}(\mathbf{b}_s, \hat{\mathbf{b}}_s)$, here the Cross Entropy (CE) loss. The loss provides a form of closeness measure between the ground truth labels, here the vectors $\mathbf{b}_s$, and the output of the adaptable function $\hat{\mathbf{b}}_s$. By minimizing the loss, the trainable parameters are adopted accordingly. The loss function is adapted by using an iterative form of stochastic gradient descent. Specifically, the key objective is to jointly train the AE structure, consisting of transmitter and receiver structures, which contains Tx- and Rx-NNs for each service s to transmit and receive symbols

$$\mathbf{b}_s \stackrel{!}{=} \hat{\mathbf{b}}_s, \tag{5}$$

The CE loss maximizes the Mutual Information between $\mathbf{b}_s$ and $\hat{\mathbf{b}}_s$ hence optimizing to (5) following the maximum likelihood objective. We summarize the overall trainable parameters as $\mathbf{\Theta} = \{\mathbf{\Theta}_1^t, \mathbf{\Theta}_2^t, \mathbf{\Theta}_1^r, \mathbf{\Theta}_2^r, \alpha_1, \alpha_2\}$. Recall that $\alpha_1 + \alpha_2 = 1$ is enforced during the training process. Recapping the overall scheme (Fig. 1), we can observe that an identity mapping is formed, as vectors $\mathbf{b}_s$ are fed into the scheme and reconstructed vectors $\hat{\mathbf{b}}_\mathbf{s}$ are the desired output, which is a key feature of an AE. We want to emphasize that our approach is fully modular, e.g., one can replace the Tx-NN$_s$ of the transmitter with regular modulation schemes and retrain our approach again. Furthermore, an extension for more than two services is straightforward. An important remark here is that we do not need to retrain our approach for each transmission, we train once for a certain Signal-to-Noise-Ratio (SNR) range and simply apply the learned NOMA-AE for inference afterward.

III. Results

To the best of our knowledge, there are no comparable NN approaches for NOMA yet. The closest we found was [5], where a 2×2 MIMO system was used in a concurrent AE fashion. As we have a 1×2 Single-Input-Multiple-Output (SIMO) system, this is not directly comparable to our case. Hence, we limit the comparison of NOMA-AE to a SotA NOMA scheme. The SotA scheme consists of $S = 2$ services using optimal power allocation factors $\boldsymbol{\alpha} = \begin{pmatrix} 0.75 & 0.25 \end{pmatrix}$ as we know this leads to an equidistant signal constellation [4], [10].

For training of our proposed NOMA-AE framework, we generate 1e6 samples of $\mathbf{b}_s$ for our dataset, a training SNR of 15 or 18dB and maximum number of 10000 epochs are used. We save the best weights with the lowest loss. The training SNR has been carefully chosen after different experiments to find a good balance between noise and reconstruction capability of the NNs to achieve the best performance. During training, we use Adam optimizer [11] with default learning rate of 0.001.

For all our comparisons, service 1 wants to transmit $k_1 = 2$ bits per symbol, yielding $M_1 = 4$ symbols and service 2 wants to transmit $k_2 = 4$ bits per symbol, i.e. $M_2 = 16$. Therefore, in the resulting superimposed transmit signal $\mathbf{x}$, $4 \cdot 16 = 64$ total symbols are used during transmission. These numbers are chosen to get a first impression of the performance of the proposed NOMA-AE scheme.

In general, there are no restrictions on M_s. For testing purposes, new test datasets are generated to make sure no overfitting occurred. To measure performance, the BLER

$$\text{BLER} = \frac{\text{\# of false classes}}{\text{\# of all classes}} \tag{6}$$

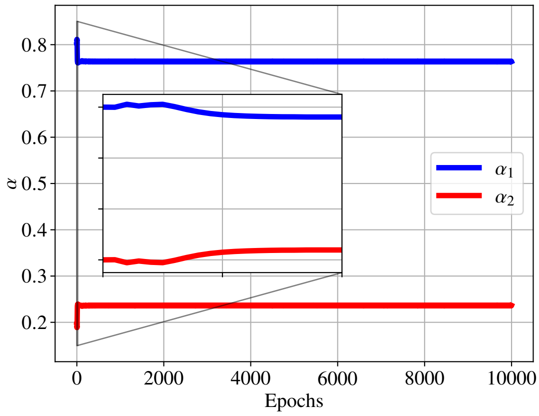

Fig. 2. **Case A**: Power allocation factors α over training epochs.

is shown over various SNR values. Each class refers to one of the one-hot-vectors in $\mathbf{b}_s$. E.g. for service 1 with $M_1 = 4$, 4 classes would exist. Please note that for inference no retraining is applied, the NOMA-AE is simply executed for each SNR and the BLER is stored.

As mentioned in the system overview, we have a modular transmitter concept where we can exchange blocks. For our analysis we will distinguish between the following cases:

- **Case A** – no Tx-NN$_s$, but 2 fixed modulations, 4-QAM and 16-QAM, where only α_s is trainable on the transmitter side
- **Case B** – one of the fixed modulations is replaced with a Tx-NN, so either 4 or 16 symbols are learned
- **Case C** – both transmit services will be handled by Tx-NN$_s$ each

We always have RX-NN$_s$ in place for each of the cases but the transmitter changes for each case. Only cases B) and C) have Tx-NN$_s$, After training, the Tx-NN$_s$ can be replaced by simple lookup tables, simply mimicking the functions of each Tx-NN$_s$.

A. Case A

We show the power allocation factors α_s over the number of training epochs in Fig. 2.

The power allocation factors end up close to the same value ($\alpha_1 \approx 0.76$ and $\alpha_2 \approx 0.24$) as the equidistant allocation of the SotA NOMA scheme we use for comparison. Therefore, the learned power factors are close to optimal [10]. As expected, the resulting learned power factors form a 64-quadrature amplitude modulation (QAM).

Finally, we show the BLER performance over the SNR and compare to classical NOMA scheme in Fig. 3. The NOMA-AE outperforms classical NOMA for all SNR and both services, as SIC is suboptimal. Another benefit of our approach is that we can estimate both bitstreams in parallel in comparison to a classical SIC, as we are not required to detect the strongest service first, but this comes at the cost of increased computational complexity.

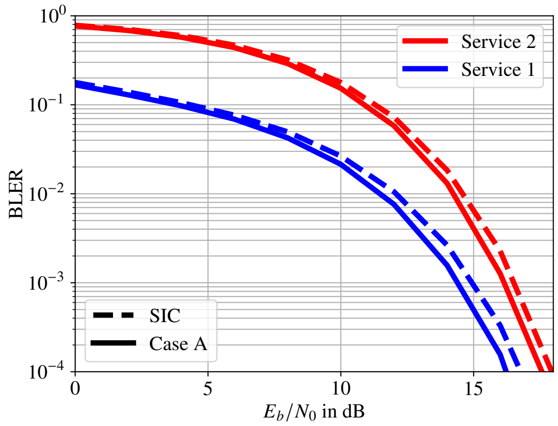

Fig. 3. **Case A**: BLER performance vs SNR for proposed NOMA-AE. The proposed scheme outperforms classical NOMA for both services and all SNR.

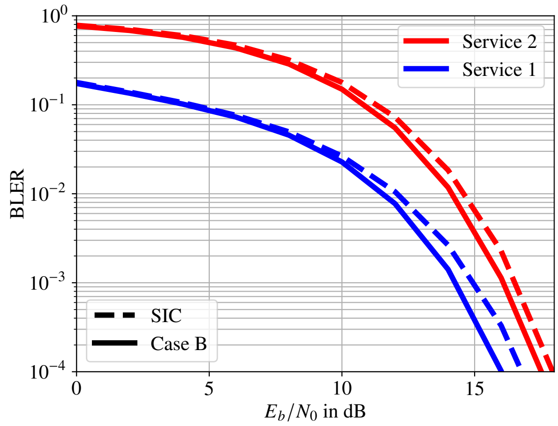

Fig. 5. **Case B**: BLER over E_b/N_0 in dB. The proposed scheme outperforms classical NOMA schemes for both services and all SNR.

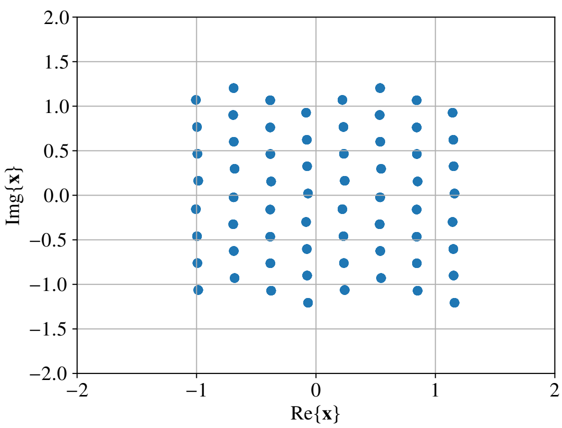

Fig. 4. **Case B**: Superimposed transmit signal **x** of 2 services, resulting in 64 total transmitsymbols. Service 1 is fixed to 4-QAM and the modulation of service 2 is learned by Tx-NN$_2$ with 16 symbols

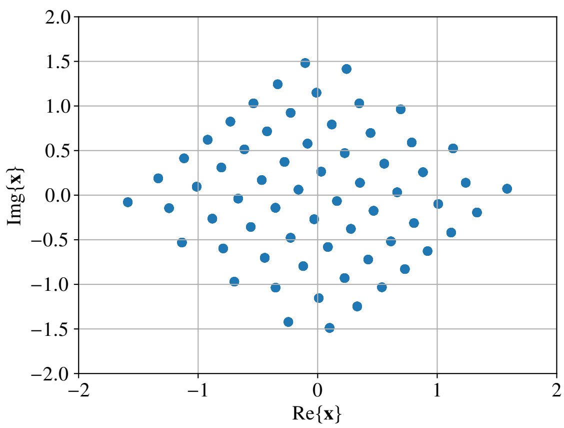

Fig. 6. **Case C**: Superimposed transmit signal **x** of 2 services, resulting in 64 total transmitsymbols. Service 1 uses Tx-NN$_1$ learning 4 symbols and service 2 uses Tx-NN2 for 16 symbols

B. Case B

Exemplary, we fix service 1 to 4-QAM and make the 16 symbols of service 2 trainable. The resulting power allocation factors α_s are very similar to case A, being again ($\alpha_1 \approx 0.75$ and $\alpha_2 \approx 0.25$), which is an expected behavior as we fix one of the modulations. In Fig. 4 the superimposed, partly learned modulation is shown.

We can see that the trainable modulation for the 16 symbols form a hexagonal modulation which is, as Forney stated, more efficient than classical QAM constellations [12]. QAM with a square shape, maximizes the minimal Euclidean distance, which enhances scalability and symmetries. However, it comes at the cost of a relatively high peak-to-average-power-ratio (PAPR). In contrast, phase shift keying (PSK) is attractive in terms of PAPR, but as the constellation size increases, the minimum Euclidean distance rapidly decreases [13]. Hexagonal constellations have the advantage of having the densest 2D packing among all existing constellations. This leads to a reduction in the PAPR of the constellation, making hexagonal constellations more power efficient compared to other constellations [14]. As a result, in the BLER performance in Fig. 5, we can see that the performance improves in comparison to case A for both services and high SNR. For low SNR, the performance is nearly the same for both services, as the noise is more dominant and no clear advantage can be gained.

C. Case C

For a first impression, we show the scatter plot of the learned superimposed constellation in Fig. 6. Again, a hexagonal modulation is learned. Furthermore, we observe that the learned modulation is nearly equidistant, but not symmetrical wrt. the axes.

In Fig. 7 the learned modulation schemes for each service is shown. We observe a 90° phase shift between service 1 and 2. This is similar to the findings in [5], where a MIMO system

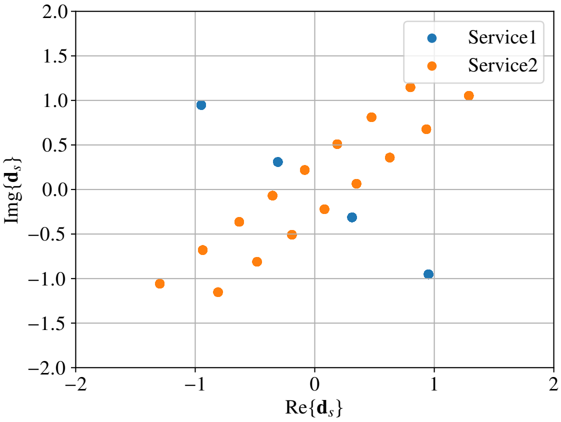

Fig. 7. **Case C**: Learned modulations for each service. Service 1 is shown in blue, Service 2 in orange. A phase shift of approximately 90° is visible.

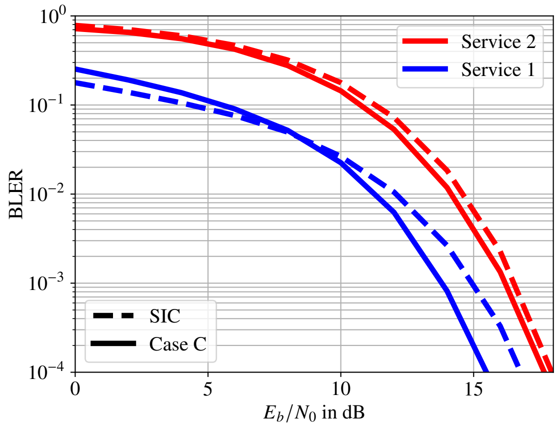

Fig. 8. **Case C**: BLER over SNR for proposed NOMA-AE and classical NOMA.

was applied. Hence, the learned modulations do not interfere destructively for certain power allocation factors, as it was for case A and B. Furthermore both α are close to 0.5 (not shown), which indicates that the learned constellations compensate for the different power allocation factors, if compared to case A and B.

Concluding the findings for this case, the BLER results in Fig 8 indicate that the NOMA-AE with combined learning of both constellations outperform traditional NOMA with SIC in case of Service 2 for all SNR. Service 1 on the other hand, is only superior for high SNR, but the gains are larger compared to the other cases. For low SNR the performance is slightly worse. This is due to the hexagonal modulations, as they are superior against noise influences and SIC is not performing as good as the Rx-NN$_s$, which can already be seen in case A. The performance loss for low SNR is explainable due the nature of smaller decision regions on the receiver side for service 1 as the modulations are phase shifted 90° to each other. For high SNR this is not an issue anymore and leads

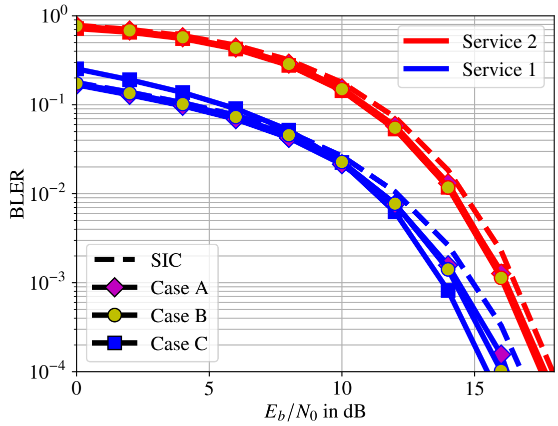

Fig. 9. Comparison of all proposed cases.

to performance gains. On top, the performance degradation is explainable with the high training SNR of 18dB, leading to degraded performance for low SNR.

D. *Comparison of all cases*

Finally, we show the performance of all cases in Fig. 9 to provide a good overview of the different settings. As expected, the performance of all cases is similar. Case C features the most flexibility and can hence perform better for high SNR, as this was the training SNR range. Although the other cases have been trained for the same SNR, the flexibility of them is reduced and hence the possible gains are limited. Case B performs a fraction better than case A which is reasonable, as in case B one modulation is fully learned and adapted. Overall, all cases outperform classical NOMA schemes for all services and SNR, except for case C and service 1 for low SNR, which shows that Rx-NN receiver yields performance gains for NOMA. The usage of Tx-NN can yield slight performance increase on top.

IV. Summary and Outlook

In this paper, we propose a novel approach to design a NOMA system by the application of Deep Learning and neural networks in an Autoencoder framework. We show that NOMA-AE outperforms existing SotA NOMA schemes in a generic setup. Furthermore, as we do not rely on SIC, no error propagation is possible. The observed gains from NOMA-AE result from the fact that classical QAM modulations are not optimal with respect to noise influence and that we perform an optimal estimation of the superimposed transmit signal. We further verified the proposed equidistant power allocation scheme of the authors of [4] with the evaluations in this paper.

In future investigations the extension to higher data rates per service, the adaptation for more than 2 services, and the transmission over more general channels is planned.

REFERENCES

[1] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[2] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.

[3] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5g systems," *IEEE Communications Letters*, vol. 20, no. 3, pp. 458–461, 2016.

[4] N. Bulk, A. Prinz, C. Bockelmann, and A. Dekorsy, "Equidistant power allocation for a service-based noma scheme," in *27. VDE-ITG-Fachtagung Mobilkommunikation (MKT'23)*, Osnabrueck, Germany, May 2023.

[5] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[6] N. Choubey, A. Trivedi, and V. S. Kushwah, "Autoencoder for end-to-end learning communication system based on NOMA," in *2022 IEEE 6th Conf. on Information and Comm Technology (CICT)*, pp. 1–5.

[7] V. Ninkovic, D. Vukobratovic, A. Pastore, and C. Anton-Haro, "A weighted autoencoder-based approach to downlink NOMA constellation design," no. arXiv:2306.13423. arXiv, 2023.

[8] T. Van Luong, N. Shlezinger, C. Xu, T. M. Hoang, Y. C. Eldar, and L. Hanzo, "Deep learning based successive interference cancellation for the non-orthogonal downlink," vol. 71, no. 11, pp. 11 876–11 888, conf. Name: IEEE Trans. on Vehicular Technology.

[9] T. Cover, "Broadcast channels," vol. 18, no. 1, pp. 2–14.

[10] Z. Yang, W. Xu, C. Pan, Y. Pan, and M. Chen, "On the Optimality of Power Allocation for NOMA Downlinks With Individual QoS Constraints," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1649–1652, Jul. 2017.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," no. arXiv:1412.6980. arXiv, 2017.

[12] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 632–647, 1984.

[13] M. Tanahashi and H. Ochiai, "A multilevel coded modulation approach for hexagonal signal constellation," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 4993–4997, Oct. 2009.

[14] P. K. Singya, P. Shaik, N. Kumar, V. Bhatia, and M.-S. Alouini, "A Survey on Higher-Order QAM Constellations: Technical Challenges, Recent Advances, and Future Trends," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 617–655, 2021.