A Weighted Autoencoder-Based Approach to Downlink NOMA Constellation Design

Vukan Ninkovic, Dejan Vukobratovic
Faculty of Technical Sciences
University of Novi Sad
Novi Sad, Serbia
{ninkovic, dejanv}@uns.ac.rs

Adriano Pastore, Carles Antón-Haro
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/iCERCA)
Parc Mediterrani de la Tecnologia (PMT), Av. Carl Friedrich Gauss 7, Building B4
08860 Castelldefels, Spain
{adriano.pastore, carles.anton}@cttc.es

Abstract—End-to-end design of communication systems using deep autoencoders (AEs) is gaining attention due to its flexibility and excellent performance. Besides single-user transmission, AE-based design is recently explored in multi-user setup, e.g., for designing constellations for non-orthogonal multiple access (NOMA). In this paper, we further advance the design of AE-based downlink NOMA by introducing weighted loss function in the AE training. By changing the weight coefficients, one can flexibly tune the constellation design to balance error probability of different users, without relying on explicit information about their channel quality. Combined with the SICNet decoder, we demonstrate a significant improvement in achievable levels and flexible control of error probability of different users using the proposed weighted AE-based framework.

Index Terms—Deep autoencoders, Non-orthogonal multiple access, Successive interference cancellation

I. INTRODUCTION

Improving the resource usage efficiency of 5G and beyond-5G wireless systems is essential to accommodate increasing user requirements. Non-orthogonal multiple access (NOMA) is a recent addition to the wireless signal processing toolset that promises further advances in achieving improved spectrum usage under additional benefits of reduced latency, increased throughput, higher connection density and improved fairness [1], [2]. The main idea behind NOMA is that multiple users can be served by the same resources if appropriate joint signal encoding (e.g., using superposition encoding principle) and efficient signal decoding (e.g., using successive interference cancellation) is applied [3], [4].

Recent trends see shifting the design of encoding and decoding procedures from conventional to machine learning (ML)-based methods. The trend is initiated in the domain of point-to-point communication systems [5], [6], but has since expanded to multi-user NOMA setup [7]–[10]. In this paper, we follow this trend, and use a deep autoencoder (AE)-based approach to design encoding and decoding solution

for downlink NOMA. Building upon the work in [8], we apply: 1) a weighted loss function to control error probability balance across different users, 2) SICNet architecture [7] to enhance deep AE-based decoding capability. Using the proposed weighted AE approach, we are able to obtain significant improvement and flexibility in the error rate performance, as evidenced by simulation experiments.

The paper is organised as follows. In Sec. II, we provide background on generic NOMA system, and recent ML-based methods we use in our work. In Sec. III, we represent NOMA system as end-to-end AE-based system and introduce our weighted AE approach combined with SICNet. Performance evaluation and comparison with the baseline AE-based NOMA is presented in Sec. IV. The paper is concluded in Sec. V.

II. BACKGROUND AND SYSTEM MODEL

A. Downlink NOMA Transmission

We consider a problem of downlink NOMA transmission of L (different) messages from a base station (BS) to L users. The BS jointly encodes user messages into a signal $\mathbf{x}$ and transmits it over n channel uses. We assume a user message $\mathbf{s}_\ell \in \mathbb{F}_2^{k_\ell}, \ell \in \{1, 2, \ldots, L\}$ is a binary (information) sequence of length k_ℓ bits.

Encoding process for the ℓth user can be described as $f_\ell : \mathbb{F}_2^{k_\ell} \rightarrow \mathbb{R}^n$. In other words, the ℓ-th user maps its information sequence $\mathbf{s}_\ell$ into a signal $\mathbf{q}_\ell$ obtained as $\mathbf{q}_\ell = f_\ell(\mathbf{s}_\ell), \mathbf{q}_\ell \in \mathbb{R}^n, \ell \in \{1, 2, \ldots, L\}$. Based on the L signals $(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_L)$, the BS generates the transmitted message $\mathbf{x}$ of length n ($\mathbf{x} \in \mathbb{R}^n$), usually by exploiting generic superposition coding function $F(\cdot)$:

$$\mathbf{x} = F(f_1(\mathbf{s}_1), \ldots, f_L(\mathbf{s}_L)) = F(\mathbf{q}_1, \ldots, \mathbf{q}_L), \quad (1)$$

where in the case of conventional superposition coding the function $F(\cdot)$ represents a weighted linear combination [4]. At the output of the encoder, $\mathbf{x}$ obeys an average power constraint, for which, $\frac{1}{M_c} \sum_{i=1}^{i=M_c} \|\mathbf{x}_i\|_2^2 = n$, where $M_c = 2^{k_1+k_2+\ldots+k_L}$.

Each of the L users has distinct channel conditions, which transforms input message $\mathbf{x} \in \mathbb{R}^n$ into L output sequences $\mathbf{y}_\ell \in \mathbb{R}^n, \ell \in \{1, 2, \ldots, L\}$ following the probabilistic channel model $p(\mathbf{y}_\ell|\mathbf{x})$. In this paper, we consider a set of independent additive white Gaussian noise (AWGN) channels between

the BS and each of the L users characterised by the set of noise variances $\{\sigma_\ell^2\}_{\ell=1,\dots,L}$ [11]. This model is suitable for downlink NOMA in OFDM-based systems, and it can be extended to a set of independent block-fading channels [8].

At the receiver side, an estimated information sequence for the ℓth user is obtained via the function $g_\ell : \mathbb{R}^n \rightarrow \hat{\mathbf{s}}_\ell$, i.e., the decoding process for each user can be defined as $\hat{\mathbf{s}}_\ell = g_\ell(\mathbf{y}_\ell), \ell \in \{1, 2, \dots, L\}$.

The goal is to design a combination of encoders $\{\{f_\ell\}_{\ell=1,\dots,L}, F\}$ and decoders $\{g_\ell\}_{\ell=1,\dots,L}$ that will optimally balance (in a Pareto optimal sense) the set of message error probabilities $\{P_{e_\ell}\}_{\ell=1,\dots,L}$:

$$P_{e_\ell} = \frac{1}{2^{k_\ell}} \sum_{\mathbf{s}_\ell \in \mathbb{F}_2^{k_\ell}} \mathbb{P}\{\hat{\mathbf{s}}_\ell \neq \mathbf{s}_\ell | \mathbf{s}_\ell\} \qquad (2)$$

across the set of L receivers.

B. Power–Domain Downlink NOMA

In the conventional downlink NOMA scheme, the function $F(\cdot)$ defined in Eq. 1 is a weighted linear combination of user signals obtained by allocating different power coefficients as the weights of the linear combination [2], [3]:

$$\mathbf{x} = F(\mathbf{q}_1, \dots, \mathbf{q}_L) = \sum_{\ell=1}^{\ell=L} \sqrt{p_\ell} \mathbf{q}_\ell, \qquad (3)$$

where p_ℓ represents power associated with user ℓ. Without loss of generality, we assume that the power is allocated to users in ascending order $p_1 < p_2 < p_3 < \dots < p_L$ [1]. From the receiver perspective, the ℓth user observes:

$$\mathbf{y}_\ell = h_\ell \mathbf{x} + \mathbf{z} = h_\ell \sum_{\ell=1}^{\ell=L} \sqrt{p_\ell} \mathbf{q}_\ell + \mathbf{z} \qquad (4)$$

where h_ℓ represents complex channel coefficient between BS and user ℓ, while additive white Gaussian noise (AWGN) with variance σ_ℓ^2 is denoted as $\mathbf{z}$. For simplicity and without loss of generality, we suppose that $h_\ell = 1, \ell \in \{1, 2, \dots, L\}$, i.e., we focus on the AWGN channel (Section II-A).

A combination of superposition coding at the BS and successive interference cancellation (SIC) decoding at the receiver is commonly used in downlink NOMA [1]. SIC exploits different powers allocated to different users at the transmitter side. The user with the power coefficient p_L (the strongest power coefficient) will be decoded first (treating other users signals as noise) and subtracted from $\mathbf{x}$. This procedure is successively repeated for the remaining users until all user messages are decoded [3], [4]. Note that the SIC performance is dependent on a perfect knowledge of channel state information (CSI) and power allocation coefficients, and may degrade if imperfect CSI is used [12].

Fig. 1. Communication system represented as a deep autoencoder: a) Conventional system - Blue blocks [5] b) NOMA downlink system with *SICNet* [7] - Red blocks; Black blocks are used by both a) and b)

C. SICNet

In order to overcome the above-mentioned problem of the conventional SIC receiver, the authors in [7] proposed a deep neural network (DNN)-based decoding algorithm, entitled *SICNet*, that incorporates DNNs to recover user messages. *SICNet* preserves original SIC decoder structure and algorithmic flow, however, each block is replaced with a DNN that performs classification task [7]. Output of each DNN can be used as a soft estimate (by using softmax activation function) and concatenated with the input of the subsequent DNN. In such a way, the input to the *SICNet* of ℓth user is its received signal $\mathbf{y}_\ell$ concatenated with soft estimates of users $\ell+1$ to L. This approach avoids the explicit need for prior CSI and power allocation knowledge [7].

D. Autoencoder–Based Communication Systems

The conventional point-to-point communication system with a transmitter and a receiver has been redesigned from the deep learning perspective in the form of an end-to-end autoencoder (AE) [5], as illustrated in Fig. 1a (blue blocks).

An input message m, which can be represented as a sequence of k bits $\mathbf{s} = (s_1, s_2, \dots, s_k)$ (where $k = \log_2 M$), is encoded as a one-hot vector $\mathbf{u} = (u_1, u_2, \dots, u_M) \in \{0, 1\}^M$. The transmitter, described by $f_{\text{AE}} : m \rightarrow \mathbf{x}$, is represented as a feed-forward neural network with H hidden layers, followed by a bottleneck layer of width n. At the output of transmitter, a batch normalization layer ensures that the average power constraint on $\mathbf{x}$ are met (Section II-A).

The AWGN channel is constructed as an additive noise layer with output $\mathbf{y} = \mathbf{x} + \mathbf{z}$ (as it is shown in Fig. 1, blue receiver), where $\mathbf{z}$ contains n independent and identically distributed samples of a Gaussian random variable with zero mean and variance σ^2.

The receiver, described as $g_{\text{AE}} : \mathbf{y} \rightarrow \hat{m}$, is implemented as a feed-forward neural network with H hidden layers and softmax activation function at its output, $\mathbf{b} = (b_1, b_2, \dots, b_M) \in$

$(0, 1)^M$, $\|\mathbf{b}\|_1 = 1$. All other hyperparameters are the same as in the AE transmitter part. The receiver takes the channel output $\mathbf{y}$ as its input and produces a message estimate $\hat{m}$ as $\hat{m} = \arg\max_i \{b_i\}$.

Except the output layer of the transmitter and receiver that have linear and softmax activation function, respectively, all other layers are activated using rectified linear units (ReLU).

The presented AE is trained in an end-to-end manner to minimize the message error probability P_e (Eq. 2), by using stochastic gradient descent with Adam optimizer [13]. In other words, f_{AE} and g_{AE} are jointly optimized. In order to optimize P_e (and parameters of the AE), the minimization of categorical–cross entropy between $\mathbf{u}$ and $\mathbf{b}$ is used as a surrogate for the message error probability [5]:

$$\ell(\mathbf{u}, \mathbf{b}) = -\sum_{i=1}^{M} u_i \log b_i, \qquad (5)$$

III. Autoencoder–Based Downlink (DL) NOMA

In this section, we present a flexible and efficient way to learn encoding and decoding strategy for NOMA downlink transmission. Conventional NOMA downlink communication system with SIC decoding, described in Sections II-A and II-B, can be implemented in a DNN fashion as an extension of the end-to-end AE-based scheme presented in Section II-D, as illustrated in Fig. 1b (red blocks) [8].

Starting from the superposition coding at the transmitter side, a composition of the function $F(\cdot)$ and single-user encoding functions $\{f_\ell\}_{\ell=1,2,\ldots,L}$ (Eq. 1), is replaced with a single learning process. More precisely, an encoder is jointly optimized with the set of L decoders using an end-to-end AE-based approach, i.e., the function pairs $(f_{\text{AE}}, \{g_{AE_\ell}\}_{\ell=1,2,\ldots,L})$ are obtained using an end-to-end training procedure. Individual users messages $\mathbf{s}_\ell \in \mathbb{F}_2^{k_\ell}, \ell \in \{1, 2, \ldots, L\}$, are represented as a message index from a set $\mathcal{M}_\ell, |\mathcal{M}_\ell| = M_\ell = 2^{k_\ell}$, and one-hot encoded, where $\mathbf{u}_\ell$ denotes one-hot encoding form of the ℓth user message $\mathbf{s}_\ell$, and $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L\}$. AE input message m is defined as a concatenation of individual user messages in their bit representations, i.e., $m = [\mathbf{s}_1; \mathbf{s}_2; \ldots; \mathbf{s}_L]$ (Fig. 1b) [8]. One-hot encoding of the AE message m maps $k_C = k_1 + k_2 + \ldots + k_L$ bits into a one-hot encoding $\mathbf{u}_C = (u_1, u_2, \ldots, u_{M_C}) \in \{0, 1\}^{M_C}$ where $M_C = 2^{k_C}$. The rest of the transmitter is the same as in Section II-D (Fig. 1, black blocks).

The transmitted message $\mathbf{x} = f_{\text{AE}}(m)$ is passed through L different AWGN channels whose outputs $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L$ are available at the respective receivers. Without loss of generality, we suppose that each of the L channels has different signal-to-noise ratio (SNR) and $SNR_1 > SNR_2 > \ldots > SNR_L$.

At the receiver side, each of the L users estimates the corresponding message through the feed-forward neural network, described in Section II-D (there are in total L DNN receivers, one per each user). In order to implement *SICNet* [7] (Section II-C), soft output of each preceding users' DNN is connected to the input of the next user (Fig. 1-red block

receiver), where the decoding process starts from the user with the most degraded channel ($\mathbf{y}_L$). The softmax output of all L users is collected into $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_L\}$, where each $\mathbf{b}_\ell$ is of length $M_\ell = 2^{k_\ell}, \ell \in \{1, 2, \ldots, L\}$.

The goal is to learn a pair $(f_{\text{AE}}, \{g_{AE_\ell}\}_{\ell=1,2,\ldots,L})$ that minimizes the individual user message error probabilities $\{P_{e_\ell}\}_{\ell=1,2,\ldots,L}$ (Eq. 2). As in Section II-D, minimization of P_{e_ℓ} is replaced with the loss function based on the categorical-cross entropy, defined for the ℓth user as:

$$\ell_\ell(\mathbf{u}_\ell, \mathbf{b}_\ell) = -\sum_{i=1}^{M_\ell} u_{\ell_i} \log b_{\ell_i}, \qquad (6)$$

The total loss function for the AE-based downlink NOMA is the sum of the individual users' loss contributions:

$$L(\mathbf{U}, \mathbf{B}) = \sum_{j=1}^{L} \ell_j(\mathbf{u}_j, \mathbf{b}_j) \qquad (7)$$

A. Weighted Autoencoder-Based Design of Downlink NOMA

In order to train AE-based downlink NOMA communication system described in the previous section, the SNR difference ΔSNR between subsequent receivers is needed as one of the hyperparameters of the training process. As channel conditions may change during the testing phase, this may lead to a potential performance degradation.

In our previous work [14], we explored a novel class of autoencoders that utilise a compound loss function in the context of single-user unequal error protection (UEP) coding. Such AE method is able to flexibly balance between error probabilities among the set of messages (message-wise UEP) or between specific subblocks of different importance classes within a single message (bit-wise UEP). Inspired with this approach, a similar weighted sum approach can be used in the AE-based downlink NOMA by introducing weights associated to L different users (Eq. 7), leading to a weighted total loss function:

$$L(\mathbf{U}, \mathbf{B}) = \sum_{j=1}^{L} \lambda_j \ell_j(\mathbf{u}_j, \mathbf{b}_j), \qquad (8)$$

where λ_j represents a weight factor associated to the user j, $\sum_{j=1}^{L} \lambda_j = 1$, and $\lambda_j \geq 0$.

Loss function defined in Eq. 7 is a special case of the weighted sum presented in Eq. 8, where $\lambda_\ell = \frac{1}{L}, \ell \in \{1, 2, \ldots, L\}$. By incorporating the above mentioned compound loss function (Eq. 8), all users can be trained assuming they experience the same SNR (thus disregarding ΔSNR as a hyperparameter) while adjusting the desirable users' performance using $\lambda = \{\lambda_1, \ldots, \lambda_L\}$. Although a simple alteration of loss function, the proposed approach introduces a single "knob" (parameter λ) one can tune to design a family of NOMA constellations that progressively balance the users' error probabilities. The influence of λ on the system performance is elaborated in the next section.

Fig. 2. Different architectures (P_{e_1}, P_{e_2}) versus SNR performance comparison for $\lambda = \{0.5, 0.55, 0.6\}$ in NOMA downlink transmission with two users ($k_1 = 2$, $k_2 = 2$, $n = 2$, $\Delta SNR = 9$ dB).

Fig. 3. λ influence on (P_{e_1}, P_{e_2}) in two users downlink NOMA transmission ($k_1 = 2$, $k_2 = 2$, $n = 2$, $SNR_1 = 12$ dB, $\Delta SNR = 9$ dB).

IV. PERFORMANCE EVALUATION OF THE WEIGHTED AE–BASED DOWNLINK NOMA

A. Training Procedure

For simplicity, and without loss of generality, the number of users in all conducted experiments is restricted to two ($\lambda = \{\lambda_1, \lambda_2\} = \{\lambda, 1 - \lambda\}$), and $k_1 = k_2 = k$ (number of messages is $M_c = 2^{k_1 + k_2} = 2^{2k}$). Apart from the introduction of the suitable loss function (Eq. 7 and 8), the same training procedure as presented in [5] is preserved, i.e., AE-based NOMA downlink system is optimized by using stohastic gradient descent with Adam optimizer [13] (learning rate $\alpha = 0.0009$, $\beta_1 = 0.9$, $\beta_2 = 0.999$). Regarding the number of bits associated to the each user, for both encoder and each of two decoders, single fully–connected hidden layer is considered ($H = 1$) with M_c neurons. Batch size is set to 3000. When Eq. 7 is used as a loss function, training is performed at $\Delta SNR = SNR_1 - SNR_2 = 9$ dB, where $SNR_1 = 15$ dB. With introduction of λ (Eq. 8) both decoders are trained on the same $SNR_1 = SNR_2 = 6$ dB. During the testing phase, ΔSNR is preserved at 9 dB for all conducted experiments. Training and test data sets contain 10^5 and 2×10^6 messages sampled at random from $\{1, 2, \ldots, M_c\}$, respectively.

B. Comparison with the State-of-the-Art

1) Case 1 - $(k, n) = (2, 2)$: In order to examine performances of the proposed approach, the two-user scenario and corresponding architecture form [8] is recreated. More precisely, BS jointly encodes $k_1 = k_2 = k = 2$ bits per user ($M_c = 2^{2k} = 16$) and sends it over $n = 2$ channel uses.

In Fig. 2 we compare performances obtained with the proposed approaches and architecture replicated from [8]. Channel conditions for the second user can be obtained as

$SNR_2 = SNR_1 - \Delta SNR$, where SNR_1 is presented on the Fig. 2 x-axis. Significant performance improvement in terms of error performance of both users can be observed with the introduction of the *SICNet* [7], compared to the baseline results in [8]. Moreover, use of compound loss function (Eq. 8) by incorporating non-uniform users weights in the loss function (parameter λ) provides for a flexible system design, where a desirable trade-off between the two users' performance can be easily controlled through manipulation of λ value.

Influence of different λ values, introduced during the training process, on system performances is illustrated in Fig. 3, where we plot (P_{e_1}, P_{e_2}) against $\lambda = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$. As the value of λ increases, the loss function (Eq. 8) begins to favour the first user. This is reflected in a graceful improvement of P_{e_1} on Fig. 3, simultaneously with the graceful degradation of P_{e_2}. By tuning λ, the system can be adapted to a different requirements and variable channel conditions (ΔSNR). The testing phase is done on $SNR_1 = 12$ dB and $\Delta SNR = 9$ dB, i.e., although the second user is tested on $SNR_2 = 3$ dB and its slope is gentler, it still outperforms the first user for λ values below 0.3 (Fig. 3).

Learned messages **x** are two-dimensional real values ($\mathbf{x} \in \mathbb{R}^2$), thus they can be visualized as symbols (or points) in 2D "constellation" (Fig. 4). Fig. 4 illustrates the influence of λ on learned constellation. We observe that the distance between messages associated to user 1 increases with increasing λ (loss function favors the first user), while the opposite happens to the second user (shape of the constellation transforms from the square ($\lambda = 0.5$) to a rectangle ($\lambda = 0.6$)).

2) Case 2 - $(k, n) = (4, 4)$: Most of the existing works targeting AE-based downlink NOMA restrict their attention to the case of two-bit user messages ($k = 2$) [8], [15]. Exploring how learned NOMA will behave in higher dimensions, we further expand the number of bits per user to $k_1 = k_2 = k = 4$ ($M_c = 2^{2k} = 256$) and the number of channel uses to $n = 4$.

Fig. 4. λ influence on learned constellations for two users NOMA downlink transmission ($k_1 = 2$, $k_2 = 2$, $n = 2$, $M_c = 2^{k_1+k_2} = 16$).

Fig. 5. Different architectures (P_{e_1}, P_{e_2}) versus SNR performance comparison for $\lambda = \{0.5, 0.55, 0.6\}$ in NOMA downlink transmission with two users ($k_1 = 4$, $k_2 = 4$, $n = 4$, $\Delta SNR = 9$ dB).

Although the number of messages grows exponentially, the proposed approach significantly outperforms an architecture from [8] (Fig. 5). More precisely, from Fig. 5 we observe that *SICNet* receiver noticeably improves the performance of the second user (when ΔSNR is known), while the first user performance remains comparable to [8]. On the other hand, with the introduction of the compound loss function (λ, Eq. 8), the proposed approach significantly outperforms [8] in terms of the first user performance (Fig. 5, green and cyan solid lines), with the price of slight performance degradation for the second user.

We note that moving to higher constellation dimensions (i.e., from $n = 2$ to $n = 4$), although incurs higher complexity ($M_c = 256$), leads to a considerable improvement of error performance of both users. It also leads to faster saturation of the second user performance with the increase of λ (i.e., already for $\lambda > 0.55$ on Fig. 5).

V. CONCLUSION

In this paper, we presented a flexible and efficient weighted AE-based method for design of downlink NOMA constellations. The method demonstrates promising performance under simplicity in training and tuning to the desired error probability balance between users. In the future work, we plan explore a combination of outer low-density parity-check codes (LDPC) with inner weighted AE-based dowlink NOMA constellations.

REFERENCES

[1] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond," *Proc. of the IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017.

[2] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

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

[4] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition Coding Strategies: Design and Experimental Evaluation," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 7, pp. 2628-2639, July 2012.

[5] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, Dec. 2017.

[6] S. Doerner, S. Cammerer, J. Hoydis, S. Ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing,* vol. 12, no. 1, pp. 132-143, January 2017.

[7] 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," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11876-11888, Nov. 2022.

[8] F. Alberge, "Constellation design with deep learning for downlink non-orthogonal multiple access," 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy, 2018, pp. 1-5.

[9] M. Kim, N. -I. Kim, W. Lee, and D. -H. Cho, "Deep Learning-Aided SCMA," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 720-723, April 2018.

[10] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep Learning for an Effective Nonorthogonal Multiple Access Scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440-8450, Sept. 2018.

[11] S. -L. Shieh, C. -H. Lin, Y. -C. Huang, and C. -L. Wang, "On Gray Labeling for Downlink Non-Orthogonal Multiple Access Without SIC," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1721-1724, Sept. 2016.

[12] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the Performance of Non-orthogonal Multiple Access Systems With Partial Channel Information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654-667, Feb. 2016.

[13] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. on Learn. Representation,* San Diego, CA, USA, May 7-9, pp. 1-41, 2015.

[14] V. Ninkovic, D. Vukobratovic, C. Haeger, H. Wymeersch, and A. Graell i Amat, "Autoencoder-Based Unequal Error Protection Codes," *IEEE Commun. Lett.,* vol. 25, no. 11, pp.3575-3579, Nov. 2021.

[15] T. Van Luong, Y. Ko, N. A. Vien, M. Matthaiou, and H. Q. Ngo, "Deep Energy Autoencoder for Noncoherent Multicarrier MU-SIMO Systems," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 6, pp. 3952-3962, June 2020.