# Learning-Based Autoencoder for Multiple Access and Interference Channels in Space Optical Communications

Abdelrahman Elfikky, Morteza Soltani, *Member, IEEE*, and Zouheir Rezki, *Senior Member, IEEE*

***Abstract*— This letter studies an autoencoder (AE) model for a multi-user space optical communications (SOC) system. In particular, a novel layered framework is proposed with batch normalization in the encoder and layer normalization in the decoders. This enhances the performance of the standard AE since the error cost function can be effectively minimized using gradient descent. Furthermore, to emulate a practical SOC channel model, a system tool kit simulator is used. The numerical results reveal that the proposed AE outperforms state-of-the-art learning frameworks by a 2 dB gain in multiple access channel and a 1.7 dB gain in interference channel, with a 20% complexity reduction.**

***Index Terms*— Space optical communications, autoencoder, multiple access channel, interference channel.**

## I. Introduction

IN RECENT years, radio frequency (RF) systems have grown rapidly and will continue to play a significant role in the future. However, due to increased demand for data-intensive wireless communications, the RF spectrum is becoming limited and expensive. Optical wireless communication (OWC) and space optical communication (SOC) are alternative methods of RF that provide various advantages, including high data speed, low implementation cost, and the utilization of license-free spectrum. These factors make OWC and SOC viable alternatives to RF communication [1].

SOC and OWC are similar in that both employ lasers as optical transmitters. The conventional photo-detector (PD) is not compatible with SOC due to the absence of a receiving telescope in its design [1]. After the laser beam travels a distance, it is collected by the receiving telescope and concentrated towards the detector [1]. In SOC, high-power laser transmitters are required due to large distances up to 36000 km range. The laser transmitter needs to have a high photon efficiency and peak power capability for the downlink from geostationary (GEO) to the ground station [1]. Moreover, a narrow bandwidth, and low modulation rates are necessary for the downlink lasers in SOC.

Researchers are increasingly exploring the use of deep learning (DL) techniques in physical layer communication networks. This integration has led to significant progress in coding and decoding, modulation classification, channel estimation, and equalization [2], [3], [4], [5]. Auto-encoders (AEs) have demonstrated strong performance as an end-to-end (E2E) approach in OWC, particularly in point-to-point space optical communication (SOC). A study on AE-based OWC systems [4] showed that AEs can outperform hamming codes for bit error rate (BER), even in the presence of imperfect channel state information (CSI). In [5], AEs were shown to optimize both transmitter and receiver components, achieving satisfactory BER performance across interference channel (IC). In [3], the feasibility of utilizing AEs for multiple access (MAC) channel in OWC systems was explored and compared to uncoded time sharing. Additionally, sparse AE is designed using sparsity regularization techniques to improve the BER performance in the MAC channel while reducing the computational complexity [6].

In [3], only an AWGN channel without fading is considered and the system is not generalized to IC. Also authors in [3] and [4] utilize similar structure as the standard AE presented in [5]. In [6], while the system has low complexity structure due to the use of sparsity, the BER performance needs to be improved compared with model-based schemes. Although the AEs described in [3] for the multiple access (MAC) channel and [5] for the IC have demonstrated promising performance in the multi-user settings, they still have certain limitations pertaining to achieving a low BER and managing higher computational complexity. On the other hand, the sparse AE approach proposed in [6] focuses primarily on reducing the computational complexity in the MAC channel. However, it does not sufficiently address the BER improvement. In [3], [4], [5], and [6], the incorporation of normalization layers in both the encoder and decoder designs was not addressed. However, the normalization layers can play a crucial role in optimizing the error cost function. Improving BER performance in symbol detection for multi-user channels using AEs in SOC is challenging and requires modifications to the standard AE structure. Existing learning frameworks have not achieved significant improvements in BER performance compared to uncoded modulations and/or state-of-the-art coded systems. Our study proposes modifications to the standard MAC-AE structure to enhance symbol detection performance while maintaining low-complexity capabilities. The proposed AE structure has also been utilized in the IC showing a significant BER improvement compared with existing designs. The main contributions can be stated as follows:

- Our proposed AE is built on a new layered framework that employs batch normalization (BN) in the encoder and layer normalization (LN) in the decoders. This approach delivers satisfactory BER performance when compared to

Manuscript received 12 August 2023; revised 29 August 2023; accepted 30 August 2023. Date of publication 6 September 2023; date of current version 11 October 2023. This work was supported by the National Science Foundation (CAREER) under Grant No. 2114779. The associate editor coordinating the review of this letter and approving it for publication was C. Gong. *(Corresponding author: Abdelrahman Elfikky.)*

Abdelrahman Elfikky and Zouheir Rezki are with the Department of Electrical and Computer Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: afikky@ucsc.edu; zrezki@ucsc.edu).

Morteza Soltani is with Qualcomm Inc., San Diego, CA 92121 USA (e-mail: msoltani@qti.qualcomm.com).

Digital Object Identifier 10.1109/LCOMM.2023.3312584

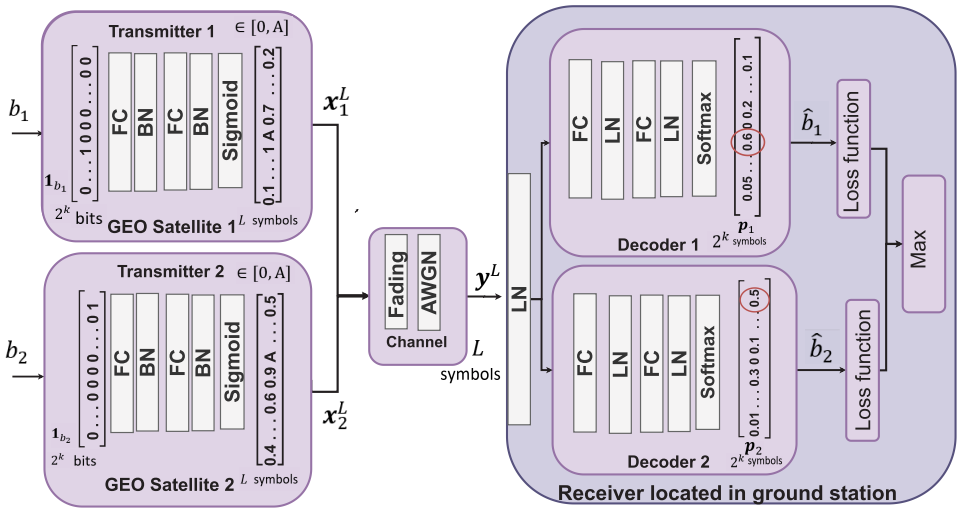



Fig. 1. Proposed AE architecture in a two-user MAC channel.

the latest, state-of-the-art learning frameworks, uncoded and Low-Density Parity-Check (LDPC) time sharing for MAC channels in SOC.

- In addition, we assess the BER performance of our model in a two-user IC and demonstrate the AE sustainability not only in MAC, but also in a more challenging channel such as IC. Our model prioritizes computational complexity while achieving superior BER performance compared to state-of-the-art frameworks. Numerical results compare our model's complexity with existing learning frameworks.

## II. SOC Channel Model

A MAC channel is established between $K$ GEO satellites that transmit signals to a shared receiver located on the ground station. The use of system tool kit (STK) simulator enables accurate modeling of the SOC channels [7]. In the system, the ground station holds the receiver antenna gimbal and avalanche photo-detector. Unlike RF coherent communication, the modulated signal in intensity-modulation and direct-detection (IM/DD) is real and non-negative. Additionally, the signal is peak-constrained in SOC for operation and safety regulations [8]. The Log-normal distribution is commonly used to describe weak atmospheric turbulence, recommended by STK for the GEO to ground SOC channel [2]. In our model, we take into consideration both additive white Gaussian noise (AWGN) and slow fading channels in the multi-user SOC channel. In this particular configuration, we assumed the Gaussian channel; however, when operating at low power levels, a better channel model would be a Poisson channel [8].

## III. Multi-User SOC Channel Based on Autoencoders

The notation $(k, L)$ is used to denote the proposed AE model, where $k$ represents the number of message bits and $L$ represents the codeword length. The proposed AE$(k, L)$ in the SOC system with rate $R = k/L$. The system consists of three distinct modules: the transmitter, receiver, and channel layers. There are $K$ independent transmitters and one shared receiver in this MAC channel. Each transmitter is located in a GEO satellite, while the common receiver is the ground station. Over a SOC channel, transmitter $i \in \{1, 2, \ldots, K\}$ send the message $b_i$ to the common receiver in the ground station, where $b_i \in \mathcal{M} = \{1, \ldots, M\}$, and $M = 2^k$.

*Transmitter:* The transmitter starts by selecting $b_i$, which is one of $M$ possible messages and then converts it into a one-hot vector $\mathbf{1}_{b_i}$ of size M, which has a 1 in the message index position and 0's elsewhere. Utilizing a one-hot encoded vector ensures equal significance for all messages since they are all in binary rather than ordinal format. Then, using the mapping function $\boldsymbol{u} : \mathcal{M} \rightarrow \mathbb{R}^L$, each transmitter converts the input one-hot vector $\mathbf{1}_{b_i}$ into the encoded vector $\boldsymbol{x}_i^L$. Obviously, each encoder applies both modulation and channel coding simultaneously. The codebook consists of all possible codewords generated by the encoder of the AE, i.e., the set $\{\boldsymbol{x}_i^L\}$, with cardinality $2^k$. When the transmitter normalization stage outputs the symbol vector $\boldsymbol{x}_i^L$, it meets the non-negativity and peak conditions for SOC. To meet the constraints $0 \leq \boldsymbol{x}_i^L(l) \leq A$ with $l = 1, \ldots, L$, each transmitter applies a normalization layer to the transmitted symbols using the weighted sigmoid function. We examine a two-user MAC channel depicted in Fig. 1, without loss of generality.Fully connected (FC) layers provide the basis of the transmitter model, with a subsequent BN layer added directly after each FC layer. The BN normalizes the layer's inputs for each mini-batch, which both shortens the training time and keeps the learning process stable. The gradient explosion issues are also lessened by the BN, according to [9]. Until the weighted sigmoid normalization step in the encoder, it is clear that no input scaling takes place. Since the input scaling does not change between training and the testing, BN parameters learned during training will lead to the same superior performance throughout testing.

*SOC channel model:* The normalized vector $\boldsymbol{x}_i^L$ for both transmitters are fed into the SOC channel. The SOC channel is composed of two components: an AWGN channel, and a Log-normal fading model with a standard deviation of $\sigma$.

The channel parameters are provided by the STK simulator discussed in Section II. The input to the neural networks on the receiver's side is represented as $\boldsymbol{y}^L$.

$$\boldsymbol{y}^L = \sum_{i=1}^{K} h_i \boldsymbol{x}_i^L + \boldsymbol{w}^L, \tag{1}$$

where $\boldsymbol{w}^L \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and the Log-normal fading coefficient, represented by $h_i$ for the transmitter $i$.

*Common Receiver:* Lastly, the receiver-based ground station processes the corrupted vector $\boldsymbol{y}^L$ and generates the estimated one-hot vector $\mathbf{1}_{\hat{b}_i}$. The common receiver includes $K$ decoders, each is composed of multiple FC layers followed by Layer Normalization (LN). Since the received symbols $\boldsymbol{y}^L$ will be scaled with different $A$ values than the one used in training, using BN in the decoder can not aid to standardize the data, since the testing input scale differs from the training due to varying $A$ values. BN's learned parameters can not work well in testing unless $A$ matches the trained value.

To prevent this issue, the receiver could wait and collect a certain amount of samples in order to produce the desired results. This implies that the BN will continue to work in the same manner during both training and testing. However, this assumption is impractical in wireless communications. Accordingly, we need to design the system for deployment to deal with received codewords independently without the need for a significant number of samples before the processing begins. Therefore, LN is introduced in the decoders, which operates sample by sample, in order to maintain efficient standardization across the entire AE system. The process of LN involves re-centering and re-scaling its input. Since LN works in the same manner in both training and testing phases, the effect of scaling the input to the decoder is mitigated, and trained decoder weights are effective in testing. Additionally, the corrupted vector $\boldsymbol{y}^L$ undergoes a single LN unit before being inputted to the decoders. This guarantees consistent scaling and enhances system performance during testing across various SNR values. The cross-entropy (CE) loss function for each AE is defined as

$$C_j = -\sum_{r=1}^{M} \mathbf{1}_{b_i}(r) \log \boldsymbol{p}_j(r), \tag{2}$$

where $\mathbf{1}_{b_i}(r) \in \{0, 1\}, i = j$ is the $r^{\text{th}}$value in the input one-hot vector $\mathbf{1}_{b_i}$. The softmax is applied at the last layer of the $j^{\text{th}}$ decoder producing the probability value $\boldsymbol{p}_j(v)$ and it can be expressed as follows

$$\boldsymbol{p}_j(v) = \frac{e^{\boldsymbol{q}_j(v)}}{\sum_{t=1}^{M} e^{\boldsymbol{q}_j(t)}}, \tag{3}$$

where $\boldsymbol{q}_j$ is the $j^{\text{th}}$ decoder final layer output. By employing the min-max algorithm, it is possible to decrease the error probability $\max\left(\Pr\left\{b_1 \neq \hat{b}_1\right\}, \Pr\left\{b_2 \neq \hat{b}_2\right\}, \ldots, \Pr\left\{b_K \neq \hat{b}_K\right\}\right)$. The use of the min-max algorithm can lead to a decrease in error probability by selecting a decision rule that aims to minimize the maximum possible loss [3]. In light of this, in a given step, only the weights corresponding to the maximum loss are updated, while all other weights are left intact. Overall, the system's loss function is given by

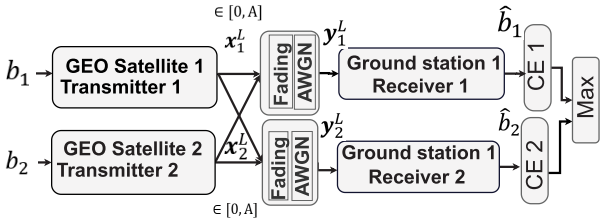



Fig. 2. The proposed AE architecture in a two-user IC setting.

$$C = \max\left(C_1, C_2, \ldots, C_K\right). \tag{4}$$

Algorithm 1 summarizes the learning strategy for the novel AE design in the MAC channel.

*Interference channel:* The idea of MAC AE can be easily expanded to incorporate multiple transmitters and receivers that utilize a shared channel. Both transmitter-receiver pairs are implemented as NNs and the difference with respect to the MAC AE is that the encoded vectors $\boldsymbol{x}_i^L, \forall i \in \{1, 2, \ldots, K\}$ now interfere at the receivers, resulting in the noisy observations

$$\boldsymbol{y}_j^L = \sum_{i=1}^{K} h_{ij} \boldsymbol{x}_i^L + \boldsymbol{w}_j^L, \forall j \in \{1, 2, \ldots, K\} \tag{5}$$

The Log-normal fading coefficient, represented by $h_{ij}$, refers to the coefficient that connects the $i^{th}$ transmitter with the $j^{th}$ receiver, where $i$ and $j \in \{1, 2, \ldots, K\}$. The target for receiver $j$ now is to estimate the message $b_i$ transmitted by the $i^{\text{th}}$ transmitter, while ignoring any other messages that may interfere with it. Furthermore, since the decoders are located at different ground stations, each decoder will have a separate LN unit at its input to negate the effect of scaling $\boldsymbol{y}_i^L$ with different SNR values during testing. Without loss of generality, we consider here the two-user IC as shown in Fig. 2. Transmitter 1 wants to communicate message $b_1 \in \mathcal{M}$ to receiver 1 while transmitter 2 wants to communicate message $b_2 \in \mathcal{M}$ to receiver 2.

## IV. Simulation Results

In this section, we present the symbol detection performance evaluated for the proposed AE for code rate $R = 1/3$ in two different scenarios: MAC and IC with two users. Also, the proposed AE architecture follows the layout given in Figs. 1 and 2. The learnable parameters are tuned using the stochastic gradient descent (SGD) algorithm with the Adam optimizer and a learning rate of $0.0001$. The model is trained using $5,000,000$ randomly generated samples for 20 epochs with the objective of minimizing the training loss. Finally, the system BER performance is evaluated over $1,000,000$ testing samples. The computational complexity of the system can be optimized by assigning the appropriate number of learnable parameters for each layer. We make our source code publicly available at https://github.com/abdo-ui. Our dataset is generated using Python as random binary data and the fading coefficients are obtained from the STK simulator. To ensure

**Algorithm 1** AE Training Algorithm in a $K$-User MAC channel

**Require:** $M$ messages $(1, 2, \ldots, M) \in \mathcal{M}$, transmitter peak intensity $A$, batch size $m$, learning rate $\eta$.

**Ensure:** $\text{argmax}(\boldsymbol{q}_j^{(t)}) = b_i^{(t)}\ \forall t \in \{1, 2, \ldots, m\}, i, j \in \{1, 2, \ldots, K\}$

$\theta_1, \ldots, \theta_K \leftarrow$ initialize encoder units parameters.
$\phi_1, \ldots, \phi_K \leftarrow$ initialize decoder units parameters.
$\upsilon \leftarrow$ initialize LN layer parameters.
**repeat**
  Draw $m$ minibatch samples for each user $((b_1^{(1)}, \ldots, b_K^{(1)}), \ldots, (b_1^{(m)}, \ldots, b_K^{(m)}))$.
  **for** $t \leftarrow 1$ to $m$ **do**
    **for** $z \leftarrow 1$ to $K$ **do**
      $\mathbf{1}_{b_z^{(t)}} \leftarrow \text{one\_hot\_vector}(b_z^{(t)})$ $\quad \{\mathbf{1}_{b_z^{(t)}} \in \{0,1\}^M\}$
      $\boldsymbol{x}_z^{(t)} \leftarrow E_{\theta_z}(\mathbf{1}_{b_z^{(t)}}, A)$ $\quad \{\boldsymbol{x}_z^{(t)} \in [0, A]^L\}$
    **end for**
    $\boldsymbol{y}^{(t)} \leftarrow \sum_{z=1}^{K} \boldsymbol{x}_z^{(t)} \boldsymbol{h}_z^{(t)} + \boldsymbol{w}^{(t)}$ $\quad \{\boldsymbol{y}^{(t)} \in \mathcal{R}^L\}$
    $\boldsymbol{y}^{(t)} \leftarrow \boldsymbol{y}^{(t)} / \hat{\boldsymbol{h}}^{(t)}$
    **for** $z \leftarrow 1$ to $K$ **do**
      $\boldsymbol{q}_z^{(t)} \leftarrow D_{\phi_z}(LN_\upsilon(\boldsymbol{y}^{(t)}))$ $\quad \{\boldsymbol{q}_z^{(t)} \in \mathcal{R}^M\}$
      $\boldsymbol{p}_z^{(t)} \leftarrow \text{Softmax}(\boldsymbol{q}_z^{(t)})$ $\quad \{\boldsymbol{p}_z^{(t)} \in [0,1]^M\}$
    **end for**
  **end for**
  **for** $z \leftarrow 1$ to $K$ **do**
    $C_z \leftarrow -\sum_{t=1}^{m} \sum_{r=1}^{M} \mathbf{1}_{b_z^{(t)}}(r) \log(\boldsymbol{p}_z^{(t)}(r))$
  **end for**
  $C = \max(C_1, C_2, \ldots, C_K)$
  **for** $\tau \leftarrow \theta_1, \ldots, \theta_K, \phi_1, \ldots, \phi_K, \upsilon$ **do**
    $\nabla_\tau C \leftarrow \frac{\partial C}{\partial \tau}$
    $\tau \leftarrow \tau - \eta \nabla_\tau C$
  **end for**
**until** convergence

TABLE I
LAYERS DESIGN OF THE AE$(k, L)$ USED IN FIG. 1

| Module | Layer | Input shape | Output shape | Number of parameters |
|---|---|---|---|---|
| Transmitter | Fully connected | $M$ | 100 | $100(M+1)$ |
| | Batch normalization | 100 | 100 | 400 |
| | Fully connected | 100 | $L$ | $101L$ |
| | Batch normalization | $L$ | $L$ | $4L$ |
| | Weighted sigmoid | $L$ | $L$ | 0 |
| Normalizer | Layer normalization | $L$ | $L$ | $2L$ |
| Decoder | Fully connected | $L$ | 100 | $100(L+1)$ |
| | Layer normalization | 100 | 100 | 200 |
| | Fully connected | 100 | $M$ | $101M$ |
| | Layer normalization | $M$ | $M$ | $2M$ |

a fair comparison, all benchmark models were trained with equal numbers of epochs and training samples. This ensures that any advantages observed in the proposed model can be attributed solely to its design. Also, note that the training is often done off-line and only once so that only the complexity during testing really matters. In addition, Tables I provides a breakdown of the relevant parameters for each layer of the proposed AE$(k, L)$ model.

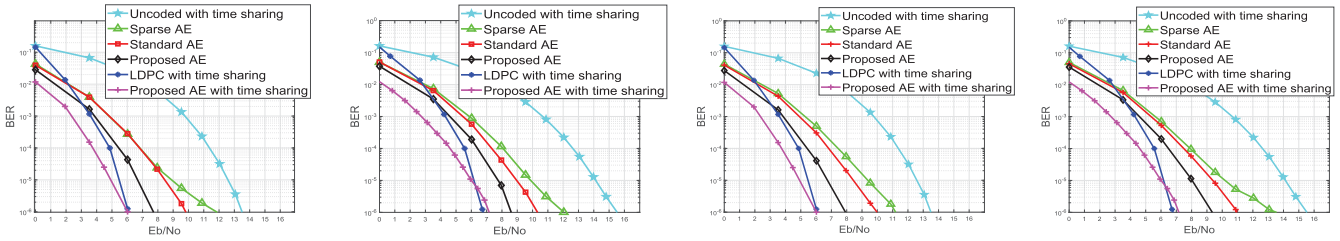
The proposed AE$(7, 21)$ is compared with the uncoded IM/DD, LDPC with time sharing, the standard AE$(7, 21)$ presented in [3], and the sparse AE$(7, 21)$ presented in [6] based on the BER performance metric. The comparison is made under two channel conditions: AWGN MAC and Log-normal fading MAC. Figures 3a and 3b present the MAC channel results. The LDPC scheme, which utilizes IM/DD modulation with time sharing settings, is characterized by a block length of $b_l = 100$. The design of this LDPC configuration of code rate $1/3$ follows the guidelines in [10]. In AWGN MAC channel illustrated in Fig. 3a, our proposed model outperforms the standard AE by 2.1 dB at BER $10^{-6}$ and is ahead of both the sparse AE and uncoded IM/DD by 4 and 5.7 dB, respectively. As shown in Fig. 3b, the proposed AE(7,21) is better than the standard AE by 1.5 dB at a BER of $10^{-6}$. In addition, the proposed AE(7,21) with time sharing settings outperforms the LDPC with time sharing at low SNR regime and has 1.3 dB gain at BER $10^{-4}$. Also, the proposed AE with time sharing has the same performance as the LDPC at BER $10^{-6}$. In the Log-normal MAC and IC channel, the standard deviation $\sigma$ is set to 0.2 [4]. The discrepancy in BER performance between the standard AE, sparse AE, and the proposed AE highlights the positive impact of integrating normalization layers and the min-max algorithm in the proposed model. In addition, we compare our system against the uncoded IM/DD, LDPC IM/DD, the sparse AE, and the AE system presented in [5] in the presence of two-user IC for Log-normal fading channel.

Figure 3c illustrates the evaluation of the proposed AE in an IC channel-based AWGN setting. The results indicate that the proposed AE outperforms the sparse AE and the uncoded IM/DD by 3 dB and 5.5 dB, respectively. In addition, Fig. 3 shows a comparison of all models in a Log-normal two-user IC channel with $\sigma = 0.2$. The results demonstrate that the proposed AE achieves better performance than the standard AE proposed in [5], with an improvement of 1.8 dB at a BER of $10^{-6}$. In addition, the proposed AE with time sharing is just inferior with 0.3 compared to the LDPC with time sharing at BER $10^{-6}$. However, the proposed AE with time sharing is better by 1.2 and 0.8 dB compared to the LDPC with time sharing at BER $10^{-3}$ and $10^{-4}$, respectively.

TABLE II
COMPUTATIONAL COMPLEXITY FOR EACH LEARNING-BASED MODEL

| Model | Number of learnable parameters |
|---|---|
| Sparse AE(7,21) [6] | 60,298 |
| Standard AE(7,21) [3], [5] | 77,098 |
| Proposed AE(7,21) | 62,220 |

Furthermore, Table II reveals that the proposed AE is more computationally efficient than the latest learning-based models. In the AWGN MAC channel, the standard AE$(7, 21)$ requires more 2.1 dB more in terms of SNR than the proposed AE$(7, 21)$ to achieve a BER of $10^{-6}$, even though it has 23% more learnable parameters. Moreover, the sparse AE$(7, 21)$ requires a 4 dB higher SNR value than the proposed AE$(7, 21)$ to achieve $10^{-6}$ BER, despite having only 3% fewer learnable parameters. Figure 4 visualizes the learned representations $\boldsymbol{x}$ of all messages as real constellation points for the proposed AE$(7, 21)$ in MAC channel. The histogram in Fig. 4 is a visualization of the learned constellations at the transmitter




(a) (b) (c) (d)

Fig. 3. BER versus $\frac{E_b}{N_o}$: (a) Two-user MAC AWGN channel, (b) Two-user MAC Log-normal channel, (c) Two-user IC AWGN channel, (d) Two-user IC Log-normal channel.

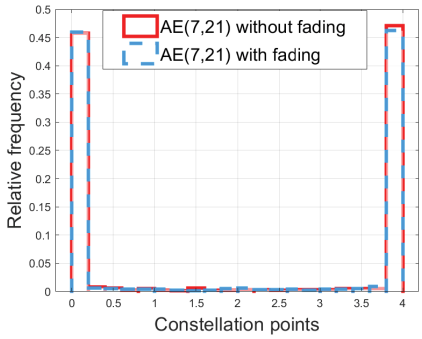



Fig. 4. The constellation points against the relative frequency generated by the proposed AE$(7, 21)$ considering a peak intensity of $A = 4$.

for all possible messages trained and tested for $A = 4$ over a two-user AWGN and log-normal fading channels. Obviously, the results in Fig. 4 are applied in the testing stage after achieving the best weights for minimizing the loss according to the min-max problem. The results in Fig. 4 are in symbol representations as the transmitter in the AE applies both channel coding and symbol mapping simultaneously considering both the peak and positivity constraints. The distribution of the encoded symbols is generated after the training stage. While Fig. 4 looks like an on-off-keying modulation and similar constellation points as represented in [3] and [4], the scattering of the points is still different and not all points are exactly located in 0 and $A$. In particular Fig. 4, verifies that the AE output follows both positivity and peak intensity intensity constraints $\in [0, A]$. It is interesting to observe that the learned constellation points for the AE(7,21) are scattered in the interval $[0, A]$ with different relative frequencies. These findings indicate that the AE effectively acquired efficient coding, modulation techniques in a MAC scenario.

## V. Conclusion

In this letter, we develop a DL AE model using a novel layered framework that incorporates BN in the encoder and LN in the decoders in SOC specifically tailored for multi-user environments. A realistic SOC channel model for fading channels is created using the STK simulator. Our model focuses not only on enhancing the BER performance but also on optimizing computational complexity making it a promising solution for improving communication reliability and efficiency in SOC systems. The proposed AE enables scalability for any number of users in a multi-user environment. The numerical results indicate that the proposed AE exhibits superior performance in terms of BER and computational complexity compared to the existing learning frameworks in both MAC and IC channels. The proposed AE results yields better results that all E2E learning frameworks currently considered in the state-of-the-art, to the best of our knowledge.

## References

[1] H. Kaushal and G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 57–96, 1st Quart., 2017.

[2] A. E. A. El-Fikky and Z. Rezki, "On the performance of autoencoder-based space optical communications," in *Proc. IEEE Global Commun. Conf.*, Rio de Janeiro, Brazil, Dec. 2022, pp. 1466–1471.

[3] M. Soltani, W. Fatnassi, A. Aboutaleb, Z. Rezki, A. Bhuyan, and P. Titus, "Autoencoder-based optical wireless communications systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[4] Z.-R. Zhu, J. Zhang, R.-H. Chen, and H.-Y. Yu, "Autoencoder-based transceiver design for OWC systems in log-normal fading channel," *IEEE Photon. J.*, vol. 11, no. 5, pp. 1–12, Oct. 2019.

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

[6] J. Lin, S. Feng, Y. Zhang, Z. Yang, and Y. Zhang, "A novel deep neural network based approach for sparse code multiple access," *Neurocomputing*, vol. 382, pp. 52–63, Mar. 2020.

[7] M. Polnik, L. Mazzarella, M. Di Carlo, D. K. Oi, A. Riccardi, and A. Arulselvan, "Scheduling of space to ground quantum key distribution," *EPJ Quantum Technol.*, vol. 7, no. 1, pp. 1–34, Jan. 2020.

[8] A. Chaaban, Z. Rezki, and M.-S. Alouini, "On the capacity of intensity-modulation direct-detection Gaussian optical wireless communication channels: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 455–491, 1st Quart., 2022.

[9] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Conf. Neural Informat. Process. Syst.*, vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., 2018, pp. 1–12.

[10] B. Tahir, S. Schwarz, and M. Rupp, "BER comparison between convolutional, turbo, LDPC, and polar codes," in *Proc. 24th Int. Conf. Telecommun. (ICT)*, May 2017, pp. 1–7.