# Noncoherent MIMO Codes Construction using Autoencoders

Mohamed A. ElMossallamy, Zhu Han, Miao Pan, Riku Jäntti, Karim G. Seddik and Geoffrey Ye Li

***Abstract*—In this paper, we examine the use of autoencoders as an optimization tool for the construction of noncoherent space-time MIMO codes. In particular, we consider the quasi-static block fading channel, where the channel state information is not available at either the transmitter or the receiver, and changes independently between transmissions. Different from traditional constructions which aim to maximize an approximation of the minimum pairwise distance of the constellation, we use the autoencoder to directly target minimizing the probability of error. We show that this different optimization goal leads to constellations with more favorable pairwise distances' distribution and better error performance at low to medium signal to noise ratios where the minimum distance is not the limiting factor. Finally, we present simulation results showing that the constructed codes outperform traditional Grassmannian codes up to a signal-to-noise ratio of 20 dB using the traditional generalized likelihood ratio test detector.**

***Index Terms*—Autoencoders, Multiple-Input Multiple-Output (MIMO), noncoherent communications, space-time codes.**

## I. Introduction

Over the last decade, machine learning techniques have made astonishing progress and found wide applications in areas where modelling and expert knowledge have struggled to make headway, such as computer vision and natural language processing. On the contrast, wireless communications is a mature domain with rich expert knowledge that has made impressive progress over the decades paving the way for the current information revolution. Hence, learning techniques will have a higher bar to overcome to make meaningful contributions to the field of wireless communications. Nonetheless, learning techniques have recently started making its way into the field of wireless communications and caught the attention of both academic and industrial sectors.

The design of wireless communications systems traditionally relied on various assumptions about the communication scenario, availability of channel state information at the transmitter and/or the receiver, and rigorous mathematical models describing the physical signal propagation, noise in the system, and hardware impairments. However, in many cases, an optimal solution does not exist or requires restrictive or unrealistic assumptions. Hence, aside from the intellectual curiosity to better understand the ability to learn communication systems components, or even communication end-to-end, there exists scenarios where learned techniques can improve upon the current state-of-the-art.

Although earlier attempts to incorporate neural network based learning in wireless communications dates back decades [1], it was only recently that research efforts in this area became main stream. Recent applications include signal detection [2]–[6], channel estimation and quantization [7]–[9], channel coding/decoding [10]–[12], and even end-to-end learning of a complete communications system [13]–[17].

In [2], the authors proposed a deep network architecture optimized for detection in spatially multiplexed multiple input multiple output (MIMO) systems, and inspired by the concept of unfolding iterative algorithms [18]. They showed that competitive performance with semidefinite relaxation is possible while achieving much lower complexity for binary transmitted vectors. In [6], the authors proposed using a neural network to estimate the initial radius of a MIMO sphere decoder which can lead to significant computational complexity savings by reducing the number of considered lattice points. In [3], it has been shown that a learned orthogonal frequency division multiplexing (OFDM) receiver that jointly estimates the channel and performs detection can be more robust to various impairments, e.g. clipping, compared to traditional methods. In [5], the authors used supervised classification methods to aid blind detection in one-bit quantized MIMO systems, where coarse quantization invalidates traditional methods. In [4], a special architecture called sliding bidirectional recurrent neural network (SBRNN) we proposed for blind detection in harsh channel models associated with molecular and free-space optical communications.

In [7], the authors presented a convolutional neural network (CNN) based low complexity channel estimator and showed it achieves competitive performance with the optimal minimum mean square error (MMSE) estimator in a variety of 3GPP channel models. In [8], the authors showed that deep learning (DL) based techniques for the challenging channel estimation in hybrid analog-digital beamspace outperformed state-of-the-art compressed sensing algorithms. DL-based techniques was also used for channel state information (CSI) quantization in [9] to facilitate CSI feedback in frequency division duplexing (FDD) systems.

Different from aforementioned works where learning-based

This work was supported in part by the U.S. National Science Foundation under grants US CNS-1350230 (CAREER), CNS-1646607, CNS-1702850, and CNS-1801925.

Mohamed A. ElMossallamy, Miao Pan and Zhu Han are with the Electrical and Computer Engineering Department, University of Houston, TX, USA. Mohamed A. ElMossallamy is also with the Department of Electrical Engineering, Future University, Cairo, Egypt. Zhu Han is also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea.

Riku Jäntti is with the Department of Communications and Networking, Aalto University, Espoo, Finland.

Karim G. Seddik is with the Electronics and Communications Engineering Department, American University in Cairo, New Cairo, Egypt.

Geoffrey Ye Li is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

elements complemented traditional systems, in [13], [19], the authors proposed replacing all communication systems block with neural networks and training the system end-to-end as a particular type of denoising autoencoder. The single-input-single-output (SISO) scenario was considered in [13], while MIMO scenarios were considered in [19]. In both scenarios competitive performance with traditional techniques was achieved. Motivated by these encouraging results, the authors in [15] implemented the same concept using software defined radios and conducted over the air testing showing performance within 1 dB of traditional techniques.

In this paper, we investigate the use of autoencoders as an optimization tool to construct noncoherent space time codes for the MIMO block fading channel. Autoenocders are a special type of neural networks capable of learning representations that are robust to specific types of corruption which make them attractive for learning signaling schemes when an optimal analytical solution does not exist. Our contributions in this paper can be summarized as follows:

- We propose using autoencoders as an optimization tool to learn noncoherent space time codes suitable for MIMO block fading channel with favorable error performance. Different from [19], we are not interested at learned encoder and decoder but just the learned space time code, since the optimal maximum likelihood (ML) detector is known for this channel model.
- We evaluate the learned codebook by constructing the histogram of pairwise distances between the codewords and compare it to traditional noncoherent Grassmannian MIMO codes. We show that the learned constellations have a more favorable distances' distribution.
- We use Monte Carlo simulations to evaluate the performance of the learned constellation and compare it to traditional Grassmannian codes. Our results show that learned codes outperforms Grassmannian for up to a relatively high signal to noise ratio of 20 dB.

The rest of the paper is organized as follows. In Section II, we present our signal model and give an quick overview of traditional noncoherent MIMO codes and autoencoders. In Section III, we introduce our autoencoder based constellation construction. In Section IV, we evaluate the leanred codebook distance distribution and error performance using traditional noncoherent Grassmannian MIMO codes as a baseline. Finally, we conclude the paper in Section V.

## II. Signal Model and Preliminaries

### A. Signal Model

We consider the scenario of noncoherent communications over richly scattered quasi-static flat-fading MIMO channels. The transmitter possess $N_t$ antennas and the receiver $N_r$ antennas. The wireless channel remains constant for a coherence interval of $T$, then changes to another independent realization. Hence, the received matrix can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{H} + \sqrt{\frac{N_t}{\rho T}}\mathbf{W}, \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{T\times N_t}$ is the transmitted matrix over the $N_t$ antennas and $T$ channel uses, $\mathbf{H} \in \mathbb{C}^{N_t\times N_r}$ and $\mathbf{W} \in \mathbb{C}^{T\times N_r}$ are the fading and noise matrices, respectively, whose entries are drawn independently from circularly symmetric complex Gaussian distribution $\mathcal{CN}(0,1)$. Finally, $\rho$ denotes the average signal-to-noise ratio (SNR) which is independent of $N_t$.

To facilitate the usage of neural networks frameworks, we adapt the following equivalent real-valued notation

$$\overline{\mathbf{Y}} = \overline{\mathbf{X}}\ \overline{\mathbf{H}} + \sqrt{\frac{N_t}{\rho T}}\overline{\mathbf{W}}, \tag{2}$$

where

$$\overline{\mathbf{Y}} = \begin{bmatrix} \Re\{\mathbf{Y}\} & -\Im\{\mathbf{Y}\} \\ \Im\{\mathbf{Y}\} & \Re\{\mathbf{Y}\} \end{bmatrix} \quad \overline{\mathbf{X}} = \begin{bmatrix} \Re\{\mathbf{X}\} & -\Im\{\mathbf{X}\} \\ \Im\{\mathbf{X}\} & \Re\{\mathbf{X}\} \end{bmatrix}$$
$$\overline{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix} \quad \overline{\mathbf{W}} = \begin{bmatrix} \Re\{\mathbf{W}\} & -\Im\{\mathbf{W}\} \\ \Im\{\mathbf{W}\} & \Re\{\mathbf{W}\} \end{bmatrix}, \tag{3}$$

and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary part, respectively.

Signalling under the assumption that the channel matrix, $\mathbf{H}$, is not known to either the transmitter or receiver was extensively studied [20]. It was found that unitary signaling is capacity-achieving at asymptotically high SNRs which led to multiple efforts to design unitary constellations. However, these codes are known to perform poorly at low-to-medium SNRs, which leaves room for improvement. In the next subsection, we give a brief overview of traditional noncoherent unitary MIMO codes.

### B. Grassmannian MIMO Codes

The set of $N_t$ dimensional subspaces in $\mathbb{C}^T$ comprises the so called complex Grassmann manifold, $\mathbb{G}_{N_t}(\mathbb{C}^T)$. Hence, for $T > N_t$, the set of $T \times N_t$ unitary matrices represent points on the complex Grassmann manifold.

The importance of the Grassmann manifold in the context of noncoherent MIMO communications becomes evident when we consider the high SNR scenario. From (9), as the effect of noise diminishes, the received signal is approximately given by $\mathbf{Y} \approx \mathbf{X}\mathbf{H}$. Multiplication by $\mathbf{H}$ alone can only rotate and scale the subspace spanned by $\mathbf{X}$ and clearly $\mathbf{Y}$ and $\mathbf{X}$ span the same subspace. Hence, by designing the constellation of $\mathbf{X}$ to comprise matrices spanning different subspaces, i.e., different points on the Grassmann manifold, we can guarantee good performance at high SNRs.

Motivated by its utility in noncoherent communications, many researchers set to find good Grassmannian constellations. There exists three main approaches to design Grassmannian codes: 1) algebraic approaches [21], 2) approaches that map coherent MIMO codes into the Grassmann manifold [22], and 3) approaches that rely on direct numerical optimization on the Grassmann manifold of some metric of performance [23]. Clearly, the latter approach will generally lead to better preforming codes since it's free to exploit the full degrees of the freedom without adhering to any specific structure.

In this paper, we use the codes designed in [23] as a baseline to judge the performance of our codes. In [23],

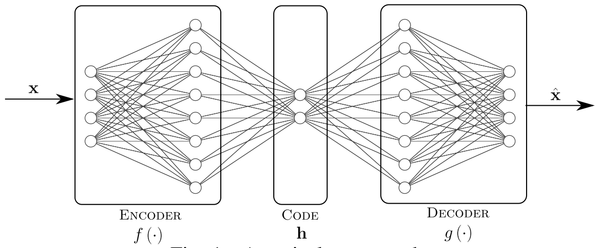


Fig. 1. A typical autoencoder.

the chordal Frobenius norm was used as metric of distance between constellation points, and an optimization problem was formulated to maximize the minimum distance of the constellation. Let $(\cdot)^\dagger$ and $\operatorname{Tr}(\cdot)$ denote the Hermitian transpose and trace operator, respectively. The chordal Frobenius norm between two $T \times N_t$ matrices, $\mathbf{X}_1$ and $\mathbf{X}_2$ is given by

$$d(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{2N_t - 2\operatorname{Tr}(\mathbf{\Sigma}_{\mathbf{X}_1,\mathbf{X}_2})} \tag{4}$$

where $\mathbf{\Sigma}_{\mathbf{X}_1,\mathbf{X}_2}$ is the diagonal matrix comprising the singular values of $\mathbf{X}_1^\dagger \mathbf{X}_2$. Hence, the following program needs to be solved to generate a constellation, $\mathcal{X}$, with maximized minimum distance:

$$\begin{aligned} &\min_{\{\mathbf{X}_k\}_{k=1}^{|\mathcal{X}|}} \max_{1 \le i,j \le |\mathcal{X}|} \operatorname{Tr}(\mathbf{\Sigma}_{\mathbf{X}_i,\mathbf{X}_j}) \\ &\text{subject to} \quad \mathbf{X}_k \in \mathbb{G}_{N_t}(\mathbb{C}^T), \quad k = 1, 2, \ldots, |\mathcal{X}|, \end{aligned} \tag{5}$$

after using the following smooth differentiable approximation:

$$\begin{aligned} &\min_{\{\mathbf{X}_k\}_{k=1}^{|\mathcal{X}|}} \left( \log \left( \sum_{i=1}^{|\mathcal{X}|-1} \sum_{j=i+1}^{|\mathcal{X}|} \mathrm{e}^{\operatorname{Tr}^n(\mathbf{\Sigma}_{\mathbf{X}_i,\mathbf{X}_j})} \right) \right)^{\frac{1}{n}} \\ &\text{subject to} \quad \mathbf{X}_k \in \mathbb{G}_{N_t}(\mathbb{C}^T), \quad k = 1, 2, \ldots, |\mathcal{X}|, \end{aligned} \tag{6}$$

where $n$ is a parameter controlling the smoothness of the approximation. This program can be solved with help of [24] in either a greedy manner, where the constellation is generated point-by-point, or a direct manner, where the entire constellation is generated at once.

### C. Autoencoders

An autoencoder, depicted in Fig. 1, consists of an encoder part, $f(\cdot)$, a hidden code layer, $\mathbf{h}$, and a decoder part, $g(\cdot)$. In essence, it is just a neural network trained to copy its input, $\mathbf{x}$ to its output, $\hat{\mathbf{x}}$. This is carried out by minimizing a given loss function:

$$\mathcal{L}(\mathbf{x}, g(f(\mathbf{x}))). \tag{7}$$

Since the desired output is just the input, what is interesting is the code representation at the hidden layer and not the output. By training the autoencoder, one can extract codes with desirable properties.

To get useful encodings with interesting properties, the autoencoder must be prevented from simply learning the trivial identity function. This is achieved by restricting the code layer in some way or adding noise. For example, undercomplete autoencoders learn a nonlinear generalization of principal component analysis (PCA) by restricting the dimension of the hidden layer, while denoising autoencoders learn codes that are robust to some form of corruption by injecting noise during the training process.

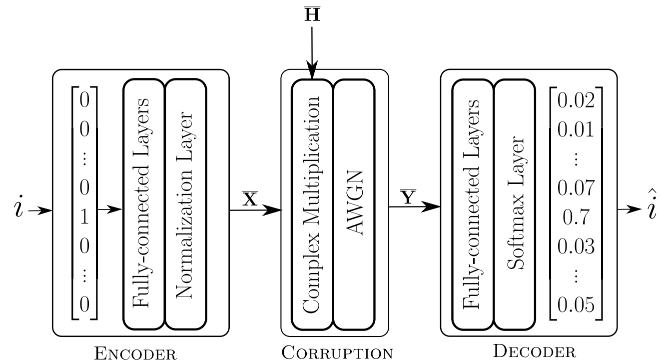


Fig. 2. Denoising autoencoder structure used for generation of noncoherent codes.

Of particular interest in the context of communications systems is denoising autoencoders (DAE). As they share the common goal of finding encodings that are robust to corruption with physical layer designs of communications systems. Although modern denosing autoencoders typically add the corruption to the input layer, earlier research on DAEs investigated injecting noise at the hidden code layer [25], which is more consistent with corruption arising in a communications system.

Recently, the utility of denoising autoencoders in the design of communications systems gained a lot of interest in the research community [13], [15], [19], [26]. Promising results were obtained in the scenarios where channel is known at the receiver or both the transmitter and receiver [19]; however, the noncoherent MIMO scenario where the channel is not known at either the transmitter or receiver was not investigated despite the fact that optimal constellation designs at low to medium SNRs are not known.

## III. Autoencoder-based Constellation Construction

In this section we propose using autoencoders as an optimization tool to generate space-time MIMO codes that perform well over the noncoherent MIMO channel. Different from [19], we are only interested in the learned codes at the hidden layer and after the learning process is done, we do not use the learned encoder and decoder functions. Recall that although an optimal constellation design for the signal model is not known except at asymptotically high SNRs, the optimal ML detector is known. Hence, there is no value in using the learned decoder unless it is offers complexity savings over the ML detector and we use the learned constellation with the known optimal detector.

As depicted in Fig. 2, a denoising autoencoder with the corruption process taking place at the hidden code layer is used. The corruption process in our scenario will be complex multiplication by a random matrix whose entries are standard complex Gaussian random variables, i.e., the channel matrix $\mathbf{H}$, and addition of complex AWGN, i.e., the noise matrix $\mathbf{W}$. Since all freely available neural network frameworks can only deal with real numbers, the equivalent model in (2) is used.

TABLE I
LAYOUT OF USED DENOISING AUTOENCODER.

| Layer | Output Dimensions |
|---|---|
| Input | $\lvert\mathcal{X}\rvert$ |
| Fully Connected + ReLU | $20\ \lvert\mathcal{X}\rvert$ |
| Fully Connected + linear | $2\ TN_t$ |
| Complex Multiplication | $2\ TN_t$ |
| AWGN | $2\ TN_t$ |
| Fully Connected + ReLU | $20\ \lvert\mathcal{X}\rvert$ |
| Fully Connected + softmax | $\lvert\mathcal{X}\rvert$ |

The encoder takes a one-hot input vector, $\mathbf{m}$, specifying which message, i.e. constellation point, is to be transmitted. The input vector is passed through several fully-connected layers then a normalization layer. The normalization layer is needed to ensure the desired SNR is maintained. The resulting codeword, $\overline{\mathbf{X}}$ is then passed through the corruption process representing our channel and noise model. Finally, the corrupted codeword, $\overline{\mathbf{Y}}$, is passed to the decoder part which consists of several fully-connected layers then a soft-max output layer. The elements of the output vector at the softmax layer, $\hat{\mathbf{m}}$, can be interpreted as the probabilities of each corresponding codeword being the actual transmitted codeword.

Different from traditional design techniques that aim to optimize some distance metric related to the probability of error. The proposed design allows us to directly target probability of error minimization by treating the problem as multi-class classification task. In particular, the categorical cross entropy loss function, $\mathcal{L}_{CE}$, is used to penalize the difference between the input vector, $\mathbf{m}$ and the prediction $\hat{\mathbf{m}}$ such that

$$\mathcal{L}_{CE}\left(\mathbf{m}, \hat{\mathbf{m}}\right) = -\sum_{i=1}^{|\mathcal{X}|} m_i \log \hat{m}_i, \tag{8}$$

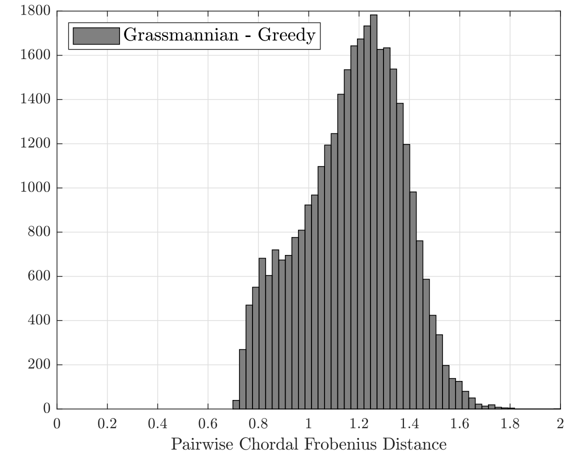
where $m_i$ and $\hat{m}_i$ are the $i$-th elements of $\mathbf{m}$ and $\hat{\mathbf{m}}$, respectively.

A variety of optimization algorithms can be used to minimize this loss function. We used the Adam optimization algorithm [27] which is a variant of stochastic gradient descent with an adaptive momentum and learning rate. The constructed autoencoder was trained to learn a codebook $\overline{\mathbf{X}} \in \mathcal{X}$ that is robust to the corruption process and achieves good error performance. In the next section, we evaluate the learned codebook and compare to traditional Grassmannian codebooks.

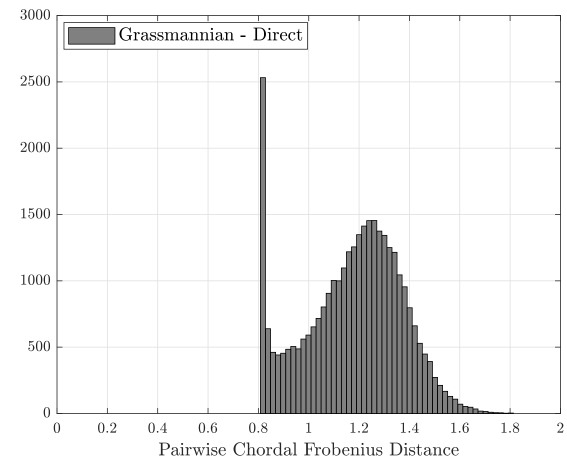
## IV. Results and Discussion

In this section, we evaluate the learned constellation by looking at distribution of pairwise distances between its points and its error performance using Monte-Carlo simulations. Traditional noncoherent Grassmannian codes designed using the techniques discussed in Section II are used as a baseline to judge the codes constructed using the proposed autoencoder-based optimization.

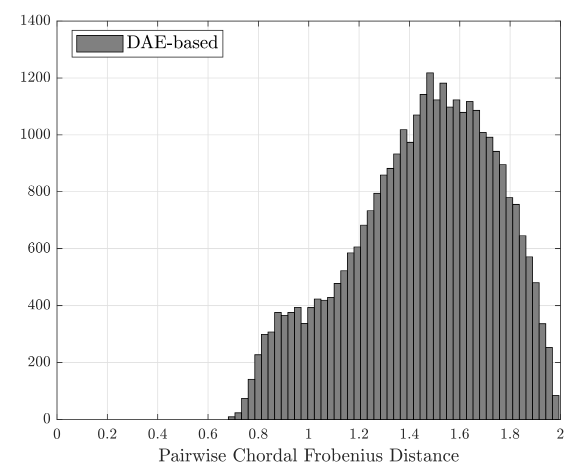
We consider two constellation sizes, namely, 256 and 512 points constellations, and assume a channel coherence




(a) Grassmannian constellation using greedy approach ($d_{min} = 0.6987$).




(b) Grassmannian constellation using direct approach ($d_{min} = 0.8080$).




(c) Proposed constellation using AE-based approach ($d_{min} = 0.6807$).

Fig. 3. Distribution of pairwise distances between constellation points for proposed codes and baselines from [23] for 256 point constellations.

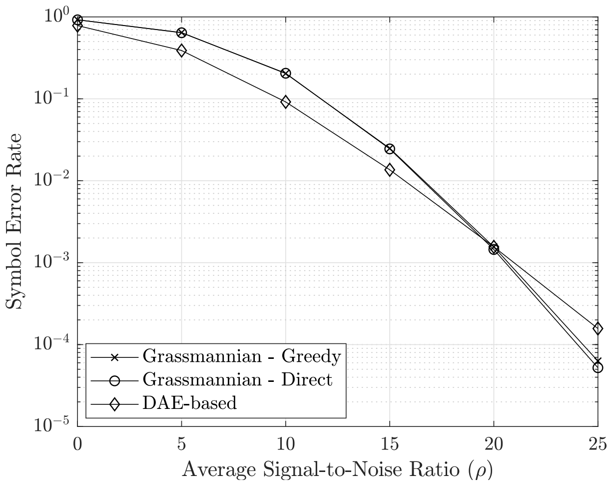



Fig. 4. Average probability of error for proposed codes and baselines from [23] using the GLRT detector. 256 points constellations. $T = 4, N_t = N_r = 2$.

interval, $T = 4$, the number of transmit antennas, $N_t = 2$, and the number of receive antennas, $N_r = 2$. Note that, the denoising autoencoder (DAE) optimized codes are constrained to be of equal energy but are not necessarily unitary. The autoencoder was trained at an SNR of 15 dB to strike a compromise between performance at low and high SNRs. Table I shows the layout of the autoencoder used. For a fair comparison we use the GLRT detector given by

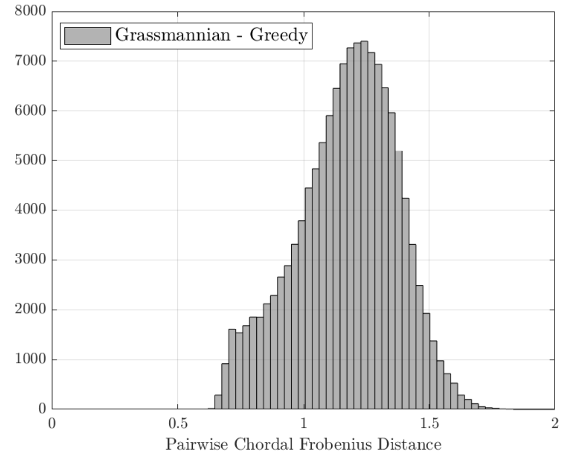
$$\arg\max_{\mathbf{X}} \mathrm{Tr}(\mathbf{Y}^\dagger \mathbf{X}\mathbf{X}^\dagger \mathbf{Y}) \tag{9}$$

for the decoding of both the baseline and proposed codes

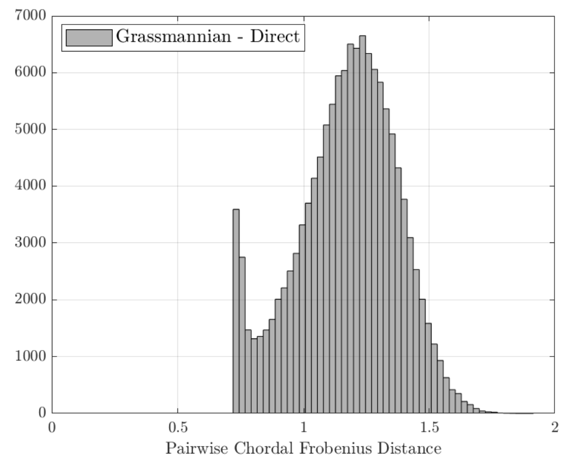
Fig. 3 shows the pairwise distance distribution for the 256 point constellations of the proposed codes and that of the baseline Grassmannian codes from both the greedy and direct approaches. We observe the proposed codes seem to have a more favorable distance distribution, i.e., higher concentration of larger distances. However, the baseline Grassmannian codes tend to have a higher minimum distance. This is expected as the optimization problem solved to generate them aims to maximize the minimum distance. The effect of this choice of optimization objective is more evident for the constellation designed using the direct approach, where the highest concentration of distances is exactly at the minimum. Although a higher minimum distance might improve performance at high SNRs, this high concentration of points close to each others can have adverse effects at lower SNRs.

Fig. 4 shows the error performance for the 256 point constellations of the proposed codes and that of the baseline Grassmannian codes from both the greedy and direct approaches. We observe the proposed codes outperform the Grassmannian codes for low-to-medium SNRs up to a moderately high SNR of 20 dB. This is consistent what we expected, at low-to-medium SNRs the more favorable distances distribution leads to better performance. However, as the SNR increases beyond 20 dB, Grassmannian codes become increasingly optimal and outperform the proposed codes.

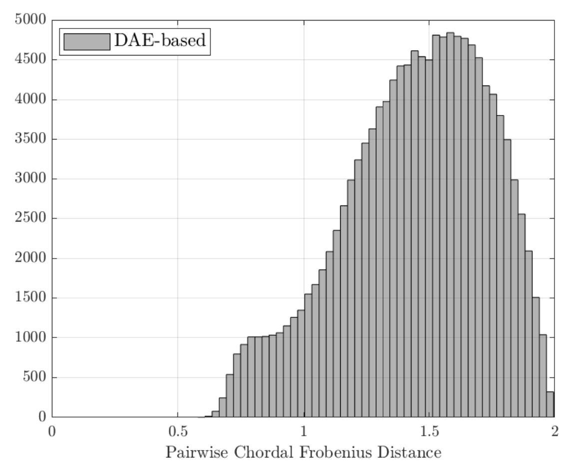
Fig. 5 shows the pairwise distance distribution for the 512 point constellations of the proposed codes and that of the




(a) Grassmannian constellation using greedy approach ($d_{min} = 0.6286$).




(b) Grassmannian constellation using direct approach ($d_{min} = 0.7348$).




(c) Proposed constellation using AE-based approach ($d_{min} = 0.5957$).

Fig. 5. Distribution of pairwise distances between constellation points for proposed codes and baselines from [23] for 512 point constellations.

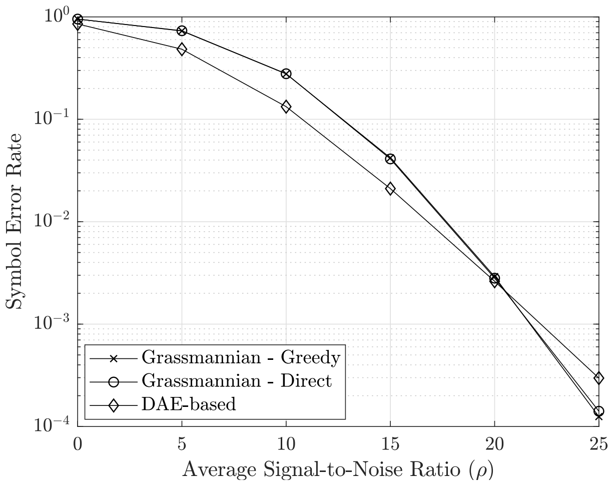



Fig. 6. Average probability of error for proposed codes and baselines from [23] using the GLRT detector. 512 points constellations. $T = 4, N_t = N_r = 2$.

baseline Grassmannian codes from both the greedy and direct approaches. From the figure, we observe similar trends to the distributions of the 256 point constellations. Grassmannian codes tend to have a higher minimum distance but the proposed DAE-based constellation have a more favorable distribution.

Fig. 6 shows the error performance of 512 point constellation designed using the proposed DAE-based approach, and again Grassmannian codes from both the greedy and direct approaches are used as a baseline. From the figure, we observe that the proposed codes still outperform the Grassmannian codes up to an SNR of 20 dB in this case as well.

## V. Conclusion

We have proposed an autoencoder-based optimization technique to generate noncoherent space time codes. In particular, a denoising autoencoder is utilized to find codes that are inherently robust to the effects of the wireless fading channel. The proposed technique forgoes the high SNR assumption usually adopted when designing noncoherent MIMO codes and directly optimize for lower error rates without relying on a surrogate distance metric. We have evaluated the resultant constellation and found it to have a more favorable distances distribution. Simulations results verified the utility of the generated constellation and showed that they outperform baseline Grassmannian constellations for SNRs up to 20 dB.

## References


[1] M. Ibnkahla, "Applications of neural networks to digital communications a survey," *Elsevier Signal Process.*, vol. 80, no. 7, pp. 1185–1215, Jul. 2000.

[2] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017.

[3] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[4] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5663–5678, Nov. 2018.

[5] Y. Jeon, S. Hong, and N. Lee, "Blind detection for MIMO systems with low-resolution ADCs using supervised learning," in *IEEE Int. Conf. on Commun. (ICC)*, Paris, France, May 2017.

[6] M. Mohammadkarimi, M. Mehrabi, M. Ardakani, and Y. Jing, "Deep learning based sphere decoding," Jul. 2018. [Online]. Available: http://arxiv.org/abs/1807.03162

[7] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, Jun. 2018.

[8] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, pp. 852–855, Oct. 2018.

[9] C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[10] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Beery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[11] T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding," in *Annu. Conf. on Inform. Sci. and Syst. (CISS)*, Baltimore, MD, Mar. 2017.

[12] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Calgary, Alberta, Canada, Apr. 2018, pp. 2326–2330.

[13] T. OShea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. on Cognitive Commun. and Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[14] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *Asilomar Conf. on Signals, Syst., and Comput.*, Pacific Grove, CA, Oct. 2018, pp. 298–303.

[15] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.

[16] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Blow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *J. of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.

[17] H. Ye, G. Y. Li, B. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in *IEEE Globecom Workshops*, Abu Dhabi, UAE, Dec. 2018.

[18] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: model-based inspiration of novel deep architectures," Sep. 2014. [Online]. Available: http://arxiv.org/abs/1409.2574

[19] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based MIMO communications," Jul. 2017. [Online]. Available: http://arxiv.org/abs/1707.07980

[20] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[21] V. Tarokh and Il-Min Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3112–3117, Dec. 2002.

[22] I. Kammoun and J. Belfiore, "A new family of Grassmann space-time codes for non-coherent MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 528–530, Nov. 2003.

[23] R. H. Gohary and T. N. Davidson, "Noncoherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, Mar. 2009.

[24] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. on Matrix Analysis and Applicat.*, vol. 20, no. 2, pp. 303–353, Jan. 1998.

[25] H. Inayoshi and T. Kurita, "Improved generalization by adding both auto-association and hidden-layer-noise to neural-network-based classifiers," in *IEEE Workshop on Machine Learning for Signal Process.*, Mystic, CT, Sep. 2005, pp. 141–146.

[26] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. t. Brink, "OFDM-Autoencoder for End-to-End Learning of Communications Systems," Mar. 2018. [Online]. Available: http://arxiv.org/abs/1803.05815

[27] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014. [Online]. Available: https://arxiv.org/abs/1412.6980v9