A DNN Architecture for the Detection of Generalized Spatial Modulation Signals

Bharath Shamasundar and Ananthanarayanan Chockalingam
Department of ECE, Indian Institute of Science, Bangalore 560012
email: bharath@iisc.ac.in, achockal@iisc.ac.in

***Abstract*—In this letter, we consider the problem of signal detection in generalized spatial modulation (GSM) using deep neural networks (DNN). We propose a novel modularized DNN architecture that uses small sub-DNNs to detect the active antennas and complex modulation symbols, instead of using a single large DNN to jointly detect the active antennas and modulation symbols. The main idea is that using small sub-DNNs instead of a single large DNN reduces the required size of the NN and hence requires learning lesser number of parameters. Under the assumption of i.i.d Gaussian noise, the proposed DNN detector achieves a performance very close to that of the maximum likelihood detector. We also analyze the performance of the proposed detector under two practical conditions: *i*) correlated noise across receive antennas and *ii*) noise distribution deviating from the standard Gaussian model. The proposed DNN-based detector learns the deviations from the standard model and achieves superior performance compared to that of the conventional maximum likelihood detector.**

Keywords *– Deep neural networks, generalized spatial modulation, signal detection, correlated noise, non-Gaussian noise.*

I. Introduction

Index modulation (IM) techniques are attracting increased research attention due to their superior bit error performance at lesser hardware complexity [1]. Spatial modulation (SM) [2]-[4] is a popular IM scheme which uses n_t transmit antennas and a single transmit radio frequency (RF) chain. In a given channel use, one of the transmit antennas is selected based on $\lfloor \log_2 n_t \rfloor$ information bits and a symbol from a conventional modulation alphabet $\mathbb{A}$ (QAM/PSK) is transmitted on the selected antenna. Thus SM achieves a rate of $\lfloor \log_2 n_t \rfloor + \log_2 |\mathbb{A}|$ bits per channel use (bpcu). The reduced hardware complexity in SM comes at the cost of the reduced throughput. This drawback is overcome by generalized SM (GSM), which allows multiple transmit antennas to be active simultaneously [5],[6]. GSM uses n_t transmit antennas and n_{rf} RF chains, $1 < n_{rf} < n_t$. In each channel use, n_{rf} out of the n_t transmit antennas are selected based on $\lfloor \log_2 \binom{n_t}{n_{rf}} \rfloor$ information bits and n_{rf} symbols from the modulation alphabet $\mathbb{A}$ are transmitted from the selected active antennas. The achieved rate in GSM is therefore $\lfloor \log_2 \binom{n_t}{n_{rf}} \rfloor + n_{rf} \log_2 |\mathbb{A}|$ bpcu. In the present work, we consider the problem of signal detection for GSM using deep neural networks (DNN).

Recently, deep learning (DL) has been employed in wireless communications for designing intelligent communication systems [7]-[12]. Specifically, in the physical layer, DL has been applied in two important ways: *i*) as a replacement to the existing communication blocks like channel coding [11] and signal detection [12], [15], and *ii*) for designing end-to-end communication systems without traditional communication blocks [10]. Both the approaches have shown promising results. DL has been applied in the context of SM in [13] to achieve link adaptation, in which the problems of transmit antenna selection (TAS) and power allocation (PA) are converted to those of data driven prediction, which are then solved using DNN-based methods. In the present work, we consider the problem of signal detection in GSM and explore the utility of DNN for detection task. Our contributions in this letter can be summarized as follows.

- We propose a novel modularized DNN architecture that uses small sub-DNNs to detect the active antennas and complex modulation symbols. This is in contrast to using a single large DNN to jointly detect the active antennas and modulation symbols. The main idea is that using small sub-DNNs reduces the required size of the NN and hence requires learning lesser number of parameters.
- We show that, under static channel conditions and i.i.d Gaussian noise across receive antennas, the proposed DNN architecture can achieve a performance very close to that of the optimum maximum likelihood detection.
- When the noise across different receive antennas are correlated, which arises in practice due to mutual coupling among the receive antennas, matching networks, etc., the DNN-based detector learns the noise correlation and achieves superior performance compared to that of using the maximum likelihood (ML) detection meant for i.i.d Gaussian noise, and a performance close to that of the true ML detector for correlated noise (which achieves the best detection performance under correlated noise). Also, when the noise is i.i.d but the distribution slightly deviates from Gaussian, the proposed DNN architecture learns a good detector for the non-Gaussian noise, and achieves superior performance compared to the ML detector meant for i.i.d Gaussian noise.
- Finally, we extend the proposed DNN-based detector to the case of varying channels and show that the proposed detector achieves a performance close to that of the optimum ML detector.

We note that, although low-complexity signal detection in GSM has been previously studied in the literature (e.g., [6],[14]), these works mainly consider GSM signal detection in the standard i.i.d Gaussian noise settings. GSM detectors under non-standard noise settings have not been reported, and the proposed DNN approach that considers GSM detection in non-standard noise settings is a novel contribution.

II. GSM system model

Consider a MIMO communication system with n_t transmit and n_r receive antennas. Let n_{rf}, $1 < n_{rf} < n_t$, be the number

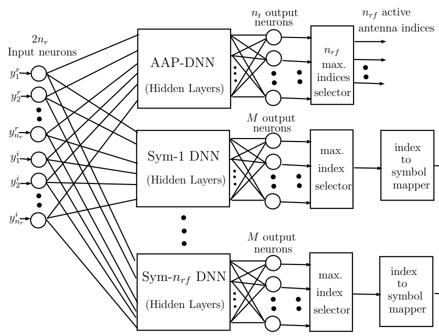


Fig. 1: Proposed DNN architecture for GSM signal detection.

of transmit RF chains at the transmitter. In GSM, in a channel use, n_{rf} out of the n_t transmit antennas are selected based on $\lfloor \log_2 \binom{n_t}{n_{rf}} \rfloor$ information bits. The selected n_{rf} antennas are called active antennas, on which n_{rf} symbols from a modulation alphabet $\mathbb{A}$ (say, QAM) are transmitted based on $n_{rf} \log_2 |\mathbb{A}|$ information bits. Let $\mathbb{A}_0 = \mathbb{A} \cup 0$. The GSM signal set is a set of n_t-length vectors given by

$$\mathbb{S} = \{\mathbf{x} | \mathbf{x} \in \mathbb{A}_0, \|\mathbf{x}\|_0 = n_{rf}, \mathbf{t}^{\mathbf{x}} \in \mathbb{T}_A\}, \quad (1)$$

where $\mathbf{t}^{\mathbf{x}}$ is the antenna activation pattern (AAP) for the GSM signal vector $\mathbf{x}$ which is an n_t-length binary vector with $\mathbf{t}_i^{\mathbf{x}} = 1$ if $\mathbf{x}_i \in \mathbb{A}$ and '0' otherwise, and $\mathbb{T}_A$ is the set of all valid AAPs. Denoting $\mathbf{H}$ to be the $n_r \times n_t$ MIMO channel matrix, the $n_r \times 1$ received signal vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{x} \in \mathbb{S}$ and $\mathbf{n}$ is an $n_r \times 1$ noise vector. Assuming perfect channel knowledge at the receiver, the maximum likelihood (ML) detection rule for GSM signal detection is given by

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{S}}{\text{argmin}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \quad (3)$$

The ML detection rule in (3) is optimal only when the noise samples across the receive antennas are i.i.d and follow Gaussian distribution. Any deviation in noise from this standard model will result in suboptimal performance when (3) is used. This key observation motivates the use of DL techniques when there is deviation from the standard model. Accordingly, in the following sections, we propose a DNN architecture for GSM signal detection and assess its performance.

III. DNN-based GSM detector

GSM signal detection involves *i*) detecting the set of n_{rf} active antennas and *ii*) detecting n_{rf} modulation symbols $s_1, s_2, \cdots, s_{n_{rf}} \in \mathbb{A}$ transmitted from the active antennas. To do this, we propose the DNN architecture shown in Fig. 1, which comprises of $n_{rf} + 1$ smaller sub-DNNs. One sub-DNN is used to detect the indices of the n_{rf} active antennas (which is shown as AAP-DNN) and n_{rf} sub-DNNs are used for detecting n_{rf} modulation symbols transmitted from the active antennas (which are shown as Sym-1 DNN, $\cdots$, Sym-n_{rf} DNN). All the sub-DNNs have $2n_r$ input neurons through which the real and imaginary parts of the received signal vector are fed as inputs.

AAP-DNN: The AAP-DNN has a set of hidden layers and an output layer with n_t neurons. Each neuron in the output layer corresponds to one transmit antenna and gives the probability of that antenna being active. We use sigmoid activation in the output layer so that the probabilities are independent across the output neurons and need not sum to one. The 'n_{rf} max. indices selector' takes the n_t probability values from the output neurons as input and declares the n_{rf} antennas corresponding to the n_{rf} highest probability values to be active.

Symbol-DNN: Each of the Sym-i, $i = 1, \cdots, n_{rf}$ DNNs has a set of hidden layers and $M = |\mathbb{A}|$ output neurons. Each output neuron of the Sym-i DNN corresponds to one symbol of $\mathbb{A}$ and gives the probability of that symbol being sent from the ith active antenna. Softmax activation is used for the output neurons of the symbol-DNNs. Hence, the probabilities in a given symbol-DNN are dependent across the output neurons and sum to one. Only one of the M neurons in each symbol-DNN will result in a high probability value, which will be declared as the transmitted symbol by the 'max. index selector' followed by the 'index to symbol mapper' blocks.

A key advantage of the proposed DNN-based detector is that it has a modular architecture where the GSM signals are detected using small sub-DNNs instead of one large DNN. For example, consider a GSM system with $n_t = 10$, $n_{rf} = 4$, and 4-QAM. The signal set for this GSM system consists of $2^{\lfloor \log_2 \binom{10}{4} \rfloor + 4 \log_2 4} = 2^{15} = 32768$ signal vectors. It is known from the DL literature that using one-hot encoding for classification leads to excellent performance. For the considered GSM system, using a single DNN to achieve signal detection with one-hot encoding requires using 32768 output neurons. Further, the required number of hidden layers and the number of neurons in each hidden layer scale in proportion to the number of neurons in the output layer. A higher number of layers and a large number of neurons in each layer requires learning a large number of parameters during the training phase. The testing (signal detection) phase also gets complicated proportionately. On the other hand, the proposed modular architecture for the considered GSM system uses five small sub-DNNs, *viz.*, one AAP-DNN and four Symbol-DNNs. The AAP-DNN requires $n_t = 10$ output neurons and the Symbol-DNNs require $|\mathbb{A}| = 4$ output neurons. Therefore, compared to using a single large DNN, using small sub-DNNs requires a lower number of output neurons for each sub-DNN which, in turn, reduces the required number of hidden layers and the number of neurons in each hidden layer. Therefore, the training phase requires learning a lesser number of parameters and the testing phase is also simplified compared to using one large DNN.

Training and Testing: We consider a static/slowly varying channel with a long coherence-time so that the detector can be trained initially with m_T labeled training examples and then subsequently be used for signal detection. In the training phase, the transmitter sends m_T pseudo-random GSM signal vectors known at both the transmitter and the receiver so that they can be used as labels for training the DNN. The received signal vectors generated according to the system model in (2) are used as inputs to train the AAP-DNN and symbol-DNNs. The number of training examples m_T is selected based on experimentation where an initial m_T of 1000 is used and is then increased gradually in steps of 1000 till a good classification (signal detection) performance is achieved. After the training phase, GSM signal vectors selected by the random information bits are transmitted in the testing phase and are detected using the trained DNN. We note that, during the training phase, the

Parameters	APP-DNN	Symbol-DNN
No. of input neurons	$2n_r = 8$	$2n_r = 8$
No. of output neurons	$n_t = 4$	$\lvert\mathbb{A}\rvert = 2$
No. of hidden layers	3	3
Hidden layer activation	ReLU	ReLU
Output layer activation	Sigmoid	Softmax
Optimizer	Adam	Adam
Loss function	Binary cross entropy	Binary cross entropy
No. of training examples	10,000	10,000
Training SNR	10 dB	10 dB
No. of epochs	20	20

TABLE I: DNN parameters of proposed detector in Fig. 2.

DNN learns the mapping from the received signal vectors to the transmitted GSM vectors, which is nothing but learning an equalizer for the channel corresponding to that coherence interval. Since the channel is static/slowly varying, the trained DNN can be used for several channel uses for signal detection until the channel is changed. Therefore, channel need not be explicitly made known at the receiver in the training process. TensorFlow and Keras framework are used for training and testing the proposed DNN architecture.

IV. Results and discussions

A. *BER with i.i.d Gaussian noise*

Figure 2 shows the BER performance of GSM using the proposed DNN-based detector, ML detector, MMSE detector, and a single-DNN-based detector (which uses a single large DNN to jointly detect active antenna indices and modulation symbols). The proposed DNN architecture has three sub-DNNs for this system setting, one sub-DNN for detecting the two active antennas and the other two sub-DNNs for detecting the two BPSK symbols transmitted by the active antennas. The DNN parameters used for the system configuration in Fig. 2 are presented in Table I. The following architectures are used: ***APP-DNN:*** i/p → 16 → ReLU → 16 → ReLU → 8 → ReLU → 4 → sigmoid. ***Symbol-DNN:*** i/p → 16 → ReLU → 16 → ReLU → 8 → ReLU → 2 → softmax. In the above architectures, numbers denote the number of neurons in a given layer, which is followed by the activation function used in that layer. The single-DNN-based detector architecture used is: ***Single-DNN:*** i/p → 32 → ReLu → 32 → ReLu → 64 → ReLu → 64 → ReLu → 32 → ReLu → 16 → Softmax. The channel is considered to be static with the channel gains taking values from an instance of Rayleigh flat fading channel. From Fig. 2, it is seen that the performance of the considered GSM system with the proposed detector is very close to that with the ML detector and is much superior to that with MMSE detector. The single-DNN-based detector performs slightly better than the proposed detector but this comes at the cost of significantly high complexity, e.g., the number of parameters to be learnt in the considered single-DNN-based detector and the proposed detector are 8128 and 1600, respectively. Further, the performance of single-DNN-based detector gets worse and inferior if fewer layers (and hence fewer parameters are to be learnt) are used.

We next consider large-scale GSM systems which use higher number of transmit and receive antennas. Figures 3a and 3b show the BER performance of two large-scale GSM systems using the proposed DNN-based detection. Combinadic encoding proposed in [16] is used in both the systems for low-complexity encoding of information bits to GSM vectors.

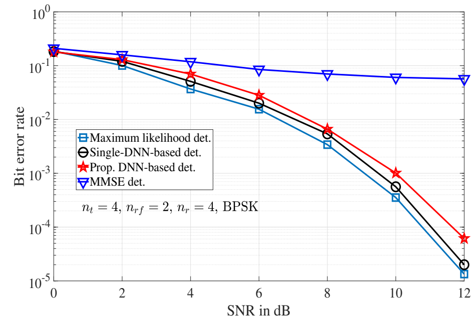
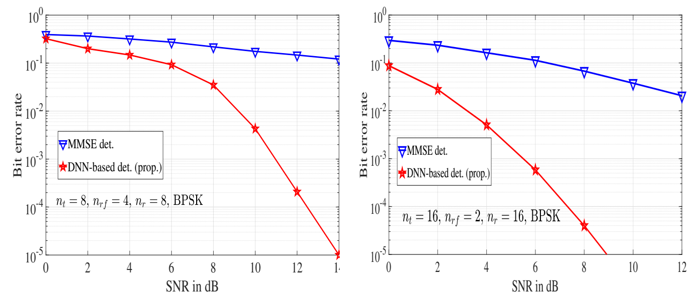


Fig. 2: BER performance of GSM with the proposed modularized DNN-based detector and single-DNN-based detector.



(a) (b)

Fig. 3: BER performance of large-scale GSM systems with the proposed DNN-based detection. Performance with MMSE detector is also shown for comparison.

The BER performance using MMSE detector is also shown for comparison. The DNN parameters used for detection are shown in Table II. The following architectures are used:

1. GSM system in Fig. 3a

APP-DNN: i/p → 128 → ReLU → 64 → ReLU → 32 → ReLU → 16 → ReLU → 8 → sigmoid.

Symbol-DNN: i/p → 32 → ReLU → 16 → ReLU → 8 → ReLU → 4 → ReLU → 2 → softmax.

2. GSM system in Fig. 3b

APP-DNN: i/p → 320 → ReLU → 160 → ReLU → 80 → ReLU → 40 → ReLU → 20 → ReLU → 16 → sigmoid.

Symbol-DNN: i/p → 128 → ReLU → 64 → ReLU → 32 → ReLU → 16 → ReLU → 8 → ReLU → 2 → softmax.

It can be seen from Fig. 3 that the proposed DNN-based detector achieves superior BER performance compared to the MMSE detector for both the GSM system settings.

We compare the complexity (in number of real operations) of the detectors considered in Figs. 3a and 3b in Table III. It can be seen that the proposed detector is computationally efficient compared to the exhaustive search based ML detection for both the systems. MMSE detector is less complex compared to the proposed detector at the cost of degraded BER performance (Fig. 3). On a machine with Intel i5 (5th gen.) processor, the training took less than 30 sec for the system considered in Fig. 2, and about 2-3 min for the systems in Fig. 3.

B. *BER with correlated noise*

Most studies on multiple antenna systems assume the noise across the receive antennas to be i.i.d Gaussian. However, this holds true only for wide antenna spacing which results

Parameters	AAP-DNN	Symbol-DNN
No. of input neurons	Fig. 3a: $2n_r = 16$ Fig. 3b: $2n_r = 32$	Fig. 3a: $2n_r = 16$ Fig. 3b: $2n_r = 32$
No. of output neurons	Fig. 3a: $n_t = 8$ Fig. 3b: $n_t = 16$	$\lvert\mathbb{A}\rvert = 2$
No. of hidden layers	Fig. 3a: 4 Fig. 3b: 6	Fig. 3a: 4 Fig. 3b: 6
Hidden layer activation	ReLU	ReLU
Output layer activation	Sigmoid	Softmax
Optimizer	Adam	Adam
Loss function	Binary cross entropy	Binary cross entropy
No. of training examples	50,000	50,000
Training SNR	Fig. 3a: 10 dB Fig. 3b: 5 dB	Fig. 3a: 10 dB Fig. 3b: 5 dB
No. of epochs	Fig. 3a: 50 Fig. 3b: 10	Fig. 3a: 50 Fig. 3b: 10

TABLE II: DNN parameters of the proposed detector for the GSM systems in Figs. 3a and 3b.

Detector	Complexity for system in Fig. 3a	Complexity for system in Fig. 3b
ML det. (exhaustive search)	327680	294912
MMSE det.	3679	25662
DNN-based det. (prop.)	34832	215876

TABLE III: Complexity (in number of real operations) of algorithms considered in Fig. 3.

in uncoupled antennas. In communication devices (e.g., user equipment), there is generally not sufficient space to maintain wide-antenna spacing to achieve independent noise. Here, we consider the performance of GSM when the noise is correlated across different receive antennas [18],[19]. A noise correlation matrix characterizing this correlation in multi-antenna systems is derived in [18] by using Nyquist's thermal noise theorem. This model depends on the receiver hardware parameters, and hence is not a general model for different hardware implementations. DNNs are relevant in this context as they can learn to map the received signal to the transmitted signal by learning the underlying model including the noise correlation specific to the receiver hardware. Accordingly, we employ the DNN-based detector proposed in Sec. II for GSM signal detection in the presence of correlated noise. For the purpose of illustration, we consider a correlation model where the noise correlation matrix $\mathbf{N}_c$ is of the form

$$\mathbf{N}_c = \begin{bmatrix} 1 & \rho_n & \rho_n^2 & \cdots & \rho_n^{n_r-1} \\ \rho_n & 1 & \rho_n & \cdots & \rho_n^{n_r-2} \\ & & \ddots & & \\ \rho_n^{n_r-1} & \rho_n^{n_r-2} & \cdots & & 1 \end{bmatrix}, \tag{4}$$

where ρ_n ($0 \leq \rho_n \leq 1$) is the correlation coefficient. With this, the correlated noise across the receive antennas is $\mathbf{n}_c = \mathbf{N}_c\mathbf{n}$, where $\mathbf{n}$ is i.i.d Gaussian with its entries from $\mathcal{CN}(0, \sigma^2)$.

Figure 4 shows the BER performance of GSM in the presence of correlated noise, with $\rho_n = 0.4$, when the proposed DNN-based detector is used. The GSM system and DNN architecture parameters considered are the same as those considered in Fig. 2. The performance with the conventional ML detection in (3) with correlated noise is also shown. Note that (3) is optimum only when noise is i.i.d Gaussian. Therefore, we have also shown the performance with modified ML detection, in which the noise correlation matrix Σ is estimated using the procedure in [17] and the following modified (optimal) ML detection rule is used:

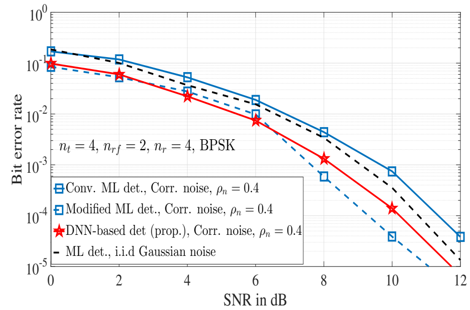


Fig. 4: BER performance of GSM with the proposed DNN-based detector in the presence of correlated noise.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{S}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{H}\mathbf{x})^H \Sigma^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{5}$$

The BER performance of ML detection in (3) with i.i.d Gaussian noise is also shown in the figure for comparison.

The following observations can be made from Fig. 4. First, it can be seen that the performance of GSM using the conventional ML detector in the presence of correlated noise degrades compared to the case with i.i.d Gaussian. This is expected because the ML detector in (3) is optimal when the noise across the receive antennas is i.i.d Gaussian, and using this detector in correlated noise leads to suboptimal detection. Whereas, the performance with the proposed DNN-based detector is better than that with the conventional ML detector and is very close to the performance with the modified (optimal) ML detector in (5). Further, the performance of the proposed detector (and the modified ML detector) with correlated noise is better than that of ML detection with i.i.d Gaussian noise. These observations can be explained as follows. Among all the noise sequences of equal average energy, i.i.d Gaussian noise (uncorrelated noise) is the worst-case noise as it has the maximum entropy [20],[21]. The correlation introduces a structure in the noise, which allows the DNN to learn the model effectively and thus achieve superior performance compared to the case of uncorrelated noise. Although the ML detector can be modified to achieve optimal performance as in (5), it is limited to the small-scale GSM systems because of its high complexity. Whereas the proposed DNN-based detector can be employed in correlated noise without any modifications in the architecture.

C. BER with non-Gaussian noise

We next consider the case when the noise samples across receive antennas are i.i.d, but deviate from Gaussian distribution. Specifically, we consider the case when the noise samples have t-distribution, parameterized by parameter ν. The t-distribution deviates more from the Gaussian pdf for smaller values of ν.

Figure 5 shows the BER performance of GSM using the proposed DNN-based detector when the noise samples are i.i.d across the receive antennas and follow t-distribution with $\nu = 10$ and $\nu = 5$. The considered GSM system and the DNN architecture are the same as those considered in Fig. 2. The performance with ML detection in (3) under i.i.d t-distributed noise as well as i.i.d Gaussian noise are also shown for comparison. From Fig. 5, it can be seen that the performance of ML detector with i.i.d t-distributed noise is inferior compared to that with i.i.d Gaussian noise. It can also be seen

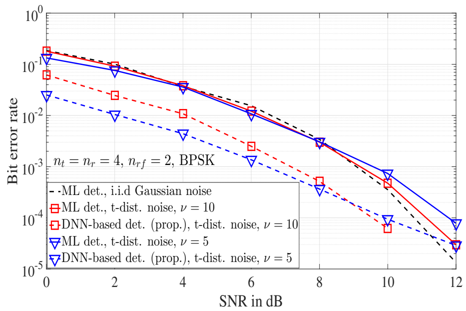


Fig. 5: BER performance of GSM using the proposed DNN-based detector with t-distributed noise.

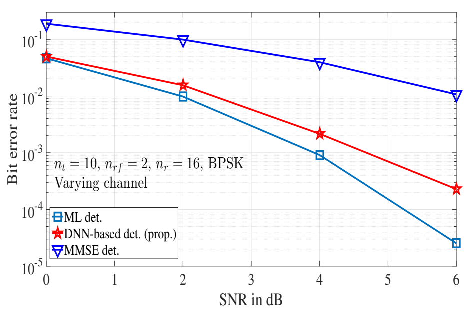


Fig. 6: BER performance of GSM using the proposed DNN-based detector in varying channel.

that smaller the value of ν (more deviation from Gaussian), more is the degradation in the performance of the ML detector. This is mainly because the ML detector in (3) is optimal when the noise is Gaussian distributed and hence deviation from Gaussian distribution results in performance degradation. Whereas, the proposed DNN based detector shows improved BER performance in t-distributed noise compared to that in Gaussian noise. As discussed earlier, the Gaussian noise is the worst-case noise among all the noise distributions for a given variance. Therefore, learning in a non-Gaussian noise is more effective, which leads to superior BER performance.

D. *Extension to varying channels*

It has been shown in [22] that, using a preprocessing on the received vector before feeding it to the DNN can reduce the input dimensions, and can enable the DNN to achieve signal detection in varying channels (VC). Here, we use the MMSE solution as the preprocessing step to achieve dimensionality reduction. We assume that the channel is known at the receiver, but the channel realizations change from one instant to the other. We feed $\mathbf{z} = (\mathbf{H}^H\mathbf{H} + \frac{1}{\text{snr}}\mathbf{I}_{n_t})\mathbf{H}^H\mathbf{y}$ as the input to the DNN architecture in Fig. 1 during both training and testing phases. Figure 6 shows the BER performance of GSM with the proposed detection architecture using MMSE preprocessing in VC. The DNN has 3 sub-DNNs and each of them uses 5 hidden layers with 320, 256, 128, 64, and 32 neurons in layers 1, 2, 3, 4, and 5, respectively. It can be seen that the GSM system with the proposed architecture achieves good performance in VC using only 5 hidden layers, unlike the detector in [22] which uses 30 layers irrespective of the input and output dimensions.

V. Conclusions

We proposed a novel modularized DNN-based GSM signal detection scheme. Due to its inherent ability to effectively learn the underlying noise models in practical receivers, the proposed detector achieved robust and better BER performance compared to ML detection performance when deviations from the standard model are witnessed. There are several potential directions for future work. These include the use of convolutional neural networks in the proposed modularized architecture, performance of the proposed detector architecture under different channel models, and extension of the detector architecture for GSM-based massive MIMO systems.

References

[1] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693-16746, Aug. 2017.

[2] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Tech.*, vol. 57, no. 4, pp. 2228-2241, 2008.

[3] M. Wen, B. Zheng, K. J. Kim, M. D. Renzo, T. A. Tsiftsis, K. C. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging wireless systems: research progresses and applications," *IEEE J. Sel. Areas in Commun.,* vol. 37, no. 9, pp. 1949-1972, Sep. 2019.

[4] L. Xiao, P. Xiao, Z. Liu, H. Haas, and L. Hanzo, "A compressive sensing assisted massive SM-VBLAST system: error probability and capacity analysis," *IEEE Trans. Wireless Commun.,* vol. 19, no. 3, pp. 1990-2005, Mar. 2020.

[5] T. Datta and A. Chockalingam, "On generalized spatial modulation," *Proc. IEEE WCNC'2013*, pp. 2716-2721, Apr. 2013.

[6] Y. Xiao, Z. Yang, L. Dan, P. Yang, L. Yin, and W. Xiang, "Low-complexity signal detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 403-406, Apr. 2014.

[7] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98-105, Apr. 2017.

[8] S. Dorner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning-based communication over the air," *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 1, pp. 132-143, Feb. 2018.

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

[10] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pp. 298-303, Oct. 2018.

[11] H. Kim, Y. Jiang, R. Rana, S. Kannan, and P. Viswanath, "Communication algorithms via deep learning," *Proc. ICLR'2018*, pp. 1-17, 2018.

[12] 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, Sep. 2018.

[13] P. Yang, Y. Xiao, M. Xiao, Y. L. Guan, S. Li, and W. Xiang, "Adaptive spatial modulation MIMO based on machine learning," *IEEE J. Sel. Areas Commun.,* vol. 37, no. 9, pp. 2117-2131, Sep. 2019.

[14] L. Xiao, P. Yang, Y. Xiao, S. Fan, M. Di Renzo, W. Xiang, and S. Li, "Efficient compressive sensing detectors for generalized spatial modulation systems," *IEEE Trans. Veh. Tech.,* vol. 66, no. 2, pp. 1284-1298, Feb. 2017.

[15] N. T. Nguyen and K. Lee, "Deep learning-aided tabu search detection for large MIMO systems," *IEEE Trans. Wireless Commun.,* vol. 19, no. 6, pp. 4262-4275, Jun. 2020.

[16] T. Lakshmi Narasimhan and A. Chockalingam, "On the capacity and performance of generalized spatial modulation," *IEEE Commun. Lett.,* vol. 20, no. 2, pp. 252-255, Feb. 2016.

[17] M. L. Ammari and P. Fortier, "Analysis of MIMO receiver using generalized least squares method in colored environments," *J. Comput. Netw. and Commun.,* pp. 1-6, Aug. 2014.

[18] S. Krusevac, P. Rapajic, and R. A. Kennedy, "Channel capacity estimation for MIMO systems with correlated noise," *Proc. IEEE GLOBECOM'05*, pp. 2812-2816, Dec. 2005.

[19] C. P. Domizioli, B. L. Hughes, K. G. Gard, and G. Lazzi, "Receive diversity revisited: correlation, coupling and noise," *Proc. IEEE GLOBECOM'2007*, pp. 3601-3606, Dec. 2007.

[20] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379-423, 1948.

[21] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd Ed.*, Wiley Publishers, Jul. 2006.

[22] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.,* vol. 67, no. 10, pp. 2554-2564, May 2019.