

Copyright Notice

©2019 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

On Unsupervised Deep Learning Solutions for Coherent MU-SIMO Detection in Fading Channels

Songyan Xue, Yi Ma, Ang Li, Na Yi, and Rahim Tafazolli

Institute for Communication Systems (ICS), University of Surrey, Guildford, England, GU2 7XH

E-mail: (songyan.xue, y.ma, ang.li, n.yi, r.tafazolli@surrey.ac.uk)

Abstract—In this paper, unsupervised deep learning solutions for multiuser single-input multiple-output (MU-SIMO) coherent detection are extensively investigated. According to the ways of utilizing the channel state information at the receiver side (CSIR), deep learning solutions are divided into two groups. One group is called equalization and learning, which utilizes the CSIR for channel equalization and then employ deep learning for multiuser detection (MUD). The other is called direct learning, which directly feeds the CSIR, together with the received signal, into deep neural networks (DNN) to conduct the MUD. It is found that the direct learning solutions outperform the equalization-and-learning solutions due to their better exploitation of the sequence detection gain. On the other hand, the direct learning solutions are not scalable to the size of SIMO networks, as current DNN architectures cannot efficiently handle many co-channel interferences. Motivated by this observation, we propose a novel direct learning approach, which can combine the merits of feedforward DNN and parallel interference cancellation. It is shown that the proposed approach trades off the complexity for the learning scalability, and the complexity can be managed due to the parallel network architecture.

I. INTRODUCTION

Multiuser multiple-input multiple-output (MU-MIMO) signal detection over noisy fading channels is mathematically an integer least-squares (ILS) problem, which aims to minimize the pairwise Euclidean distance between the transmitted signal multiplied by channel matrix and the received signal [1]. Concerning the maximum-likelihood solution to be computationally too expensive, it is usual practice to employ linear algorithms such as the matched filter (MF), zero forcing (ZF), and linear minimum mean-square error (LMMSE) to trade off the optimality for lower computational complexity. Concerning the linear algorithms often too sub-optimum due to their uses of symbol-by-symbol detection, enormous research efforts have been paid in the last two decades, to achieve the best performance-complexity trade-off through the use of nonlinear sub-optimal algorithms such as V-BLAST [2], LMMSE-SIC [3], fixed-complexity sphere decoding (SD) [4], lattice-reduction (LR) aided detection algorithms [5], and so forth. A quite comprehensive survey of the MIMO detection algorithms can be found in [6]; and we can reach two conclusions: 1) nonlinear algorithms are too complex to afford by the current DSP technology, and 2) they do not support well parallel computing which is however the trend of the future DSP technology.

Recent advances towards the multiuser-MIMO detection problem lie in the use of deep learning. Notable achievements

so far include model-driving coherent MIMO detection [7]–[9] and deep learning-aided iterative detection [10], where deep learning plays a central role of signal detection after the channel equalization. Deep learning has also been employed for joint MU-MIMO transmitter and receiver design in [11], which extends the autoencoder-based approach originally proposed in [12] for single-input single-output (SISO) systems. In addition to the application for MIMO detection, there are growing activities of employing deep learning for the physical layer design, which include recent works on millimeter-wave channel estimation [13], multiuser-frequency synchronization [14], FEC decoding [15], blind sequence detection [16] and channel state information feedback [17], [18]. A relatively comprehensive survey of deep learning techniques for wireless communications can be found in [19], [20].

With specific to the MU-MIMO detection, deep learning-aided solutions have demonstrated their potential of offering computational complexity close to linear receivers' in the communication procedure, without compromising the detection performance. Moreover, many deep learning solutions are parallel computing ready. Despite their advantages, major challenges of deep learning solutions lie in their learning complexity as well as training data sets, which scale poorly in multiuser MIMO fading channels.

The aim of this paper is to investigate fundamental behaviors of deep learning for MU-MIMO detection through intensive study of the deep learning gain, training complexity and data sets scalability. According to the ways of utilizing channel state information at the receiver side (CSIR), we classify deep learning solutions into two categories: channel equalization and learning (CE-L) and direct learning (Direct-L), with the following definition.

Definition 1: It is called the CE-L mode when deep learning modules are applied after the channel equalization, and they do not utilize CSIR to conduct the signal classification (or equivalently the detection); see Fig. 1. When CSIR and received signals are both the input to deep learning modules to enable the signal classification, such is called the Direct-L mode; see Fig. 2.

Major contributions of this paper lie in:

1) The analysis of deep learning gain in the CE-L mode, where deep learning modules are employed after the MF, ZF, or LMMSE channel equalizers. Such combinations are rather trivial and not novel. However, we show that the deep learning gain comes mainly from the sequence detection gain, which

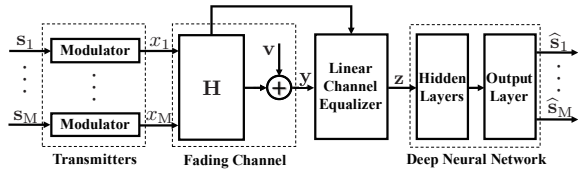


Fig. 1. Block diagram of the CE-L mode for MU-SIMO detection

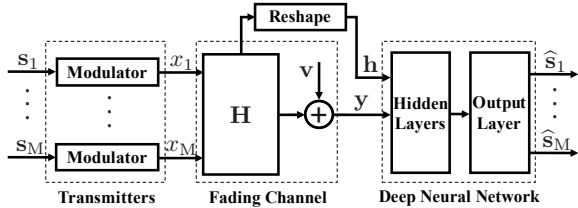


Fig. 2. Block diagram of the Direct-L mode for MU-SIMO detection.

can largely improve the performance of the MF-based receiver (around 8 dB), but not for the ZF or LMMSE ones.

2) An extensive study of the Direct-L mode in terms of their performances and scalability. It is shown that the Direct-L mode is able to achieve the maximum-likelihood performance for very small MIMO (i.e., a system with 2 users) in fading channels. However, the Direct-L mode is not scalable to the size of MIMO networks due to the growth of dynamic interferences.

3) The development of a novel Direct-L approach, which efficiently combines the feed-forward deep neural network (FF-DNN) with the parallel interference cancellation (PIC) to scale up the deep learning solutions. It is shown that the proposed DNN-PIC approach can achieve near-optimum performances, with their MIMO detection complexity growing in a linear order of the number of users in the network.

In addition, it is observed that the Direct-L approaches can work with the partial CSIR (i.e., the CSIR is known only for several user-to-receiver links but not all). The users with known CSIR can enjoy much better performances than those without the CSIR. This interesting result is the evidence showing that deep learning utilizes the CSIR as the user signature to assist the signal classification.

II. MU-SIMO UPLINK MODEL AND MACHINE LEARNING-AIDED DETECTION

A. MU-SIMO System Model and Optimum Detection

Consider a MU-SIMO system¹, where M UTs simultaneously talk to the uplink receiver equipped with N receive antennas ($N \geq M$). The discrete-time equivalent model of such MU-SIMO system can be expressed by the following matrix form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v} \quad (1)$$

where $\mathbf{x} \triangleq [x_0, \dots, x_{M-1}]^T$ stands for a block of transmitted complex-valued symbols with covariance $\sigma_x^2 \mathbf{I}$, with each

¹We assume each user having a single transmit-antenna to focus our presentation on the key ideas. An extension to MU-MIMO network is trivial.

symbol independently drawn from a finite-alphabet set \mathcal{A} of the size $L = |\mathcal{A}|$, $\mathbf{y} \triangleq [y_0, \dots, y_{N-1}]^T$ for the spatial-domain received block, \mathbf{H} for a $(N) \times (M)$ MU-SIMO random channel matrix, and \mathbf{v} for the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma_v^2 \mathbf{I}$. Moreover, the superscript $[\cdot]^T$ stands for the matrix/vector transpose, and \mathbf{I} for the identity matrix.

The MU-SIMO detection problem is to form the decision $\hat{\mathbf{x}}$ based upon the received block \mathbf{y} as well as the channel matrix \mathbf{H} . The maximum-likelihood solution (or equivalently the sphere decoding) is ILS-optimum by achieving the following objective

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{A}^M} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \quad (2)$$

where $\|\cdot\|$ denotes the Euclidean norm.

B. Machine Learning-Aided Detection and Challenges

In theory, a machine is able to learn a Bayesian-optimum solution (equivalent to the maximum-likelihood solution as far as the AWGN case is concerned) to the ILS problem (2). The basic principle of machine learning can be stated by:

Proposition 1 (See [11]): Given the channel matrix \mathbf{H} and the training data set $\mathbf{x} \in \mathcal{A}$, machine learning will establish the link between \mathbf{y} and $\mathbf{H}\mathbf{x}$ according to the maximum a posteriori probability $p(\mathbf{H}\mathbf{x}|\mathbf{y})$.

For a fixed channel matrix \mathbf{H} , we have the maximum a posteriori probability $p(\mathbf{H}\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\mathbf{y})$; and in this case, the training data set grows exponentially with the number of users M . Nevertheless, the size of training data set is finite, and as usual practice, the machine learning complexity can be managed through parallel computing.

The major challenge arises when the channel matrix \mathbf{H} is randomly varying, as such effectively renders the training data set infinite. More seriously, the randomness of \mathbf{H} can result in channel ambiguity, i.e., the received signal \mathbf{y} might have multiple combinations of \mathbf{H} and \mathbf{x} even in the noiseless case, in which machine learning will fail to conduct the signal classification. Theoretically, the channel ambiguity can be resolved if the machine is given the full channel knowledge as a part of the training data, as we have the relationship $p(\mathbf{H}\mathbf{x}|\mathbf{y}) = p(\mathbf{x}|\mathbf{y}, \mathbf{H})$ held. However, the \mathbf{H} -defined training data grows in a square-order of M , while the information data \mathbf{x} grows only linearly. We call such a phenomenon ‘CSIR flooding’, which will confuse the machine at the learning stage. This problem, if not get appropriately resolved, will induce poor scalability to the machine learning procedure. In Section III, we will study how and to what extent the machine learning-aided techniques (as a promising machine learning technology) can tackle this problem.

III. MACHINE LEARNING-AIDED MU-SIMO DETECTION

A. The CE-L Mode

Definition 1 indicates that the CE-L mode has two steps:

$$\text{Channel equalization: } \mathbf{z} = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{H}\mathbf{x} + \mathbf{W}\mathbf{v} \quad (3)$$

$$\text{Detection: } \hat{\mathbf{x}} = \Gamma(\mathbf{z}) \quad (4)$$

where \mathbf{z} stands for the channel equalized signal block, $\Gamma(\cdot)$ for the decision function which can be replaced by a machine-learned function if appropriate, \mathbf{W} for a linear filter specified by

$$\mathbf{W} \triangleq \begin{cases} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H, & \text{ZF equalizer} \\ (\mathbf{H}^H \mathbf{H} + \sigma_v^2 \mathbf{I})^{-1} \mathbf{H}^H, & \text{LMMSE equalizer} \\ \mathbf{H}^H, & \text{MF equalizer} \end{cases} \quad (5)$$

and the superscript $[\cdot]^H$ denotes the matrix/vector Hermitian transpose. Now, the signal detection problem becomes

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{A}^M} \|\mathbf{z} - \mathbf{W}\mathbf{H}\mathbf{x}\|^2 \quad (6)$$

The signal detection performance depends on the ways of making decision. When the decision is made upon each individual user, the performance can be evaluated by the signal-to-interference and noise ratio (SINR)

$$\text{sinr}_m = \frac{\gamma_0 [\mathbf{W}\mathbf{H}\mathbf{H}^H \mathbf{W}^H]_m \det(\mathbf{Q})}{\det(\tilde{\mathbf{Q}})}, \quad m=0, \dots, M-1 \quad (7)$$

$$\mathbf{Q} = \mathbf{W}\mathbf{W}^H + \gamma_0 (\mathbf{W}\tilde{\mathbf{H}})(\mathbf{W}\tilde{\mathbf{H}})^H \quad (8)$$

where sinr_m stands for the instantaneous SINR for the m^{th} user, $[\cdot]_m$ for the $(m, m)^{\text{th}}$ diagonal entry of a matrix, $\tilde{\mathbf{H}}$ for the matrix formed by removing the m^{th} column of \mathbf{H} , $\tilde{\mathbf{Q}}$ for the matrix formed by removing the m^{th} column of \mathbf{Q} , and $\gamma_0 \triangleq (\sigma_x^2)/(\sigma_v^2)$. Then, the system capacity is given by

$$C_s = \sum_{m=0}^{M-1} \log(1 + \text{sinr}_m) \quad (9)$$

When the decision is made upon the sequence \mathbf{z} , the system capacity is given by

$$C_{\text{seq}} = \log \left(\frac{\det(\mathbf{W}\mathbf{W}^H + \gamma_0 \mathbf{W}\mathbf{H}\mathbf{H}^H \mathbf{W}^H)}{\det(\mathbf{W}\mathbf{W}^H)} \right) \quad (10)$$

The sequence detection gain can be evaluated by

$$G = C_{\text{seq}} - C_s \quad (11)$$

Proposition 1 shows that machine learning maps the sequence \mathbf{z} onto the sequence \mathbf{x} according to the the maximum a posteriori probability and follows exactly the sequence detection principle. Hence, we can also call G the machine learning gain.

1) *The ZF-L mode:* Considering \mathbf{W} to be the linear ZF filter, we will have $[(\mathbf{W}\tilde{\mathbf{H}})(\mathbf{W}\tilde{\mathbf{H}})^H]_m = 0$ and $[(\mathbf{W}\mathbf{H})(\mathbf{W}\mathbf{H})^H]_m = 1$, and can easily obtain $C_{\text{seq}} = C_s$ through (7) and (10). This basically means no machine learning gain as $G = 0$.

2) *The LMMSE-L mode:* The only difference between the LMMSE-L mode and the ZF-L mode lies in the regularization factor $(\sigma_v^2 \mathbf{I})$ in (5). For the high SNR regime where the regularization factor becomes negligible, we can reach the same conclusion as the ZF-L mode, i.e., $G = 0$. For the low SNR regime, sequence detections do not have considerable gain, as they take advantage of more degrees-of-freedom (DoF) due to signal correlations, which dominate the performance only at high SNRs.

3) *The MF-L mode:* Given $\mathbf{W} = \mathbf{H}^H$, the SINR term in (7) becomes

$$\text{sinr}_m = \frac{\gamma_0 [\mathbf{H}^H \mathbf{H}\mathbf{H}^H \mathbf{H}]_m \det(\mathbf{Q}_{\text{MF}})}{\det(\tilde{\mathbf{Q}}_{\text{MF}})} \quad (12)$$

$$\mathbf{Q}_{\text{MF}} = \mathbf{H}^H \mathbf{H} + \gamma_0 (\mathbf{H}^H \tilde{\mathbf{H}})(\tilde{\mathbf{H}}^H \mathbf{H}) \quad (13)$$

and $\tilde{\mathbf{Q}}_{\text{MF}}$ is the corresponding version of \mathbf{Q}_{MF} . At high SNRs, (13) approximates to

$$\mathbf{Q}_{\text{MF}} \approx \gamma_0 \mathbf{H}^H \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H \mathbf{H} \quad (14)$$

Applying (14) into (12) and (9) leads to

$$C_s \approx \sum_{m=0}^{M-1} \log \left(1 + \frac{[\mathbf{H}^H \mathbf{H}\mathbf{H}^H \mathbf{H}]_m}{\det(\tilde{\mathbf{Q}}_{\text{MF}})/\det(\mathbf{H}^H \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H \mathbf{H})} \right) \quad (15)$$

Similarly, we can apply $\mathbf{W} = \mathbf{H}^H$ into (10) and obtain

$$C_{\text{seq}} = \log \left(\frac{\det(\mathbf{H}^H \mathbf{H} + \gamma_0 \mathbf{H}^H \mathbf{H}\mathbf{H}^H \mathbf{H})}{\det(\mathbf{H}^H \mathbf{H})} \right) \quad (16)$$

$$= \sum_{m=0}^{M-1} \log(1 + \gamma_0 \lambda_m) \quad (17)$$

where λ_m denotes the m^{th} eigenvalue of $(\mathbf{H}^H \mathbf{H})$. With the increase of γ_0 , C_{seq} grows in a logarithmic scale, and C_s keeps constant. It is trivial to justify $C_{\text{seq}} > C_s$ at high SNRs. This means that there is a potential machine-learning gain in the MF-L mode.

B. The Direct-L Mode

According to *Definition 1*, the Direct-L mode applies the machine-learned detection function directly on the received block \mathbf{y} , i.e., $\hat{\mathbf{x}} = \Gamma(\mathbf{y}|\mathbf{H})$. The data processing inequality implies: $I(\mathbf{y}; \mathbf{x}|\mathbf{H}) \geq I(\mathbf{z}; \mathbf{x}|\mathbf{H})$, where $I(\cdot)$ denotes the mutual information. Therefore, the Direct-L mode should outperform the CE-L mode from the information-theoretic point of view. In Section IV, we will study how to employ deep learning to realize the MU-SIMO detection function, and what are the pros/cons for both the CE-L and Direct-L modes.

IV. DEEP LEARNING FOR MU-SIMO DETECTION

Deep learning is a special paradigm of machine learning which employs a well-trained DNN to replace the decision function $\Gamma(\mathbf{z})$ in the CE-L mode or $\Gamma(\mathbf{y}|\mathbf{H})$ in the Direct-L mode. The learning results are stored in form of weighting coefficients and biases of neural networks.

A. Deep Learning for the CE-L Mode

Fig. 1 illustrates the block diagram of the CE-L mode for the MU-SIMO detection. There are a set of independent user terminals (UTs) at the transmitter side. Each UT modulates their information bits, \mathbf{s}_m , into the symbol, x_m , and transmits them through the mobile fading channel \mathbf{H} . At the receiver side, the received signal \mathbf{y} is first fed into the linear channel equalizer, of which the output, \mathbf{z} , is then fed into the DNN-based decision module, $\Gamma(\mathbf{z})$. The output of $\Gamma(\mathbf{z})$ is the binary decision, $\hat{\mathbf{s}}_m$, instead of the modulation symbol, x_m . This

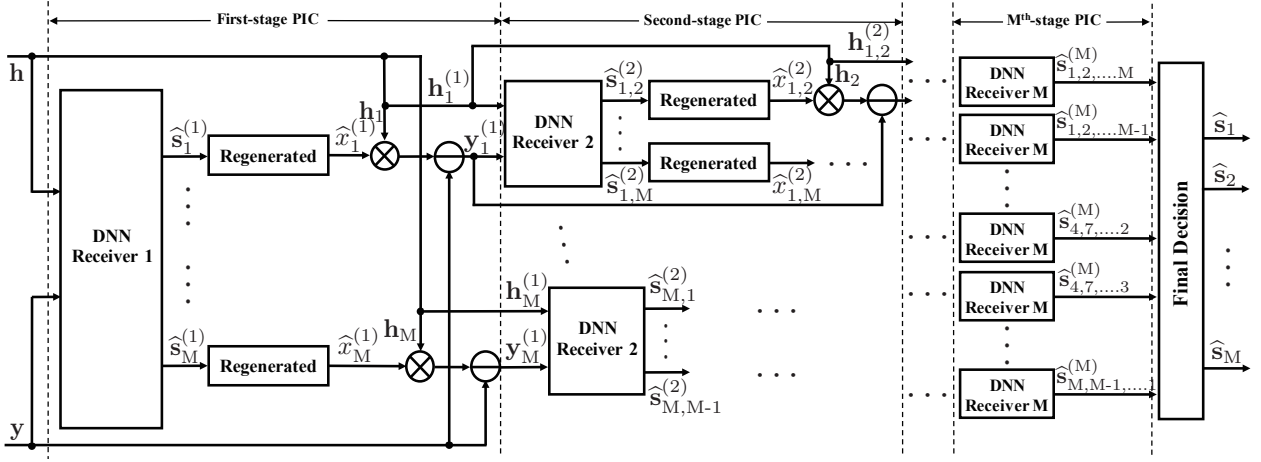


Fig. 3. Block diagram of the novel Direct-L algorithm combining DNN with PIC.

is because binary bits are the ultimate goal of the receiver procedure. It is perhaps worth noting that communication signals are often complex valued, and current neural networks are based on real-value operations. Thus, a $(M) \times (1)$ complex signal block must be represented by a $(2M) \times (1)$ real-valued block (see [1]) to facilitate the learning and communication procedure.

The deep learning procedure can be found in Section V (Experiment 1). In the CE-L mode, the training data is \mathbf{z} , which is the superposition of \mathbf{x} and the Gaussian noise $(\mathbf{W}\mathbf{v})$. As also discussed in Section III, such forms the major advantage of the CE-L mode, as the training data set is effectively a noisy version of \mathbf{x} , which is drawn from a finite-alphabet set. By such means, we can avoid or largely mitigate the channel model overfitting problem².

B. Deep Learning for the Direct-L Mode

Fig. 2 illustrates the block diagram of the Direct-L mode for MU-SIMO detection. It differs from the CE-L mode mainly at the receiver side, where there are two inputs to the DNN-based decision module, i.e., \mathbf{y} and \mathbf{H} . Concerning the input layer of DNN often a column vector, we suggest reshape the channel matrix \mathbf{H} into a $(NM) \times (1)$ column vector denoted by \mathbf{h} , and then form a super vector $[\mathbf{y}^T, \mathbf{h}^T]^T$ to serve as the input to the DNN in both the learning and communication procedure. Although \mathbf{h} is involved in the learning procedure, the learning objective is to minimize the difference (e.g., the cross-entropy) between the original information bits and the decoded bits only.

Simulation results in Section V show that the above design is able to mitigate the channel ambiguity and offer better detection performance than the CE-L mode. Moreover, this approach could work even in the absence of full CSIR, i.e., only part of \mathbf{h} is available at the receiver side; see detail in Section V (Experiment 2).

²The channel model overfitting problem occurs when a DNN trained for a special channel model is not suitable for another channel model.

Despite those remarkable advantages, the Direct-L mode is not scalable to the number of UTs due to the so called ‘‘CSIR flooding’’ problem, highlighted in Section I, which largely reduce the inter-user interference cancellation ability of the DNN. This problem motivates us to develop a novel approach combining the merits of direct learning and PIC.

C. The DNN-PIC Approach

Fig. 3 illustrates the block diagram of the proposed DNN-PIC receiver. The entire receiver consists of M cascaded PIC stages, with each employing a group of identical pre-trained DNNs for signal detection. All of the DNNs share the same structure as the DNN used in the Direct-L mode (see Fig. 2). The input to the first-stage consists of the full CSIR, \mathbf{h} , and the received signal, \mathbf{y} . The output of the first stage is the estimated information bits $\hat{\mathbf{s}}_m^{(1)}$, $1 \leq m \leq M$, where the superscript $[\cdot]^{(1)}$ stands for the stage number. Denote \mathbf{h}_m to the channel vector between the m^{th} UT to the receiver, and $\hat{\mathbf{x}}_m^{(1)}$, $1 \leq m \leq M$ to the regenerated transmitted symbols. We can subtract the user interference from the received signal by

$$\mathbf{y}_m^{(1)} = \mathbf{y} - \mathbf{h}_m \hat{\mathbf{x}}_m^{(1)}, \quad 1 \leq m \leq M \quad (18)$$

which, together with the corresponding CSIR, $\mathbf{h}_m^{(1)}$, serves the input to the second-stage DNNs. Note that $\mathbf{h}_m^{(1)}$ is formed by removing \mathbf{h}_m from \mathbf{h} . Repeating the interference cancellation for M times, we will be able to obtain $M!$ estimates of the information-bearing symbols as depicted in Fig. 3. The final decision is made by taking average of the relevant estimates.

The proposed DNN-PIC receiver has a parallel computing structure, which is an advantage to reduce the computational latency. Despite, the computational complexity of the ‘‘tree’’ traversal seems to be still too expensive. According to the system model in Section II-A, the MU-SIMO channel is randomly generated, and \mathbf{x} is drawn independently from a finite-alphabet set with the equal probability. The Direct-L mode achieves the same average error probability for each independent user. Therefore, the DNN-PIC receiver can be simplified into two ‘‘branches’’ with no performance penalty.

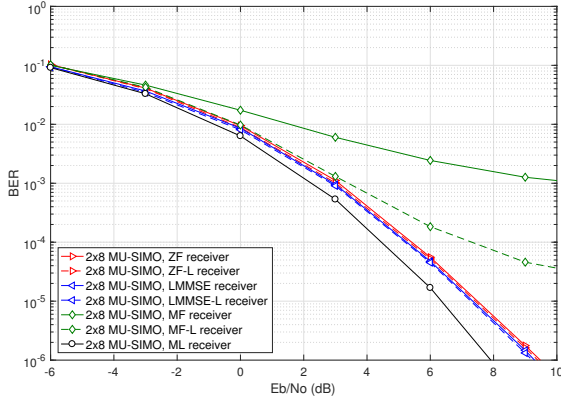


Fig. 4. BER as a function of E_b/N_0 for the CE-L mode in uncoded 2-by-8 MU-SIMO system.

V. COMPUTER SIMULATIONS AND DISCUSSION

Our computer simulations are structured into three experiments with respect to the CE-L mode, the Direct-L mode and the proposed DNN-PIC approach. The performance is evaluated using the bit error rate (BER) averaging over sufficient Monte-Carlo trials of block Rayleigh fading channels, and compared to widely used baselines under different channel conditions. The SNR is defined as the average received bit-energy to noise ratio (E_b/N_0).

Experiment 1 (The CE-L Mode): In this experiment, we evaluate the BER performance of the CE-L mode under BPSK modulation according to the setup specified in Section IV-A and Fig. 1. Training is conducted using mini-batch stochastic gradient descent with Adam optimizer at the learning rate $\alpha = 0.001$ (see [21] for the detailed description). The loss function is categorical cross-entropy. The mini-batch size of each epoch is 1000, and the DNN is trained at $E_b/N_0 = 5$ dB; as this configuration is found to provide the best performance.

Fig. 4 shows the BER performance of the CE-L mode with different channel equalizers. The baseline for performance comparison is obtained using conventional symbol-by-symbol detection algorithm. It is shown that deep learning modules largely improve the performance of the MF-based receiver (around 8 dB at BER of 10^{-3}). The performance gain mainly comes from the sequence detection as we discussed in Section III-A. For both ZF and LMMSE receiver, the deep learning gain is almost negligible since there is no correlation among the equalized symbols.

Experiment 2 (The Direct-L Mode): In this experiment, we evaluate the BER performance of the Direct-L mode under BPSK modulation according to the setup specified in Section IV-B and Fig. 2. Instead of using linear equalizer, we give the DNN receiver full channel knowledge as a part of the training data to resolve channel ambiguity. Moreover, we demonstrate that the Direct-L approach can work with partial CSIR (i.e., the CSIR is known for part of the user-to-receiver links). The DNN is trained at $E_b/N_0 = 0$ dB, and the rest of the network setup remains the same as Experiment 1.

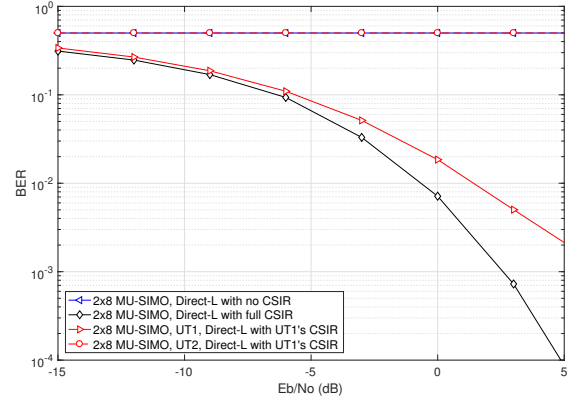


Fig. 5. BER as a function of E_b/N_0 for the Direct-L mode in uncoded 2-by-8 MU-SIMO system.

Fig. 5 shows the BER performance of the Direct-L mode with different levels of CSIR. The baseline for performance comparison is the maximum likelihood receiver. For the uncoded 2-by-8 MU-SIMO communications, it is found that DNN receiver fails to detect the transmitted waveform without CSIR since the received signals might have multiple combinations of channel and transmitted signals. Consider the receiver has full CSIR, the Direct-L approach achieves near-optimal performance almost throughout the whole SNR range. The most interesting phenomenon is that the Direct-L approach can work with partial CSIR. Consider the CSIR is known only for UT1-to-receiver link, we observe that UT1 enjoys a much better performance than UT2. The performance gap between UT1 and the ML receiver is caused by the co-channel interference from UT2. This result shows that deep learning utilizes the CSIR as the user signature to assist the signal classification.

Experiment 3 (The proposed DNN-PIC receiver): In this experiment, we conduct the proposed DNN-PIC receiver in uncoded 4-by-8 MU-SIMO system based on the system configurations specified in Section IV-C and Fig. 3. With the spatial-domain user increasing, we observe that the BER performance is highly influenced by the training SNR point. For all the BER performance plots in this experiment, the training results from three different E_b/N_0 points (e.g., 0 dB, 5 dB and 10 dB) are merged to yield the best performance throughout the whole SNR range.

Fig. 6 shows the BER performance of the proposed DNN-PIC approach in uncoded 4-by-8 MU-SIMO system. The baselines for performance comparison includes: the coherent ZF receiver, the maximum likelihood receiver as well as the Direct-L approach. It is shown that the performance of the Direct-L approach quickly moves away from optimum with increasing spatial-domain user load. This is potentially caused by the CSIR flooding which confuses the machine at the learning stage. Meanwhile, the proposed DNN-PIC approach largely improves the detection performance of the Direct-L approach which is around 1.5 dB for BPSK and 3 dB for

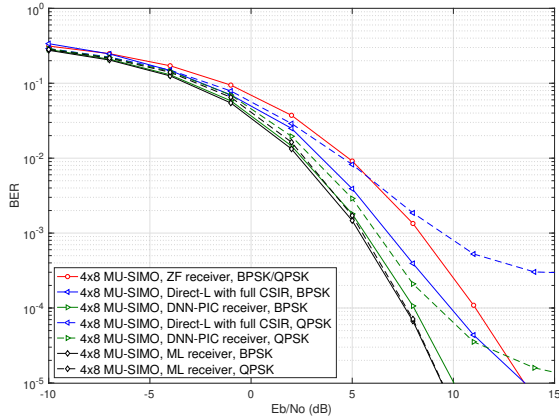


Fig. 6. BER as a function of E_b/N_0 for the proposed DNN-PIC receiver in uncoded 4-by-8 MU-SIMO system.

QPSK at BER of 10^{-3} . In addition, the performance gap between the proposed approach and the maximum-likelihood receiver is around 0.2 dB and 1.5 dB at BER of 10^{-3} for BPSK and QPSK, respectively.

VI. CONCLUSION

In this paper, we have conducted an extensive study of deep learning solutions for MU-SIMO coherent detection with specific to the EC-L mode and the Direct-L mode. It has been shown that the machine learning (or deep learning) gain mainly comes from the nature of sequence detection (or classification) inherent in the learning procedure. This conclusion is justified through theoretical analysis and confirmed by computer simulations. It has been shown that deep learning could not bring extra gain to the ZF or LMMSE MU-SIMO receiver due to their nature of enabling the symbol-by-symbol detection; however, it could contribute significant gain (8 dB in E_b/N_0) to the MF receiver due to a better use of the residual interferences between users to form the sequence detection. Given the same principle, the Direct-L mode could enjoy fully the sequence detection gain in theory, but face the practical challenge of DNN scalability to the number of users. To tackle this problem, a novel DNN-PIC approach has been proposed to enrich the Direct-L mode. The proposed approach exhibited near-optimum performances with linearly growing complexity to the number of users.

ACKNOWLEDGEMENT

The work was supported in part by European Commission under the framework of the Horizon2020 5G-Drive project, and in part by 5G Innovation Centre (5GIC) HEFEC grant.

REFERENCES

- [1] J. C. D. L. Ducoing, Y. Ma, N. Yi, and R. Tafazolli, "A real-complex hybrid modulation approach for scaling up multiuser MIMO detection," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3916–3929, Sept 2018.
- [2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct 1998, pp. 295–300.
- [3] T. Liu and Y. Y. Liu, "Modified fast recursive algorithm for efficient MMSE-SIC detection of the V-BLAST system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3713–3717, October 2008.
- [4] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug 2005.
- [5] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proceedings 2003 IEEE Information Theory Workshop*, March 2003, pp. 345–348.
- [6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [7] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *IEEE 18th Int. Workshop on SPAWC*, July 2017, pp. 1–5.
- [8] H. He, C.-K. Wen, S. Jin, and G. Li, "A model-driven deep learning network for MIMO detection," *CoRR*, vol. abs/1809.09336, 2018.
- [9] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *CoRR*, vol. abs/1805.07631, 2018.
- [10] M. Imanishi, S. Takabe, and T. Wadayama, "Deep learning-aided iterative detector for massive overloaded MIMO channels," *CoRR*, vol. abs/1806.10827, 2018.
- [11] S. Xue, Y. Ma, N. Yi, and R. Tafazolli, "Unsupervised deep learning for MU-SIMO joint transmitter and noncoherent receiver design," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2018.
- [12] 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.
- [13] 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, no. 5, pp. 852–855, Oct 2018.
- [14] A. Li, Y. Ma, S. Xue, N. Yi, and R. Tafazolli, "A carrier-frequency-offset resilient OFDMA receiver designed through machine deep learning," in *IEEE PIMRC*, Sept. 2018, pp. 1–5.
- [15] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Beery, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 119–131, Feb 2018.
- [16] 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.
- [17] 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.
- [18] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2018.
- [19] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, Nov 2017.
- [20] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *CoRR*, vol. abs/1803.04311, 2018.
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.