Fast Beamforming Design via Deep Learning

Hao Huang, *Student Member, IEEE*, Yang Peng, *Student Member, IEEE*, Jie Yang, *Member, IEEE*, Wenchao Xia, *Student Member, IEEE*, and Guan Gui, *Senior Member, IEEE*

***Abstract*—Beamforming is considered as one of the most important techniques for designing advanced multiple-input and multiple-output (MIMO) systems. Among existing design criterions, sum rate maximization (SRM) under a total power constraint is a challenge due to its nonconvexity. Existing techniques for the SRM problem only obtain local optimal solutions but require huge amount of computation due to their complex matrix operations and iterations. Unlike these conventional methods, we propose a deep learning based fast beamforming design method without complex operations and iterations. Specifically, we first derive a heuristic solution structure of the downlink beamforming through the virtual equivalent uplink channel based on optimum MMSE receiver which separates the problem into power allocation and virtual uplink beamforming (VUB) design. Next, beamforming prediction network (BPNet) is designed to perform the joint optimization of power allocation and VUB design. Moreover, the BPNet is trained offline using two-step training strategy. Simulation results demonstrate that our proposed method is fast while obtains the comparable performance to the state-of-the-art method.**

***Index Terms*—Beamforming design, sum rate maximization, deep learning, beamforming prediction network.**

I. Introduction

Higher transmission rates and higher spectrum efficiency are the inevitable trend in future wireless communications. Multiple-Input and Multiple-Output (MIMO) technology has made great success to improve system capacity and spectrum efficiency [1], [2]. As one of the effective transceiver methods for MIMO systems, beamforming technologies have been developed in various ways in the past decades, i.e., interference balancing problem, power minimization problem, and sum rate maximization (SRM) problem [3]–[6]. Since the first two problems are convex, it is easy to get their optimal solutions. However, the SRM is non-convex, many existing methods are required complicate iterative operations involving matrix decomposition and inversion to approach the optimal solution.

This paper focuses on the optimization problem of SRM under a total power constraint. Due to the non-convexity, existing methods suffer from high computational burden. In order to solve this problem, zero-forcing (ZF) method is proposed using channel diagonalized [7]. However, ZF method has poor performance when the noise power is large and thus ZF method is not suitable for the SRM problem.

Moreover, channel duality between uplink and downlink has been proven to be useful for optimum transceiver design. The channel duality can share the same mean square error (MSE) region [4] where the SRM problem is solved by iterations between uplink and downlink with joint optimization of power allocation and transceiver beamforming. However, the MSE-duality method is often with high complexity resulting from its inner iterations. To improve this, S. S. Christensen *et al.* proposed an low complexity weighted minimization mean square error (WMMSE) which is based on mutual information and MMSE [6] getting rid of the inner iterations. Although WMMSE can achieve the performance of MSE-duality [4], the computational complexity of WMMSE becomes higher when meeting the high-dimensional problem due to its iteration operation [8], [9]. The complexity of these methods is mainly caused by matrix inversion and decomposition in each iteration. Therefore, reducing the number of iterations and avoiding the complicated matrix inversion and decomposition are considered as the possible solutions to meet the real-time requirement.

Deep learning (DL), also known as the deep neural network technology, that consists of simple linear and nonlinear operations is an effective mean to achieve this goal [10], [11]. Its unique network structure can approximate any functions under specific conditions. Therefore, using DL to solve optimization problems such as power allocation [12] and beamforming design can get rid of complexity caused by excessive iterations to realize real-time calculation. From real-time routing [13], [14] to real-time resource allocation [8], [15]–[17], DL has achieved great success in improving computational efficiency and performance in both regression [18]–[22] and classification tasks [23]. Hence, it is a natural way to adopt DL for fast beamforming design and the DL based beamforming design method has been studied in mmWave systems [24], [25] and MISO systems [26], [27]. Inspired by these works, we first virtualize an equivalent uplink transmission model and show the equivalent part of the uplink and downlink. Next, the solution structure of MIMO downlink beamforming is derived which separates the downlink beamforming design into power allocation and virtual uplink beamforming (VUB) design. Then, we propose the beamforming prediction network (BPNet) for power allocation and uplink beamforming prediction. After that, a two-step training strategy based on supervised and unsupervised learning is used to train the BPNet.

The reminder of the rest paper is organized as follows. Section II introduces the downlink transmission model and problem formulation. In Section III, we propose fast beamforming design method based on DL. Simulation results are provided to confirm the proposed method in IV and our work is concluded in Section V.

Notation: The $\overline{\mathbf{M}}$ represents the column normalized matrix of $\mathbf{M}$. $\mathbf{M}^{\mathrm{T}}/\mathbf{M}^{\mathrm{H}}$ denotes transpose/conjugate transpose of a matrix $\mathbf{M}$. $\mathbf{I}_K$ denotes $K \times K$ identity matrix. $\mathbb{E}[\cdot]$ is statistical expectation.

II. System Model and Problem Formulation

A. System Model

We consider the downlink transmission scenario in a typical multiuser MIMO system where a base station (BS) equipped with N

Manuscript received July 18, 2019; revised September 25, 2019; accepted October 16, 2019. Date of publication October 23, 2019; date of current version January 15, 2020. This work was supported in part by the Project Funded by the National Science and Technology Major Project of the Ministry of Science and Technology of China under Grant TC190A3WZ-2, in part by the Jiangsu Specially Appointed Professor under Grant RK002STP16001, Innovation and Entrepreneurship of Jiangsu High-level Talent under Grant CZ0010617002, in part by Summit of the Six Top Talents Program of Jiangsu under Grant XYDXX-010, in part by 1311 Talent Plan of Nanjing University of Posts and Telecommunications. The review of this article was coordinated by Dr. L. Zhao. *(Corresponding author: Guan Gui.)*

The authors are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: 1017010502@njupt.edu.cn; 1019172207@njupt.edu.cn; jyang@njupt.edu.cn; 2015010203@njupt.edu.cn; guiguan@njupt.edu.cn).

Digital Object Identifier 10.1109/TVT.2019.2949122

antennas serves K decentralized users, each with Q receive antennas. The channel between user k and BS is denoted as $\mathbf{H}_k = d_k\tilde{\mathbf{H}}_k$ where d_k and $\tilde{\mathbf{H}}_k \in \mathbb{C}^{[N\times Q]}$ are the large-scale fading factor and the small-scale fading, respectively. The overall channel matrix can be denoted as $\mathbf{H}_{[N\times QK]} = [\mathbf{H}_1, \ldots, \mathbf{H}_k]$. Assume that the transmit symbols in $\mathbf{x} = [\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_K^{\mathrm{T}}]^{\mathrm{T}} = [x_1, \ldots, x_M]^{\mathrm{T}}$ are of independent unity-power, i.e., $E\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \mathbf{I}$, where $\mathbf{x}_k \in \mathbb{C}^{[Q\times 1]}$ is the data vector to be transmitted to the kth user. The total number of the transmit data streams is $M = KQ$ and each stream is assumed to be independent. The zero-mean complex white Gaussian noise is denoted by $\mathbf{n} = [\mathbf{n}_1^{\mathrm{T}}, \ldots, \mathbf{n}_K^{\mathrm{T}}] \sim \mathcal{CN}(0, \sigma_n^2\mathbf{I})$, which is independent of the data streams.

For convenience, we separate the transmit and receive beamforming into unity-norm columns matrices and diagonal matrices. Specifically, the transmit beamforming is $\mathbf{W} = \overline{\mathbf{W}}\mathbf{P}^{1/2}$ and the receive beamforming for user k is $\mathbf{B}_k^{\mathrm{H}} = \mathbf{P}_k^{-1/2}\boldsymbol{\alpha}_k\overline{\mathbf{B}}_k$, where the diagonal matrix $\mathbf{P} = \mathrm{blkdiag}\{\mathbf{P}_1, \ldots, \mathbf{P}_{\mathrm{K}}\}$ consists of the transmit power of all the users. The matrices $\overline{\mathbf{W}}_{[P\times QK]} = [\overline{\mathbf{W}}_1, \ldots, \overline{\mathbf{W}}_K]$ and $\overline{\mathbf{B}}_{[QK\times QK]} = \mathrm{blkdiag}\{\overline{\mathbf{B}}_1, \ldots, \overline{\mathbf{B}}_{\mathrm{K}}\}$ are with normalized columns, i.e., $\|\overline{\mathbf{w}}_m\|_2 = 1$ and $\|\overline{\mathbf{b}}_m\|_2 = 1$. Moreover, the matrix $\boldsymbol{\alpha} = \mathrm{blkdiag}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{\mathrm{K}}\}$ is a diagonal weighted matrix with positive entries. The system model is given by:

$$\widehat{x}_m = \frac{\alpha_m}{\sqrt{p_m}}\overline{\mathbf{b}}_m^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\sum_{i=1}^{M}\overline{\mathbf{w}}_i\sqrt{p_i}x_i + \frac{\alpha_m}{\sqrt{p_m}}\overline{\mathbf{b}}_m^{\mathrm{H}}\mathbf{n}. \tag{1}$$

In this work we assume that the perfect channel state information (CSI) is available at the transmitter and the channel matrices are constant in a transmission duration.

B. Problem Formulation

Our main objective is to design the normalized linear tranceiver filters $\overline{\mathbf{W}}_1, \ldots, \overline{\mathbf{W}}_k$ and $\overline{\mathbf{B}}_1, \ldots, \overline{\mathbf{B}}_k$ as well as the transmit power vector $\mathbf{p} = \mathrm{diag}\{\mathbf{P}\}$ to maximize the sum rate of all users. This utility maximization problem can be formulated as:

$$\begin{aligned} &\text{maximize} \sum_{k=1}^{K} R_k \quad \text{s. t. } \|\mathbf{p}\|_1 = P_{\max}, \\ &\qquad\qquad \|\overline{\mathbf{w}}_m\|_2 = 1, \quad \|\overline{\mathbf{b}}_m\|_2 = 1,\ \forall m, \end{aligned} \tag{2}$$

where $P_{\max}$ is the total power constraint. The achievable rate of each user can be expressed as:

$$R_k = \log\det\left(\mathbf{I}_k + \mathbf{W}_k^{\mathrm{H}}\mathbf{H}_k\mathbf{J}_{\tilde{v}_k\tilde{v}_k}^{-1}\mathbf{H}_k^{\mathrm{H}}\mathbf{W}_k\right), \quad \forall k, \tag{3}$$

where $\mathbf{J}_{\tilde{v}_k\tilde{v}_k}$ represents the effective noise matrix of user k:

$$\mathbf{J}_{\tilde{v}_k\tilde{v}_k} = \mathbf{I}_k + \sum_{i=1, i\neq k}^{K}\mathbf{H}_k^{\mathrm{H}}\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\mathbf{H}_k, \quad \forall k. \tag{4}$$

III. Our Proposed Method

A. Solution Structure

The proposed method is inspired from the perspective of the equivalent uplink. To this end, we first virtualize an equivalent uplink channel by switching the normalized transmit and receive beamforming as shown in Fig. 1. The transmit beamforming of user k is $\mathbf{B}_k = \overline{\mathbf{B}}_k\mathbf{U}_k^{1/2}$ and the receive beamforming of BS is $\mathbf{W}^{\mathrm{H}} = \mathbf{U}^{-1/2}\boldsymbol{\alpha}\overline{\mathbf{W}}^{\mathrm{H}}$. The quantities $\overline{\mathbf{W}}$ and $\overline{\mathbf{B}}$ in the virtual uplink model are the same as in the downlink model. Therefore, the effective channel and transceiver in the virtual uplink transmission are the transpose of the role in the downlink. The power allocations are considered to be different in the up/downlink. Additionally, we use the diagonal matrix to denote $\mathbf{U} = \mathrm{blkdiag}\{\mathbf{U}_1, \ldots, \mathbf{U}_{\mathrm{K}}\}$ and its vector form is $\mathbf{u} = \mathrm{diag}\{\mathbf{U}\}$. Moreover, we assume that both links are fulfilled with the same total power constraint, i.e., $\|\mathbf{u}\|_1 = \|\mathbf{p}\|_1 = P_{\max}$.

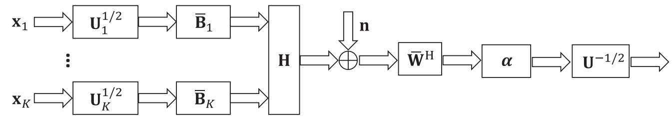

Fig. 1. Virtual equivalent uplink transmission model.

With the help of virtual uplink channel model, we can analyze the downlink transmission model from the perspective of equivalent virtual uplink under the optimum MMSE receiver with nomalized noise power spectral density [28]. The normalized downlink transmit beamforming which is equal to the normalized virtual uplink receive beamforming can be expressed as the function of VUB and virtual uplink power:

$$\overline{\mathbf{W}}_k\boldsymbol{\alpha}_k\mathbf{U}_k^{-1/2} = \left(\sum_{i=1}^{K}\mathbf{H}_i\overline{\mathbf{B}}_i^{\mathrm{H}}\mathbf{U}_i\overline{\mathbf{B}}_i\mathbf{H}_i^{\mathrm{H}} + \mathbf{I}\right)^{-1}\mathbf{H}_k\overline{\mathbf{B}}_k^{\mathrm{H}}\mathbf{U}_k^{1/2}, \quad \text{for } k = 1, \ldots, K \tag{5}$$

where $\boldsymbol{\alpha}_k$ is the diagonal matrix with positive entries and we denote $\mathbf{D}_k = (\sum_{i=1}^{K}\mathbf{H}_i\overline{\mathbf{B}}_i^{\mathrm{H}}\mathbf{U}_i\overline{\mathbf{B}}_i\mathbf{H}_i^{\mathrm{H}} + \mathbf{I})^{-1}\mathbf{H}_k\overline{\mathbf{B}}_k^{\mathrm{H}}$. To this end, we can obtain the 'beamforming direction' $\overline{\mathbf{W}}_k = \overline{\mathbf{D}}_k$ through normalizing each column of $\mathbf{D}_k$. Finally, the downlink beamforming can be denoted as:

$$\begin{aligned} \mathbf{W}_k &= \overline{\mathbf{W}}_k\mathbf{P}_k^{1/2} = \overline{\mathbf{D}}_k\mathbf{P}_k^{1/2} \\ &= \underbrace{\overline{\left(\sum_{i=1}^{K}\mathbf{H}_i\overline{\mathbf{B}}_i^{\mathrm{H}}\mathbf{U}_i\overline{\mathbf{B}}_i\mathbf{H}_i^{\mathrm{H}} + \mathbf{I}\right)^{-1}\mathbf{H}_k\overline{\mathbf{B}}_k^{\mathrm{H}}}}_{=\overline{\mathbf{D}}_k=\text{beamforming direction}}\ \underbrace{\mathbf{P}_k^{1/2}}_{\text{beamforming power}}, \\ &\text{for } k = 1, \ldots, K \end{aligned} \tag{6}$$

Therefore, the downlink beamforming solution structure can be simply expressed as the function of downlink transmit power $\mathbf{P}_k$, virtual uplink power $\mathbf{U}_k$, and normalized VUB $\overline{\mathbf{B}}_k$. In other words, the beamforming optimization problem is decomposed into three subproblems: downlink power allocation, virtual uplink power allocation, and virtual uplink beamforming. Therefore, the variables in the solution structure that are needed to be predicted for downlink power allocation and virtual uplink power allocation are both QK and for virtual uplink beamforming are $2Q^2$ (take the imaginary part as real number). However, the number of variables that needed to be optimized in the original problem are $2QNK$. Benefiting from this solution structure, the number of variables to be predicted is reduced from $2QNK$ to $2Q(K+Q)$. Under the fact that NK is larger than $K+Q$ in most MIMO systems, the method based on the solution structure has less variables needed to be optimized compared with the prediction of downlink beamforming directly.

B. BPNet

The BPNet designed in this work is based on the CNN architecture and integrated learning method, as shown in Fig. 2. The BPNet consists of two subnets: Power Prediction Network (PowerNet) and Uplink Beamforming Prediction Network (UBNet), which respectively includes the input layer, convolution (CONV) layers, batch normalization (BN) layers, flatten layers, full connected (FC) layers and output layers. Among them, CONV layers and FC layers are in the same structures of

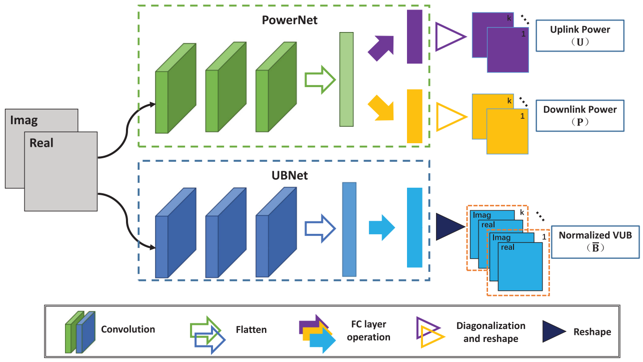

Fig. 2. BPNet structure.

the hidden layers as commonly used in CNN and DNN, respectively. In each CONV layer and FC layer, the activation function is ‘elu’. Furthermore, a BN layer is used to prevent over-fitting and speed up training after each CONV layer. The flatten layers are used to reshape the output of the CONV layers into the vector as the input of the FC layers. Finally, the output layer is divided into three parts including virtual uplink power vector, downlink power vector, and the VUB. The activations of two parts in the PowerNet output layer each adopts ‘softmax’ functions which is in order to ensure the sum of the output results to be ‘1’, meanwhile, the activation function of the output layer of UBNet is ‘tanh’.

C. Training the BPNet

A two-step training based on supervised pre-training and unsupervised re-training method is proposed to train the BPNet with integrated structure. First, randomly generate channel samples and implement the WMMSE algorithm to obtain uplink/downlink power and virtual uplink beamforming. Then, based on the MSE criterion, the pre-training is implemented with the loss function to achieve the results that are close to the WMMSE algorithm, i.e.,

$$Loss = \frac{1}{2SK}\sum_{s=1}^{S}\sum_{k=1}^{k=K}\Big(\|\widetilde{\boldsymbol{p}}_k^{(s)} - \hat{\boldsymbol{p}}_k^{(s)}\|_2^2 + \|\widetilde{\boldsymbol{u}}_k^{(s)} - \hat{\boldsymbol{u}}_k^{(s)}\|_2^2 + \|\widetilde{\mathbf{B}}_k^{(s)} - \hat{\mathbf{B}}_k^{(s)}\|_F^2\Big), \tag{7}$$

where S is the number of the samples, and $\widetilde{\boldsymbol{p}}_k^{(s)}$ and $\widetilde{\boldsymbol{u}}_k^{(s)}$ are the power vectors by running the WMMSE algorithm, $\hat{\boldsymbol{p}}_k^{(s)}$ and $\hat{\boldsymbol{u}}_k^{(s)}$ are the predicted power vectors of the BPNet. In order to further improve the performance of learning, we perform unsupervised learning on the basis of supervised learning for re-training step, i.e.,

$$Loss = -\frac{1}{2SK}\sum_{s=1}^{S}\sum_{k=1}^{k=K} R_k^{(s)}. \tag{8}$$

When performing the re-training step, it is necessary to project the predicted $\widehat{\mathbf{B}}_k^{(s)}$ and $\widehat{\mathbf{D}}_k^{(s)}$ into the row/column normalized domain after each forward propagation. It is worth pointing out that when the transmit power is high, it is not necessary to perform the pre-training step. The unsupervised learning can significantly improve the performance compared with that of only supervised learning.

IV. Numerical Results and Analysis

The simulation environment is based on Python 3.6.5 with TensorFlow 1.1.0 and Keras 2.2.2 on a computer with 8 Intel i7-7820x CPU Cores, one NVIDIA GTX 1080Ti GPU, and 32 GB of memory. The WMMSE algorithm and the DNN are programmed using MATLAB and Python, respectively. In the simulation experiment, the generation of data and the results and analysis are as follows.

A. Data Generation and Setup

We consider a downlink transmission scenario where the BS equipped with N antennas and its coverage is a disc with a radius of 500 m. There are K users each with Q receive antennas and these users are distributed uniformly within the coverage of the BS. Moreover, we set that none of the users is closer to the BS than 100 m. The pathloss between the user and the BS is set as $128.1 + 37.6\log_{10}(\omega)[\mathrm{dB}]$ where ω is the distance in kilometer [29]. The noise power spectral density is $\sigma^2 = -174$ dBm/Hz and the total system banwith is 20 MHz. Besides, we assume perfect CSI is available at the BS.

Before fed into the neural network, the input data transformations of the channel $\mathbf{H}_{[P\times QK]} = [\mathbf{H}_1, \ldots, \mathbf{H}_k]$ can be separated into the in-phase component $\Re(\mathbf{H})$ and the quadrature component $\Im(\mathbf{H})$ which is referred to as **I/Q** transformation. The $\Re(\mathbf{H})$ and $\Im(\mathbf{H})$ are the real and imaginary part of each element in $\mathbf{H}$, which is also regarded as the two channels of the input data. Meanwhile, the labels for the supervised learning are obtained through the WMMSE algorithm [6] which iterates up to 10 times with randomly initialization. In order to make the WMMSE algorithm better adapt to the solution structure, we let the MMSE-matrices be kept diagonal in the process of the WMMSE implementation.

There are 100000 training samples and 2000 testing samples prepared in the simulation. The PowerNet and the UBNet share the same input layer but the other layers are separated. The PowerNet includes 5 CONV layers, 1 FC layer and 2 output layers, while, the UBNet has 3 CONV layers, 1 FC layer and 1 output layers. Besides, each CONV layer has 8 kernels of size 3×3 and each FC layer has 64 neurons. Moreover, Adam optimizer is used with the proposed loss function to update the weights of the BPNet.

B. Results and Analysis

In this subsection, we evaluate the performance of the BPNet for the SRM problem. Our experimental results under three different variables prove the effectiveness of the BPNet. All the simulation results show that our proposed BPNet with two step learning stratege can lead to the performance comparable to the WMMSE algorithm, meanwhile, supervised learning which directly learn the power and virtual beamforming is the worst.

We first study the performance versus number of users each with $Q = 2$ receive antennas under different transmit power in Fig. 3. For $K = 1$, i.e., the single user case, the problem becomes convex and therefore it is easier for BPNet to achieve the near-optimal than the WMMSE using diagonal MSE-matrices suboptimal structure. For $K = \{2, 6\}$, in the case of low transmit power, the proposed BPNet solution is almost equivalent to the WMMSE solution performance. When the transmit power is high, the performance of the BPNet has an advantage than the WMMSE solution.

Then, we further analyze the sum rate when each user equipped with different receive antennas in Fig. 4. The number of transmit antennas is fixed as 8 and the transmit power is $p = 40$ dBm under two user case. For $Q = 1$, which is the MISO scene, the VUB matrix for each user becomes the constant 1. In this case, the sum rate only depends on the allocation of the downlink and virtual uplink power. Moreover, power allocation is not important when the number of transmit antenna is much more than the total number of user receive antenna ($N >>$

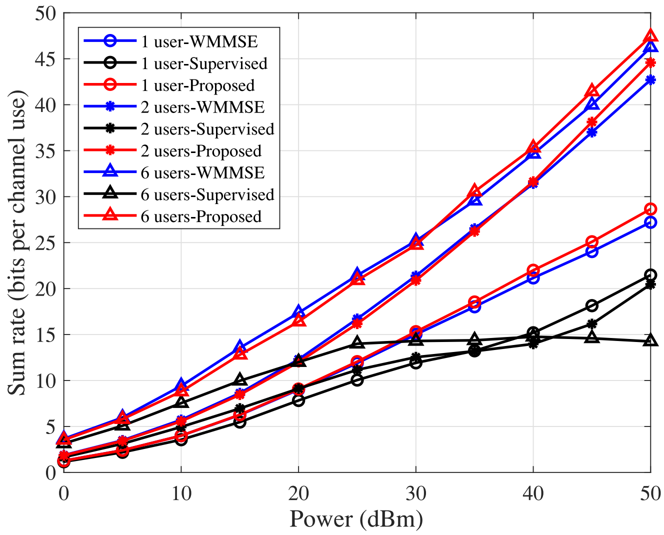

Fig. 3. Sum rate performance averaged over 2000 random channels with $N = 4$ transmit antennas and $K \in \{1, 2, 6\}$ users with dual receive antennas ($Q = 2$) under different transmit power.

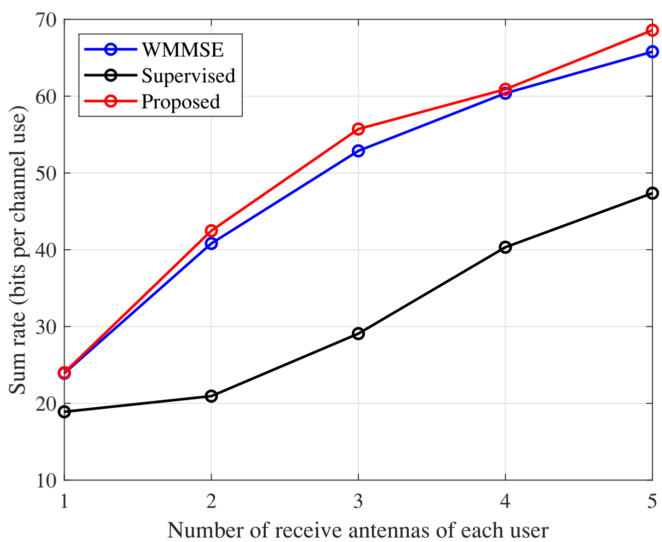

Fig. 4. Sum rate performance versus receive antennas of each $K = 2$ users on 2000 samples with transmit number of antennas is 8 and the transmit power is 40 dBm.

KQ), so the performance of the proposed BPNet is comparable to the performance of the WMMSE. As the number of receive antennas increases, the proposed BPNet is better than the WMMSE due to the limitation of iterations, but a slight decrease in performance gain occurs when $N = KQ$.

Finally, the impact on the number of transmit antenna to the performance is also considered in Fig. 5. When the $N \leq KQ$ (i.e. $N \leq 4$) which means that both the allocation of power and the design of the VUB matrix are necessary, the supervised based learning strategy is always poor and the proposed two-step learning works well compared to the WMMSE. When the $N > KQ$ (i.e. $N > 4$) which means that the meaning of power allocation is of little significance, the performance of the proposed method is also comparable to the WMMSE.

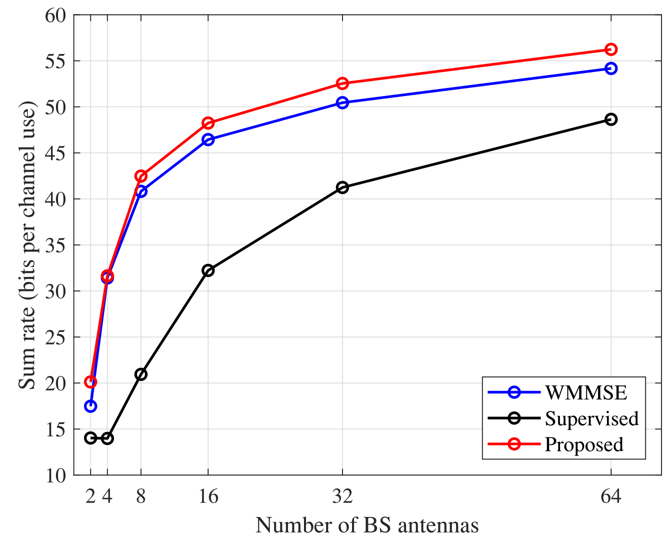

Fig. 5. Sum rate performance under different number of transmit antennas with $K = 2$ users each with dual receive antennas under transmit power 40 dBm.

V. Conclusion

This paper proposes a fast downlink beamforming design method to maximizing the sum rate under the total transmit power constraint for MIMO systems. Based on virtual equivalent uplink channel, the proposed method converts the downlink beamforming to power and virtual uplink beamforming solution structure form. An integrated neural network BPNet is proposed to predict power allocation and virtual beamforming design simultaneously. Simulation results confirm that the proposed fast beamforming design method can achieve sum rate comparable to the WMMSE method.

References

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Sep. 1999.

[2] P. He and L. Zhao, "Optimal power allocation for maximum throughput of general MU-MIMO multiple access channels with mixed constraints," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1042–1054, Mar. 2016.

[3] E. Bjrnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[4] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[5] P. He, L. Zhao, and Z. Liao, "Improved and extended sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," in *Proc. Int. Conf. Wireless Algorithms, Syst. Appl.*, 2011, pp. 211–222.

[6] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[7] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[8] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[9] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, Jan. 2019.

[10] H. Huang *et al.*, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, to be published, doi: 10.1109/MWC.2019.1900027.

[11] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, Nov. 2017.

[12] P. He, S. Zhang, L. Zhao, and X. Shen, "Energy-efficient power allocation with individual and sum power constraints," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5353–5366, Aug. 2018.

[13] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

[14] F. Tang *et al.*, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 154–160, Feb. 2018.

[15] Y. Zhou, Z. M. Fadlullah, B. Mao, and N. Kato, "A deep-learning-based radio resource assignment technique for 5G ultra dense networks," *IEEE Netw.*, vol. 32, no. 6, pp. 28–34, Nov. 2018.

[16] W. Lee, M. Kim, and D. Cho, "Deep power control: Transmit power control scheme based on convolutional neural network," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1276–1279, Jun. 2018.

[17] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2019.

[18] M. Liu, T. Song, G. Gui, J. Hu, and H. Sari, "Deep cognitive perspective: Resource allocation for NOMA based heterogeneous IoT with imperfect SIC," *IEEE Int. Things J.*, vol. 6, no. 2, pp. 2885–2894, Apr. 2019.

[19] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[20] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surv. Tuts.*, vol. 19, no. 4, pp. 2432–2455, Apr. 2017.

[21] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[22] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.

[23] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[24] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[25] T. Lin and Y. Zhu, "Beamforming design for large-scale antenna arrays using deep learning," 2019, *arXiv:1904.03657*.

[26] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, "A deep learning framework for optimization of MISO downlink beamforming," 2019, *arXiv:1901.00354*.

[27] A. Balatsoukas-Stimming, O. Castañeda, S. Jacobsson, G. Durisi, and C. Studer, "Neural-network optimized 1-bit precoding for massive mu-mimo," 2019, *arXiv:1903.03718*.

[28] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Hoboken, NJ, USA: Wiley, 2009.

[29] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.