Deep Learning Based Beamforming for MISO Systems with Dirty-Paper Coding

Xingliang Lou[1], Wenchao Xia[1], Zhao Haitao[1], Wanli Wen[2], Xiaohui Li[3], and Bin Wang[4]

[1]Nanjing University of Posts and Telecommunications
[2]Chongqing University
[3]Taiyuan University of Technology
[4]Hangzhou Hikvision Digital Technology Co Ltd

October 20, 2022

Abstract

Beamforming technique can effectively improve the spectrum utilization of multi-antenna systems, while the dirty-paper coding (DPC) technique can reduce inter-user interference. In this letter, we aim to maximize the weighted sum-rate under power constraint in a multiple-input-single-output (MISO) system with the DPC. However, the existing methods of beamforming optimization mainly rely on customized iterative algorithms, which have high computational complexity. To address this issue, by utilizing the deep learning technique and the uplink-downlink duality, and carefully exploring the optimal solution structure, we devise a beamforming neural network (BFNNet), which includes a deep neural network module and a signal processing module. Besides, we use the modulus of the channel coefficients as the input of deep neural network, which reduces the input size. Simulation results show that a well-trained BFNNet can achieve near-optimal solutions, while significantly reducing computational complexity

Deep Learning Based Beamforming for MISO Systems with Dirty-Paper Coding

Xingliang Lou, Wenchao Xia, Wanli Wen, Haitao Zhao✉, Xiaohui Li and Bin Wang✉

Beamforming technique can effectively improve the spectrum utilization of multi-antenna systems, while the dirty-paper coding (DPC) technique can reduce inter-user interference. In this letter, we aim to maximize the weighted sum-rate under power constraint in a multiple-input-single-output (MISO) system with the DPC. However, the existing methods of beamforming optimization mainly rely on customized iterative algorithms, which have high computational complexity. To address this issue, by utilizing the deep learning technique and the uplink-downlink duality, and carefully exploring the optimal solution structure, we devise a beamforming neural network (BFNNet), which includes a deep neural network module and a signal processing module. Besides, we use the modulus of the channel coefficients as the input of deep neural network, which reduces the input size. Simulation results show that a well-trained BFNNet can achieve near-optimal solutions, while significantly reducing computational complexity.

Introduction: Beamforming technique can improve the spectrum efficiency of multi-antenna systems while the dirty-paper coding (DPC) technique [1] can reduce inter-user interference. Thus, beamforming strategies using the DPC technique are a potential way to maximize the weighted sum-rate under power constraint in a multi-antenna system. However, finding the optimal beamforming to maximize the weighted sum-rate is a non-convex problem. There have been some methods of beamforming design studied in existing literature. For example, the weighted minimum mean square error (WMMSE) algorithm was proposed in [2, 3]. Since the uplink-downlink duality was proved in [4], the downlink sum-rate maximization problem can be solved by considering the dual uplink problem. [5] has found the achievable rate of multi-antenna downlink, and [6, 7] have established the conversion relationship between the uplink and downlink transmission. [8] used iterative water-filling (IWF) algorithm to find the optimal solution of the uplink transmission, then with which the optimal solution of downlink transmission was inferred. Nevertheless, these algorithms are iterative algorithms in general, which leads to high computational complexity, especially when the problem size is huge. The delay caused by the iterative process also makes the beamforming scheme unable to adapt to high-reliability and low-latency scenarios in 5G/B5G wireless networks.

Deep learning (DL) is regarded as a promising technique which can balance system delay and performance. This is because DL trains the neural network model offline, which includes the most computational complexity, and then predicts the beamforming matrix online with some linear and nonlinear calculations. [9] used the DL technique to predict the pilot sequence in a quantized codebook. Different from finding the optimal solution in a limited space, [9, 10, 11] directly used deep neural network to predict the beamforming vector, which may cause significantly high complexity as the numbers of transmitting antennas and users increase. In [12], the authors proposed a DL framework to solve three kinds of beamforming optimization problems. This framework exploits the uplink-downlink duality and the existing optimal solution structure to reduce the prediction complexity. Using the power budget as side information, [13] investigated the influence of power constraint on beamforming optimization. However, these works mentioned above did not consider the DPC technique which can reduce inter-user interference and enables better performance.

In this letter, we consider a MISO system with the DPC technique and formulate a sum-rate maximization problem under a total power constraint. By utilizing the DL technique and the uplink-downlink duality, we devise a beamforming neural network (BFNNet), which includes a deep neural network module that predicts key feature vectors and a signal processing module that uses expert knowledge to recover beamforming solutions. Note that the signal processing module is designed based on the optimal solution structure of the sum-rate maximization problem. Finally, simulation results show that a well-trained BFNNet can find near-optimal solutions with a significantly lower computational complexity.

System Model: We consider a downlink transmission scenario, where there is a BS with M antennas and K single-antenna users. The channel between user i and BS is expressed as $\mathbf{h}_i \in \mathbb{C}^{M\times 1}$. Then, the received signal at user i is given by

$$y_i = \mathbf{h}_i^H \sum_{k=1}^{K} \mathbf{u}_k x_k + n_i, \tag{1}$$

where $\mathbf{u}_i$ represents the beamforming vector for user i, $x_i \sim \mathcal{CN}(0,1)$ is the transmitted symbol from the BS to user i, and $n_i \sim \mathcal{CN}(0,\sigma^2)$ denotes the additive Gaussian white noise (AWGN) with zero mean and variance σ^2.

Assume a pre-coding order $K...1$ when using the DPC. Because decoding/encoding is performed in sequence, the interference of user k ($k > i$) has no effect on the demodulated received SINR of user i. Thus, the received SINR at user i can be expressed as

$$\mathrm{SINR}_i^{\mathrm{DL}} = \frac{|\mathbf{h}_i^H \mathbf{u}_i|^2}{\sum_{k=1}^{i-1} |\mathbf{h}_i^H \mathbf{u}_k|^2 + \sigma^2}. \tag{2}$$

Define $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K]$, then the downlink sum-rate maximization problem under power constraint is formulated as

$$\begin{aligned} \mathbf{P1:} \max_{\mathbf{U}} & \sum_{i=1}^{K} \log_2(1 + \mathrm{SINR}_i^{\mathrm{DL}}) \\ \text{s.t.} & \sum_{i=1}^{K} ||\mathbf{u}_i||^2 \le P_m, \end{aligned} \tag{3}$$

where P_m is the power budget. Note that, **P1** is a challenging non-convex problem, which can be solved using the WMMSE algorithm or the uplink-downlink duality based algorithms [6, 7]. But these algorithms relying on iterative processes are difficult to meet the implementation requirements. Thus, we propose to solve it using a DL-based beamforming framework, which will be described in next Sections.

Expert Knowledge: Before giving the DL based beamforming framework, we first establish a concept of expert knowledge [4] for the purpose of reducing prediction complexity.

Lemma 1: The achievable uplink sum-rate is equal to the achievable downlink sum-rate, i.e.,

$$C_{sum}^{DL} = C_{sum}^{UL}, \tag{4}$$

where

$$\begin{aligned} C_{sum}^{DL} = \max_{\tilde{\mathbf{U}},\mathbf{p}} & \sum_{i=1}^{K} \log_2(1 + \mathrm{SINR}_i^{DL}) \\ \text{s.t.} & \ ||\mathbf{p}||_1 \le P_m, \\ & \ ||\tilde{\mathbf{u}}_i||_2 = 1, \forall i, \end{aligned} \tag{5}$$

and

$$\begin{aligned} C_{sum}^{UL} = \max_{\tilde{\mathbf{U}},\mathbf{q}} & \sum_{i=1}^{K} \log_2(1 + \mathrm{SINR}_i^{UL}) \\ \text{s.t.} & \ ||\mathbf{q}||_1 \le P_m, \\ & \ ||\tilde{\mathbf{u}}_i||_2 = 1, \forall i, \end{aligned} \tag{6}$$

with

$$\mathrm{SINR}_i^{DL} = \frac{p_i |\mathbf{h}_i^H \tilde{\mathbf{u}}_i|^2}{\sum_{k=1}^{i-1} p_k |\mathbf{h}_i^H \tilde{\mathbf{u}}_k|^2 + \sigma^2}, \tag{7}$$

and

$$\mathrm{SINR}_i^{UL} = \frac{q_i |\tilde{\mathbf{u}}_i^H \mathbf{h}_i|^2}{\sum_{k=i+1}^{K} q_k |\tilde{\mathbf{u}}_i^H \mathbf{h}_k|^2 + \sigma^2}, \tag{8}$$

in which $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \ldots, \tilde{\mathbf{u}}_K]$ is the normalized beamforming, $\mathbf{p} = [p_1, ..., p_K]^T$ and $\mathbf{q} = [q_1, ..., q_K]^T$ are downlink and uplink power allocation vectors, respectively.

Proof: The proof is similar to the proof of Theorem 2 in [4] and thus is omitted here. In addition, [4] also proves that the optimal normalized beamforming of the uplink is also optimal for the downlink. ■

Note that the problem in (5) equals to **P1** with $\mathbf{U}=\tilde{\mathbf{U}}\sqrt{\mathbf{P}}$, where $\mathbf{P}=\text{diag}(\mathbf{p})$. It is well known the uplink problem in (6) is easier to handle. Therefore, we can first obtain the optimal uplink power vector $\mathbf{q}^*$ and the normalized beamforming matrix $\tilde{\mathbf{U}}^*$ of the uplink problem in (6), then with which infer the optimal downlink power vector $\mathbf{p}^*$, finally obtain the optimal beamforming matrix $\mathbf{U}^*$ of problem **P1**.

To find solutions to (6), we first simplify (6) as

$$
\begin{aligned}
C_{sum}^{\text{UL}} &= \max_{\tilde{\mathbf{U}},||\mathbf{q}||\le P_m,} \sum_{i=1}^{K} \log_2(1+\text{SINR}_i^{\text{UL}}) \\
&= \max_{||\mathbf{q}||\le P_m} \log_2|\mathbf{I}+\frac{1}{\sigma^2}\sum_{k=1}^{K} q_k \mathbf{h}_k \mathbf{h}_k^H|, \qquad (9)
\end{aligned}
$$

where the second equation is obtained due to $||\tilde{\mathbf{u}}_i||_2=1,\forall i$. The problem in (9) can be solved using the IWF algorithm [8] until convergence. Knowing the optimal $\mathbf{q}^*$, the optimal beamforming vectors are given as the MMSE solutions [3], i.e.,

$$
\tilde{\mathbf{u}}_i^* = \frac{(\sigma^2\mathbf{I}+\sum_{k=i+1}^{K} q_k^*\mathbf{h}_k\mathbf{h}_k^H)^{-1}\mathbf{h}_i}{||(\sigma^2\mathbf{I}+\sum_{k=i+1}^{K} q_k^*\mathbf{h}_k\mathbf{h}_k^H)^{-1}\mathbf{h}_i||_2}. \qquad (10)
$$

Then, we can find the optimal power allocation vector $\mathbf{p}^*$ of the downlink problem in (5) according to the following lemma.

Lemma 2: Given the optimal transmit power vector $\mathbf{q}^*$ and beamforming matrix $\tilde{\mathbf{U}}^*$ of the uplink problem in (6)*, then we can obtain the optimal transmit power vector $\mathbf{p}^*$ of the downlink problem in* (5) *as*

$$
\begin{aligned}
p_1^* &= B_1^{-1/2}\overline{q_1^*}B_1^{-1/2}, \\
&\ldots \\
p_i^* &= B_i^{-1/2}\overline{A_i^{1/2}q_i^*A_i^{1/2}}B_i^{-1/2}, \qquad (11) \\
&\ldots \\
p_K^* &= \overline{A_K^{1/2}q_K^*A_K^{1/2}}.
\end{aligned}
$$

where $B_i=\sigma^2+\sum_{k=i+1}^{K} q_k^*\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_k\mathbf{h}_k^H\tilde{\mathbf{u}}_i^*$ *and* $A_i=\sigma^2+\sum_{k=1}^{i-1} p_k^*\mathbf{h}_i^H\tilde{\mathbf{u}}_k^*\tilde{\mathbf{u}}_k^{*H}\mathbf{h}_i$ *represent the interference experienced by user i in the uplink and the interference experienced by user i in the downlinks, respectively.*

Proof: The achievable rate of user i in the uplink is given by

$$
\text{R}_i^{\text{UL}} = \log_2(1+B_i^{-1}q_i^*\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_i\mathbf{h}_i^H\tilde{\mathbf{u}}_i^*), \qquad (12)
$$

Using matrix knowledge, we have the simplified formula as

$$
\begin{aligned}
&\text{R}_i^{\text{UL}} = \\
&\log_2(1+B_i^{-1/2}\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_iA_i^{-1/2}A_i^{1/2}q_i^*A_i^{1/2}A_i^{-1/2}\mathbf{h}_i^H\tilde{\mathbf{u}}_i^*B_i^{-1/2}), \\
&\hfill (13)
\end{aligned}
$$

Treating $B_i^{-1/2}\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_iA_i^{-1/2}$ as the effective channel of the system, we flip the channel and get

$$
\begin{aligned}
&\text{R}_i^{\text{UL}} = \\
&\log_2(1+A_i^{-1/2}\mathbf{h}_i^H\tilde{\mathbf{u}}_i^*B_i^{-1/2}\overline{A_i^{1/2}q_i^*A_i^{1/2}}B_i^{-1/2}\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_iA_i^{-1/2}). \\
&\hfill (14)
\end{aligned}
$$

Now, consider the achievable rate of user i in the downlink and we have

$$
\begin{aligned}
\text{R}_i^{\text{DL}} &= \log_2(1+A_i^{-1}\mathbf{h}_i^H\tilde{\mathbf{u}}_i^*p_i^*\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_i) \\
&= \log_2(1+A_i^{-1/2}\mathbf{h}_i^H\tilde{\mathbf{u}}_i^*p_i^*\tilde{\mathbf{u}}_i^{*H}\mathbf{h}_iA_i^{-1/2}), \qquad (15)
\end{aligned}
$$

By setting the downlink transmit power as in (11), then we have $\text{R}_i^{\text{UL}}=\text{R}_i^{\text{DL}}$. [6] has proved that $\sum_{i=1}^{K} q_i^*=\sum_{i=1}^{K} p_i^*$. ■

Note that p_i^* only depends on $p_1^*,...,p_{i-1}^*$, thus the transmit power can be calculated sequentially in ascending order.

BFNNet Structure: The proposed BFNNet for the sum-rate maximization problem is shown in Fig. 1, which includes a deep neural network module and a signal processing module.

The deep neural network module includes an input layer, multiple hidden layers, and an output layer. The channel coefficients

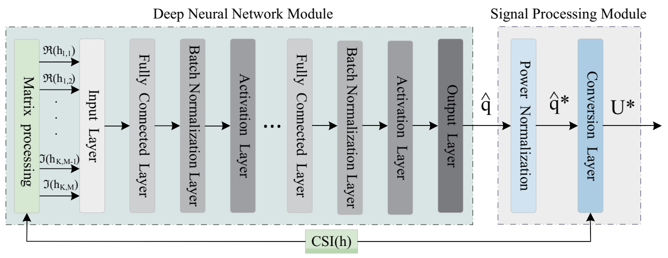

Fig. 1. BFNNet for the sum-rate maximization problem.

$\mathbf{H}=\{\mathbf{h}_1^T,\mathbf{h}_2^T,...,\mathbf{h}_K^T\}^T\in\mathbb{C}^{MK\times 1}$ is converted to a vector $[\Re(\mathbf{H}),\Im(\mathbf{H})]^T\in\mathbb{C}^{2\times MK}$ as the first layer input, where $\Re(\mathbf{H})$ and $\Im(\mathbf{H})$ contain the real and imaginary parts of each element in h, respectively. The fully connected layers are used as hidden layers to perform feature extraction from the input data.

The signal processing module is to restore the beamforming matrix based on the key features predicted by the output layer. Due to the existence of output prediction errors, the output of output layer is almost impossible to guarantee to meet the power constraints. Therefore, the result $\hat{\mathbf{q}}$ of the output layer is normalized in the power normalization layer as $\hat{\mathbf{q}}^*=\frac{P_m}{||\hat{\mathbf{q}}||_1}\hat{\mathbf{q}}$. Finally, we recovery the downlink beamforming matrix $\mathbf{U}$ using the conversion layer, which includes the following process:

1 Calculate $\tilde{\mathbf{u}}_i^*$ using (10).
2 Calculate $\mathbf{p}^*$ using (11).
3 Output the downlink beamforming vectors $\mathbf{u}_i^*=\sqrt{p_i^*}\tilde{\mathbf{u}}_i^*,\forall i$, as the final results.

Simulation Results: In this section, we use the scene in [12] to conduct some numerical simulations to evaluate the performance of the proposed BFNNet. In order to train the deep neural network module, we use the IWF algorithm [8] to generate 20000 training samples and 5000 testing samples, respectively. In our simulations, We use a network with three hidden layers, one input layer and one output layer for the deep neural network module. The first hidden layer contains 256 neurons and the second hidden layer contains 128 neurons and the third hidden layer contains 64 neurons. For comparison, several baseline solutions are introduced, including zero-forcing (ZF) beamforming, regularized ZF (RZF) beamforming [14], and the WMMSE algorithm with the RZF initialization [2]. Moreover, the DPC used for all the baseline solutions.

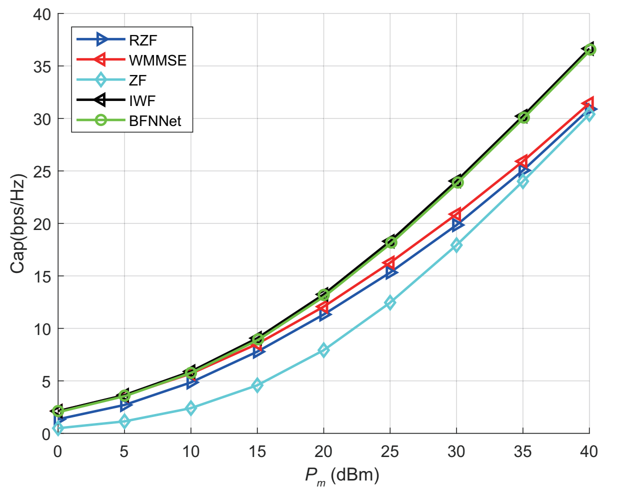

Fig. 2 *Sum-rate performance averaged over 5000 samples under* $\{K=4,M=4\}$.

Fig. 2 shows that with the increase of normalized transmission power, the sum-rate performance of all solutions are improved. We observe that the performance of the BFNNet solution is very close to the IWF algorithm, and better than the WMMSE algorithm which can find the locally optimal solution to the sum-rate maximization problem. This is because that the BFNNet is trained using the samples generated by the IWF algorithm which can achieve the optimal solution to problem **P1**.

Fig. 3 shows the sum-rate performance of the five beamforming solutions, where $P_m = 30$ dBm and $M = K$. In Fig. 3(a), as the number of transmitting antennas increases, the sum-rate performance of the five schemes increases at the same time and the BFNNet solution outperforms the other solutions except the IWF algorithm. In addition, as the number of transmit antennas increases, the performance gap becomes greater. In Fig. 3(b), the computational complexity, in terms of the execution time, of the BFNNet solution is higher than that of the ZF beamforming solution as well as the RZF beamforming solution. The reason is that ZF beamforming and RZF beamforming solutions do not require any iterative process, the BFNNet solution needs to perform neural network operations and conversion processes. The WMMSE algorithm as well as the IWF algorithm, consumes more time than the BFNNet solution due to its iterative process. The above observations validate that the BFNNet solution provides a good balance between system performance and computational complexity to the sum-rate maximization problems under a total power constraint.

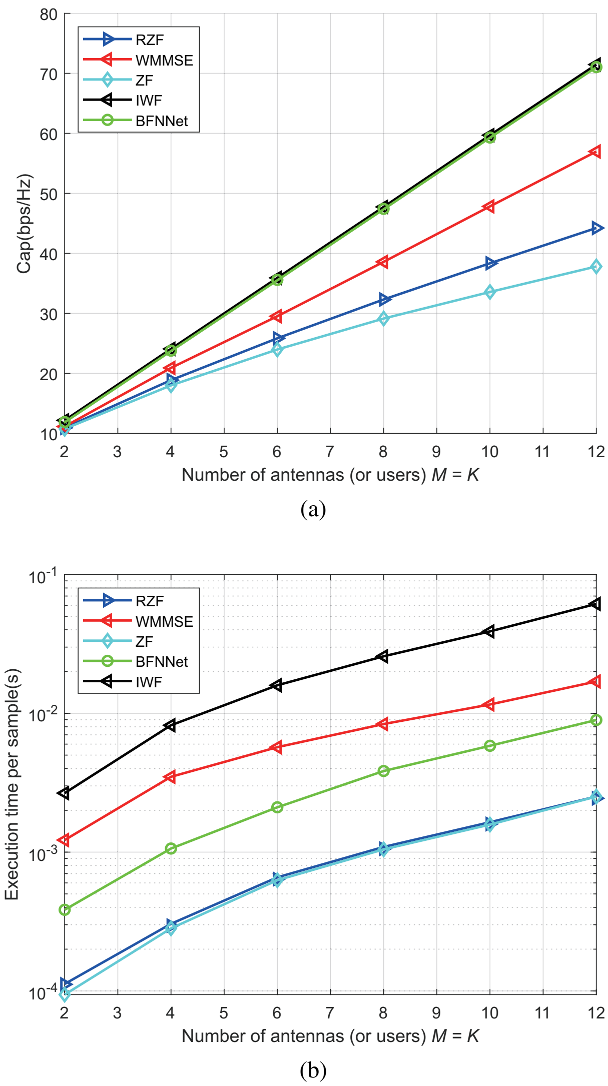

Fig. 3 *Comparison of five different beamforming solution: (a) sum-rate performance and (b) execution time of each sample averaged over 5000 samples under* $\{K = M, P_m = 30\,\text{dBm}\}$.

Conclusion: In this letter, we considered the MISO system with the DPC and formulated the problem of sum-rate maximization under a total power constraint. By utilizing the DL technique and the uplink-downlink duality, and carefully exploring the optimal solution structure, we devise the BFNNet to find the near-optimal solutions. The simulation results showed that, compared with the existing algorithms, the BFNNet has achieved a good balance between performance and computational complexity.

Acknowledgment: This work was supported by the National Key Research and Development Program (No. 2020YFB1806608); the Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu under Grant BK20212001; the National Natural Science Foundation of China (No. 92067201,61871446); the Natural Science Research Project of Jiangsu Higher Education Institutions (No. 21KJB510034, 21KJB510027); Future Network Scientific Research Fund Project (FNSRFP-2021-ZD-8, FNSRFP-2021-YB-31).

X. Lou, W. Xia, H. Zhao(*Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing, China*)

W. Wen (*The School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China*)

X. Li (*Taiyuan University of Technology, Taiyuan, China*)

B. Wang (*College of Electrical Engineering, Zhejiang University, Hangzhou, China*)

✉E-mail: zhaoht@njupt.edu.cn, wangbin2@hikvision.com

References

1 M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.
2 S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.
3 Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.
4 H. Boche and M. Schubert, “A general duality theory for uplink and downlink beamforming,” in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 1, 2002, pp. 87–91 vol.1.
5 G. Caire and S. Shamai, “On the achievable throughput of a multiantenna gaussian broadcast channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
6 S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.
7 P. Viswanath and D. Tse, “Sum capacity of the vector gaussian broadcast channel and uplink–downlink duality,” *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.
8 N. Jindal, S. Jafar, S. Vishwanath, and A. Goldsmith, “Sum power iterative water-filling for multi-antenna gaussian broadcast channels,” in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, vol. 2, 2002, pp. 1518–1522 vol.2.
9 A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, “Deep learning coordinated beamforming for highly-mobile millimeter wave systems,” *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.
10 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, 2019.
11 H. J. Kwon, J. H. Lee, and W. Choi, “Machine learning-based beamforming in two-user miso interference channels,” in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 496–499.
12 W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, “A deep learning framework for optimization of miso downlink beamforming,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1866–1880, 2020.
13 J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, “Deep learning methods for universal miso beamforming,” *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1894–1898, 2020.
14 D. H. Nguyen and T. Le-Ngoc, “Mmse precoding for multiuser miso downlink transmission with non-homogeneous user snr conditions,” *Eurasip Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 85, 2014.