Deep Learning-Based Downlink Channel Prediction for FDD Massive MIMO System

Yuwen Yang [iD], Feifei Gao [iD], *Senior Member, IEEE*, Geoffrey Ye Li, *Fellow, IEEE*, and Mengnan Jian

Abstract— In a frequency division duplexing (FDD) massive multiple-input multiple-output (MIMO) system, the acquisition of downlink channel state information (CSI) at base station (BS) is a very challenging task due to the overwhelming overheads required for downlink training and uplink feedback. In this letter, we reveal a deterministic uplink-to-downlink mapping function when the position-to-channel mapping is bijective. Motivated by the universal approximation theorem, we then propose a sparse complex-valued neural network (SCNet) to approximate the uplink-to-downlink mapping function. Different from general deep networks that operate in the real domain, the SCNet is constructed in the complex domain and is able to learn the complex-valued mapping function by off-line training. After training, the SCNet is used to directly predict the downlink CSI based on the estimated uplink CSI without the need of either downlink training or uplink feedback. Numerical results show that the SCNet achieves better performance than general deep networks in terms of prediction accuracy and exhibits remarkable robustness over complicated wireless channels, demonstrating its great potential for practical deployments.

Index Terms— FDD, massive MIMO, downlink CSI prediction, deep learning, complex-valued neural network.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) has been widely recognized as a promising technique in future wireless communication systems for its high spectrum and energy efficiency, high spatial resolution, and large

Y. Yang is with the State Key Lab of Intelligent Technologies and Systems, Beijing National Research Center for Information Science and Technology (BNRist), Department of Automation, Institute for Artificial Intelligence Tsinghua University (THUAI), Tsinghua University, Beijing 100084, China (e-mail: yyw18@mails.tsinghua.edu.cn).

F. Gao is with the State Key Lab of Intelligent Technologies and Systems, Beijing National Research Center for Information Science and Technology (BNRist), Department of Automation, Institute for Artificial Intelligence Tsinghua University (THUAI), Tsinghua University, Beijing 100084, China, and also with the Key Laboratory of Digital TV System of Guangdong Province and Shenzhen City, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China (e-mail: feifeigao@ieee.org).

G. Y. Li is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: liye@ece.gatech.edu).

M. Jian is with the Department of Algorithm, Wireless Product R&D Institute, ZTE Corporation, Beijing 100029, China (e-mail: jian.mengnan@zte.com.cn).

Digital Object Identifier 10.1109/LCOMM.2019.2934851

beamforming gains [1]. To embrace these benefits, accurate downlink channel state inforamtion (CSI) is usually required at both the base station (BS) (for beamforming, user scheduling, etc) and the user side (for signal detection). However, the acquisition of downlink CSI is a very challenging task for frequency division duplexing (FDD) massive MIMO systems due to the prohibitively high overheads associated with downlink training and uplink feedback.

In fact, there are two important observations that can help reduce the overheads. First, the wireless channels between BS and users only have a small angular spread (AS) as demonstrated in [2]–[4]. Due to the small AS and the large dimension of the channels, massive MIMO channels exhibit sparsity in the angular domain. Secondly, there exists angular reciprocity between the uplink and the downlink channels since the uplink and the downlink share the common physical paths [4]. Since the acquisition of the uplink CSI is convenient in massive MIMO systems, many studies have suggested to extract partial information of the downlink CSI from the uplink CSI, thereby reducing the downlink training overhead or to employ compressive sensing (CS) based algorithms to reduce the overhead of the uplink feedback [5], [6]. For example, in [5], the downlink channel covariance matrix (CCM) is first estimated from the uplink CCM and then the eigen-beamforming is used to reduce the overhead for the downlink training when AS is less than $5°$. In [6], the channels are first parameterized by distinct paths, each characterized by path delay, angle, and gain. Then, the frequency-independent parameters, i.e., path delays and angles, are extracted from the uplink CSI to help reduce the downlink training. Nevertheless, the method in [6] is applicable as long as the propagation paths are distinguishable and the path number is small. Besides, several CS-based channel feedback schemes for massive MIMO have been proposed to reduce the feedback overhead but are sensitive to the model errors and suffer from high complexity.

Due to its excellent performance and low complexity [7], deep learning has been introduced recently to the wireless physical layer and has achieved superior performance over various topics, such as channel estimation [8], detection [9], CSI feedback [10], etc. In [11], a convolutional neural network (CNN) is trained to predict the downlink CSI based on the CSI of multiple adjacent uplink subcarriers for single-antenna FDD systems. In [12], a fully-connected neural network (FNN) is trained for uplink/downlink channel calibration for massive MIMO systems. In this letter, we propose a sparse complex-valued neural network (SCNet) for the downlink CSI prediction in FDD massive MIMO systems. Due to the richer representational capacity offered by complex rep-

resentations, the SCNet can further improve the performance of channel prediction. Our contributions are summarized as follows.

1) Inspired by [14], we reveal a deterministic uplink-to-downlink mapping function for a given communication environment when the position-to-channel mapping is bijective. Then, we prove that the uplink-to-downlink mapping function can be approximated with an arbitrarily small error by a feedforward network.
2) We propose a SCNet for downlink CSI prediction in FDD massive MIMO systems, which is applicable to complex-valued function approximation with complex-valued representations. Moreover, sparse network structure is adopted to reduce the complexity and improve the robustness.
3) Experiment results demonstrate that SCNet outperforms the FNN of [12] in terms of prediction accuracy and exhibits remarkable robustness over the number of paths.

II. SYSTEM MODEL

Fig. 2 illustrates an FDD massive MIMO system, where the BS is equipped with $M \gg 1$ antennas in the form of uniform linear array (ULA)[1] and the user is equipped with a single antenna. Since the proposed approach works for different users separately, we only need to illustrate for a single user. The channel between the user and the BS is assumed to consist of P rays and can be expressed as[2] [13],

$$\boldsymbol{h}\left(f\right) = \sum_{p=1}^{P} \alpha_p e^{-j2\pi f \tau_p + j\phi_p} \boldsymbol{a}\left(\theta_p\right), \qquad (1)$$

where f is the carrier frequency, and α_p, ϕ_p, τ_p and θ_p are the attenuation, phase shift, delay, and direction of arrival (DOA) of the p-th path, respectively. Moreover, $\boldsymbol{a}\left(\theta_p\right)$ is the the array manifold vector defined as,

$$\boldsymbol{a}\left(\theta_p\right) = \left[1, e^{-j\chi \sin \theta_p}, \cdots, e^{-j\chi (M-1) \sin \theta_p}\right]^T, \qquad (2)$$

where $\chi = 2\pi df/c$, d is the antenna spacing, and c is the speed of light. According to [2]–[4], the incident AS with mean DOA θ seen by the BS is limited in a certain region, i.e., $\theta_p \in [\theta - \Delta\theta/2, \theta + \Delta\theta/2]$.

Note that α_p depends on (i) the distance between the user and the BS, denoted by D, (ii) the transmitter and receiver antenna gains, (iii) the carrier frequency, and (iv) the scattering environment. The phase ϕ_p depends on the scatterer materials and wave incident/impinging angles at the scatterers. The delay τ_p depends on the distance travelled by the signal along the p-path [4].

---

[1]We adopt the ULA model here for simpler illustration, however, the proposed approach does not restrict the specifical array shape, and therefore is applicable for an array of arbitrary geometry.

[2]Note that we have ignored spatial- and frequency- wideband effects [1] as most of the literature. In fact, an accurate mathematical model is unnecessary for the proposed approach since the network can be trained by data from practical systems.
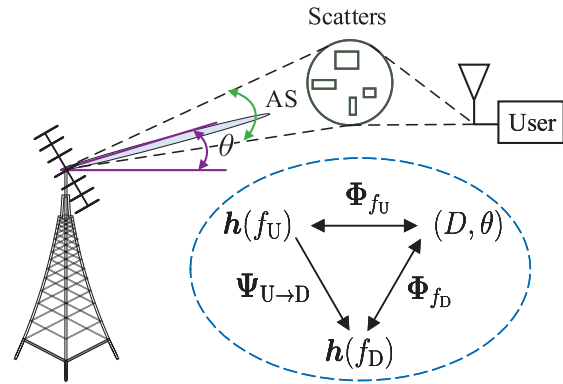

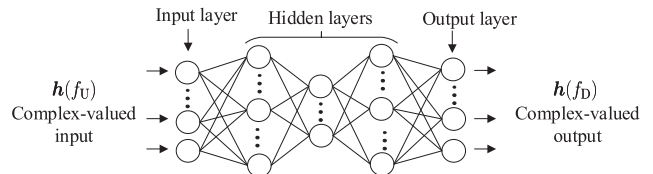Fig. 1. Downlink CSI prediction for FDD Massive MIMO systems.

Fig. 2. The SCNet architecture.

III. CHANNEL MAPPING FORMULATION

Denote $\boldsymbol{h}\left(f_U\right)$ and $\boldsymbol{h}\left(f_D\right)$ as the uplink and the downlink channels of the user with f_U and f_D being the uplink and the downlink frequencies, respectively. As indicated in Eq. (1), $\boldsymbol{h}\left(f_D\right)$ cannot be simply obtained from the $\boldsymbol{h}\left(f_U\right)$ for FDD systems. However, since the downlink and the uplink experience the same propagation environment with the common physical paths and the spatial propagation characteristics of the wireless channels are nearly unchanged within certain bandwidth [4], there is an intrinsic relation between the uplink and the downlink CSI.

In the following, we first define an uplink-to-downlink mapping function, following the approach in [14], and prove its existence. Then, we leverage deep learning to find the mapping function.

A. Existence of Uplink to Downlink Mapping

Consider the channel model in Eq. (1), where the channel function $\boldsymbol{h}(f)$ is completely determined by the parameters α_p, ϕ_p, τ_p, P, $\Delta\theta$, and θ. As discussed at the end of Section II, α_p, ϕ_p, τ_p, P, and $\Delta\theta$ are the functions of the communication environment (including the antenna gains, scatterers, etc.), mean DOA θ and distance D.

Definition 1: The position-to-channel mapping $\boldsymbol{\Phi}_f$ can be written as follows,

$$\boldsymbol{\Phi}_f : \{(D, \theta)\} \to \{\boldsymbol{h}(f)\}, \qquad (3)$$

where the sets $\{(D, \theta)\}$ and $\{\boldsymbol{h}(f)\}$ are the domain and codomain of the mapping $\boldsymbol{\Phi}_f$, respectively.

Then, we adopt the following assumption for further analysis.

Assumption 1 [14]: The position-to-channel mapping function, $\boldsymbol{\Phi}_f : \{(D, \theta)\} \to \{\boldsymbol{h}(f)\}$, is bijective.

The *Assumption 1* means that every user position has a unique channel function $\boldsymbol{h}(f)$, and vice versa. Although it

cannot be proved analytically, the probability that $\mathbf{\Phi}_f$ is bijective is actually very high in practical wireless communication scenarios, and approaches 1 as the number of antennas at the BS increases [14]. Therefore, it is reasonable to adopt *Assumption 1* in massive MIMO systems.

Under *Assumption 1*, the channel-to-position mapping, i.e., the inverse mapping of $\mathbf{\Phi}_f$, exists, which can be written as follows:

$$\mathbf{\Phi}_f^{-1} : \{\boldsymbol{h}(f)\} \to \{(D, \theta)\}. \tag{4}$$

Next, we investigate the existence of the uplink-to-downlink mapping, as given in *Proposition 1*.

Proposition 1: With *Assumption 1*, the uplink-to-downlink mapping exists for a given communication environment, and can be written as follows,

$$\mathbf{\Psi}_{\text{U}\to\text{D}} = \mathbf{\Phi}_{f_{\text{D}}} \circ \mathbf{\Phi}_{f_{\text{U}}}^{-1} : \{\boldsymbol{h}(f_{\text{U}})\} \to \{\boldsymbol{h}(f_{\text{D}})\}, \tag{5}$$

where $\mathbf{\Phi}_{f_{\text{D}}} \circ \mathbf{\Phi}_{f_{\text{U}}}^{-1}$ represents the composite mapping related to $\mathbf{\Phi}_{f_{\text{D}}}$ and $\mathbf{\Phi}_{f_{\text{U}}}^{-1}$.

Proof: From the *Definition 1*, we have the mappings $\mathbf{\Phi}_{f_{\text{D}}} : \{(D, \theta)\} \to \{\boldsymbol{h}(f_{\text{D}})\}$ and $\mathbf{\Phi}_{f_{\text{U}}} : \{(D, \theta)\} \to \{\boldsymbol{h}(f_{\text{U}})\}$ exist in position candidate set $\{(D, \theta)\}$. Under *Assumption 1*, the mapping $\mathbf{\Phi}_{f_{\text{U}}}^{-1}$ exists with its codomain equal to the domain of $\mathbf{\Phi}_{f_{\text{D}}}$. Therefore, the composite mapping $\mathbf{\Phi}_{f_{\text{D}}} \circ \mathbf{\Phi}_{f_{\text{U}}}^{-1}$ exists for any possible position (D, θ). A more general proposition can be found in [14]. □

B. Deep Learning for Uplink-to-Downlink Mapping

Proposition 1 proves the existence of the uplink-to-downlink mapping function. However, the function cannot be depicted by known mathematical models, which motivates us to resort to deep learning algorithms. Based on the universal approximation theorem [15], we obtain *Theorem 1* as following.

Theorem 1: For any given small error $\varepsilon > 0$, there always exists a positive constant N large enough such that

$$\sup_{\boldsymbol{x}\in\mathbb{H}} \|\text{NET}_N(\boldsymbol{x}, \mathbf{\Omega}) - \mathbf{\Psi}_{\text{U}\to\text{D}}(\boldsymbol{x})\| \leq \varepsilon, \quad \mathbb{H} = \{\boldsymbol{h}(f_{\text{U}})\}, \tag{6}$$

where $\text{NET}_N(\boldsymbol{x}, \mathbf{\Omega})$ is the output of a three-layer feedforward network with $\boldsymbol{x}$, $\mathbf{\Omega}$ and N denoting the input data, network parameters, and the number of hidden units, respectively.

Proof: (i) Since $\boldsymbol{h}(f_{\text{U}})$ is bounded and closed, $\mathbb{H}$ is a compact set; (ii) Since $\mathbf{\Phi}_{f_{\text{D}}}$ and $\mathbf{\Phi}_{f_{\text{U}}}^{-1}$ are continuous mapping and the composition of continuous mappings is still a continuous mapping, we know that for $\forall \boldsymbol{x} \in \mathbb{H}$ such that $\mathbf{\Psi}_{\text{U}\to\text{D}}(\boldsymbol{x})$ is a continuous function. Based on (i), (ii), and the universal approximation theorem [15, Theorem 1], *Theorem 1* is proved. □

According to *Theorem 1*, the uplink-to-downlink mapping function can be approximated with an arbitrarily small error by a feedforward network with a single hidden layer. Thus, we can train a network to predict the downlink CSI from the uplink CSI and can significantly reduce the overhead required for downlink training and uplink feedback at the cost of off-line training.

IV. SCNet BASED DOWNLINK CSI PREDICTION

In this section, we will first introduce the architecture of the SCNet. Then, we discuss how to train and deploy it in massive MIMO systems.

A. SCNet Architecture

Although it has been proven in *Theorem 1* that a three-layer network is able to predict the downlink CSI, we propose the SCNet instead of the three-layer network for practical considerations as follows: (i) A deep network with an appropriate number of layers learns better than a three-layer network; (ii) A spare network can reduce the network parameters, and therefore is easier to train and is more robust; (iii) Compared with the real-valued networks, the complex ones have richer representational capacities and therefore are more powerful in learning complex-valued functions [16].

As shown in Fig. 2, the input of the SCNet is the uplink CSI $\boldsymbol{h}(f_{\text{U}})$. The output of the SCNet is a cascade of nonlinear transformation of $\boldsymbol{h}(f_{\text{U}})$, i.e.,

$$\hat{\boldsymbol{h}}(f_{\text{D}}) = \text{NET}(\boldsymbol{h}(f_{\text{U}}), \mathbf{\Omega}) = \boldsymbol{f}^{(L-1)}\left(\cdots \boldsymbol{f}^{(1)}\left(\boldsymbol{h}\left((f_{\text{U}})\right)\right)\right), \tag{7}$$

where L is the number of layers and $\mathbf{\Omega} \triangleq \left\{\boldsymbol{W}^{(l)}, \boldsymbol{b}^{(l)}\right\}_{l=1}^{L-1}$ is the network parameters to be trained. Moreover, $\boldsymbol{f}^{(l)}$ is the nonlinear transformation function of the l-th layer and can be written as,

$$\boldsymbol{f}^{(l)}(\boldsymbol{x}) = \begin{cases} \boldsymbol{g}\left(\boldsymbol{W}^{(l)}\boldsymbol{x} + \boldsymbol{b}^{(l)}\right), & 1 \leq l < L-1; \\ \boldsymbol{W}^{(l)}\boldsymbol{x} + \boldsymbol{b}^{(l)}, & l = L-1, \end{cases} \tag{8}$$

where $\boldsymbol{g}$ is the activation function and is given by

$$\boldsymbol{g}(\boldsymbol{z}) = \max\{\Re[\boldsymbol{z}], \mathbf{0}\} + j\max\{\Im[\boldsymbol{z}], \mathbf{0}\} \tag{9}$$

with $\Re[\cdot]$ and $\Im[\cdot]$ being the real and imaginary parts of the vectors, respectively.

We set the number of neurons in the middle hidden layer to be much fewer than that in the output layer, which forces the SCNet to compress the representation of the input. We would like to emphasize that the compression task would be very difficult if the elements of input $\boldsymbol{x}$ are independent of each other. However, since there exists the sparse structure[3] in the uplink channel $\boldsymbol{h}(f_{\text{U}})$, the SCNet is able to discover the intrinsic sparsity of $\boldsymbol{h}(f_{\text{U}})$ in massive MIMO systems. As a result, the SCNet can not only reduce the redundancy of network parameters but also become more functional and robust [17].

B. Training and Deployment

The proposed downlink CSI prediction has two stages, i.e., the off-line training and the on-line deployment stages. In the off-line training stage, the BS collects both the downlink and the uplink CSI as training samples to train the SCNet. Specifically, during a coherence time period, the downlink CSI

---

[3]Since AS is narrow, the massive MIMO channels exhibit sparsity in the angular domain. See more details in [1].

(a) NMSE performance versus AS $\Delta\theta$

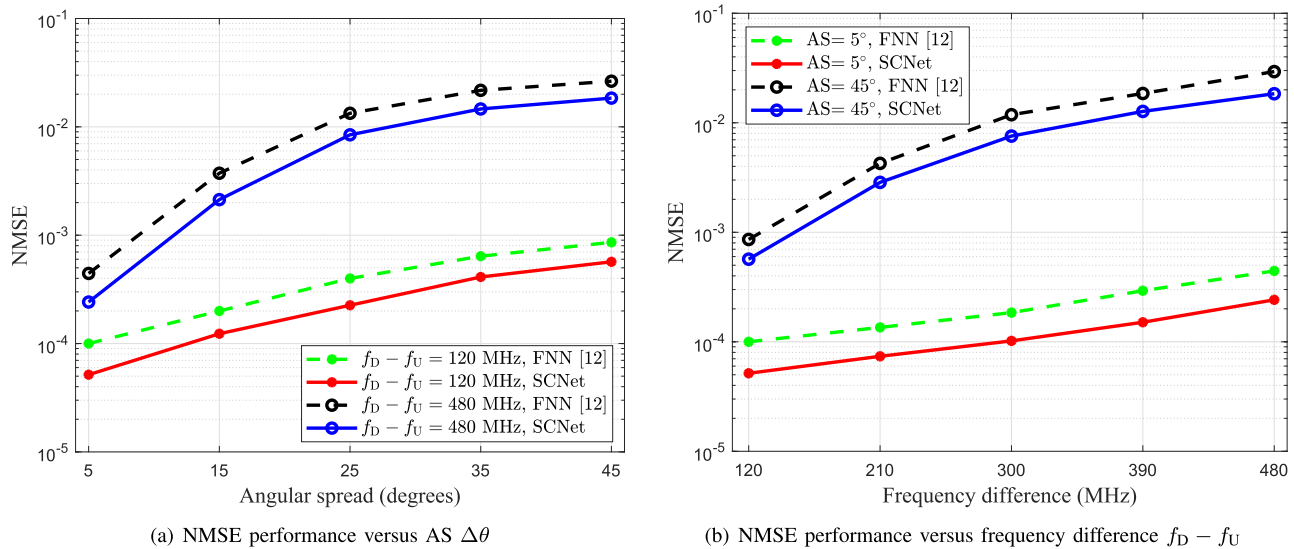(b) NMSE performance versus frequency difference $f_D - f_U$

Fig. 3.   The NMSE performance of the SCNet and the FNN based downlink CSI predictors versus AS (a) and the frequency difference $f_D - f_U$ (b).

is first estimated at the user side by downlink training and then fed back to the BS. The uplink CSI is estimated at the BS by uplink training. The SCNet is trained to minimize the difference between the output $\hat{h}(f_D)$ and the supervise label $h(f_D)$. The loss function is

$$\text{Loss}\,(\boldsymbol{\Omega}) = \frac{1}{VN_h} \sum_{v=0}^{V-1} \left\| \hat{\boldsymbol{h}}(f_D)^{(v)} - \boldsymbol{h}(f_D)^{(v)} \right\|_2^2, \quad (10)$$

where V is the batch size,[4] the superscript (v) denotes the index of the v-th training sample, $\|\cdot\|_2$ denotes the ℓ_2 norm, and N_h is the length of the vector $\boldsymbol{h}(f_D)$. The loss function Loss $(\boldsymbol{\Omega})$ is minimized by the complex designed adaptive moment estimation (ADAM) algorithm [16] until the SCNet converges.

In the deployment stage, the parameters of the SCNet are fixed. The SCNet directly generates the prediction of the downlink CSI $\hat{\boldsymbol{h}}(f_D)$ based on the uplink CSI $\boldsymbol{h}(f_U)$.

C. Complexity Analysis

Denote n_l as the number of neurons in the l-th layer. The required number of floating point operations (FLOPs) is used as the metric of complexity. For real-valued network, the total number of FLOPs required is $\sum_{l=1}^{L-1} n_{l-1}n_l$. However, as a complex multiplication is 4 times of its real counterpart, the total number of FLOPs required in the SCNet is $4\sum_{l=1}^{L-1} n_{l-1}n_l$. Nevertheless, it should be noted that in a real-valued network, the input complex data are typically separated to real and imaginary parts and then fed to the network. Therefore, the size of the real-valued network is lager than that of a complex-valued network.

V. SIMULATION RESULTS

Unless otherwise specified, the system parameters are set as follows: the BS is equipped with 128 antennas; the uplink frequency follows the 3GPP R15 standard, i.e., $f_U = 2.5$ GHz.

In the simulation of Section V-A, the attenuation of each path follows Rayleigh distribution. The phase and delay of each path follow uniform distribution over $[-\pi, \pi)$ and $[0, 10^{-4}]s$. The number of paths is 200 in both the training and deployment stages. While in the simulation of Section V-B, the parameters of each path are generated according to the ray-tracing simulator [18]. The number of paths is 200 in the training stage and varies in the deployment stage.

An FNN in [12] is originally designed for uplink/downlink channel calibration for massive MIMO systems, which can also be used for the downlink channel prediction in the FDD massive MIMO systems. Therefore, the FNN is used as a benchmark in this letter. Keras 2.2.0 is employed as the deep learning framework for both the SCNet and the FNN. We choose the number of neurons in the hidden layer as (128, 64, 128) by trails and adjustments. The initial learning rate of the ADAM algorithm is 0.001. The batch size is 128. The parameters of the SCNet are initialized as complex distribution with normalized variance.[5] The uplink CSI fed to the SCNet is estimated by the minimum mean-squared-error (MMSE) algorithm when the signal-to-noise ratio (SNR) is 25 dB. The network is trained for each AS degree and each downlink frequency separately. The number of training samples is 102,400, and the number of epochs is 400.

A. Prediction Accuracy Versus AS and Frequency Difference

Normalized MSE (NMSE) is used to measure the prediction accuracy, which is defined as

$$\text{NMSE} = E\left[\left\| \boldsymbol{h}_D - \hat{\boldsymbol{h}}_D \right\|_2^2 / \left\| \boldsymbol{h}_D \right\|_2^2 \right], \quad (11)$$

where $E[\cdot]$ represents the expectation operation.

Fig. 3 depicts the NMSE performance of the SCNet and the FNN based downlink CSI predictors versus AS $\Delta\theta$ and frequency difference $f_D - f_U$, respectively. Fig. 3(a) shows

---

[4]Batch size is the number of samples in one training batch.

[5]The weights of neurons in the l-th layer are initialized as complex normal variables with variance $1/n_l$.
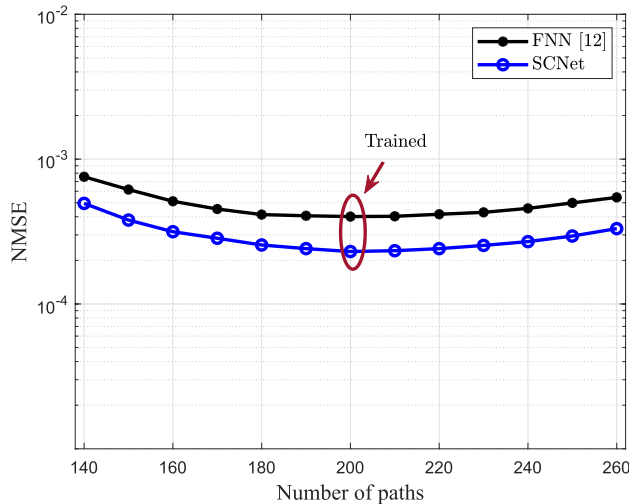
Fig. 4. The NMSE performance of the SCNet and the FNN [12] based downlink CSI predictors when mismatches occur between the training and deployment stages.

that the NMSE performance of both the SCNet and the FNN degrades as AS increases while the slope of the NMSE curve decreases as AS increases. This is because that as AS increases, the sparsity of channels in the angular domain decreases, and thus it is harder for the networks to learn the structure of the channels and to accurately predict the downlink CSI. The networks are less sensitive to AS in the wider AS case,[6] which accounts for the decrease of the slope in the wide AS case. Fig. 3(b) shows that the NMSE performance of both the SCNet and the FNN degrades as the frequency difference increases. This is because the correlation of CSI between the uplink and the downlink tends to vanish as the frequency difference increases. As shown in Fig. 3, the proposed the SCNet outperforms the FNN in all scenarios, which validates that the SCNet can benefit from the rich representational capacity offered by complex representations.

B. Robustness Analysis

In Sections V-A, the channels are generated based on Eq. (1) with the same statistics. However, channels in real-world may be more complicated and the statistics mismatches between the training and deployment stages are also inevitable. To test the robustness of both the SCNet and the FNN, data generated from Wireless InSite [18] under different scenarios are used to train and test. As shown in Fig. 4, the number of paths in the training stage is 200 while it varies in the deployment stages. The results show that the variations on statistics of channel degrade the performance, but the SCNet and the FNN still exhibit remarkable prediction accuracy, which validates the excellent generalization ability of deep neural networks.

---

[6]More specifically, AS increases by 100% from $5°$ to $10°$ while it increases by 25% from $20°$ to $25°$. The decreasing proportion renders the SCNet less sensitive to AS in wide AS case.

VI. CONCLUSION

In this letter, we revealed the existence of a deterministic uplink-to-downlink mapping function for a given communication environment. Then, we proposed the SCNet for the downlink CSI prediction in FDD massive MIMO systems. Simulation results have demonstrated that the SCNet performs better than the existing network in terms of prediction accuracy. Furthermore, the remarkable robustness of the SCNet with respect to the statistic characteristics of wireless channels has shown its great potential in real-world applications.

REFERENCES

[1] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[2] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, Aug. 2012, pp. 811–815.

[3] Y. Zhou, M. Herdin, A. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2007, vol. 5, pp. 10–15.

[4] K. Hugl, K. Kalliola, and J. Laurila, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," in *Proc. COST*, vol. 273, 2002, p. 066.

[5] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.

[6] Y. Han, Q. Liu, C.-K. Wen, S. Jin, and K.-K. Wong, "FDD massive MIMO based on efficient downlink channel reconstruction," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4020–4034, Jun. 2019.

[7] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[8] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, Mar. 2019.

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

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

[11] M. S. Safari and P. Pourahmadi, "Deep UL2DL: Channel knowledge transfer from uplink to downlink," 2018, *arXiv:1812.07518*. [Online]. Available: https://arxiv.org/abs/1812.07518

[12] C. Huang, G. C. Alexandropoulos, A. Zappone, C. Yuen, and M. Debbah, "Deep learning for UL/DL channel calibration in generic massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, Mar. 2019, pp. 1–6.

[13] Y. Han, T. Hsu, C. Wen, K. Wong, and S. Jin, "Efficient downlink channel reconstruction for FDD multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3161–3176, Jun. 2019.

[14] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," 2019, *arXiv:1905.03761*. [Online]. Available: https://arxiv.org/abs/1905.03761

[15] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[16] C. Trabelsi *et al.*, "Deep complex networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, May 2018.

[17] A. Ng, "Sparse autoencoder," *CS294A Lect. Notes*, vol. 72, no. 2, pp. 1–19, 2011.

[18] *Remcom Wireless Insite*. Accessed: Apr. 2019. [Online]. Available: https://www.remcom.com/wireless-insite-em-propagation-software