Overcoming the Channel Estimation Barrier in Massive MIMO Communication via Deep Learning

Zhenyu Liu, Lin Zhang, and Zhi Ding

Abstract

A new wave of wireless services, including virtual reality, autonomous driving and Internet of Things, is driving the design of new generations of wireless systems to deliver ultra-high data rates, massive numbers of connections and ultra-low latency. Massive multiple-input multiple-output (MIMO) is one of the critical underlying technologies that allow future wireless networks to meet these service needs. This article discusses the application of deep learning (DL) for massive MIMO channel estimation in wireless networks by integrating the underlying characteristics of channels in future high-speed cellular deployment. We develop important insights derived from the physical radio frequency (RF) channel properties and present a comprehensive overview on the application of DL for accurately estimating channel state information (CSI) with low overhead. We provide examples of successful DL application in CSI estimation for massive MIMO wireless systems and highlight several promising directions for future research.

Introduction

Current and future generations of wireless networks must cope with the continuous and rapid growth of applications and data traffic to deliver ultra-high data rate over a wide coverage area for a massive number of connected devices and to support low-latency applications. One of the most important technical advances at the radio frequency (RF) physical layer is the emergence of massive multiple-input multiple-output (MIMO) transceivers. By exploiting spatial diversity and multiplexing gains, massive MIMO can help improve spectrum efficiency and robustness of wireless communication systems under limited bandwidth and channel fading. To fully utilize their potential gains, massive MIMO transmitters require sufficiently accurate channel state information (CSI) on the forward link.

However, the need for accurate CSI in massive MIMO systems poses serious challenges to the traditional channel estimation and feedback techniques. On the one hand, owing to the large number of antennas and the wide bandwidth, channel estimation in massive MIMO systems suffers from high signal acquisition costs and large training overhead. On the other hand, different uplink and downlink frequency bands in frequency division duplex (FDD) mode lead to weaker reciprocity between the two channels. Consequently, gNB (or gNodeB) transmitters in FDD networks would require user equipment (UE) to provide downlink CSI feedback frequently, which can consume a staggering amount of uplink channel capacity in massive MIMO communication.

Recently, deep learning (DL) has emerged as a powerful tool for learning the underlying structures from large measurement of data, and has achieved notable success in areas including computer vision and natural language processing. Although still in a nascent stage, DL has recently found several interesting applications in the physical layer of wireless communications, including low rate CSI feedback, channel estimation and signal detection, among others. However, how to effectively apply DL techniques to enhance the estimation accuracy and efficiency by exploiting RF channel properties remains an open research issue, as many existing works do not explicitly utilize the physical RF characteristics and provide insufficient physical insights despite apparent successes.

In wireless communications, there exists a wealth of expert knowledge on various channel models for designing and achieving fast and reliable data links. Specifically, MIMO channels exhibit a number of important physical characteristics including spatial correlation, spectral correlation and temporal correlation. Based on these special characteristics of physical wireless channels, appropriately designed DL architectures and algorithms can potentially help increase the accuracy and robustness of CSI estimation in massive MIMO links. Clearly, how to configure, adapt and improve such tools for accurate CSI estimation by massive MIMO gNB represents an important technical barrier, as well as an exciting and promising research issue.

In this article, we elaborate the insight of integrating channel correlation with DL-based channel estimation, present an overview of the integration of DL in massive MIMO systems, and outline some future research directions. The rest of this article is organized as follows. The next section introduces the challenges in massive MIMO channel estimation using conventional methods. Then, we introduce the important basis of DL and three correlations in wireless channels for reducing the

Digital Object Identifier:
10.1109/MWC.001.1900413

*Zhenyu Liu and Lin Zhang are with Beijing University of Posts and Telecommunications;
Zhi Ding (corresponding author) is with University of California, Davis.*

overhead required in channel estimation. Following that, we present an overview of integrating channel correlations with DL in massive MIMO CSI acquisition. We then discuss open research issues in DL-based channel estimation before this article is concluded.

Challenges in Channel Estimation

In massive MIMO systems, an accurate acquisition of CSI is essential to achieve spectrum and energy efficiency. One of the most widely used approaches in massive MIMO channel estimation is employing pilot signals to estimate the channel. With the received data and the knowledge of training symbols, least square (LS) and minimum mean square error (MMSE) can be used to infer the CSI. Although the LS method is well known for its low complexity, the estimation accuracy is usually not satisfying. MMSE can achieve accurate CSI estimation, but it requires channel correlation matrix and noise variance as prior information, and also suffers from high computation complexity.

The complex channel conditions and low-cost hardware also raise the situations where the traditional methods generate limited performance. On the one hand, pilot contamination can significantly degrade the system's performance due to the noise and the limited number of orthogonal pilots, which hence have to be reused in adjacent cells. On the other hand, due to the fast time-varying and non-stationary characteristics, it is hard for massive MIMO systems to design an efficient channel estimation algorithm in high mobility environments. In addition, the requirements of millimeter-wave (mmWave) massive MIMO also bring new challenges. Due to the limited physical space with closely placed antennas and prohibitive power consumption in mmWave massive MIMO systems, it is difficult to equip a dedicated RF chain for each antenna. Consequently, channel estimation under a limited number of RF chains and low-resolution analog-to-digital converters (ADCs) becomes essential.

Even if the UEs have succeeded in acquiring accurate downlink CSI, the feedback of CSI to the gNB is still challenging with affordable overhead in FDD massive MIMO systems. Due to massive antennas at the gNB, each UE has to feedback the CSI associated with hundreds of transmit antennas and large bandwidth, which results in prohibitive feedback overhead. While codebook-based CSI quantization is time-efficient, the resulting compression often consumes too much uplink bandwidth to be practical.

To overcome challenges in massive MIMO systems, compressive sensing (CS) has been studied for channel estimation and channel feedback. CS-based approaches exploit the spatial coherence and channel sparsity that stem from the limited scattering characteristics of signal propagation and strong spatial correlation inherent in the antennas at the gNB, and can formulate a compressed representation of CSI matrices. However, there are also limitations. For example, most CS-based approaches impose strong channel sparsity in some domain which may not hold exactly. When the CSI matrix is not exactly sparse on the channel sampling grids, it may lead to degraded performance due to the power leakage effect around the recovered discrete CSI samples. Furthermore, CS-based approaches are often iterative, which can cause additional delay. Consequently, new methods are required to enhance the CSI estimation performance in massive MIMO systems.

Due to massive antennas at the gNB, each UE has to feedback the CSI associated with hundreds of transmit antennas and large bandwidth, which results in prohibitive feedback overhead. While codebook-based CSI quantization is time-efficient, the resulting compression often consumes too much uplink bandwidth to be practical.

Deep Networks for Channel Estimation

As a powerful tool to extract the underlying features from the measured data, DL is expected to provide improved solutions to the complicated problems in massive MIMO communications. In this section, we introduce the basics of several relevant neural networks for DL-based channel estimation.

Fully Connected Deep Neural Network (DNN): Fully connected DNNs can extract appropriate features for classification and regression, as shown in Fig. 1a. Starting by sending measurement data to the input layer, each successive layer attempts feature extraction from the input data, gradually accentuating features that affect decision making while suppressing irrelevant features. Through the optimization of network parameters, DNN can be trained to capture underlying data structures and models despite outliers and noises.

Convolutional Neural Networks (CNNs): CNNs have been widely applied in problems such as image processing. As shown in Fig. 1b, CNNs specialize in processing data with grid-like structures and include special layers for functions such as convolution and pooling. By stacking the convolutional and pooling layers alternately, CNN can progressively learn complex models. Considering the correlation in spatial, temporal, or spectral domains, CSI matrices of massive MIMO systems can often be viewed as two-dimensional images. CNNs have strong potential for success in channel estimation.

Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM): RNN is a class of neural networks that exploit sequential information by using earlier outputs as part of inputs in the later time. Unlike the above two neural networks, RNNs use internal state to store the previous information that has been calculated. An RNN consisting of LSTM units is often known as an LSTM network. LSTM networks can handle exploding and vanishing gradients in traditional RNNs, and work well in prediction and classification according to time series data. Exploiting the temporal correlation of CSI, RNN and LSTM can further improve the accuracy of CSI estimation.

Autoencoder: An autoencoder is a neural network trained to efficiently regenerate its input. As shown in the structure of Fig. 1d, modern autoencoders have generalized the idea of an encoder and a decoder beyond deterministic functions to stochastic mappings. An autoencoder can acquire compressed but robust representations of its input and can be highly effective and efficient in dimension reduction or feature learning. From a DL perspective, CSI feedback within a wireless communication system can be viewed as a particular type of autoencoder, which aims to recover the downlink CSI at the gNB based on its received CSI that was compressed and sent by the UE.

Exploiting Wireless Channel Correlations

Learning and exploiting channel correlations in massive MIMO can substantially benefit CSI estimation and feedback for massive MIMO. In our

With the help of temporal correlation, UE can encode and feed back the CSI variations, instead of the full CSI. The gNB can combine the new feedback with its previously estimated CSI within coherence time for subsequent CSI reconstruction. The temporal correlation feature is a good match with the capabilities of RNN and LSTM that are effective in sequential data processing.

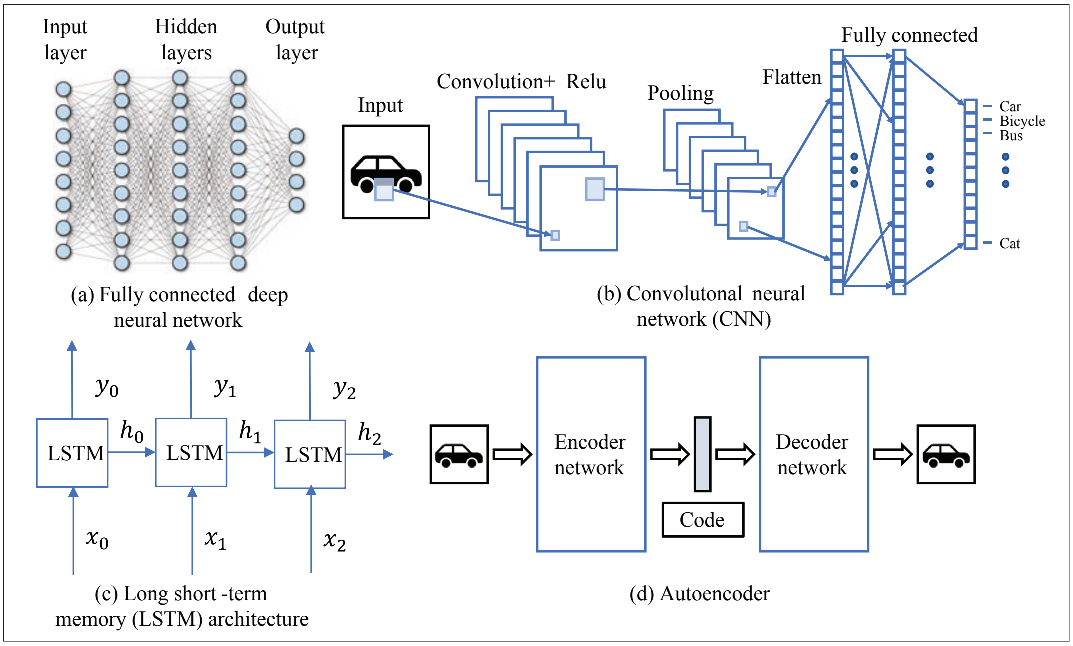

FIGURE 1. Commonly utilized deep learning architectures.

work [1], we have presented a DL-based CSI feedback solution for massive MIMO communications. We achieved high efficiency by exploiting the underlying spatial correlation and the correlation between uplink and downlink. Stimulated by this and other preliminary successes, we investigate the underlying CSI data structure in terms of spatial and spectral correlation, temporal correlation, and bi-directional correlation of channels for channel estimation in massive MIMO systems. Note that, however, these physical characteristics also apply for conventional MIMO systems and can reduce the amount of necessary CSI feedback from UEs.

Spatial and Spectral Correlation

Spatial and spectral correlation of CSI has been commonly exploited in CS-based feedback and CSI estimation. Physical RF propagation elements such as multipaths and scatters provide the foundation of spatial and spectral correlation. For a given gNB, channels are governed by cell-specific attributes such as the buildings, ground and vehicles. Smaller antenna separation leads to higher spatial correlation in massive MIMO. Since the physical size of antenna arrays at gNB is small compared to path distance, paths between different gNB-UE antenna pairs share some common properties [2]. On the other hand, spectral coherence measures channel similarity across frequency. For subcarriers within channel coherence bandwidth, their channels exhibit strong correlation. Influenced by the spatial and spectral correlation, massive MIMO channels exhibit some sparsity in the corresponding transform domains including virtual angular domain and delay domain, which can be leveraged in channel estimation and feedback to formulate a sparse signal recovery problem. Unlike the traditional approaches that often require reasonably accurate correlation models, DL can learn and exploit these underlying channel correlation structures to reduce the CSI feedback overhead and increase the estimation accuracy.

Temporal Correlation

It is well known that even for the mobile environment under severe Doppler effect, RF channel responses are temporally correlated in typical massive MIMO configurations. For mobile users, coherence time can measure the temporal channel variations and describes the Doppler effect caused by mobility. Since gNB and UE can store their previous CSI estimates, temporal correlation of massive MIMO channels can be exploited to reduce the amount of pilots and UE feedback in downlink CSI estimation. With the help of temporal correlation, UE can encode and feed back the CSI variations, instead of the full CSI. The gNB can combine the new feedback with its previously estimated CSI within coherence time for subsequent CSI reconstruction. The temporal correlation feature is a good match with the capabilities of RNN and LSTM that are effective in sequential data processing.

Bi-Directional Correlation

Historically, the uplink-downlink channel reciprocity has predominantly been utilized by time division duplex (TDD) gNB to infer downlink CSI from its own uplink CSI estimate. For FDD systems, however, RF components that positively superimpose in one frequency band may cancel each other in another. Hence, FDD uplink-downlink channels do not exhibit direct reciprocity. Nevertheless, because of the shared propagation environment, correlation still exists between the two. For example, the angles of arrival of signals in the uplink transmission are almost the same as the angles of departure of signals in the downlink transmission. When the band gap between uplink and downlink channels is moderate, both links should share similar propagation characteristics including scattering effects. Such correlation can be exploited by gNB for CSI acquisition in massive MIMO systems.

Existing works have demonstrated certain levels of correlation between bi-directional channels for

FDD systems. The directional properties of uplink and downlink FDD channels have been shown as correlated [3]. Downlink channel covariance estimation can also benefit from the observed uplink covariance. Similarly, although channel responses could vary for different downlink and uplink frequency bands, their multipath delays remain physically the same. In [4], the multipath delay reciprocity and angle reciprocity have been verified by the channel measurements. Thus, to utilize the bi-directional channel correlation, channel response matrix from the frequency domain should be transformed to the delay domain using inverse Fourier transform. Compared with the frequency domain, the reciprocity in the delay domain is evident owing to the shared multipath delays.

To demonstrate the bi-directional correlation, we illustrate the correlation coefficient between uplink and downlink FDD channels obtained through numerical tests in Fig. 2. The COST 2100 channel model [5] is utilized to generate the uplink and downlink CSI matrices. The COST 2100 channel model is a geometry-based stochastic channel model that can capture the stochastic properties of MIMO channels over frequency, space and time. Specifically, random scatterer positions are generated according to a probability density function before generating the corresponding channel response. In this way, this model can generate a large set of random channels typically encountered by wireless networks in the real world [5]. 5.1 GHz uplink and 5.3 GHz downlink channel responses are generated with the bandwidths of 20 MHz. We place gNB at the center of a square area with lengths of 20 meters for indoor coverage and randomly place UEs within the coverage area. The gNB uses uniform linear array (ULA) with 32 antennas and 1024 subcarriers. After transforming the channel matrix into the delay domain, only the first 32 rows are retained due to sparsity. 10,000 items are generated for the correlation evaluation. As shown in Fig. 2, the correlation coefficients between uplink CSI and downlink CSI using the "original" (real/imaginary) format are quite erratic. Since the CSI is complex-valued, their real part and imaginary part correlations are evaluated. The test results appear to show weak correlation between corresponding downlink-uplink channel responses.

A closer examination of the physics reveals that in FDD, CSIs of two carriers in different frequencies may have uncorrelated phases. However, based on FDD multipath channel models, the CSI magnitudes in the delay domain should exhibit much stronger correlation. To confirm, we transform the CSI elements into polar coordinates to separately consider their magnitude and phase correlations. Figure 2 shows that the corresponding magnitudes of uplink and downlink CSIs exhibit strong correlation whereas their corresponding phases show very weak correlation. In fact, even by removing the signs from the CSI's real and imaginary parts, Fig. 2 shows that the absolute values (ABS) of uplink and downlink CSI coefficients are also strongly correlated. However, their signs show little correlation. These results demonstrate some shared characteristics between uplink and downlink channels in the delay domain. This observation provides the basic principle for utilizing magnitude correlation between uplink and downlink channels in the delay domain for estimating CSI of massive MIMO systems.

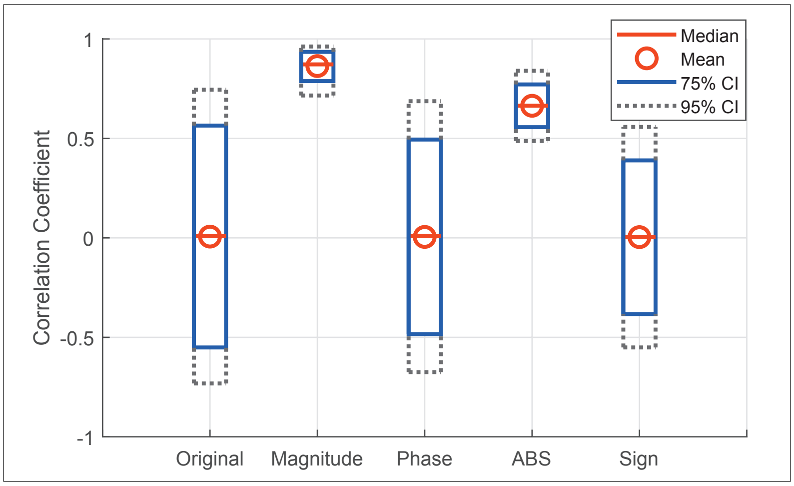

FIGURE 2. Distribution of correlation coefficient between uplink and downlink CSI at various levels of Confidence Interval (CI) [1].

DL-Based CSI Estimation

To present a clearer perspective and context, in this section we introduce the applications of deep networks and underlying channel characteristics in designing and improving channel estimation and feedback systems.

Channel Estimation

Massive MIMO has been viewed as an important technique for 5G and beyond, while there are still some channel estimation barriers to be overcome.

Orthogonal frequency division multiplexing (OFDM), as a technology used in 5G, exploits the received pilots in some sparsely and evenly placed grids to infer the CSIs in the overall grids. However, the performances of LS and MMSE estimators are not satisfying due to the accuracy and the requirement for prior information, respectively. To conduct a capable estimator, channel estimation using the CSIs in the pilot locations can be modeled as a low-resolution image processing problem in DL. With the help of channel correlations, DL networks can be used to interpolate the CSIs in the missing locations and reconstruct a high-resolution CSI matrix. In [6], a spatial pilot-reduced CNN (SPR-CNN) is designed to save spatial pilot overhead for massive MIMO-OFDM channel estimation. Specifically, CNN is utilized to exploit the spatial and spectral correlation to enhance the channel estimation accuracy, while a channel estimation unit with memory is utilized to group and cache channels in successive coherence intervals and incorporate the temporal correlation. Numerical results demonstrate that the SPR-CNN based approach can save two thirds of spatial pilot overhead with minor performance loss. Also, DL can be used to reduce the influence of quantization distortion from the low-resolution ADCs by mapping the signals from high-resolution ADC antennas to those of low-resolution ADC antennas with the help of spatial correlation [7].

CSI estimation in massive MIMO is more difficult in high mobility environments due to the fast time-varying and non-stationary channel characteristics. To adapt the characteristics of fast time-vary-

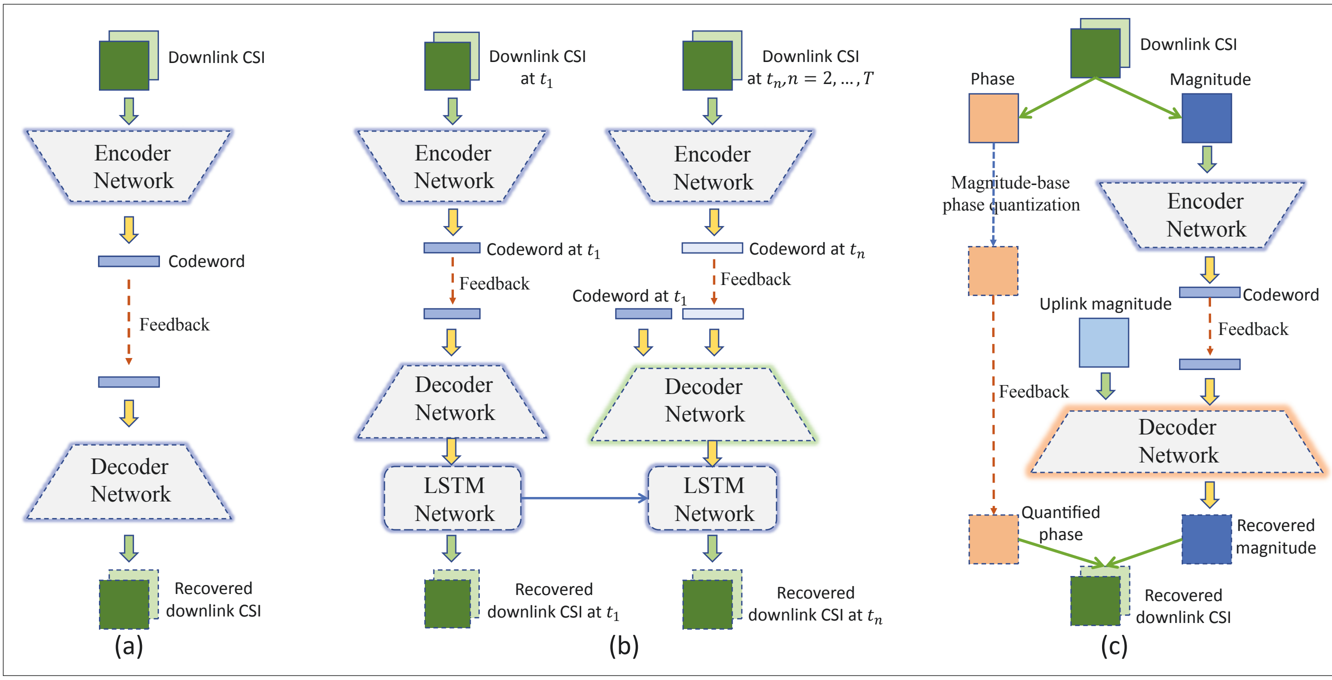

FIGURE 3. CSI feedback architectures: a) architecture of CsiNet in [12]; b) architecture of CsiNet-LSTM in [13]; c) architecture of DualNet-MAG in [1].

ing channels, LSTM networks, which have inherent memory cells to keep the previously extracted information, can be used to learn the temporal correlation and make the CSI prediction. In [8], a DL network has been proposed to handle the massive MIMO channel estimation problem in high-speed mobile scenarios by exploiting the temporal correlation as well as the spatial and spectral correlation. Specifically, the CSI data between all pairs of antennas in the frequency domain and time domain are treated as a 4-dimensional tensor as the input to the DL network. Then, CNN is used to mimic the interpolation processes for the CSI in no-pilot locations in the spatial and frequency domains, and LSTM units are used to learn the temporal channel variation for CSI prediction. Simulation results show that this DL-based method is superior to LS and linear minimum mean squared error (LMMSE) method in high-speed mobile scenarios.

Another major issue of CSI estimation in massive MIMO systems is the estimation error caused by channel noise and pilot contamination (i.e., interference among pilot symbols transmitted by users in neighboring cells). As DL techniques have benefited from image denoising, a CNN-based denoising network can mitigate noise and interference effects. With the help of features extracted from spatial, spectral and temporal correlations, denoising can substantially improve the performance of conventional and DL-based solutions. In [9], taking the received signal in the spatial, frequency, and time domains as a 3-dimensional input data, a CNN-based deep image prior network designed to denoise can extract the received signal before channel estimation for multi-cell interference-limited massive MIMO systems. Next, by applying denoised data with LS, this estimator can approach the MMSE estimator performance for high-dimensional signals. In [10], a denoising CNN has been designed to enhance the performance of the approximate message passing (AMP) algorithm, which can recover the high-dimensional beamspace massive MIMO channel that exhibits sparsity with low computational complexity. The results show that this learning-based denoising AMP scheme can outperform the AMP schemes using conventional denoisers.

By exploiting channel correlations, DL can be exploited to develop new channel estimation methods and to enhance the performance of conventional ones. The integration of DL and channel correlations in CSI estimation can be summarized as follows. First, owing to the physical correlations in the spatial, frequency and time domains, CNN-based techniques that have been successful in high-resolution image restoration can be helpful in pilot reduction and CSI interpolation. Also, RNN can improve the estimation and prediction accuracy for fast time-varying channels. Second, using CNN to exploit the inherent CSI correlations, CSI estimates can be more robust against channel noise and quantization error effects. Furthermore, based on many successful applications of DL network in capturing nonlinear functional relationships, DL can better incorporate bi-directional correlation to further improve TDD channel calibration accuracy in massive MIMO systems [11]. We note, however, that the optimization on the allocation and placement of pilots for DL-based channel estimation in a specific environment remains an open problem under investigation, particularly since CSI correlations vary with the environment.

Channel Feedback

By integrating the DL with channel properties, there are research works showing substantial performance improvement for downlink CSI acquisition (Table 1) in FDD massive MIMO systems with limited feedback resources.

The authors of [12] proposed a DL-based CsiNet, as illustrated in Fig. 3a, to reduce UE feed-

	Spatial and spectral correlation	Temporal correlation	Bi-directional correlation
Channel estimation	Accuracy enhancement [6], denoising [9, 10], pilot reduction [6], mixed-resolution ADCs [7]	Pilot reduction [6], mobility environment [8]	Channel calibration [11]
Channel feedback	Low-overhead feedback [12]	Low-overhead feedback [13]	Channel prediction [14, 15], low-overhead feedback [1]

TABLE 1. Works in DL-based CSI acquisition.

back overhead in massive MIMO systems. Under the structure of autoencoder, CsiNet uses the encoder for CSI compression and decoder for CSI reconstruction. CNN is used in both encoder and decoder to exploit the spatial and spectral correlation, in a similar way to image processing. Owing to the path delay range limitation among multiple paths and spectral correlation, CSI matrix can be transformed into the delay domain and truncated to slots that are with distinct non-zero values first, which can reduce the model size and complexity. The CSI matrix is then separated to real and imaginary parts, which correspond to the two sets of input in this neural network. CsiNet shows substantial performance gain and efficiency over some compressive sensing methods.

To further improve the feedback efficiency, temporal correlation can be utilized by decreasing the redundancy in the adjacent feedback codewords. In [13], a DL network named CsiNet-LSTM has been designed for time-varying massive MIMO channels. To exploit temporal channel correlation, as shown in Fig. 3b, the codewords at t1 are utilized as an auxiliary input of the decoder networks for the later moments to eliminate the corresponding information in the new codewords, and LSTM networks are adopted to process sequential CSI matrices for extracting the temporal relationship therein. Using these two ways, the redundancy in codewords owing to the temporal correlation can be degraded. For CsiNet-LSTM, only the first MIMO CSI matrix of the time sequence is compressed under a moderate compression ratio (CR) and reconstructed by CsiNet. The ensuing CSI matrices are encoded at a high compression ratio owing to the information saving from previous correlated CSIs. Consequently, CsiNet-LSTM can achieve a better compression ratio and reduced average feedback payload.

Similar to the temporal correlation, the correlated uplink CSI due to the bi-directional correlation can also be used to enhance the downlink CSI feedback efficiency. In [1], we proposed a DL-based CSI feedback framework to exploit bi-directional channel correlation characteristics. Unlike CsiNet and CsiNet-LSTM, DualNet in [1] exploits the available uplink CSI at gNB as an auxiliary input for the decoder network to help estimate the downlink CSI from UE feedback in massive MIMO systems. We designed two DL architectures, DualNet-MAG and DualNet-ABS, to improve the UE feedback accuracy. DualNet-MAG and DualNet-ABS can utilize the bi-directional channel correlation of the magnitude and the absolute value of the CSI coefficients, respectively. As shown in Fig. 3c, the decoder in DualNet-MAG reconstructs the downlink CSI magnitudes based on the uplink CSI magnitudes and its received UE feedback codewords. We further developed a magnitude-dependent phase quantization method to reduce the UE phase feedback overhead. Our work in [1] shows better performance by DualNet than the DL architecture CsiNet relying only on UE feedback.

We can further investigate the influence of bi-directional channel band gap and RF channel bandwidth on the performance of DL-based CSI feedback in FDD systems. We test the performance of CsiNet, DualNet-MAG, and DualNet-ABS in different bandwidths and band gaps. The central downlink frequencies are set to 5.3 GHz and 930 MHz for indoor and outdoor scenarios, respectively.

For the indoor scenario, the band gap of 180 MHz and bandwidth of 20 MHz are selected as the baselines for comparison. For the outdoor scenario, the band gap of 75 MHz and bandwidth of 5 MHz are set as the baseline. We compare the downlink CSI estimates under feedback compression ratios of 1/8, 1/12, and 1/16. A smaller ratio implies higher compression in CSI feedback and is more efficient. To test the bandwidth effect, we double the channel bandwidth by maintaining the same band gap and test the CSI performance. To test the band gap effect, we reduce the band gap to 0 without changing the bandwidth. As shown in Fig. 4, DualNet-ABS and DualNet-MAG can improve the downlink CSI recovery accuracy by reducing the band gap and/or increasing the bandwidth.

In addition to compressing the CSI feedback payload on uplink, an alternative is to estimate the downlink channel without downlink pilots and uplink feedback by utilizing the reciprocity between uplink and downlink channels (mostly for TDD links). To exploit this bi-directional correlation in FDD, DL networks can be trained to approximate an uplink-to-downlink channel mapping function, which attempts to extract propagation information from uplink CSI to a latent domain before transferring the derived information from the latent domain to estimate the downlink channel. The authors of [14] trained such CNN to estimate downlink CSI based on the CSI from multiple adjacent uplink subcarriers in a single-antenna FDD system. Additionally, [14] also utilized temporal correlation to predict downlink CSI in future time slots. Another work [15] trained a fully connected DNN to predict FDD downlink CSI based on uplink CSI in single-carrier massive MIMO systems.

Exploiting physical CSI correlations with DL-based CSI feedback can substantially reduce the feedback payload. On the one hand, the resulting channel sparsity can be utilized in training and data preprocessing to lower model complexity and increase feedback efficiency. For example, CSI matrices can be transformed from the frequency domain to the delay domain to reduce the size of the input and corresponding DL model [12]. On the other hand, temporal and bi-directional CSI estimates can serve as an auxiliary input to reduce the amount of CSI feedback [1, 13]. In fact, bi-directional CSI correlation allows on the gNB transmitter

For CsiNet-LSTM, only the first MIMO CSI matrix of the time sequence is compressed under a moderate compression ratio (CR) and reconstructed by CsiNet. The ensuing CSI matrices are encoded at a high compression ratio owing to the information saving from previous correlated CSIs. Consequently, CsiNet-LSTM can achieve a better compression ratio and reduced average feedback payload.

Despite a number of preliminary successes, the high complexity of DL models and the diverse characteristics of practical wireless channels make massive MIMO downlink CSI estimation and prediction highly challenging. The design of appropriate DL architecture of sufficiently low complexity for wideband CSI estimation in massive MIMO remains an exciting and open problem.

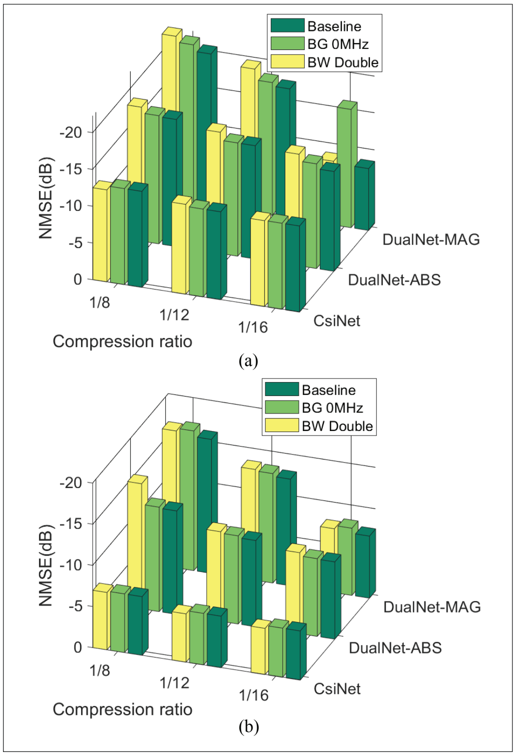

FIGURE 4. CSI feedback performance under the influence of band gap (BG) and bandwidth (BW): a) indoor; b) outdoor.

to learn the mapping function between the naturally available uplink CSI to directly estimate the downlink CSI in both FDD and TDD links. Despite a number of preliminary successes, the high complexity of DL models and the diverse characteristics of practical wireless channels make massive MIMO downlink CSI estimation and prediction highly challenging. The design of appropriate DL architecture of sufficiently low complexity for wideband CSI estimation in massive MIMO remains an exciting and open problem.

Open Issues

We have demonstrated the benefits of integrating DL to exploit inherent channel correlations for downlink CSI estimation in massive MIMO communications. We now discuss several research directions to further improve DL-based CSI estimation for future wireless networks.

DL Under Different System Settings: In wireless communications, various system parameters including antenna settings and spectral settings can potentially affect the efficacy of DL-based CSI methods. As DL techniques are poised to play a more prominent role in future wireless systems, one crucial work is to assess how effective DL-based architectures can be under different system settings. Investigating the impact of system parameters on DL-based solutions can provide better design insights for future development.

Specialized DL Models Combined with Channel Features: More specialized DL architectures for wireless communications are still under study and should be developed to incorporate wireless channel features and other domain knowledges. Most existing DL networks in the physical layer of wireless communications directly adopt known DL architectures or algorithms. Application-specific architectures for wireless networks would depend on channel models, signal features, redundancy, and other domain-specific knowledge.

As an example of combining bi-directional CSI correlation to infer the downlink CSI using uplink CSI, three architectures are tested.[1] CSI matrices are transformed to the delay domain from the frequency domain first, and then the small fraction of large components are kept to reduce the model complexity. Unlike DualNet-ABS and DualNet-MAG, U2D-ABS and U2D-MAG respectively recover the absolute values and magnitudes of downlink CSI from the corresponding uplink CSI directly without feedback. U2D-ORG divides the downlink CSI into real and imaginary parts without separating their signs as the DL network input. Using the data set from above, we consider perfect knowledge of phases and signs of downlink CSI at gNB. The performance of Fig. 5 shows improved downlink CSI accuracy from U2D-ABS and U2D-MAG for reduced band gap and increased bandwidth. U2D-MAG is superior in all cases. By reducing the feedback payload, U2D-ABS and U2D-MAG show more considerable promise for better downlink CSI estimation efficiency.

Low Complexity Channel Estimation: Existing works and designs have demonstrated the power of integrating channel correlations with DL in massive MIMO channel estimation of FDD systems by reducing the cost in terms of training and UE side feedback payload. These results and new findings are expected to impact the future development of massive wireless communications. On the one hand, exploiting channel correlations requires more inputs and parameters in the underlying DL architecture, particularly with respect to time-varying channels. Consequently, there is a price to pay in terms of the DL model complexity and training time. To limit model complexity without sacrificing CSI estimation accuracy, experienced engineers should explore other physical insights and investigate effective DL methods by leveraging domain knowledge. We also note that individual UE node, with limited battery capacity and computation power, should only be with lower complexity DL tasks. Specifically, the encoder module at UE for the CSI feedback must be kept simple.

Transfer Learning-Based Channel Estimation: Existing works related to massive MIMO channel estimation can achieve satisfactory performance under a given environment, but it is hard to adapt to channels in a new environment with different correlations directly. By avoiding model learning from scratch, transfer learning can often simplify training in a new configuration and allows DL networks to achieve good CSI estimates even without access or time to process large volumes of training data. Practically, there are many active UEs and gNBs in wireless networks. Therefore, by leveraging prior knowledge, transfer learning is a potentially valuable direction in practical implementation of DL in massive MIMO systems.

Unsupervised Learning-Based Channel Estimation: Most DL-based channel estimation methods require ground truth data to train the mapping from measured pilot signals to CSI. However, ground truth channel data is not always available in the practical deployment of massive MIMO networks. To eliminate the requirement of

[1] Details and related codes are provided on: https://ieee-collabratec.ieee.org/app/workspaces/6247/activities.

labeled CSI or ground truth CSI, deep unfolding networks which can reconstruct the signal from noisy measurements can be used for massive MIMO channel estimation. Similar to the structure of RNN, the deep unfolding network takes an iterative CS algorithm with limited iterations, unfolds its structure, and introduces a number of trainable parameters. Unlike the conventional CS algorithms, deep unfolding networks experience much less delay with the help of a graphics processing unit (GPU) and higher recovery accuracy by adjusting the trainable parameters.

Conclusion

DL has recently emerged as an exciting design tool in developing future wireless communication systems. In this article, we introduce the basic principles of applying DL for improving RF wireless network performance through the integration of underlying physical channel characteristics in practical massive MIMO deployment. We provide important insights into how DL benefits from physical RF channel properties and present a comprehensive overview on the application of DL for accurately estimating CSI in massive MIMO communications. We provide examples of successful DL application in CSI estimation and feedback for massive MIMO wireless systems and outline several promising directions for future research.

Acknowledgments

The work of Z. Liu and L. Zhang was supported in part by the State Major Science and Technology Special Projects under Grant 2018ZX03001024; and in part by the China Scholarship Council. The work of Z. Ding was supported by the National Science Foundation under Grant 2029027.

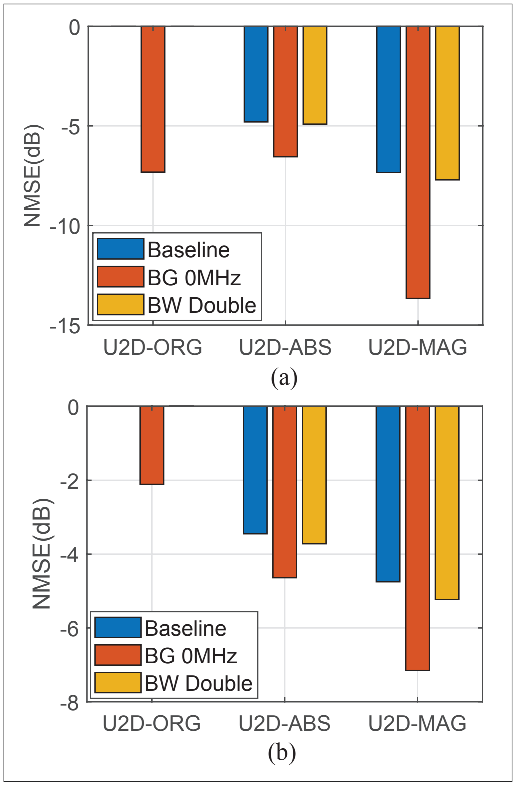

FIGURE 5. CSI recovery performance comparison in different band gap (BG) and bandwidth (BW): a) indoor; b) outdoor.

To limit model complexity without sacrificing CSI estimation accuracy, experienced engineers should explore other physical insights and investigate effective DL methods by leveraging domain knowledge. We also note that individual UE node, with limited battery capacity and computation power, should only be with lower complexity DL tasks.

References

[1] Z. Liu, L. Zhang, and Z. Ding, "Exploiting Bi-Directional Channel Reciprocity in Deep Learning for Low Rate Massive MIMO CSI Feedback," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, June 2019, pp. 889–92.

[2] Z. Gao *et al.*, "Structured Compressive Sensing-Based Spatio-Temporal Joint Channel Estimation for FDD Massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, Feb. 2016, pp. 601–17.

[3] K. Hugl, K. Kimmo, and J. Laurila, "Spatial Reciprocity of Uplink and Downlink Radio Channels in FDD Systems," *COST 273 TD 066*, Dec. 2002, pp. 1–6.

[4] Z. Zhong, L. Fan, and S. Ge, "FDD Massive MIMO Uplink and Downlink Channel Reciprocity Properties: Full or Partial Reciprocity?" arXiv preprint arXiv:1912.11221, 2019.

[5] L. Liu *et al.*, "The Cost 2100 MIMO Channel Model," *IEEE Wireless Commun.*, vol. 19, no. 6, Dec. 2012, pp. 92–99.

[6] P. Dong *et al.*, "Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, Sept. 2019, pp. 989–1000.

[7] S. Gao *et al.*, "Deep Learning Based Channel Estimation for Massive MIMO with Mixed-Resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, Nov. 2019, pp. 1989–93.

[8] Y. Liao, Y. Hua, and Y. Cai, "Deep Learning Based Channel Estimation Algorithm for Fast Time-Varying MIMO-OFDM Systems," *IEEE Commun. Lett.*, vol. 24, no. 3, Mar. 2020, pp. 572–76.

[9] E. Balevi, A. Doshi, and J. G. Andrews, "Massive MIMO Channel Estimation with an Untrained Deep Neural Network," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, Mar. 2020, pp. 2079–90.

[10] H. He *et al.*, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, Oct. 2018, pp. 852–55.

[11] C. Huang *et al.*, "Deep Learning for UL/DL Channel Calibration in Generic Massive MIMO Systems," *Proc. 2019 IEEE ICC*, May 2019, pp. 1–6.

[12] C. Wen, W. Shih, and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, Oct. 2018, pp. 748–51.

[13] T. Wang *et al.*, "Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, Apr. 2019, pp. 416–19.

[14] M. S. Safari, V. Pourahmadi, and S. Sodagari, "Deep UL2DL: Datadriven Channel Knowledge Transfer from Uplink to Downlink," *IEEE Open J. Veh. Technol.*, vol. 1, 2020, pp. 29–44.

[15] Y. Yang *et al.*, "Deep Learning-Based Downlink Channel Prediction for FDD Massive MIMO System," *IEEE Commun. Lett.*, vol. 23, no. 11, Nov. 2019, pp. 1994–98.

Biographies

Zhenyu Liu (lzyu@bupt.edu.cn) is working toward the Ph.D. degree at Beijing University of Posts and Telecommunications. From 2017 to 2019, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, University of California, Davis. His research interests include deep learning in wireless communications and vehicular communications.

Lin Zhang (zhanglin@bupt.edu.cn) is a professor at Beijing University of Posts and Telecommunications. He has authored more than 120 papers in referenced journals and international conferences. His current research interests include ultra-wideband bio-radar imaging and vital signal detection, AI driven information processing and Internet of Vehicles.

Zhi Ding (zding@ucdavis.edu) is a distinguished professor at the University of California, Davis. He is a Fellow of IEEE and has been an active member of IEEE, serving on technical programs of several workshops and conferences. He has published over 400 refereed papers and book chapters. His current research interests include signal processing and machine learning in wireless communications.