Enabling Feedback-Free MIMO Transmission for FD-RAN: A Data-Driven Approach

Jingbo Liu, Jiacheng Chen, *Member, IEEE*, Zongxi Liu, *Student Member, IEEE*, and Haibo Zhou, *Senior Member, IEEE*

***Abstract*—To enhance flexibility and facilitate resource cooperation, a novel fully-decoupled radio access network (FD-RAN) architecture is proposed for 6G. However, the decoupling of uplink (UL) and downlink (DL) in FD-RAN makes the existing feedback mechanism ineffective. To this end, we propose an end-to-end data-driven MIMO solution without the conventional channel feedback procedure. Data-driven MIMO can alleviate the drawbacks of feedback including overheads and delay, and can provide customized precoding design for different BSs based on their historical channel data. It essentially learns a mapping from geolocation to MIMO transmission parameters. We first present a codebook-based approach, which selects transmission parameters from the statistics of discrete channel state information (CSI) values and utilizes nearest neighbor interpolation for spatial inference. We further present a non-codebook-based approach, which 1) derives the optimal precoder from the singular value decomposition (SVD) of the channel; 2) utilizes variational autoencoder (VAE) to select the representative precoder from the latent Gaussian representations; and 3) exploits Gaussian process regression (GPR) to predict unknown precoders in the space domain. Extensive simulations are performed on a link-level 5G simulator using realistic ray-tracing channel data. The results demonstrate the effectiveness of data-driven MIMO, showcasing its potential for application in FD-RAN and 6G.**

***Index Terms*—6G, fully-decoupled RAN, MIMO, codebook, variational autoencoder.**

I. Introduction

COMMERCIALIZATION of the fifth generation (5G) mobile network is ongoing worldwide, yet some fundamental challenges have not been fully addressed. First, the scarce spectrum resources are still not efficiently utilized. Second, it is still difficult to guarantee users' quality of experience (QoE). Third, the costs of network infrastructure and operation are very high. To address the above issues, the fully-decoupled radio access network (FD-RAN) architecture is proposed for the upcoming sixth generation (6G) mobile network [1]. It decouples the base stations (BSs) into control BSs (C-BSs) and data BSs, while data BSs are further decoupled into downlink BSs (DL-BSs) and uplink BSs (UL-BSs) [2], [3], [4]. With control, UL, and DL fully separated, control and resource allocation flexibility can be exploited to meet the requirements of 6G [5], [6], [7], [8].

However, such a decoupled architecture makes the existing feedback schemes employed in 5G ineffective. In 5G, take DL transmission as an example, BS first transmits channel state information reference signals (CSI-RSs) to user equipment (UE). Then, UE estimates the channel and calculates the transmission parameters, which are fed back to the BS for carrying out multiple-input and multiple-output (MIMO) transmission. In FD-RAN, DL-BS is unable to obtain the CSI feedback from UE directly, since UL-BSs and DL-BSs are physically separated. It is also not efficient to use the C-BS for CSI feedback, as the resource consumption for all UEs is high, and the additional delay incurred by C-BS will make the feedback inaccurate. Thus, how to implement MIMO transmission without CSI feedback is a significant challenge for FD-RAN.

On the other hand, the state-of-the-art 5G MIMO techniques also suffer from issues caused by feedback. 5G adopts closed-loop spatial multiplexing (CLSM) as its main MIMO scheme [9], due to its higher capacity than open-loop spatial multiplexing (OLSM) and transmit diversity, as well as its ability to exploit massive antennas. Specifically, the CSI feedback includes precoder, the number of spatial streams, and modulation and coding scheme (MCS), which are necessary transmission parameters for MIMO. Precoding in MIMO systems is a technique where the transmitted signal at each antenna is pre-adjusted using CSI to maximize transmission efficiency. It enhances signal strength, reduces interference, and increases data throughput, crucial for achieving high spectral efficiency and robustness in wireless communications like 5G. However, when the number of transmit antenna increases, the overheads of pilots, namely CSI-RSs, and CSI feedback become much higher, reducing the actual data rate. The overheads are determined by the chosen transmission mode and the number of transmit antennas. In 5G CLSM,

Received 19 May 2024; revised 13 September 2024; accepted 5 November 2024. Date of publication 11 November 2024; date of current version 5 February 2025. This work was supported in part by the National Natural Science Foundation Original Exploration Project of China under Grant 62250004, in part by the Natural Science Foundation of China (NSFC) under Grant 62271244, in part by the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067, in part by the High-level Innovation and Entrepreneurship Talent Introduction Program Team of Jiangsu Province under Grant JSSCTD202202, and in part by the Major Key Project of PCL under Grant PCL2024A03. Recommended for acceptance by M. Chen. *(Corresponding authors: Jiacheng Chen; Haibo Zhou.)*

Jingbo Liu is with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with Pengcheng Laboratory, Shenzhen 518000, China (e-mail: liujingbo@sjtu.edu.cn).

Jiacheng Chen is with the Department of Strategic and Advanced Interdisciplinary Research, Pengcheng Laboratory, Shenzhen 518000, China (e-mail: chenjch02@pcl.ac.cn).

Zongxi Liu and Haibo Zhou are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: zongxiliu@smail.nju.edu.cn; haibozhou@nju.edu.cn).

Digital Object Identifier 10.1109/TMC.2024.3495719

to maintain orthogonality, different CSI-RSs are required to estimate the channels at different transmit antennas, causing the number of CSI-RSs to increase linearly with the number of transmit antennas. Additionally, the number of feedback bits scales linearly with the number of transmit antennas N_{tx} and is calculated as $(N_{tx}-1)\times \mathrm{SNR}/3$ [10], where SNR denotes the signal-to-noise ratio. Also, feedback delay will affect the transmission performance, especially when the channel propagation environment changes fast.

To this end, we propose a totally different solution, namely data-driven MIMO. Instead of relying on feedback to acquire necessary channel information for MIMO, we exploit the knowledge from historical channel data and only use UE's geolocation to determine transmission parameters. Intuitively, data-driven MIMO is feasible because the channel and the corresponding transmission parameters are strongly correlated with geolocation since the signal propagation environment at a certain location will not dramatically change. Although the channel is generally time-varying, we expect performance loss can be mitigated by saving overheads of pilots and CSI feedback. Data-driven MIMO falls in the broader field of zero-feedback transmission, which has been investigated in literature [11], [12], [13].

In this paper, we attempt to realize the same capability as CLSM, namely MIMO spatial multiplexing. We adopt an end-to-end approach, so we need to construct a mapping from location to transmission parameters. Since we do not have any channel feedback, the mapping is fixed, or we can say it is static in the time domain. Given that the mapping can only be trained with existing channel data, we need to find the transmission parameters of all the possible locations. Then, inference in the space domain is required. It is also similar for the frequency domain if multiple subcarriers are considered.

Data-driven MIMO is very challenging since the performance loss incurred by fixed transmission parameters is definite and should be reduced as much as possible, to make the approach meaningful. One major challenge is to deal with the channel characteristics in time, frequency, and space domains. Specifically, in time domain we need to find the "representative" transmission parameters considering the channel fluctuations, and in space and frequency domains we need to infer the transmission parameters considering the channel correlations. Another key challenge is how to jointly choose transmission parameters, namely precoder, rank indicator (RI), and channel quality indicator (CQI). Generally, the precoder and RI are determined at the same time, since RI indicates the rank of the precoder. Then, proper CQI can be derived based on the chosen precoder and RI. Any improper choice of these three parameters would lead to performance degradation, especially considering RI and CQI are quantized discrete values. On the one hand, low RI and CQI values will lead to loss in transmission performance. On the other hand, high RI values will bring more interference among data streams while high CQI values can lead to increase in bit error ratio (BER), both of which will lower the performance. Existing works mainly concentrate on beamforming and use only one stream. [14] studies the beamforming problem without feedback, and [15] proposes a deep learning framework for beam selection. Different from them, we consider multi-stream transmission and jointly select the three interdependent transmission parameters, as in the current feedback-based transmission system.

To address the above challenges, we propose two approaches to realize data-driven MIMO for FD-RAN DL transmission. The first approach is based on 5G Type I codebook [16], which basically quantizes the precoder and assigns each precoder with a precoding matrix indicator (PMI). Thus, all the transmission parameters are discrete values and we can use a statistic-based method to choose the fixed transmission parameters and further utilize integer interpolation for spatial prediction. However, the quantization of the codebook also leads to the inaccuracy of the precoder. Based on the fact that singular value decomposition (SVD) of the channel matrix can optimally calculate the precoder, we further consider a second approach that starts processing from the optimal precoders. To fix the transmission parameters in the time domain, we exploit variational autoencoder (VAE) [17] to discover the latent representations of optimal precoders in lower dimensions. The encoder of VAE returns a distribution over the latent space, and close sample points are expected to have similar representations in the latent space, which makes the results of the generative process of decoder more meaningful. Given the latent representations are Gaussian variables, we further utilize Gaussian process regression (GPR) for spatial prediction. Note that in this paper, we consider transmission on narrow bands, and do not infer in the frequency domain, thus the same transmission parameters are applied to all subcarriers. To summarize, the contributions of this paper are listed as follows:

- We propose a novel channel feedback free solution to realize the advanced MIMO spatial multiplexing for both FD-RAN and future 6G. The proposed data-driven MIMO learns from the historical channel data and only utilizes geolocation to determine the transmission in an end-to-end way. The data-driven approach essentially customizes the MIMO transmission at each BS. To the best of our knowledge, we are the first to study feedback-free multi-stream transmission, which needs to jointly determine three corresponding transmission parameters, as in the current feedback-based transmission system.
- We propose an easy-to-implement codebook-based approach for data-driven MIMO. Transmission parameters are selected using statistics of CSI values, and integer interpolation is used to handle the correlation of channel in the space domain.
- We further propose an SVD-based approach for performance improvement. VAE is used to select the representative precoder in time domain from latent Gaussian representations. Then, GPR is used to predict the unknown precoders in space domain. Natural neighbor interpolation (NNI) is utilized to infer CQIs at unknown locations.
- We perform extensive simulations based on a 5G-compatible link-level simulator and realistic ray-tracing channel data. Simulation results demonstrate the effectiveness of the proposed two approaches, and also the superiority of the SVD-based approach in terms of spatial inference. Compared with state-of-the-art 5G CLSM, we demonstrate

the potential of data-driven MIMO, considering the saved overheads of pilots and feedback.

The remainder of this paper is organized as follows. Related works are introduced in Section II. The system model and problem formulation are given in Section III. The codebook-based approach and SVD-based approach are presented in Sections IV and V, respectively. Simulation results and discussions are given in Section VI. Finally, Section VII concludes the whole paper.

II. Related Works

Due to heavy overheads of feedback in massive MIMO, a lot of efforts have been made to reduce the overheads to unleash the potential of 5G [18], [19], [20], [21], [22], [23], [24]. A comprehensive review of deep learning-based CSI feedback schemes is presented in [18]. In [19], a CSI compressing and recovering mechanism called CsiNet is first proposed for reducing the CSI feedback. It treats the channel in angle and delay domains as an image, and exploits an encoder to compress the channel at UE side and a decoder to recover it at BS side. To deal with the time-varying characteristics of channel, long short-term memory (LSTM) is employed for CsiNet [20], which can achieve a better tradeoff between complexity and compression ratio. Then, convolutional neural networks are further used for compressing CSI [21], with quantization of it taken into consideration. In [22], a novel channel reconstruction network (CRNet) based on inception block is proposed. It aims to extract multi-resolution CSI features and includes an advanced training scheme to improve performance. To cope with the existing feedback mechanism based on codebook, [23] proposes a bidirectional LSTM-based CSI feedback architecture to achieve a fine mapping between the precoder and corresponding PMI, where Type I and Type II codebooks of 5G are taken into account. To reduce the complexity of optimal neural network (NN) design for CSI feedback, [24] proposes Auto-CsiNet, which uses neural architecture search (NAS) to automatically generate the NN architecture. This approach leverages implicit scene knowledge to tailor the network to specific scenarios in a data-driven manner, resulting in performance that surpasses that of manually designed models. Different from the works mentioned above, we study the data-driven MIMO that requires no channel feedback at all.

Furthermore, a number of studies have been made to eliminate the channel feedback [25], [26], [27], [28], [29], [30], [31], [32], [33]. In [25], a system called R2-F2 that enables BS to predict the downlink channel based on the estimated uplink channel is introduced. R2-F2 essentially builds up a channel-to-path transform that captures the underlying physical paths to infer the channels at different frequency bands, which provides a 0.7 dB of beamforming gain. The problem of channel mapping in space and frequency domains is studied in [26]. Utilizing a fully connected network architecture, the channel at a certain location and frequency band can be mapped into the one at a different location and frequency band. In [27], a radio environment map (REM) of one BS is built on the measurements from all grids of beams. Then, the reference signal received powers utilizing different beamforming schemes can be derived based on REMs, facilitating the interference coordination between the serving and neighboring cells. The problem of joint beamforming, power control, and interference coordination is considered in [28]. A deep reinforcement learning architecture is proposed at both sub-6 GHz and millimeter wave frequency bands, where UE merely sends back its received signal to interference plus noise ratio (SINR) and coordinates to BS. In [29], the authors propose a destination-feedback-free distributed beamforming scheme, where location information is required. One radio acts as a leader and guides the others to point toward the destination. [30] introduces a deep learning-based radio map that utilizes the location information of UE to determine the beamforming vector. The neural network's loss function can be tailored to accommodate various communication needs. In [31], the authors propose a radio map-based complex-valued precoding network (RMCPNet) to facilitate feedback-free transmission in FD-RAN. The model not only utilizes location information but also incorporates statistical channel information at the edge cloud, such as the transmit or receive correlation matrix and the channel energy matrix, to enhance transmission performance. The concept of the channel knowledge map (CKM) is first introduced in [32] to facilitate environment-aware communications. CKM utilizes the locations of transmitters and receivers to provide additional channel knowledge and assist the transmission. A practical example of this idea is beam selection which uses the channel path map (CPM) to reconstruct the channel from the three most significant paths. For spatial interpolation, CPM employs the inverse distance weighting (IDW) method of the three nearest neighbors to interpolate the channel at locations without collected data. In [33], FIRE, an end-to-end approach empowered by VAE, is proposed to obtain the downlink channel in frequency division duplex (FDD) system without any feedback from UE. Instead of recovering the inputs like traditional VAE, FIRE learns the mapping between the uplink channel and the corresponding downlink channel, thereby enabling the acquisition of the downlink channel at BS. Different from the above works, we aim to realize MIMO spatial multiplexing and determine all three essential MIMO transmission parameters. Additionally, we exploit VAE to reduce the dimension of precoders to find the representative precoder in time domain and generate meaningful precoders for spatial inference.

III. System Model and Problem Formulation

In this section, we first introduce the system model of DL MIMO transmission in FD-RAN. Then, we analyze and formulate the problem of data-driven MIMO.

A. System Model

1) Architecture and Operation of FD-RAN: The FD-RAN architecture decouples BS functions into C-BS, UL-BS, and DL-BS. The C-BS manages control signaling over a large coverage area through the bi-directional control link, while the UL-BS and DL-BS are dedicated to signal receiving and transmitting, respectively. All three units are connected to the edge cloud via high-capacity, low-latency optical fibers. However, the DL-BS cannot directly obtain CSI feedback from the UE due to

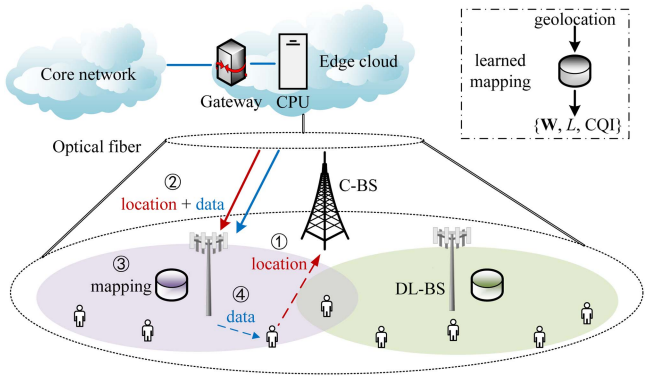


Fig. 1. The feedback-free downlink MIMO transmission in FD-RAN.

the physical separation of UL-BSs and DL-BSs in FD-RAN. Additionally, using the C-BS for CSI feedback is inefficient due to the high resource consumption for all UEs and the extra delay introduced by the C-BS. Therefore, one of the key challenges in FD-RAN is how to realize feedback-free MIMO transmission.

In this paper, we study the DL single-user MIMO (SU-MIMO) transmission of FD-RAN. For a DL-BS, the mapping from any location within its coverage area to the appropriate transmission parameters needs to be initially generated at the edge cloud and then stored at the DL-BS. Fig. 1 illustrates the system model, where a C-BS coordinates with multiple DL-BSs and UEs. The C-BS obtains the locations reported by users and informs the DL-BS. Then, from the mapping, DL-BS can find out transmission parameters and carry out MIMO transmission based on UE's geolocation. Although the feedback of location information through C-BS also has delay, the influence on the transmission performance is limited. On the one hand, the user's location will not significantly change in seconds in typical scenarios. Also, the fixed transmission parameters will not change dramatically at close locations. On the other hand, the user's location is usually easy to predict when it changes fast, e.g., when the user is inside a car or high-speed train. Meanwhile, integrated sensing and communication (ISAC) [34] can be utilized to sense the location of UE. In this paper, we only focus on the performance at static user locations. Meanwhile, this work can be extended to multi-user cases by allocating orthogonal subcarriers to different users, a common approach of SU-MIMO transmission mode for multi-user scenarios.

2) Downlink MIMO Transmission: The whole signal processing procedure is based on orthogonal frequency division multiplexing (OFDM) and MIMO technologies. OFDM is a multicarrier modulation mode employed in 4G long-term evolution (LTE) and is still a major modulation scheme in 5G new radio (NR). Leveraging the discrete Fourier transform (DFT), OFDM transforms the frequency selective broadband channel into K flat narrowband channels, and $k \in \mathcal{K} = \{1, \ldots, K\}$ refers to the k-th orthogonal subcarrier. The DL-BS and UE are assumed to be equipped with N_{tx} transmit antennas and N_{rx} receive antennas, respectively. Then the received symbol vector $\boldsymbol{c}_k \in \mathbb{C}^{N_{rx}\times 1}$ on the subcarrier k at any sampling time instant can be expressed as

$$\boldsymbol{c}_k = \mathbf{H}_k \mathbf{W} \boldsymbol{s}_k + \boldsymbol{n}_k, \quad k \in \mathcal{K}, \tag{1}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_{rx}\times N_{tx}}$ is the channel matrix that the transmitted symbol vector $\boldsymbol{s}_k \in \mathbb{C}^{L\times 1}$ with normalized unit power experiences on the subcarrier k. $\boldsymbol{n}_k \sim \mathcal{CN}(\mathbf{0}, \varsigma_n^2 \mathbf{I}_{N_{rx}})$ is the additive zero means complex-valued white Gaussian noise with variance ς_n^2, where $\mathbf{I}_{N_{rx}}$ is the identity matrix of dimension $N_{rx} \times N_{rx}$. $\mathbf{W} \in \mathbb{C}^{N_{tx}\times L}$ is the employed precoding matrix and L denotes the number of layers employed for spatial multiplexing, where $L \in \{1, \ldots, \Upsilon\}$. The maximum value of L is defined by $\Upsilon = \min\{N_{rx}, N_{tx}\}$, which represents the minimum of the number of receive and transmit antennas. Note that since we do not infer in the frequency domain, the same precoder is used for all subcarriers. The method of determining the appropriate precoder will be introduced in the following sections. RI corresponds to the rank of precoder $\mathbf{W}$, which equals L.

At the receiver side, $\mathbf{H}_k\mathbf{W}$ and noise variance $\boldsymbol{n}_k$ are estimated by demodulation reference signals (DM-RSs). Then, the received symbol vector $\boldsymbol{c}_k$ is equalized by an equalizer, which is denoted by $\mathbf{E}_k \in \mathbb{C}^{L\times N_{rx}}$. The post-equalization received symbol vector $\boldsymbol{m}_k \in \mathbb{C}^{L\times 1}$ is written as

$$\boldsymbol{m}_k = \mathbf{E}_k \boldsymbol{c}_k = \mathbf{E}_k \mathbf{H}_k \mathbf{W} \boldsymbol{s}_k + \mathbf{E}_k \boldsymbol{n}_k, \quad k \in \mathcal{K}. \tag{2}$$

We denote the multiplication of three matrices $\mathbf{E}_k\mathbf{H}_k\mathbf{W}$ as $\mathbf{G}_k \in \mathbb{C}^{L\times L}$, which can be seen as the equivalent channel between transmitter and receiver. Then SINR of ℓ-th layer after equalization is given as

$$\mathrm{SINR}_{k,\ell} = \frac{\|[\mathbf{G}_k]_{\ell,\ell}\|_F^2}{\sum_{i=1,i\neq\ell}^{L} \|[\mathbf{G}_k]_{\ell,i}\|_F^2 + \varsigma_n^2 \sum_{i=1}^{N_{rx}} \|[\mathbf{E}_k]_{\ell,i}\|_F^2}, \tag{3}$$

where $k \in \mathcal{K}, \ell \in \mathcal{L} = \{1, \ldots, L\}$ and $\|\cdot\|_F$ denotes the Frobenius norm. $[\mathbf{G}_k]_{\ell,i}$ denotes the element of the equivalent channel matrix $\mathbf{G}_k$ that lies in the ℓ-th row and i-th column. Thus, the numerator of (3) represents the desired signal power of layer ℓ on subcarrier k. The first term of the denominator refers to the inter-layer interference and the second term is the enhanced noise.

Then, the post-equalization SINRs over all subcarriers $\mathcal{K}$ and all layers $\mathcal{L}$ are mapped into one effective SNR of an equivalent single input and single output (SISO) system with AWGN channel[35], which has similar transmission performance with the MIMO OFDM system. The mapping is defined as

$$\mathrm{SNR}_{\mathrm{eff}} = \alpha f^{-1}\left(\frac{1}{KL}\sum_{k\in\mathcal{K}}\sum_{\ell\in\mathcal{L}} f\left(\frac{\mathrm{SINR}_{k,\ell}}{\alpha}\right)\right), \tag{4}$$

where f denotes the mapping function and f^{-1} is its inverse. Here, the mutual information effective SNR mapping (MIESM) is adopted, and f is the bit interleaved coded modulation (BICM) capacity [36]. α is the adjustment factor that adapts the block error rate (BLER) performance of the equivalent channel to approximate that of the original channel as much as possible. At last, the CQI value is chosen to be the maximal MCS value that keeps BLER lower than 0.1. A high MCS value means that the high-order modulation and high coding rate scheme are employed, which brings high spectrum efficiency and throughput.

In the multi-user MIMO (MU-MIMO) setup, the BS serves multiple UEs at the same time and frequency, thus interference among UEs should be considered. Generally, the selected UEs should be positioned orthogonally to each other so that their directional beams do not interfere. For a specific user group, each user's precoder can be first determined using CLSM as in SU-MIMO, where RI corresponds to the rank of the precoder. The precoder can either be a PMI from the codebook or an SVD precoder derived from the channel. The equivalent MU precoder for MU-MIMO transmission is then formed by combining each user's precoder and normalizing the power. Finally, the SNR_{eff} for each user is calculated by replacing $\mathbf{W}$ in the numerator of (3) with the respective user's component from the equivalent MU precoder and considering the MU interference from other users in the denominator of (3). This derived SNR_{eff} is then used to determine the CQI. With the calculated transmission parameters, the proposed approaches are feasible to implement in MU-MIMO. However, implementing MU-MIMO in practice is relatively complicated due to user pairing and grouping, along with managing MU interference in time-varying channels. For this reason, it is not considered a key scenario in this study.

B. Problem Formulation

In essence, the problem of data-driven MIMO is to estimate a mapping from geolocation to transmission parameters, i.e., $\boldsymbol{\varrho} = (x, y, z) \rightarrow \{\mathbf{W}, L, \text{CQI}\}$ based on the channel data. x, y and z are Cartesian coordinates in euclidean space. The objective is to counter performance loss caused by fixing transmission parameters in the time domain and inferring unknown transmission parameters in the space domain.

We consider the coverage area of a DL-BS and we assume that channel variations result from the dynamic small-scale environmental changes around the user. Specifically, the existing channel data is a set of channel matrices $\mathcal{H}$, where every element is a channel matrix $\mathbf{H}_{t,k,q}$ with dimension $N_{rx} \times N_{tx}$. $t \in \mathcal{T} = \{1, \ldots, T\}$ refers to index in the time domain. Note that t does not need to be an actual time point, since we will not do any inference in the time domain. Different values of t only indicate that the channel matrices are different due to the time-varying characteristics of the channel. k denotes the k-th subcarrier, which relates to the channel in the frequency domain. $q \in \mathcal{Q} = \{1, \ldots, Q\}$ is the index of location, which corresponds to the channel sample in the space domain. Therefore, $|\mathcal{H}| = T \times K \times Q$.

In data-driven MIMO, utilizing the mapping derived from existing channel data in $\mathcal{H}$, we aim to maximize the throughput at any time and location sample in the coverage area of DL-BS. The problem can be formulated over an arbitrary set of channel matrices, e.g., $\mathcal{H}'$:

$$\max_{\{\mathbf{W}_q, L_q, \text{CQI}_q\}} \sum_{t \in \mathcal{T}'} \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}'} \mathfrak{T}(\mathbf{H}_{t,k,q}, \{\mathbf{W}_q, L_q, \text{CQI}_q\})$$

$$\text{s.t.} \begin{cases} \mathbf{H}_{t,k,q} \in \mathcal{H}', & \forall t \in \mathcal{T}', \forall k \in \mathcal{K}, \forall q \in \mathcal{Q}' \\ \|\mathbf{W}_q\|_F^2 = 1, & \forall q \in \mathcal{Q}' \\ L_q = r(\mathbf{W}_q) \in \{1, \ldots, \Upsilon\}, & \forall q \in \mathcal{Q}' \\ \text{CQI}_q \in \{1, \ldots, 15\}, & \forall q \in \mathcal{Q}', \end{cases} \tag{5}$$

where $\mathfrak{T}$ denotes the throughput, and it depends on the channel and transmission parameters. Here, throughput is not calculated on the theoretical formula. $\|\mathbf{W}_q\|_F^2 = 1$ indicates the power of precoder is normalized to 1. $r(\cdot)$ refers to the rank of a matrix. CQI is an integer, ranging from 1 to 15 [16]. Since we fix the transmission parameters in time and frequency domains, we only use the subscript q to denote the transmission parameters at location q, which are the same for any time point t and subcarrier k.

Collecting the training channel may present a challenge in data-driven MIMO, while mobile crowdsensing offers a potential solution. Users can collect downlink channel data at various times and locations and upload them to the operator's edge cloud. Furthermore, operators could offer incentives to encourage users to provide channel feedback to the edge cloud. In practice, imperfect CSI can result from channel estimation error and transmission delay. The former factor is influenced by channel estimation algorithms and pilot configurations, which can be mitigated by a dedicated channel data collection process. In this paper, we assume that the channel data are accurately collected. The latter factor can significantly affect the accuracy of the measured channel. However, in our approach, channel data do not need to be time-labeled, thus our approach remains unaffected by imperfect CSI due to delay.

Although the use of location data may raise privacy concerns, our proposed data-driven approach can prioritize user privacy and data security by utilizing advanced technologies such as blockchain, homomorphic encryption, and zero-knowledge proofs to safeguard and anonymize user location data. By integrating these methods with robust encryption for data transmission and granting users control over their location data, we ensure comprehensive privacy protection and adherence to legal standards.

IV. Codebook-Based Approach

In this section, we first introduce Type I codebook used in 5G CLSM and then present the approach based on the statistic of historical CSI.

A. Codebook-Based 5G CLSM

According to 5G NR standard [16], Type I codebook is employed for SU-MIMO and varies when BS employs single-panel or multi-panel antennas. In this paper, we assume DL-BS is equipped with a single-panel antenna array of configuration (N_1, N_2) with cross-polarization, where N_1 and N_2 represent the number of horizontal and vertical antenna ports in the direction of one polarization, respectively. Thus, with two polarizations combined together, $N_{tx} = 2N_1N_2$.

Considering the tradeoff between the overheads of feedback and link performance, Type I codebook is designed in accordance with the selection of two-dimensional (2D) DFT beams. In addition to the original (N_1, N_2) spatial beams, Type I codebook oversamples the horizontal and vertical beams by factors O_1 and O_2 to obtain a finer granularity. Then, the spatial orthonormal bases of horizontal dimension $\boldsymbol{a}_1$ and vertical dimension $\boldsymbol{a}_2$ can

be expressed as

$$\boldsymbol{a}_1 = \left[1, e^{j\frac{2\pi\theta_1}{N_1O_1}}, \ldots, e^{j\frac{2\pi\theta_1(N_1-1)}{N_1O_1}}\right]^{\mathrm{T}},$$
$$\boldsymbol{a}_2 = \left[1, e^{j\frac{2\pi\theta_2}{N_2O_2}}, \ldots, e^{j\frac{2\pi\theta_2(N_2-1)}{N_2O_2}}\right]^{\mathrm{T}}, \quad (6)$$

where $\theta_1 \in \{0, 1, \ldots, N_1O_1 - 1\}$ and $\theta_2 \in \{0, 1, \ldots, N_2O_2 - 1\}$ represent the beam indexes in horizontal and vertical dimensions, respectively. $[\cdot]^{\mathrm{T}}$ denotes the transpose. Therefore, $\boldsymbol{a}_1 \in \mathbb{C}^{N_1\times 1}$ and $\boldsymbol{a}_2 \in \mathbb{C}^{N_2\times 1}$.

The supported configuration with respect to (N_1, N_2) and (O_1, O_2) can be found in [16]. Then, the set of spatial beams $\mathcal{B}$ based on 2D DFT is given by

$$\mathcal{B} = \{\boldsymbol{b}_{\theta_1,\theta_2} | \boldsymbol{b}_{\theta_1,\theta_2} = \boldsymbol{a}_1 \otimes \boldsymbol{a}_2\}, \quad (7)$$

where $\otimes$ indicates the Kronecker product, and $\boldsymbol{b}_{\theta_1,\theta_2} \in \mathbb{C}^{N_1N_2\times 1}$. Therefore, the whole candidate beam set contains $N_1N_2O_1O_2$ spatial beams. In this paper, Type I codebook in codebookmode 1 is considered, where the optimal wideband beam is selected from the candidate beam set for the whole system bandwidth.

The antenna arrays of different polarizations employ the same spatial beam, but a co-phasing factor ϕ is used to quantize the phase difference between them. Take the precoder of rank 4 in Type I codebook when the number of transmit antenna is less than 16 as an example, which can be defined as

$$\mathbf{W} = \frac{1}{\sqrt{4N_{tx}}}\begin{bmatrix} \boldsymbol{b}_{\theta_1,\theta_2} & \boldsymbol{b}_{\theta_1',\theta_2'} & \boldsymbol{b}_{\theta_1,\theta_2} & \boldsymbol{b}_{\theta_1',\theta_2'} \\ \phi\boldsymbol{b}_{\theta_1,\theta_2} & \phi\boldsymbol{b}_{\theta_1',\theta_2'} & -\phi\boldsymbol{b}_{\theta_1,\theta_2} & -\phi\boldsymbol{b}_{\theta_1',\theta_2'} \end{bmatrix}, \quad (8)$$

where the fraction is to normalize the precoder's power to 1 and the coefficient in the denominator has a 4 since $L = 4$. $\phi \in \{1, j\}$, while a phase offset $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2]$ is set to horizontal and vertical beams. Then, $\theta_1' = \theta_1 + \boldsymbol{\epsilon}_1$ and $\theta_2' = \theta_2 + \boldsymbol{\epsilon}_2$. The values of $\boldsymbol{\epsilon}$ vary in different settings of (N_1, N_2) and (O_1, O_2), which can be obtained by looking up to [16]. Hence, $\mathbf{W} \in \mathbb{C}^{N_{tx}\times L}$ depends on four variables, i.e., θ_1, θ_2, $\boldsymbol{\epsilon}$ and ϕ. Each column in $\mathbf{W}$ is orthogonal to each other.

5G supports up to 8 layers for DL transmission, thus the set of Type I codebook $\mathcal{W}_{\text{codebook}}$ contains precoders from rank 1 to 8. Clearly, more antenna ports are more likely to provide more throughput, since the beams are quantized with finer granularity to direct the energy toward the multipath of the channel. The appropriate precoder can be chosen to maximize the sum mutual information over all subcarriers $\mathcal{K}$ and all layers $\mathcal{L}$ by an exhaustive search in codebook [37]. The mutual information when precoder $\mathbf{W}$ is employed can be expressed as

$$I_{k,\ell}(\mathbf{W}) = \log_2(1 + \text{SINR}_{k,\ell}(\mathbf{W})), \quad (9)$$

where $k \in \mathcal{K}$, $\ell \in \mathcal{L}$, and $\text{SINR}_{k,\ell}(\mathbf{W})$ denotes the $\text{SINR}_{k,\ell}$ defined in (3) if precoder $\mathbf{W}$ is adopted. Then, the best precoder in the codebook at any time and location indexes is chosen as follows:

$$\mathbf{W}_{\text{PMI}^*} = \arg\max_{\mathbf{W}_{\text{PMI}}} \sum_{k\in\mathcal{K}}\sum_{\ell\in\mathcal{L}} I_{k,\ell}(\mathbf{W}_{\text{PMI}})$$
$$\text{s.t. } \mathbf{W}_{\text{PMI}} \in \mathcal{W}_{\text{codebook}}, \ \forall\ \text{PMI} \in \{1, \ldots, |\mathcal{W}_{\text{codebook}}|\}, \quad (10)$$

where PMI indicates the index of the precoder in the codebook. Then, the index of the chosen precoder is PMI^*, which is reported by UE to BS in 5G CLSM. RI is the rank of $\mathbf{W}_{\text{PMI}^*}$, i.e., $L = r(\mathbf{W}_{\text{PMI}^*})$. CQI is mapped from the effective SNR defined in (4) given the precoder and RI according to the CQI table in [16]. In 5G CLSM, with the help of CSI-RSs sent by BS, UE calculates the proper transmission parameters and reports $\{\text{PMI}^*, r(\mathbf{W}_{\text{PMI}^*}), \text{CQI}\}$ to BS for MIMO transmission.

B. Statistic-Based Solution

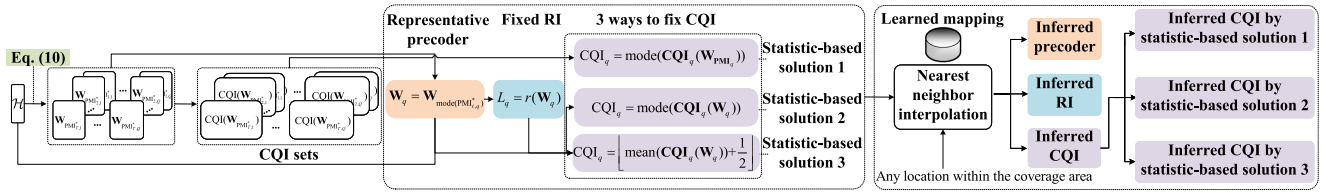
In this subsection, we first analyze how to deal with channel data in time, frequency, and space domains to obtain the mapping, and then present the statistical solution based on 5G Type I codebook for data-driven MIMO.

In time domain, it is hard to infer transmission parameters since we do not utilize any channel feedback. Fixed RI should be first determined, which indicates the rank of the precoder. Then, we need to find the "representative" channel to get the precoder or obtain the "representative" precoder directly. Since the precoder represents different spatial beams, we choose to fix it in time domain, which is more interpretable. For CQI, it can be decided as the mode or mean value of the CQI set at one location according to the fluctuations of channels in time domain.

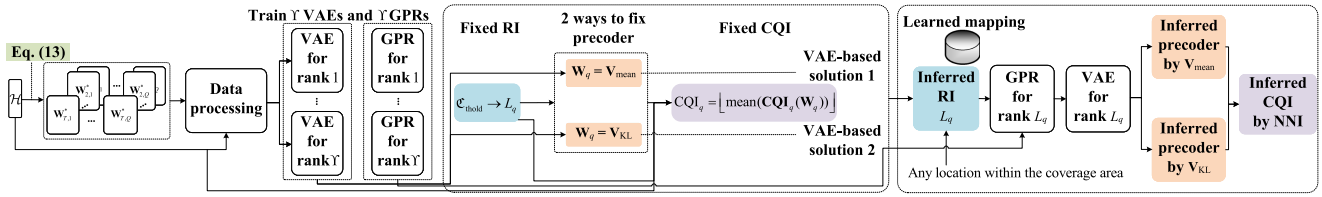
In frequency domain, channels on adjacent subcarriers will present correlation to some extent. Thus, it is proper to handle the channels on grouped adjacent subcarriers by dividing K channels into several groups, so as to obtain multiple sets of transmission parameters in frequency domain. In this paper, we consider data transmission on narrow bands and employ one set of transmission parameters, i.e., the same precoder, RI, and CQI apply to all subcarriers.

In space domain, we need to infer the transmission parameters at unknown locations in $\mathcal{H}'$. The fixed parameters in time and frequency domains derived from $\mathcal{H}$ facilitate the prediction in space domain, otherwise, we would obtain different inputs at the same location sample. Also, the channels are somewhat correlated in space due to the similar signal propagation environment. To determine the transmission parameters at unknown locations, RI should also be first chosen such that the fixed precoders derived from existing channel data can be utilized for inference. Then, the integer interpolation method can be exploited to infer CQIs at unknown locations.

For the codebook-based approach, the three transmission parameters are all quantized discrete numbers. Therefore, the statistical solution can be employed to fix transmission parameters as the values that occur most frequently from historical CSI based on the codebook of 5G, which is illustrated in Fig. 2(a). Denote $\mathbf{PMI}_q = \{\text{PMI}_{t,q}^* | t \in \mathcal{T}\}$ as the historical PMI set at location q, the fixed precoder $\mathbf{W}_q$ is then chosen as $\mathbf{W}_{\text{mode}(\mathbf{PMI}_q)}$, where $\text{mode}(\cdot)$ indicates the mode value of the set. L_q is given by the rank of $\mathbf{W}_q$. In statistic-based solution 1, CQI_q is determined



(a) Codebook-based approach.



(b) SVD-based approach.

Fig. 2. Flow charts of the proposed approaches for data-driven MIMO.

by the mode of the CQI set $\mathbf{CQI}_q(\mathbf{W}_{\mathbf{PMI}_q})$ from CLSM, where $\mathbf{CQI}_q(\mathbf{W}_{\mathbf{PMI}_q}) = \{\mathrm{CQI}_{t,q}(\mathbf{W}_{\mathrm{PMI}^*_{t,q}}) | t \in \mathcal{T}\}$. Besides, the CQI set can be obtained from the results of fixed PMI and RI, which is denoted as $\mathbf{CQI}_q(\mathbf{W}_q) = \{\mathrm{CQI}_{t,q}(\mathbf{W}_{\mathrm{mode}(\mathbf{PMI}_q)}) | t \in \mathcal{T}\}$. In statistic-based solution 2 or 3, the fixed CQI is decided as the mode of $\mathbf{CQI}_q(\mathbf{W}_q)$ or the rounded mean value of $\mathbf{CQI}_q(\mathbf{W}_q)$ due to small performance fluctuations of Type I codebook in time domain. Thus, different statistic-based solutions vary in the way to fix CQI. Since adjacent PMI values can correspond to precoders with different beams or ranks, we use nearest neighbor interpolation to predict transmission parameters at any location. This method prevents PMI interpolation from introducing precoders that direct beams in completely incorrect directions.

The proposed approach includes two implementation phases: training and inference. Note that the training process can be performed offline. Specifically, during the training phase, the codebook-based approach calculates transmission parameters at known locations using the CLSM method. After the training process, Q sets of transmission parameters are stored at the DL-BS. In the inference process, the codebook-based approach uses nearest neighbor interpolation to obtain interpolated transmission parameters like a lookup table. If binary search is employed, the time complexity is $\mathcal{O}(\log Q)$ in terms of searching within the stored table of Q sets of transmission parameters.

V. SVD-Based Approach

The quantization of spatial beams in the codebook brings in loss of accuracy of the precoder, yet the optimal precoder can be calculated from the SVD of the channel. Therefore, we present another approach that starts by dealing with the optimal SVD precoders. To capture the underlying meaning of high-dimensional precoders, we turn to VAE to find the representative one from the lower-dimensional latent representations. Considering that the latent representations are Gaussian variables, we leverage GPR to generate new precoders for spatial prediction. VAE and GPR offer dual benefits: effective data compression and generation. First, the VAE's encoder compresses high-dimensional SVD precoders into lower-dimensional latent Gaussian variables. This transformation simplifies the selection process from choosing a specific SVD precoder to selecting a representative Gaussian variable. Second, the Gaussian nature of these latent variables makes GPR an ideal method for predicting their values at unknown locations. Since VAE's latent space is regular, a precoder at any location can be generated by inputting the sampled values of the interpolated Gaussian variable into the decoder. In this section, we first introduce how to obtain the optimal precoder from the SVD of the channel and then exploit VAE and GPR to implement data-driven MIMO based on SVD.

The optimal precoding matrix can be derived from SVD of the channel matrix, as it eliminates interference among layers. The SVD of channel matrix $\mathbf{H}_k$ can be expressed as

$$\mathbf{H}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{V}_k^H, \tag{11}$$

where $\mathbf{U}_k \in \mathbb{C}^{N_{rx} \times N_{rx}}$ are the left unitary matrix and $\mathbf{V}_k \in \mathbb{C}^{N_{tx} \times N_{tx}}$ are the right one. $(\cdot)^H$ denotes the Hermitian transpose operation. $\mathbf{\Lambda}_k \in \mathbb{R}^{N_{rx} \times N_{tx}}$ is a rectangular matrix whose diagonal elements are non-negative real numbers, otherwise zeros. Diagonal elements in $\mathbf{\Lambda}_k$ are named as singular values of $\mathbf{H}_k$, which are denoted as λ_i and given in descending order, i.e., $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_\Upsilon$.

Then, the optimal precoder can be obtained as the first L columns of $\mathbf{V}_k$, which is denoted as $[\mathbf{V}_k]_{:,:L}$. Meanwhile, if the selected equalizer is the first L columns of $\mathbf{U}_k^H$, i.e., $\mathbf{E}_k = [\mathbf{U}_k^H]_{:,:L}$, the post-equalization received symbol vector $\boldsymbol{m}_k$ in (2) can be rewritten as

$$\begin{aligned}
\boldsymbol{m}_k &= [\mathbf{U}_k^H]_{:,:L} \mathbf{H}_k [\mathbf{V}_k]_{:,:L} \boldsymbol{s}_k + [\mathbf{U}_k^H]_{:,:L} \boldsymbol{n}_k \\
&= \left([\mathbf{U}_k^H]_{:,:L} \mathbf{U}_k\right) \mathbf{\Lambda}_k \left(\mathbf{V}_k^H [\mathbf{V}_k]_{:,:L}\right) \boldsymbol{s}_k + [\mathbf{U}_k^H]_{:,:L} \boldsymbol{n}_k \\
&= \begin{bmatrix} \mathbf{I}_L & \mathbf{0} \end{bmatrix} \mathbf{\Lambda}_k \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0} \end{bmatrix} \boldsymbol{s}_k + [\mathbf{U}_k^H]_{:,:L} \boldsymbol{n}_k \\
&= [\mathbf{\Lambda}_k]_{:L,:L} \boldsymbol{s}_k + [\mathbf{U}_k^H]_{:,:L} \boldsymbol{n}_k,
\end{aligned} \tag{12}$$

where $[\mathbf{\Lambda}_k]_{:L,:L}$ corresponds to the first L rows and first L columns of $\mathbf{\Lambda}_k$. Hence, the MIMO transmission can be viewed as L parallel SISO transmissions due to SVD of channel matrix. With waterfilling power allocation, channel capacity can be achieved under certain constraints [38].

Similar to (10), the best precoder from SVD of the channel at any time and location indexes is chosen as follows:

$$\begin{aligned}\mathbf{W}^* &= \arg\max_{\mathbf{W}_i} \sum_{k\in\mathcal{K}} \sum_{\ell\in\mathcal{L}} I_{k,\ell}(\mathbf{W}_i) \\ \text{s.t.}\quad \mathbf{W}_i &= \frac{[\mathbf{V}_k]_{:,:L}}{\|[\mathbf{V}_k]_{:,:L}\|_F} \in \mathcal{W}_{\text{SVD}},\ \ \forall k \in \mathcal{K}, \\ &\forall L \in \{1,\ldots,\Upsilon\},\ \forall\, i \in \{1,\ldots,|\mathcal{W}_{\text{SVD}}|\}, \end{aligned} \tag{13}$$

where $\mathcal{W}_{\text{SVD}}$ is the set of precoders obtained from the SVD of the channel at a certain location, and $|\mathcal{W}_{\text{SVD}}| = K \times \Upsilon$. $\mathbf{W}^*$ is the optimal precoder for the whole system bandwidth. Note that $[\mathbf{V}_k]_{:,:L}$ is a unitary matrix and its power equals to L. Hence, it is divided by $\|[\mathbf{V}_k]_{:,:L}\|_F$ to satisfy the power constraint in (5). RI and CQI can then be determined as introduced before.

A. VAE-Based Solution

AE consists of an encoder and a decoder, which are constructed by multiple layers of neural networks. The encoder compresses the initial input data and the decoder recovers them from lower-dimensional latent space. However, traditional AE fails to generate new meaningful contents, since inputs are encoded into discrete points, leading to the irregularity of latent space. Therefore, VAE is first introduced in [17] to make the latent space continuous. Instead of directly mapping data into points, VAE's encoder returns a distribution over the latent space. Then, the decoder uses points randomly sampled from latent space to recover the initial data. Since the latent space of VAE is regular, we can sample any point from it to generate new data via decoder, which fits well for our purpose of understanding the high dimensional precoders and inferring them in space domain.

Specifically, $\boldsymbol{d}_i$ is denoted as the i-th independent and identically distributed sample of dataset $\mathcal{D}_{\text{VAE}}$ for VAE, where $i \in \{1,\ldots,|\mathcal{D}_{\text{VAE}}|\}$. It is assumed that outputs of VAE are generated by a random process, corresponding to the latent variable $\boldsymbol{\varpi}$ with dimension $N_{lv} \times 1$. The encoder network's parameters are denoted as $\boldsymbol{\psi}$, and the decoder network's as $\boldsymbol{\xi}$. $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ are unknown parameters, which require to be learned from $\mathcal{D}_{\text{VAE}}$. The objective is to maximize the marginal likelihood distribution $p_{\boldsymbol{\xi}}(\boldsymbol{d})$ such that the generative process can accurately recover the inputs.

Nevertheless, the integral of marginal likelihood has no closed-form solution. Thus, the variational inference is exploited in VAE to handle the intractability of the marginal likelihood by introducing the variational distribution $\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d})$ to approximate the true posterior distribution $p_{\boldsymbol{\xi}}(\boldsymbol{\varpi}|\boldsymbol{d})$ [17]. Then, the evidence lower bound (ELBO) for the log-likelihood distribution $\log p_{\boldsymbol{\xi}}(\boldsymbol{d}_i)$ can be derived as

$$\begin{aligned}\text{ELBO}(\boldsymbol{\psi},\boldsymbol{\xi},\boldsymbol{d}_i) &= \mathbb{E}_{\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d}_i)}\left(\log p_{\boldsymbol{\xi}}(\boldsymbol{d}_i|\boldsymbol{\varpi})\right) \\ &\quad - \text{KL}(\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d}_i)||p_{\boldsymbol{\xi}}(\boldsymbol{\varpi})), \end{aligned} \tag{14}$$

where KL is the Kullback-Leibler divergence between two distributions, which measures the difference between them in terms of distribution. The ELBO depends on the parameters of both decoder and encoder networks, along with the input data point. The marginal log-likelihood $\log p_{\boldsymbol{\xi}}(\boldsymbol{d}_i,\ldots,\boldsymbol{d}_{|\mathcal{D}_{\text{VAE}}|}) = \sum_{i=1}^{|\mathcal{D}_{\text{VAE}}|} \log p_{\boldsymbol{\xi}}(\boldsymbol{d}_i)$. Then, maximizing the marginal log-likelihood is equivalent to maximizing the ELBO.

In VAE, the prior $p_{\boldsymbol{\xi}}(\boldsymbol{\varpi})$ and posterior approximation $\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d}_i)$ are both assumed to be Gaussian distribution. $p_{\boldsymbol{\xi}}(\boldsymbol{\varpi}) \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ is a standard multivariate Gaussian distribution, and $\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d}_i) \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\psi},\boldsymbol{d}_i), \boldsymbol{\sigma}^2(\boldsymbol{\psi},\boldsymbol{d}_i)\mathbf{I})$ is a multivariate Gaussian distribution with a diagonal covariance structure, where $\boldsymbol{\mu}(\boldsymbol{\psi},\boldsymbol{d}_i)$ is the mean vector and $\boldsymbol{\sigma}^2(\boldsymbol{\psi},\boldsymbol{d}_i)$ is the variance vector that lies in the diagonal of the covariance matrix. Thus, the encoder maps the input $\boldsymbol{d}_i$ into vectors of mean and variance to obtain the posterior approximation $\widehat{p}_{\boldsymbol{\psi}}(\boldsymbol{\varpi}|\boldsymbol{d}_i)$. Since the sampling operation of latent space lacks gradient information, the reparameterization trick is used in VAE to make gradient decent possible, which can be expressed as

$$\tilde{\boldsymbol{\varpi}}(\boldsymbol{\psi},\boldsymbol{d}_i) = \boldsymbol{\mu}(\boldsymbol{\psi},\boldsymbol{d}_i) + \boldsymbol{\sigma}(\boldsymbol{\psi},\boldsymbol{d}_i) \odot \boldsymbol{\varepsilon}, \tag{15}$$

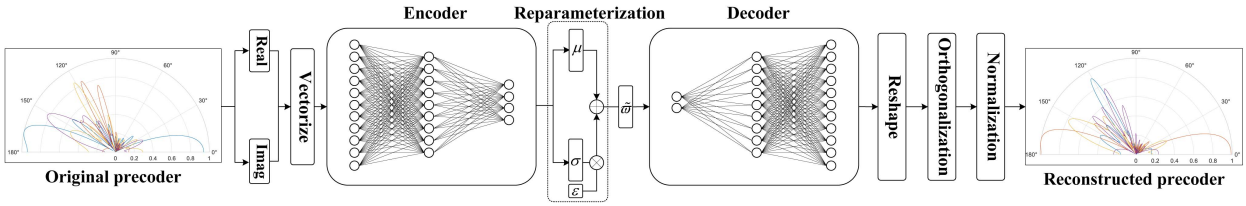
where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ and $\odot$ denotes the element-wise product. The sampled latent variable $\tilde{\boldsymbol{\varpi}}(\boldsymbol{\psi},\boldsymbol{d}_i)$ is the function of parameters of encoder and input data point, so are $\boldsymbol{\mu}(\boldsymbol{\psi},\boldsymbol{d}_i)$ and $\boldsymbol{\sigma}(\boldsymbol{\psi},\boldsymbol{d}_i)$. For simplicity, we leave out $(\boldsymbol{\psi},\boldsymbol{d}_i)$ in the following part when referring to these three variables.

We denote the output of VAE as $\widehat{\boldsymbol{d}}_i, i \in \{1,\ldots,|\mathcal{D}_{\text{VAE}}|\}$, which can be viewed as the estimate of input $\boldsymbol{d}_i$. The negative ELBO can be chosen as the loss function, which is minimized during the training process of VAE and can be defined as

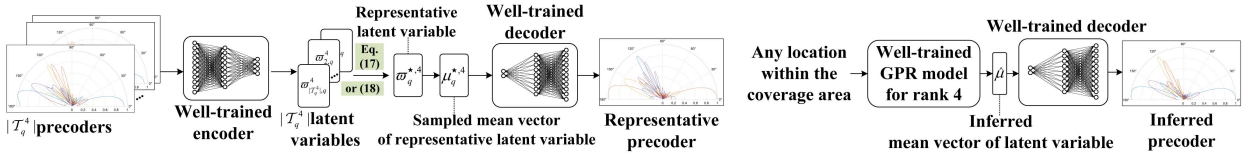
$$\begin{aligned}\text{Loss}(\boldsymbol{\psi},\boldsymbol{\xi},\boldsymbol{d}_i) &= \text{SE}(\boldsymbol{d}_i,\widehat{\boldsymbol{d}}_i) \\ &\quad + \beta\text{KL}\left(\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\sigma}^2\mathbf{I})||\mathcal{N}(\mathbf{0},\mathbf{I})\right) \\ &= ||\boldsymbol{d}_i - \widehat{\boldsymbol{d}}_i||^2 \\ &\quad + \frac{\beta}{2}\sum_{j=1}^{N_{lv}}(\boldsymbol{\mu}_j^2 + \boldsymbol{\sigma}_j^2 - \log(\boldsymbol{\sigma}_j^2) - 1), \end{aligned} \tag{16}$$

where the first term indicates the reconstruction loss in squared error form. The second term refers to the KL divergence loss between the posterior approximation and the prior of latent variable, which is integrated analytically in the Gaussian case. To minimize the loss function is to reduce the reconstruction error and keep the posterior approximation close to standard Gaussian distribution at the same time, where the latter can be considered as a regularisation term to the latent space. β is an adjustable hyperparameter introduced in [39] to balance the tradeoff between reconstruction loss and KL divergence loss. During the training process, a minibatch of sampled data $\{\boldsymbol{d}_i,\widehat{\boldsymbol{d}}_i,\boldsymbol{\mu}_i,\boldsymbol{\sigma}_i\}_{i=1}^{N_b}$ of size N_b is utilized to compute the loss in (16). Subsequently, the parameters $\boldsymbol{\psi}$ of encoder network and $\boldsymbol{\xi}$ of decoder network are updated by an optimization algorithm based on the gradient $\nabla_{\boldsymbol{\psi},\boldsymbol{\xi}}\text{Loss}(\boldsymbol{\psi},\boldsymbol{\xi},\boldsymbol{d}_i)$.

In the following parts of this subsection, we introduce how to implement the SVD-based approach leveraging VAE. Fig. 2(b) illustrates the flow chart of the SVD-based approach.



(a) The proposed VAE architecture for precoders.



(b) Selecting the representative precoder in time domain utilizing VAE (with precoders of rank 4 as an example).

(c) Inferring the precoder at any location utilizing VAE (with precoders of rank 4 as an example).

Fig. 3. Illustration of the proposed VAE architecture for precoders, fixing and inferring precoder utilizing VAE.

We first describe how to generate the dataset for VAE. Then, RI is fixed according to a percentage count threshold. We propose two methods to choose the representative precoders at different locations in $\mathcal{H}$, and then the fixed CQI can be determined from the CQI set of the fixed precoder and RI. At last, we utilize GPR to infer the precoder at any location in $\mathcal{H}'$ and exploit NNI for CQI interpolation. The details of how we utilize VAE for precoders are demonstrated in Fig. 3.

1) Selection of Transmission Parameters: In this paper, we utilize VAEs to reduce the dimensionality of the precoder and capture its underlying characteristics, represented by the latent Gaussian variables. The representative precoder at different locations is then obtained by selecting the representative latent variable and inputting the sampled values of it into the well-trained decoder network. Since precoders of different ranks vary in dimension, Υ distinct VAEs need to be trained to accommodate each unique input dimension.

However, preparing the input precoder dataset for VAE remains a challenge. First, the optimal precoders $\mathbf{W}_{t,q}^{*}$ for specific time t and location q based on existing channel data can be derived from (13). It should be noted that the rank $r(\mathbf{W}_{t,q}^{*})$ may vary across different time and location indexes, reflecting the dynamic nature of channel conditions. By setting the rank of the precoder obtained from the SVD of the channel to $\mathfrak{r}$ in (13), we can obtain "optimal" precoders of rank $\mathfrak{r}$, denoted as $\mathbf{W}^{*,\mathfrak{r}}$. Then, the precoder dataset of rank $\mathfrak{r}$ for the VAE $\mathcal{D}_{\text{VAE}}^{\mathfrak{r}}$ can be collected from $\mathbf{W}_{t,q}^{*,\mathfrak{r}}$. This is done by applying constraints on the time t and location q, which are derived from the original optimal precoders $\mathbf{W}^{*}$. In principle, the time and location indexes selected for the higher-rank precoder dataset are also applicable to the lower-rank one, resulting in the latter containing more elements. Thus, we denote the set of rank choices for the precoder dataset of rank $\mathfrak{r}$ as $\mathfrak{L}$, where $\mathfrak{L} = \{\mathfrak{r}, \ldots, \Upsilon\}$ and $\mathfrak{r}$ ranges from 1 to Υ. Then, the constraints on location and time indexes are derived as follows: a) The location index set $\mathcal{Q}^{\mathfrak{r}}$ for precoders of rank $\mathfrak{r}$ is selected as $\mathcal{Q}^{\mathfrak{r}} = \{q|\text{mode}(\mathcal{R}_q(\mathbf{W}_q^{*})) \in \mathfrak{L}, q \in \mathcal{Q}\}$, where $\mathcal{R}_q(\mathbf{W}_q^{*}) = \{r(\mathbf{W}_{t,q}^{*})|t \in \mathcal{T}\}$ represents the rank set of optimal precoders at location index q; b) The time index set $\mathcal{T}_q^{\mathfrak{r}}$ for precoders of rank $\mathfrak{r}$ at location q is chosen as $\mathcal{T}_q^{\mathfrak{r}} = \{t|r(\mathbf{W}_{t,q}^{*}) \in \mathfrak{L}, t \in \mathcal{T}\}, \forall q \in \mathcal{Q}^{\mathfrak{r}}$. Finally, the precoder dataset for the VAE of rank $\mathfrak{r}$ is formulated by combining these two constraints: $\mathcal{D}_{\text{VAE}}^{\mathfrak{r}} = \{\mathbf{W}_{t,q}^{*,\mathfrak{r}}|q \in \mathcal{Q}^{\mathfrak{r}}, t \in \mathcal{T}_q^{\mathfrak{r}}\}$.

Since the values of $\mathbf{W}^{*} = \mathbf{V}^{*}/\|\mathbf{V}^{*}\|_F$ are small due to the power normalization, the optimal precoder without power normalization $\mathbf{V}^{*}$ is employed in the VAE dataset. $\mathbf{V}^{*}$ is first reshaped into a vector of size $N_{tx} \times L \times 2$, where 2 accounts for the real and imaginary parts of the complex-valued precoder. The output from the VAE is then reshaped back to the original dimensions. Furthermore, to ensure the recovered precoder $\widehat{\mathbf{V}}^{*}$ conforms to the unitary matrix property, an orthogonalization operation is performed via QR decomposition on it: $\widehat{\mathbf{V}}^{*} = \widehat{\mathbf{V}}_{\text{Q}}^{*}\widehat{\mathbf{V}}_{\text{R}}^{*}$. Here, $\widehat{\mathbf{V}}_{\text{Q}}^{*} \in \mathbb{C}^{N_{tx} \times L}$ is an orthogonal matrix with columns as unit vectors and $\widehat{\mathbf{V}}_{\text{R}}^{*} \in \mathbb{C}^{L \times L}$ is an upper triangular matrix. Finally, $\widehat{\mathbf{V}}_{\text{Q}}^{*}/\|\widehat{\mathbf{V}}_{\text{Q}}^{*}\|_F$ is used as the generated or recovered precoder, ensuring power normalization and column orthogonality.

In this paper, we fix the transmission parameters at known locations in $\mathcal{H}$ to facilitate inference at any location. Initially, RI needs to be fixed to choose the representative precoder corresponding to that rank. To this end, we define a percentage count threshold $\mathfrak{C}_{\text{thold}}$ for selecting the fixed RI. If no fewer than $\mathfrak{C}_{\text{thold}}$ percent of T RIs obtained from (13) at location q belong to $\mathfrak{L}$ when the mode of $\mathcal{R}_q(\mathbf{W}_q^{*})$ is $\mathfrak{r}$, then L_q is fixed to $\mathfrak{r}$; otherwise, it is set to $\mathfrak{r}-1$. Thus, apart from locations where the highest rank of precoder is employed, the set of locations $\mathcal{Q}_{\text{fixed}}^{\mathfrak{r},\mathfrak{C}_{\text{thold}}}$ where the fixed RI is $\mathfrak{r}$ for a specific threshold $\mathfrak{C}_{\text{thold}}$ consists of two parts: the location indexes that meet the constraints described above when the mode of $\mathcal{R}_q(\mathbf{W}_q^{*})$ is $\mathfrak{r}$ and those that do not when the mode of $\mathcal{R}_q(\mathbf{W}_q^{*})$ is $\mathfrak{r}+1$. $\mathcal{Q}_{\text{fixed}}^{\mathfrak{r},\mathfrak{C}_{\text{thold}}} \subseteq \mathcal{Q}^{\mathfrak{r}}$ and $\sum_{\mathfrak{r}=1}^{\Upsilon}|\mathcal{Q}_{\text{fixed}}^{\mathfrak{r},\mathfrak{C}_{\text{thold}}}| = |\mathcal{Q}| = Q$. A higher value of $\mathfrak{C}_{\text{thold}}$ indicates a more conservative choice of fixed RI.

The representative latent variable at location q for precoders of rank $\mathfrak{r}$, denoted as $\boldsymbol{\varpi}_q^{\star,\mathfrak{r}}$, can be chosen based on its proximity to the mean values of mean and variance vectors of all latent variables at that location. The time index of the representative

latent variable at location q can be selected as

$$\arg\min_{t} \left[(\boldsymbol{\mu}_{t,q}^{\mathfrak{r}} - \bar{\boldsymbol{\mu}}_{q}^{\mathfrak{r}})^{\mathrm{T}}(\boldsymbol{\mu}_{t,q}^{\mathfrak{r}} - \bar{\boldsymbol{\mu}}_{q}^{\mathfrak{r}}) + ((\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}})^2 - (\bar{\boldsymbol{\sigma}}_{q}^{\mathfrak{r}})^2)^{\mathrm{T}}((\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}})^2 - (\bar{\boldsymbol{\sigma}}_{q}^{\mathfrak{r}})^2)\right], \forall t \in \mathcal{T}_q^{\mathfrak{r}}, \quad (17)$$

where $\bar{\boldsymbol{\mu}}_{q}^{\mathfrak{r}} = \sum_{t\in\mathcal{T}_q^{\mathfrak{r}}} \boldsymbol{\mu}_{t,q}^{\mathfrak{r}}/|\mathcal{T}_q^{\mathfrak{r}}|$ is the averaged mean vector at location q, and $(\bar{\boldsymbol{\sigma}}_{q}^{\mathfrak{r}})^2 = \sum_{t\in\mathcal{T}_q}(\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}})^2/|\mathcal{T}_q^{\mathfrak{r}}|$ is the averaged variance vector. Since the probability of obtaining the mean of a Gaussian variable is highest, we take the mean vector $\boldsymbol{\mu}_q^{\star,\mathfrak{r}}$ of $\boldsymbol{\varpi}_q^{\star,\mathfrak{r}}$ as its sampled values. This mean vector is then input into the well-trained decoder of VAE to obtain the representative precoder. Since this method identifies the latent variable closest to the statistical mean value, we denote the recovered or generated precoder as $\mathbf{V}_{\text{mean}}$.

Another method for choosing the representative latent variable is selecting it based on its distributional proximity to others. The chosen time index of the representative latent variable at location q can be determined as

$$\arg\min_{t} \sum_{t'\in\mathcal{T}_q^{\mathfrak{r}}\setminus t} \begin{cases} \mathrm{KL}(D(\boldsymbol{\varpi}_{t,q}^{\mathfrak{r}})||D(\boldsymbol{\varpi}_{t',q}^{\mathfrak{r}})), \text{if } t < t' \\ \mathrm{KL}(D(\boldsymbol{\varpi}_{t',q}^{\mathfrak{r}})||D(\boldsymbol{\varpi}_{t,q}^{\mathfrak{r}})), \text{if } t' < t \end{cases}, \forall t \in \mathcal{T}_q^{\mathfrak{r}}, \quad (18)$$

where $D(\cdot)$ represents the distribution of the variable. We use two cases to ensure that KL divergence is consistently calculated from the distribution of the lower-indexed latent variable to the higher-indexed one such that the summation can accurately reflect the total distributional distance. $\mathrm{KL}(D(\boldsymbol{\varpi}_{t,q}^{\mathfrak{r}})||D(\boldsymbol{\varpi}_{t',q}^{\mathfrak{r}}))$ in Gaussian case is given by

$$\frac{1}{2}\sum_{j=1}^{N_{lv}}\left(\frac{([\boldsymbol{\mu}_{t,q}^{\mathfrak{r}}]_j - [\boldsymbol{\mu}_{t',q}^{\mathfrak{r}}]_j)^2 + [\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}}]_j^2}{[\boldsymbol{\sigma}_{t',q}^{\mathfrak{r}}]_j^2} + \log[\boldsymbol{\sigma}_{t',q}^{\mathfrak{r}}]_j^2 - \log[\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}}]_j^2 - 1\right), \quad (19)$$

where $[\boldsymbol{\mu}_{t,q}^{\mathfrak{r}}]_j$ and $[\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}}]_j$ denote the j-th elements of vectors $\boldsymbol{\mu}_{t,q}^{\mathfrak{r}}$ and $\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}}$, respectively. Since this method selects the representative latent variable based on KL divergence, we refer to the resulting precoder as $\mathbf{V}_{\text{KL}}$.

When employing the representative precoder and RI obtained from VAE at location q, the CQI set $\mathbf{CQI}_q(\mathbf{W}_q)$ can be established using $\mathbf{W}_q = \mathbf{V}_{\text{mean}}$ or $\mathbf{V}_{\text{KL}}$. The fixed CQI is then determined as the floor of the mean value of $\mathbf{CQI}_q(\mathbf{W}_q)$ since fixed SVD precoder may lead to more performance fluctuations in time domain.

2) Inference of Transmission Parameters: In data-driven MIMO, the performance of selected transmission parameters is first evaluated on $\mathcal{H}$. Then, the performance of inferred transmission parameters at any location is tested on $\mathcal{H}'$. Thus, $\mathcal{H}'$ can be viewed as the testing dataset.

RI should be first determined such that the representative precoders of different layers can be utilized for prediction at any location. Since higher RI would lead to more interference among data streams, we have a conservative choice of RI and determine the value of RI at any location in $\mathcal{H}'$ as the minimal value of fixed RIs at N_{RI} closest locations in $\mathcal{H}$. Since the latent representations are multivariate Gaussian variables, we use GPR to predict the latent variable at any location and then feed the sampled values of it into the decoder of VAE to generate the inferred precoder.

GPR is a non-parameter regression model based on probability [40]. It assumes that input data are the samples of the Gaussian process, which obey the Gaussian distribution and so do the latent variables in VAE. In this paper, similar to VAE, we need to train Υ GPRs. Consider a dataset $\mathcal{D}_{\text{GPR}}^{\mathfrak{r}} = \{(\boldsymbol{\varrho}_q, \boldsymbol{\mu}_q^{\star,\mathfrak{r}})|q \in \mathcal{Q}^{\mathfrak{r}}\}$ for GPR to infer precoders of rank $\mathfrak{r}$, where $\boldsymbol{\mu}_q^{\star,\mathfrak{r}}$ is the mean vector of the representative latent variable and is also the desirable output corresponding to input location $\boldsymbol{\varrho}_q$. GPR fits the unknown function by utilizing a kernel function $\mathfrak{K}$, which measures the similarity between two input locations. Close inputs are expected to have similar outputs in GPR. The most common kernel function is radial basic function (RBF), which is defined as

$$\mathfrak{K}(\boldsymbol{\varrho}_q, \boldsymbol{\varrho}_{q'}) = \gamma^2 e^{-\frac{(\boldsymbol{\varrho}_q - \boldsymbol{\varrho}_{q'})(\boldsymbol{\varrho}_q - \boldsymbol{\varrho}_{q'})^{\mathrm{T}}}{2\zeta^2}}, \quad (20)$$

where γ and ζ are hyperparameters concerning width and characteristic length-scale, respectively.

Let $\boldsymbol{\Gamma}$ be the combined observed location matrix with each column as $\boldsymbol{\varrho}_q$, and $\boldsymbol{\Delta}$ be the combined known output matrix with each column as $\boldsymbol{\mu}_q^{\star,\mathfrak{r}}$, where $q \in \mathcal{Q}^{\mathfrak{r}}$. The regression model $p(\boldsymbol{\Delta}|\boldsymbol{\Gamma})$ is assumed to be a multivariate Gaussian distribution with the mean vector of zeros and covariance matrix $\boldsymbol{\Xi}(\boldsymbol{\Gamma},\boldsymbol{\Gamma})$. $\boldsymbol{\Xi}(\boldsymbol{\Gamma},\boldsymbol{\Gamma}) \in \mathbb{R}^{|\mathcal{Q}^{\mathfrak{r}}|\times|\mathcal{Q}^{\mathfrak{r}}|}$ characterises the correlation between observed location points $\boldsymbol{\Gamma}$, and its element in i-th row and j-th column is $\mathfrak{K}(\boldsymbol{\varrho}_i, \boldsymbol{\varrho}_j)$. During the training process, the log marginal likelihood $\log p(\boldsymbol{\Delta}|\boldsymbol{\Gamma})$ is maximized by adjusting the hyperparameters of the kernel function, where

$$\log p(\boldsymbol{\Delta}|\boldsymbol{\Gamma}) = -\frac{1}{2}\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\boldsymbol{\Delta} - \frac{1}{2}\log\det(\boldsymbol{\Xi}) - \frac{|\mathcal{Q}^{\mathfrak{r}}|}{2}\log 2\pi. \quad (21)$$

Here, $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\boldsymbol{\Gamma},\boldsymbol{\Gamma})$ and $\det(\boldsymbol{\Xi})$ indicates its determinant. Then, the optimal hyperparameters of RBF are determined. Based on the properties of multivariate Gaussian, the posterior probability $p(\hat{\boldsymbol{\Delta}}|\boldsymbol{\Delta})$ is maximized when the predicted mean vectors $\hat{\boldsymbol{\Delta}}$ of latent variables at unobserved locations $\hat{\boldsymbol{\Gamma}}$ equal to $\boldsymbol{\Xi}(\hat{\boldsymbol{\Gamma}},\boldsymbol{\Gamma})\boldsymbol{\Xi}(\boldsymbol{\Gamma},\boldsymbol{\Gamma})^{-1}\boldsymbol{\Delta}$. $\boldsymbol{\Xi}(\hat{\boldsymbol{\Gamma}},\boldsymbol{\Gamma})$ describes the correlation between unobserved locations $\hat{\boldsymbol{\Gamma}}$ and observed ones $\boldsymbol{\Gamma}$, while $(\cdot)^{-1}$ indicates the inverse operation. By inputting $\hat{\boldsymbol{\Delta}}$ to the well-trained decoder network of VAE, we can obtain the inferred precoders at unobserved locations.

As to CQI, NNI can be used as the integer interpolation method to obtain interpolated CQI at any location in $\mathcal{H}'$ based on the fixed CQI at locations in $\mathcal{H}$. NNI is a spatial interpolation method based on the Voronoi diagram [41], which is a partition of a space obtained from the perpendicular bisectors of points given the known sample set. Specifically, we can obtain the original Voronoi diagram based on fixed CQIs at locations in $\mathcal{H}$. New location point q' in $\mathcal{H}'$ will lead to the change in the Voronoi cells of points close to it, then interpolated CQI at location q' is given by

$$\mathrm{CQI}_{q'} = \sum_{q\in\mathcal{Q}(q')} \frac{\mathfrak{V}_q \cap \mathfrak{V}_{q'}}{\mathfrak{V}_{q'}}\mathrm{CQI}_q, \quad \forall q' \in \mathcal{Q}', \quad (22)$$

where $\mathcal{Q}(q')$ is the set of original locations that are close to new location sample q' and CQI_q is the fixed CQI at location q. The volume of new Voronoi cell centered at location q' is denoted as $\mathfrak{V}_{q'}$, and the intersected volume between the original cell centered at q and the new cell centered at q' is denoted as $\mathfrak{V}_q \cap \mathfrak{V}_{q'}$. Therefore, $\mathfrak{V}_q \cap \mathfrak{V}_{q'}/\mathfrak{V}_{q'}$ serves as the interpolation weight, where $0 < \mathfrak{V}_q \cap \mathfrak{V}_{q'}/\mathfrak{V}_{q'} \leqslant 1$ and $\sum_{q\in\mathcal{Q}(q')} \mathfrak{V}_q \cap \mathfrak{V}_{q'}/\mathfrak{V}_{q'} = 1, \forall q' \in \mathcal{Q}'$.

In summary, the representative high-dimensional SVD precoder can be efficiently determined through the VAE's lower-dimensional latent variable, which follows a Gaussian distribution. Subsequently, GPR is used to interpolate these latent variables at unknown locations. These interpolated values are then input into the VAE to generate corresponding precoders. By incorporating the interpolated CQI, the transmission parameters at any location can be determined.

The training process for the SVD-based approach involves more steps, including calculating SVD precoders and training both the VAE and GPR. After the training phase, the parameters of the VAE's decoder and GPR are saved at the DL-BS. The parameters of the decoder are determined by the neural network's design, while the space complexity of GPR is known as $\mathcal{O}(Q^2)$ in terms of the number of data points fitted. In the inference phase, the complexity primarily stems from predicting precoders, where the latent Gaussian variable interpolated by GPR is fed into the decoder to infer the precoder. The time complexity of the former procedure, involving the calculation of the inverse of the covariance matrix in GPR, is $\mathcal{O}(Q^3)$. The time complexity of the latter step, which involves a forward propagation of the decoder, is influenced by the number of layers in the decoder and the number of neurons in each layer. In this paper, the decoder is constructed with four layers of sizes $10L$, 128, 400, and $2N_{tx}L$, which is a relatively small neural network model. In practical implementations of the proposed approaches, the inference process will not be frequent when the user's location remains unchanged. Even when the user moves and the location needs updating, the reporting period of location is not required to be as short as CSI feedback at the millisecond level. Considering the edge computing capability of 5G BSs, hardware limitation will not be an issue for our proposed approaches.

The proposed approaches are applicable in practical networks since both the feedback-free transmission and 5G require the same MIMO transmission parameters. However, each proposed approach offers distinct advantages and disadvantages. The codebook-based approach is easy to implement in 5G BS by collecting statistics of CSI feedback from users. However, it exhibits poor performance in spatial prediction. Conversely, the SVD-based approach provides consistent performance within the coverage area, at the cost of the dedicated process of collecting raw channel data for subsequent processing of SVD precoders, VAE, and GPR. The proposed approaches are designed in alignment with the 5G system, allowing their application in both FD-RAN and 5G in a variety of scenarios. Once the learned mapping is deployed, the existing 5G BS can effectively implement these methods for static and moving users, with their locations reported or predicted. One of the static user scenarios is the industrial IoT, where the locations of sensors, machines, and devices are generally fixed. For moving user scenarios, the user's location is usually easy to predict when the user is inside a moving car or on a high-speed train.

Considering the possibilities for blockages to obstruct the line-of-sight (LOS) path from BS to UE, an additional set of transmission parameters for each location can be introduced to our proposed approaches. The system performance can then be enhanced through a lightweight feedback mechanism from UE to the edge cloud. This feedback allows the system to switch between transmission parameters that either consider or disregard the LOS path, depending on the BLER of the currently employed parameters.

VI. Simulation Results and Discussions

In this section, we introduce the simulation setup of data-driven MIMO and present the simulation results and discussions of proposed solutions for two different approaches.

A. Simulation Setup

In this paper, throughput is obtained by signal processing in the physical layer based on the Vienna 5G link level simulator [42]. BS is equipped with $N_{tx} = 16$ dual polarized transmit antennas with configuration $(N_1, N_2) = (8, 1)$ and UE is equipped with $N_{rx} = 4$ single polarized receive antennas. Υ is given by $\min\{N_{rx}, N_{tx}\} = 4$. The center frequency is 3.5 GHz and the subcarrier spacing is 15 kHz, with a transmission time interval (TTI) of 1 ms. In a single TTI or subframe, the resource element (RE) grid for MIMO transmission consists of 1008 REs, obtained from 72 subcarriers $\times$ 14 symbols. Low-density parity-check (LDPC) channel coding is employed and the equalizer is chosen based on minimum mean squared error (MMSE).

Meanwhile, simulations are conducted based on the realistic ray-tracing dataset: DeepMIMO [43]. Using accurate ray-tracing data obtained from Remcom Wireless InSite, clustered delay line (CDL) channels between BS and UE are generated according to standard [44], which serve as the dataset in this paper. The DeepMIMO dataset has many candidate BSs and UEs with different coordinates to choose from, while all BSs' height is 6m and that of UEs is 2m. In this paper, we focus on using *BS1* to transmit data to UEs in *User Grid 1* of DeepMIMO's *O1* scenario. The training set consists of channels generated between *BS1* and UEs of 126 rows from *R715* to *R1465* in *O1* with equal spacing, where each row contains 30 user locations. Thus, $|\mathcal{Q}| = 3780$. Similarly, the testing set is made up of channels generated between *BS1* and UEs of 63 rows from *R718* to *R1462* in *O1* with equal spacing, where each row contains 15 user locations. Then, $|\mathcal{Q}'| = 945$, which corresponds to the 80-20 split of training and testing sets. When generating the channel, the maximum number of channel paths is set to 10, though the actual number varies at different locations based on ray-tracing results. Besides, we simulate time-varying channels by considering 100 different channel samples in the time domain, corresponding to channels in arbitrary 100 TTIs. Specifically, we assume that the channel path loss remains constant at a specific location.

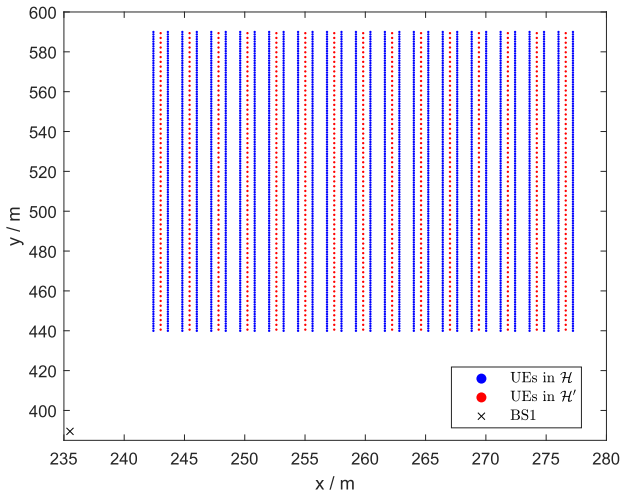


Fig. 4. Locations of selected UEs and BS in DeepMIMO's *O1* scenario in the x and y directions.

Since the *O1* scenario is static, we adopt its 5G NR-compatible version by importing the ray-tracing results into the CDL model to generate channel data for each TTI with random variations. The random seeds are set from 1 to 100 to reflect the small-scale environmental variations for both training and testing channel datasets. This setup yields a channel coherence time of 1 ms, which is relatively short compared with the CSI-RS period that ranges from 1 to 640 ms as specified in [45]. Therefore, each sample in the training and testing datasets falls outside the channel coherence time interval. Detailed locations of selected UEs and BS are shown in Fig. 4.

Although a higher accuracy of the recovered precoder can be obtained if the network in VAE gets deeper, we found that the loss wouldn't reduce too much after more than four layers are employed for both encoder and decoder networks. Hence, in this paper, they are both designed as four-layer fully connected neural networks. For the encoder, the input dimension is $N_{tx} \times L \times 2 = 32L$. The following three layers for the encoder are of size 400, 128, and $N_{lv} \times 2$, where 2 indicates the means and log variances of the latent variable. N_{lv} is set to $10L$. In this paper, we output the log variance of the latent variable to ensure the variance is always positive. Leaky rectified linear unit (ReLU) is chosen to be the activation function for the encoder. For the decoder, the input dimension is N_{lv}. Conversely, the following three layers for the decoder are of size 128, 400, and $32L$. The activation functions for the decoder are also leaky ReLU, except that the last layer is Tanh to obtain the outputs whose values are between -1 and 1. During the training process, a minibatch of $N_b = 128$ samples is used to calculate the gradients. β in (16) is chosen to be 0.01. Then, the parameters of networks in VAE are updated for 100 epochs using ADAM [46] optimization algorithm with 10^{-3} learning rate. Despite fine-tuning may be helpful, we found that the proposed VAE solution works well under many values of hyperparameters.

Currently, there are no existing studies on feedback-free MIMO that consider multi-stream transmission and jointly select all three transmission parameters. Therefore, we benchmark the performance of our proposed schemes against feedback-based 5G CLSM, which is recognized for achieving the highest capacity among various transmission modes in 5G. However, according to [32], CPM-enabled training-free beamforming eliminates the overhead and processing delays associated with channel training and feedback. In essence, CPM reconstructs the channel using information from the three strongest channel paths. It stores channel path parameters at locations with collected ray-tracing data and uses these parameters to interpolate and construct the channel at any location. Transmission parameters are then determined based on the reconstructed channel. Nevertheless, CPM does not account for channel variations over time. Thus, in our dataset, we reconstruct the channel using the channel path parameters at the first time index. Subsequently, we determine the transmission parameters using both 5G Type I codebook and SVD, denoted as CPM-codebook and CPM-SVD, respectively.

During both the training and inference processes, CPM-codebook and CPM-SVD calculate the transmission parameters from one representative channel based on the codebook and SVD, respectively. In the training phase, the predefined precoder in the codebook can be directly used to calculate the appropriate transmission parameters, while deriving the SVD precoder from the channel requires extra computation. In terms of the number of sets of calculated transmission parameters, the time complexity of CPM is constant. On the contrary, the time complexity of the proposed approaches is $\mathcal{O}(|\mathcal{T}|)$ since they consider transmission parameters across all time indexes. Hence, the time complexities are ranked as follows: CPM-codebook $<$ CPM-SVD $<$ codebook-based approach $<$ SVD-based approach. The space complexity of CPM is $\mathcal{O}(Q)$, in terms of the number of sets of channel path parameters to be stored, which is equivalent to that of the codebook-based approach. Consequently, the space complexities are ranked as follows: CPM-codebook $=$ CPM-SVD $=$ codebook-based approach $<$ SVD-based approach. In the inference phase, the codebook-based approach obtains interpolated transmission parameters similar to referencing a lookup table, whereas CPM still requires computation for these parameters. Additionally, the SVD-based approach involves the combination of VAE and GPR, which adds further computation. Thus, the time complexities are ranked as follows: codebook-based approach $<$ CPM-codebook $<$ CPM-SVD $<$ SVD-based approach.

B. Evaluation on Time and Frequency Domains

In this subsection, we first evaluate the codebook-based approach under imperfect CSI and analyze its performance when transmitting single and multiple streams. Then, we verify the SVD-based approach at locations suitable for transmitting four data streams and across all locations in $\mathcal{H}$. Here, location samples suitable for transmitting four streams refer to location points that utilize precoders of rank 4 at all time indexes based on the results of SVD obtained from (13), i.e., $\mathcal{Q}_{\text{fixed}}^{4,100}$.

1) Evaluation of Codebook-Based Approach: Given that the process of collecting channel data may introduce errors, we evaluate the proposed approach under imperfect CSI caused by

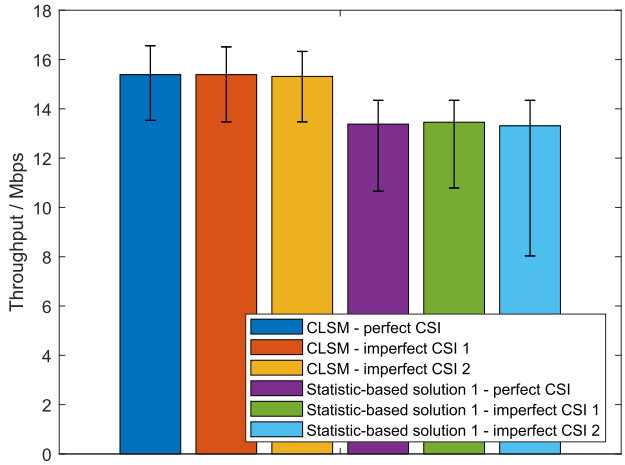


Fig. 5. Average throughput comparison among different transmission schemes under perfect and imperfect CSI at 20 locations samples in $\mathcal{H}$.

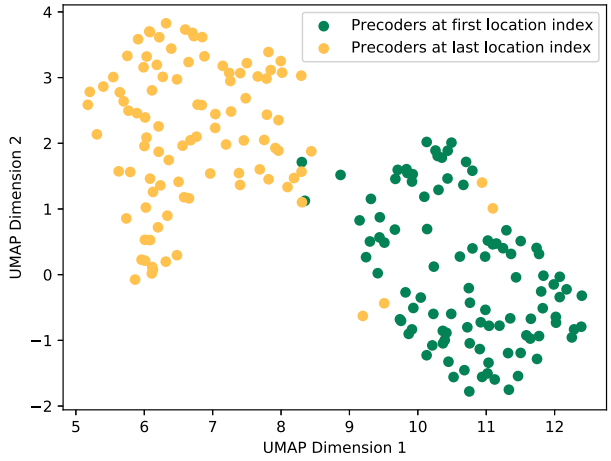
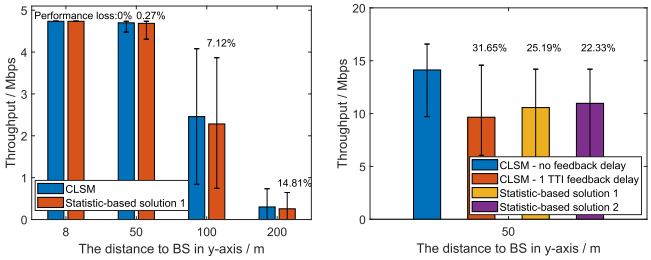


Fig. 7. Visualization of latent variables of VAE based on UMAP corresponding to the optimal SVD precoders at the first and last location indexes in $\mathcal{H}$.



(a) Single stream. (b) Multiple streams.

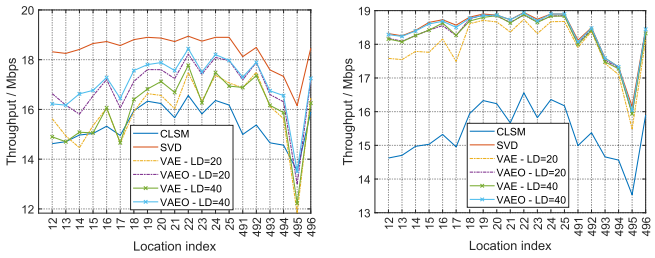
Fig. 6. Average throughput comparison between CLSM and statistic-based solutions when transmitting single and multiple streams.

least square channel estimation errors at 20 location samples in $\mathcal{H}$, as illustrated in Fig. 5. We allocate 768 and 64 pilots across time and frequency on the RE grid of 1008 REs to estimate the channel, respectively. This allocation leads to small and large channel estimation errors, denoted as imperfect CSI 1 and 2. Fig. 5 shows that the average throughput of the proposed approach is almost the same under both perfect and imperfect CSI, demonstrating its robustness comparable to CLSM. The error bar shows the maximal and minimal throughput at these locations. Imperfect CSI resulting from channel estimation in practice has little impact on performance, as the calculated transmission parameters remain largely unchanged. Thus, we will focus on evaluating the performance of various transmission schemes without channel estimation error in the following. Note that the performance loss of proposed solutions for data-driven MIMO does not take the saved overheads of pilots and feedback into consideration.

Fig. 6 shows the average throughput comparison between CLSM and statistic-based solutions when transmitting single and multiple streams. Throughput is first averaged in the time domain and then is averaged corresponding to all locations that have the same distance to BS in the y-axis. The simulation setup for Fig. 6(a) models an open space environment that includes only the LOS path and one non-LOS (NLOS) path. Note that throughput is determined from the 5G-compatible link level simulator, which runs the full physical layer processing chain including adopting quantized CQI instead of the theoretical Shannon formula. Under these conditions, CQI for the UE closest to the BS consistently achieves the maximum value of 15, resulting in no performance loss for statistic-based solution 1 compared with CLSM. It can be seen from Fig. 6(a) that the performance of both transmission schemes lowers as the distance to BS increases. However, the gap between the statistic-based solution and CLSM is negligible, since they are both numerically small. It should be noted that the channel variations are obvious, resulting in an average throughput decrease of over 30% for CLSM with just 1 TTI feedback delay, which can be seen from Fig 6(b). Though outperforming CLSM with feedback delay, statistic-based solutions suffer large performance degradation when multiple data streams are transmitted, with up to 25.19% and 22.33% throughput loss, respectively. Thus, it is of greater value to study the multi-stream transmission of data-driven MIMO, which is the focus of this paper.

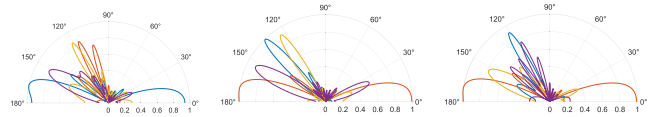
2) Verification of SVD-Based Approach: We first assess the performance of VAE-based solutions for the SVD-based approach at 20 location samples suitable for transmitting four data streams, followed by an evaluation across all samples at those locations. Before that, we demonstrate the interpretability of the VAE-based solution. Specifically, we use the uniform manifold approximation and projection (UMAP) method to visualize the latent variables of VAE [47]. UMAP is widely used for dimension reduction and can project the latent variables into a 2D space, which is shown in Fig. 7. Each dot represents the latent representations of the optimal precoder at a certain time index, thus 100 dots of the same color refer to those at one location index. In the projected 2D space of UMAP, precoders at the same location are mostly entangled with each other while precoders among different locations present diverse representations. These properties assist to choose the representative latent variables at observed locations and infer them at unobserved locations.

Fig. 8 shows the average throughput comparison at 20 location samples in $\mathcal{H}$ when the precoder dataset is made up of



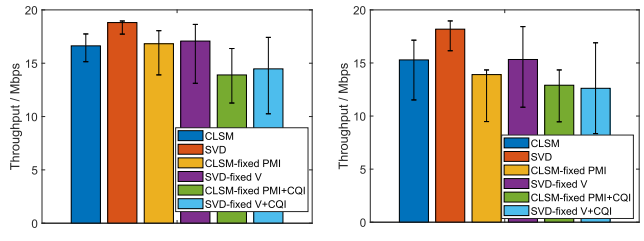
(a) Non-unitary matrices. (b) Unitary matrices.

Fig. 8. Average throughput comparison among different transmission schemes at 20 location samples in $\mathcal{H}$ when the precoder dataset consists of non-unitary or unitary matrices.



(a) Original SVD precoder. (b) Recovered preocder with dataset of non-unitary matrices. (c) Recovered preocder with dataset of unitary matrices.

Fig. 9. Radiation pattern comparison among the original SVD precoder and its recovered ones from VAEO when different kinds of precoder datasets are employed.



(a) Same fixed RI obtained from CLSM and SVD. (b) Different fixed RIs obtained from CLSM and SVD.

Fig. 10. Average throughput comparison among different transmission schemes at location samples in $\mathcal{Q}_{\text{fixed}}^{4,100}$.

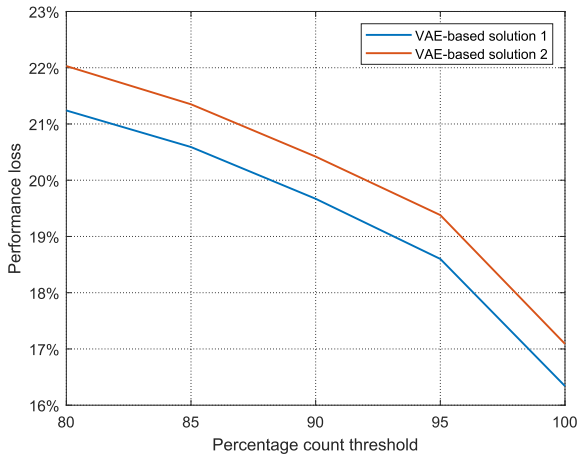


Fig. 11. Performance loss of VAE-based solutions compared with CLSM at locations in $\mathcal{H}$ when $\mathfrak{C}_{\text{thold}}$ is changing.

non-unitary matrices or unitary matrices. Note that the precoder recovered from the VAE at each time index with CQI feedback is implemented to demonstrate the VAE's reconstruction performance. In Fig. 8, LD refers to the latent dimension, SVD denotes SVD-based precoding with CQI feedback, and VAEO represents VAE with the orthogonalization operation. Employing precoders of unitary matrices as the precoder dataset shows significantly better reconstruction performance of the original SVD precoders than directly using the energy normalized precoders since the values of normalized ones are too small. VAE of higher latent dimension has greater reconstruction performance for it can better capture the characteristics of precoders. Besides, the orthogonalization operation offers large performance gains since it keeps the multiple streams orthogonalized to each other to reduce the interference among them. The radiation patterns in Fig. 9 intuitively illustrate that using precoders consisting of unitary matrices along with orthogonalization can reconstruct the original beams well. Then we focus on all the location samples that employ precoders of rank 4 for all time indexes from the results of SVD. We find that RIs derived from the statistic-based solution are not always four at these locations, thus we divide them into two parts: locations where the fixed RI are both four from statistic-based and VAE-based solutions and the rest of them.

5G Type I codebook quantizes the optimal SVD precoder, resulting in the inaccuracy of the corresponding beams and potentially reducing the number of data streams for transmission, which causes performance loss, as shown in Fig. 10. The fixed PMI and CQI of the codebook-based approach from CLSM in Fig. 10 is chosen according to statistic-based solution 3. $\mathbf{V}_{\text{mean}}$ is chosen to be the method to fix the SVD precoder in Fig. 10. As demonstrated in Fig. 10(a), the performance of the VAE-based solution outperforms the statistic-based solution in average throughput if their fixed RIs are the same, since the former can better form the beams towards the multipath signals. CLSM based on the codebook exhibits a conservative preference in the selection of transmission parameters, employing RIs of 3 and 4 in Fig. 10(a). This leads to relatively smaller throughput fluctuations with fixed transmission parameters. The VAE-based solution is expected to outperform the statistic-based solution if the former has a higher fixed RI since it has more data streams to transmit. However, the simulation results prove the other way around, as shown in Fig. 10(b). Although the VAE-based solution outperforms if only precoders are fixed, it has severe performance degradation when CQI is then fixed because it brings higher CQI fluctuations due to higher RI. It raises the crucial procedure of VAE-based solution: selecting the appropriate fixed RI for each location. The conservative choice made by the codebook-based approach can reduce the performance loss brought by fixed transmission parameters when the channel state is not ideal.

Since the proper choice of fixed RI is critical to VAE-based solutions, we evaluate their performance loss compared with CLSM when the criterion for selecting the fixed RI changes, which is shown in Fig. 11. VAE-based solution 1 indicates that the representative SVD precoder is chosen as $\mathbf{V}_{\text{mean}}$, and solution

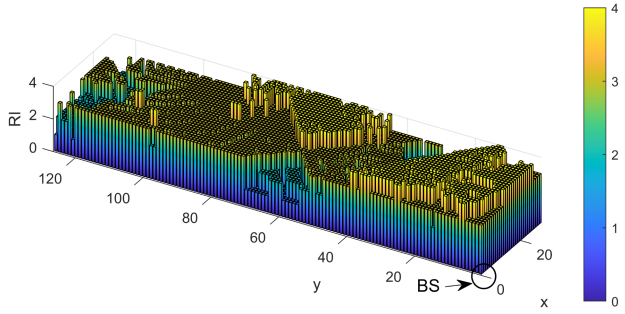


Fig. 12. Fixed RIs at locations in $\mathcal{H}$ when employing VAE-based solutions and $\mathfrak{C}_{\text{thold}} = 100$.

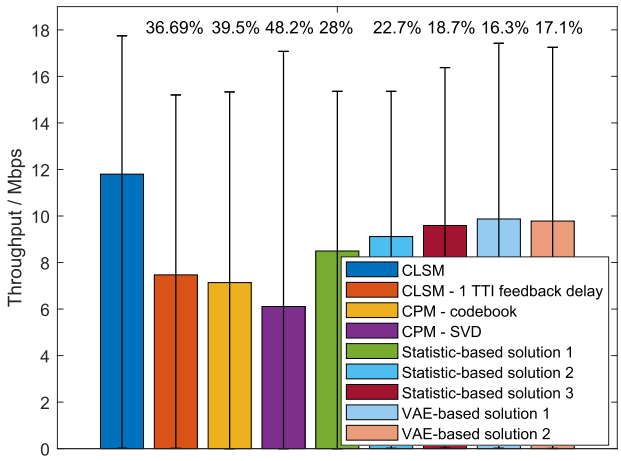
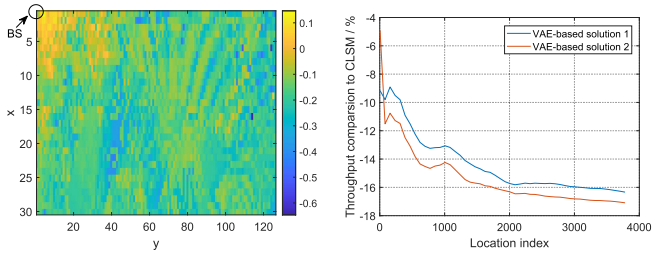


Fig. 13. Average throughput comparison among different transmission schemes at locations in $\mathcal{H}$ when $\mathfrak{C}_{\text{thold}}$ for VAE-based solutions is 100.

2 is $\mathbf{V}_{\text{KL}}$. It can be seen from Fig. 11 that a more conservative choice leads to less performance loss compared with CLSM. As a result, we set $\mathfrak{C}_{\text{thold}}$ to 100 to choose the fixed RI to obtain the best performance of VAE-based solutions. Fig. 12 shows the fixed RIs at locations of UEs in the training dataset when we set $\mathfrak{C}_{\text{thold}}$ to 100. From now on, BS is at coordinate origin to demonstrate the relative locations of UEs to it. In general, it is more likely to realize higher data stream transmission if UE's location is closer to BS. However, UEs which are located right under BS are unable to fully utilize the maximal streams to transmit data, since the channel matrices tend to be ill-conditioned.

3) Comparison at All Locations in Training Dataset: Fig. 13 shows the average throughput comparison of CLSM, CPM, and proposed solutions for data-driven MIMO at all locations in $\mathcal{H}$. CPM suffers from significant performance degradation since it fails to account for the time-varying characteristics of channels. Furthermore, although the SVD precoder is optimal for a specific channel, its highly precise beams can lead to greater fluctuations in time domain compared with the codebook when used in feedback-free MIMO transmission, if not carefully managed. This difference is evident in the performance of CPM-codebook and CPM-SVD. These results and analyses demonstrate the drawback of CPM-based approaches, which require to reconstruct the channel first. Instead, our proposed approaches



(a) Heatmap. (b) Line chart.

Fig. 14. Average throughput gap ratio of VAE-based solutions compared with CLSM at locations in $\mathcal{H}$.

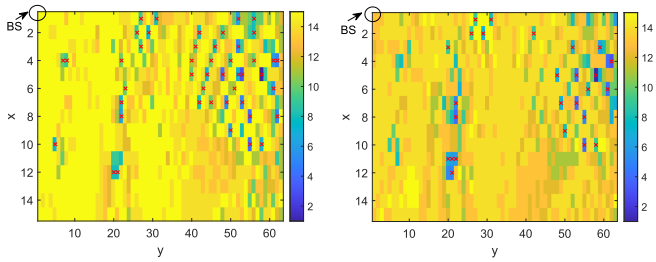
determine the transmission parameters in an end-to-end way. Due to the fixed transmission parameters in time domain, the best VAE-based solution exhibits the performance loss of 16.34%, while 18.7% for the best statistic-based solution. VAE-based solutions show superiority after a conservative choice of RI. Moreover, the performance improvement of VAE-based solutions at locations with the same fixed RI as statistic-based solutions surpasses the reduction at locations with higher RIs. Regarding training data requirements, the proposed codebook-based approach requires historical PMI, RI, and CQI data, whereas the SVD-based approach requires raw channel data to achieve enhanced performance. UEs can collect these data and upload them to the edge cloud during the actual communication, as the proposed approaches are designed to be compatible with the 5G system. In contrast, CPM needs ray-tracing data for the three strongest channel paths, including path gain, phase, elevation angle of departure (AoD), and azimuth AoD. This requires the digital twin of the signal propagation environment, which is not available in most places.

Fig. 14(a) presents the average throughput gap ratio of VAE-based solution 1 to CLSM, which intuitively shows that VAE-based solution can outperform CLSM if UE is close to BS, although it endures large performance loss at locations where channel states are poor. The location index starts at the one closest to BS and increases along the x-axis first and then along the y-axis. It can be seen from Fig. 14(b) that the farther the UE is to BS, the worse the performance of VAE-based solutions is compared with CLSM. In general, VAE-based solution 1 achieves approximately 1% higher average throughput than solution 2, except for locations near the BS.

C. Evaluation on Space Domain

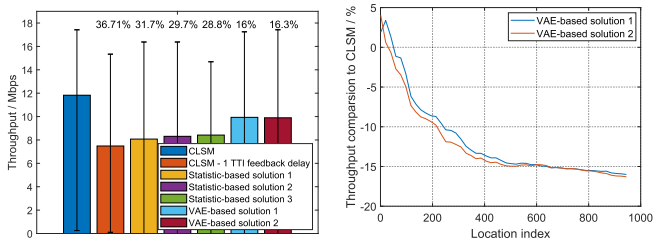
In this subsection, we evaluate the proposed solutions for data-driven MIMO in terms of spatial inference, namely to predict transmission parameters at location samples in the testing dataset $\mathcal{H}'$.

Fig. 15 shows the heatmap of the rounded mean value of true CQIs if we employ the precoders inferred by statistic-based solution 3 and VAE-based solution 1. We choose the fixed RIs at 4 closest observed locations in $\mathcal{H}$ to determine the interpolated RI at unobserved locations in $\mathcal{H}'$, i.e., $N_{\text{RI}} = 4$. In general, the fluctuation of CQI is localized and stable; however, anomalies



(a) Statistic-based solution 3. (b) VAE-based solution 1.

Fig. 15. Heatmap of the rounded mean value of true CQIs at locations in $\mathcal{H}'$ when employing different approaches for data-driven MIMO.



(a) Bar chart. (b) Line chart.

Fig. 17. Average throughput comparison among different transmission schemes at locations in $\mathcal{H}'$.

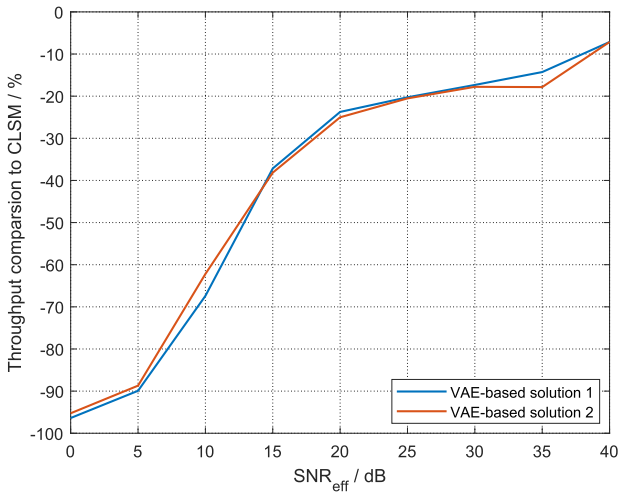


Fig. 16. Average throughput comparison between VAE-based solutions and CLSM at locations in $\mathcal{H}'$ with different SNRs_{eff}.

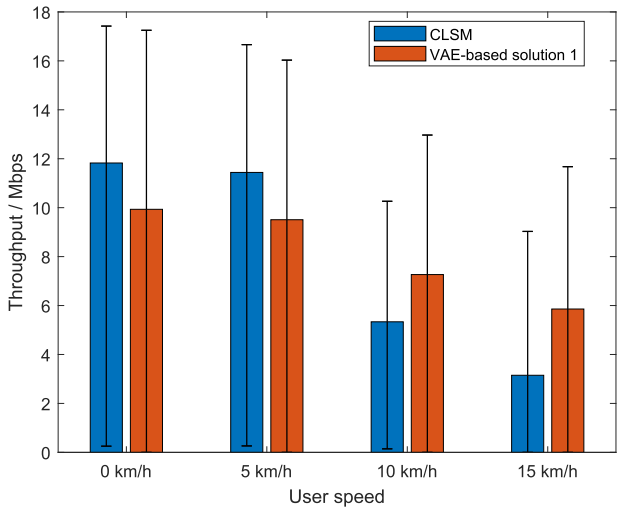


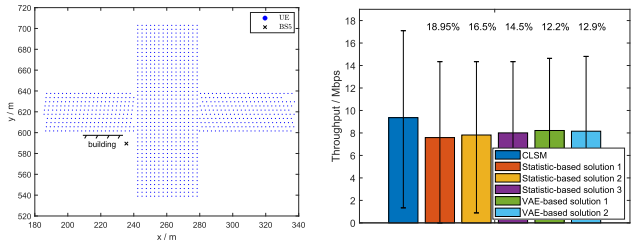
Fig. 18. Average throughput comparison between VAE-based solution 1 (trained with channel dataset of zero user speed) and CLSM at locations in $\mathcal{H}'$ when the user speed varies.

caused by sudden blockages makes it challenging to interpolate, even resulting in zero throughput at certain points, which are marked with a red cross in Fig. 15. The VAE-based solution utilizes GPR for spatial prediction on precoders, effectively capturing the underlying relationships among precoders at different locations. Thus, this solution leads to fewer zero-throughput locations than the statistic-based solution. As shown in Fig. 16, for unknown locations in the testing dataset $\mathcal{H}'$, the performance gap between the proposed approaches and CLSM is well acceptable when SNR_{eff} is high. However, these approaches face limitations at a few low SNR_{eff} locations. As illustrated in Fig. 15, these locations usually experience severe path loss due to abrupt blockages. Such anomalies are difficult to track for interpolation methods at the current spatial granularity of the original dataset. In our future work, to deal with the issue, these rare low-performance locations can be specially treated or fine-tuned. Additionally, instead of using spatial multiplexing, spatial diversity techniques such as space-time coding can be used, which are also feedback-free.

As illustrated in Fig. 17(a), VAE-based solutions exhibit significant performance gains over statistic-based solutions in predicting transmission parameters at $\mathcal{Q}'$. These solutions achieve an improvement of over 12% compared with the best statistic-based solution, due to the effective spatial beam inference at unknown locations enabled by the combination of VAE and GPR. Additionally, the performance loss of VAE-based solutions in the testing set (Fig. 17(a)) is comparable to that in the training set (Fig. 13), indicating their superiority over statistic-based solutions in spatial inference. The throughput comparison to CLSM in Fig. 17(b) shows a similar trend as that in Fig. 14(b). Overall, the proposed VAE-based solutions experience a performance loss of approximately 16% compared with 5G CLSM. Considering the saved overheads of pilots and feedback, VAE-based solutions for data-driven MIMO can potentially achieve comparable or even higher performance than 5G CLSM.

D. Robustness and Generalization

We expand the testing channel set $\mathcal{H}'$ to incorporate channels of various user speeds to validate the robustness of the proposed approach. Fig. 18 illustrates the throughput comparison as user speed varies. Note that VAE-based solution 1 is trained using a channel dataset at zero user speed, while the results for CLSM are obtained using channel datasets at various user speeds. When the user is stationary, the channels remain consistent across all 14 symbols in a subframe. However, as



(a) Locations of UEs and BS. (b) Throughput comparison.

Fig. 19. Locations of selected UEs and BS in the new channel dataset and corresponding average throughput comparison among different transmission schemes.

user speed increases, variations among the channels at different symbols lead to performance degradation for both CLSM and the proposed approach. Since CLSM computes transmission parameters based on channels at pilot symbols, these parameters fail to match throughout the subframe when user speed becomes large, causing a significant drop in performance. Conversely, the proposed VAE-based solution, despite being trained with a dataset at a fixed user speed, demonstrates robustness across various user speeds due to its use of representative transmission parameters. These results illustrate the strong support for user mobility provided by our proposed approach, whereas CPM is only suitable for static or periodically changing environments [32].

We further introduce a new channel dataset that uses *BS5* to serve 1138 UEs at the crossroads in DeepMIMO's *O1* scenario, as depicted in Fig. 19(a). In this scenario, a building obstructs the LOS path of a specific area and more NLOS paths exist. We apply the proposed approaches to derive feedback-free transmission parameters for BS5, with the results displayed in Fig. 19(b). The performance trends of the proposed approaches demonstrate similar results as in Fig. 13, indicating their strong generalization ability across diverse environments. This consistency is expected, as the data-driven approach can effectively learn the underlying propagation environment from the data.

VII. Conclusion

In this paper, we have studied how to implement end-to-end data-driven MIMO for both FD-RAN and future 6G. It requires no channel feedback and only utilizes geolocation to determine all the transmission parameters based on the mapping learned from historical channel data. First, an approach based on 5G Type I codebook is proposed, which selects mode values of historical PMIs, RIs, and CQIs and uses nearest neighbor interpolation for spatial inference. Then, an SVD-based approach is proposed for performance improvement, where VAE is utilized to choose the representative precoder in time domain. GPR and NNI are exploited to enable higher spatial prediction accuracy with respect to the precoder and CQI. Simulation results based on a 5G-compatible link-level simulator and realistic ray-tracing channel data manifest the effectiveness of the proposed approaches, the superiority of the SVD-based approach in spatial inference, and the potential of data-driven MIMO. In our future work, we will consider the multi-user scenario and utilize other generative artificial intelligence techniques to further realize the capabilities of data-driven MIMO.

References

[1] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tut.*, vol. 25, no. 2, pp. 905–974, Second Quarter 2023.
[2] Q. Yu et al., "A fully-decoupled RAN architecture for 6G inspired by neurotransmission," *J. Commun. Inf. Netw.*, vol. 4, no. 4, pp. 15–23, Dec. 2019.
[3] K. Yu et al., "Fully-decoupled radio access networks: A flexible downlink multi-connectivity and dynamic resource cooperation framework," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 4202–4214, Jun. 2023.
[4] J. Zhao et al., "Fully-decoupled radio access networks: A resilient uplink base stations cooperative reception framework," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5096–5110, Aug. 2023.
[5] B. Qian et al., "Enabling fully-decoupled radio access with elastic resource allocation," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 4, pp. 1025–1040, Aug. 2023.
[6] J. Liu, J. Chen, C. He, and H. Zhou, "Leveraging load-aware dynamic pricing for cell-level demand-supply equilibrium," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6902–6906, May 2023.
[7] Y. Xu, B. Qian, K. Yu, T. Ma, L. Zhao, and H. Zhou, "Federated learning over fully-decoupled RAN architecture for two-tier computing acceleration," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 789–801, Mar. 2023.
[8] J. Xue, K. Yu, T. Zhang, H. Zhou, L. Zhao, and X. Shen, "Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 8, pp. 8143–8157, Aug. 2024.
[9] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures (Release 16)," 3GPP TS 36.213 V16.2.0, Jun. 2022.
[10] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
[11] A. Bletsas, A. Lippman, and J. N. Sahalos, "Simple, zero-feedback, distributed beamforming with unsynchronized carriers," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1046–1054, Sep. 2010.
[12] K. Alexandris, G. Sklivanitis, and A. Bletsas, "Reachback WSN connectivity: Non-coherent zero-feedback distributed beamforming or TDMA energy harvesting?," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4923–4934, Sep. 2014.
[13] U. D. Suleiman, M. Esa, N. N. N. Abd Malik, K. M. Yusof, M. F. M. Yusoff, and M. R. Hamid, "A review on frequency synchronization in collaborative beamforming: A practical approach," *J. Adv. Res. Appl. Mech.*, vol. 31, no. 1, pp. 1–15, May 2017.
[14] G. Sklivanitis, K. Alexandris, and A. Bletsas, "Testbed for non-coherent zero-feedback distributed beamforming," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 2563–2567.
[15] T. T. Nguyen and K.-K. Nguyen, "A deep learning framework for beam selection and power control in massive MIMO-millimeter-wave communications," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4374–4387, Aug. 2023.
[16] 3GPP, "NR; physical layer procedures for data (Release 16)," 3GPP TS 38.214 V16.10.0, Jun. 2022.
[17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
[18] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8017–8045, Dec. 2022.
[19] 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.
[20] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.
[21] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, Apr. 2020.
[22] Z. Lu, J. Wang, and J. Song, "Multi-resolution CSI feedback with deep learning in massive MIMO system," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

[23] M. Chen, J. Guo, C.-K. Wen, S. Jin, G. Y. Li, and A. Yang, "Deep learning-based implicit CSI feedback in massive MIMO," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 935–950, Feb. 2022.
[24] X. Li, J. Guo, C.-K. Wen, and S. Jin, "Auto-CsiNet: Scenario-customized automatic neural network architecture generation for massive MIMO CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14759–14775, Oct. 2024, doi: 10.1109/TWC.2024.3418907.
[25] D. Vasisht, S. Kumar, H. Rahul, and D. Katabi, "Eliminating channel feedback in next-generation cellular networks," in *Proc. ACM SIGCOMM Conf.*, 2016, pp. 398–411.
[26] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *Proc. 53rd Asilomar Conf. Signals Syst. Comput.*, 2019, pp. 1465–1470.
[27] W. B. Chikha, M. Masson, Z. Altman, and S. B. Jemaa, "Radio environment map based inter-cell interference coordination for massive-MIMO systems," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 785–796, Jan. 2024.
[28] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1581–1592, Mar. 2020.
[29] S. Hanna, E. Krijestorac, and D. Cabric, "Destination-feedback free distributed transmit beamforming using guided directionality," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5858–5869, Oct. 2023.
[30] W. Wang, B. Yang, and W. Zhang, "Deep learning-based radio map for MIMO-OFDM downlink precoding," *J. Commun. Inf. Netw.*, vol. 8, no. 3, pp. 203–211, Sep. 2023.
[31] J. Zhao, J. Chen, Z. Sun, Y. Shi, H. Zhou, and X. S. Shen, "Channel-feedback-free transmission for downlink FD-RAN: A radio map based complex-valued precoding network approach," *China Commun.*, vol. 21, no. 4, pp. 10–22, Apr. 2024.
[32] Y. Zeng and X. Xu, "Toward environment-aware 6G communications via channel knowledge map," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 84–91, Jun. 2021.
[33] Z. Liu, G. Singh, C. Xu, and D. Vasisht, "FIRE: Enabling reciprocity for FDD MIMO systems," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, 2021, pp. 628–641.
[34] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
[35] K. Brueninghaus et al., "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE 16th Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2005, pp. 2306–2311.
[36] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," *Electron. Lett.*, vol. 32, no. 12, pp. 1060–1061, 1996.
[37] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *Proc. 6th Conf. Wireless Adv.*, 2010, pp. 1–6.
[38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[39] I. Higgins et al., "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. Int. Conf. Learn. Representation*, 2016, pp. 1–22.
[40] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[41] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: A survey," *J. Electron. Imag.*, vol. 11, no. 2, pp. 157–176, Apr. 2002.
[42] S. Pratschner et al., "Versatile mobile communications simulation: The Vienna 5G link level simulator," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, pp. 1–17, Sep. 2018.
[43] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," 2019, *arXiv: 1902.06435*.
[44] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," 3GPP TR 38.901 V16.1.0, Dec. 2019.
[45] 3GPP, "NR; radio resource control (RRC); protocol specification (Release 17)," 3GPP TS 38.331 V17.0.0, Mar. 2022.
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[47] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv: 1802.03426*.

Jingbo Liu received the BS degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2020. He is currently working toward the PhD degree with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. He is also an intern with Pengcheng Laboratory. His current research interests include resource management in wireless networks, reinforcement learning, and fully-decoupled radio access network.

Jiacheng Chen (Member, IEEE) received the PhD degree in information and communications engineering from Shanghai Jiao Tong University, Shanghai, China, in 2018. From 2015 to 2016, he was a visiting scholar with BBCR Group, University of Waterloo, Canada. Currently, he is an assistant researcher with Pengcheng Laboratory, Shenzhen, China. His research interests include fully-decoupled radio access network technologies. He has served as the guest editor of the *IEEE Internet of Things Journal*, and the Workshop co-chair for IEEE/CIC ICCC from 2021 to 2024. He was the recipient of the *Journal of Communications and Information Networks* (JCIN) Best Paper Award in 2016, and the Chinese Institute of Electronics (CIE) Outstanding Scientific Paper in the field of electronic information in 2020.

Zongxi Liu (Student Member, IEEE) received the BS degree in electrical engineering and automation from Nanjing Normal University, Nanjing, China, in 2022. He is currently working toward the MS degree in communication engineering with Nanjing University, Nanjing. His research interests include transmission scheme for B5G/6G networks and machine learning for wireless communications.

Haibo Zhou (Senior Member, IEEE) received the PhD degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a postdoctoral fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo. He is currently a full professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. He was a recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award, 2023-2024 IEEE ComSoc distinguished lecturer, and 2023-2025 IEEE VTS distinguished lecturer. He served as track/symposium co-chair for IEEE/CIC ICCC 2019, IEEE VTC-Fall 2020, IEEE VTC-Fall 2021, WCSP 2022, IEEE GLOBECOM 2022, IEEE ICC 2024, IEEE GLOBECOM 2024. He is currently an associate editor of *IEEE Transactions on Wireless Communications*, *IEEE Internet of Things Journal*, *IEEE Network Magazine*, and *Journal of Communications and Information Networks*. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks.