A Variational Autoencoder Enabled Feedback-Free MIMO Transmission Approach for FD-RAN

Jingbo Liu*†, Jiacheng Chen†, Zongxi Liu‡, Yunting Xu‡, Jiawen Kang§, and Haibo Zhou‡
*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240 China
†Department of Strategic and Advanced Interdisciplinary Research, Pengcheng Laboratory, Shenzhen 518000, China
‡School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
§School of Automation, Guangdong University of Technology, Guangzhou 510006, China
E-mail: liujingbo@sjtu.edu.cn, chenjch02@pcl.ac.cn,
{zongxiliu,yuntingxu}@smail.nju.edu.cn, kavinkang@gdut.edu.cn, haibozhou@nju.edu.cn

GLOBECOM 2024 - 2024 IEEE Global Communications Conference | 979-8-3503-5125-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBECOM52923.2024.10901000

***Abstract*—To enhance flexibility and facilitate resource cooperation, a novel fully-decoupled radio access network (FD-RAN) architecture has been proposed. The decoupling of uplink and downlink in FD-RAN renders the current channel feedback mechanism ineffective, particularly when factoring in the feedback overheads and delays inherent in realistic scenarios. To this end, we investigate the feedback-free MIMO spatial multiplexing transmission in FD-RAN. Specifically, we generate a mapping from geolocation to MIMO transmission parameters from the historical channel data. We first obtain optimal precoders from singular value decomposition (SVD) of channel data. Then, a variational autoencoder (VAE) is trained using the optimal precoders, and the representative precoder for each geolocation is selected from the latent Gaussian representations of VAE. Simulations are performed on a link-level simulator using ray-tracing channel data, and the results demonstrate the effectiveness of our scheme, showcasing its feasibility for adoption in FD-RAN.**

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

I. Introduction

COMMERCIALIZATION of the fifth generation (5G) mobile network is ongoing worldwide, yet some fundamental challenges have not been fully addressed. Firstly, the scarce spectrum resources are still not efficiently utilized. Secondly, it is still difficult to guarantee users' quality of experience (QoE). Thirdly, the costs of network infrastructure and operation are very high. To address the above issues, the fully-decoupled radio access network (FD-RAN) architecture has been proposed for 5G and the upcoming sixth generation (6G) mobile network [1]. It decouples the base station (BS) functionalities into control BS (C-BS), downlink BS (DL-BS), and uplink BS (UL-BS). Then, control and resource allocation flexibility can be exploited to meet the requirements of 6G.

However, such a decoupled architecture makes the existing feedback schemes employed in 5G ineffective. For example, in conventional DL transmission, BS first transmits channel state information (CSI) reference signals (CSI-RSs) to user equipment (UE). Then, UE estimates the channel and calculates the transmission parameters, namely the CSI, which are fed back to the BS for carrying out multiple-input and multiple-output (MIMO) transmission. In closed-loop spatial multiplexing (CLSM) transmission mode [2], CSI includes precoder, the number of spatial streams, and modulation and coding scheme (MCS). In FD-RAN, DL-BS is unable to obtain the CSI feedback from UE directly, since UL-BSs and DL-BSs are physically separated. The feedback delay will severely affect the accuracy of CSI and MIMO performance. Furthermore, as the number of transmit antenna increases, the overheads of CSI-RSs and CSI become much higher, effectively reducing the resources for transmitting user data. Thus, implementing MIMO transmission without CSI feedback is a significant challenge for FD-RAN.

To this end, we investigate the feedback-free MIMO transmission in FD-RAN. Instead of relying on feedback to acquire channel information, we exploit the knowledge learned from historical channel data and only use UE's geolocation to determine transmission parameters. Intuitively, such a data-driven approach is feasible because the channel and the corresponding transmission parameters are strongly correlated with geolocation, when the signal propagation environment does not change dramatically. Although the channel is generally time-varying, we expect the performance loss can be mitigated by saving resources from pilots and feedback. Other implicit feedback and zero feedback transmission schemes have also been studied in [3]–[5].

In this paper, we aim to realize the same capability as CLSM, namely MIMO spatial multiplexing, using transmission parameters derived solely from geolocation. Existing works mainly concentrate on beamforming and use only one data stream [6]–[8]. For this purpose, we need to jointly choose transmission parameters, namely precoder, rank indicator (RI), and channel quality indicator (CQI). The precoder and RI can be determined at the same time, since RI indicates the rank of precoder. Then, proper CQI can be derived based on the chosen precoder and RI. An improper choice of these parameters would lead to severe performance degradation, considering RI and CQI are quantized discrete values.

To address the above challenges, we propose a non-codebook-based approach to realize feedback-free MIMO transmission for FD-RAN DL. Although 5G Type I codebook [9] provides a set of pre-defined precoders, they are not optimal due to quantization. Thus, we utilize the optimal precoders calculated from singular value decomposition (SVD) of the channel matrix. Then, we exploit variational autoencoder (VAE)

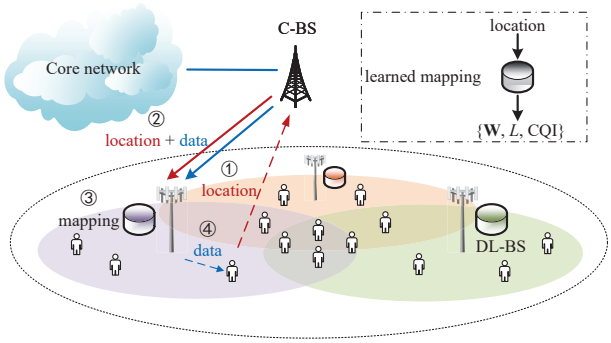


Fig. 1. The feedback-free downlink MIMO transmission in FD-RAN.

[10] to discover the latent representations of optimal precoders in lower dimensions, and find the appropriate transmission parameters used for each location. Note that we consider transmission on narrow bands, 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 feedback-free approach to realize MIMO spatial multiplexing for FD-RAN. The approach is data-driven in that it learns from the historical channel data of BS and only utilizes geolocation to determine all the transmission parameters.
- We formulate the problem of feedback-free MIMO transmission, and a VAE-based spatial multiplexing transmission approach is proposed, in which VAE is used to select the transmission parameters from latent Gaussian representations for each geolocation.
- We perform simulations based on a high-fidelity link-level simulator and realistic ray-tracing channel data. The results demonstrate the effectiveness of the proposed approach. Compared with feedback-based 5G CLSM, we showcase the same level of throughput performance, considering the saved overheads of pilots and feedback.

The remainder of this paper is organized as follows. The system model and problem formulation are given in Section II. The VAE-based approach is presented in Section III. Simulation results and discussions are given in Section IV. Finally, Section V concludes the whole paper.

II. System Model and Problem Formulation

A. System Model

In this paper, we study the DL single-user MIMO (SU-MIMO) transmission of FD-RAN. Fig. 1 illustrates the system model, where a control BS coordinates with multiple DL-BSs and UEs. The control BS obtains the locations reported by users and informs the DL-BS. Then, exploiting 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 control 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) can be utilized to sense the location of UE. In this study, we only focus on the performance at static user locations.

The whole signal processing procedure is based on orthogonal frequency division multiplexing (OFDM) and MIMO technologies. 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. $\boldsymbol{n}_k \sim \mathcal{CN}(\mathbf{0}, \varsigma_n^2 \mathbf{I})$ is the additive zero means complex-valued white Gaussian noise with variance ς_n^2. $\mathbf{W} \in \mathbb{C}^{N_{tx} \times L}$ is the employed precoding matrix and $L \in \{1, \ldots, \Upsilon\}$ is the number of employed layers for spatial multiplexing. 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. RI corresponds to the rank of precoder $\mathbf{W}$, which is equal to 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 $\mathbf{G}_k$ 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 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, which has similar transmission performance with the MIMO OFDM system. The mapping is defined as

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

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. 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.

B. Problem Formulation

In essence, the problem of feedback-free MIMO is to estimate a mapping from geolocation to transmission parameters, i.e., $\boldsymbol{\varrho} = (x, y, z) \rightarrow \{\mathbf{W}, L, \mathrm{CQI}\}$ 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.

We consider the coverage area of a DL-BS and we assume that the environment will not change significantly over the studied long time span. Specifically, the 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 location index, corresponding to the channel sample at a certain geolocation. Therefore, $|\mathcal{H}| = T \times K \times Q$.

In feedback-free MIMO, utilizing the mapping derived from the channel data, we aim to maximize the throughput at all time and location samples in $\mathcal{H}$. The problem can be formulated as

$$\max_{\{\mathbf{W}_q, L_q, \mathrm{CQI}_q\}} \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}} \mathfrak{T}(\mathbf{H}_{t,k,q}, \{\mathbf{W}_q, L_q, \mathrm{CQI}_q\})$$
$$\text{s.t.} \begin{cases} \mathbf{H}_{t,k,q} \in \mathcal{H}, & \forall t \in \mathcal{T}, \forall k \in \mathcal{K}, \forall q \in \mathcal{Q} \\ \|\mathbf{W}_q\|_F^2 = 1, & \forall q \in \mathcal{Q} \\ L_q = r(\mathbf{W}_q) \in \{1, \ldots, \Upsilon\}, & \forall q \in \mathcal{Q} \\ \mathrm{CQI}_q \in \{1, \ldots, 15\}, & \forall q \in \mathcal{Q}, \end{cases} \quad (5)$$

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 [9]. 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.

III. VAE-based Approach

A. Optimal Precoders Based on SVD

The quantization of spatial beams in the 5G Type I codebook brings in loss of accuracy of the precoder, yet the optimal precoder can be calculated from the SVD of the channel for it eliminates the interference among layers. The SVD of $\mathbf{H}_k$ can be expressed as $\mathbf{H}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{V}_k^H$, where $\mathbf{U}_k$ and $\mathbf{V}_k$ are the left and right unitary matrices, respectively. $(\cdot)^H$ denotes the Hermitian transpose operation. Then, the optimal precoder $\mathbf{V}_k^L$ can be obtained as the first L columns of $\mathbf{V}_k$.

The mutual information when precoder $\mathbf{W}$ is employed can be expressed as $I_{k,\ell}(\mathbf{W}) = \log_2(1 + \mathrm{SINR}_{k,\ell}(\mathbf{W}))$, where $k \in \mathcal{K}$, $\ell \in \mathcal{L}$, and $\mathrm{SINR}_{k,\ell}(\mathbf{W})$ indicates the $\mathrm{SINR}_{k,\ell}$ defined in (3) if precoder $\mathbf{W}$ is adopted. For 5G CLSM, 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 [11]. Similar to CLSM, the best precoder from SVD of the channel at any time and location indexes can be chosen as follows:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}_i} \sum_{k\in\mathcal{K}}\sum_{\ell\in\mathcal{L}} I_{k,\ell}(\mathbf{W}_i)$$
$$\text{s.t.} \quad \mathbf{W}_i = \frac{\mathbf{V}_k^L}{\|\mathbf{V}_k^L\|_F} \in \mathcal{W}_{\mathrm{SVD}}, \ \forall\, i \in \{1, \ldots, |\mathcal{W}_{\mathrm{SVD}}|\}, \quad (6)$$
$$\forall k \in \mathcal{K}, \ \forall L \in \{1, \ldots, \Upsilon\},$$

where $\mathcal{W}_{\mathrm{SVD}}$ is the set of precoders obtained from the SVD of the channel at a certain location, and $|\mathcal{W}_{\mathrm{SVD}}| = K \times \Upsilon$. $\mathbf{W}^*$ 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 is the rank of $\mathbf{W}^*$. CQI is then mapped from the effective SNR defined in (4) given the precoder and RI.

B. Selecting Transmission Parameters

Autoencoder consists of an encoder and a decoder, which are constructed by multiple layers of neural networks. The encoder compresses the initial inputs and the decoder recovers them from lower-dimensional latent space. VAE is introduced in [10] 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 sampled from latent space to recover the inputs. 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 understanding the high dimensional precoders and fixing them in time domain. The proposed VAE architecture is shown in Fig. 2.

Specifically, $\boldsymbol{d}_i$ is denoted as the i-th independent and identically distributed sample of dataset $\mathcal{D}_{\mathrm{VAE}}$ for VAE, where $i \in \{1, \ldots, |\mathcal{D}_{\mathrm{VAE}}|\}$. The encoder network's parameters are denoted as $\boldsymbol{\psi}$, and the decoder network's as $\boldsymbol{\xi}$. The latent space of VAE is represented by the latent variable $\boldsymbol{\varpi}$ with dimension $N_{lv} \times 1$. The encoder maps the input $\boldsymbol{d}_i$ into vectors of mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\sigma}_i$ 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

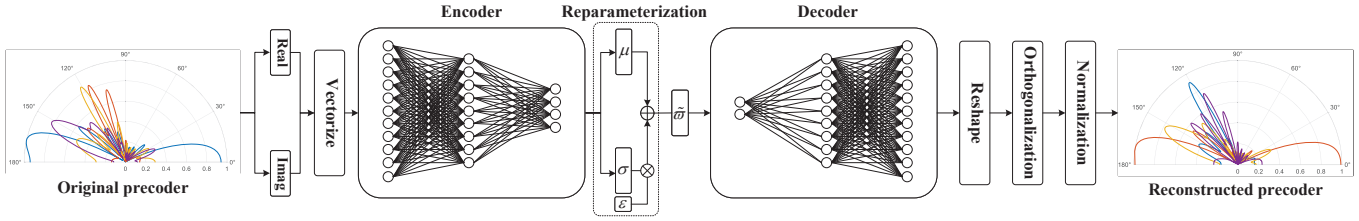


Fig. 2. The proposed VAE architecture to deal with precoders.

trick is used in VAE to make the back-propagation of the error terms possible. Then, the sampled latent variable can be expressed as $\tilde{\boldsymbol{\varpi}} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\odot$ denotes the element-wise product. Denote the output of VAE as $\widehat{\boldsymbol{d}}_i$, which can be viewed as the estimate of input $\boldsymbol{d}_i$. Utilizing the evidence lower bound on the log-likelihood [10], the loss function of VAE can be defined as

$$\text{Loss} = \text{Loss}_{\text{recon}} + \beta \text{Loss}_{\text{KL}}, \tag{7}$$

where the first term indicates the reconstruction loss and equals to $||\boldsymbol{d}_i - \widehat{\boldsymbol{d}}_i||^2$ in squared error form. The second term refers to the Kullback-Leibler (KL) divergence loss between the posterior approximation and the prior of latent variable, which yields $\frac{\beta}{2}\sum_{j=1}^{N_{lv}}((\boldsymbol{\mu}_i)_j^2 + (\boldsymbol{\sigma}_i)_j^2 - \log((\boldsymbol{\sigma}_i)_j^2) - 1)$ in Gaussian case, where $(\boldsymbol{\mu}_i)_j$ and $(\boldsymbol{\sigma}_i)_j$ are j-th element of vectors $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$. β is an adjustable hyperparameter to balance the tradeoff between reconstruction loss and KL divergence loss.

Since precoders of different layers have different dimensions, Υ VAEs are required to be trained corresponding to distinct input dimensions. However, how to prepare the input precoder dataset remains a problem. Basically, the optimal precoders $\mathbf{W}_{t,q}^*$ at any time t and location q can be derived from (6). If we fix the rank of precoder L obtained from SVD of channel to $\mathfrak{r}$ in (6), we can obtain "optimal" precoders of rank $\mathfrak{r}$, denoted as $\mathbf{W}^{*,\mathfrak{r}}$. Then, the chosen indexes of time and location for precoders with higher ranks are also applicable to precoders with lower ranks. The downward compatible set of layer choice for precoders of rank $\mathfrak{r}$ is denoted as $\mathfrak{L}$, where $\mathfrak{L} = \{\mathfrak{r}, \ldots, \Upsilon\}$ and $\mathfrak{r} \in \{1, \ldots, \Upsilon\}$. Specifically, the time and location indexes for the selected precoders $\mathbf{W}^{*,\mathfrak{r}}$ can be determined by applying constraints to the original optimal precoders $\mathbf{W}^*$. First, the location index set $\mathcal{Q}^{\mathfrak{r}}$ for precoders of rank $\mathfrak{r}$ is chosen as $\mathcal{Q}^{\mathfrak{r}} = \{q|\text{mode}(\mathcal{R}_q(\mathbf{W}_{\mathcal{T},q}^*)) \in \mathfrak{L}, q \in \mathcal{Q}\}$, where $\mathcal{R}_q(\mathbf{W}_{\mathcal{T},q}^*) = \{r(\mathbf{W}_{t,q}^*)|t \in \mathcal{T}\}$ is the rank set of optimal precoders at location index q and mode$(\cdot)$ indicates the mode value of the set. Next, time index set $\mathcal{T}_q^{\mathfrak{r}}$ at location q for precoders of rank $\mathfrak{r}$ is selected 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 dataset of VAE $\mathcal{D}_{\text{VAE}}^{\mathfrak{r}}$ for precoders of rank $\mathfrak{r}$ can be expressed as $\mathcal{D}_{\text{VAE}}^{\mathfrak{r}} = \{\mathbf{W}_{t,q}^{*,\mathfrak{r}}|q \in \mathcal{Q}^{\mathfrak{r}}, t \in \mathcal{T}_q^{\mathfrak{r}}\}$.

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 VAE dataset. Besides, an orthogonalization operation is performed to the recovered precoder $\widehat{\mathbf{V}}^*$ to make it satisfy the property of the unitary matrix. The orthogonalization operation is done by QR decomposition of $\widehat{\mathbf{V}}^*$, which is given by $\widehat{\mathbf{V}}^* = \widehat{\mathbf{V}}_{\text{Q}}^*\widehat{\mathbf{V}}_{\text{R}}^*$. $\widehat{\mathbf{V}}_{\text{Q}}^*$ is an orthogonal matrix and $\widehat{\mathbf{V}}_{\text{R}}^*$ is an upper triangular matrix. Finally, $\widehat{\mathbf{V}}_{\text{Q}}^*/\|\widehat{\mathbf{V}}_{\text{Q}}^*\|_F$ is decided as the generated or recovered precoder, with power normalized and columns orthogonalized.

In this paper, we fix transmission parameters at locations in $\mathcal{H}$ to evaluate the performance loss of feedback-free MIMO. We need to first fix RI such that the representative precoder of a specific rank can be chosen. Specifically, we define a percentage count threshold $\mathfrak{C}_{\text{thold}}$. If no fewer than $\mathfrak{C}_{\text{thold}}$ percents of T RIs obtained from (6) at location q are elements in $\mathfrak{L}$ when the mode of $\mathcal{R}_q(\mathbf{W}_{\mathcal{T},q}^*)$ is $\mathfrak{r}$, then L_q is fixed to $\mathfrak{r}$, otherwise $\mathfrak{r}-1$. Thus, apart from locations where the highest rank of precoder is employed, the set of location $\mathcal{Q}_{\text{fixed}}^{\mathfrak{r},\mathfrak{C}_{\text{thold}}}$ where the fixed RI is $\mathfrak{r}$ consists of two parts: the indexes that satisfy the constraints described above when the mode of $\mathcal{R}_q(\mathbf{W}_{\mathcal{T},q}^*)$ is $\mathfrak{r}$ and those that don't when the mode of $\mathcal{R}_q(\mathbf{W}_{\mathcal{T},q}^*)$ 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.

Exploiting the latent variables obtained by the encoder of VAE, the precoder at location q can be fixed, as illustrated in Fig. 3. Specifically, the representative latent variable $\boldsymbol{\varpi}_q^{\star,\mathfrak{r}}$ at location q for precoders of rank $\mathfrak{r}$ can be chosen based on its proximity to the mean values of the mean and variance vectors of all latent variables at location q. The time index for the representative latent variable can be selected as follows:

$$\begin{aligned}\arg\min_t [&(\boldsymbol{\mu}_{t,q}^{\mathfrak{r}} - \bar{\boldsymbol{\mu}}_q^{\mathfrak{r}})^{\text{T}}(\boldsymbol{\mu}_{t,q}^{\mathfrak{r}} - \bar{\boldsymbol{\mu}}_q^{\mathfrak{r}}) + \\ &((\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}})^2 - (\bar{\boldsymbol{\sigma}}_q^{\mathfrak{r}})^2)^{\text{T}}((\boldsymbol{\sigma}_{t,q}^{\mathfrak{r}})^2 - (\bar{\boldsymbol{\sigma}}_q^{\mathfrak{r}})^2)], \forall t \in \mathcal{T}_q^{\mathfrak{r}},\end{aligned} \tag{8}$$

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. Finally, $\boldsymbol{\mu}_q^{\star,\mathfrak{r}}$ is input into the well-trained decoder to obtain the representative precoder.

At location q, the CQI set $\mathbf{CQI}_q(\mathbf{W}_q) = \{\text{CQI}_{t,q}(\mathbf{W}_q)|t \in \mathcal{T}\}$ can be determined when the fixed RI and precoder obtained from VAE are employed. Then, the fixed CQI is decided as the floor of the mean value of $\mathbf{CQI}_q(\mathbf{W}_q)$ since fixed SVD precoder may lead to more fluctuations in time domain.

IV. Simulation Results and Discussions

A. Simulation Setup

1) Environment: In this study, throughput is obtained by signal processing in the physical layer based on the Vienna

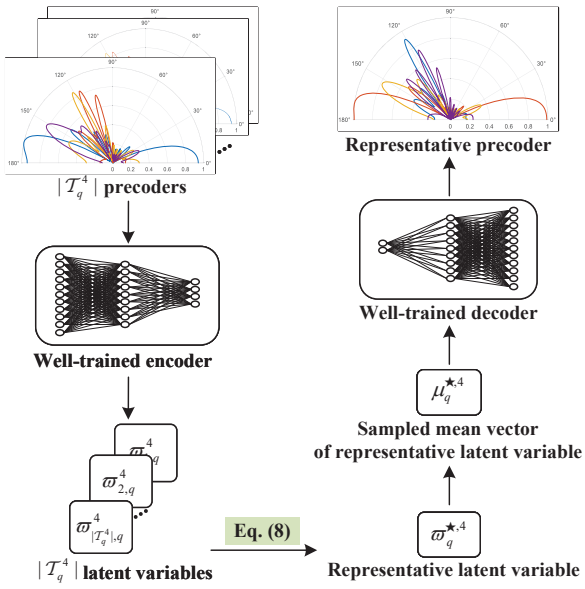


Fig. 3. Illustration of selecting the representative precoder in time domain utilizing VAE (with precoders of rank 4 as an example).

5G link level simulator [12]. BS is equipped with $N_{tx} = 16$ dual polarized transmit antennas with horizontal and vertical configurations $(N_1, N_2) = (8, 1)$ and UE is equipped with $N_{rx} = 4$ single polarized receive antennas. Υ is given by $\min\{N_{tx}, N_{rx}\} = 4$. A total of $K = 72$ subcarriers are used to carry MIMO transmission for UE. The center frequency is 3.5 GHz and the subcarrier spacing is 15 kHz. Low-density parity-check (LDPC) channel coding is employed and the equalizer is chosen based on minimum mean square error (MMSE).

Meanwhile, simulations are conducted based on the realistic ray-tracing dataset: DeepMIMO [13]. Using accurate ray-tracing data obtained from Remcom Wireless InSite, cluster delay line (CDL) channels between BS and UE are generated, serving as the dataset in this paper. We focus on using *BS1* to transmit data to UEs in *User Grid 1* of *O1* scenario in DeepMIMO. 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 distinct user locations. Thus, $|\mathcal{Q}| = 3780$. Besides, 100 different samples of channels in time domain are considered.

2) VAE: 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, 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 study, 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.

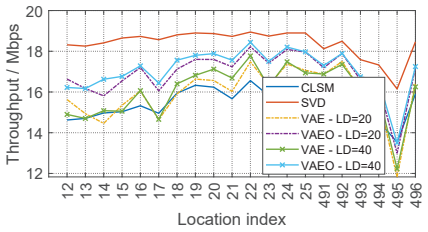
During the training process, a minibatch of 128 samples is used to calculate the gradients. β in (7) is chosen to be 0.01. Then, the parameters of networks in VAE are updated for 100 epochs using the adaptive moment estimation (ADAM) optimization algorithm with 10^{-3} learning rate. Despite fine-tuning may be helpful, we found that the proposed VAE approach works well under many values of hyperparameters.

B. Evaluations and Discussions

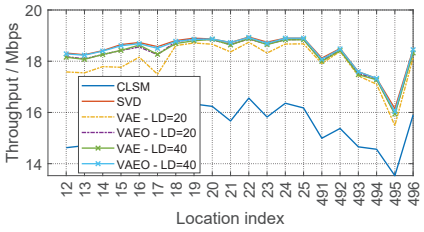
In the following, we first verify the proposed VAE-based approach at location samples suitable for transmitting four data streams and then extend to all locations in $\mathcal{H}$. Here, samples suitable for transmitting four streams refer to locations that utilize precoders of rank 4 at all time indexes based on the results of SVD obtained from (6). We compare the performance of the proposed VAE-based approach with feedback-based 5G CLSM, which is recognized for achieving the highest capacity among various transmission modes in 5G.

1) Verification of VAE-based Approach: Fig. 4 shows the average throughput comparison when the VAE dataset is made up of non-unitary matrices or unitary ones. Note that the SVD-based precoding and the proposed scheme with CQI feedback are implemented in Fig. 4 to demonstrate the reconstruction performance of VAE, where *LD* refers to latent dimension, and *VAEO* indicates VAE with orthogonalization. Employing precoders of unitary matrices as the VAE dataset shows significantly better reconstruction performance than directly using the energy normalized precoders since the values of normalized ones are too small. Besides, the orthogonalization operation offers large performance gains since it keeps the multiple streams orthogonalized to each other to reduce interference. As stated before, the proper choice of RI is critical to this problem. In Fig. 5, we assess the performance loss of the VAE-based approach compared to Type I codebook-based CLSM when the criterion for RI selection changes. As a result, we set $\mathfrak{C}_{\text{thold}}$ to 100 to fix RI so that the VAE-based approach can obtain the best performance. Note that the performance loss does not take the saved overheads of pilots and feedback into consideration.

2) Comparison at All Locations in the Dataset: Fig. 6(a) demonstrates the average throughput comparison of CLSM and the proposed VAE-based approach for feedback-free MIMO at all locations in $\mathcal{H}$, while Fig. 6(b) details the cumulative distribution function of the average throughput. Note that channel variations are obvious, as evidenced by the performance loss in CLSM with only 1 transmission time interval (TTI) feedback delay. Due to the fixed transmission parameters in time domain, the VAE-based approach exhibits a performance loss of 16.34% after a conservative choice of RI, yet still outperforms CLSM with feedback delay. Fig. 6(c) presents the average throughput gap ratio of the VAE-based approach to CLSM, which intuitively shows that the proposed approach can outperform CLSM if UE is close to BS. Considering the saved



(a) Non-unitary matrices



(b) Unitary matrices

Fig. 4. Average throughput comparison among different transmission schemes at 20 location samples in $\mathcal{H}$ when the precoder dataset consists of non-unitary matrices or unitary matrices.

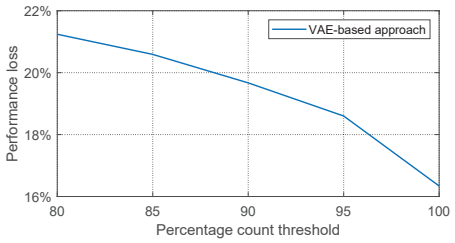
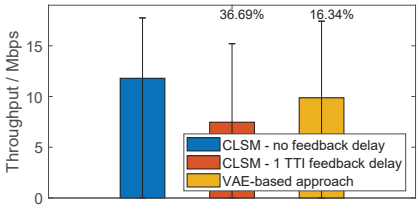
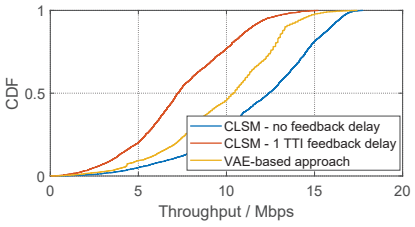


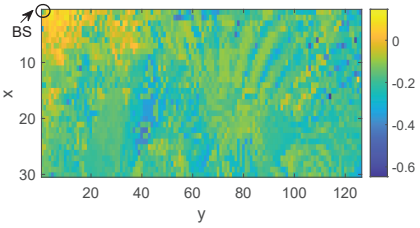
Fig. 5. The performance loss of the VAE-based approach compared with 5G CLSM when $\mathfrak{C}_{\text{thold}}$ is changing.



(a) Average throughput comparison



(b) Cumulative distribution function comparison of average throughput



(c) Average throughput gap ratio

Fig. 6. Comparisons between VAE-based approach and CLSM at all locations in $\mathcal{H}$ when $\mathfrak{C}_{\text{thold}} = 100$.

overheads of pilots and feedback, the proposed VAE-based approach for feedback-free MIMO can potentially achieve comparable or even higher performance than 5G CLSM.

V. Conclusion and Future Works

In this paper, we have studied feedback-free MIMO for FD-RAN, which only utilizes geolocation to determine all the transmission parameters based on the knowledge learned from historical channel data. A VAE-based approach is proposed, where the representative precoder for each location is selected from the latent Gaussian variables. Simulation results based on a link-level simulator and realistic ray-tracing channel data manifest the effectiveness of the proposed approach and its feasibility for application in FD-RAN. In our future work, we will consider the inference of transmission parameters in space domain and extend to multi-user scenarios.

Acknowledgment

This work was supported in part by the National Natural Science Foundation Original Exploration Project of China under Grant 62250004, the Natural Science Foundation of China (NSFC) under Grant 62271244, the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067, the High-level Innovation and Entrepreneurship Talent Introduction Program Team of Jiangsu Province under Grant JSSCTD202202, and The Major Key Project of PCL (PCL2024A03).

References

[1] 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.

[2] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2020, Version 16.2.0.

[3] 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, Dec. 2021.

[4] 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.

[5] 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, May. 2014.

[6] G. Sklivanitis, K. Alexandris, and A. Bletsas, "Testbed for non-coherent zero-feedback distributed beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2013, pp. 2563–2567.

[7] 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, Mar. 2022.

[8] Z. Liu et al., "Leveraging deep reinforcement learning for geolocation-based MIMO transmission in FD-RAN," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, 2023, pp. 1–6.

[9] 3GPP, "NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.214, 2022, Version 16.10.0.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*.

[11] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *Proc. IEEE Conf. Wireless Adv. (WiAD)*, 2010, pp. 1–6.

[12] 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.

[13] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," 2019, *arXiv:1902.06435*.