# Learning-Based Joint Channel Prediction and Antenna Selection for Massive MIMO with Partial CSI

Ke He, Thang X. Vu, Symeon Chatzinotas, and Björn Ottersten

The Interdisciplinary Centre for Security, Reliability and Trust (SnT) - University of Luxembourg,
L-1855 Luxembourg. E-mail: {ke.he, thang.vu, symeon.chatzinotas, bjorn.ottersten}@uni.lu

***Abstract*—This paper investigates the massive multi-input multi-output (MIMO) system in practical deployment scenarios, in which, to balance the economic and energy efficiency with the system performance, the number of radio frequency (RF) chains is smaller than the number of antennas. The base station employs antenna selection (AS) to fully harness the spatial multiplexing gain. Conventional AS techniques require full channel state information (CSI), which is time-consuming as the antennas cannot be simultaneously connected to the RF chains during the channel estimation process. To tackle this issue, we propose a novel joint channel prediction and AS (JCPAS) framework to reduce the CSI acquisition time and improve the system performance under temporally correlated channels. Our proposed JCPAS framework is a fully probabilistic model driven by deep unsupervised learning. The proposed framework is able to predict the current full CSI, while requiring only a historical window of partial observations. Extensive simulation results show that the proposed JCPAS can significantly improve the system performance under temporally correlated channels, especially for very large-scale systems with highly correlated channels.**

***Index Terms*—Channel estimation, antenna selection, partial CSI, massive MIMO, deep learning.**

## I. Introduction

Massive multiple-input multiple-out (MIMO) system has been considered as a most promising technology which offers significant improvements on both spectral and energy efficiencies for the next-generation communication system [1]. Massive MIMO is able to serve multiple users with the same time-frequency resources by deploying a large-scale antenna array at the base station (BS). Since the expensive equipment is only required to be deployed at the BS, the user devices can thereby be relatively inexpensive. Moreover, the performance of massive MIMO systems is generally less sensitive to unfavorable propagation environments [2]. These potential benefits of massive MIMO, however, heavily rely on the knowledge of the complete channel state information (CSI) via transmission of known pilot signal in the channel estimation phase [1]–[3]. During the channel estimation process, the antennas need to be connected to radio frequency (RF) chains for signal detection and measurement. In massive MIMO systems with a very large number of antennas, equipping each antenna with a dedicated RF chain is inefficient from both economical and energy efficiency perspectives [1]. Therefore, in practice a massive MIMO BS usually has a number of RF chains smaller than the number of antennas, which is also the scenario of our interest. In order to fully reap the spatial multiplexing gain, the BS employs antenna selection (AS) before applying precoding techniques to serve the users.

AS has been proposed to significantly reduce the required number of RF chains, by activating only a small subset of all the available antennas at each instant [4]. Specifically, AS adopts simple RF switches to achieve a low hardware cost and power consumption while still benefiting from the spatial diversity gain of antenna arrays [1]. Low-complexity AS algorithms have recently been extensively studied for massive MIMO systems [4]–[7]. For instance, a self-supervised learning based Monte Carlo tree search (MCTS) method was proposed in [5], which solves the AS problem for large-scale systems with achieving near-optimal performance. In [4], a learning-based joint AS and precoding design was proposed to maximize the system sum-rate subject to a transmit power constraint and quality of service (QoS) requirements.

It is worth noting that the above-mentioned AS techniques are only applicable when full CSI is known at the BS, meaning the channel states are complete and fully observable. Getting full CSI in massive MIMO can be a prohibitive task, especially when the number of antennas exceeds the number of the RF chains. One may consecutively switch the RF chains to a subset of antennas during channel estimation phase. This method, however, incurs extra channel estimation overhead and results in a less effective data transmission phase.

In order to tackle this issue, one promising approach is to incorporate channel prediction into the channel estimation phase to avoid extra channel estimation overhead. In practical massive MIMO environments, wireless channels are often dominated by a small number of propagation paths. For low mobility users, the channels will have strong temporal correlation which can be exploited for channel prediction. [8]–[10]. From this perspective, many channel prediction methods have recently been investigated in [11]–[13]. A machine learning (ML)-based time-division duplex scheme was proposed in [11], where full CSI is obtained by leveraging the temporal channel correlation that is applied to both low and high mobility scenarios. To predict channels smart high-speed railway communication networks, the authors of [13] proposed a channel prediction scheme based on convolutional neural network

(CNN) and long short-term memory network (LSTM) which predicts full CSI in a multi-step ahead manner by exploiting the channel correlations. Nevertheless, the prediction methods in [11]–[13] require the fully observed channels history and are not applicable to the case of partial observation, especially when the number of antennas exceeds the number of RF chains. Recently, AS with partial CSI has been proposed in [14] based on multi-armed bandit (MAB) and Thompson sampling technique to reduce the channel estimation overhead. This approach, as we will show later, performs poorly when the number of RF chains is significantly smaller than the number of antennas, since the history of partial observations is not exploited.

In this paper, we investigate the massive MIMO downlink in practical scenarios, in which the channel coefficients are temporally correlated and the number of antennas surpasses the number of the RF chains. In order to select antennas with partially observed channel states while keeping a minimum channel estimation time, we propose a novel joint channel prediction and AS (JCPAS) framework to simultaneously minimize the channel estimation time and harness the spatial multiplexing gain. The core idea of the proposed JCPAS is to exploit the channel temporal correlation so that the current full CSI can be predicted based on the past incomplete observations. To summarize our work, the main contributions of this paper are as follows,

- The proposed JCPAS framework is a unsupervised learning based probabilistic model. Thus, only channel samples are needed rather than the statistics to capture the potential correlations.
- In contrast to the existing literature, our approach does not require the knowledge of full CSI. Instead, we only need to estimate partial CSI at each instant, thus reducing the channel estimation overhead and improving the effective achievable rate. To the best of our knowledge, our proposed approach is the first attempt in the literature to predict channel states only based on the past incomplete observations.
- The performance of the proposed JCPAS is demonstrated via extensive simulation results, which can save about 45% in energy cost in very large-scale systems and achieve nearly the performance bound with highly correlated channels.

## II. System Model and Problem Formulation

### A. System Model

We consider a massive MIMO system downlink consisting of a BS, equipped with $N_t$ antennas and $N_f \leq N_t$ RF chains, and $N_u$ single-antenna users, as in Fig.

For each frame block, we denote $\boldsymbol{H} \in \mathbb{C}^{N_u \times N_t}$ as the full CSI matrix, and let $\boldsymbol{h}_k \in \mathbb{C}^{1 \times N_t}$ be the corresponding full channel vector for user $k$. Since the number of RF chains is smaller than the number of antennas, BS has to select $N_f$ out of $N_t$ antennas for transmitting data. For convenience, let $\boldsymbol{a} = [a_1, a_2, \cdots, a_j, \cdots, a_{N_t}]$ be the AS

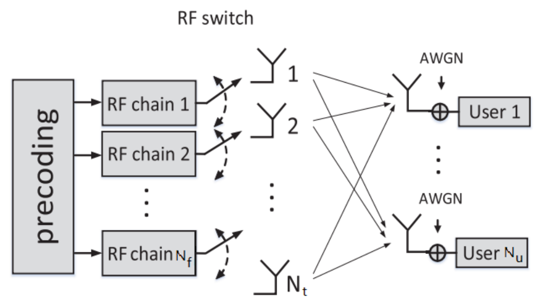


Fig. 1. Structure of the considered massive MIMO system.

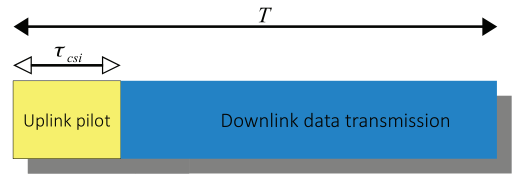


Fig. 2. Block diagram of one transmission block.

vector satisfying $a_j \in \{0,1\}$ and $\sum_{j=1}^{N_t} a_j = N_f$, with $a_j = 1$ indicating the $j$-th antenna is selected and $a_j = 0$ otherwise. Denote $\mathcal{A}$ as the set of all possible subsets of $N_f$ antennas such as $|\mathcal{A}| = \binom{N_t}{N_f}$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient. For each AS vector $\boldsymbol{a} \in \mathcal{A}$, we denote $\boldsymbol{h}_k(\boldsymbol{a}) = [h_{k,a_1}, \cdots, h_{k,a_j}, \cdots, h_{k,a_{N_t}}]$ as the active channel vector of user $k$, where $h_{k,a_j}$ represents the $a_j$-th element of $\boldsymbol{h}_k$. Furthermore, let $\boldsymbol{w}_k(\boldsymbol{a}) \in \mathbb{C}^{N_f \times 1}$ denote the precoding vector for user $k$ with antenna combination $\boldsymbol{a}$. The received signal vector at user $k$ can be expressed as

$$y_k = \boldsymbol{h}_k(\boldsymbol{a})\boldsymbol{w}_k(\boldsymbol{a})x_k + \sum\nolimits_{j \neq k} \boldsymbol{h}_j(\boldsymbol{a})\boldsymbol{w}_j(\boldsymbol{a})x_k + n_k, \quad (1)$$

where $x_k$ is the transmit symbol for user $k$ and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2$. By considering the inter-user interference as well as the noise, the effective signal-to-interference-plus-noise ratio (SINR) is given as $\texttt{SINR}_k = \frac{|\boldsymbol{h}_k(\boldsymbol{a})\boldsymbol{w}_k(\boldsymbol{a})|^2}{\sum_{j \neq k} |\boldsymbol{h}_k(\boldsymbol{a})\boldsymbol{w}_k(\boldsymbol{a})|^2 + \sigma^2}$. Then, the effective achievable rate for user $k$ with antenna combination $\boldsymbol{a}$ is given by

$$R_k(\boldsymbol{a}) = \left(1 - \frac{\tau_{csi}}{T}\right) B \log_2 \left(1 + \texttt{SINR}_k\right), \quad (2)$$

where $B$ is the channel bandwidth. For each frame, the effective system capacity with antenna combination $\boldsymbol{a}$ can be bounded by $\sum_{k=1}^{N_u} R_k(\boldsymbol{a})$, and the total power consumption for transmitting data at each frame is $\sum_{k=1}^{N_u} \|\boldsymbol{w}_k(\boldsymbol{a})\|^2$.

### B. Antenna Selection with Incomplete CSI

Because there are only $N_f \leq N_t$ RF chains, the BS must select a best subset of $N_f$ antennas for downlink data transmission. For the AS problem, a common objective is to optimize a generic objective function $\mathcal{F}_k(\boldsymbol{a})$ while obeying

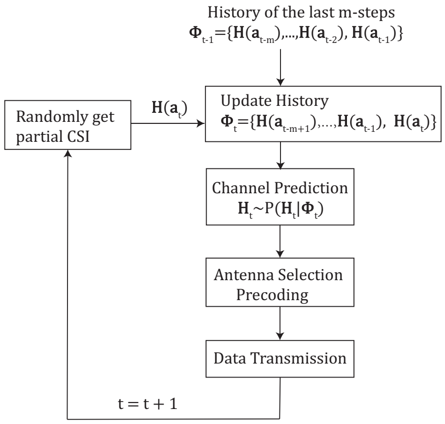


Fig. 3. Diagram of the proposed joint channel prediction and antenna selection framework.

the total transmit power and minimimum QoS requirements. Mathematically, the AS problem can be formulated as

$$\begin{aligned} &\underset{\boldsymbol{a}\in\mathcal{A}}{\text{minimize}} \quad \sum\nolimits_{k=1}^{N_u} \mathcal{F}_k(\boldsymbol{a}) && (3)\\ &\text{subject to} \quad \sum\nolimits_{k=1}^{N_u} \|\boldsymbol{w}_k(\boldsymbol{a})\|^2 \le P_{tot}; \quad R_k(\boldsymbol{a}) \ge \eta_k, \forall k, \end{aligned}$$

where $\eta_k$ is the QoS requirement for user $k$, $P_{tot}$ is the total power, and $\mathcal{F}_k(\boldsymbol{a})$ is the objective function of interest. For example, $\mathcal{F}_k(\boldsymbol{a}) = \|\boldsymbol{w}_k(\boldsymbol{a})\|^2$ in the energy minimization problem, and $\mathcal{F}_k(\boldsymbol{a}) = -R_k(\boldsymbol{a})$ in the sum rate maximization problem.

Solutions to problem (3) have been well studied in the literature [4], [6], [7], [11]–[13] under the *complete CSI* assumption, meaning that the wireless channel is fully observable. Since the number of RF chains is less than the number of transmit antennas at BS, obtaining complete CSI requires $\tau_{csi}^{full} = N_u(\lfloor \frac{N_t}{N_f} \rfloor + 1)$ c.u. in the channel estimation phase. With fixed $N_f$ and $N_u$, we can see that the channel estimation overhead increases significantly for large $N_t$ if we obtain the full CSI. This extra pilot overhead can be very large in massive MIMO and result in reduced effective transmision rate, as shown in (2). For this reason, obtaining the full CSI is thereby a very inefficient strategy for massive MIMO systems, which motivates us to study the AS problem in the presence of only incomplete (or partial) CSI.

## III. Proposed Learning-Based Joint Channel Prediction and Antenna Selection Framework

In this section, we will introduce the proposed learning-based joint channel prediction and AS framework, which can operate without fully estimating the channel states.

### *A. Channel Prediction with Incomplete CSI*

As we have mentioned before, a key observation on the practical massive MIMO system is that the wireless channel is often temporally correlated, which indicates that predicting the full CSI based on the incomplete observation history becomes a possible approach to reduce the channel estimation overhead. Inspired by this, we propose to employ a probabilistic generative neural network (GNN) to predict the full CSI from the history (of partial CSI). Mathematically, a GNN that predicts the full CSI based on the history can be regarded as a probabilistic model $p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$ conditioned on the history $\boldsymbol{\Phi}_t \triangleq \{\boldsymbol{H}_{t-m+1}(\boldsymbol{a}_{t-m+1}), \cdots, \boldsymbol{H}_{t-1}(\boldsymbol{a}_{t-1}), \boldsymbol{H}_t(\boldsymbol{a}_t)\}$ with a fixed length $m$. Note that the missing history will be replaced by zero matrices if $t < m$. Though it is hard to know the unknown distribution $p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$ exactly, one can still use the maximum likelihood approximation method to estimate it.

In order to estimate the unknown distribution $p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$, we first need to choose an appropriate model $q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$, which is parameterized by $\boldsymbol{\theta}$, where $\boldsymbol{\Phi}_t$ is the history before $t$. Then, we collect a training data set $\mathcal{D} = \{\boldsymbol{H}_t, \boldsymbol{\Phi}_t\}_{t=1}^{L}$. Our training objective is to maximize the likelihood of the collected training samples on the chosen model, which is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \frac{1}{L} \sum\nolimits_{t=1}^{L} -\log q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t). \tag{4}$$

Obviously, (4) quantifies how well the chosen distribution fits the samples drawn from the unknown distribution $\boldsymbol{H}_t \sim p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$. In particular, this objective achieves its minimum if $q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$ perfectly approximates $p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$. On the other hand, the objective enlarges if the chosen model deviates from the unknown distribution. Unfortunately, choosing an appropriate model $q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$ requires prior knowledge on the unknown distribution, which is typically difficult in practice. Therefore, we employ a *conditional normalizing flow* to approximate the underlying distribution.

As a kind of GNN, normalizing flow can efficiently infer the latent variables. Instead of directly computing the log-likelihood, it computes the corresponding log-likelihood using the rule of change of variable [15]. Given a complete observation $\boldsymbol{H}_t \sim p(\boldsymbol{H}_t|\boldsymbol{\Phi}_t)$, we assume that it relies on a latent variable $\boldsymbol{Z}_t$ with entries $z_{i,j}$ following a tractable distribution conditioned on the history. Thus, the latent space is also conditioned on the history $\boldsymbol{Z}_t \sim p_{\mathcal{Z}}(\boldsymbol{Z}_t|\boldsymbol{\Phi}_t)$, and the generative model is given by

$$\boldsymbol{Z}_t \sim p_{\mathcal{Z}}(\boldsymbol{Z}_t|\boldsymbol{\Phi}_t) \text{ and } \boldsymbol{H}_t = g_{\boldsymbol{\theta}}(\boldsymbol{Z}_t), \tag{5}$$

where $g_{\boldsymbol{\theta}}(\cdot)$ is a invertible function with parameter $\boldsymbol{\theta}$. The latent variables can be obtained efficiently by applying the inversion $\boldsymbol{Z} = f_{\boldsymbol{\theta}}(\boldsymbol{H}) = g_{\boldsymbol{\theta}}^{-1}(\boldsymbol{H})$. Then, the unknown distribution can be approximated by

$$\log q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t) = \log p_{\mathcal{Z}}(f_{\boldsymbol{\theta}}(\boldsymbol{H}_t)|\boldsymbol{\Phi}_t) + \log\Big|\det\Big(\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{H}_t}\Big)\Big|, \tag{6}$$

where $\det\left(\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{H}}\right)$ denotes the determinant on the Jacobian matrix. Assume that the invertible function $f(\cdot)$ can be further factorized as $L$ invertible sub-functions as $f(\cdot) = f_1(\cdot) \otimes f_2(\cdot) \otimes \cdots f_k(\cdot) \cdots \otimes f_L(\cdot)$. Then, latent variables $\boldsymbol{Z}$ can be computed by

$$\boldsymbol{H} \xrightarrow{f_1} \boldsymbol{U_1} \xrightarrow{f_2} \boldsymbol{U_2} \cdots \xrightarrow{f_k} \boldsymbol{U}_l \cdots \xrightarrow{f_L} \boldsymbol{Z}. \tag{7}$$

Denoting $\boldsymbol{U_0} \triangleq \boldsymbol{H}_t$ and $\boldsymbol{U_L} \triangleq \boldsymbol{Z}_t$, (6) can be rewritten as

$$\begin{aligned}\log q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{\Phi}_t) = &\log p_{\mathcal{Z}}(f_{\boldsymbol{\theta}}(\boldsymbol{H}_t)|\boldsymbol{\Phi}_{t<t_0})\\ &+ \sum\nolimits_{l=1}^{L} \log\Big|\det\Big(\frac{\mathrm{d}\boldsymbol{U}_l}{\mathrm{d}\boldsymbol{U}_{l-1}}\Big)\Big|. \end{aligned} \tag{8}$$

**Algorithm 1** The Proposed joint Channel Prediction and Antenna Selection (JCPAS) Algorithm

1: Set $\mathbf{\Phi}_0 = \{\mathbf{0}_{N_u \times N_t}, ..., \mathbf{0}_{N_u \times N_t}\}$;
2: **for** $t = 1, 2, \cdots T$ **do**
3: Uniformly select a random AS vector $\boldsymbol{a}_t \in \mathcal{A}$ to estimate the current partial CSI $\boldsymbol{H}_t(\boldsymbol{a}_t)$;
4: $\mathbf{\Phi}_t = \{\boldsymbol{H}_{t-m+1}(\boldsymbol{a}_{t-m+1}), \cdots, \boldsymbol{H}_{t-1}(\boldsymbol{a}_{t-1}), \boldsymbol{H}_t(\boldsymbol{a}_t)\}$
5: $\widehat{\boldsymbol{H}}_t \sim q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\mathbf{\Phi}_t)$
6: ANTENNASELECTION($\widehat{\boldsymbol{H}}_t$)
7: PRECODING($\widehat{\boldsymbol{H}}_t$)
8: DATATRANSMISSION($\widehat{\boldsymbol{H}}_t$)
9: **end for**

Thus, each sub-function becomes a small step of the complete flow, and thereby we build a conditional normalizing flow for approximating the unknown distribution.

### B. The Proposed Learning-Based Framework

Since now we have $q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\mathbf{\Phi}_t) \approx p(\boldsymbol{H}_t|\mathbf{\Phi}_t)$, it is straightforward to select antennas for data transmission after predicting the full channel based on the history. In order to fulfill it, we hereby introduce a general joint channel prediction and antenna selection (JCPAS) framework, which jointly uses the GNN $q_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\mathbf{\Phi}_t)$ together with any existing AS algorithms to improve the system performance under temporally correlated channels. The structure of the proposed JCPAS framework is illustrated in Fig. 3. As shown in this figure, we firstly randomly obtain the partial observation of the current channel, denoted by $\boldsymbol{H}_t(\boldsymbol{a}_t)$. Then, we update the observation history by setting $\mathbf{\Phi}_t = \{\boldsymbol{H}_{t-m+1}(\boldsymbol{a}_{t-m+1}), \cdots, \boldsymbol{H}_{t-1}(\boldsymbol{a}_{t-1}), \boldsymbol{H}_t(\boldsymbol{a}_t)\}$. After that, we predict the full channel via the well-trained GNN $\widehat{\boldsymbol{H}}_t \sim q_{\theta}(\boldsymbol{H}_t|\mathbf{\Phi}_t)$. Since we have the estimation of the full channel, we just need to employ a specific antenna selection algorithm to select antennas for both precoding and data transmission. To summarize the process, the pseudo code of the proposed framework is detailed in Algorithm 1. It should be noted that the proposed framework is a general framework which aims to help reduce the channel estimation overhead for massive MIMO systems. The choice of AS algorithms and precoding designs are determined based on the available resources for different scenarios.

### C. Implementation Details

The network structure employed by the proposed framework is illustrated in Fig. 4. As we have introduced in Sec. III-A, the most important thing to implement a GNN is to ensure that the sub-functions represented by each building block are fully invertible. Generally, such invertible sub-functions are implemented by normalization layers, invertible convolutional layers and affine coupling layers, where the details can be found in [16] and [15], respectively. By using these invertible layers, we are able to construct a invertible network for inferring latent variables. Specifically, we construct the invertible network with $L$ flow steps, and each flow step contains three

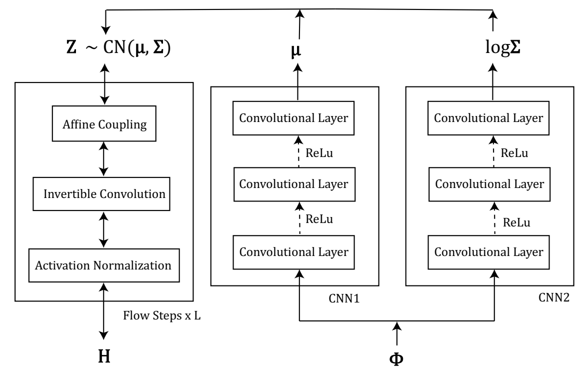


Fig. 4. The network structure for the proposed JCPAS framework.

layers: activation normalization layer, invertible convolutional layer and affine coupling layer.

Since the latent random variable $\boldsymbol{Z}$ is conditioned on the history $\mathbf{\Phi}$, we still need to implement the conditional distribution $p_{\mathcal{Z}}(\boldsymbol{Z}_t|\mathbf{\Phi}_t)$. In order to fulfill it, we consider $\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{Z}; \boldsymbol{\mu}(\mathbf{\Phi}_t), \mathbf{\Sigma}(\mathbf{\Phi}_t))$, where the mean $\boldsymbol{\mu}(\mathbf{\Phi}_t)$ and variance $\mathbf{\Sigma}(\mathbf{\Phi}_t)$ are determined by the history $\mathbf{\Phi}_t$. Specifically, the mean and variance are computed by two separated CNNs, which can be expressed as$\boldsymbol{\mu}(\mathbf{\Phi}) = \text{CNN1}(\mathbf{\Phi})$ and $\mathbf{\Sigma}(\mathbf{\Phi}_t) = \exp(\text{CNN2}(\mathbf{\Phi}))$. The two CNNs have the same network structure, and they are only composed by convolutional layers and rectified linear units (ReLU) [15]. It should be noted that we use zero-padding to keep each $\boldsymbol{H}_t(\boldsymbol{a}_t)$ having the shape of $N_u \times N_t$ in order to retain the spatial information.

### D. Computational Complexity

The normalizing flow is composed by three different types of invertible layers, and the computational complexity of these invertible layers is dominated by element-wise operations and log-determinants [3], [15], [16]. Hence, for the computational complexity of normalizing flow, it relies on the input size, which is given by $\mathcal{O}((LN_tN_u)$. For a CNN with $L_c$ layers, we denote the convolutional kernel size and number of convolutional kernels at the $i$-th layer as $s_i$ and $n_i$. Therefore, the computational complexity of CNN is given by $\mathcal{O}(\sum_{i=1}^{L_c} n_{i-1}s_i^2N_tN_un_i)$ [17]. As to the total computational complexity of the proposed framework, it not only depends on the normalizing flow and CNNs, but also depends on the chosen AS and precoding algorithm. If we denote the computational complexity of the chosen AS and precoding algorithms as $\mathcal{O}(P)$. Then, the total computational complexity of the proposed framework is given by $\mathcal{O}(LN_tN_u + \sum_{i=1}^{L_c} n_{i-1}s_i^2N_tN_un_i + P)$.

## IV. PERFORMANCE EVALUATION

### A. Environment Setup

In order to demonstrate the performance of the proposed framework, we perform simulations on various massive MIMO systems with different system scales and different channel conditions. In addition, we employ a conventional AS algorithm which successively select antennas based on the column norms of the provided CSI matrix [18]. Moreover, we adopt the zero-forcing precoding algorithm in the simulation [4].

As to the initial state, we assume that users are randomly located around BS. The channels are time-varying and we apply the Jakes model [19] to generate the channel matrix with the normalized Doppler frequency $f_D = 0.1$. In addition, the following temporal correlation model is employed, which is given by

$$\boldsymbol{H}_t = \sqrt{\rho}\boldsymbol{H}_{t-1} + \sqrt{1-\rho}\boldsymbol{G}_t, \tag{9}$$

where $\rho \in [0, 1]$ is the correlation coefficient, and $\boldsymbol{G}_t$ is a time independent random matrix whose entries $g_{i,j} \sim \mathcal{CN}(0, 1)$. In particular, $\rho = 1$ implies that the channel is completely correlated while $\rho = 0$ implies that the channel is independent for different time slots.

As to the network structure, we employ a normalizing flow with $L = 16$ flow steps, and each CNN contains $L_c = 6$ layers where the number of convolutional kernels and the kernel size of each layer are $\{64, 32, 32, 16, 64, 128\}$ and $\{3, 9, 3, 3, 3, 9\}$, respectively. Other parameters are as follows: $N_f = 16, N_u = 10, P_{tot} = 10$W, $m = 64$, $\sigma = 1$, $B = 100$ MHz, $T = 512$ c.u., $N_t \in [32, 160]$, and the QoS $\eta_k = \eta \in [350, 400]$ Mbps, $\forall k$.

### B. Competing Algorithms

In order to show the effectiveness of the proposed framework, we compare the proposed JCPAS framework with various competitive algorithms. Before discussing the simulation results, we first need to introduce the following abbreviations,

- **JCPAS**: The proposed JCPAS framework, which randomly estimates partial CSI $\boldsymbol{H}(\boldsymbol{a}) \in \mathbb{C}^{N_u \times N_f}$ at the channel estimation stage of each transmission block, and recovers the full CSI $\boldsymbol{H} \in \mathbb{C}^{N_u \times N_t}$ based on the history.
- **OAS [14]**: The online antenna selection (OAS) algorithm introduced in [14]. Instead of exploiting the history, this algorithm obtains partial CSI by modeling the problem as a MAB problem.
- **Perfect**: This scheme performs simulations by considering that the full CSI can be perfectly recovered from the history. The performance of this scheme can serve as the performance upper bound of the proposed JCPAS.
- **Full**: This scheme uses maximum channel estimation overhead to obtain full CSI.
- **Random**: This scheme randomly selects antennas for data transmission, which can be seen as the lower bound.

It should be noted that the above schemes employ the same AS and precoding algorithms for a fair comparison.

### C. Simulation Results and Discussion

Fig. 5 illustrates the power consumption comparison versus the temporal correlation coefficient $\rho$ for the aforementioned algorithms, where $N_t = 128, \eta_k = 400$MHz$, \forall k$. Besides, the correlation coefficient $\rho$ varies from 0 to 1.0. It is shown that the proposed JCPAS outperforms the other schemes when $\rho \geq 0.5$, and the performance gap between the proposed JCPAS framework and the "Full" scheme enlarges with the increasing

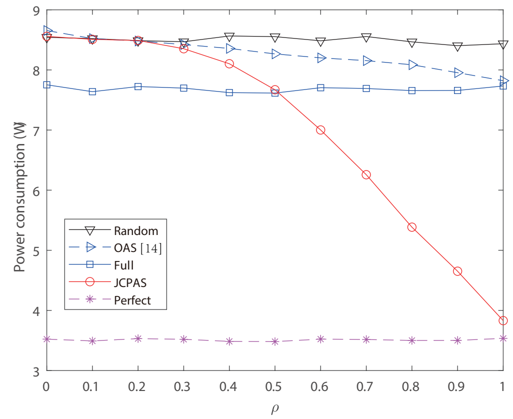


Fig. 5. System performance comparison versus the temporal correlation coefficient $\rho$, where $N_t = 128$, $N_f = 16$ and $N_u = 10$.

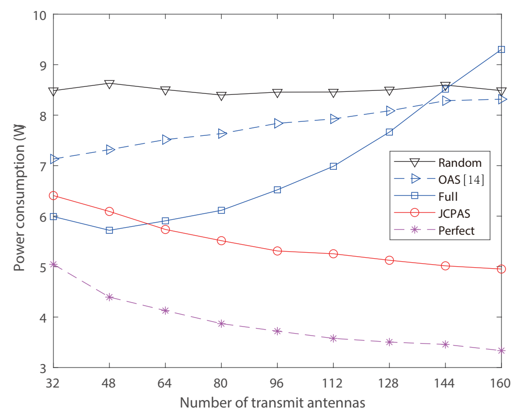


Fig. 6. System performance comparison versus the system scale, where $\rho = 0.8$, $N_f = 16$ and $N_u = 10$.

correlation level. Specifically, when $\rho = 0.9$ where the channel is highly correlated, the proposed JCPAS framework can save nearly 45% and 50% in energy costs with comparison to the "Full" and "Random" schemes. Moreover, we can also find that the proposed JCPAS framework can approach the upper bound performance in highly correlated scenarios, which demonstrates the effectiveness of the proposed framework.

Fig. 6 demonstrates the power consumption comparison versus the number of antennas $N_t$, where $\rho = 0.8$ and $\eta_k = 400$ MHz. From this figure, we can observe that the power consumptiosn of the "Full" scheme raises rapidly with the increasing number of transmit antennas. This is because the channel estimation overhead will increase rapidly for very large-scale systems, where the effective ratio for getting the full CSI decreases from around 0.941 to 0.785. This indicates that the system needs to allocate much more energy to satisfy the QoS requirements, which is obviously inefficient for very-large scale systems. In contrast, the effective data transmission time of the partial CSI is fixed as 0.98, since the $\tau_{csi}$ remains the same for different system scales. On the other hand, the proposed JCPAS consumes less power as $N_t$ increases, which confirms the effectiveness of the CSI prediction of JCPAS. In addition, we can find that it performs bad as the

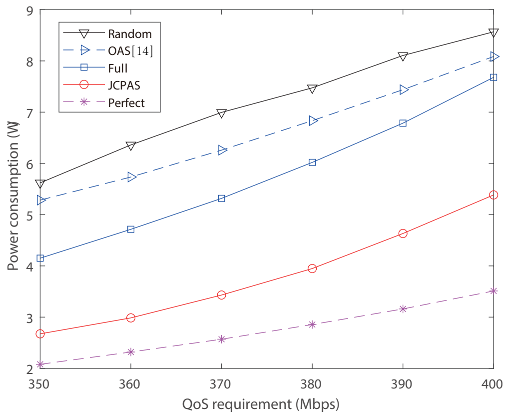


Fig. 7. System performance comparison versus QoS requirements, where $\rho = 0.8$, $N_t = 128$, $N_f = 16$ and $N_u = 10$.

system scale growing, and JCPAS still outperforms the OAS algorithm as well as the "Full" and "Random" schemes. The performance gap between the proposed JCPAS framework and the "Perfect" scheme remains unchanged for very-large scale systems. Since OAS does not exploit the history, it performs bad when the channel is non-static. This indicates that the proposed JCPAS can efficiently and accurately recover the full CSI from the partially observed history, which further verifies the effectiveness of the proposed framework.

Fig. 7 depicts the energy consumption comparison versus the QoS requirements. It is shown from Fig. 7 that the proposed JCPAS outperforms the OAS algorithm as well as the "Full" and "Random" schemes for different QoS requirements. Specifically, for the QoS requirement of $400$ Mbps, JCPAS is able to save around $32\%$, $36\%$ and $40\%$ in energy costs with respect to the "Full", OAS and "Random" schemes. These results show that for very large-scale systems with partial CSI, the proposed JCPAS can reduce the energy cost and still satisfy different QoS requirements, which further verifies the effectiveness of the proposed JCPAS framework.

## V. Conclusion

In this paper, we have investigated the AS problem for massive MIMO systems by considering only a portion of the full CSI is available for each transmission block. In order to reduce the channel estimation overhead for massive MIMO systems, we have proposed to employ a deep conditional normalizing flow to recover the full channel from the history of partial observations. By utilizing the proposed conditional normalizing flow, we have further established a general joint channel prediction and antenna selection framework, which can help improve the performance of other AS algorithms in the presence of massive MIMO systems and practical CSI. Since the practical systems often suffer from temporal correlations, we believe that the proposed JCPAS framework can effectively improve the robustness of massive MIMO systems in practical scenarios.

## Acknowledgement


This work is supported in parts by project ERC AGNOSTIC, ref. EC/H2020/ERC2016ADG/742648/AGNOSTIC and project FNR DisBUS, ref. FNR/BRIDGES19/IS/13778945/DISBuS.


## References


[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742–758, 2014.

[2] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[3] K. He, L. He, L. Fan, Y. Deng, G. K. Karagiannidis, and A. Nallanathan, "Learning-based signal detection for MIMO systems with unknown noise statistics," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3025–3038, 2021.

[4] T. X. Vu, S. Chatzinotas, V.-D. Nguyen, D. T. Hoang, D. N. Nguyen, M. Di Renzo, and B. Ottersten, "Machine learning-enabled joint antenna selection and precoding design: From offline complexity to online performance," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3710–3722, 2021.

[5] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using monte carlo tree search," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5380–5390, 2019.

[6] Q. Hu, Y. Liu, Y. Cai, G. Yu, and Z. Ding, "Joint deep reinforcement learning and unfolding: Beam selection and precoding for mmWave multiuser MIMO with lens arrays," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2289–2304, 2021.

[7] Z. Liu, Y. Yang, F. Gao, T. Zhou, and H. Ma, "Deep unsupervised learning for joint antenna selection and hybrid beamforming," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1697–1710, 2022.

[8] G. J. Byers and F. Takawira, "Spatially and temporally correlated MIMO channels: Modeling and capacity analysis," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 634–643, 2004.

[9] K. Huang, R. W. Heath, and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Transactions on Signal Processing*, vol. 57, no. 5, pp. 1959–1975, 2009.

[10] Z. Gao, L. Dai, Z. Lu, C. Yuen, and Z. Wang, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE communications letters*, vol. 18, no. 7, pp. 1266–1269, 2014.

[11] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive mimo with channel aging," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2960–2973, 2020.

[12] C. Wu, X. Yi, Y. Zhu, W. Wang, L. You, and X. Gao, "Channel prediction in high-mobility massive mimo: From spatio-temporal autoregression to deep learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 1915–1930, 2021.

[13] T. Zhou, H. Zhang, B. Ai, C. Xue, and L. Liu, "Deep-learning based spatial-temporal channel prediction for smart high-speed railway communication networks," *IEEE Transactions on Wireless Communications*, 2022.

[14] Z. Kuai and S. Wang, "Thompson sampling-based antenna selection with partial csi for tdd massive MIMO systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7533–7546, 2020.

[15] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," *Advances in neural information processing systems*, vol. 31, 2018.

[16] L. Dinh, D. Krueger, and Y. Bengio, "Nice: Non-linear independent components estimation," *arXiv preprint arXiv:1410.8516*, 2014.

[17] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[18] T.-W. Ban and B. C. Jung, "A practical antenna selection technique in multiuser massive MIMO networks," *IEICE transactions on communications*, vol. 96, no. 11, pp. 2901–2905, 2013.

[19] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Wiley-IEEE press, 1994.