Learning-Based Remote Channel Inference: Feasibility Analysis and Case Study

Sheng Chen, Zhiyuan Jiang, Sheng Zhou, Zhisheng Niu, *Fellow, IEEE*, Ziyan He, Andrei Marinescu, and Luiz A. DaSilva, *Fellow, IEEE*

***Abstract*— Channel state information (CSI) plays a vital role in wireless communication systems. However, the CSI acquisition overhead is an enormous obstacle to realize the system performance improvements promised by massive connectivity and massive multiple-input-multiple-output (MIMO). To alleviate this overhead, this paper proposes a remote channel inference framework by probing the channels occupied by a source base station (BS) and inferring the channels of target BSs at geographically separated sites. The work generalizes existing literature which mainly focuses on utilizing the CSI linear correlations of adjacent antennas, by adopting a model-free deep learning framework to investigate non-linear dependence among remote CSI. The existence of such cross-BS CSI dependence is first shown by calculating the mutual information between remote channels, and the Cramér-Rao lower bound of remote CSI inference performance based on a one-ring channel model. Inspired by this finding, modern deep learning approaches are leveraged to perform remote channel inference in heterogeneous networks for both single user and multi-user scenarios. The simulation results based on ray tracing data show evident performance advantages over conventional methods, under both homogeneous and heterogeneous frequency coverage. The proposed framework achieves beamformer inference accuracy within 4.6% of the genie-aided optimum at the cost of sweeping only two beams.**

***Index Terms*— Channel state information, MIMO, deep neural networks, channel models, beamforming.**

Manuscript received October 24, 2018; revised February 12, 2019 and April 17, 2019; accepted April 22, 2019. Date of publication May 15, 2019; date of current version July 10, 2019. This work was supported in part by the Nature Science Foundation of China under Grant 61861136003, Grant 61871254, Grant 91638204, Grant 61571265, and Grant 61621091, in part by the National Key R&D Program of China under Grant 2018YFB0105005, in part by the Intel Collaborative Research Institute for Intelligent and Automated Connected Vehicles, and in part by the Science Foundation Ireland under Grant 17/NSFC/5224. This paper was presented in part at the *IEEE GLOBECOM* 2017 [1] and the *IEEE GLOBECOM* 2018 [2]. The associate editor coordinating the review of this paper and approving it for publication was S. Wang. *(Corresponding author: Zhiyuan Jiang.)*

S. Chen, S. Zhou, and Z. Niu are with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: chen-s16@tsinghua.edu.cn; sheng.zhou@tsinghua.edu.cn; niuzhs@tsinghua.edu.cn).

Z. Jiang is with the Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China (e-mail: zhiyjiang@foxmail.com).

Z. He is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: heziyan@gatech.edu).

A. Marinescu and L. A. DaSilva are with CONNECT, Trinity College Dublin, Dublin 2, D02 PN40 Ireland (e-mail: marinesa@tcd.ie; dasilval@tcd.ie).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2019.2915683

I. Introduction

CHANNEL state information (CSI) plays a crucial role in wireless communication systems. Advanced signal processing techniques for the physical layer in 5G and beyond systems require accurate and timely CSI for coherent data decoding and high-dimensional spatial signal operations to guarantee adequate system performance. High-layer operations also need CSI for, e.g., user scheduling and mobility management. On the flip side, due to the high reliance on CSI of new techniques in multi-antenna use, the channel acquisition overhead is increasing dramatically, which is a key limiting factor for system scalability. In particular, deploying a large number of antennas at the base station (BS), namely massive multi-input multi-output (MIMO) [3], can achieve a significant spatial multiplexing gain and diversity gain with simplified signal processing methods thanks to the channel hardening effect. However, multiple antennas at both the BS and user equipment (UE) sides require the estimation of a high dimensional channel matrix, which adds to the CSI overhead. Moreover, ultra-dense network (UDN) is proposed to increase frequency reuse [4]; it can provide a higher coverage probability, spectral efficiency (SE) and energy efficiency, but the dense deployment of cells brings about the challenge of high pilot overhead density. In addition, directional communication is considered to be essential in small cells and millimeter wave (mm-wave) based systems to mitigate interference and enhance signal coverage, which requires CSI between a massive number of transceivers. Recently, vehicular networks and air-to-ground networks have been extensively studied [5], [6], wherein the vehicles or unmanned aerial vehicles (UAVs) are considered as UEs or mobile BSs. Therein, the high mobility also increases the time variability of CSI, resulting in a more frequent CSI acquisition requirement. In summary, future trends of wireless communications entail substantial CSI acquisition overhead, and hence novel, more efficient methods are needed.

Conventionally, in time-division-duplex (TDD) massive MIMO systems, BSs can obtain the downlink (DL) channel directly from the uplink (UL) pilots sent by users thanks to channel reciprocity. In this fashion, the pilot overhead scales with the number of users [7]. In frequency-division-duplex (FDD) systems, instantaneous CSI reciprocity does not exist, as different frequency bands are used for the UL and DL. In this case, the BS needs to transmit DL pilots and UEs estimate the DL channels and feedback the estimations. As a result, the overhead scales with

the number of BS antennas. Therefore, the acquisition of instantaneous CSI in full digital beamforming systems poses a severe challenge, and hence novel beamforming methods, such as hybrid beamforming, have been proposed [8]. Hybrid beamforming architectures were first proposed for reducing the number of radio frequency (RF) chains in massive MIMO systems, as well as reducing CSI overhead as in [9], wherein a joint spatial division and multiplexing (JSDM) scheme consisting of a two-stage beamformer was proposed by first training the analog beamformer based on the second-order CSI, and then training the digital beamformer based on the dimensionality-reduced instantaneous CSI. Recently, analog beamforming for non-orthogonal multiple access was discussed in [10], wherein the proposed sub-optimal solution for power allocation and beamforming can achieve close-to-bound sum-rate performance.

Faced with the dilemma between the limited spectrum resources and the increasing bandwidth requirements, researchers have become interested in *mm-wave* bands. Channels in mm-wave bands show sparsity in angular and delay domains, hence a line of work exploits methods based on compressed sensing to recover the sparse MIMO CSI from on a relatively small number of observations [11]. These approaches can accurately recover the CSI when the multi-path components (MPCs) are sparse, at a cost of large computation overhead. Another set of works consider using beam search methods to estimate the mm-wave channels [12], [13]. Specifically, the authors build a hierarchical search framework to reconstruct the CSI, wherein the BSs transmit the beamformed reference signals from a predefined hierarchical codebook and the UEs feed the signal power back to the BSs. The candidate beams become narrower during the training procedure, and finally the exact discrete beam directions can be found. However, these hierarchical methods suffer from large beam training time and can only support channels with a limited number of MPCs. To reduce the channel acquisition overhead in MIMO systems, a large body of work considers exploiting the channel correlations. Correlation models among channels have been studied for over a decade. The work in [9], [14]–[16], proposes to leverage spatial channel correlations among co-located BS antennas to reduce the overhead and shows considerable performance gain. Authors in [17] utilize the temporal channel correlation to help find the current channel support based on the previous channel observations in mm-wave systems. Research in [18] manages to exploit the UL channel estimations to help learn the DL channels, essentially relying on the channel correlations in the frequency domain. In general, these works reduce the channel estimation overhead based on the channel correlation in different domains, mainly focusing on exploiting linear correlations of CSI by leveraging the second-order CSI statistics. Recently, with the rapid development of deep learning, deep neural networks (DNNs) have started to be applied in wireless communications. A model-driven approach for physical layer operations is proposed in [19]. The work in [20] predicts the channel fading with the help of a complex-valued DNN. The work in [21] learns the coordinated beamforming weights from the UL reference signals with the help of a convolutional neural network (CNN). Authors in [22] leverage machine learning techniques to decide the beam directions from the context information, i.e., the locations of transceivers and scatterers.

In this paper, we study the CSI dependence among remote BSs from the same UE and exploit these dependencies for CSI inference. To the best of our knowledge, the existing literature seldom explores the possibility that the CSI of one BS can be inferred by the CSI of remote BSs, since it is generally assumed that the CSI of geographically separated (beyond wavelengths) BSs is independent. Our previous work [23] explored this potential, in which we used supervised learning to make cross-BS channel inference for the purpose of BS selection. We extend the work to beamforming design in a scenario where the BSs are equipped with multiple antennas and the CSI of BSs with no assigned pilot resources, namely target BSs, can be inferred based on the CSI of source BSs with pilot resources. In general, the relationship between remote channels is non-linear and cannot be easily handled with linear operations. Hence DNNs are introduced to make the inference, due to their universal approximating properties on any measurable functions providing adequate and appropriate training data [24]. As a result, the pilot resources can be significantly reduced at the cost of training a DNN offline.[1] The main contributions of this work are as follows:

1) We show dependence among remote CSIs exists, based on calculating the mutual information from the remote ray-tracing based CSI data and the Cramér-Rao lower bound (CRLB) of remote channel inference based on a one-ring channel model.
2) We propose a remote channel inference framework, which can reduce the pilot overhead by replacing the pilot-aided channel estimation procedure by the CSI inference from an offline-trained DNN.
3) We study beamforming for small cells in heterogeneous networks as a use case for the framework, and propose the DNN architectures based on corresponding application scenarios and purposes, e.g., whether multi-user spatial multiplexing is considered. Simulation results based on ray-tracing data show significant performance gains over conventional methods. The proposed framework can achieve a performance loss of 4.60% compared with an optimal beamformer, at the cost of sweeping between only two beams.

The rest of this paper is organized as follows. Section II shows the dependence among remote channels. In Section III, the remote channel inference framework is introduced, and beamforming for small cells in heterogeneous networks is studied as a use case. Section IV depicts the performance of the proposed framework on ray-tracing based channel data. Finally, Section V presents our concluding remarks and discusses potential avenues for future work.

[1] In this work, some statistical features of CSI can be inferred, e.g. angular power spectrum (APS), while the inference of the instantaneous CSI is saved for future work. Therefore, reference signal (RS) for beamforming can be saved, while RS for demodulation is still required.

II. Dependence Among Remote Channels

Before diving into details about the proposed remote channel inference schemes, this section is dedicated to show that dependence exists between geographically separated BSs; this finding justifies the feasibility of the channel inference performed in our proposed framework.

A. Mutual Information Calculation on Remote Channels

Most existing literature assumes the CSIs from geographically separated BSs (larger than the order of wavelength) to be statistically independent, by showing that their linear correlation is approximately zero. However, in this work, we show that they actually exhibit significant mutual information, based on ray-tracing CSI data.

Towards proving the existence of this dependence, we first provide some high-level intuition. In general, the CSI can be considered as a function of UE locations and other factors. This function mapping changes with time due to changes in the environment, but it can be assumed that the CSI is quasi-static as long as the estimation time is smaller than the channel coherence time. Moreover, the average CSI (CSI statistics) is expected to be strongly related to UE locations, especially when a line-of-sight (LoS) path exists. Additionally, UE locations are strongly related to the observable CSI at a source BS which is also available for the UE. Under the condition that source BS antennas form a large antenna array, existing literature shows that the channel responses from different UE locations are asymptotically orthogonal [3]. Thus the mapping from UE location to the source BS array response is invertible. Therefore, the CSI at target BSs can be inferred by first estimating the UE location from source BS channel and then using the estimated location to generate the average CSI. However, in our approach, these two steps are implicitly combined by learning the mapping between the channels of source BS and target BS directly. In general, this implicit relationship is non-linear, and thus it does not exhibit any significant linear correlations. Of course in general scenarios, these mappings are much more complex and hence a data-driven approach is reasonable.

To further illustrate this point, this dependence can be shown clearly by calculating the mutual information, which measures the degree of relevance between two variables (vectors), between the local CSI of source BSs and the remote CSI at a target BS. The calculation is based on a data-driven approach, whereby channel parameters are generated according to a ray-tracing software named Wireless InSite [25]. As shown in Fig. 1, a source BS and target BS are placed at a distance of 122 meters in an urban outdoor scenario, equipped with 128 and 32 antennas, respectively. Single-antenna UEs are uniformly distributed under the coverage of the target BS. Assuming that both the source BS and the target BS are equipped with uniform linear antenna arrays, a steering vector can be expressed as

$$[\boldsymbol{a}(\theta)]_i = e^{-j\frac{2\pi}{\lambda} i\delta \sin\theta} \, (i = 0, 1, \cdots, M-1), \tag{1}$$

where λ denotes the wavelength, δ denotes the antenna spacing, M denotes the number of antennas and θ denotes the angle of departure (AoD) for transmitters or the angle of arrival (AoA) for receivers. Considering narrowband signals, the DL CSI between BSs and UEs can be expressed as

$$\boldsymbol{H} = \sum_{i=1}^{N_{\mathrm{P}}} \alpha_i \boldsymbol{a}_{\mathrm{B}}(\theta_{\mathrm{d}i}) \boldsymbol{a}_{\mathrm{U}}^T(\theta_{\mathrm{a}i}), \tag{2}$$

where α_i, $\theta_{\mathrm{d}i}$, $\theta_{\mathrm{a}i}$ denote the complex impulse response, AoD, AoA of the i-th propagation path between transceivers, respectively, and N_{P} denotes the total number of propagation paths. Since UEs are equipped with a single antenna in the simulations, $\boldsymbol{a}_{\mathrm{U}}$ equals 1 and CSI $\boldsymbol{H}$ becomes a vector $\boldsymbol{h}$. The CSI of the target BS is quantized to an index in the discrete-Fourier-transform (DFT) based codebooks or random vector quantization (RVQ) based codebooks. As for the source BS, we use a scalar quantization scheme with reverse water filling bit-loading proposed in [14]. Fig. 2 shows the calculated mutual information under different degrees of quantization. It can be seen that the mutual information between the local CSI of the source BS and the remote CSI at the target BS (based on either the DFT codebook or RVQ codebook) is close to the information entropy of the latter one provided 40 bits for local CSI quantization, which indicates that the optimal beam pattern at the target BS is almost certain given the local CSI of source BS. Besides, results from canonical-correlation analysis show that the linear correlation between remote channel vectors is approximately zero, suggesting that, combined with mutual information results, remote channel vectors are uncorrelated but dependent.

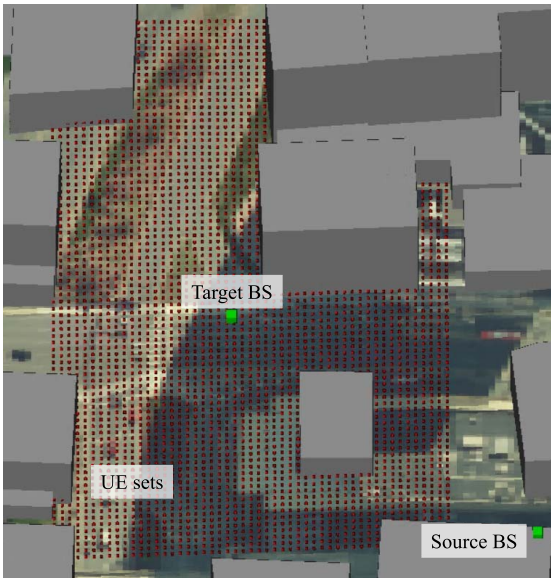
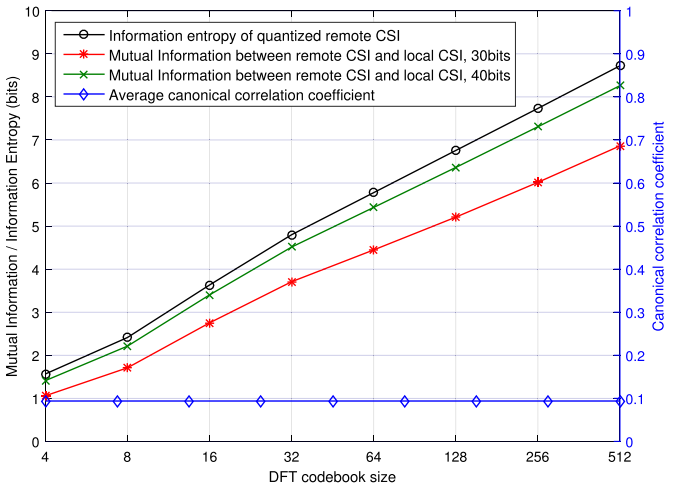


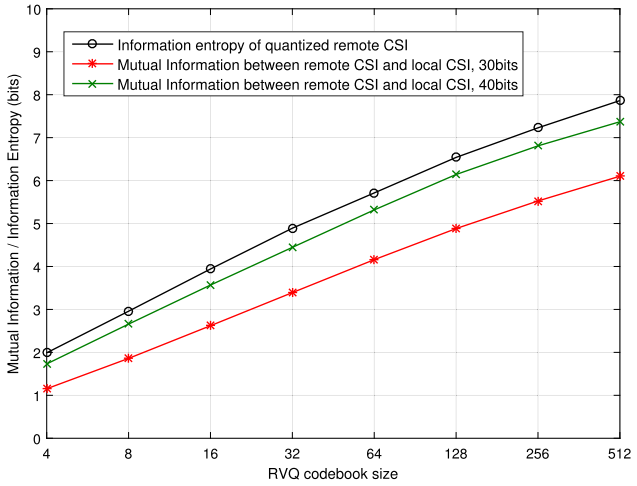
Fig. 1. Scenario for ray-tracing based simulations. Two BSs are placed at different sites, and UEs are uniformly distributed in the outdoor square range covered by the target BS.

B. CRLB Analysis Based on One-Ring Channel Model

To gain more insight into the channel inference problem and performance, a theoretical analysis based on widely-adopted channel models is conducted in this subsection. Consider a communication setup (see Fig. 3) where the UE is surrounded by local scatterers. Let $\boldsymbol{h}_{\mathrm{s}}$ be the CSI between the UE and



(a)



(b)

Fig. 2. Analysis on the relationship between local CSI and remote CSI. The remote CSI is quantized by (a) DFT based codebook and (b) RVQ based codebook, while the local CSI is discretized according to a scalar quantization scheme proposed in [14]. The mutual information between the local observable CSI and the remote CSI is close to the information entropy of the latter provided 40 bits for local CSI quantization, while the average canonical correlation coefficient between remote channel vectors is relatively small.

source BSs. Let there be a target BS, with CSI $\boldsymbol{h}_\mathsf{t}$. The UE has a single antenna element, while the BSs have antenna arrays of size M_s for the source BSs and M_t for the target BS, respectively. The goal is to infer $\boldsymbol{h}_\mathsf{t}$ from the knowledge of $\boldsymbol{h}_\mathsf{s}$. Such an inference problem from $\boldsymbol{h}_\mathsf{s}$ to $\boldsymbol{h}_\mathsf{t}$ might be ill-posed, in the sense that any of the following conditions is not satisfied:

- For any $\boldsymbol{h}_\mathsf{s}$, there exists a unique $\boldsymbol{h}_\mathsf{t}$ that corresponds to the channel realization.
- The mapping from $\boldsymbol{h}_\mathsf{s}$ to $\boldsymbol{h}_\mathsf{t}$ is steady, i.e., given a small error of $\boldsymbol{h}_\mathsf{s}$, the corresponding error of $\boldsymbol{h}_\mathsf{t}$ is also limited.

In practice, we can usually relax the first condition to that if the mapping is not unique, then the mapping error is, to some extent, acceptable. The inference problem is studied by considering the physical propagation environment. Define the propagation channel as a vector of parameters $\boldsymbol{z}_\mathsf{p}$, consisting of, e.g., scatterer locations, reflection attenuation

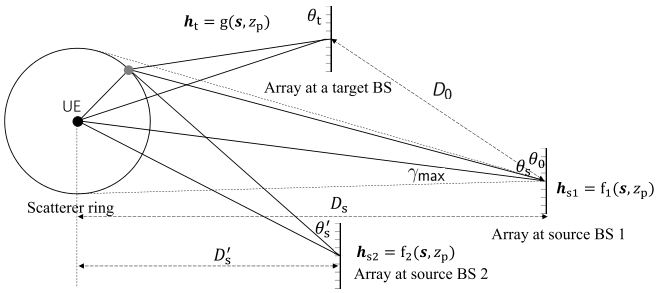


Fig. 3. A one-ring model based CSI inference scenario illustration. Two source BSs with known CSI are described; the CSI at the target BS is unknown and to be inferred.

factors, etc. Then both the CSIs at the source or target BS can be expressed as a function of $\boldsymbol{z}_\mathsf{p}$ and UE location $\boldsymbol{l}$ based on the same methodology of ray-tracing models, i.e.,

$$\boldsymbol{h}_\mathsf{s} = \mathrm{f}(\boldsymbol{l}, \boldsymbol{z}_\mathsf{p}) \quad \text{and } \boldsymbol{h}_\mathsf{t} = \mathrm{g}(\boldsymbol{l}, \boldsymbol{z}_\mathsf{p}), \tag{3}$$

where $\mathrm{f}(\cdot)$ and $\mathrm{g}(\cdot)$ denote the function mappings from the physical environment and UE location to the CSIs at the source BSs and target BS, respectively.

The general model described in (3) is implicit and therefore infeasible for theoretical analysis. Towards this end, we adopt a model-based approach which essentially transforms the CSI inference problem to a parameter extraction problem based on a well-defined channel model; specifically, the one-ring ray-tracing channel model is used where the scatterers are assumed to be placed within a ring of radius $r_{\max}$ (only single-scattering is considered). The UE is at the center of the scatterer-ring. The received signal $\boldsymbol{y} = [y_1, \ldots, y_M]^\mathsf{T}$ at a BS can be written by (denote the channel vector by $\boldsymbol{h} = [h_1, \ldots, h_M]^\mathsf{T}$)

$$\begin{aligned} y_i &= \sqrt{P_\mathrm{tx}} h_i + n_i, \\ h_i &= \sum_{k=1}^{N_\mathrm{p}} g_{ki} \exp\left(-\frac{j2\pi}{\lambda} d_{ki}\right), \end{aligned} \tag{4}$$

where k is the index of MPC going through the k-th scatterer, $g_{ki} \triangleq \frac{\lambda}{4\pi d_{ki}}$ denotes the channel gain due to pathloss based on Friis' law, the Gaussian additive noise is denoted by n_i with variance of σ^2 (the sounding signal is omitted for simplification and assumed with effective transmit power of P_tx), and the path distance of the k-th MPC received at the i-th antenna is denoted by d_{ki} which equals

$$d_{ki} = \xi_{\mathsf{t}k} + \xi_{ki} \tag{5}$$

with $\xi_{\mathsf{t}k}$ denoting the distance from the UE to the k-th scatterer and ξ_{ki} denoting the distance from the k-th scatterer to the i-th receive antenna. Assuming that the angle spread at the UE is relatively small and the UE is in the far-field, i.e., $D \gg r_{\max}$ and $D \gg \frac{2\ M_\mathsf{s}^2 \delta^2}{\lambda}$ where δ is antenna spacing [26], the channel gains of all MPCs are therefore approximately identical, i.e., $g_{ki} = g$, $\forall 1 \le k \le N_\mathrm{p}, 1 \le i \le M$, and

$$d_{ki} \approx \xi_k + i\delta \cos\gamma_k. \tag{6}$$

Denote by ξ_k the distance from the UE to the antenna array (a reference point such that the above equation is upheld) passing

through the k-th scatterer, and γ_k is the AoA of the k-th MPC. Based on this approximation, (4) can be re-written as

$$y_i = \sum_{k=1}^{N_{\mathrm{P}}} g\sqrt{P_{\mathrm{tx}}} \exp\left(-\frac{j2\pi\delta i}{\lambda}\cos\gamma_k + j\phi_k\right) + n_i, \quad (7)$$

where $\phi_k = -\frac{j2\pi}{\lambda}\xi_k$, and $g = \frac{\lambda}{4\pi D}$ represents the pathloss.

The APS seen at a site is defined as

$$S(\gamma) = g\mu(\gamma)p(\gamma), \quad (8)$$

where γ is the AoA to the site and $S(\gamma)$ is normalized such that

$$\int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} S^2(\gamma)\mathsf{d}\gamma = g^2. \quad (9)$$

The scatterer angular distribution is characterized by $\mu(\gamma)$, and the probability that the MPC with AoA γ is observable (not blocked) at the site is $p(\gamma)$. For example, if we assume that the scatterers are continuously placed on the ring with radius r_{max}, then [27]

$$\mu_{\mathsf{ring}}(\gamma) = \frac{2}{\sqrt{\gamma_{\mathsf{max}}^2 - (\gamma-\theta)^2}}, \quad (10)$$

where $\theta - \gamma_{\mathsf{max}} \le \gamma \le \theta + \gamma_{\mathsf{max}}$ and θ is the mean AoA of the UE and γ_{max} denotes the maximum angular spread. A typical form of $p(\gamma)$ is e.g., uniform on the disk with a radius of r_{max}. The channel array response is written as

$$y_i = \int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} \sqrt{P_{\mathrm{tx}}} S(\gamma) \times \exp\left(-\frac{j2\pi\delta i}{\lambda}\cos(\gamma+\theta) + j\phi_\gamma\right)\mathsf{d}\gamma + n_i. \quad (11)$$

Combing with (3), we denote

$$\begin{aligned} h_{\mathsf{s},i} &= \mathrm{f}\left(\theta_{\mathsf{s}}, S_{\mathsf{s}}(\gamma), \phi_{\mathsf{s},\gamma}, r_{\mathsf{max,s}}\right), \\ h_{\mathsf{t},i} &= \mathrm{g}\left(\theta_{\mathsf{t}}, S_{\mathsf{t}}(\gamma), \phi_{\mathsf{t},\gamma}, r_{\mathsf{max,t}}\right), \end{aligned} \quad (12)$$

where the subscripts $(\cdot)_{\mathsf{s}}$ and $(\cdot)_{\mathsf{t}}$ denote the source BS and the target BS, respectively. The function mapping of $\mathrm{f}(\cdot)$ and $\mathrm{g}(\cdot)$ are substantiated by (11). Note that the model of (12) allows, e.g., different angular spreads, different visibility of scatterers, different APSs at the source BS and the target BS, by distinct r_{max}, p_γ, $\mu(\gamma)$, respectively. It is thus significantly more general than a model valid within a stationarity region, which assumes that only the phases of MPCs vary (due to phase shifts related to the different run lengths).

Based on the above model, the channel inference task can be stated concretely below:

P1: Estimate $\boldsymbol{h}_{\mathsf{t}}$, given $\boldsymbol{h}_{\mathsf{s}}$, s.t., Eq. (12) is satisfied. (13)

It is observed that the implicit channel inference problem in (3) is transformed into a parameter extraction and estimation problem in **P1**, based on the adopted channel models; this allows us to derive the CSI inference CRLB, which indicates the CSI inference accuracy. Equivalently, the inverse of the CRLB, representing the Fisher information, indicates the amount of information that the local CSI carries about the remote CSI.

To further simplify the model and focus on the main goal of CSI inference, two reasonable assumptions are made, which describe the capability boundary of CSI inference, i.e., parameters that can be inferred and those that cannot. Specifically,

- The phase of an arrival MPC, i.e., ϕ_γ, is random (i.i.d. among MPCs) and cannot be inferred. This is a practical and realistic consideration given that the phase of an electromagnetic wave shifts dramatically even with a slight movement of the UE (several wavelengths).
- The observable scatterers seen at the target BS, i.e., $S_{\mathsf{t}}(\gamma)$, $r_{\mathsf{max,t}}$, cannot be inferred based on the observation at the source BS, which provides little information about whether an MPC is obstructed seen at the target BS. Instead, this information can be obtained by using relatively infrequent probing signals by the target BS given that the scattering environment is constant inside the stationarity region (typical size of tens of meters in urban areas).[2]

Given these assumptions, we focus on analyzing the inference performance with respect to the AoA at the target BS θ_{t} to obtain theoretical results. The end goal is to derive the CRLB of the estimation of θ_{t}, towards which we first solve the inverse problem of $h_{\mathsf{s},i} = \mathrm{f}\left(\theta_{\mathsf{s}}, S_{\mathsf{s}}(\gamma), \phi_{\mathsf{s},\gamma}, r_{\mathsf{max,s}}\right)$, and then relates to the AoA of θ_{t} based on geometry.

Theorem 1: Define $\boldsymbol{z} = \{S(\gamma), \gamma_{\mathsf{max}}, \theta_{\mathsf{s}}\}$, then the Fisher information matrix (FIM) of $\boldsymbol{z}$ can be approximated by:

$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{ij} = K\mathsf{tr}\left[\boldsymbol{C}_{y_{\mathsf{s}}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_{\mathsf{s}}}^{-1}\boldsymbol{D}_i\right], \quad (14)$$

where K is the number of sampling points, $\boldsymbol{C}_{y_{\mathsf{s}}} = \mathbb{E}[\boldsymbol{y}_{\mathsf{s}}\boldsymbol{y}_{\mathsf{s}}^{\mathsf{H}}]$ and $\boldsymbol{D}_i = \frac{\partial \boldsymbol{C}_{y_{\mathsf{s}}}}{\partial z_i}$.

Proof: The proof is given in Appendix A. □

Based on (14), the CRLB of AoA (θ_{s}) and UE distance D_{s} can be obtained by extracting the diagonal entries of the inverse FIM. The CRLB of θ_{t} can be therefore obtained in the following. Without loss of generality, assume the coordinates of the source BS and target BS are (0, 0) and $(D_0\cos\theta_0, D_0\sin\theta_0)$ in a two-dimensional Cartesian coordinate system, respectively. Based on θ_{s} and D_{s}, the coordinate of the UE can be expressed in two ways as follows:

$$l_{\mathsf{x}} = D_{\mathsf{s}}\cos\theta_{\mathsf{s}} = D_0\cos\theta_0 + D_{\mathsf{t}}\cos\theta_{\mathsf{t}}, \quad (15)$$

$$l_{\mathsf{y}} = D_{\mathsf{s}}\sin\theta_{\mathsf{s}} = D_0\sin\theta_0 + D_{\mathsf{t}}\sin\theta_{\mathsf{t}}. \quad (16)$$

The AoA at the target BS can be hence represented by D_{s} and D_{t} as follows

$$\theta_{\mathsf{t}} = \arctan\left(\frac{D_{\mathsf{s}}\sin\theta_{\mathsf{s}} - D_0\sin\theta_0}{D_{\mathsf{s}}\cos\theta_{\mathsf{s}} - D_0\cos\theta_0}\right). \quad (17)$$

[2]The requirement for pilot signals to probe the MPC distributions is mainly assumed for theoretical analysis; based on the deep learning implementation, these pilots are usually unnecessary to obtain good CSI inference performance.

The CRLB of θ_{t} can be expressed as:

$$\begin{aligned}\mathsf{CRB}(\theta_{\mathsf{t}}) &= \left(\frac{\partial\theta_{\mathsf{t}}}{\partial D_{\mathsf{s}}}\right)^2 \mathsf{CRB}(D_{\mathsf{s}}) + \left(\frac{\partial\theta_{\mathsf{t}}}{\partial\theta_{\mathsf{s}}}\right)^2 \mathsf{CRB}(\theta_{\mathsf{s}}) \\ &= \frac{D_0^2\sin^2(\theta_0-\theta_{\mathsf{s}})}{D_{\mathsf{t}}^4}\mathsf{CRB}(D_{\mathsf{s}}) \\ &\quad + \frac{D_{\mathsf{s}}^2\left(D_{\mathsf{s}}-D_0\cos(\theta_0-\theta_{\mathsf{s}})\right)^2}{D_{\mathsf{t}}^4}\mathsf{CRB}(\theta_{\mathsf{s}}), \quad (18)\end{aligned}$$

where $\mathsf{CRB}(D_{\mathsf{s}})$ and $\mathsf{CRB}(\theta_{\mathsf{s}})$ are obtained from (14).

The calculation of $\boldsymbol{D}_i$ in (14) depends on the channel model, e.g., scattering distribution inside the ring. It seems elusive to calculate closed-form expressions for general models. To gain some insights, we consider several special cases in the following, where the angular spreads $\gamma_{\mathsf{max,s}}$ and $\gamma_{\mathsf{max,t}}$ are small and approach zero, i.e., a LoS MPC only.

Corollary 1: In the scenario of one source BS and one LoS MPC only, the CRLB of θ_{t} satisfies

$$\mathsf{CRB}_1(\theta_{\mathsf{t}}) \sim \frac{1}{M_{\mathsf{s}}}. \quad (19)$$

Proof: The proof is given in Appendix B. □

Therefore the inference error is reciprocal with M_{s} in this scenario, while the following CRLB analysis, which accounts for two separate local sites (with known CSI) to infer the CSI at a remote site, shows that by using the observations at two geographically separated sites, the inference error can be significantly reduced.

Corollary 2: In the scenario of two source BSs and one LoS MPC each, the CRLB of θ_{t} satisfies

$$\mathsf{CRB}_2(\theta_{\mathsf{t}}) \sim \frac{1}{M_{\mathsf{s}}^3}. \quad (20)$$

Proof: The proof is given in Appendix C. □

Remark 1: The CRLB analysis shows that the accurate estimation of the AoA at the target BS can be obtained provided enough channel observations at the source BS, e.g. deploying large antenna arrays. As for the non-line-of-sight (NLoS) scenario, obtaining closed-form expressions for the CRLB is a challenging task, and the preceding asymptotic laws may not hold in that case.

Remark 2: While this paper focuses on inference based solely on spatial domain signals, one direction that is worth studying is the CSI inference performance incorporating wideband signals such that the time-of-arrival (ToA) can also be estimated.

III. Remote Channel Inference Framework

Inspired by the dependence among remote channels, we proposed a remote channel inference framework [28] in order to reduce the pilot overhead. Fig. 4 shows a typical urban communication scenario, where massive connectivity and high mobility of users must be supported, which brings significant challenges for traditional pilot-aided CSI acquisition methods. In order to provide high data rate and low latency for mobile users, small BSs are densely deployed, each with multiple antennas. Assume that the BSs adopt a closed-loop CSI acquisition scheme, i.e., first broadcasting pilot resources for DL CSI estimation and then receiving UL CSI feedback. However, limited pilot resources cannot satisfy the large pilot demand of those densely-deployed cells, and therefore the large pilot overhead is a major obstacle to further improve the system quality of service. To address this issue, a remote channel inference framework is proposed, where only source BSs are assigned with pilot resources. The CSI of target BSs is inferred, based on CSI estimations at source BSs.

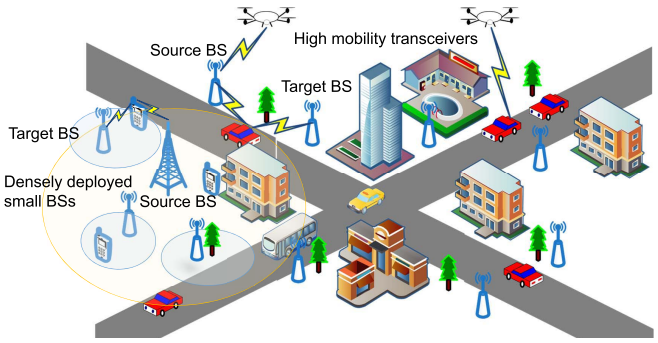


Fig. 4. A typical urban communication scenario, wherein the CSI estimation at source BSs can help infer the CSI of target BSs.

In the proposed framework, source BSs are responsible for UE attachment. Consequently, a UE connects to a source BS and feeds the estimated DL CSI back to the source BS. After that, the source BS can decide whether to serve the UE itself or assign a target BS to serve the UE. If the latter happens, the source BS will infer the CSI of the target BS by the proposed approaches (specified later), and transmit the inferred CSI to the target BS via backhaul links. Based on the inferred CSI, the target BS can beamform to the UE.

We study beamforming for small cells in heterogeneous networks as a use case. The first subsection proposes to infer the beamforming directions of the small cells (target BSs) based on the CSI of the macro BS (source BS). Then the second subsection extends the results to a multi-user scenario, wherein the APS of the target BS is inferred to implement user grouping in order to mitigate the multi-user interference. In practice, the specific NN architecture to be applied depends on the application scenarios, e.g., whether multi-user spatial multiplexing is considered.

A. Beamforming for Small Cells in Two-Tier Heterogeneous Networks

Network densification and software-defined virtualization are key technologies for future wireless systems [4]. The basic idea behind the UDN concept is to reduce the transmission distance between users and BSs, which leads to less pathloss, more spectrum reuse and hence higher data rate. Meanwhile, the *hyper-cellular network* (HCN) is proposed in [29], based on which the control- and data-coverage are separated to provide uniform high quality of service with elastic data-coverage. In HCN, the control-BS (CBS) is responsible for control coverage and several traffic-BSs (TBSs) take care of data coverage. This architecture allows agile sleeping of

TBSs, and thus the energy consumption of the network can be drastically reduced, while the CBS can handle the control signaling to provide seamless coverage.

A major technical challenge for ultra-dense HCN is the strong interference among densely deployed TBSs [30]. Coordinated beamforming is an effective way to mitigate interference [31]. However, timely and accurate CSI acquisition may present a big challenge. Generally, the system exploits pilot signals to acquire the CSI. However, limited pilot resources cannot satisfy the demand of a huge number of densely deployed TBSs (assuming FDD systems are considered, where channel reciprocity cannot be exploited). Furthermore, searching over all possible beam directions of TBSs leads to increased UE energy consumption and delay overhead.

1) Approach: The optimal maximum-ratio-transmission (MRT) beamforming structure, or eigen-based beamforming scheme where only channel statistics are available [32], requires accurate channel estimations to derive the beamforming vector (pattern). In the proposed framework, we use remote channel inference to find the optimal beamforming vector for TBSs. To simplify the approach, a finite and prescribed codebook is involved, and the task is to select the beamforming vector (codeword) from the codebook. Thereby, the problem is to find the optimal beam pattern for TBSs from the prescribed codebook according to the observable CSI of the CBS. Here we define the beamforming pattern that has the maximum signal-to-noise ratio (SNR) to the desired user as the optimal one (note that a single-user scenario is considered in this use case).

We propose a remote beamforming inference framework to address this problem. In this scheme, in order to reduce pilot overhead, only the CBS sends DL reference signals. Consequently, the UE connects to the CBS and sends estimated DL CSI feedback to the CBS. After that, the central controller at the CBS will decide which TBS should serve the UE. Then the selected TBS will exploit beamforming to transmit DL data. We resort to a data-driven channel learning approach and exploit the CSI between the CBS and UEs to infer the CSI between TBSs and UEs. Specifically, the CBS infers the CSI for the TBS and thereby sends its suggestion for the optimal beamforming vector to the TBS via the backhaul link. Then the TBS creates an initial connection to the UE using the selected beam pattern. If the connection fails in the last stage, the TBS can ask for an updated beam pattern from the CBS or use a high-overhead default beamforming mechanism to finally establish the link. The above situation occurs when CSI changes during the prediction time or the predicted beamforming pattern deviates too much from the optimal one.

NN architectures are known to achieve tremendous success in approximating non-linear functions [24], which are suitable for CSI inference. The training of the NN is done as follows. For a given cellular network, channel estimations for both the CBS and TBSs can be collected at sampling points, e.g., by conducting an offline CSI data collection or crowdsourcing to certain users at random locations; the data are then utilized for training of the NN.

An NN with four hidden layers is adopted, and the input of the NN is the pre-processed observable CSI at the CBS. Preprocessing is needed to extract features from raw CSI so as to speed up the training process. Specifically, we transform channel vectors into the angular domain by fast Fourier transform (FFT), then extract the amplitude information, take the logarithm, and normalize for each angular bin. The first step is to obtain a better description on the physical environment by transforming the input CSI the into angular domain. The following steps are to generate Gaussian-like inputs to the NN and scale the range of the inputs, which can help improve the training speed. The number of input features is the same as the number of CBS antennas. The outputs are the selection probabilities of all beam patterns at a TBS from a DFT-based beamforming codebook. The size of the codebook can be larger than the number of TBS antennas in order to obtain a higher angular resolution. Specifically, assuming N_{b} codewords, with N_{b} larger than the number of TBS antennas N_{T}, the codewords are the first N_{T} columns of an $N_{\mathrm{b}} \times N_{\mathrm{b}}$ DFT matrix. Since the CSI of the TBS is known for each training sample, it is straightforward to find the optimal beam pattern from the codebook based on the training data. The cost function is the cross-entropy [33] function, expressed as

$$L = -\frac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} \sum_{j=1}^{N_{\mathrm{b}}} u_{ij} \log(v_{ij}). \tag{21}$$

Here N_{s} is the number of samples, and u_{ij} and v_{ij} represent the ground truth and the predicted value of the selection probability for the j-th beamforming vector at the i-th sample, respectively. We use a back-propagation algorithm to calculate gradients and obtain the weight update based on those gradients, which is implemented based on the Tensorflow architecture [34]. The activation function for the hidden layers is *Relu*, while the activation function for the output layer is *softmax*. Dropout [35] is utilized to prevent the network from overfitting. We also adopt the batch normalization [36] technique to accelerate the training process.

After the NN has been trained, predictions can be made to select optimal beam patterns. We feed extracted features into the NN and the output vector represents the selection probabilities of all the beam patterns. The beamforming vector with the largest probability is the output. The NN is trained offline, while the low-complexity beamforming inference is made online, and thus the proposed scheme only introduces small delay overhead in comparison with other schemes such as beam sweeping; the impact of delay on performance is studied and resolved in our related work [37]. When major environmental changes happen, the training data can be re-collected by crowd-sourcing techniques or in an online manner with pilot-aided channel estimation.

The aforementioned beamforming inference method is proposed for beamforming at the transmitter side; this method can also be applied similarly for receive beamforming at the UE side. In this case, the output of the NN denotes the selection probability of each receiver beam direction and the input of the NN should be a pre-processed channel matrix. Two-dimensional DFT (the Kronecker product of two simple

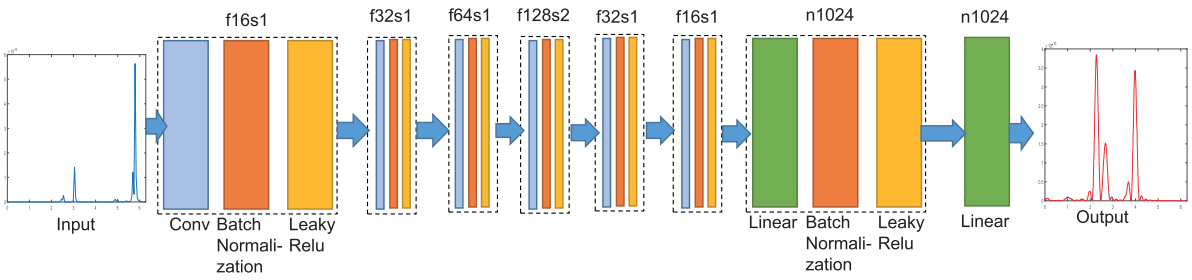


Fig. 5. The architecture of the CNN-based APS inference framework. f and s stand for the number of feature maps (output of a convolutional layer) and the stride length for convolutional layers, respectively. The number of hidden units for fully-connected layers is denoted by n.

DFTs) is adopted to transform the channel matrix into the angular domain and for simplicity, the pre-processed channel matrix is reshaped into a vector and fed into the NN. Exploiting similar NN architectures and training methods as transmit beamforming, the receive beamforming vector at the UE for the TBS can be inferred.

2) Enhancements for Prediction Accuracy: First, note that UE locations have a big influence on the system performance. Specifically, UEs that have shorter distances to TBSs tend to have worse performance. The reason is that the estimation error of AoD is inversely proportional to the distances between UEs and TBSs given certain estimation error of UE locations. In extreme cases where UEs are very close to TBSs and far from the CBS, beam pattern selection can be a really difficult decision based on our framework. For these UEs, a wider beam may be a better choice to trade angular resolution for the robustness of estimation.

Therefore, providing more information about UE locations can potentially help to improve prediction accuracy. Inspired by Section II-B, we propose to include more CBSs into the inference framework. In long-term-evolution (LTE) standards, UEs can measure the CSI of neighboring BSs and send CSI reports to the serving BS. In that way, the CBS can have CSIs of UEs to other CBSs and thus it can exploit this information to improve beamforming inference accuracy and hence the system performance.

A set of candidate beam patterns can be obtained through the learning approach by selecting the S candidate beams with the highest probabilities. Specifically, the output of the NN is the soft information indicating the probabilities for selecting each of the beam patterns. If however the predicted beam pattern fails to work, the beam with the second highest selection probability is most likely to work. After a set of candidate beams is fed to the TBS, it can combine the inference with conventional beam sweeping to further select the optimal beam with much less overhead, since only the beams in the candidate list need to be further evaluated. Either way, our proposed scheme can reduce, if not eliminate completely, the training overhead and training delay.

B. APS Inference for User Grouping in Downlink Multi-User MIMO Systems

The aforementioned learning-based beamforming can achieve near-optimal beamforming gain for single-user systems, but the multi-user interference cannot be eliminated on condition that only the optimal beam direction towards a certain user is known. One simple solution is to allocate orthogonal resources for different users, at the cost of losing the potential spatial multiplexing gain. To further improve the performance of the proposed channel inference framework, extensions are made to infer the APS of the TBS, which characterizes the channel power distribution in the angular domain, so as to suppress the multi-user interference. Based on the inferred APS, users with small overlap in terms of APS can be grouped to transmit together in the same resource blocks (RBs), since they are spatially separable. Different user groups can be assigned with orthogonal resources to suppress interference. In general, the proposed method can reduce the multi-user interference and achieve spatial multiplexing gain with very little pilot overhead.

For simplicity, we assume that user devices are equipped with a single antenna, and hence the APS is a one-dimensional vector. Even then, the inference of APS is still much more challenging compared with inferring the strongest beam directions. The good news is that the APS of the MIMO channel shows block-sparsity, i.e., APS is typically sparse in the angular domain, which inspires the usage of CNN. Specifically, as shown in Fig. 5, a DNN with 6 convolutional layers and 2 fully-connected layers is trained, where the input and output of the DNN are the APS of CBS and TBS, respectively. The size of the one-dimension convolutional kernel is 5×1. The activation function of each layer is *leaky Relu* [38], and dropout is utilized to prevent the model from overfitting. The optimization target is to minimize the mean square error (MSE) of the APS of the target BS, while the Adam optimizer is applied for updating the gradients of the parameters. Additionally, batch-normalization is adopted for accelerating the training speed of the DNN.

To deal with multi-user interference, the conventional zero-forcing or minimum MSE precoder relies on accurate instantaneous CSI, which entails enormous CSI acquisition overhead. The proposed APS inference can reduce the multi-user interference by user-grouping only based on the inferred APS. The basic idea is to group users with small overlap in terms of APS. Users in the same group can be served simultaneously, for the reason that their spatial directions are non-overlapped.

Inspired by our previous work [39], a graph-coloring-based user grouping algorithm is applied to divide the users into spatial compatible sets. The first step is to detect the dominant

directions for all users. Given an APS, the dominant directions are defined as the components with significant values, i.e.,

$$\mathcal{S}_k = \{i | S_k(i) \geq \epsilon\}, \tag{22}$$

where $S_k(i)$ denotes the i-th components in the APS of the k-th user and $\mathcal{S}_k$ denotes the set of the dominant directions of the k-th user. ϵ is an adaptive threshold, which satisfies the constraint

$$\frac{\int_{\mathcal{A}} S_k(i)di - \int_{\mathcal{S}_k} S_k(i)di}{\int_{\mathcal{A}} S_k(i)di} = \eta. \tag{23}$$

Here $\mathcal{A}$ denotes the set of all directions, and η is a predefined threshold and can be used to control the size of different groups. A larger η leads to a larger set of dominant directions, and hence a smaller group size. Users with no overlap in terms of the dominant directions experience small interference towards each other, so a graph-coloring algorithm is adopted on the detected dominant directions. Specifically, an undirected conflict graph is built wherein users can be regarded as vertices. An edge connects two vertices if and only if the dominant directions of the two corresponding users overlap. Consequently, the user-grouping problem can be transformed into a graph-coloring problem, and the non-overlap constraint is equivalent to the requirement that the connected vertices should be assigned with different colours. Graph-coloring is a well-studied problem and for simplicity, a greedy algorithm [40] is applied here. As shown in Alg. 1, the coloring set is first initialized by a vector with all zero elements, and the color set is $\varnothing$. After sorting the users based on the smallest index of the dominant directions of APS, the sorted users are colored in a greedy manner, avoiding the same color for neighboring vertices. The complexity of the algorithm is $\mathcal{O}(N_{\mathsf{u}}^2)$, where N_{u} denotes the number of users.

Algorithm 1: Greedy Coloring User Grouping (GCUP)

Input: Dominant directions of the APS $\mathcal{S}_k$, $k = 1, 2, \cdots, N_{\mathsf{u}}$
Output: Grouping results $\boldsymbol{c}$
Initialize: set $\boldsymbol{c} = [0, 0, \ldots, 0]$, the corresponding color set $\mathcal{C} = \varnothing$;
Generate the conflict graph $\mathcal{G}$ based on $\mathcal{S}_k$, $k = 1, 2, \cdots, N_{\mathsf{u}}$;
Sorting: sort all users in an ascending order according to the smallest index of the dominant directions of APS;
Coloring:
for $k = 1, 2, \cdots, N_{\mathsf{u}}$ **do**
  Color the k-th vertex with a color different to the neighboring vertices from $\mathcal{C}$;
  if *no color is available* **then**
    Assign a new color for the vertex and add the color to $\mathcal{C}$;
  end
  $\boldsymbol{c}_k = m$, m denotes the color of the k-th vertex;
end

Users in the same group occupy the same RBs, while different user groups occupy orthogonal resources. Consider that the total transmission power is P_0, and there are totally N_{u} users divided into N_{g} groups. In each group, an equal power allocation scheme is adopted. Consequently, the transmit signal for the m-th group can be expressed as:

$$\boldsymbol{s}_m = \sum_{i=1}^{k_m} \sqrt{\frac{P_0}{k_m}} \boldsymbol{\omega}_{im} x_{im}, \tag{24}$$

where k_m denotes the number of users in the m-th group and satisfies $\sum_{m=1}^{N_{\mathsf{g}}} k_m = N_{\mathsf{u}}$. $\boldsymbol{\omega}_{im}$ and x_{im} denote the beamforming vector and transmitted symbols of the i-th user in the m-th group. $\boldsymbol{\omega}_{im}$ is from a predefined DFT-based codebook standing for the optimal beam direction based on the inferred APS. The transmitted symbols of different users are assumed to be uncorrelated. In other words, x_i satisfies $\mathbb{E}(\bar{x_i} x_i) = 1$ and $\mathbb{E}(\bar{x_i} x_j) = 0$, for any $i \neq j$.

The received signal for the i-th user can be expressed as:

$$\boldsymbol{y}_{im} = \boldsymbol{h}_{im}^{\mathsf{H}} \boldsymbol{s}_m + \boldsymbol{n} = \sum_{i=1}^{k_m} \sqrt{\frac{P_0}{k_m}} \boldsymbol{h}_{im}^{\mathsf{H}} \boldsymbol{\omega}_{im} x_{im} + \boldsymbol{n}, \tag{25}$$

where $\boldsymbol{h}_{im}$ denotes the DL CSI of the i-th user in the m-th group.

The signal to interference plus noise ratio (SINR) of the i-th user can be expressed as:

$$\text{SINR}_{im} = \frac{\frac{P_0}{k_m} |\boldsymbol{h}_{im}^{\mathsf{H}} \boldsymbol{\omega}_{im}|^2}{\frac{P_0}{k_m} \sum_{j=1, j\neq i}^{k_m} |\boldsymbol{h}_{im}^{\mathsf{H}} \boldsymbol{\omega}_{jm}|^2 + N_0}, \tag{26}$$

where N_0 denotes the noise power.

Provided that the resources occupied by different user groups are proportional to the group size, the constrained sum rate can be expressed as:

$$R = \frac{1}{N_{\mathsf{u}}} \sum_{m=1}^{N_{\mathsf{g}}} k_m \sum_{i=1}^{k_m} \log_2(1 + \text{SINR}_{im}) \mathbb{1}(\text{SINR}_{im} \geq \zeta), \tag{27}$$

where ζ denotes the minimal SINR for effective transmission; note that an SNR below ζ does not allow any meaningful data rate in practice, since a limited modulation-coding selection is considered.

IV. Simulation Results

In this section, ray-tracing based simulations are conducted to evaluate the performance of the proposed framework. Specifically, we generate CSI data from Wireless InSite in a typical urban outdoor scenario. The carrier frequency is 2.6 GHz unless otherwise stated. A CBS with uniform linear array and half wavelength antenna spacing is deployed at a height of 35 m. Without loss of generality, a TBS with uniform linear array and half wavelength antenna spacing is deployed at a height of 5 m within the coverage of the CBS. The number of antennas for the CBS and TBS are, unless stated specifically, 128 and 32, respectively. The UEs are uniformly distributed in a rectangular area under the coverage of the TBS, equipped with a 1×8 antenna array. If not specified, the sizes of DFT-based beamforming codebook for the TBS and the UE are 256 and 8, respectively. There are totally around 5×10^4

samples, and 20% of them are test sets; the rest are used for training.[3]

A. Simulation Results for Remote Beamforming Inference

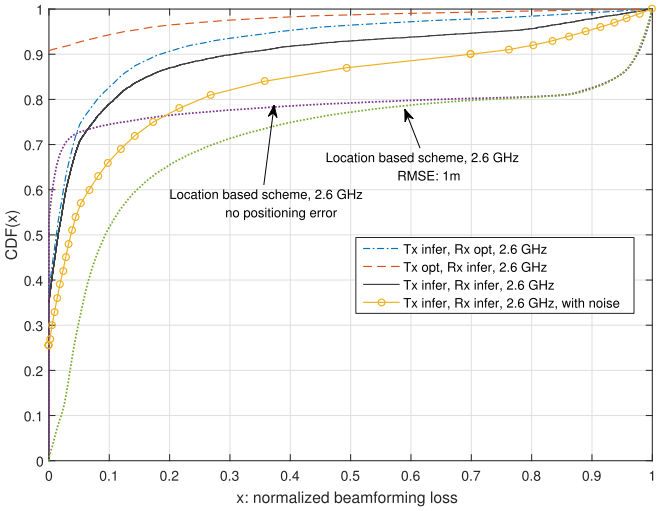
In the single user case, the performance metric shown in the figures is the normalized beamforming loss, which can be expressed as

$$L_{\mathrm{b}} = 1 - \frac{\|\boldsymbol{w}_{\mathrm{r,infer}}^{T}\boldsymbol{H}\boldsymbol{w}_{\mathrm{t,infer}}\|}{\|\boldsymbol{w}_{\mathrm{r,opt}}^{T}\boldsymbol{H}\boldsymbol{w}_{\mathrm{t,opt}}\|}, \tag{28}$$

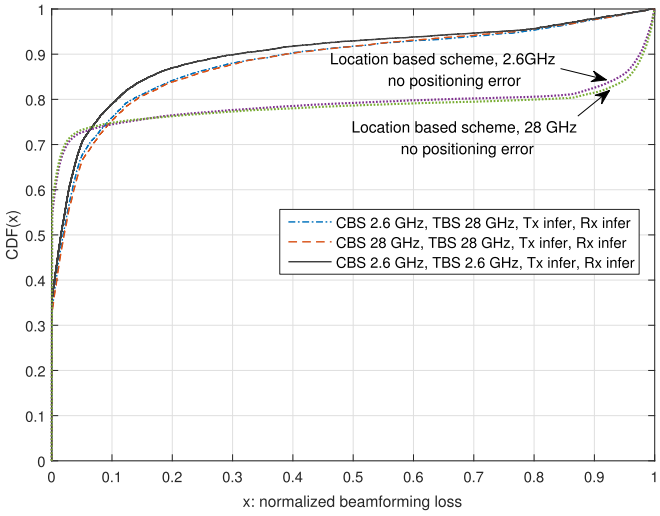
where $\boldsymbol{w}_{\mathrm{r,infer}}$ and $\boldsymbol{w}_{\mathrm{t,infer}}$ denote the inferred receive and transmit beamforming patterns, respectively, while $\boldsymbol{w}_{\mathrm{r,opt}}$ and $\boldsymbol{w}_{\mathrm{t,opt}}$ denote the optimal receive and transmit beamforming patterns from the codebook. We take this as our performance metric, along with prediction accuracy, because there is a chance that UEs have almost the same received signal strength using two different beamforming vectors, i.e., when these two beamforming vectors are very close such that the impact of choosing either one does not affect the performance. The optimal beamforming pattern in this case can be, therefore, either one.

Fig. 6 shows the cumulative distribution function (CDF) of the normalized beamforming loss of different algorithms. The red dashed line reflects the results of inferring the receive beamforming while the transmit precoding is optimal by exhaustive sweeping. The gap between the red dashed line and the line $y = 1$ represents the performance loss of receive beamforming inference. Similarly, the gap between the black solid line and the line $y = 1$ represents the performance loss of inferring the beam directions at both the transmitter and the receiver side. The average normalized beamforming loss in that case is 10.05%, which is acceptable given massive beam pattern choices. When the channel estimation error of the CBS is considered and a 5% Gaussian noise is added to the input CSI vectors, an average normalized beamforming loss of 16.64% can be achieved by the proposed method. Compared with location based beamforming, where the beam directions are calculated based on the locations of transceivers and assuming LoS propagation channels, the proposed method shows evident improvement and avoids the extra overhead of gathering positioning information [41]. The average beamforming loss of location based beamforming is 21.12% even without any positioning error, while the loss increases to 28.74% when the root mean square error (RMSE) of positioning error is 1 m. It can also be observed that the location-based beamforming with exact location information has a larger probability to obtain the optimal beam directions compared with our proposed method, but more than 20% of the users suffer from very poor beamforming performance. Simulation results in different frequency bands are exhibited in Fig. 6(b). Although the remote beamforming inference scheme exhibits a sight performance decline in the mm-wave band compared with the sub 6 GHz band, it still shows notable superiority over location-based beamforming. Besides, the proposed scheme can easily provide comparative beamforming gain in a typical scenario that CBS and TBS work in different bands, owing to its learning-based model-free feature.

[3]The corresponding channel data and code files can be found on https://gitlab.com/xgwx/Channel-Infer.



(a)



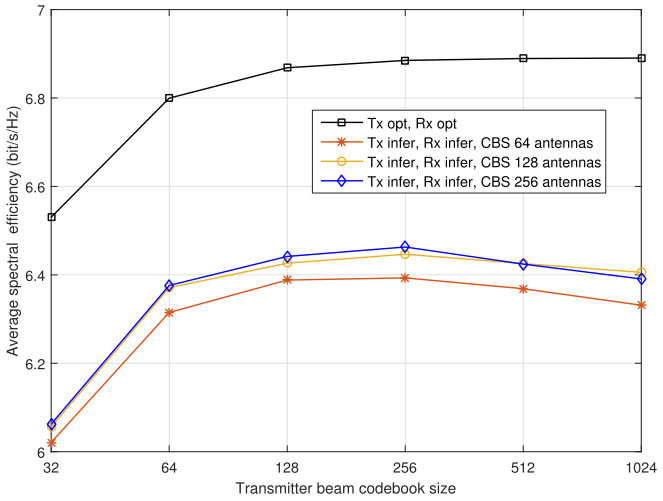
(b)

Fig. 6. CDF of normalized beamforming loss. Tx and Rx denote transmit precoding and receive beamforming, respectively, while infer and opt stand for inferring from remote BS and the optimal one from the codebook.

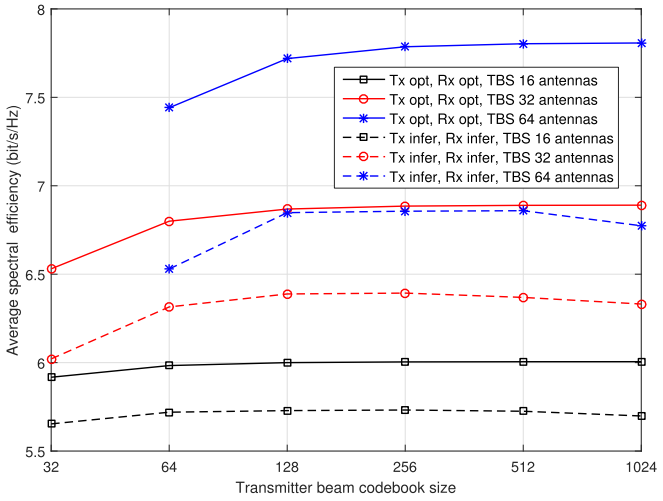
Fig. 7 shows the relationship between the average SE and the beamforming codebook size. The average SE is calculated by the Shannon formula, which can be expressed as:

$$\mathbb{E}(S_{\mathrm{e}}) = \frac{1}{N}\sum_{i=1}^{N}\log_2\left(1 + \frac{P\|\boldsymbol{w}_{\mathrm{ri}}^{T}\boldsymbol{H}_i\boldsymbol{w}_{\mathrm{ti}}\|^2}{N_0}\right), \tag{29}$$

where N, P and N_0 denote the number of test samples, transmitting power and noise power, respectively. It is straightforward to see that the optimal beamforming performance increases with the beam codebook size in Fig. 7(a). However, when it comes to the proposed channel inference method, the average SE first increases with the beam codebook size and then drops. The explanation is that when the beam codebook size increases, the accuracy of inferring the optimal beam directions decreases; the performance improvement that arises

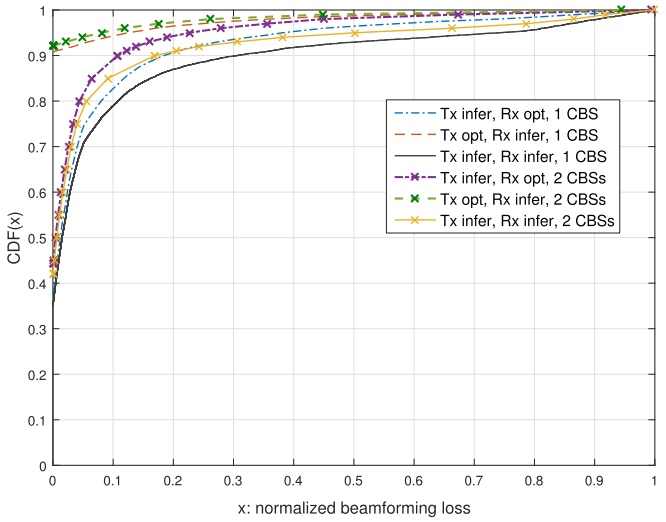


(a)

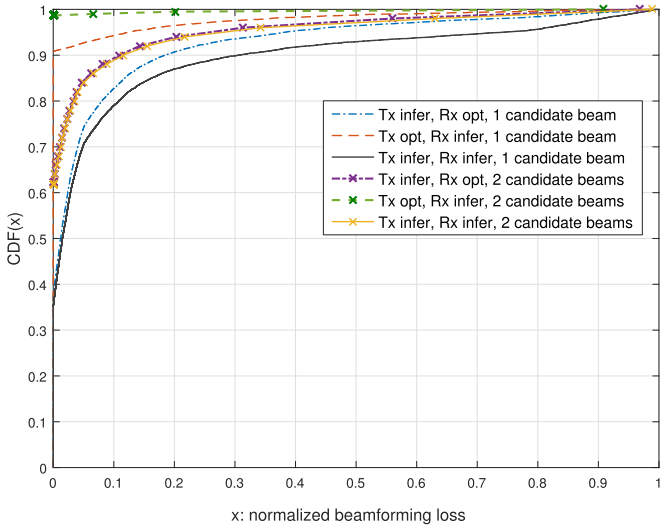


(b)

Fig. 7. Average SE versus beam codebook size with different number of (a) CBS antennas. (b) TBS antennas.



(a)

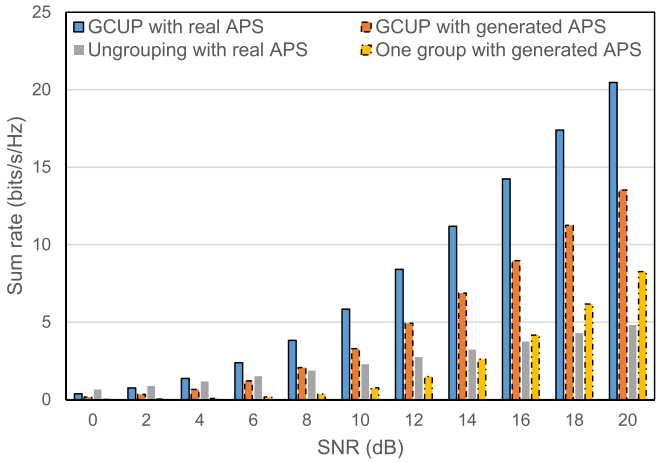


(b)

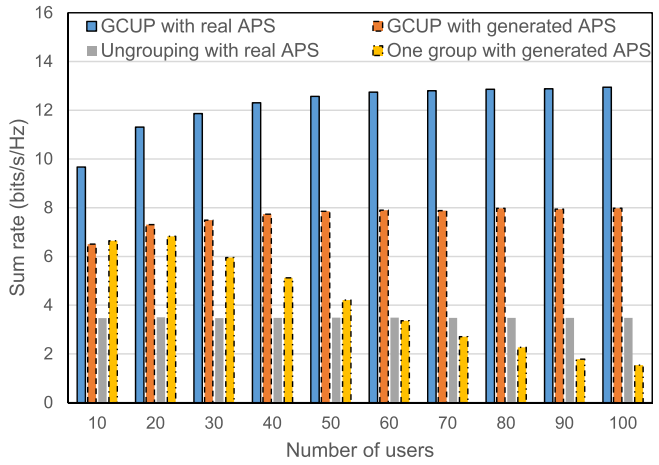
Fig. 8. (a) CDF of normalized beamforming loss with multiple CBSs. (b) CDF of normalized beamforming loss with multiple candidate beams.

from increasing the beam direction resolution cannot counteract the loss from increasing the labels of the learning task when the beam codebook size is relatively large, and hence the average SE drops after a certain point. Accordingly, there exists a *trade-off* between the codebook quantization loss and the learning performance. Under the current setting of 32 antennas at the TBS, the preferable transmission codebook size is 256. When the number of the CBS antennas grows from 64 to 128, the average SE increases, for the reason that a larger number of CBS antennas can provide more observations on the physical propagation environment. However, the average SE shows little improvement when the number of the CBS antennas increases from 128 to 256, indicating that a CBS with 128 antennas provides enough information for beam inference at the TBS under current settings. Fig. 7(b) shows the average SE with different number of TBS antennas. It is observed that the performances of both the optimal beamforming and the proposed method increase with the number of TBS antennas, and the aforementioned trade-off still exists when the number of TBS antennas changes. In addition, the performance gap between the optimal beamforming and the proposed scheme grows with the number of TBS antennas, resulting from the fact that narrower beams lead to a larger loss of beamforming gain when not aligned perfectly.

Fig. 8 shows the performance of the remote beamforming inference framework with the proposed enhancements. It can be observed that by jointly learning from two CBSs, the performance of the proposed scheme can be improved from a loss of 10.05% to 7.21%, resulting from a better characterization of the physical environments providing the view of two CBSs. Fig. 8(b) shows that combining our proposed framework with conventional beam sweeping can significantly improve the system performance. For instance, by sweeping only two candidate beams, an accuracy of 98.67% for the receive beamforming can be achieved. Sweeping two candidate beams at both the receiver and the transmitter sides can reduce the normalized beamforming loss from 10.05% to 4.60%. Accordingly, the performance of the proposed scheme



(a)



(b)

Fig. 9. (a) Sum rate versus SNR with 60 users. (b) Sum rate versus number of users with an SNR of 15 dB.

can be easily improved at a cost of very small extra overhead (2 candidate beams are sufficient); the overhead is much smaller compared with conventional beam sweeping methods.

B. Simulation Results for APS inference and User Grouping

In the multi-user scenario, users are grouped based on the APS, which is inferred by an offline-fitted CNN. Here the APS of each user is an 1024-dimension vector in order to provide high angular resolution. Fig. 9 shows the performance of the proposed user grouping schemes, together with the proposed APS inference schemes. Two baseline algorithms are considered here. One serves all users simultaneously, while the other allocates orthogonal resources for different users. A DFT-based analog beamformer is used for all the schemes, based on either the real APS or the inferred APS. Since the TBS only has 32 antennas, the beamforming vectors are the first 32 dimensions of the DFT vectors with a size of 1024. The minimal threshold η for SINR is set at 0.2. As shown in Fig. 9(a), the proposed user grouping method outperforms the method that serves all the users simultaneously. Besides, the proposed algorithm has a significant performance improvement over allocating orthogonal resources for different users in a high SNR regime, benefiting from the spatial multiplexing gain. For instance, the constraint sum rate of the proposed approach is 1.8 times larger than that from serving all 60 users separately at an SNR of 20 dB. Fig. 9(b) shows the constrained sum rate versus the number of users, when the SNR is set at 15 dB. It can be seen that the proposed scheme can guarantee a stable performance when serving a large number of users. In general, the proposed graph coloring based user grouping can achieve an effective trade-off between the beamforming gain and the spatial multiplexing gain.

V. Conclusions

In this paper, the dependence among remote channels was shown by the calculation of the mutual information and CRLB. Based on this finding, the idea of remote channel inference framework was proposed, and beamforming for small cells in hyper-cellular networks was studied as a use case. Different NN architectures were exploited to infer the beamforming directions or the APS of target BSs based on the CSI of source BSs. The simulation results on ray-tracing channel data show significant performance gain over conventional schemes.

An interesting question to ask is whether instantaneous CSI can be inferred, since the current approach can only infer CSI features such as AoAs and receive power. Closed-form expressions for the remote CSI dependence in the general case would definitely be desirable, although obtaining such expressions seems a daunting task given the complicated wireless propagation environment. In addition, the generalization to wideband CSI could be a promising path, because wideband signals give us great resolution in the delay domain, and thus contain much more information that can be taken advantage of.

Appendix A

Based on the first assumption in Section II-B, the phase is random and hence the received signal in (11) can be viewed as a zero-mean circular complex Gaussian process (assuming a large number of scatterers), whose probability distribution function (pdf) is completely characterized by its covariance matrix (sufficient statistics)

$$\boldsymbol{C}_y = \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}] = P_{\text{tx}}\mathbb{E}\left[\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\right] + \boldsymbol{C}_n, \tag{30}$$

where $\boldsymbol{C}_n$ is the noise covariance, and its estimation as

$$\hat{\boldsymbol{C}}_y = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{y}(k)\boldsymbol{y}^{\mathsf{H}}(k), \tag{31}$$

where K is the number of sampling points, $\boldsymbol{y}(k)$ denotes the k-th sample, and furthermore

$$\begin{aligned}\left\{\mathbb{E}\left[\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\right]\right\}_{ml} &= \mathbb{E}\left[h_m h_l^{\mathsf{H}}\right] \\ &= \int_{-\gamma_{\max}}^{\gamma_{\max}} S(\gamma)\exp\left(-\frac{j2\pi\delta m}{\lambda}\cos(\gamma+\theta)+j\phi_\gamma\right)\mathrm{d}\gamma \\ &\quad\times\int_{-\gamma_{\max}}^{\gamma_{\max}} S(\gamma)\exp\left(\frac{j2\pi\delta l}{\lambda}\cos(\gamma+\theta)-j\phi_\gamma\right)\mathrm{d}\gamma\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{=} \int_{-\gamma_{\max}}^{\gamma_{\max}} S^2(\gamma)\exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos(\gamma+\theta)\right)\mathrm{d}\gamma\\
&\overset{(b)}{=} \exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos\theta\right)\\
&\quad\times\int_{-\gamma_{\max}}^{\gamma_{\max}} S^2(\gamma)\exp\left(\frac{j2\pi\delta(m-l)}{\lambda}\sin\theta\gamma\right)\mathrm{d}\gamma\\
&\overset{(c)}{=} \exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos\theta\right)\\
&\quad\times\mathcal{F}^{-1}\left\{S^2(\gamma)\mathsf{rect}\left(\frac{\gamma}{2\gamma_{\max}}\right)\Big|_{f=\gamma}\right\}\Big|_{t=\frac{(m-l)\delta\sin\theta}{\lambda}}.
\end{aligned}\tag{32}$$

The equality of (a) is based on the fact that the arrival phases of MPCs are assumed i.i.d. and hence the cross terms in the integral are averaged out. The equality of (b) is based on the approximation that $r_{\max}$ is small, thereby $\sin\gamma\approx\gamma$ and $\cos\gamma\approx 1$. The equality of (c) is obtained by employing the inverse Fourier transform and $\mathsf{rect}(\cdot)$ denotes the rectangular function. The channel covariance matrix can be obtained with each entry given in (32). Thereby, we are ready to derive the CRLB of channel inference. Given the observations at the source BS $\boldsymbol{C}_{y_\mathsf{s}}$ in (30), the log-likelihood function can be written as

$$\mathcal{L}(\boldsymbol{z}) = -K\log|\boldsymbol{C}_{y_\mathsf{s}}| - K\mathsf{tr}\left[\boldsymbol{C}_{y_\mathsf{s}}^{-1}\hat{\boldsymbol{C}}_{y_\mathsf{s}}\right] + \mathsf{const.} \tag{33}$$

where $\boldsymbol{z} = \{S(\gamma), \gamma_{\max}, \theta_\mathsf{s}\}$. Its derivative can be calculated as

$$\mathrm{d}\mathcal{L}(\boldsymbol{z}) = -K\mathsf{tr}\left[\boldsymbol{C}_{y_\mathsf{s}}^{-1} - \boldsymbol{C}_{y_\mathsf{s}}^{-1}\hat{\boldsymbol{C}}_{y_\mathsf{s}}\boldsymbol{C}_{y_\mathsf{s}}^{-1}\right]\mathrm{d}\boldsymbol{C}_{y_\mathsf{s}}. \tag{34}$$

The FIM is given by

$$\begin{aligned}
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{ij} &= -\mathbb{E}\left[\frac{\partial^2\mathcal{L}(\boldsymbol{z})}{\partial\boldsymbol{z}_i\partial\boldsymbol{z}_j}\right]\\
&= K\mathbb{E}\left[\frac{\partial\left(\mathsf{tr}\left[\boldsymbol{I}_{M_\mathsf{s}} - \boldsymbol{C}_{y_\mathsf{s}}^{-1}\hat{\boldsymbol{C}}_{y_\mathsf{s}}\right]\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_i\right)}{\partial\boldsymbol{z}_j}\right]\\
&= K\mathsf{tr}\left[\boldsymbol{C}_{y_\mathsf{s}}^{-1}\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_\mathsf{s}}\right]\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_i\right]\\
&\quad + K\mathsf{tr}\left[\boldsymbol{I}_{M_\mathsf{s}} - \boldsymbol{C}_{y_\mathsf{s}}^{-1}\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_\mathsf{s}}\right]\right]\frac{\partial\left(\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_i\right)}{\partial\boldsymbol{z}_j}\\
&\overset{(a)}{=} K\mathsf{tr}\left[\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_i\right],
\end{aligned}\tag{35}$$

where we use $\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_\mathsf{s}}\right] = \boldsymbol{C}_{y_\mathsf{s}}$ in equality (a), and $\boldsymbol{D}_i = \frac{\partial\boldsymbol{C}_{y_\mathsf{s}}}{\partial z_i}$.

APPENDIX B

Considering the LoS case, $\gamma_{\max} = 0$, $S(\gamma) = \Delta(\gamma)$ ($\Delta(x) = 0, \forall x\neq 0$, and $\int_{-\infty}^{\infty}\Delta(x) = 1$), and hence

$$\boldsymbol{y}_{\mathsf{s},\mathsf{LOS}} = \rho_\mathsf{s}\exp(j\phi)\boldsymbol{e} + \boldsymbol{n}, \tag{36}$$

where $\rho_\mathsf{s} = \sqrt{P_\mathsf{tx}M_\mathsf{s}}\frac{\lambda}{4\pi D_\mathsf{s}}$, $\{\boldsymbol{e}\}_i = \frac{e^{\frac{-j2\pi i\delta}{\lambda}\cos\theta_\mathsf{s}}}{\sqrt{M_\mathsf{s}}}$. Due to the fact that there is only one LoS MPC, the Gaussianity of the array response is lost; specifically the first term on the right-hand side of the equation is deterministic and therefore we have

$$\boldsymbol{y}_{\mathsf{s},\mathsf{LOS}} \sim \mathcal{CN}(\rho_\mathsf{s}\exp(j\phi)\boldsymbol{e}, \sigma^2\boldsymbol{I}_{M_\mathsf{s}}). \tag{37}$$

A small modification to (14) is required to account for non-zero mean, which reads ($\boldsymbol{z} = [\rho_\mathsf{s}, \tau_\mathsf{s}, \phi]$)

$$\begin{aligned}
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{ij} = K\mathsf{tr}\left[\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_\mathsf{s}}^{-1}\boldsymbol{D}_i\right]\\
+\frac{2K}{\sigma^2}\left(\frac{\partial\boldsymbol{u}}{\partial z_j}\frac{\partial\boldsymbol{u}}{\partial z_i} + \frac{\partial\boldsymbol{v}}{\partial z_i}\frac{\partial\boldsymbol{v}}{\partial z_j}\right),
\end{aligned}\tag{38}$$

where in this case $\boldsymbol{m} = \rho_\mathsf{s}\exp(j\phi)\boldsymbol{e} \triangleq \boldsymbol{u} + j\boldsymbol{v}$, $\boldsymbol{C}_{y_\mathsf{s}} = \sigma^2\boldsymbol{I}_{M_\mathsf{s}}$ and therefore $\boldsymbol{D}_i = \boldsymbol{0}$, $\forall i$. Denote $\tau_\mathsf{s} \triangleq -\frac{2\pi\delta\cos\theta_\mathsf{s}}{\lambda}$, then

$$\begin{aligned}
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{11} &= \frac{2K}{\sigma^2}\boldsymbol{e}^\mathsf{H}\boldsymbol{e} = \frac{2K}{\sigma^2},\\
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{12} &= \{\boldsymbol{J}_{\boldsymbol{z}}\}_{21} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{13} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{31} = 0,\\
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{22} &= \frac{2K\rho_\mathsf{s}^2}{\sigma^2}\left(\frac{\partial\boldsymbol{e}}{\partial\tau_\mathsf{s}}\right)^\mathsf{H}\frac{\partial\boldsymbol{e}}{\partial\tau_\mathsf{s}} = \frac{K\rho_\mathsf{s}^2(M_\mathsf{s}-1)(2M_\mathsf{s}-1)}{3\sigma^2},\\
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{23} &= \{\boldsymbol{J}_{\boldsymbol{z}}\}_{32} = \frac{M_\mathsf{s}-1}{\sigma^2}K\rho_\mathsf{s}^2,\\
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{33} &= \frac{2\ K\rho_\mathsf{s}^2}{\sigma^2}.
\end{aligned}\tag{39}$$

The CRLBs can be readily derived as

$$\begin{aligned}
\mathsf{CRB}(\rho_\mathsf{s}) &= \left\{\boldsymbol{J}_{\boldsymbol{z}}^{-1}\right\}_{11} = \frac{\sigma^2}{2K},\\
\mathsf{CRB}(\tau_\mathsf{s}) &= \left\{\boldsymbol{J}_{\boldsymbol{z}}^{-1}\right\}_{22} = \frac{6\sigma^2}{K\rho_\mathsf{s}^2(M_\mathsf{s}^2-1)M_\mathsf{s}},
\end{aligned}\tag{40}$$

respectively. Similar with (18), we can then obtain the CRLBs of D_s and θ_s as:

$$\begin{aligned}
\mathsf{CRB}(D_\mathsf{s}) &= \frac{8\pi^2\ D_\mathsf{s}^4\sigma^2}{\lambda^2\ KP_\mathsf{tx}M_\mathsf{s}},\\
\mathsf{CRB}(\theta_\mathsf{s}) &= \frac{24\sigma^2D_\mathsf{s}^2}{KM_\mathsf{s}(M_\mathsf{s}^2-1)P_\mathsf{tx}\delta^2\sin^2\theta_\mathsf{s}}.
\end{aligned}\tag{41}$$

Denote the effective receive signal-to-noise ratio (SNR) as $\mathsf{SNR} = \frac{P_\mathsf{tx}\left(\frac{\lambda}{4\pi D_\mathsf{s}}\right)^2}{\sigma^2}$, and $\delta = \lambda/2$, then

$$\begin{aligned}
\mathsf{CRB}_1(\theta_\mathsf{t}) = \frac{D_\mathsf{s}^2}{D_\mathsf{t}^4}\frac{1}{K\mathsf{SNR}}\Bigg(\underbrace{\frac{c_1\ D_0^2\sin^2(\theta_0-\theta_\mathsf{s})}{M_\mathsf{s}}}_{\mathcal{M}_1}\\
+\underbrace{\frac{c_2\left(D_\mathsf{s}-D_0\cos(\theta_0-\theta_\mathsf{s})\right)^2}{M_\mathsf{s}(M_\mathsf{s}^2-1)\sin^2\theta_\mathsf{s}}}_{\mathcal{M}_2}\Bigg),
\end{aligned}\tag{42}$$

where $c_1 = \frac{1}{2}$, $c_2 = \frac{6}{\pi^2}$. It is noted that

$$\mathcal{M}_1 \sim \frac{1}{M_\mathsf{s}}, \quad \text{and } \mathcal{M}_2 \sim \frac{1}{M_\mathsf{s}^3}, \tag{43}$$

and hence

$$\mathsf{CRB}_1(\theta_\mathsf{t}) \sim \frac{1}{M_\mathsf{s}}. \tag{44}$$

APPENDIX C

Inheriting the denotations in Appendix B, denote the location of the other source BS as, without loss of generality, $(D_\mathsf{s}', 0)$ with $D_\mathsf{s}' > 0$, and denote the AoA at the other source

BS as θ_{s}'. The AoA at the target BS, i.e., θ_{t}, can be expressed by

$$\theta_{\mathsf{t}} = \arctan\left(\frac{D_{\mathsf{s}}' \sin\theta_{\mathsf{s}}' \sin\theta_{\mathsf{s}} - D_0 \sin\theta_0 \sin\left(\theta_{\mathsf{s}}' - \theta_{\mathsf{s}}\right)}{D_{\mathsf{s}}' \sin\theta_{\mathsf{s}}' \cos\theta_{\mathsf{s}} - D_0 \cos\theta_0 \sin\left(\theta_{\mathsf{s}}' - \theta_{\mathsf{s}}\right)}\right). \tag{45}$$

The CRLB of θ_{t} with known CSI at two source BSs is

$$\mathsf{CRB}_2(\theta_{\mathsf{t}}) = \frac{6\ D_{\mathsf{s}}^2}{\pi^2 D_{\mathsf{t}}^4 M_{\mathsf{s}}(M_{\mathsf{s}}^2-1)} \frac{1}{K\mathsf{SNR}} \times \frac{\omega_1 + \omega_2}{\sin^2\left(\theta_{\mathsf{s}} - \theta_{\mathsf{s}}'\right)\sin^2\theta_{\mathsf{s}}}, \tag{46}$$

where

$$\begin{aligned} \omega_1 &= D_0^2 \sin^2\theta_{\mathsf{s}} \sin^2(\theta_0 - \theta_{\mathsf{s}}) \\ \omega_2 &= \sin^2\theta_{\mathsf{s}}' \left(D_{\mathsf{s}}' \sin\theta_{\mathsf{s}}' - D_0 \sin(\theta_{\mathsf{s}}' - \theta_0)\right)^2. \end{aligned} \tag{47}$$

It follows that

$$\mathsf{CRB}_2(\theta_{\mathsf{t}}) \sim \frac{1}{M_{\mathsf{s}}^3}. \tag{48}$$

REFERENCES

[1] S. Chen *et al.*, "Remote channel inference for beamforming in ultra-dense hyper-cellular network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[2] Z. Jiang, Z. He, S. Chen, A. F. Molisch, S. Zhou, and Z. Niu, "Inferring remote channel state information: Cramér-Rae lower bound and deep learning implementation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeter-wave communication: Potentials and approaches," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 66–73, May 2016.

[6] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. Shen, "Software defined space-air-ground integrated vehicular networks: Challenges and solutions," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 101–109, Jul. 2017.

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

[8] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[9] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[10] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X.-G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May 2018.

[11] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[12] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1–6, Oct. 2009.

[13] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[14] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[15] Z. Jiang, S. Zhou, R. Deng, Z. Niu, and S. Cao, "Pilot-data superposition for beam-based FDD massive MIMO downlinks," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1357–1360, Jun. 2017.

[16] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[17] X. Rao and V. K. N. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

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

[19] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu. (Sep. 2018). "Model-driven deep learning for physical layer communications." [Online]. Available: https://arxiv.org/abs/1809.06059

[20] T. Ding and A. Hirose, "Fading channel prediction based on combination of complex-valued neural networks and chirp Z-transform," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1686–1695, Sep. 2014.

[21] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[22] Y. Wang, M. Narasimha, and R. W. Heath, "MmWave beam prediction with situational awareness: A machine learning approach," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[23] J. Liu, R. Deng, S. Zhou, and Z. Niu, "Seeing the unobservable: Channel learning for wireless communication networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

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

[25] Remcom. *Wireless InSite*. Accessed: May 14, 2019. [Online]. Available: http://www.remcom.com/wireless-insite

[26] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.

[27] T. L. Fulghum, K. J. Molnar, and A. Duel-Hallen, "The Jakes fading model for antenna arrays incorporating azimuth spread," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 968–977, Sep. 2002.

[28] Z. Jiang, S. Chen, A. F. Molisch, R. Vannithamby, S. Zhou, and Z. Niu, "Exploiting wireless channel state information structures beyond linear correlations: A deep learning approach," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 28–34, Mar. 2019.

[29] S. Zhou, T. Zhao, Z. Niu, and S. Zhou, "Software-defined hyper-cellular architecture for green and elastic wireless access," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 12–19, Jan. 2016.

[30] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[31] G. Bartoli *et al.*, "Beamforming for small cell deployment in LTE-advanced and beyond," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 50–56, Apr. 2014.

[32] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

[33] P. Golik, P. Doetsch, and H. Ney, "Cross-entropy vs. squared error training: A theoretical and experimental comparison," in *Proc. Interspeech*, vol. 13, Aug. 2013, pp. 1756–1760.

[34] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.

[37] S. Chen, Z. Jiang, S. Zhou, and Z. Niu, "Time-sequence channel inference for beam alignment in vehicular networks," in *Proc. GlobalSIP*, Nov., 2018, pp. 1199–1203.

[38] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, vol. 30, Jun. 2013, p. 3.

[39] R. Deng, S. Zhou, and Z. Niu, "Scalable non-orthogonal pilot design for massive MIMO systems with massive connectivity," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.
[40] U. I. Gupta, D. T. Lee, and J. Y.-T. Leung, "Efficient algorithms for interval graphs and circular-arc graphs," *Network*, vol. 12, no. 4, pp. 459–467, 1982.
[41] P. Kela *et al.*, "Location based beamforming in 5G ultra-dense networks," in *Proc. Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–7.

Sheng Chen received the B.E. degree from the Electronic Engineering Department, Tsinghua University, Beijing, in 2016, where he is currently pursuing the Ph.D. degree with Niulab. His research interests include channel estimation in MIMO systems, millimeter wave beamforming, and machine learning applications in wireless communications.

Zhiyuan Jiang received the B.E. and Ph.D. degrees from the Electronic Engineering Department, Tsinghua University, in 2010 and 2015, respectively. He visited the University of Southern California from 2013 to 2014 and from 2017 to 2018. He was an Experienced Researcher and a Wireless Signal Processing Scientist with Ericsson and Intel Labs, from 2015 to 2016 and in 2018, respectively. He is currently an Associate Professor with the School of Communication and Information Engineering, Shanghai University. His main research interests include sequential decision making with applications in wireless networks and multi-antenna communication systems.

Sheng Zhou received the B.E. and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2005 and 2011, respectively. In 2010, he was a Visiting Student with the Wireless System Lab, Department of Electrical Engineering, Stanford University, Stanford, CA, USA. From 2014 to 2015, he was a Visiting Researcher with the Central Research Lab, Hitachi Ltd., Japan. He is currently an Associate Professor with the Department of Electronic Engineering, Tsinghua University. His research interests include cross-layer design for multiple antenna systems, mobile edge computing, vehicular networks, and green wireless communications. He received the IEEE ComSoc Asia–Pacific Board Outstanding Young Researcher Award in 2017.

Zhisheng Niu (M'98–SM'99–F'12) received the B.E. degree from Beijing Jiaotong University, China, in 1985, and the M.E. and D.E. degrees from the Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. From 1992 to 1994, he was with Fujitsu Laboratories Ltd., Japan. In 1994, he joined Tsinghua University, Beijing, China, where he is currently a Professor with the Department of Electronic Engineering. His major research interests include queuing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communication and networks. He received the Outstanding Young Researcher Award from the Natural Science Foundation of China in 2009 and the Best Paper Award from the IEEE Communication Society Asia–Pacific Board in 2013. He has served as a Chair for the Emerging Technologies Committee from 2014 to 2015, the Director for Conference Publications from 2010 to 2011, and the Director for Asia–Pacific Board in the IEEE Communication Society, from 2008 to 2009. He is currently serving as the Director for Online Contents from 2018 to 2019 and as an Area Editor for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He was also selected as a Distinguished Lecturer of the IEEE Communication Society from 2012 to 2015 and the IEEE Vehicular Technologies Society from 2014 to 2018. He is a fellow of IEICE.

Ziyan He received the B.S. degree in electronic engineering from Tsinghua University in 2018. He is currently pursuing the Ph.D. degree in electrical engineering with the Georgia Institute of Technology, Atlanta, GA, USA. His research interests include massive multiple-input multiple-output systems and machine learning for signal processing in communications.

Andrei Marinescu received the Ph.D. degree from the Trinity College Dublin in 2015, after researching distributed autonomic control solutions for dynamic environments as part of the Distributed Systems Group at the School of Computer Science and Statistics. His research work involves the application of machine learning techniques for optimization in the fields of wireless networks and smart grids. He has been a Research Fellow with CONNECT, Ireland's Research Centre for Future Networks and Communications, until 2019. He is currently a Lead Data Scientist at Eaton. His research interests include reinforcement learning, multi-agent systems, predictive analytics, dynamic fractional frequency reuse, and energy demand forecasting.

Luiz A. DaSilva (S'97–M'98–SM'00–F'16) was a tenured Professor with the Bradley Department of Electrical and Computer Engineering, Virginia Tech. He holds the Chair of Telecommunications at Trinity College Dublin, where he is currently the Director of CONNECT, the Science Foundation Ireland Research Centre for Future Networks and Communications. His research interests focus on distributed and adaptive resource management in wireless networks, and in particular radio resource sharing and the application of game theory to wireless networks. He is a Principal Investigator on research projects funded by the Science Foundation Ireland and the European Commission. He is a fellow of the Trinity College Dublin, for contributions to cognitive networks and to resource management in wireless networks.