Inferring Remote Channel State Information: Cramér-Rao Lower Bound and Deep Learning Implementation

Zhiyuan Jiang[1], Ziyan He[1], Sheng Chen[1], Andreas F. Molisch[2], *Fellow, IEEE*,
Sheng Zhou[1], Zhisheng Niu[1], *Fellow, IEEE*

[1]{zhiyuan@, hezy14@mails., chen-s16@mails., sheng.zhou@, niuzhs@}tsinghua.edu.cn, Tsinghua University, Beijing, China
[2]molisch@usc.edu, University of Southern California, Los Angeles, USA

Abstract—Channel state information (CSI) is of vital importance in wireless communication systems. Existing CSI acquisition methods usually rely on pilot transmissions, and geographically separated base stations (BSs) with non-correlated CSI need to be assigned with orthogonal pilots which occupy excessive system resources. Our previous work adopts a data-driven deep learning based approach which leverages the CSI at a local BS to infer the CSI remotely, however the relevance of CSI between separated BSs is not specified explicitly. In this paper, we exploit a model-based methodology to derive the Cramér-Rao lower bound (CRLB) of remote CSI inference given the local CSI. Although the model is simplified, the derived CRLB explicitly illustrates the relationship between the inference performance and several key system parameters, e.g., terminal distance and antenna array size. In particular, it shows that by leveraging multiple local BSs, the inference error exhibits a larger power-law decay rate (w.r.t. number of antennas), compared with a single local BS; this explains and validates our findings in evaluating the deep-neural-network-based (DNN-based) CSI inference. We further improve on the DNN-based method by employing dropout and deeper networks, and show an inference performance of approximately 90% accuracy in a realistic scenario with CSI generated by a ray-tracing simulator.

Index Terms—Channel state information, multiple-input multiple-output, deep neural network, Cramér-Rao lower bound

I. INTRODUCTION

Channel state information (CSI) plays a pivotal role in wireless communication systems, especially with the wide adoption of massive multiple-input multiple-output (MIMO) technology in the 5G and beyond systems. The knowledge of CSI can facilitate, and is necessary for, beamforming, spatial multiplexing, user scheduling and spatial diversity.

The acquisition of CSI usually relies on pilot transmissions, i.e., known signals transmitted only for probing the propagation channel. Broadly speaking, without considering pilots for particular usage such as demodulation and phase offset correction, the pilot-based CSI acquisition involves either uplink pilots utilizing channel reciprocity [1] or downlink pilots and uplink CSI feedback [2], both entailing a severe signalling overhead in the system design. Traditional methods for overhead-reduced CSI acquisition usually involves leveraging the CSI *linear correlations* in spatial, time and frequency

domains. The majority of the related works focus on the spatial domain correlation. By transforming the CSI to the angular domain where most of the angular bins carry negligible energy, the dimensionality of CSI is significantly reduced [2]–[6]. The time domain correlation is often modeled by the Gauss-Markov process, and work has been done considering the heterogeneity of channel coherence among users [7]. There is also recent work [8] proposing a more efficient pilot packing which is enabled by orthogonal time frequency space (OTFS) modulation. In OFDM systems, the CSI correlation that can be exploited in the frequency domain stems from the sparse multipath component (MPC) delay profile of the channel impulse response; such correlations can be exploited based on the compressive sensing framework [9].

In the spatial domain, the linear correlation among co-located antennas proves helpful in reducing the CSI overhead; however, the distance between two remote base stations (BSs) renders this simple structure powerless since channels of remote sites (with distance much larger than wavelength, cf. Fig. 1) are considered linearly independent. More sophisticated data structures in CSI, which are not obvious so far due to the complicated propagation environment, should be explored for CSI overhead reduction. Our previous work [10], [11] adopted a data-driven approach, particularly the deep neural network (DNN), to address this issue and showed promising results in inferring the CSI at a remote site based on observations at a local site. However, the work uses a completely model-free approach and hence no theoretical analysis was given.

In this paper, a deeper understanding of the CSI structure is desired, towards which end we first describe the general remote CSI inference problem and accordingly present a model-based Cramér-Rao lower bound (CRLB) analysis. Although the employment of channel models, e.g., one-ring model [12], [13] and line-of-sight (LoS) model in this paper, or perhaps any other existing channel models in the literature, cannot fully describe the real-world propagation environment, they are useful in obtaining insightful and closed-form results. Afterwards, for realistic implementation we improve upon our deep learning based approach and achieve good and robust (w.r.t. propagation channels) performance. The main
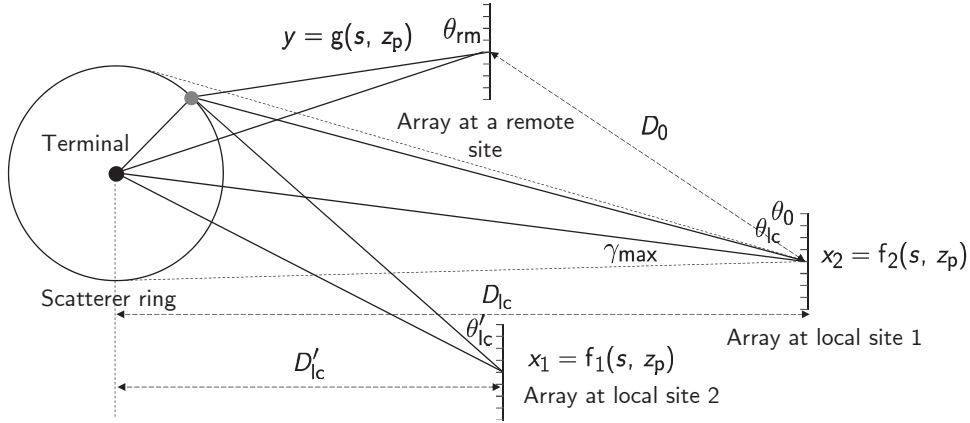
Fig. 1. A one-ring model based CSI inference scenario illustration. Two local sites with known CSI are described; the CSI at the remote site is unknown and to be inferred.

contributions include that we derive the closed-form CRLB of remote CSI inference when the channel is LoS; we find that the inference CRLB when considering a LoS and 2D scenario scales down as $1/M_{lc}$ where M_{lc} denotes the number of local BS antennas when CSI of one local BS is used, and $1/M_{lc}^3$ with CSI from more than one BSs; a novel DNN is proposed which exhibits improved performance compared with previous work, and more importantly, can be applied universally in real-world scenarios.

II. System Model and Problem Formulation

Consider a communication setup (see Fig. 1) where the terminal (user equipment, UE) is surrounded by local scatterers. Let x be the CSI between the terminal and a number of BSs (henceforth called "local sites"). Let there be another BS, henceforth called "remote site", with CSI y. The terminal has a single antenna element, while the BSs have antenna arrays of size M_{lc} for the local sites and M_{rm} for the remote site, respectively. The goal is to infer y from the knowledge of x. Such an inference problem from x to y might be ill-posed, in the sense that any of the following conditions is not satisfied:

- For any x, there exists a unique y that corresponds to the channel realization.
- The mapping from x to y is steady, i.e., given a small error of x, the corresponding error of y 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 z_p, consisting of, e.g., scatterer locations, reflection attenuation factors and etc. Then both the CSIs at the local or remote site can be expressed as a function of z_p and terminal location s based on the same methodology of ray-tracing models, i.e.,

$$x = f(s, z_p) \text{ and } y = g(s, z_p), \quad (1)$$

where $f(\cdot)$ and $g(\cdot)$ denotes the function mappings from the physical environment and terminal location to the CSIs at

the local sites and remote site, respectively. Based upon this, the inference problem is by and large solved if the physical environment and the terminal location are perfectly known. In particular, it has been shown [14], [15] that if a high-resolution estimation of the MPC parameters (angles, amplitudes) is available, then a prediction of the instantaneous channel impulse response can be made as long as the separation between local and remote sites is less than the stationarity region of the channel. However, in realistic situations we cannot invert the function mapping of $x = f(s, z_p)$ due to e.g., modelling errors, limited sampling points, insufficient observation accuracy and etc, rendering the whole inference problem ill-posed.

On the other hand, we argue that as M_{lc} increases, the inverse problem of $x = f(s, z_p)$ becomes less ill-conditioned. Furthermore, in the following sections we will show that with growing M_{lc}, the inference problem can be addressed with more and more accuracy. This phenomenon is referred to as *CSI manifestation*, as the propagation channel manifests from the CSI as the CSI spatial dimensionality grows. One of the main objectives of the paper is to obtain the scaling result of, e.g., the inference error, with M_{lc}.[1]

III. CRLB Analysis Based on One-Ring Channel Model

To gain insights into the channel inference problem and performance, as well as the CSI manifestation phenomenon, a theoretical analysis based on widely-adopted channel models is conducted in this section. The general model described in (1) 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

---

[1]In this paper, we assume the uniform linear antenna arrays at the BS sites and the system topology is two-dimensional.

terminal is at the center of the scatterer-ring. The received signal $\boldsymbol{y} = [y_1, ..., y_M]^\mathsf{T}$ at a site can be written by (denote the channel vector by $\boldsymbol{h} = [h_1, ..., h_M]^\mathsf{T}$)

$$y_i = \sqrt{P_{\text{tx}}} h_i + n_i,$$
$$h_i = \sum_{k=1}^{K} g_{ki} \exp\left(-\frac{j2\pi}{\lambda} d_{ki}\right), \tag{2}$$

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

$$d_{ki} = \xi_{\text{t}k} + \xi_{ki} \tag{3}$$

with $\xi_{\text{t}k}$ denoting the distance from the terminal to the k-th scatterer and ξ_{ki} the distance from the k-th scatterer to the i-th receive antenna. Assuming that the angle spread at the terminal is relatively small and the terminal is in the far-field, i.e., $D \gg r_{\text{max}}$ and $D \gg \frac{2M_{\text{lc}}^2 \delta^2}{\lambda}$ where δ is antenna spacing [16], the channel gains of all MPCs are therefore approximately identical, i.e., $g_{ki} = g, \forall 1 \le k \le K, 1 \le i \le M$, and

$$d_{ki} \approx \xi_k + i\delta \cos \gamma_k. \tag{4}$$

Denote by ξ_k the distance from the terminal 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 angle-of-arrival (AoA) of the k-th MPC. Based on this approximation, (2) can be re-written as

$$y_i = \sum_{k=1}^{K} g\sqrt{P_{\text{tx}}} \exp\left(-\frac{j2\pi\delta i}{\lambda}\cos\gamma_k + j\phi_k\right) + n_i, \tag{5}$$

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

The angular power spectrum (APS) seen at a site is defined as

$$S(\gamma) = g\mu(\gamma)p(\gamma), \tag{6}$$

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

$$\int_{-\gamma_{\text{max}}}^{\gamma_{\text{max}}} S^2(\gamma)\mathsf{d}\gamma = g^2. \tag{7}$$

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 [17]

$$\mu_{\text{ring}}(\gamma) = \frac{2}{\sqrt{\gamma_{\text{max}}^2 - (\gamma - \theta)^2}}, \tag{8}$$

where $\theta - \gamma_{\text{max}} \le \gamma \le \theta + \gamma_{\text{max}}$ and θ is the mean AoA of the terminal 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

$$y_i = \int_{-\gamma_{\text{max}}}^{\gamma_{\text{max}}} \sqrt{P_{\text{tx}}} S(\gamma)$$
$$\times \exp\left(-\frac{j2\pi\delta i}{\lambda}\cos(\gamma + \theta) + j\phi_\gamma\right)\mathsf{d}\gamma + n_i. \tag{9}$$

Combing with (1), we denote

$$h_{\text{lc},i} = \mathsf{f}\left(\theta_{\text{lc}}, S_{\text{lc}}(\gamma), \phi_{\text{lc},\gamma}, r_{\text{max,lc}}\right),$$
$$h_{\text{rm},i} = \mathsf{g}\left(\theta_{\text{rm}}, S_{\text{rm}}(\gamma), \phi_{\text{rm},\gamma}, r_{\text{max,rm}}\right), \tag{10}$$

where the subscripts $(\cdot)_{\text{lc}}$ and $(\cdot)_{\text{rm}}$ denote the local site and the remote site, respectively. The function mapping of $\mathsf{f}(\cdot)$ and $\mathsf{g}(\cdot)$ are substantiated by (9). Note that the model of (10) allows, e.g., different angular spreads, different visibility of scatterers, different APSs at the local site and the remote site, 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 $\boldsymbol{h}_{\text{rm}}$, given $\boldsymbol{h}_{\text{lc}}$, s.t., Eq. (10) is satisfied. (11)

A. Parameter Extraction and Inference: CRLB Analysis

It is observed that the implicit channel inference problem in (1) is transformed to 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 cannot based on realistic rationality. 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 the fact that the phase of an electromagnetic wave shifts dramatically even with a slight movement of the terminal (several wavelengths).
- The observable scatterers seen at the remote site, i.e., $S_{\text{rm}}(\gamma)$, $r_{\text{max,rm}}$, cannot be inferred based on the observation at local site which provides little information about whether an MPC is obstructed seen at the remote site. Instead, this information can be obtained by using a relatively infrequent probing signals by the remote site given the fact that the scattering environment is constant inside the stationarity region (typical size of tens of meters in urban areas).[2]

[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 a good CSI inference performance.

Based on these assumptions, we focus on analyzing the inference performance with respect to the AoA at the remote site θ_{rm} to obtain theoretical results. The end goal is to derive the CRLB of the estimation of θ_{rm}, towards which we first solve the inverse problem of $h_{\mathsf{lc},i} = \mathsf{f}\left(\theta_{\mathsf{lc}}, S_{\mathsf{lc}}(\gamma), \phi_{\mathsf{lc},\gamma}, r_{\mathsf{max,lc}}\right)$, and then relates to the AoA of θ_{rm} based on geometry. Based on the first assumption, the phase is random and hence the received signal in (9) 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)

$$\boldsymbol{C}_y = \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}] = P_{\mathsf{tx}}\mathbb{E}\left[\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\right] + \boldsymbol{C}_n, \tag{12}$$

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

$$\hat{\boldsymbol{C}}_y = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{y}(k)\boldsymbol{y}^{\mathsf{H}}(k), \tag{13}$$

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

$$
\begin{aligned}
&\left\{\mathbb{E}\left[\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\right]\right\}_{ml} = \mathbb{E}\left[h_m h_l^{\mathsf{H}}\right]\\
&= \int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} S(\gamma)\exp\left(-\frac{j2\pi\delta m}{\lambda}\cos(\gamma+\theta)+j\phi_{\gamma}\right)\mathsf{d}\gamma\\
&\quad\times\int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} S(\gamma)\exp\left(\frac{j2\pi\delta l}{\lambda}\cos(\gamma+\theta)-j\phi_{\gamma}\right)\mathsf{d}\gamma\\
&\overset{(a)}{=} \int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} S^2(\gamma)\exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos(\gamma+\theta)\right)\mathsf{d}\gamma\\
&\overset{(b)}{=} \exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos\theta\right)\\
&\quad\times\int_{-\gamma_{\mathsf{max}}}^{\gamma_{\mathsf{max}}} S^2(\gamma)\exp\left(\frac{j2\pi\delta(m-l)}{\lambda}\sin\theta\gamma\right)\mathsf{d}\gamma\\
&\overset{(c)}{=} \exp\left(-\frac{j2\pi\delta(m-l)}{\lambda}\cos\theta\right)\\
&\quad\times\mathcal{F}^{-1}\left\{\left.S^2(\gamma)\mathsf{rect}\left(\frac{\gamma}{2\gamma_{\mathsf{max}}}\right)\right|_{f=\gamma}\right\}\Bigg|_{t=\frac{(m-l)\delta\sin\theta}{\lambda}}
\end{aligned}
\tag{14}
$$

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 and hence

$$\sin\gamma \approx \gamma \quad \text{and} \quad \cos\gamma \approx 1. \tag{15}$$

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 (14). Thereby, we are ready to derive the CRLB of channel inference. Given the observations at the local site $\boldsymbol{C}_{y_{\mathsf{lc}}}$ in (12), the log-likelihood function can be written as

$$\mathcal{L}(\boldsymbol{z}) = -K\log|\boldsymbol{C}_{y_{\mathsf{lc}}}| - K\mathsf{tr}\left[\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\right] + \mathsf{const.} \tag{16}$$

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

$$\mathsf{d}\mathcal{L}(\boldsymbol{z}) = -K\mathsf{tr}\left[\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1} - \boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\right]\mathsf{d}\boldsymbol{C}_{y_{\mathsf{lc}}}. \tag{17}$$

The Fisher information matrix FIM is given by

$$
\begin{aligned}
\{\boldsymbol{J}_{\boldsymbol{z}}\}_{ij} &= -\mathbb{E}\left[\frac{\partial^2\mathcal{L}(\boldsymbol{z})}{\partial z_i\partial z_j}\right]\\
&= K\mathbb{E}\left[\frac{\partial\left(\mathsf{tr}\left[\boldsymbol{I}_{M_{\mathsf{lc}}} - \boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\right]\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_i\right)}{\partial z_j}\right]\\
&= K\mathsf{tr}\left[\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\right]\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_i\right]\\
&\quad + K\mathsf{tr}\left[\boldsymbol{I}_{M_{\mathsf{lc}}} - \boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\right]\right]\frac{\partial\left(\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_i\right)}{\partial z_j}\\
&\overset{(a)}{=} K\mathsf{tr}\left[\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_j\boldsymbol{C}_{y_{\mathsf{lc}}}^{-1}\boldsymbol{D}_i\right],
\end{aligned}
\tag{18}
$$

where we use $\mathbb{E}\left[\hat{\boldsymbol{C}}_{y_{\mathsf{lc}}}\right] = \boldsymbol{C}_{y_{\mathsf{lc}}}$ in equality (a), and $\boldsymbol{D}_i = \frac{\partial\boldsymbol{C}_{y_{\mathsf{lc}}}}{\partial z_i}$. Based on (18), the CRLB of AoA (θ_{lc}) and terminal distance D_{lc} can be obtained by extracting the diagonal entries of the inverse FIM.

The CRLB of θ_{rm} can be therefore obtained in the following. Without loss of generality, assume the coordinates of the local site and remote site 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_{lc} and D_{lc}, the coordinate of the terminal can be expressed in two ways as follows:

$$x_{\mathsf{t}} = D_{\mathsf{lc}}\cos\theta_{\mathsf{lc}} = D_0\cos\theta_0 + D_{\mathsf{rm}}\cos\theta_{\mathsf{rm}}, \tag{19}$$

$$y_{\mathsf{t}} = D_{\mathsf{lc}}\sin\theta_{\mathsf{lc}} = D_0\sin\theta_0 + D_{\mathsf{rm}}\sin\theta_{\mathsf{rm}}. \tag{20}$$

The AoA at the remote site can be hence represented by D_{lc} and D_{rm} as follows

$$\theta_{\mathsf{rm}} = \arctan\left(\frac{D_{\mathsf{lc}}\sin\theta_{\mathsf{lc}} - D_0\sin\theta_0}{D_{\mathsf{lc}}\cos\theta_{\mathsf{lc}} - D_0\cos\theta_0}\right). \tag{21}$$

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

$$
\begin{aligned}
\mathsf{CRB}(\theta_{\mathsf{rm}}) &= \left(\frac{\partial\theta_{\mathsf{rm}}}{\partial D_{\mathsf{lc}}}\right)^2\mathsf{CRB}(D_{\mathsf{lc}}) + \left(\frac{\partial\theta_{\mathsf{rm}}}{\partial\theta_{\mathsf{lc}}}\right)^2\mathsf{CRB}(\theta_{\mathsf{lc}})\\
&= \frac{D_0^2\sin^2(\theta_0 - \theta_{\mathsf{lc}})}{D_{\mathsf{rm}}^4}\mathsf{CRB}(D_{\mathsf{lc}})\\
&\quad + \frac{D_{\mathsf{lc}}^2\left(D_{\mathsf{lc}} - D_0\cos(\theta_0 - \theta_{\mathsf{lc}})\right)^2}{D_{\mathsf{rm}}^4}\mathsf{CRB}(\theta_{\mathsf{lc}})
\end{aligned}
\tag{22}
$$

where $\mathsf{CRB}(D_{\mathsf{lc}})$ and $\mathsf{CRB}(\theta_{\mathsf{lc}})$ are obtained from (18).

The calculation of $\boldsymbol{D}_i$ in (18) 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 $\gamma_{\mathsf{max,lc}}$ and $\gamma_{\mathsf{max,rm}}$ are small and approach zero, i.e., a LoS MPC only.

B. Special Case: LoS Scenario and One Local Site

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

$$\boldsymbol{y}_{y_{\mathsf{lc}},\mathsf{LOS}} = \rho_{\mathsf{lc}}\exp(j\phi)\boldsymbol{e} + \boldsymbol{n}, \tag{23}$$

where $\rho_{\mathsf{lc}} = \sqrt{P_{\mathsf{tx}}M_{\mathsf{lc}}}\frac{\lambda}{4\pi D_{\mathsf{lc}}}$, $\{\boldsymbol{e}\}_i = \frac{e^{\frac{-j2\pi i\delta}{\lambda}\cos\theta_{\mathsf{lc}}}}{\sqrt{M_{\mathsf{lc}}}}$. 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

$$\boldsymbol{y}_{y_{\text{lc}},\text{LOS}} \sim \mathcal{CN}(\rho_{\text{lc}} \exp (j\phi)\boldsymbol{e}, \sigma^2 \boldsymbol{I}_{M_{\text{lc}}}). \tag{24}$$

A small modification to (18) is required to account for non-zero mean, which reads ($\boldsymbol{z} = [\rho_{\text{lc}}, \tau_{\text{lc}}, \phi]$)

$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{ij} = K\text{tr}\left[\boldsymbol{C}_{y_{\text{lc}}}^{-1} \boldsymbol{D}_j \boldsymbol{C}_{y_{\text{lc}}}^{-1} \boldsymbol{D}_i\right]$$
$$+ \frac{2K}{\sigma^2}\left(\frac{\partial \boldsymbol{u}}{\partial z_j}\frac{\partial \boldsymbol{u}}{\partial z_i} + \frac{\partial \boldsymbol{v}}{\partial z_i}\frac{\partial \boldsymbol{v}}{\partial z_j}\right), \tag{25}$$

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

$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{11} = \frac{2K}{\sigma^2}\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e} = \frac{2K}{\sigma^2},$$
$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{12} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{21} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{13} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{31} = 0,$$
$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{22} = \frac{2K\rho_{\text{lc}}^2}{\sigma^2}\left(\frac{\partial \boldsymbol{e}}{\partial \tau_{\text{lc}}}\right)^{\mathsf{H}}\frac{\partial \boldsymbol{e}}{\partial \tau_{\text{lc}}} = \frac{K\rho_{\text{lc}}^2(M_{\text{lc}} - 1)(2M_{\text{lc}} - 1)}{3\sigma^2},$$
$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{23} = \{\boldsymbol{J}_{\boldsymbol{z}}\}_{32} = \frac{M_{\text{lc}} - 1}{\sigma^2}K\rho_{\text{lc}}^2,$$
$$\{\boldsymbol{J}_{\boldsymbol{z}}\}_{33} = \frac{2K\rho_{\text{lc}}^2}{\sigma^2}. \tag{26}$$

The CRLBs can be readily derived as

$$\text{CRB}(\rho_{\text{lc}}) = \{\boldsymbol{J}_{\boldsymbol{z}}^{-1}\}_{11} = \frac{\sigma^2}{2K},$$
$$\text{CRB}(\tau_{\text{lc}}) = \{\boldsymbol{J}_{\boldsymbol{z}}^{-1}\}_{22} = \frac{6\sigma^2}{K\rho_{\text{lc}}^2(M_{\text{lc}}^2 - 1)M_{\text{lc}}}, \tag{27}$$

respectively. Similar with (22), we can then obtain the CRLBs of D_{lc} and θ_{lc} as:

$$\text{CRB}(D_{\text{lc}}) = \frac{8\pi^2 D_{\text{lc}}^4 \sigma^2}{\lambda^2 K P_{\text{tx}} M_{\text{lc}}},$$
$$\text{CRB}(\theta_{\text{lc}}) = \frac{24\sigma^2 D_{\text{lc}}^2}{KM_{\text{lc}}(M_{\text{lc}}^2 - 1)P_{\text{tx}}\delta^2\sin^2\theta_{\text{lc}}}. \tag{28}$$

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

$$\text{CRB}_1(\theta_{\text{rm}}) = \frac{D_{\text{lc}}^2}{D_{\text{rm}}^4}\frac{1}{K\text{SNR}}\left(\underbrace{\frac{c_1 D_0^2 \sin^2(\theta_0 - \theta_{\text{lc}})}{M_{\text{lc}}}}_{\mathcal{M}_1}\right.$$
$$\left.+ \underbrace{\frac{c_2(D_{\text{lc}} - D_0\cos(\theta_0 - \theta_{\text{lc}}))^2}{M_{\text{lc}}(M_{\text{lc}}^2 - 1)\sin^2\theta_{\text{lc}}}}_{\mathcal{M}_2}\right), \tag{29}$$

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

$$\mathcal{M}_1 \sim \frac{1}{M_{\text{lc}}}, \text{ and } \mathcal{M}_2 \sim \frac{1}{M_{\text{lc}}^3}, \tag{30}$$

and hence

$$\text{CRB}_1(\theta_{\text{rm}}) \sim \frac{1}{M_{\text{lc}}}. \tag{31}$$

Therefore the inference performance bottleneck is at $\mathcal{M}_1$, i.e., the inference error related to distance estimation which is reciprocal with M_{lc}. On the other hand, the inference error associated with AoA estimation scales inversely with M_{lc}^3. Based on this observation, the performance can be improved by leveraging CSI at multiple sites and correspondingly multiple AoAs to make the inference. We present the following CRLB analysis which accounts for two separate local sites (with known CSI) to infer the CSI at a remote cite.

C. Special Case: LoS Scenario with Known CSI at Two Sites

In this subsection, we will show that by using the AoAs at two geographically separated sites, the inference error can be significantly reduced; such a scenario presents itself in densely deployed cellular systems and, moreover, the cost of acquiring CSI at one other site is affordable.

Inheriting the denotations in the last subsection, denote the location of the other local site as, without loss of generality, $(D'_{\text{lc}}, 0)$ with $D'_{\text{lc}} > 0$, and denote the AoA at the other local site as θ'_{lc}. The AoA at the remote site, i.e., θ_{rm}, can be expressed by

$$\theta_{\text{rm}} = \arctan\left(\frac{D'_{\text{lc}}\sin\theta'_{\text{lc}}\sin\theta_{\text{lc}} - D_0\sin\theta_0\sin(\theta'_{\text{lc}} - \theta_{\text{lc}})}{D'_{\text{lc}}\sin\theta'_{\text{lc}}\cos\theta_{\text{lc}} - D_0\cos\theta_0\sin(\theta'_{\text{lc}} - \theta_{\text{lc}})}\right). \tag{32}$$

The CRLB of θ_{rm} with known CSI at two local sites is

$$\text{CRB}_2(\theta_{\text{rm}}) = \frac{6D_{\text{lc}}^2}{\pi^2 D_{\text{rm}}^4 M_{\text{lc}}(M_{\text{lc}}^2 - 1)}\frac{1}{K\text{SNR}}$$
$$\times \frac{\omega_1 + \omega_2}{\sin^2(\theta_{\text{lc}} - \theta'_{\text{lc}})\sin^2\theta_{\text{lc}}}, \tag{33}$$

where

$$\omega_1 = D_0^2 \sin^2\theta_{\text{lc}}\sin^2(\theta_0 - \theta_{\text{lc}})$$
$$\omega_2 = \sin^2\theta'_{\text{lc}}\left(D'_{\text{lc}}\sin\theta'_{\text{lc}} - D_0\sin(\theta'_{\text{lc}} - \theta_0)\right)^2. \tag{34}$$

It follows that

$$\text{CRB}_2(\theta_{\text{rm}}) \sim \frac{1}{M_{\text{lc}}^3}. \tag{35}$$

Remark 1: While this paper focuses on inference based solely on spatial domain signals, one direction that is certainly worth studying is the CSI inference performance incorporating wideband signals such that the time-of-arrival (ToA) can also be estimated. □

IV. CSI INFERENCE BASED ON DNN

CSI inference in real-world signal propagation scenarios is extremely difficult and intractable, and moreover existing model-based approaches are inadequate to address this issue. Therefore, inspired by the recent advance in deep learning field, we develop a DNN to accomplish this task. A DNN with 4 hidden layers is adopted. The sizes of layers are M_{lc}, M_{lc}, 64 and 32, respectively. The non-linear function for the hidden layers is LeakyReLU, while the non-linear function for the output layer is sigmoid. The input features are quantized CSI in the angular domain (take modulus and logarithm) which are normalized by Z-score normalization method while the output
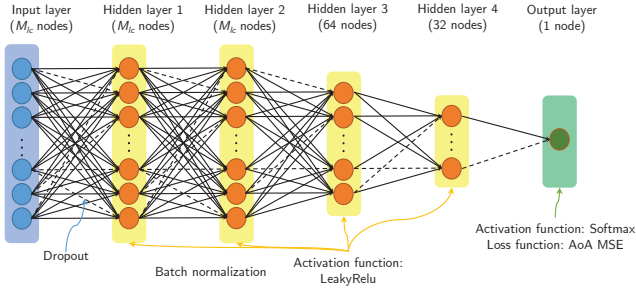
Fig. 2. The proposed DNN architecture.

is normalized AOA (AOA divided by π). Here the prescribed codebook is the discrete-Fourier-transform (DFT) codebook, and the size of the codebook is M_{lc}.

In order to avoid overfitting, we apply dropout [18] in our neural network and the keep probability is set to be 0.7. The cost function of the neural network is mean-squared-error (MSE) of AOA and the MSE is also our performance metric. The Adam gradient-based optimizer is used with step size of 10^{-4}. We randomly divide data set into two parts for training (90%) and test (10%). The trained model runs for 10 times and the average is taken.

V. SIMULATION RESULTS

In Fig. 3 and 4, we present simulation results based on the one-ring channel model. The two local BSs (LBSs) are located at $(-100, 0)$ and $(100, 0)$ (the distance unit is meter), the remote BS (RBS) whose CSI is to be inferred is at $(0, 50)$ and the terminal is randomly located on a semi-disk (to avoid AoA phase ambiguity of ULA) centered at the RBS with a radius of 50 m (not allowed to be within 5 m to the RBS). The DNN is as described in Section IV. The results, which are plotted on logarithmic scale, indicate that the scaling law derived for the LoS scenario, i.e., the CRLB scales with $1/M_{lc}$ and $1/M_{lc}^3$ with one LBS and two LBSs respectively, is upheld for the one-ring channel model. More importantly, it is found that the inference performance by DNN also exhibits such behaviour, while the exact scaling factors may vary due to the unclear noise structure of the black-box of DNN. We should emphasize that it may be unfair to compare the performance of DNN and CRLB since DNN is applied to a much more general scenario while the CRLB is concerned with the optimal channel estimator in this specific channel model. However, the insights given by analyzing the CRLB are still useful as it is shown that the qualitative results are consistent between the DNN performance and analytical CRLB.

The DNN-based approach is tested in a more realistic scenario in Fig. 5. We use the Wireless Incite® ray-tracing channel simulator to generate the CSI. The system layout is depicted in Fig. 5(a). The antenna arrays are two-dimensional with sizes of 4×100 and 4×20 for LBS and RBS respectively; both arrays are located with heights of 20 m. The carrier frequency is 28 GHz and the system bandwidth is 10 MHz. 20000 samples are obtained for training and test. In this
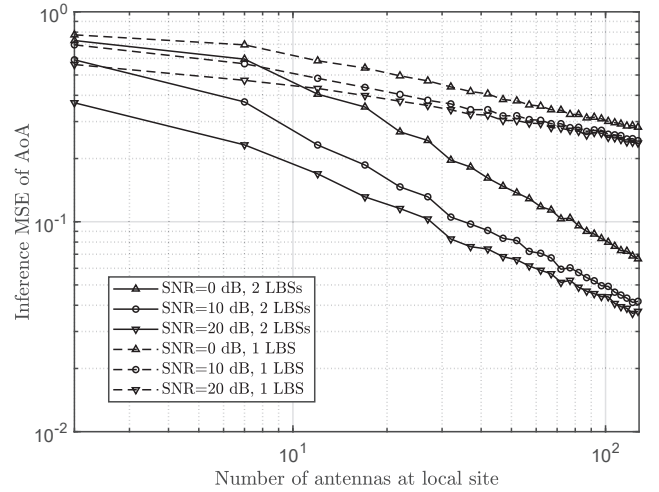

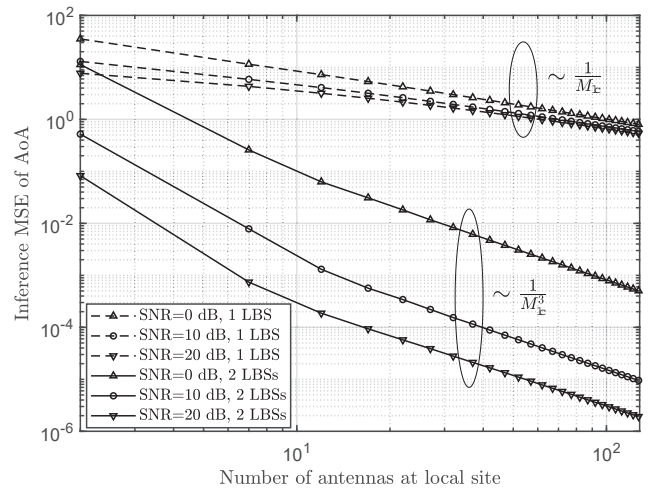Fig. 3. DNN-based CSI inference performance.

Fig. 4. CRLB of inference error for one-ring channel model.
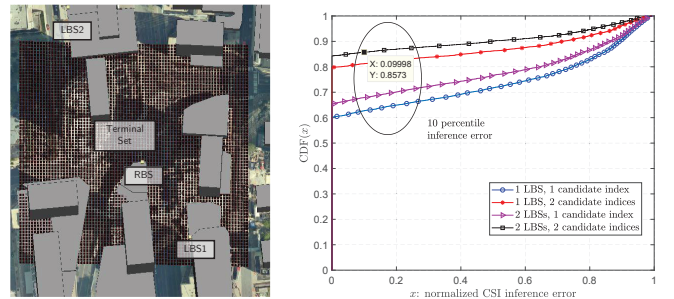

Fig. 5. DNN-based CSI inference performance with a realistic ray-tracing channel simulator.

case where the propagation channel may be non-LoS, the beamforming vector (DFT codebook) with the largest power, instead of the AoA which is not well-defined in this case, is the inference objective; therefore the output layer is modified to have the same number of neurons as the number of codewords in the DFT codebook, and the loss function is changed to the

cross-entropy function compared with the AoA MSE in Fig. 2. The performance metric is still the inference error normalized in the DFT codebook (to values within $[0, 1]$). Concretely, the normalized inference error is defined as

$$e \triangleq 1 - \frac{\hat{h}^{\dagger} h}{h_{\text{opt}}^{\dagger} h}, \tag{36}$$

where $\hat{h}$ is the inferred CSI from the DFT codebook, h_{opt} is the optimal DFT beamforming vector in the codebook and the CSI obtained by the ray-tracing simulator is denoted by h. It is observed from Fig. 5(b) that about 85% of the time, the optimal beamforming vector of RBS can be inferred with an error less than 10% based on observations at LBSs in a realistic scenario.

VI. CONCLUSIONS

The remote CSI inference problem is considered both theoretically and practically. To establish the relevance between CSI of LBS and RBS, we use the physical propagation environment as the bridge, however the inference problem is identified as ill-posed with insufficient observation capability. With proper simplifications by adopting the one-ring channel model, we transform the inference problem to a parameter extraction problem and derive the CRLB thereof. For a special LoS case, the closed-form expressions of CRLB can be obtained, showing that the CRLB scales inversely with M_{lc} and M_{lc}^3 by using one and two LBSs respectively. An improved DNN with dropout to thwart overfitting and a deeper network is proposed, showing robustly good performance in real-world scenarios. It is found by simulations that the theoretical CRLB analysis, although significantly simplified, provides insightful performance observations and meaningful implications.

REFERENCES

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

[2] Z. Jiang, A. 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.

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

[4] X. Rao and V. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[5] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propagat.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[6] H. Lin, F. Gao, S. Jin, and G. Y. Li, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Select. Areas Commun.*, vol. 35, no. 10, pp. 2268–2280, Oct 2017.

[7] R. Deng, Z. Jiang, S. Zhou, and Z. Niu, "How often should CSI be updated for massive MIMO systems with massive connectivity?" in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2017, pp. 1–6.

[8] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar 2017, pp. 1–6.

[9] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb 2016.

[10] S. Chen, Z. Jiang, J. Liu, R. Vannithamby, S. Zhou, Z. Niu, and Y. Wu, "Remote channel inference for beamforming in ultra-dense hyper-cellular network," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2017, pp. 1–6.

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

[12] W. C. Y. Lee, "Effects on correlation between two mobile radio base-station antennas," *IEEE Trans. Veh. Tech.*, vol. 22, no. 4, pp. 130–140, Nov 1973.

[13] D. shan Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[14] A. Molisch, "A generic model for MIMO wireless propagation channels in macro- and microcells," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 61–71, Jan. 2004.

[15] R. S. Thomä, D. Hampicke, M. Landmann, A. Richter, and G. Sommerkorn, "Measurement-based parametric channel modelling (MBPCM)," in *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2003, pp. 779–782.

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

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

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.