Deep Learning for Massive MIMO With 1-Bit ADCs: When More Antennas Need Fewer Pilots

Yu Zhang, Muhammad Alrabeiah, and Ahmed Alkhateeb

***Abstract*—This letter considers uplink massive MIMO systems with 1-bit analog-to-digital converters (ADCs) and develops a deep-learning based channel estimation framework. In this framework, the prior channel estimation observations and deep neural networks are leveraged to learn the non-trivial mapping from quantized received measurements to channels. For that, we derive the sufficient length and structure of the pilot sequence to guarantee the existence of this mapping function. This leads to the interesting, and *counter-intuitive*, observation that when more base-station antennas are employed, our proposed deep learning approach achieves better channel estimation performance, for the same pilot sequence length. Equivalently, for the same channel estimation performance, this means that when more antennas are employed, fewer pilots are required. This observation is also analytically proved for some special channel models. Simulation results confirm our observations and show that more antennas lead to better channel estimation in terms of the normalized mean squared error and the receive signal-to-noise ratio per antenna.**

***Index Terms*—Deep learning, massive MIMO, channel estimation, 1-bit ADCs.**

I. Introduction

USING low-resolution analog-to-digital converters (ADCs) in massive MIMO systems has the potential of reducing the power consumption while maintaining good achievable rate performance. These gains attracted increasing research interest in the last few years [1], [2]. One main challenge in these systems, however, is estimating the channels from highly quantized measurements. While several channel estimation techniques have been proposed in the literature [1], [2], these solutions generally require very large training overhead (long pilot sequences). This highly impacts the feasibility of employing low-resolution ADCs in practical millimeter wave (mmWave) and massive MIMO systems. This letter targets overcoming these challenges and enabling low-resolution ADCs in large-scale MIMO systems.

Contribution: In this letter, we propose a deep-learning based framework for the channel estimation problem in massive MIMO systems with 1-bit ADCs. In this framework, the prior channel estimation observations and deep neural networks are exploited to learn the mapping from the received quantized measurements to the channels. Learning this mapping, however, requires its existence in the first place. For that, we derive the sufficient length and structure of the pilot sequence that guarantees the existence of this quantized measurement to channel mapping. Then, we make the interesting observation that for the same set of candidate user locations, more antennas require fewer pilots to guarantee the mapping existence. This means that increasing the number of base station antennas reduces the number of required pilots to estimate their channels, which may seem counter-intuitive. The intuition justifying this observation, however, is that with more antennas, the quantized measurement vectors become more unique for the different channels. Hence, they can be efficiently mapped to their corresponding channels with less error probability. This observation is also proved analytically for the case of single-path channels. Simulation results highlight the promising gains of the proposed deep learning approach; they confirm that more antennas lead to better channel estimates, in terms of normalized mean-squared error (NMSE) and per-antenna signal-to-noise ratio (SNR).

Prior Work: To the best of our knowledge, no prior work has addressed deep learning based channel estimation for massive MIMO systems with 1-bit ADCs, or revealed the interesting observation that more antennas need fewer pilots. Relevant problems, however, have attracted a lot of research interest in the last few years [1]–[6]. In [1], [2], classical signal processing tools have been leveraged to design low-complexity near-maximum likelihood (ML) data detector [1] and develop efficient channel estimation techniques for massive MIMO and mmWave systems [2]. A common problem in these solutions is the need for very long pilot sequences to achieve good data detection or channel estimation performance.

In [3]–[6], machine learning techniques were developed to address several problems with 1-bit ADCs. For example, [3] developed a machine learning framework for data detection, but not for channel estimation, and [4] considered the channel estimation problem in OFDM systems, but only for systems with single antennas. In [5], deep learning solutions for data detection and channel equalization were developed for MIMO systems, but their approach works only for low-dimensional MIMO regimes as the proposed network does not converge for large numbers of antennas. In [6], systems with mixed full-resolution and low-resolution ADCs were adopted, but only the received signals from full-resolution ADCs were used to estimate the channels, which is in fact equivalent to the channel mapping concept we proposed in [7].

II. System and Channel Models

We consider the system shown in Fig. 1 where a massive MIMO base station (BS) with M antennas is communicating

Manuscript received March 23, 2020; accepted April 4, 2020. Date of publication April 14, 2020; date of current version August 7, 2020. This work was supported by the National Science Foundation under Grant 1923676. The associate editor coordinating the review of this article and approving it for publication was C.-K. Wen. *(Corresponding author: Ahmed Alkhateeb.)*

The authors are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: y.zhang@asu.edu; malrabei@asu.edu; alkhateeb@asu.edu).

Digital Object Identifier 10.1109/LWC.2020.2987893

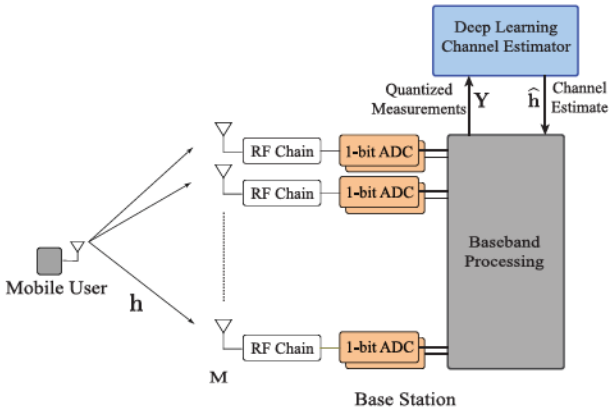

Fig. 1. The adopted massive MIMO system where the base station receiver uses 1-bit ADCs. The uplink quantized received measurement matrix $\mathbf{Y}$ is fed to a deep learning model that predicts the channel vector $\widehat{\mathbf{h}}$.

with a single-antenna user. The BS employs only 1-bit analog-to-digital converters (ADCs) in its receive chains. Further, we adopt a time division duplexing (TDD) system operation where the channel is estimated through an uplink training and used for downlink data transmission as summarized below.

Uplink Training: If the user transmits an uplink pilot sequence $\mathbf{x} \in \mathbb{C}^{N\times 1}$, where N is the length of pilot sequence, then the received signal at the BS after the ADC quantization can be expressed as

$$\mathbf{Y} = \text{sgn}(\mathbf{h}\mathbf{x}^{\text{T}} + \mathbf{N}), \tag{1}$$

where $\mathbf{h} \in \mathbb{C}^{M\times 1}$ is the channel vector between the mobile user and the BS antennas, $\mathbf{N}$ is the receive noise matrix at the BS with independent and identically distributed (i.i.d.) elements drawn from $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, and the transmitted pilot sequence is satisfying $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = P_t\mathbf{I}$ with P_t denoting the average transmit power per symbol. The element-wise operator $\text{sgn}(\cdot)$ is the signum function, and it is applied separately to the real and imaginary part of its argument. Finally, $\mathbf{Y}$ is the $M \times N$ quantized receive measurement matrix that consists of the received pilot signals after quantization.

Channel Model: We adopt a general geometric channel model for $\mathbf{h}$. Assume that the signal propagation between the user and BS consists of L paths. Each path ℓ has a complex gain α_ℓ and an angle of arrival ϕ_ℓ, then we define

$$\mathbf{h} = \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}(\phi_\ell), \tag{2}$$

where $\mathbf{a}(\phi_\ell)$ is the array response vector of the BS.

Channel Estimation: The quantized receive measurement matrix $\mathbf{Y}$ will be processed using a channel estimator to construct an estimated channel vector $\widehat{\mathbf{h}}$. This letter focuses on leveraging deep learning models for this channel estimation task as will be discussed shortly in Sections III and IV.

Downlink data transmission: Based on the estimated channel vector, the downlink beamforming $\mathbf{f}$ is constructed as a conjugate beamforming, i.e., $\mathbf{f} = \widehat{\mathbf{h}}^* / \|\widehat{\mathbf{h}}\|$. With this design, the downlink receive SNR per *transmit* antenna can be written as

$$\text{SNR}_{\text{ant}} = \frac{\rho}{M} \frac{\left|\widehat{\mathbf{h}}^H \mathbf{h}\right|^2}{\|\widehat{\mathbf{h}}\|^2}, \tag{3}$$

where ρ denotes the average receive SNR before beamforming.

III. Problem Definition

In this letter, we investigate the design of efficient channel estimation strategy that can construct the channel $\widehat{\mathbf{h}}$ from the highly quantized received signal $\mathbf{Y}$. More specifically, assuming that the pilot sequence $\mathbf{x}$ is known to the BS (the receiver), our objective is to develop a channel estimation strategy that minimizes the normalized mean-squared error (NMSE) between the estimated channel and the original channel vectors, defined as

$$\text{NMSE} = \mathbb{E}\left[\frac{\|\mathbf{h} - \widehat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}\right]. \tag{4}$$

Traditionally, channel estimation algorithms attempt to process the quantized signal $\mathbf{Y}$ to estimate the channel $\mathbf{h}$. Since $\mathbf{Y}$ is highly quantized, though, very long pilot sequences normally need to be used to achieve a reasonable channel estimation quality. To overcome this challenge, we propose to exploit deep learning models to learn how to efficiently estimate the channels from the quantized measurements while requiring only small pilot sequences.

IV. Deep Learning Based Channel Estimation

Classical channel estimation techniques for massive MIMO systems with low resolution ADCs, such as [1], [2], attempt to estimate the channel only from the quantized received signal, without using prior observations. The channels, however, are intuitively some functions of the different elements of the environment, such as the environment geometry, materials, transmitter/receiver positions, etc. [8]. This means that the BSs deployed in a certain environment will likely experience similar channels more than once. Hence, prior experience could be leveraged to learn the underlying relation between the quantized received signals and the channels. This has the potential of significantly reducing the pilot length. With that in mind, we propose to utilize deep learning to learn the mapping from the quantized received measurement matrix $\mathbf{Y}$ to the channel $\mathbf{h}$. Next, we first establish the conditions under which this mapping exists then highlight an interesting observation about how scaling the number of antennas up scales the number of required pilots down.

A. Mapping Quantized Measurements to Channels

Now, we investigate the existence of the mapping from quantized measurements to channels and highlight the motivation to leverage deep learning models. We also briefly explain why our proposed approach has the potential of reducing the number of pilots. First, consider a setup (indoor or outdoor environment) where a massive MIMO BS is serving a single-antenna user, as described in Section II. Let $\{\mathbf{h}\}$ denote the

set of candidate channels for the user, which depends on the candidate user positions as well as the surrounding environment. Further, let $\{\mathbf{Y}\}$ represent the corresponding quantized measurement matrices for the channel set $\{\mathbf{h}\}$ and a pilot sequence $\mathbf{x}$. We can then define the mapping from quantized measurements to channels, $\mathbf{\Phi}(.)$, as

$$\mathbf{\Phi} : \{\mathbf{Y}\} \rightarrow \{\mathbf{h}\}. \tag{5}$$

Note that if this mapping exists and is known, it can be used to predict the channel vector $\mathbf{h}$ from the quantized receive matrix $\mathbf{Y}$. Next, we establish existence of this mapping in Proposition 1 and then discuss how we can learn it.

Proposition 1: Consider the system and channel model in Section II with $\mathbf{N} = 0$ and a set of candidate channels $\{\mathbf{h}\}$. Define the angle α as

$$\alpha = \min_{\substack{\forall \mathbf{h_u}, \mathbf{h_v} \in \{\mathbf{h}\} \\ u \neq v}} \max_{\forall m} |\angle[\mathbf{h_u}]_{\mathrm{m}} - \angle[\mathbf{h_v}]_{\mathrm{m}}|. \tag{6}$$

If the pilot sequence $\mathbf{x}$ is constructed to have a length N satisfying $N \geq \lceil \frac{\pi}{2\alpha} \rceil$, with the angles of the pilot complex symbols uniformly sampling the range $]0, \frac{\pi}{2}]$, then the mapping function $\mathbf{\Phi}(.)$ exists.

Proof: See the Appendix. ∎

Proposition 1 means that once the pilot sequence is designed with the specific structure described in the proposition, then there exists a one-to-one mapping $\mathbf{\Phi}(.)$ that can map the quantized measurement matrix $\mathbf{Y}$ to the channel $\mathbf{h}$, i.e., it can use $\mathbf{Y}$ to predict $\mathbf{h}$. Interestingly, as we will see in Section V, only a few pilot symbols (very small N) are needed in massive MIMO systems to make this mapping $\mathbf{\Phi}(.)$ exist with high probability. This has the potential of significantly reducing the channel training overhead compared to classical 1-bit ADC channel estimation techniques. To be able to leverage this mapping function, however, we need to know it. Characterizing this mapping analytically is very non-trivial mainly because of the non-linear quantization. With this motivation, we propose to exploit the powerful learning capabilities of deep neural networks to learn this mapping and harvest the promising gains in reducing the channel training overhead. Before describing the adopted deep learning model in Section IV-C, we first highlight in the next subsection an interesting gain of using the proposed deep learning based 1-bit ADC channel estimation approach in massive MIMO systems.

B. Discussion: More Antennas Need Fewer Pilots

As shown in Proposition 1 and its proof, the desired pilot sequence should have a length that guarantees that every two different channels in $\{\mathbf{h}\}$ result in two unique quantized measurement matrices. Intuitively for the same uplink pilot sequence length, the more antennas deployed at the BS the more likely they lead to unique measurement matrices. Interestingly, this means that more antennas will lead to better channel estimation as will be shown in Section V. This also implies that when more antennas are employed at the BS, fewer pilots are required to guarantee that the mapping from $\{\mathbf{h}\}$ to $\{\mathbf{Y}\}$ is bijective (one-to-one). This interesting relation can be analytically characterized for several channel models. In the next corollary, we consider the LOS channel model, and prove that more antennas need fewer pilots.

Corollary 1: Consider a BS employing a ULA with half-wavelength antenna spacing and a single-path channel model ($L = 1$). Let $\delta\phi$ denote the smallest difference between any two angles of arrival, $\phi_1, \phi_2 \in [0, \pi[$, of any two users. If the pilot sequence is constructed according to Proposition 1, then the required pilot sequence length to guarantee that the mapping $\mathbf{\Phi}(.)$ exists is

$$N = \left\lceil \frac{1}{(M-1)(4\sin^2(\delta\phi/2))} \right\rceil. \tag{7}$$

Proof: The proof follows from Proposition 1 and is omitted because of the limited space. ∎

Corollary 1 captures clearly the interesting gain in our proposed deep learning approach, where more antennas require fewer pilots to ensure the existence of $\mathbf{\Phi}(.)$, i.e., to have the same channel estimation quality. This interesting finding will also be validated using numerical simulations in section Section V.

C. Proposed Deep Learning Model

To map back quantized received signal to complex-valued channels, we chose to utilize the expressive ability of deep learning [9], more specifically fully-connected neural networks. These networks are known to be good function approximators [10], and therefore, we design and train a dense neural network to learn the mapping from quantized measurements to channels.

Network architecture: The designed network has three dense stacks of layers. The first two are very wide and comprise a sequence of fully-connected layer, non-linearity layer, and dropout layer. The number of neurons in each fully-connected layer is L_{NN}, and they are followed by Rectified Linear Unit (ReLU) non-linearities. The last stack, which is the output, has only a fully-connected layer with $2M$-neurons.

Network training: As our target is estimating users' channels, we pose our learning problem as a regression problem in a supervised leaning setting; the network is trained to minimize a loss function measuring the accuracy of the predictions using a distance measure to some *desired* outputs. Given the channel estimation problem formulation in Section III, we choose NMSE as the loss function. Training is implemented using ADAM optimizer, and, therefore, we choose to minimize the average NMSE over a training mini-batch of size B users.

Data pre-processing: Before any training takes place, the inputs and outputs of the network must be pre-processed for efficient training, see [11]. The first stage of pre-processing normalizes those channels, whether in the training or testing datasets, to the range $[-1, 1]$ using the maximum absolute channel value obtained from the training set. We have found such normalization to be very useful in prior work [7], [8]. The second stage of the pre-processing is vectorizing the received quantized measurement matrices to have dimensions of M $N \times 1$. Finally, since popular deep learning software frameworks mainly support real-valued computations, channel and measurement vectors are decomposed into real and imaginary

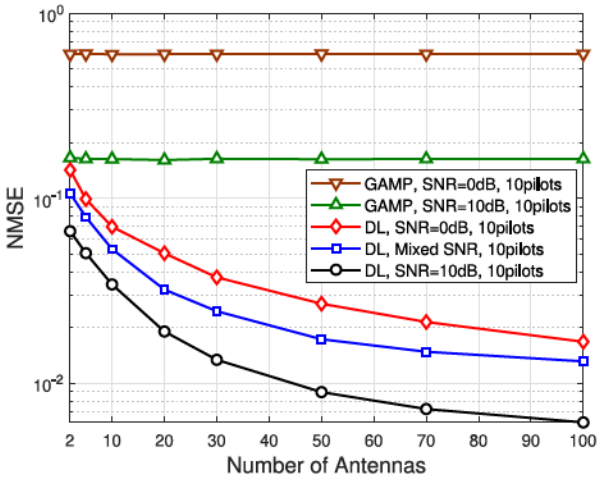

Fig. 2. The NMSE of the predicted channel using the proposed deep learning based approach and EM-GM-GAMP versus the number of BS antennas, for different pilot sizes and SNRs.

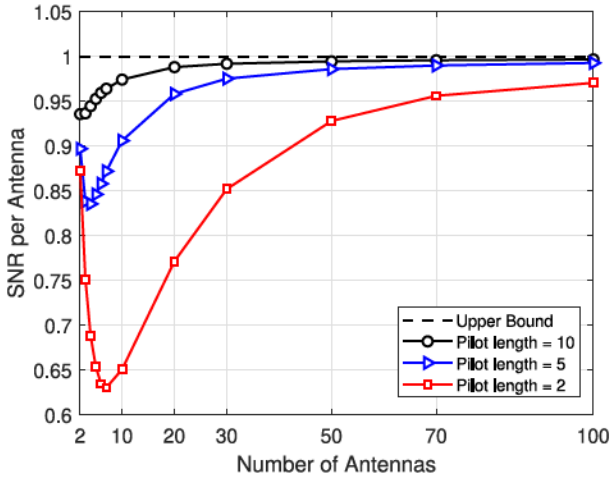

Fig. 3. Achievable SNR per antenna for the proposed deep learning based channel estimation approach and the upper bound for different BS antenna numbers and pilot sizes.

components and flattened into a $(2M \times 1)$ vectors for the channels and $2M$ N-dimensional vectors for the measurement.

V. Simulation Results

In this section, we evaluate the performance of our proposed deep learning based channel estimation solution for massive MIMO systems with 1-bit ADCs. We first describe the adopted scenario, dataset and the deep learning parameters used in our simulations and then discuss the interesting results.

A. Scenario and Dataset

In our simulations, we consider the indoor massive MIMO scenario 'I1_2p4' which is offered by the DeepMIMO dataset [12] and is generated based on the accurate 3D ray-tracing simulator Wireless InSite [13]. This scenario depicts a 10 m $\times$ 10 m room with two tables and users distributed across two x-y grids.

Given this ray-tracing scenario, we generate the DeepMIMO dataset which contains the channels between every candidate user location and every antenna at the BS. We adopt the following DeepMIMO parameters: (1) Scenario name: I1_2p4, (2) Active BSs: 32, (3) Active users: Row 1 to 502, (4) Number of BS antennas in (x, y, z): (1, 100, 1), (5) System bandwidth: 0.01 GHz, (6) Number of OFDM sub-carriers: 1 (single-carrier), (7) Number of multipaths: 10 in Fig. 2 and 1 in Fig. 3. To form training and testing datasets, we first shuffle the elements of the generated DeepMIMO dataset, and split it into 70% training set and 30% testing set. These datasets are then used to train the deep learning model and evaluate the performance of the proposed solution.

B. Model Training and Testing

The fully-connected network adopted in our simulation has two hidden layers, with 8192 neurons each.[1] The dimension of input and output layers depends on the number of antennas at BS and the number of pilots used for estimation. For example, if there are 100 antennas and 10 pilot symbols, then the input and output sizes will be 2000 and 200 respectively. We organize our training samples by $(\mathbf{y}, \mathbf{h})$, with $\mathbf{y}$ stands for the input and $\mathbf{h}$ for the target channel. Each sample corresponds to one user and all those samples are randomly drawn from the two grids.

The network is trained with 105981 samples for 100 epochs, and the performance of the trained network is evaluated with 45421 unseen samples. We use NMSE as our performance metric. To guarantee a fair comparison, the structure of the network and training parameters are kept unchanged in all our simulations, except for the input and output dimensions.

C. Performance Evaluation

Now, we evaluate the performance of our proposed deep learning based channel estimation approach, adopting the described uplink massive MIMO scenario in Sections V-A and V-B. In Fig. 2, we plot the NMSE versus the number of antennas (M) for different SNRs (0dB, 10dB, and mixed SNRs in the range of 0dB-10dB). This is implemented by adding noise samples to the measurement matrices used to train and test the deep neural network. Fig. 2 first shows that the proposed deep learning approach requires only a few pilots to have very accurate channel predictions at different SNRs. This is in contrast with the very long sequences normally required by classical (non-machine learning) channel estimation approaches [1], [2], such as expectation maximization Gaussian-mixture generalized approximate message passing (EM-GM-GAMP) [2]. As shown in Fig. 2, with the very short pilot sequences, the proposed deep learning approach has a clear gain over the EM-GM-GAMP solution..

Furthermore, Fig. 2 illustrates that the NMSE performance of the proposed solution improves significantly when more antennas are deployed at the BS, which confirms our results in Section IV-B. This interesting result is consistent with Proposition I. The minimum α (in rad) defined by (6) of the adopted dataset is 3.07×10^{-5} for the system with 2 antennas

[1]The code files that implement our deep learning based channel estimation solution are available in [14].

and 0.2476 for the 100 antenna system. Accordingly, based on Proposition I, the minimum pilot lengths guaranteeing the uniqueness of all the channels (in the form of quantized measurements) are 51166 and 7, respectively. This clearly shows the potential of our deep learning based approach that requires very small pilot sequences for massive MIMO systems. Further, even though Proposition 1 suggests that the number of pilots required for full bijectiveness (i.e., to be able to distinguish between any two channels) is very large for the case of 2 antennas, we can show that the *percentage* of these channels that need long pilots to be distinguishable is negligible. For example, with only 5 pilots, 98% of the *complex* channels in the dataset are distinguishable. This percentage increases to 99.5% with 10 pilots. All that clearly explains the good performance achieved by the proposed solution with only few pilots.

In Fig. 3, we repeat the same experiment of Fig. 2 and compare the SNR *per antenna* defined in (3), for different pilot length and antenna numbers. In the figure, the SNR of the received measurement matrices is fixed at 0 dB. Fig. 3 shows that for a fixed pilot length, despite a dip in the region of small number of antennas, the achievable per-antenna SNR approaches the upper bound as the antenna number increases. This upper bound is achieved by designing the conjugate beamformer based on the exact channel knowledge. Interestingly, this is the case even when only 2 pilots are used, i.e., with $N = 2$. The reason that we get the dip (in case of $N = 2, 5$) is mainly because the improvement on the total SNR does not match the rate at which antenna number increases. However, we notice that when increasing the pilot length or antenna numbers, this dip gradually vanishes. This can be explained by the insights in Sections IV-A and IV-B, which conclude that the mapping from the quantized measurements to channels becomes more bijective as more pilots are sent or more antennas are employed.

VI. Conclusion

In this letter, we developed a deep learning based channel estimation framework for massive MIMO systems with 1-bit ADCs. We derived the structure and length of the pilot sequence that guarantee the existence of the mapping from quantized measurements to channels. We then showed that this existence requires fewer pilots for larger antenna numbers. This was confirmed using both analytical and simulation results which showed that only few pilots are required to efficiently estimate massive MIMO channels. Further, the results showed that the achievable SNR per antenna approach the upper bound with increasing the number of antennas, which highlights the promising gains for massive MIMO systems. For future work, it is interesting to extend the proposed approach for broadband systems with frequency-selective channels.

Appendix

Proof of Proposition 1: For the mapping $\mathbf{\Phi} : \{\mathbf{Y}\} \rightarrow \{\mathbf{h}\}$ to exist, the (inverse) mapping $\mathbf{\Psi} : \{\mathbf{h}\} \rightarrow \{\mathbf{Y}\}$ has to be bijective. To guarantee the bijectiveness of the mapping $\mathbf{\Psi}$, any two channels in $\{\mathbf{h}\}$ have to lead to two unique received quantized measurements in $\{\mathbf{Y}\}$. Consider any two channels $\mathbf{h}_\mathrm{u}, \mathbf{h}_\mathrm{v} \in \{\mathbf{h}\}$, and a pilot sequence $\mathbf{x}$, the following condition needs then to be satisfied

$$\mathrm{sgn}\Big(\mathbf{h}_\mathrm{u}\mathbf{x}^\mathrm{T}\Big) \neq \mathrm{sgn}\Big(\mathbf{h}_\mathrm{v}\mathbf{x}^\mathrm{T}\Big). \tag{8}$$

This can be achieved if we guarantee that there exists at least one pair of corresponding elements in (8) that satisfies

$$\mathrm{sgn}([\mathbf{h}_\mathrm{u}]_\mathrm{m}[\mathbf{x}]_\mathrm{n}) \neq \mathrm{sgn}([\mathbf{h}_\mathrm{v}]_\mathrm{m}[\mathbf{x}]_\mathrm{n}), \tag{9}$$

for any $m \in 1, \ldots, M, n \in 1, \ldots, N$. Since the elements of the channel and pilot vectors are complex values, we can view these elements as vectors in the complex plane. If we write $[\mathbf{x}]_\mathrm{n}$ as $|x_\mathrm{n}|e^{j\theta_\mathrm{n}}$, then multiplying the channel element $[\mathbf{h}_\mathrm{u}]_\mathrm{m}$ by $[\mathbf{x}]_\mathrm{n}$ simply rotates $[\mathbf{h}_\mathrm{u}]_\mathrm{m}$ by θ_n and scales it by $|x_\mathrm{n}|$. In this sense, achieving (9) is by ensuring that $[\mathbf{h}_\mathrm{u}]_\mathrm{m}$ and $[\mathbf{h}_\mathrm{v}]_\mathrm{m}$ are rotated by $[\mathbf{x}]_\mathrm{n}$ to lie in two different quadrants of the complex plane. This can happen if we have a pilot sequence of length $N \geq \lceil \frac{\pi}{2\overline{\alpha}} \rceil$, where

$$\overline{\alpha} = \max_{\forall m} |\angle[\mathbf{h}_\mathrm{u}]_\mathrm{m} - \angle[\mathbf{h}_\mathrm{v}]_\mathrm{m}|, \tag{10}$$

and draw these N pilots such that their angles evenly sample the range $]0, \frac{\pi}{2}]$. Finally, to ensure that any two channels in $\{\mathbf{h}\}$ satisfy the condition in (8), we find the minimum $\overline{\alpha}$ defined in (10) among all channel pairs in $\{\mathbf{h}\}$, that is

$$\alpha = \min_{\substack{\forall \mathbf{h}_\mathrm{u}, \mathbf{h}_\mathrm{v} \in \{\mathbf{h}\} \\ u \neq v}} \max_{\forall m} |\angle[\mathbf{h}_\mathrm{u}]_\mathrm{m} - \angle[\mathbf{h}_\mathrm{v}]_\mathrm{m}|. \tag{11}$$

This leads to $N \geq \lceil \frac{\pi}{2\alpha} \rceil$, which concludes the proof.

References

[1] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

[2] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[3] Y.-S. Jeon, S.-N. Hong, and N. Lee, "Supervised-learning-aided communication framework for MIMO systems with low-resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7299–7313, Aug. 2018.

[4] E. Balevi and J. G. Andrews, "One-bit OFDM receivers via deep learning," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4326–4336, Jun. 2019.

[5] A. Klautau, N. González-Prelcic, A. Mezghani, and R. W. Heath, "Detection and channel equalization with deep learning for low resolution MIMO systems," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Oct. 2018, pp. 1836–1840.

[6] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1989–1993, Nov. 2019.

[7] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," May 2019. [Online]. Available: arXiv:1905.03761.

[8] 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.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

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

[11] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Heidelberg, Germany: Springer, 2012, pp. 9–48.

[12] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.

[13] *Wireless Insite*, Remcom, State College, PA, USA. [Online]. Available: http://www.remcom.com/wireless-insite

[14] [Online]. Available: https://github.com/YuZhang-GitHub/1-Bit-ADCs.