Recurrent Neural Network Channel Estimation Using Measured Massive MIMO Data

Termeh Faghani*, Arman Shojaeifard†, Kai-Kit Wong§, and A. Hamid Aghvami*

*King's College London, London WC2R 2LS, UK

†BT Labs, Adastral Park, Ipswich IP5 3RE, UK

§University College London, London WC1E 7JE, UK

*termeh.faghani@kcl.ac.uk †arman.shojaeifard@bt.com §kai-kit.wong@ucl.ac.uk *hamid.aghvami@kcl.ac.uk

Abstract—In this work, we develop a novel channel estimation method using recurrent neural networks (RNNs) for massive multiple-input multiple-output (MIMO) systems. The proposed framework alleviates the need for channel-state-information (CSI) feedback and pilot assignment through exploiting the inherent time and frequency correlations in practical propagation environments. We carry out the analysis using empirical MIMO channel measurements between a 64T64R active antenna system and a state-of-the-art multi-antenna scanner for both mobile and stationary use-cases. We also capture and analyze similar MIMO channel data from a legacy 2T2R base station (BS) for comparison purposes. Our findings confirm the applicability of utilising the proposed RNN-based massive MIMO channel acquisition scheme particularly for channels with long time coherence and hardening effects. In our practical setup, the proposed method reduced the number of pilots used by 25%.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is the de facto air-interface technology in modern mobile communication systems. Massive MIMO provides several key capabilities including spatial-multiplexing – the ability to serve multiple users over the same time and frequency resources. However, as the number of spatial layers is increased, the resources required for pilots and channel-state-information (CSI) feedback per coherence interval increases accordingly, especially in the case of frequency-division-duplex (FDD) operation. Even with adequate radio resources in place, pilot contamination and non-full-reciprocity of the end-to-end link may affect the channel estimation adversely. As a result, enhancements in channel acquisition for massive MIMO continues to be an active area of research (e.g., 3GPP RAN1 Working Item on Rel. 16 NR MIMO Enhancements).

Artificial Intelligence (AI) has in recent years entered the classical signal processing layers of wireless communications [1]. The authors in [2] developed an end-to-end communication system based on autoencoders. In another work in [3], the authors studied the application of deep learning for channel estimation. Their results, based on a deep neural network (DNN), performs joint channel estimation and detection to recover the received signals. In [4], the feedback overhead for FDD massive MIMO systems was investigated. The proposed CsiNet in [4] exploits deep learning to convert the complex channel matrices to optimal codewords, with the downlink

channel assumed to be perfectly estimated via pilot-based modes of transmission.

The original CsiNet concept was extended in [5], where the authors utilised a long short-term memory (LSTM) network, providing better performance by extracting time correlation information between the channel matrices. In [6], the authors proposed a channel estimation method for massive MIMO in millimeter wave (mmWave) bands using iterative signal detection algorithms and deep learning. All the above works focus on reducing the CSI feedback by means of learning. In [7], the authors propose a cascaded DNN estimation method, which eliminates the need for CSI feedback entirely in time-division-duplex (TDD) massive MIMO. Their deep leaning-based system estimates the channel for all downlink subcarriers based on the uplink sparse channel matrix.

Most of the deep learning-based related work on channel estimation rely on CSI feedback or pilot assignment. In addition, the learning from all these works is carried out based on simulated theoretical channel models. Here, we propose a novel recurrent neural network LSTM-RNN-based massive MIMO channel estimation technique which exploits the inherent correlations of practical fading channels for recovering the current channel information from previous time slots. This is achieved by utilising empirical MIMO channel measurements between a state-of-the-art multi-antenna scanner and both a 64T64R active antenna system and a baseline legacy 2T2R base station (BS) for both stationary and mobile use-cases. We thoroughly analyze the data and demonstrate that in practice the complex MIMO channel matrices are correlated in time. Our findings demonstrate the validity of utilising the proposed massive MIMO channel estimation approach, which alleviates the need for CSI feedback and pilot assignment, particularly for hardened channels with long time coherence.

II. SYSTEM MODEL AND ASSUMPTIONS

We consider a massive MIMO system with M antennas at the BS and K single-antenna user equipments (UEs). The system is operating in long-term-evolution (LTE) mode. In this work, we utilise empirical MIMO channel measurements obtained from measurement campaigns in which a state-of-the-art scanning device is used to capture LTE reference signals from a 64T64R massive MIMO cell cite. Moreover, for comparison purposes, we also obtain empirical MIMO

channel measurements from a legacy 2T2R BS. In both cases, we conduct separate tests for mobile and stationary radio conditions. In the former, the results correspond to a test drive with a fixed speed of 10 km/s around the cell cites, whereas in the latter, the distance between the corresponding BS and scanner is fixed.

The total system radio resource is divided into N_{rb} resource blocks, each being 0.5 ms wide in time (a slot) and 180 kHz wide in frequency (12 subcarriers with a spacing of 15 kHz). We consider all channels to be full rank hence enabling spatial-multiplexing; this is confirmed through analyzing the data from our measurement campaigns. The channel condition number, defined as the ratio between maximum and minimum sigular values of the MIMO channel matrix, provides an additional indication of the air-interface suitability for MIMO operation. Extracting and utilizing empirical MIMO channel measurements is a key contribution of this work, providing the means to understand the impact of parameters and algorithms in the real-world.

In particular, we utilise the measured MIMO channel data to investigate the following concepts:

1) Channel Hardening: It is well-understood in the literature that for a massive MIMO system, with M antennas at the BS, as M grows larger, spatial diversity leads to the channel hardening phenomenon [8]. This means the wireless channel variations are closer to the mean which results in the fading channel behaving in a deterministic fashion within a coherence interval. This concept has inspired many blind channel estimation methods [9]. In this work, we train our RNN-based estimator on data from a massive MIMO BS (64T64R) as well as a legacy BS (2T2R) for both mobile and stationary use-cases in order to test whether channel hardening exits in practice and to understand its impact on channel estimation.

2) Channel Time Coherence: Although in theory the time-varying fading channel is typically considered to be random from one time slot to another, that is not exactly the case in practice. Specially in low-mobility conditions, channel matrices can be fairly correlated for long periods. Here, we aim to exploit the time-correlated behavior of practical fading channels in our RNN-based estimation method. We use the LSTM network to estimate the future channel based on previous channel measurement. Based on our measurements, the method outperforms the conventional blind estimation, and in some cases even those with pilot-assisted estimation.

3) Stationary vs. Mobile Conditions: For the stationary user, channel matrices are expected to be correlated for longer periods (i.e., larger time coherence interval). However, for mobile use-cases, due to mobility of the user and Doppler spread, channel matrices are expected to be correlated on a smaller time scale in comparison. Let the time coherence interval to be denoted with t_c, which can be calculated as $\lambda/2v$ where λ is wavelength of the carrier signal and v is the velocity of the UE. We train RNN-based estimator with channel instances sampled with Δt intervals. Here, we consider the cases where $\Delta t \geq t_c$ and $\Delta t < t_c$.

III. RNN-BASED CHANNEL ESTIMATION

Let $H^{(t_i)}$ be the channel matrix of n_{RB}-th resource block at time t_i. H is sampled every Δt. We train a LSTM network, which estimates the channel matrix at time t_i based on the T previous channel matrices sampled every Δt. The problem can be formulated as in the following:

$$\hat{H}^{(t_i)} = f_{\text{est}}(H^{(t_i-\Delta t)}, H^{(t_i-2\Delta t)}, ..., H^{(t_i-T\Delta t)}) \qquad (1)$$

where f_{est} is the LSTM estimator and T is the number of time steps or number of previous values of H that are considered for the estimation of current value of H.

Finding the optimal value of T and Δt is of high importance in training the channel estimator. Note that different values of T and Δt result in different accuracy values for the estimator. Since we are considering time-correlation property of practical channels, if channel matrices are correlated in the interval $(T+1)\Delta t$ the estimator is expected to be highly accurate. In the mobility scenario, the term can be compared with the correlation time of the channel (t_c), which was defined previously. In theory, the channel is frequency-flat in an interval t_c, and completely random in the next interval. However, in practice, this is not the case. Subsequent channel matrices become less and less random especially in cases with long time coherence and channel hardening present. If T grows larger in a way that $(T+1)\Delta t >> t_c$, the randomness of the channel outweighs the time correlation and thus results in low accuracy. If T is relatively small the estimator might not properly capture the time correlation between the channel matrices. We train the LSTM network with different values for T and choose the one which results in highest accuracy between the empirical and predicted channel.

A. LSTM network architecture

An LSTM cell is consisted of an Input, an Output, and a Forget gate. The Forget gate decides which information the cell should throw away. The Input gate determines what new information should be stored while updating the cell state, and Output gate decides what information should be output based on the cell state.

A single cell of an LSTM network can be formulated as:

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \qquad (2)$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \qquad (3)$$

$$\tilde{c}_t = \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}) \qquad (4)$$

$$c_t = f_t\Delta c_{t-1} + i_t\Delta\tilde{c}_t \qquad (5)$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \qquad (6)$$

$$h_t = o_t.\tanh(c_t) \qquad (7)$$

where W_{fh}, W_{fx}, W_{ih}, W_{ix}, $W_{\tilde{c}h}$, $W_{\tilde{c}x}$, W_{oh}, and W_{ox} are weight parameters. b_f, b_i, $b_{\tilde{c}}$, and b_o are bias values, and σ is the Sigmoid function. In (2), if the value of f_t (Forget cell) is equal to 1, all the information will be kept in the cell state and a value of 0 means it will get rid off all the information. Similarly, (3) acts for the input gate and new information. c_t in

(5) is the current cell state. o_t in (6) decides what information should be output and finally h_t denotes the output of the cell. Each layer of the LSTM network contains several number of cells. In order to fully exploit the channel matrix features, we stack multiple LSTM Layers followed by dense layer(s). We implement the aforementioned network architecture using Keras library. We use grid search for hyper parameter tuning and use Dropout in each layer to prevent over-fitting.

B. Data pre-processing

For each resource block, we train the channel estimator separately. Consider the resource block number n_{RB}. We pick N_s channel matrices of resource block n_{RB} with average condition number $cn_{n_{RB}}$ shown as H^{t_1}, H^{t_2},..., H^{t_i}, where $t_i - t_{i-1} = \Delta t$ and $i = 0, 1, ..., N_s$. We train the network for the resource block with a very good condition number $cn_{n_{RB}} \to 1$ and draw the estimation accuracies.

1) Feature scaling: After distribution fitting, we came to realize that the H^{t_i} values in the data set follows a Gaussian distribution with $\mathscr{CN}(\mu, \nu)$ where $\mu \to 0$ and ν in $(-40dB, -50dB)$. We scale the channel matrices (input features) in a way that the distribution become $\mathscr{CN}(\mu \to 0, \nu \to 1)$.

2) Input/Output Structure: Consider X to be the input array and y the out put array of the LSTM network. X is of shape $(N_s, T, 2 \times K \times M)$, where N_s is total number of samples, T is the number of time steps, and $2 \times K \times M$ is the size of feature vector or the flatten real equivalent of the channel matrix.

To illustrate, consider a 2×2 MIMO system. Assume T is equal to four, which means to estimate the current channel matrix, we take into account four previous MIMO channel matrix values. A training example (X_i, y_i) can be given as:

$$X_i = \begin{bmatrix} h_{11}^{(t_{i-4})} & h_{12}^{(t_{i-4})} & h_{21}^{(t_{i-4})} & h_{22}^{(t_{i-4})} \\ h_{11}^{(t_{i-3})} & h_{12}^{(t_{i-3})} & h_{21}^{(t_{i-3})} & h_{22}^{(t_{i-3})} \\ h_{11}^{(t_{i-2})} & h_{12}^{(t_{i-2})} & h_{21}^{(t_{i-2})} & h_{22}^{(t_{i-2})} \\ h_{11}^{(t_{i-1})} & h_{12}^{(t_{i-1})} & h_{21}^{(t_{i-1})} & h_{22}^{(t_{i-1})} \end{bmatrix} \quad (8)$$

$$y_i = \begin{bmatrix} h_{11}^{(t_i)} & h_{12}^{(t_i)} & h_{21}^{(t_i)} & h_{22}^{(t_i)} \end{bmatrix}. \quad (9)$$

Note that X_i and X_{i+1} or any two subsequent training examples have $T - 1$ row in common. To feed the data to the network, first, we have to convert the complex MIMO channel matrices to an equivalent real representation:

$$X_i = \begin{bmatrix} \mathscr{R}(\mathbf{X_i}) & \mathscr{I}(\mathbf{X_i}) \end{bmatrix}, y_i = \begin{bmatrix} \mathscr{R}(\mathbf{y_i}) & \mathscr{I}(\mathbf{y_i}) \end{bmatrix}. \quad (10)$$

C. Loss function and metrics

Since this is a regression problem and the probability of the data is Gaussian we use mean square error (MSE) as the loss function to be minimized. Adam optimizer is used for this purpose. We use MSE as the metric for comparing the predicted channel and the actual value. This performance measure is defined as:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (11)$$
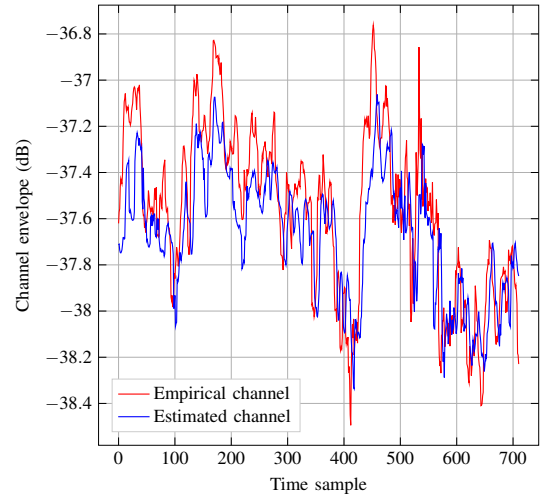
**Fig. 1:** Channel estimation between the BS antenna #1 and the UE #2 for the stationary 2T2R system.

where y_i, and $\hat{y}_i$ are the actual and predicted values of the channel, respectively.

IV. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we investigate the performance of the proposed LSTM-RNN channel estimator. Given our scheme requires a fewer number of pilots assigned, we compare the proposed model with two schemes: (i) a blind estimator, which utilizes the mean of the channel as the estimation in each time stamp and (ii) a conventional LS estimator which requires pilots in every time coherence interval. The analysis is carried out based on measured complex MIMO channel data for stationary and mobile conditions taken from both a 64T64R active antenna system and a legacy 2T2R BS.

Fig. 1 shows the channel estimation for the stationary 2T2R case between antenna #1 of the BS and UE #2. Note that the scanner time stamps are 30 ms apart. As it can be seen from the figure, even with less pilot assignment, LSTM-RNN estimator can successfully track the actual channel variations. This shows although in theory the channel is typically considered to be memoryless or with short time coherence, in practice and based on the our measurement campaigns, the MIMO channels behave differently. The LSTM-RNN looks back at $T = 3$ channel realizations to estimate the current channel matrix. Thus, in each $30 \times 4 = 120$ ms interval, channel values are highly correlated in time. Note that in the previous $T = 3$ time slots channel estimation is done via pilot transmission and in the fourth time slot , with the aid of LSTM-RNN estimator, channel value can be obtained with no pilot transmission. To measure the goodness of the proposed estimator we use MSE as metric (in dB) and compare the result with the baseline blind estimator.
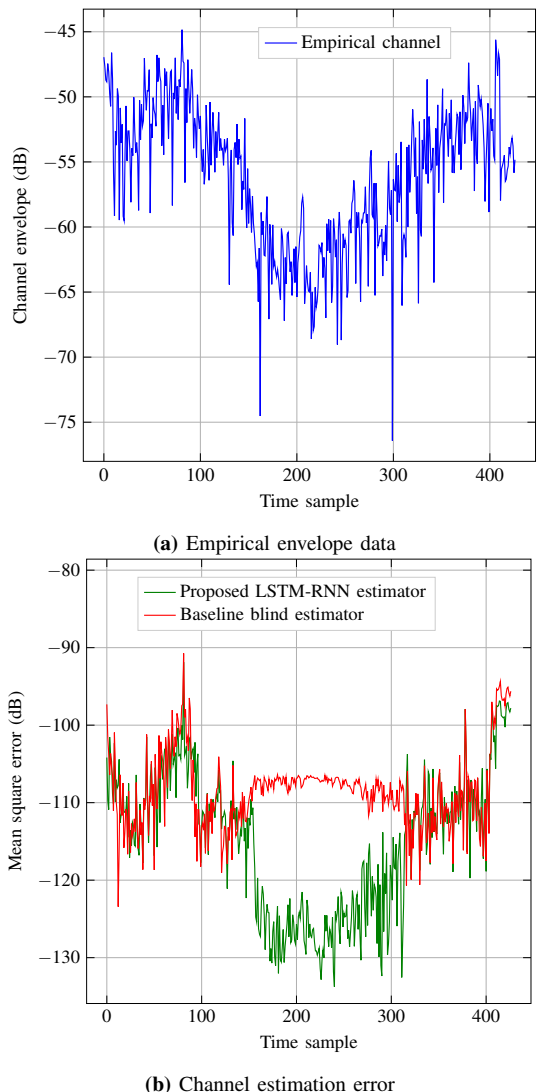
Fig. 2a shows the MSE in dB for both methods of estimation on the stationary 2T2R data. As it can be seen, the proposed method performs better by achieving a reduced error of at least 10 dB. For comparison, we trained the LSTM-RNN estimator

(a) 2T2R BS, stationary UEs

(b) 64T64R, stationary UEs
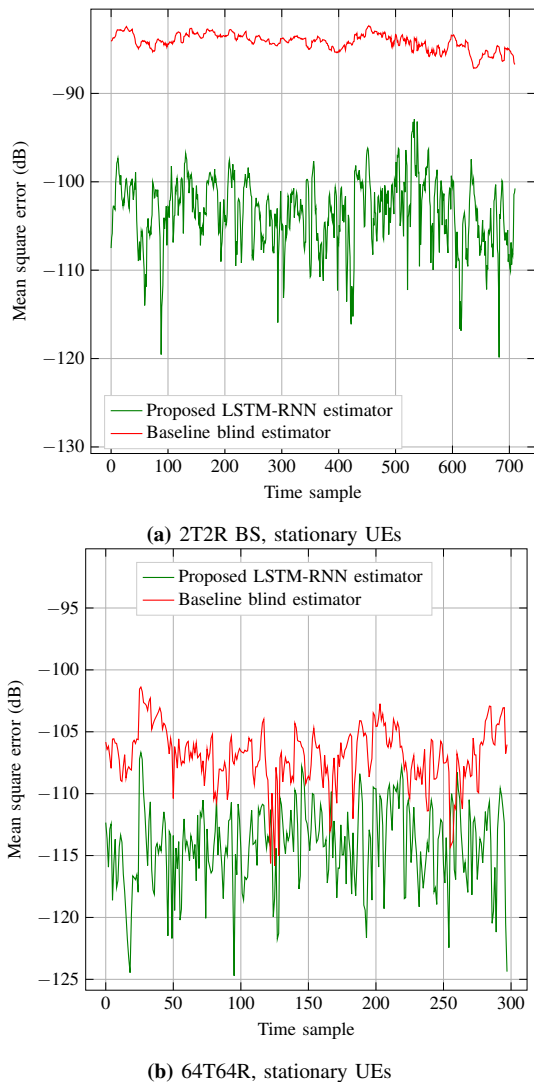
Fig. 2: Channel estimation error under stationary radio conditions.

(a) Empirical envelope data

(b) Channel estimation error

Fig. 3: (a) variations of the channel in time, and (b) channel estimation error for the 64T64R BS with mobile UEs.

for a stationary 64T64R massive MIMO system and drew the results in fig. 2b. Considering fig. 2a and fig. 2b, for the stationary use-case, we observe the following trends:

- When the number of antennas grows much larger ($64 \gg 2$), the estimation error becomes noticeably smaller and the gap between the proposed LSTM-RNN estimator and the blind estimator becomes smaller. This is because of the channel hardening effect from having larger number of antennas in massive MIMO. When the channel hardens, it behaves almost deterministic (close to mean), hence, a mean (blind) estimator performance gets closer to the proposed method.

- To train the estimator for the 64T64R active antenna system, time stamps are separated with an interval of 88 ms and the estimator looks at the previous $T = 2$ samples to estimate the current channel value. This results in a semi-coherence interval of $3 \times 88 = 264$ ms. This time is much larger than the actual coherence time of the channel especially in mobile condition, however, the results show

that although channel is not flat in this interval it can be estimated with high accuracy without pilot transmission.

To compare stationary and mobile conditions, consider fig. 2b, and fig. 3b. We expected the mobile channel to have higher estimation error since for the mobile user, channel varies more frequently than the stationary case. The average MSE over all time samples are higher in mobile conditions than the stationary case. Note that the LSTM-RNN estimator seems to capture channel variation quite well and leads to have an estimation error in the range of a stationary condition for time stamps 200-300. Comparing with the blind estimator, in fig. 3b for some time stamps, both methods perform almost the same, this suggests in some channel environments it is better to use the blind estimator rather than the RNN one. We provide empirical channel measurements in fig. 3a, which explains the high estimation error in some time stamps in fig. 3b results from high variations in channel (time stamps 100-
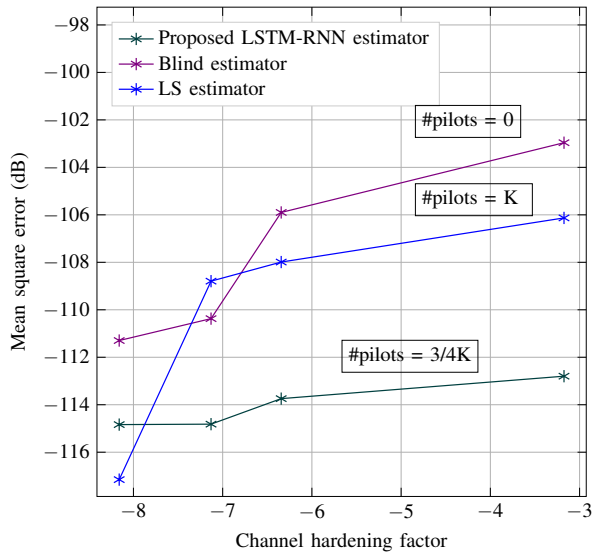
(a) 64T64R, stationary case
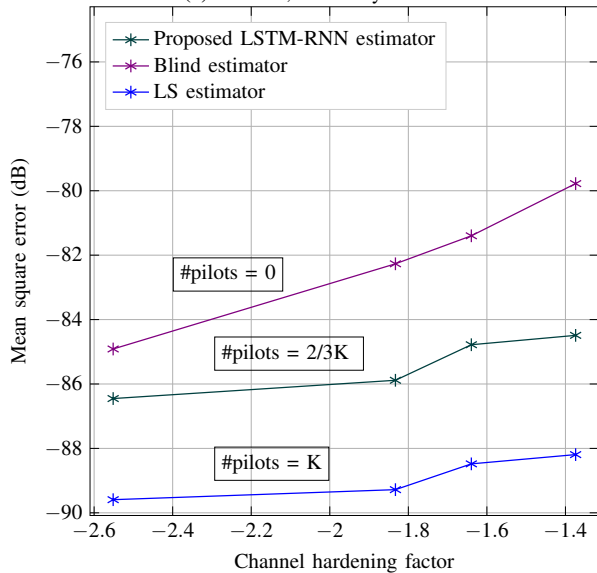
(b) 2T2R, mobile case

Fig. 4: Impact of channel hardening effects on the performance of the different estimation techniques.

200 in fig. 3a). We investigate this trend in figures 4b and 4a. Therein, the channel hardening effect in massive MIMO is studied. We define the channel hardening factor (CH), which is the channel standard deviation (std) (channel variation from the mean) divided by mean of the channel in dB [10]. Thus, when CH$\to -\infty$ the channel acts as almost deterministic and is assumed to be hardened.

In fig. 4a, for the 64T64R system, the proposed LSTM-RNN estimator performs better than both the LS estimator and the blind estimator. This suggests that the proposed estimator performs better when the MIMO channel is hardened. Moreover, the LSTM-RNN estimator is using less pilots than LS estimator in a way that for each 12 time slots, LS uses 24 pilots for two users, LSTM-RNN here uses 18 pilots for

two users (25% less control overhead). However, in fig. 4b, for the 2T2R system, the proposed method error is higher than the LS and lower than the blind estimator. Nevertheless, the proposed method is using less pilots than LS estimator. For each 12 time slots with two users, LS uses 24 pilots, LSTM-RNN uses 16 pilots for two users (33% less control overhead). The higher error for the proposed model in fig. 4b roots in the higher variation of the channel for the 2T2R mobile case.

In fig. 4a, for the 64T64R system, as CH gets higher it is expected for error to get higher which is indeed the case. In fig. 4b for the 2T2R system, the error is much higher than fig. 4a since CH is high and channel has not hardened. Note that in both figures as the CH gets lower and channel is behaving more deterministically, the gap between the different methods becomes smaller. The gap becomes larger as the variation of the channel is higher in the right side of both graphs. This suggests that, the proposed estimator performs notably better in fading environments with significant channel variations.

V. CONCLUSION

We designed a LSTM-RNN channel estimation method using measured complex MIMO channel data from a 64T64R active antenna system and a 2T2R legacy BS for two stationary and mobile use-cases. Our findings demonstrated that the proposed scheme, which alleviates per coherence interval CSI feedback or pilot assignment, outperforms the conventional blind detection approach and LS estimator in massive MIMO environments with channel hardening. For mobile use-cases, LSTM-RNN also proved to be a more efficient implementation, albeit higher estimation errors were recorded due to the shorter channel time coherence and higher fading variations.

REFERENCES

[1] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[2] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based mimo communications," *arXiv preprint arXiv:1707.07980*, 2017.

[3] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, Feb 2018.

[4] C. Wen, W. Shih, and S. Jin, "Deep learning for massive mimo csi feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, Oct 2018.

[5] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based csi feedback approach for time-varying massive mimo channels," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 416–419, April 2019.

[6] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, Oct 2018.

[7] N. Athreya, V. Raj, and S. Kalyani, "Beyond 5g: Leveraging cell free tdd massive mimo using cascaded deep learning," *arXiv preprint arXiv:1910.05705*, 2019.

[8] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[9] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in tdd massive mimo," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2921–2935, May 2017.

[10] S. Gunnarsson, J. Flordelis, L. Van der Perre, and F. Tufvesson, "Channel hardening in massive mimo-a measurement based analysis," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.