

# Deep Learning Based Predictive Beamforming Design

Juping Zhang, Gan Zheng, *Fellow, IEEE*, Yangyishi Zhang, Ioannis Krikidis, *Fellow, IEEE*, and Kai-Kit Wong, *Fellow, IEEE*

**Abstract**—This paper investigates deep learning techniques to predict transmit beamforming based on only historical channel data without current channel information in the multiuser multiple-input-single-output downlink. This will significantly reduce the channel estimation overhead and improve the spectrum efficiency especially in high-mobility vehicular communications. Specifically, we propose a joint learning framework that incorporates channel prediction and power optimization, and produces prediction for transmit beamforming directly. In addition, we propose to use the attention mechanism in the Long Short-Term Memory Recurrent Neural Networks to improve the accuracy of channel prediction. Simulation results using both a simple autoregressive process model and the more realistic 3GPP spatial channel model verify that our proposed predictive beamforming scheme can significantly improve the effective spectrum efficiency compared to traditional channel estimation and the method that separately predicts channel and then optimizes beamforming.

**Index Terms**—channel prediction, beamforming, deep learning.

## I. INTRODUCTION

Timely and accurate channel state information (CSI) is essential to exploit the full potential of multiuser multi-antenna systems by designing the optimal transmission strategies such as beamforming, but it is challenging to obtain in practice. Traditionally downlink CSI is obtained at a base station (BS) via either feedback from users, or channel estimation via uplink pilots by using the channel reciprocity. Both methods introduce significant overhead, extra error and latency, and as a result the CSI at the BS becomes outdated for beamforming design especially in high-mobility scenarios, e.g., unmanned aerial vehicle and vehicle-to-everything communications.

A more efficient channel acquisition method is to predict channels based on historical CSI data by exploiting the temporal correlation. There has been a large body of research on

channel prediction. Early works assume the accurate channel model such as the autoregressive (AR) process or the long-term channel statistics is available, and typically Kalman filtering [1] is employed to estimate the AR coefficients. However, practical channels may not be characterized by analytical models and they could be non-stationary, which degrade the performance of model-based prediction methods.

Recently deep learning based channel prediction has received much attention for its ability to learn CSI from data without prior knowledge about channel models. Starting from a single-antenna system, an efficient long-short term memory (LSTM) network, a type of recurrent neural network (RNN), is proposed in [2] for CSI prediction and adaptation in non-stationary changing channels. Going further, a data-driven receiver architecture that reduces the pilot overhead is designed in [3], following RNN-based channel prediction. The work in [4] begins by designing and conducting a measurement campaign to collect IQ samples of the IEEE 802.11p transmission and extract CSI in various real-world vehicular environments. The LSTM method is then employed to predict future CSI and received signal levels which is verified by trace-based evaluation. Deep learning based channel prediction is extended in [5] to massive multiple-input-multiple-output (MIMO) systems, which improves the channel prediction quality for both low and high mobility scenarios. A comparative study on a vector Kalman filter (VKF) predictor and a deep learning based predictor for massive MIMO systems using the spatial channel model (SCM) is carried out in [6], and it is shown that both can achieve substantial gain over the outdated channel in terms of the channel prediction accuracy and data rate. Following channel prediction, the beamforming optimization has also been studied. A deep learning approach is adopted in [7] to predict the angles between the unmanned aerial vehicle (UAV) and the user equipment in the presence of jittering due to the inherent random wind gusts, such that the UAV and the UE can prepare the transmit and receive beams in advance. A versatile unsupervised deep learning based predictive beamforming design is proposed in [8] in vehicular networks, which implicitly learns the features of historical channels and directly predicts the beamforming matrix to be adopted for the next time slot. The proposed method not only guarantees the required sensing performance, but also achieves a satisfactory sum-rate. Predictive beamforming is studied for dual-functional radar-communication (DFRC) systems in vehicular networks [9], in which a novel message passing algorithm based on factor graph is proposed to estimate the motion parameters of vehicles, and the beamformers are then

This work was supported in part by the CELTIC-Next Project C2019/2-5 Artificial Intelligence-enabled Massive Multiple-input multiple-output (AIMM), the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation, under the project INFRASTRUCTURES/1216/0017 (IRIDA), and the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/T015985/1.

J. Zhang and K.-K. Wong are with the Department of Electronic and Electrical Engineering, University College London, London, WC1E 6BT, UK (Email: {juping.zhang,kai-kit.wong}@ucl.ac.uk).

G. Zheng is with the School of Engineering, University of Warwick, Coventry, CV4 7AL, UK (Email: gan.zheng@warwick.ac.uk).

Y. Zhang is with the British Telecom, Adastral Park, Martlesham Heath, Ipswich, IP5 3RE. (E-mail: yangyishi.zhang@bt.com).

I. Krikidis is with the Department of Electrical and Computer Engineering, University of Cyprus, 1678 Nicosia Cyprus. (E-mail: krikidis@ucy.ac.cy).

For the purpose of open access, the authors have applied for a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

designed based on the predicted angles for establishing the communication links.

Different from the aforementioned works that focus on channel prediction only or have assistance from radar, we propose a deep learning framework that takes historical CSI data as input and directly predicts the beamforming solution for future channels to maximize the sum rate performance of a multiuser multi-antenna system. The framework incorporates an LSTM-based channel prediction module and a power optimization module which helps reconstruct the beamforming vectors using hybrid supervised-unsupervised learning. Furthermore, we propose to use the attention mechanism in the LSTM network such that the impact of historical channels in different coherent intervals will be correctly reflected in the channel prediction and this thus improves the performance of beamforming prediction.

The remainder of this paper is organized as follows. Section II introduces the system model and the problem formulation. Section III presents the proposed deep learning framework for predictive beamforming. Simulation results are given to validate the proposed method in Section IV and we conclude our work in Section V.

*Notation:* All boldface letters indicate vectors (lower case) or matrices (upper case). The superscripts  $(\cdot)^\dagger$  and  $(\cdot)^{-1}$  denote the conjugate transpose and the matrix inverse, respectively.  $\text{vec}(\cdot)$  denotes the vector operation of a matrix. The identity matrix is denoted by  $\mathbf{I}$ .  $\|\mathbf{z}\|$  denotes the  $L_2$  norms of a complex vector  $\mathbf{z}$ .

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

We consider a multi-input single-output (MISO) downlink system in which a BS with  $N_t$ -antennas serves  $K$  single-antenna users that employ single-user detection. Suppose  $s_k$  is the transmit signal to the user  $k$  with unit power and the BS transmits with a total power  $P_T$ . The transmitted data symbol  $s_k$  is mapped onto the antenna array elements by multiplying the beamforming vector  $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ . The received signal at the user  $k$  can be expressed as

$$y_k = \mathbf{h}_k^\dagger \mathbf{w}_k s_k + \sum_{j \neq k} \mathbf{h}_k^\dagger \mathbf{w}_j s_j + n_k, \quad (1)$$

where  $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$  is the channel between the BS and the user  $k$ , the second term represents the interference and  $n_k$  denotes the additive white Gaussian noise (AWGN) component with zero mean and variance  $\sigma_k^2$ . Therefore, the signal-to-interference-plus-noise ratio (SINR) that measures quality of the data detection at the  $k$ -th user is given by

$$\Gamma_k = \frac{|\mathbf{h}_k^\dagger \mathbf{w}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^\dagger \mathbf{w}_j|^2 + \sigma_k^2}. \quad (2)$$

We choose the sum rate as the system performance metric to maximize and the resulting problem is expressed as

$$\max_{\{\mathbf{w}_k\}} R \triangleq \sum_{k=1}^K \log(1 + \Gamma_k), \text{ s.t. } \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_T. \quad (3)$$

The sum rate optimization problem in (3) is nonconvex and the standard approach to find its suboptimal solution is to use the weighted minimum mean squared error (WMMSE) algorithm [10] assuming the CSI  $\{\mathbf{h}_k\}$  is available which normally relies on pilot-based channel estimation and introduces substantial overhead.

### B. Problem Formulation

In this paper, we adopt a hybrid channel-estimation-prediction scheme to solve the problem (3), in which we have CSI estimation of a certain number of channels and then use it to predict CSI of some future channels. Specifically, we assume the time horizon is divided into frames of  $N + P$  coherent intervals, and within each frame, the CSI of the first  $N$  coherent intervals is available, and the CSI of the rest  $P$  coherent intervals will be predicted without estimation.

The known CSI estimation is written as

$$\hat{\mathbf{h}}_k(n) = \mathbf{h}_k(n) + \mathbf{e}_k(n), \forall n = 1, \dots, N, \quad (4)$$

where  $\mathbf{h}_k(n)$  is the true CSI of user  $k$  at the  $n$ -th coherent interval and  $\mathbf{e}_k(n)$  is the channel estimation error that follows the complex Gaussian distribution with zero mean and variance matrix of  $\sigma_{e,k}^2 \mathbf{I}$ . We assume the least-square channel estimation is used, so  $\sigma_{e,k}^2$  depends on pilot transmission power.

The CSI prediction is expressed as

$$\tilde{\mathbf{h}}_k(N+m) = \mathbf{M}(\{\hat{\mathbf{h}}_k(n)\}, \forall n = 1, \dots, N), \forall m = 1, \dots, P, \quad (5)$$

with  $\mathbf{M}(\cdot)$  being the mapping from the known channel estimation to the channel prediction.

With the hybrid scheme, our aim is to solve the problem (3) by predicting the beamforming solutions directly for those  $P$  unknown channels given the  $N$  known channels.

## III. PREDICTIVE BEAMFORMING SOLUTION

### A. The General framework

Our proposed deep learning based framework to predict the beamforming solution is illustrated in Fig. 1 below.

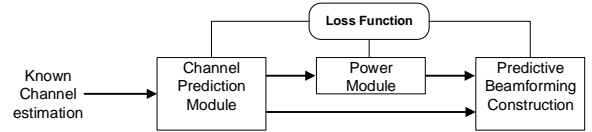


Fig. 1. The proposed framework for predictive beamforming.

The proposed framework includes three key modules. The first module is a neural network (NN) for channel prediction, which realizes the mapping  $\mathbf{M}(\cdot)$  from the input known channel estimations  $\hat{\mathbf{h}}_k(n), n = 1, \dots, N$  to the predicted channel  $\tilde{\mathbf{h}}_k(N+m), m = 1, \dots, P$ . Supervised learning is adopted and the associated loss function is

$$L_H = \frac{1}{2LK} \sum_{l=1}^L \sum_{k=1}^K \sum_{m=1}^P \left( \|\mathbf{h}_k(N+m)^{(l)} - \tilde{\mathbf{h}}_k(N+m)^{(l)}\|^2 \right), \quad (6)$$

where  $L$  is the number of batches, and the superscript  $(l)$  denotes the index of the batch. Details of the channel prediction module will be introduced in the next subsection.

The second module is an NN to predict power vectors in order to facilitate the prediction of beamforming solution. With the predicted channel, the beamforming solution could be inferred directly using supervised learning, but the high dimensional beamforming will make the training challenging and reduce the accuracy of inference. Instead, we exploit the following parameterized structure of the beamforming solution for the sum rate maximization problem:

$$\tilde{\mathbf{w}}_k = \sqrt{P_k} \frac{\left(\mathbf{I} + \sum_{j=1}^K \frac{q_j}{\sigma_j^2} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^\dagger\right)^{-1} \tilde{\mathbf{h}}_k}{\left\| \left(\mathbf{I} + \sum_{j=1}^K \frac{q_j}{\sigma_j^2} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^\dagger\right)^{-1} \tilde{\mathbf{h}}_k \right\|}, \quad (7)$$

where  $\{\mathbf{q}\}$  and  $\{\mathbf{p}\}$  are the uplink and downlink power vectors, respectively, with the same total sum power of  $P_T$ . This structure is adapted from the result for perfect CSI [11]. It can be seen that once the channel prediction is obtained, we can employ an NN to predict the power vectors  $\{\mathbf{q}\}$  and  $\{\mathbf{p}\}$  using supervised learning, and then reconstruct the beamforming solution using (7). The labelled data can be obtained using the WMMSE algorithm [10]. For simplicity, we use fully connected layers for the power module with details provided in Section IV. Suppose the associated loss is given by

$$L_P = \frac{1}{2LK} \sum_{l=1}^L \sum_{m=1}^P \left( \left\| \mathbf{q}^{(N+m,l)} - \tilde{\mathbf{q}}^{(N+m,l)} \right\|^2 \right) \quad (8)$$

$$+ \left( \left\| \mathbf{p}^{(N+m,l)} - \tilde{\mathbf{p}}^{(N+m,l)} \right\|^2 \right), \quad (9)$$

where  $\tilde{\mathbf{q}}$  and  $\tilde{\mathbf{p}}$  are the output of the power NN.

The next module will leverage the predicted channel and power to construct the predictive beamforming solution using (7). To improve the end performance of maximizing the sum rate, we also consider to incorporate the sum rate expression  $R(\{\tilde{h}_k, \tilde{\mathbf{w}}_k\})$  in (3) into the loss function for training the overall NN. The overall loss function is thus given as a weighted sum below:

$$L = L_H + L_P - bR(\{\tilde{h}_k, \tilde{\mathbf{w}}_k\}), \quad (10)$$

where  $b$  is a positive coefficient. The novel design of the loss function (10) reflects the fact that our proposed solution is a joint learning framework that incorporates channel prediction and power optimization, while it produces prediction for transmit beamforming directly. The rest of this section is devoted to the development of the channel prediction module.

### B. Attention-based LSTM Channel Prediction

Channel prediction is to predict future CSI given historical CSI data by exploiting the temporal correlation between them. The RNN is a well-suited machine learning technology to predict time series data [12]. However, standard RNNs have the issues of vanishing and exploding gradients during back-propagation, which makes predicting long time series data sequences challenging. LSTM is one of the most successful

variants of RNNs to predict the correlate time series data and can solve the issues of vanishing and exploding gradients [12], so it is adopted in this paper to predict the temporal correlated channel.

An LSTM is composed of a memory cell which can store data for long periods. The flow of information into and out of the cell is managed by three gates. Specifically, the forget gate determines what information from the previous state cell will be memorized and what information will be removed that is no longer useful; the input gate determines which information should enter the cell state; and the output gate determines and controls the outputs.

Most existing works in channel prediction use a simple LSTM and only its last hidden state is fed into a fully connected layer to produce the predicted channels. In this paper we propose an improved solution by introducing the attention mechanism to allow the algorithm to put focus on different historical channels. Specifically, the attention scheme will put higher weights on more recent channels to improve the future channel prediction. This is intuitive since not all historical CSI data have the same impact on a future channel, and it is mostly influenced by more recent channels. The advantage of the proposed attention scheme is that the weights will be optimized by training and do not need to be predetermined manually.

In this paper we propose an improved solution by introducing the attention mechanism to allow the algorithm to put focus on different historical channels. Specifically, the attention scheme will assign higher weights on more recent channels to improve the future channel prediction. This is intuitive since not all historical CSI data have the same impact on a future channel, and it is mostly influenced by more recent channels. The attention mechanism has gained remarkable success in sequence-to-sequence tasks like language translation and handwritten word recognition [13], but has not been used for channel prediction.

The proposed channel prediction module using LSTM with the attention mechanism is depicted in Fig. 2. It has  $N$  time steps each corresponding to the known channel estimation in one coherent interval. In order to best make use of the historical data we create an attention layer, which is located between the LSTM and the fully connected layer. This layer assigns a weight  $c_n$  to the hidden state output of each time step  $n$  with  $\sum_{n=1}^N c_n = 1$ , and then combines the weighted sum of the original hidden states as the LSTM output state. This will allow the importance of channel estimation in different coherent intervals to be correctly characterized. The new output state is then fed into fully connected layers to produce the predicted channels. Instead of treating the weights  $\{c_n\}$  as hyperparameters, one distinct advantage of the proposal scheme is to incorporate them in the overall neural network training, so there is no need to adjust them manually.

### C. Effective Sum Rate

In this subsection we define and analyze the effective sum rate  $R_E$  as a performance metric that incorporates the sum

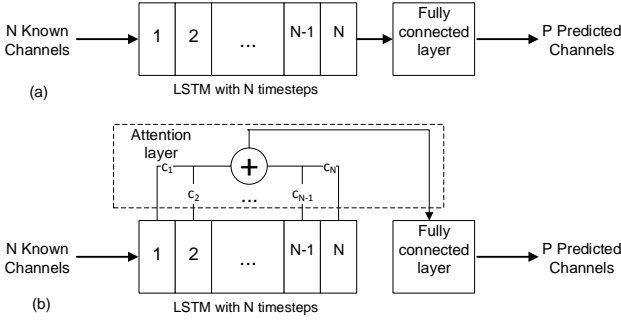


Fig. 2. (a) The traditional channel prediction using LSTM without attention; (b) The proposed channel prediction using LSTM with an attention layer.

rates of estimated channels and predicted channels, which can be written as

$$R_E = \frac{(1 - \alpha)N}{N + P} R_e + \frac{P}{N + P} R_p, \quad (11)$$

where  $\alpha$  as the portion of the channel estimation overhead in an coherent interval and therefore it satisfies  $\alpha < 1$ ,  $R_e$  is the average sum rate for the estimated channels, which can be obtained using the existing neural network method such as that in [14] as if the CSI is perfect, while  $R_p$  is the average sum rate for the predicted channels, which is obtained using the proposed predictive beamforming method. Note that compared to the traditional channel estimation based optimization method, our proposed method reduces the channel estimation overhead by  $\alpha P$ . Intuitively, the signalling overhead caused by channel estimation reduces the effective sum rate, so choosing a large  $N$  may be beneficial. However, the quality of channel prediction also relies on the number of known channel estimations. Therefore, the effective sum rate is in general not monotonically varying according to  $N$  and  $P$ , and a balance between low overhead and high-quality prediction needs to be achieved in practice.

## IV. NUMERICAL RESULTS

### A. Simulation setup

In this section, we provide numerical results to validate and evaluate the performance of the proposed deep learning-based predictive beamforming solution. Unless otherwise specified, we consider a MISO downlink consisting of  $N_t = 4$  transmit antennas and  $K = 3$  users. The transmit signal-to-noise (SNR)  $P_T$  and the variance of channel estimation error is  $\sigma_{e,k}^2 = 1/P_T, \forall k$ . The channel estimation overhead is set to be  $\alpha = 0.1$ . Unless otherwise specified, we assume  $N = 20$  known CSI data are available and  $P_T$  is 20 dB. In our simulation, we generate 30,000 training labels and 1,000 testing samples, respectively, using the WMMSE algorithm in [10]. For the LSTM layer, we use  $2N_tKN$  units while three fully connected layers each with  $30N_tK$  neurons, relu activation are used for the power NN.  $b$  is chosen to be 0.001 in (10). We use Python and Keras in Tensorflow to train the proposed deep learning model. All simulation results are generated by using a computer with an Intel i7-7700 CPU and an NVIDIA

Titan Xp GPU. The normalized MSE (NMSE) defined below is used as the performance metric for channel prediction:

$$\text{NMSE} = \mathbb{E} \left[ \frac{\|\mathbf{H} - \tilde{\mathbf{H}}\|^2}{\|\mathbf{H}\|^2} \right]. \quad (12)$$

The following benchmark schemes are considered for comparison:

- Predict beamforming without attention. It is the same as the proposed method except that the channel prediction module does not employ the attention mechanism.
- Separate optimization, i.e., to use the proposed LSTM-based method with attention to predict the channel, and the zero-forcing (ZF) beamforming is used to optimize the sum rate of both the predicted channels and the known channels.
- Channel estimation followed by ZF beamforming. This is the traditional estimation scheme with pilot overhead in which no channel prediction is used. It may not always be feasible due to the latency.
- Kalman filtering. This scheme uses Kalman filtering for channel prediction and ZF for optimizing the sum rate.

### B. Channel models

We consider two different scenarios for modelling the channel dynamics.

- The first scenario is the first-order AR process. Suppose we collect all users' channel as  $\mathbf{H}$ . In this scenario, the temporal evolution of the channel is given by

$$\text{vec}(\mathbf{H}_n) = \beta \text{vec}(\mathbf{H}_{n-1}) + \mathbf{u}_n, \quad (13)$$

where  $\beta = \mathcal{J}_0(2\pi f_D T_s)$ ,  $\mathcal{J}_0(\cdot)$  is the zeroth-order Bessel function of the first kind,  $f_D$  is the maximum Doppler frequency shift and  $T_s$  is the sampling duration.  $\mathbf{u}_n$  is the zero-mean Gaussian excitation noise with covariance matrix  $\sigma_u^2 \mathbf{I}$  with  $\sigma_u^2 = 1 - \beta^2$ . The composite term  $f_D T_s$  denotes the normalized Doppler rate. In the simulation, we choose  $f_D T_s = 0.005$ , so  $\beta = 0.9998$ , which corresponds to a slow user velocity of 2.7 km/h at a frequency of 2 GHz and sampling duration of 1 ms. For the Kalman filtering, an AR model with the prediction order of one in (13) is used together with the measure data in (4) to estimate the channels  $\{\mathbf{h}_n\}, 1 \leq n \leq N$ . While to predict the channels without measurement data, the simple state evolution is used:  $\mathbf{h}_{N+m} = \beta \mathbf{h}_{N+m-1}, 1 \leq m \leq P$ . Note that we have assumed that the model in (13) with the parameter  $\beta$  is known when designing the Kalman filtering, while for the proposed prediction method, neither the model nor the parameter is available and it will learn directly from the data.

- The second scenario is the 3GPP urban micro SCM which represents a more realistic but challenging channel environment [15]. We employ the default SCM parameters in [15] except that NumBsElements is  $N_t$ , NumMsElements is 1, and NumPaths is set to 1.

### C. Results

We first consider the first scenario of AR model. The NMSE results of the channel prediction are depicted in Fig. 3(a) against the number of predicted channels  $P$ . In general, the more channels to predict, the higher the NMSE is. It is observed that our proposed channel prediction with the attention mechanism is superior to the counterpart without attention, and both achieve much lower NMSE than the traditional channel estimation scheme. Note that this is because our proposed prediction exploits the temporal correlation of the channel, while the traditional channel estimation does not and it simply uses pilots to estimate the current channel. As the number of predicted channel becomes larger, the performance of prediction will unavoidably degrade and become worse than the traditional channel estimation. In the sequel for our proposed solution we assume attention is always used. Kalman filtering achieves the lowest NMSE among all considered schemes by making use of the model information and known parameters, which are not available to the proposed prediction method. The accuracy of the first  $N$  channel estimation results will affect the subsequent channel prediction performance, and therefore in Fig. 3(b) we show the results of NMSE of predicted channels versus the training SNR for the estimated channels when  $P = 20$ . As can be seen, the NMSE of the traditional channel estimation keeps decreasing as the SNR increases; while for the prediction methods, the NMSE saturates when the SNR is above a certain threshold. This is because the performance of channel prediction is limited by the number of predicted channels no matter how accurate the channel estimation is.

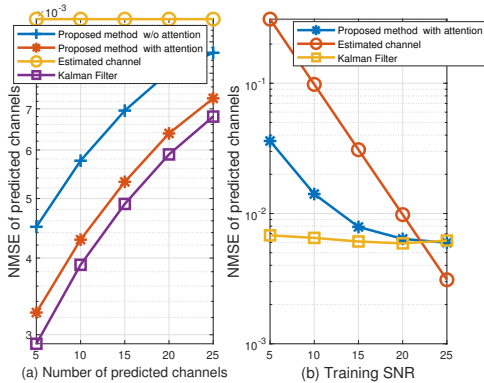


Fig. 3. The NMSE of the predicted channels versus the number of predicted channels; and (b) SNR,  $P = 20$ .

Next, we plot the sum rate and the effective sum rate normalized by the sum rate achieved by the WMMSE solution against the number of predicted channels in Fig. 4(a) and Fig. 4(b), respectively. It can be seen from Fig. 4(a) that our proposed solution achieves the highest sum rate while the sum rate with the traditional channel estimation is the lowest. The separate solution that first predicts the channel and then uses ZF beamforming achieves slightly lower sum rate than that of Kalman filtering. The same trend can be observed from the Fig. 4(b) about the effective sum rate. Because the channel estimation overhead is taken into account, the

achievable effective sum rates are lower than those in Fig. 4(a). For our proposed solution, the sum rate performance shows ceiling effect as the number of predicted channels increases. As per the analysis in Section III.C, this is because more future channels will reduce the prediction accuracy, and consequently the effective sum rate.

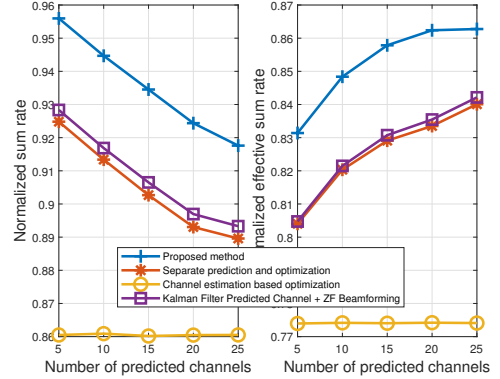


Fig. 4. The normalized sum rate results against the number of predicted channels.

Next, we examine the effect of the user velocity on the normalized sum rate and the results are provided in Fig. 5. As expected, the sum rate decreases for all schemes as user mobility increases. Kalman filtering performs well at low velocity, but it cannot track the change of channel dynamics well at high user mobility. Our proposed method achieves much higher effective sum rate than Kalman filter and the separate approach. The channel estimation method clearly outperforms others at high user mobility in theory, but it may not be practical to obtain the channel estimation in time.

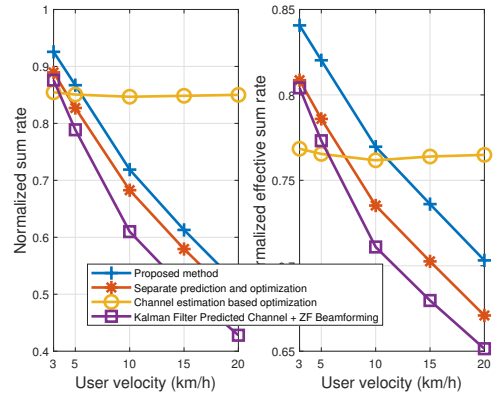


Fig. 5. The normalized sum rate results against the user velocity.

We then consider the second scenario of the urban micro SCM model. The normalized sum rate and effective sum rate results are shown in Fig. 6. Kalman filtering is not included in the comparison because in the simulation we found its performance is not satisfactory and this may be because the channel dynamic is too complex for Kalman filtering to predict. We can see from Fig. 6(a) that as the number of predicted channels increases, the sum rate of the separate optimization degrades quickly and is even much worse than the traditional channel estimation based optimization. Our proposed solution

still achieves the highest sum rate although as the number of predicted channels increases, the performance gap with the traditional solution becomes smaller. This highlights the importance of end to end learning of the predictive beamforming. Fig. 6(b) depicts the effective sum rate results. As expected, our proposed solution achieves superior performance, while the performance of the separate solution is worse than the estimation-based optimization when the number of predicted channels is high. The sum rates of both our proposed solution and the separate solution demonstrate the trend of first increasing and then decreasing. This again validates our analysis in Section III.C that a balance between the reduced overhead and high-quality channel prediction is necessary. For instance, for our proposed solution and the separate solution, the optimal numbers of predicted channels are 10 and 5, respectively.

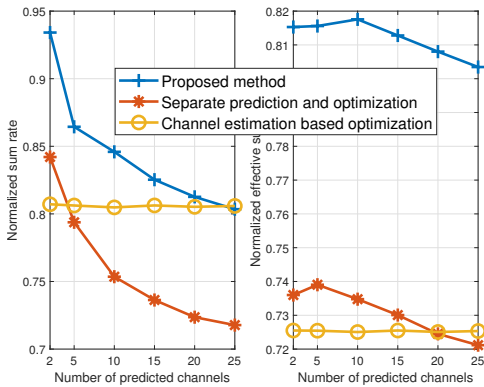


Fig. 6. The normalized sum rate results against the number of predicted SCM channels.

Finally, we assess the impact of the number of predicted channels  $P$  given a total frame length of  $N + P = 40$  in Fig. 7. There is no quantitative criterion on how to choose the optimal  $N$  and  $P$ . Intuitively, when  $N$  is larger, we have more channel information available to predict future channels more accurately, but this also causes higher overhead which will reduce the effective sum rate. A good tradeoff can be obtained by empirical study for a specific scenario. As can be seen from Fig. 7, for the AR-model, the channel is relatively easy to predict, i.e., with a small number of known channels  $N$ , the proposed algorithm can predict a large number ( $P = 30$ ) of future channels. While for the SCM model, it is more challenging to track the channel evolution, so only a small number ( $P = 10$ ) of future channels are predicted in order to achieve a high effective sum rate.

## V. CONCLUSIONS

In this paper, we have studied the predictive beamforming using the deep learning approach in the multiuser MISO downlink. A general framework that predict the beamforming solution to maximize the sum rate with historical channel measurement data was proposed. An LSTM with an attention layer was devised to improve the performance of channel prediction. Simulation results have shown that the proposed deep learning based solution achieves significantly higher effective sum rate over the traditional channel estimation based

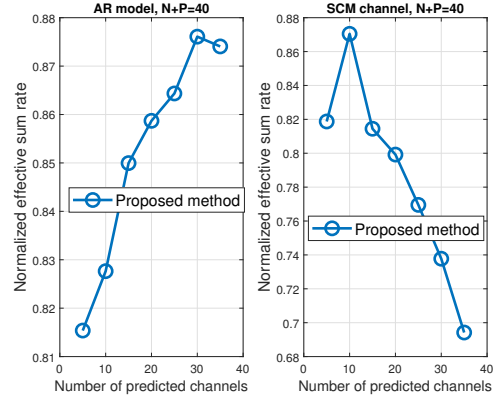


Fig. 7. The impact of the number of predicted channels on the normalized sum rate results given a fixed frame length of  $N + P = 40$ .

optimization and the separate prediction and then optimization scheme.

## REFERENCES

- [1] A. Duel-Hallen, S. Hu and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 62-75, May 2000.
- [2] Y. Zhu, X. Dong and T. Lu, "An Adaptive and Parameter-Free Recurrent Neural Structure for Wireless Channel Prediction," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8086-8096, Nov. 2019.
- [3] S. R. Mattu, L. N. Theagarajan and A. Chockalingam, "Deep Channel Prediction: A DNN Framework for Receiver Design in Time-Varying Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6439-6453, June 2022.
- [4] J. Joo, M. C. Park, D. S. Han and V. Pejovic, "Deep Learning-Based Channel Prediction in Realistic Vehicular Communications," *IEEE Access*, vol. 7, pp. 27846-27858, 2019.
- [5] J. Yuan, H. Q. Ngo and M. Matthaiou, "Machine Learning-Based Channel Prediction in Massive MIMO With Channel Aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960-2973, May 2020.
- [6] H. Kim et al., "Massive MIMO Channel Prediction: Kalman Filtering vs. Machine Learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518-528, Jan. 2021.
- [7] W. Yuan, C. Liu, F. Liu, S. Li and D. W. K. Ng, "Learning-Based Predictive Beamforming for UAV Communications With Jittering," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1970-1974, Nov. 2020.
- [8] C. Liu et al., "Learning-Based Predictive Beamforming for Integrated Sensing and Communication in Vehicular Networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2317-2334, Aug. 2022.
- [9] W. Yuan, F. Liu, C. Masouros, J. Yuan, D. W. K. Ng and N. González-Prelcic, "Bayesian Predictive Beamforming for Vehicular Networks: A Low-Overhead Joint Radar-Communication Approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1442-1456, Mar. 2021.
- [10] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792-4799, Dec. 2008.
- [11] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142-148, Jul. 2014.
- [12] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Computer Science*, article number: 420 (2021), Aug. 2021.
- [13] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, Dec. 2017.
- [14] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang and A. P. Petropulu, "A Deep Learning Framework for Optimization of MISO Downlink Beamforming," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1866-1880, Mar. 2020.
- [15] J. Salo, G. Del Galdo, J. Salmi, P. Kyösti, M. Milojevic, D. Lalselva, and C. Schneider, MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996), [Online]. Available: <http://www.tkk.fi/Units/Radio/scm/>, Jan. 2005.