

# LSTM-based Beam Tracking for mmWave Vehicular Networks

Chen Wang, XXX, XXX, XXX

*Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications*

Beijing, China

wangchen@bupt.edu.cn

**Abstract**—The use of millimeter wave (mmWave) frequency bands for transmission can improve data transmission rate with the help of beamforming technology to overcome the high path and penetration losses. However for vehicle, the high mobility of vehicle results in extremely frequent beam alignment and significant overhead. In this paper, a long short-term memory (LSTM)-based beam tracking scheme was proposed for reducing overhead brought by beam alignment in mmWave Vehicular Networks, by predicting beam angles at next time step through known beam angles at a certain number of consecutive time steps as features. To train this network, a time series array antenna channel data was set up by statistical channel model using time series vehicle information generated from road traffic simulation software named “Simulation of Urban MObility (SUMO)”. Simulation results show that the proposed LSTM-based scheme can reduce a significant overhead with an acceptable loss in spectral efficiency compared with exhaustive search scheme.

**Index Terms**—beam tracking, mmWave vehicular networks, LSTM

## I. INTRODUCTION

MmWave is one of the important technology for the future development of communications [1], relies on abundant spectrum resources. In order to apply mmWave better, the high path and penetration loss caused by high frequency of mmWave need to be overcome [2]. Thanks to the short wavelength of mmWave, it is possible to integrate array antennas into small chip and use beamforming technique to constrained signal power in specified direction for high gain [3].

Unlike conventional omni-directional antennas, array antennas are directional [4], which means that if high antenna gain needs to be achieved, it is important to ensure that the beams generated by array antennas at the receiver and transmitter are aligned well with each other. Otherwise the misalignment of the beam pairs will result in serious gain reduction, degrading communications performance. A hierarchical codebook design was proposed in [5], which essentially exploit binary search algorithm to reduce overhead by using different widths of beams at different stages to perform beam search. However, this algorithm requires a high quantization level phase shifter and potentially multiple RF chains to implement, which results in high cost. In [6], Zhang et al. proposed a Kalman filter based scheme for beam tracking using exhaustive search. However, the number of measurements and protocol overhead for exhaustive search required increases with the number of elements in antenna. On the other hand, [7] exploits extended Kalman

filter and requires only a single measurement, making it more suitable for beam tracking in fast-changing environments.

Providing reliable beam alignment for high speed vehicles is more challenging than for low speed terminals, due to vehicles move out of the beam coverage faster. To ensure that the vehicles are always covered by high gain beams, high frequency beam alignment is required, and it will further increase system overhead, especially when high resolution code books are used. Fortunately, due to the sparsity of mmWave channel [8] and the correlation between the angle of arrival (AoA) and angle of departure (AoD) of a millimeter-wave channel and the location of the receiver and transmitter [9], the beam angles of a vehicle traveling along a road under LOS condition can be well predicted.

In this paper, a LSTM-based beam tracking scheme is proposed. By using LSTM network to learn the historical channel state information (CSI) which was obtained from exhaustive beam search, AoA and AoD of current channel between vehicle and BS can be predicted to reduce system overhead. Meanwhile, a road traffic simulation software is used to generate time seires vehicle information (e.g. position, speed, etc.) to set up time seires CSI with a statistical channel model for training the network. Compared with prior work, this paper differs in the following ways: 1) More realistic channel model; 2) Simultaneous prediction of AoA and AoD; 3) The outage judgment is based on the SNR of the receiver instead of the angle difference; 4) Fewer beam measurements required on average in a tracking cycle. 5) In addition, beam search is performed more realistic by using a quantized phase shifter based beamforming.

The following notation will be used in this paper. Matrices, vectors and scalars are denoted by bold uppercase letters (e.g.  $\mathbf{A}$ ), bold lowercase letters (e.g.  $\mathbf{a}$ ) and lowercase letters (e.g.  $a$ ), respectively.  $(\cdot)^T$  denote transpose and  $(\cdot)^H$  denote conjugate transpose (Hermitian).  $[\mathbf{A}]_{m,:}$ ,  $[\mathbf{A}]_{:,n}$  and  $[\mathbf{A}]_{m,n}$  denote the  $m$ th row,  $n$ th column and the  $m$ th row  $n$ th column entry of  $\mathbf{A}$ , respectively.  $[a]_n$  denote the  $n$ th entry of  $a$ . Besides,  $\|\cdot\|_2$  denote  $\ell_2$ -norm of a vector.  $\mathbb{C}$  denote the set of complex number. Gaussian, complex Gaussian, wrapped Gaussian and exponential distribution are denoted by  $\mathcal{N}$ ,  $\mathcal{CN}$ ,  $\mathcal{WN}$  and  $\mathcal{E}$ , respectively.

## II. SYSTEM MODEL

A mmWave vehicular networks scenario including a vehicle with  $M_r$  antenna as receiver and a base station (BS) with  $M_t$  antenna as transmitter are considered. Both of them equips with uniform linear array (ULA) of half wave interval antenna as shown in Fig. 1, and adopted analog beamforming used quantized phase shifter which connect with single analog radio frequency (RF) chain. The array response vector of a uniform linear array with  $M$  half wave interval antenna elements is given by:

$$\mathbf{a}(M, \varphi) = \frac{1}{\sqrt{M}} [1, e^{j\pi \cos(\varphi)}, \dots, e^{j\pi(M-1)\cos(\varphi)}]^T \quad (1)$$

where  $\varphi$  is the arrival angle of the signal. When AoA of the vehicle is  $\phi$  and AoD of the BS is  $\theta$ , the array response vectors of both are  $\mathbf{a}_r(\phi) = \mathbf{a}(M_r, \phi)$  and  $\mathbf{a}_t(\theta) = \mathbf{a}(M_t, \theta)$ , respectively.

The statistical 28 GHz mmWave channel model in [10] was used, which is modeled by real experimental data collected in New York City. The narrowband time-varying  $L$  subpaths channel matrix between the vehicle and BS at  $n$ th time step is  $\mathbf{H}_n \in \mathbb{C}^{M_r \times M_t}$ , which is given as:

$$\mathbf{H}_n = \frac{1}{\sqrt{L}} \sum_{l=1}^L g_{ln} \mathbf{a}_r(\phi_{ln}) \mathbf{a}_t^H(\theta_{ln}) \quad (2)$$

where  $g_{ln} \in \mathbb{C}$  is complex small-scale fading gain on subpath  $l$  at  $n$ th time step,  $\phi_{ln}$  is AoA of the  $l$ th subpath signal received by the vehicle at  $n$ th time step, and  $\theta_{ln}$  is AoD of the  $l$ th subpath signal transmitted by the BS at  $n$ th time step.  $\phi_{ln} = \phi_n + \Delta\phi_{ln}$  and  $\theta_{ln} = \theta_n + \Delta\theta_{ln}$ , where  $\phi_n$  and  $\theta_n$  is the AoA and AoD of the cluster center at  $n$ th time step,  $\Delta\phi_{ln} \sim \mathcal{WN}(\delta_{rn}^2)$ ,  $\Delta\theta_{ln} \sim \mathcal{WN}(\delta_{tn}^2)$ ,  $\delta_{tn} \sim \mathcal{E}(\lambda_t)$  and  $\delta_{rn} \sim \mathcal{E}(\lambda_r)$ . The complex small-scale fading gain is given by:

$$g_{ln} = \bar{g}_{ln} e^{j2\pi t_n f_{Dn,max} \cos(\omega_{ln})}, \bar{g}_{ln} \sim \mathcal{CN}(0, 10^{-0.1PL_n}) \quad (3)$$

where  $t_n$  is the signal transmission time at  $n$ th time step,  $f_{Dn,max} = v_n/\lambda_c$  is maximum doppler shift at  $n$ th time step,  $v_n$  is the speed of vehicle at  $n$ th time step,  $\lambda_c$  is the carrier wavelength,  $\omega_{ln} = \phi_{ln} - \pi/2$  is the AoA of subpath  $l$  relative to the direction of vehicle at  $n$ th time step, and  $PL_n$  is omnidirectional path loss at  $n$ th time step, which is defined as:

$$PL_n = 61.4 + 20 \log(d_n) + \xi \text{ [dB]} \quad (4)$$

where  $d_n$  is distance between the vehicle and BS at  $n$ th time step, and  $\xi \sim \mathcal{N}(0, \sigma_p^2)$ . Note that only the case of single cluster under line-of-sight (LOS) condition is considered in this paper. In addition, the cluster angle depends on the geometric position of the vehicle and base station.

For a transmitted signal  $x_n$  ( $|x_n| = 1$ ) from the BS, received signal  $y_n$  in the vehicle at  $n$ th time step is:

$$y_n = \mathbf{w}_n^H \mathbf{H}_n \mathbf{f}_n x_n + \mathbf{w}_n^H \mathbf{v}_n \quad (5)$$

where  $\mathbf{w}_n$  and  $\mathbf{f}_n$  are combining vector and beamforming vector at  $n$ th time step, respectively.  $\mathbf{v}_n$  is Gaussian noise at time step  $n$  and  $\mathbf{v}_n \sim \mathcal{CN}(0, \sigma_v^2 \mathbf{I}_{M_r})$ .

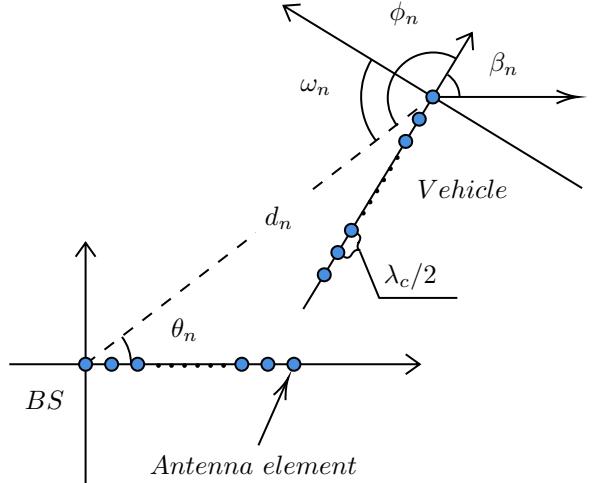


Fig. 1. Channel geometry

## III. LSTM-BASED BEAM TRACKING SOLUTION

### A. The LSTM Network

The purpose of beam tracking is to obtain beam angles ( $\phi_{ln}$  and  $\theta_{ln}$ ) with as few beam measurements as possible for reducing overhead brought by beam search. Since the vehicle must travel in a particular direction on the road and the mmWave beams are correlated with the position of vehicles, the problem of obtaining beam angles can be formulated as a time series regression problem after the historical CSI of vehicles is collected.

An artificial neural network named LSTM network was used to solve this regression problem. It has been implemented successfully in several fields such as machine translation, image captioning, speech recognition, and even stock prices prediction in economy. As one of the recurrent neural networks (RNNs), LSTM are generally considered to be more robust to long time series than simpler vanilla RNN implementations [11]. It is powerful for handling time series data, because relying on three gates which are forget, input and output gate composed of neural networks (NNs) in LSTM cells, it can learn long-term relations between features.

Fig. 2 show the structure of the network, which is a sequential model comprises LSTM layer with 50 hidden units, a dropout layer with 0.2 probability, a LSTM layer with 50 hidden units, a dropout layer with 0.2 probability and a fully connected layer with 2 nodes. The output of the first LSTM layer is a sequence of length  $n_{LSTM}$ , which is used as the input of the second LSTM layer. And two dropout layers are set to prevent overfitting.

### B. Loss Function

Mean Square Error (MSE) is the most commonly used regression loss function for quantifying the error of prediction. In [12], considering the problem of angle ambiguity due to the periodicity of the beam angle results in unnecessary loss for large difference of beam angle when used MSE. D. Burghal

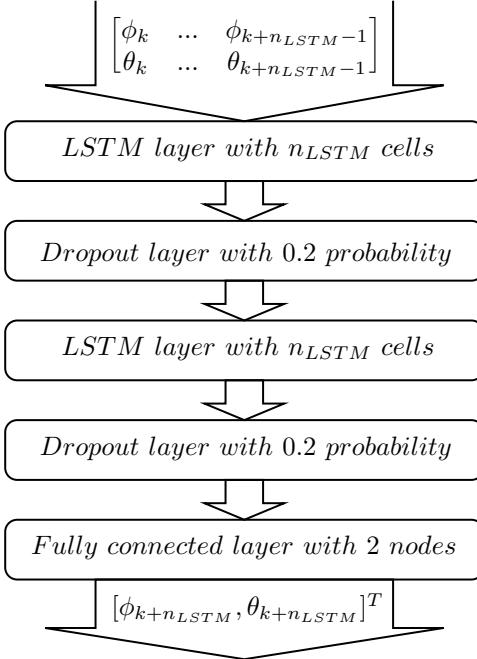


Fig. 2. Structure of network

et al. proposed to use a shifted cosine function as the loss function. However, when using this formula as a loss function, the training progress may be unstable due to the presence of multiple global minima, and the responses may not converge in a reasonable interval.

In addition, another problem is that the codewords whose beam angle near 0 or  $\pi$  have close gains at two different angles, resulting in unnecessary loss in some cases. For instance, in the case of the codeword whose beam angle is  $\pi$ , there are same gains at 0 and  $\pi$ . The loss should be small, even zero, if estimated angle is 0 and predicted angle is  $\pi$ , although the absolute difference of them is  $\pi$ . A straightforward idea is taking the negative of the dot product of the array response vector of estimated angle and predicted angle plus a positive constant as the loss function. Calculating this loss, however, requires array response vectors of estimated angle and predicted angle at the corresponding time step, which will increase the computational cost when training network. On the other hand, considering this function contains complex numbers, it will be difficult to perform backpropagation (BP).

Hence, to avoid the problem of angle ambiguity and close gains, the AoAs and AoDs are limited in  $(0, \pi]$  by simple conversion before training. Considering complexity of computation and training performance, a modified MSE is used as loss function, which is given as:

$$Loss = \frac{1}{N} \sum_{i=1}^N [(\phi_i^p - \phi_i^e)^2 + (\theta_i^p - \theta_i^e)^2] \quad (6)$$

where  $N$  is the number of training samples,  $\phi_i^e$ ,  $\theta_i^e$ ,  $\phi_i^p$ ,  $\theta_i^p$ , are the estimated AoA, estimated AoD, predicted AoA, and

predicted AoD, respectively, for training sample  $i$ .

### C. Procedure of Beam Tracking

Before the beam tracking, the BS demands to collect historical CSI, which are estimated by the vehicles using exhaustive beam search, for training the LSTM network. For time step  $k$ , the network takes the AoAs and AoDs from time step  $(k - n_{LSTM})$  to  $(k - 1)$  as features and the AoA and AoD at time step  $k$  as responses. The dataset is consisted of time series AoAs and AoDs of the vehicles from initial access to disconnection.

After the LSTM network has been trained, at the beginning of a round of beam tracking, a vehicle takes  $n_{LSTM}$  consecutive exhaustive beam search to obtain initial estimated AoAs and AoDs, and the network will take the estimated AoAs and AoDs to predict AoA and AoD at next time step. Then, the predicted AoA and AoD at  $k$ th time steps will be used to compose new feature which from  $(k - n_{LSTM} + 1)$ th to  $k$ th time steps to predict the AoA and AoD at  $(k+1)$  time step. For every time step, SNR under current predicted AoA and AoD will be calculated and compares with a threshold  $SNR_T$ . The network will repeat to use predicted result to compose new features to predict AoA and AoD, when the calculated SNR is greater than  $SNR_T$ . Otherwise, a new round of estimation and prediction will be repeated.

It makes more sense to use the SNR of received signal rather than the absolute difference of the angles as a threshold, because there is no way to obtain an accurate beam angle for comparing with the threshold until it is accurately measured or predicted.

## IV. DATA GENERATION

### A. Channel Matrix

A road traffic simulation software named “SUMO” was used to generate time series information of the vehicle in the scene mentioned in Section II, which including a BS and a vehicle driving on curve road as shown in Fig. 3. The vehicle departs from  $(0, 100)$  at first time step, and arrives  $(100, 0)$  at final time step. Note that the speed of vehicle is non-linear and depends on the road conditions and behavior of the driver which defined by SUMO.

The generated vehicle time series information is composed of position,  $v_n$  and directional angle  $\beta_n$  of the vehicle at each time step. For  $n$ th time step, the position of vehicle is given by  $(x_{rn}, y_{rn})$ ,  $v_n$  is used to determine maximum doppler shift in (3), and  $\beta_n$  is used to determine  $\phi_{ln}$  and  $\omega_{ln}$ .

Once the position of the vehicle and the BS are obtained,  $\theta_n$  and  $\phi_n$  can be calculated by simple geometric methods as following:

$$\begin{cases} \theta_n = \arg(\Gamma_{xn} + j\Gamma_{yn}) \\ \phi_n = \arg(\Gamma_{xn} + j\Gamma_{yn}) - \beta_n \end{cases} \quad (7)$$

where  $\Gamma_{xn} = x_{rn} - x_t$ ,  $\Gamma_{yn} = y_{rn} - y_t$ ,  $(x_t, y_t)$  are coordinates of the BS,  $\beta_n$  is directional angle of the vehicle. Furthermore, the signal transmission time  $t_n = d_n/c$  can be calculated, where the distance between the BS and the vehicle is  $d_n =$

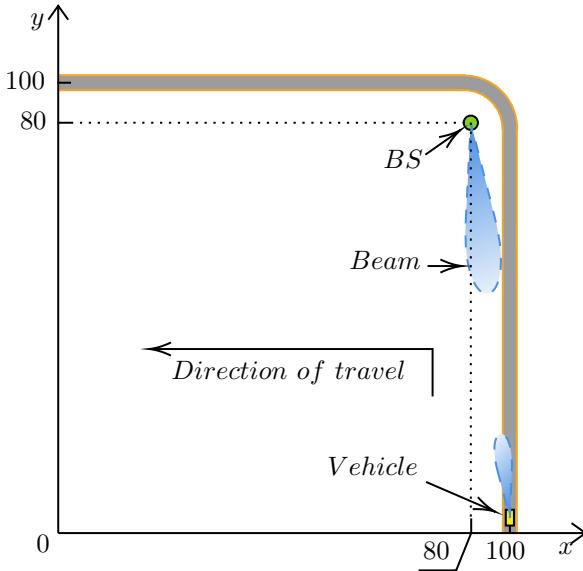


Fig. 3. Scene of mmWave vehicular networks for data generation

$\sqrt{\Gamma_{xn}^2 + \Gamma_{yn}^2}$ , and  $c$  is light speed. Now the channel matrix at each time step can be obtained from (2), (3), (4) and (7).

#### B. Beam Search

In this paper, Downlink (DL) configuration scheme is considered for beam management. The beam management procedure is as follows. The reference signals are sent to the vehicle from the BS, the vehicle measures and determines the optimal beam for communication, then transmits beam reporting back to the BS.

AoAs and AoDs measured from exhaustive beam search scheme is used to train the network, rather than using the generated CSI directly. Although AoAs and AoDs are in range  $[0, 2\pi]$ , only the angles in range  $[0, \pi]$  are considered, since the cosine function of the angles is symmetric around  $\pi$ . To reduce the overhead of beamforming, a discrete resolution  $b$ -bit codebook is adopted to perform beam search, where  $b$  is the number of bits for quantized phase shifters. For a  $b$ -bit codebook, it is consisted of  $2^b$  beam patterns. In this paper,  $b$  is assumed to be  $2\log_2 M$  for a  $M$ -element antenna terminal. For a terminal with  $M$ -element antenna, the candidate beam angles vector  $\mathbf{b}_M$  is given by:

$$\mathbf{b}_M = \left[ \frac{1}{2M}\pi, \frac{2}{2M}\pi, \dots, \pi \right] \quad (8)$$

So, there are  $2M_r$  beam patterns for the vehicle, and  $2M_t$  beam patterns for the BS. Note that another codebook which achieves uniform maximum gain in all direction is not used for this solution, because its low spatial resolution near 0 and  $\pi$  resulting in network underfitting. The beam codebooks of the vehicle and the BS are:

$$\begin{aligned} \mathbf{W} &= [\mathbf{a}_r([\mathbf{b}_{M_r}]_1), \mathbf{a}_r([\mathbf{b}_{M_r}]_2), \dots, \mathbf{a}_r([\mathbf{b}_{M_r}]_{2M_r})] \\ \mathbf{F} &= [\mathbf{a}_t([\mathbf{b}_{M_t}]_1), \mathbf{a}_t([\mathbf{b}_{M_t}]_2), \dots, \mathbf{a}_t([\mathbf{b}_{M_t}]_{2M_t})] \end{aligned} \quad (9)$$

where  $\mathbf{W} \in \mathbb{C}^{M_r \times 2M_r}$  is the codebook matrix of vehicle, and  $\mathbf{F} \in \mathbb{C}^{M_t \times 2M_t}$  is the codebook matrix of BS. Each column of the codebook matrix represents a beam pattern, each entry in the column is phase rotation for corresponding antenna element to generate directional beam.

Assume  $x_n = 1$  is sent as reference signal, the observation matrix comprises all measurement of exhaustive search now can be given by:

$$\mathbf{Y}_n = \mathbf{W}^H \mathbf{H}_n \mathbf{F} + \mathbf{V}_n \quad (10)$$

where  $\mathbf{V}_n \in \mathbb{C}^{2M_r \times 2M_t}$  is Gaussian noise matrix composed of independent identically distribution (i.i.d.) elements, which are the noise part of (5).  $\mathbf{Y}_n \in \mathbb{C}^{2M_r \times 2M_t}$ ,  $[\mathbf{Y}_n]_{i,j}$  is the received signal of using  $[\mathbf{W}]_{:,i}$  as combining vector and  $[\mathbf{F}]_{:,j}$  as beamforming vector.

For each time step, both of the vehicle and the BS use their beam codebook to perform exhaustive beam search to find the optimal beam pair. The indexs of optimal beam pair can be obtained by solving a optimization problems as following:

$$\begin{aligned} (i_n^*, j_n^*) &= \underset{i_n, j_n}{\operatorname{argmax}} \|[\mathbf{Y}_n]_{i_n, j_n}\|_2 \\ \text{s.t. } i_n &\in [1, 2M_r], \\ j_n &\in [1, 2M_t] \end{aligned} \quad (11)$$

A simple nested loop algorithm can be exploited to solve this problem. Then, at  $n$ th time step, the optimal combining vector  $\mathbf{w}_n^*$  and the optimal beamforming vector  $\mathbf{f}_n^*$  are  $\mathbf{w}_n^* = [\mathbf{W}]_{:,i_n^*}$  and  $\mathbf{f}_n^* = [\mathbf{F}]_{:,j_n^*}$ , respectively. And the estimated AoA  $\phi_n^e$  and AoD  $\theta_n^e$  used to training network are given by:

$$\begin{cases} \phi_n^e = \frac{i_n^*\pi}{2M_r} \\ \theta_n^e = \frac{j_n^*\pi}{2M_t} \end{cases} \quad (12)$$

The advantage of training with radian instead of the codeword indices is that it allows the proposed beam tracking solution to be generalized under terminals with different antenna configuration and beamforming techniques, without additional computation and transmitted data.

Note that due to the limited resolution of the beam codebook and the noise received at the vehicle, there will be some errors for the channel estimation, which is reflected in fluctuations in the estimated time series AoAs and AoDs. To reduce the fluctuations, a first order Savitzky-Golay filter [13] with frame length 11 was used to smooth the estimated AoAs and AoDs for training the LSTM network better.

#### V. SIMULATION RESULTS

In this section, the performances of proposed LSTM-based beam tracking solution is presented by analyzing the SNR of received signal, number of measurements and outage probability. The outage probability for a simulation is given by:

$$P_o = \frac{\sum_{i=1}^N \mathbf{I}(SNR_i \leq SNR_T)}{N} \quad (13)$$

where  $SNR_i$  is the SNR of received signal for  $i$ th sample and  $\mathbf{I}$  is an indicator function.

TABLE I  
SIMULATION PARAMETERS

Parameters	Values
Carrier frequency $f_c$	$28 \times 10^9$ Hz
Coordinates of BS $(x_t, y_t)$	(80, 80)
Number of antenna elements for BS $M_t$	16
Number of antenna elements for Vehicle $M_r$	4
Number of subpaths $L$	20
Expectation of BS rms angular spread $\lambda_t$	10.2°
Expectation of Vehicle rms angular spread $\lambda_r$	15.5°
Path loss standard deviation $\sigma_p$	5.8 dB
Noise variance of received signal $\sigma_v^2$	-140 dBW
Sequence length of LSTM $n_{LSTM}$	5
Threshold of SNR $SNR_T$	5 dB

In this paper, the sample rate of series channel data is 20 Hz (20 samples per second) and there are total 613 samples for training and 401 samples for testing. Table I show the common simulation parameters. For each simulation, parameters default to the values in Table I, except for those mentioned specifically.

Fig. 4 show the performances of algorithms versus noise variance of received signal. It illustrates that for both exhaustive search and LSTM-based scheme, the average SNR of received signal is decreasing and the probability of outages is increasing with the increasing of  $\sigma_v^2$ . The difference in performance between the LSTM-based scheme and the exhaustive search scheme is small, up to about 6 dB. When  $\sigma_v^2 = -140$  dBW, the average SNR of received signal in exhaustive search is 54.67 dB and in LSTM-based scheme is 52.32 dB, there is only 2.35 dB loss in SNR, and about 4.3% loss in spectral efficiency.

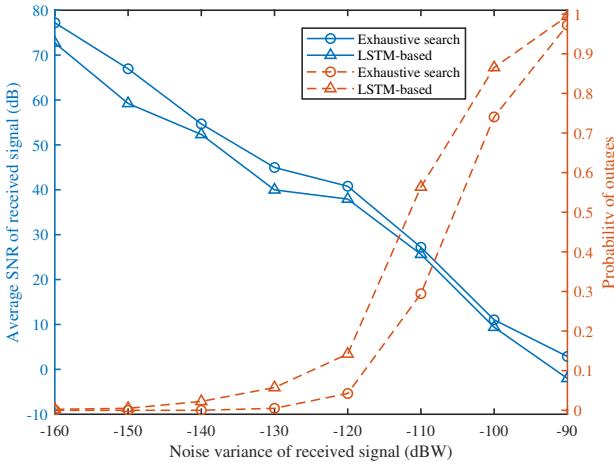


Fig. 4. Average SNR and outages probability versus  $\sigma_v^2$

Since the SNR performance of exhaustive search is considered to be the best, there is a high probability that if there are outages when using exhaustive search, there will also be outages when using other schemes at the same time. Fig. 4

show that there is not any outages when the noise variance of received signal is lower than -140 dBW. Therefore, the default noise variance of received signal is defined as -140 dBW for following simulations to ensure that there are not any outages when using exhaustive search to acquire AoAs and AoDs.

Table II show comparison of the number of measurements in different schemes. Compared with the exhaustive search scheme, proposed LSTM-based scheme can reduce about 94.3% measurement overhead. Considering the significant reduction in measurement overhead, the 4.3% loss of spectral efficiency in Fig. 4 seems to be acceptable. Note that the proposed LSTM-based scheme obtain initial beam angles for prediction by exhaustive search method in this paper, and it is obvious that the measurement overhead can be further reduced by utilizing a less complex beam search method (e.g. hierarchical method).

TABLE II  
COMPARISONS OF THE NUMBER OF MEASUREMENTS

Schemes	Complexity	Number of Measurements
LSTM-based	$N_s + N_f M_t M_r$ <sup>1</sup>	2240
Exhaustive search	$N M_t M_r$	39552

<sup>1</sup>  $N_s$  is the number of successful prediction samples and  $N_f = N - N_s$ .

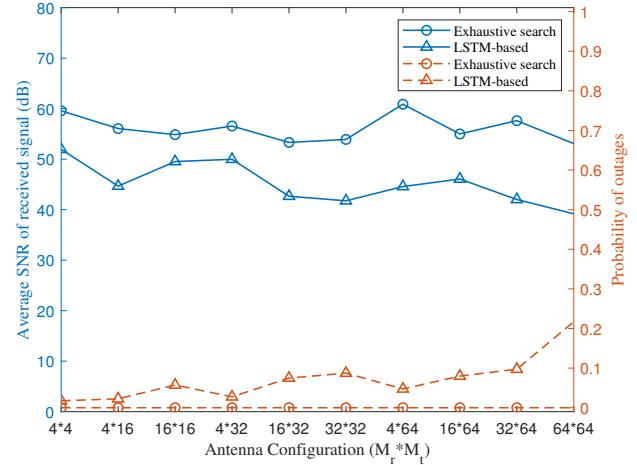


Fig. 5. Average SNR and outages probability versus  $M_t \times M_r$

## VI. CONCLUSIONS

## ACKNOWLEDGMENT

Thanks, guy.

## REFERENCES

- [1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s11276-015-0942-z>

- [2] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, “28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city,” *IEEE International Conference on Communications*, no. Icc, pp. 5163–5167, 2013.
- [3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.
- [4] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A tutorial on beam management for 3GPP NR at mmwave frequencies,” *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 173–196, 2019.
- [5] S. Noh, M. D. Zoltowski, and D. J. Love, “Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, 2017.
- [6] C. Zhang, D. Guo, and P. Fan, “Tracking angles of departure and arrival in a mobile millimeter wave channel,” *2016 IEEE International Conference on Communications, ICC 2016*, 2016.
- [7] V. Va, H. Vikalo, and R. W. Heath, “Beam tracking for mobile millimeter wave communication systems,” *2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2016 - Proceedings*, no. 1, pp. 743–747, 2017.
- [8] J. Lee, G. T. Gil, and Y. H. Lee, “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications,” *2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 3326–3331, 2014.
- [9] J. C. Aviles and A. Kouki, “Position-aided mm-wave beam training under NLOS conditions,” *IEEE Access*, vol. 4, pp. 8703–8714, 2016.
- [10] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.
- [11] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, 1997.
- [12] D. Burghal, N. A. Abbasi, and A. F. Molisch, “A Machine Learning Solution for Beam Tracking in mmWave Systems,” *Conference Record - Asilomar Conference on Signals, Systems and Computers*, vol. 2019-Novem, pp. 173–177, 2019.
- [13] A. Savitzky and M. J. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” *Analytical Chemistry*, 1964.