

Wireless Communications and Mobile Computing

# Channel Characterization and Modeling for 5G and Future Wireless System Based on Big Data

Lead Guest Editor: Liu Liu

Guest Editors: Jianhua Zhang, Sana Salous, and Tommi Jamsa





---

# **Channel Characterization and Modeling for 5G and Future Wireless System Based on Big Data**


Wireless Communications and Mobile Computing

---

## **Channel Characterization and Modeling for 5G and Future Wireless System Based on Big Data**

Lead Guest Editor: Liu Liu

Guest Editors: Jianhua Zhang, Sana Salous, and Tommi Jamsa



---

Copyright © 2018 Hindawi. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

- Javier Aguiar, Spain  
Ghufran Ahmed, Pakistan  
Wessam Ajib, Canada  
Muhammad Alam, China  
Eva Antonino-Daviu, Spain  
Shlomi Arnon, Israel  
Leyre Azpilicueta, Mexico  
Paolo Barsocchi, Italy  
Alessandro Bazzi, Italy  
Zdenek Becvar, Czech Republic  
Francesco Benedetto, Italy  
Olivier Berder, France  
Ana M. Bernardos, Spain  
Mauro Biagi, Italy  
Dario Bruneo, Italy  
Jun Cai, Canada  
Zhipeng Cai, USA  
Claudia Campolo, Italy  
Gerardo Canfora, Italy  
Rolando Carrasco, UK  
Vicente Casares-Giner, Spain  
Luis Castedo, Spain  
Ioannis Chatzigiannakis, Italy  
Lin Chen, France  
Yu Chen, USA  
Hui Cheng, UK  
Ernestina Cianca, Italy  
Riccardo Colella, Italy  
Mario Collotta, Italy  
Massimo Condoluci, Sweden  
Daniel G. Costa, Brazil  
Bernard Cousin, France  
Telmo Reis Cunha, Portugal  
Igor Curcio, Finland  
Laurie Cuthbert, Macau  
Donatella Darsena, Italy  
Pham Tien Dat, Japan  
André de Almeida, Brazil  
Antonio De Domenico, France  
Antonio de la Oliva, Spain  
Gianluca De Marco, Italy  
Luca De Nardis, Italy  
Liang Dong, USA  
Mohammed El-Hajjar, UK  
Oscar Esparza, Spain  
Maria Fazio, Italy  
Mauro Femminella, Italy  
Manuel Fernandez-Veiga, Spain  
Gianluigi Ferrari, Italy  
Ilario Filippini, Italy  
Jesus Fontecha, Spain  
Luca Foschini, Italy  
A. G. Fragkiadakis, Greece  
Sabrina Gaito, Italy  
Óscar García, Spain  
Manuel García Sánchez, Spain  
L. J. García Villalba, Spain  
José A. García-Naya, Spain  
Miguel Garcia-Pineda, Spain  
A.-J. García-Sánchez, Spain  
Piedad Garrido, Spain  
Vincent Gauthier, France  
Carlo Giannelli, Italy  
Carles Gomez, Spain  
Juan A. Gomez-Pulido, Spain  
Ke Guan, China  
Antonio Guerrieri, Italy  
Daojing He, China  
Paul Honeine, France  
Sergio Ilarri, Spain  
Antonio Jara, Switzerland  
Xiaohong Jiang, Japan  
Minho Jo, Republic of Korea  
Shigeru Kashihara, Japan  
Dimitrios Katsaros, Greece  
Minseok Kim, Japan  
Mario Kolberg, UK  
Nikos Komninos, UK  
Juan A. L. Riquelme, Spain  
Pavlos I. Lazaridis, UK  
Tuan Anh Le, UK  
Xianfu Lei, China  
Hoa Le-Minh, UK  
Jaime Lloret, Spain  
Miguel López-Benítez, UK  
Martín López-Nores, Spain  
Javier D. S. Lorente, Spain  
Tony T. Luo, Singapore  
Maode Ma, Singapore  
Imadeldin Mahgoub, USA  
Pietro Manzoni, Spain  
Álvaro Marco, Spain  
Gustavo Marfia, Italy  
Francisco J. Martinez, Spain  
Davide Mattera, Italy  
Michael McGuire, Canada  
Nathalie Mitton, France  
Klaus Moessner, UK  
Antonella Molinaro, Italy  
Simone Morosi, Italy  
Kumudu S. Munasinghe, Australia  
Enrico Natalizio, France  
Keivan Navaie, UK  
Thomas Newe, Ireland  
Wing Kwan Ng, Australia  
Tuan M. Nguyen, Vietnam  
Petros Nicopolitidis, Greece  
Giovanni Pau, Italy  
Rafael Pérez-Jiménez, Spain  
Matteo Petracca, Italy  
Nada Y. Philip, UK  
Marco Picone, Italy  
Daniele Pinchera, Italy  
Giuseppe Piro, Italy  
Vicent Pla, Spain  
Javier Prieto, Spain  
Rüdiger C. Prys, Germany  
Sujan Rajbhandari, UK  
Rajib Rana, Australia  
Luca Reggiani, Italy  
Daniel G. Reina, Spain  
Abusayeed Saifullah, USA  
Jose Santa, Spain  
Stefano Savazzi, Italy  
Hans Schotten, Germany  
Patrick Seeling, USA  
Muhammad Z. Shakir, UK  
Mohammad Shojafar, Italy  
Giovanni Stea, Italy  
Enrique Stevens-Navarro, Mexico  
Zhou Su, Japan  
Luis Suarez, Russia



---




Ville Syrjälä, Finland  
Hwee Pink Tan, Singapore  
Pierre-Martin Tardif, Canada  
Mauro Tortonesi, Italy  
Federico Tramarin, Italy

Reza Monir Vaghefi, USA  
Juan F. Valenzuela-Valdés, Spain  
Aline C. Viana, France  
Enrico M. Vitucci, Italy  
Honggang Wang, USA



Jie Yang, USA  
Sherali Zeadally, USA  
Jie Zhang, UK  
Meiling Zhu, UK

# Contents

## **Channel Characterization and Modeling for 5G and Future Wireless System Based on Big Data**

Liu Liu , Jianhua Zhang , Sana Salous , and Tommi Jamsa  
Editorial (2 pages), Article ID 1046836, Volume 2018 (2018)

## **Predicting Wireless MmWave Massive MIMO Channel Characteristics Using Machine Learning Algorithms**

Lu Bai, Cheng-Xiang Wang , Jie Huang, Qian Xu, Yuqian Yang, George Goussetis, Jian Sun ,  
and Wensheng Zhang  
Research Article (12 pages), Article ID 9783863, Volume 2018 (2018)




## **A Request-Based Handover Strategy Using NDN for 5G**

Fan Jia  and Xiaolin Zheng  
Research Article (9 pages), Article ID 4513070, Volume 2018 (2018)

## **Channel Characteristics of Rail Traffic Tunnel Scenarios Based on Ray-Tracing Simulator**

Jinmeng Zhao , Lei Xiong , Danping He, and Jiadong Du  
Research Article (9 pages), Article ID 9284639, Volume 2018 (2018)




## **A Survey on Machine Learning-Based Mobile Big Data Analysis: Challenges and Applications**

Jiyang Xie , Zeyu Song, Yupeng Li, Yanting Zhang, Hong Yu, Jinnan Zhan, Zhanyu Ma ,  
Yuanyuan Qiao, Jianhua Zhang , and Jun Guo  
Review Article (19 pages), Article ID 8738613, Volume 2018 (2018)


## **Analysis of Nonstationary Characteristics for High-Speed Railway Scenarios**

Tao Zhou , Cheng Tao , and Kai Liu  
Research Article (7 pages), Article ID 1729121, Volume 2018 (2018)



## **Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments**

Yan Zhang , Jinxiao Wen , Guanshu Yang , Zunwen He , and Xinran Luo  
Research Article (9 pages), Article ID 8489326, Volume 2018 (2018)

## **A Full Duplex D2D Clustering Resource Allocation Scheme Based on a K-Means Algorithm**

Xu Huang , Mengjia Zeng, Jing Fan, Xiangxiang Fan, and Xuefeng Tang  
Research Article (8 pages), Article ID 1843083, Volume 2018 (2018)

## **MU-MIMO Downlink Capacity Analysis and Optimum Code Weight Vector Design for 5G Big Data Massive Antenna Millimeter Wave Communication**

Adam Mohamed Ahmed Abdo, Xiongwen Zhao , Rui Zhang, Zhenyu Zhou , Jianhua Zhang ,  
Yu Zhang, and Imran Memon   
Research Article (12 pages), Article ID 7138232, Volume 2018 (2018)

## Editorial

# Channel Characterization and Modeling for 5G and Future Wireless System Based on Big Data

Liu Liu <sup>1</sup>, Jianhua Zhang <sup>2</sup>, Sana Salous <sup>3</sup> and Tommi Jamsa<sup>4</sup>

<sup>1</sup>*Institute of Broadband Wireless Mobile Communications, Beijing Jiaotong University, Beijing 100044, China*

<sup>2</sup>*Beijing University of Post and Communications, Beijing 100876, China*

<sup>3</sup>*Durham University, UK*

<sup>4</sup>*Huawei Technologies Sweden AB, Sweden*

Correspondence should be addressed to Liu Liu; [liuliu@bjtu.edu.cn](mailto:liuliu@bjtu.edu.cn)

Received 12 November 2018; Accepted 13 November 2018; Published 20 December 2018

Copyright © 2018 Liu Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Big Data is attracting more and more attention in various fields nowadays, changing the way we live greatly, and proving to be one of the hottest research topics. The base of big data lies in the huge volume and scale of data, hence making it better for us to make decisions if we can extract more useful information from the huge amount of data.

The tremendous development of 5G, especially with the combination of Internet and Internet-of-Thing (IoT), boosts the amount and type of wireless data dramatically. Compared with the 4G, the bandwidth (over hundreds of MHz), central frequency (centimeter and millimeter wave band), amount of antennas (3-dimensional and massive MIMO), number of sensors (IoT), and application scenarios expand enormously, leading to the rapid growth in the amount of data. To be more specific, as for the IoT, the density of sensors is increasing rapidly in 5G systems and sensors are distributed everywhere in a typical scenario which is leading to the exponential growth in wireless links, making it difficult for conventional channel characterization methods to handle it.

It is crucial to characterize the wireless channel accurately to guarantee the demand of 5G system. The preliminary performance of big data provides a promising prospect in various fields. By using big data, we can mine the characteristic of wireless channel deeply and parameterize the channel more precisely, which has never been studied. In this special issue, we cordially invite some researchers to contribute papers that discuss the channel modeling and simulation for 5G systems using big data and machine learning, as well as other artificial

intelligence theories. And this special issue provides the state-of-art research in this field.

The paper “A Survey on Machine Learning-Based Mobile Big Data Analysis: Challenges and Applications” investigates how to identify the requirement and the development of machine learning-based mobile big data (MBD) analysis through discussing the insights of challenges in the MBD and reviewing state-of-the-art application of data analysis in the area of MBD. The paper introduced the development of MBD and reviewed the frequently applied data analysis methods. Three typical applications of MBD analysis, namely, wireless channel modeling, human online and offline behavior analysis, and speech recognition and verification in the Internet of vehicles, are introduced, respectively. Finally the paper proposed five main challenges including large-scale and high-speed-M-Internet, overfitting and underfitting problems, generalization problem, cross-modal learning, and extended channel dimensions.

The paper “A Full Duplex D2D Clustering Resource Allocation Scheme Based on a K-means Algorithm” talks about the Device-to-Device (D2D) technology problem of resource allocation and control in a single-cell scene. The concept of a restricted D2D communication area and a restricted D2D user-reusage area is put forward to reduce the complexity and interference intensity of resource allocation. And under the premise of satisfying the QoS (Quality of Service) demands of every system user, the resource allocation algorithm is improved, the optimal allocation of resources is carried



out, and the algorithm's processes are given in detail. The simulation result shows a good performance in eliminating interference and improving the spectrum efficiency and the system fairness.

As the unmanned aerial vehicle (UAV) plays an important role in many applications due to its high flexibility and low cost, more and more research is implemented to explore the channel characterization of Air-to-Air (AA) scenario. The paper titled "Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments" proposes path loss models for the UAV AA scenario based on machine learning. A ray-tracing software has been utilized to generate the data for an urban AA scenario. And the models have been learned by two machine learning algorithms, Random Forest and k-Nearest Neighbor (KNN). The test data have been used to evaluate the accuracy performance of these machine-learning based models and two empirical models, SUI model and COST231-W-I model. It has been demonstrated that machine learning provides a flexible modeling approach based on the training data for such complex environment, and Random Forest has the best prediction performance. In addition, the importance of five input features for the path loss in the AA scenario is analyzed. Results have confirmed that the path visibility is the dominant factor. Propagation distance and elevation angle have also shown great influences.

The paper "Predicting Wireless MmWave Massive MIMO Channel Characteristics Using Machine Learning Algorithms" deals with the topic of the prediction of channel statistical characteristics based on the well-known machine learning algorithm, convolutional neural network (CNN), for three dimensional millimeter wave massive multiple-input multiple output indoor channels. The ray tracing software, Wireless InSite, is used in this paper to build the measurement datasets. A complete description of the process for creating and training a CNN-based model is presented with special emphasis on the training process. The results show good fittings between the predicted channel statistical characteristics and the real channel statistical characteristics.

As we know, tunnel scenario is a major and significant communication scenario attracting more and more attention with the rapid development of high-speed railway. The paper titled "Channel Characteristics of Rail Traffic Tunnel Scenarios Based on Ray-Tracing Simulator" provides a good understanding of channel characterization of tunnel scenarios based on ray-tracing. The channel characteristics in different carrier frequencies and tunnel cross sections were analyzed and some important conclusions are drawn: the channel experiences a severe and stable fading in long arched tunnels compared to other tunnel scenarios. The presence of the vehicle body introduces additional 35 dB of the path loss, which leads to the fluctuation and instability of the channel. K-factor changes severely when the distance between Tx and Rx is smaller than 100m and then decrease smoothly in far region.

The paper "Analysis of Nonstationary Characteristics for High-Speed Railway Scenarios" presents the analysis of nonstationary characteristics in typical high-speed railway (HSR) scenarios involving rural, station, and suburban,

according to passive long-term evolution (LTE) based channel measurements. Additionally, a four-state Markov chain model (MCM) is established to characterize the birth-death (B-D) process of multipath components (MPCs), and the corresponding state transition probability matrix and steady-state probability are provided. The results provide helpful information for nonstationary channel modeling of HSR communication systems.

The paper titled "A Request-Based Handover Strategy Using NDN for 5G" explores the small cell base stations (SBS) handover problem in ultra-dense networks (UDN) for 5G. Request-Based Handover Strategy (RBHS) is presented to improve the user experience in performance and obtain the optimal allocation of resources, and a caching mechanism based on the users' requests is introduced for it. The proposed caching mechanism and access network selection mechanism were validated utilizing ndnSim. The simulation results demonstrate that the proposed strategy achieves around 30% higher cache hit rate and 20% more traffic reduction, compared with the access network selection base on SINR.

The paper "MU-MIMO Downlink Capacity Analysis and Optimum Code Weight Vector Design for 5G Big Data Massive Antenna Millimeter Wave Communication" discusses the design of the optimum beam-vector for each user to minimize interference from other users in multiuser multiple input multiple output (MU-MIMO) wireless communication system. The nonlinear sum-rate analysis using dirty paper coding (DPC) in Ricean fading channels based on signal-to-leakage plus noise ratio (SLNR) was undertaken. And a new method is proposed to find an optimum beam weight vector by exploring the power iteration method using eigenvector approximation. The proposed method achieves higher performance regarding mean achievable sum-rate capacity per user, proving to provide significant system capacity enhancement compared with the SVD method.

In conclusion, this special issue brings new insights into the intricate characterization and modeling of wireless channel based on big data. We hope that this information will be helpful to the development of 5G and provide some novel methods to resolve some wireless channel problems.

## Conflicts of Interest

The editors declare that they have no conflicts of interest regarding the publication of this Special Issue.

*Liu Liu  
Jianhua Zhang  
Sana Salous  
Tommi Jamsa*

## Research Article

# Predicting Wireless MmWave Massive MIMO Channel Characteristics Using Machine Learning Algorithms

Lu Bai,<sup>1</sup> Cheng-Xiang Wang ,<sup>2</sup> Jie Huang,<sup>1</sup> Qian Xu,<sup>3</sup> Yuqian Yang,<sup>1</sup> George Goussetis,<sup>2</sup> Jian Sun ,<sup>1</sup> and Wensheng Zhang<sup>1</sup>

<sup>1</sup>Shandong Provincial Key Lab of Wireless Communication Technologies, School of Information Science and Engineering, Shandong University, Qingdao, Shandong 266237, China

<sup>2</sup>Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

<sup>3</sup>School of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

Correspondence should be addressed to Cheng-Xiang Wang; [cheng-xiang.wang@hw.ac.uk](mailto:cheng-xiang.wang@hw.ac.uk)

Received 31 March 2018; Accepted 16 July 2018; Published 23 August 2018

Academic Editor: Tommi Jamsa

Copyright © 2018 Lu Bai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a procedure of predicting channel characteristics based on a well-known machine learning (ML) algorithm and convolutional neural network (CNN), for three-dimensional (3D) millimetre wave (mmWave) massive multiple-input multiple-output (MIMO) indoor channels. The channel parameters, such as amplitude, delay, azimuth angle of departure (AAoD), elevation angle of departure (EAoD), azimuth angle of arrival (AAoA), and elevation angle of arrival (EAoA), are generated by a ray tracing software. After the data preprocessing, we can obtain the channel statistical characteristics (including expectations and spreads of the above-mentioned parameters) to train the CNN. The channel statistical characteristics of any subchannels in a specified indoor scenario can be predicted when the location information of the transmitter (Tx) antenna and receiver (Rx) antenna is input into the CNN trained by limited data. The predicted channel statistical characteristics can well fit the real channel statistical characteristics. The probability density functions (PDFs) of error square and root mean square errors (RMSEs) of channel statistical characteristics are also analyzed.

## 1. Introduction

The fifth generation (5G) wireless communication networks have lots of novel requirements, such as the 1000 times the system capacity with respect to the fourth generation (4G) networks, wide frequency range (covering millimetre wave (mmWave) bands, e.g., 450 MHz–100 GHz), increased data rate, reduced latency, energy, and cost [1–6]. To satisfy the above-mentioned requirements, several advanced technologies, such as mmWave and massive multiple-input multiple-output (MIMO), have been proposed and brought new challenges on channel modeling. Since the performance bound of wireless communication systems is determined by channel characteristics [7], an accurate channel model plays an important role in designing, evaluating, and developing wireless communication systems. The 5G wireless communication channel models, such as mobile and

wireless communications Enablers for the Twenty-twenty Information Society (METIS) channel model [8], Millimetre-Wave Evolution for Backhaul and Access (MiWEBA) channel model [9], ITU-R IMT-2020 channel model [10], COST 2100 channel model [11, 12], IEEE 802.11 ay channel models [13], millimetre-wave based mobile radio access network for fifth generation integrated communications (mmMAGIC) channel model [14], quasi deterministic radio channel generator user manual and documentation (QuaDRiGa) channel model [15, 16], and a general three-dimensional (3D) nonstationary 5G channel model [17], can be classified as deterministic and stochastic channel models. As the most important technologies of 5G wireless communication networks, massive MIMO and mmWave have also attracted great attentions. According to the massive MIMO and mmWave indoor channel measurement in [18], authors in [19] did the massive MIMO and mmWave channel parameter estimation.

Authors in [20] summarized recent massive MIMO channel measurements and models. The above-mentioned models are complex and hard to use. So a revolutionary channel model is necessary.

The explosive increase of frequencies/bandwidths, antennas, and new services/scenarios will generate massive data and bring the research of 5G wireless communications to the era of artificial intelligence (AI) [21, 22]. Machine learning (ML), as an important branch of AI, has received extensive attentions due to its capability of digging the valuable and hidden rules from enormous unknown channel information. It can take advantages of both the low complexity of stochastic channel models and the accuracy of deterministic channel models. As a conventional ML algorithm, convolutional neural network (CNN) exhibits excellent performance on compressing and processing redundant channel information [23].

Until now, there are two kinds of applications of AI to 5G wireless communication channels. One is measurement data preprocessing based on statistical learning methods, e.g., clustering algorithms. The Kernel-Power-Density algorithm proposed in [24] used the kernel density and only considered the neighboring points when computing the density. Authors in [25] proposed a novel clustering framework based on Kernel-Power-Density algorithm and took elevation angles into consideration. The Kuhn–Munkres algorithm was proposed to solve the tracking problem in [26]. The Kalman filter in [27] was used to track the clusters and to predict the cluster positions. Furthermore, several other algorithms were used for clusters identification in measurement data preprocessing, such as KPowerMeans algorithm [28] and hierarchical tree [29]. The above-mentioned clustering algorithms play a significant role in conventional cluster-based stochastic channel models, such as COST 2100 channel model and WINNER channel models, but it cannot predict the channel characteristics. The other one is to predict the channel characteristics based on ML algorithms which can dig the mapping relationship between physical environment information and the channel characteristics. The function between frequency, distance, and path loss (PL) was modeled by two types of artificial neural networks (ANNs), i.e., multilayer perceptron (MLP) and radial basis function (RBF) [30–34]. In [35], PL was also modeled as a mapping relationship between delay and the atmosphere by MLP. Authors in [36] and [37] modeled Doppler frequency shift by RBF and MLP, respectively. The mapping relationship between channel characteristics and geographical location was modeled by a feed-forward network (FFN) in [38] and a DeepFi architecture in [39]. In-vehicle wireless channels at 60 GHz were modeled by a FFN and a RBF network [40, 41]. Author in [42] proposed a three-layer structure based on ML –“wave, cluster-nuclei, and channel”. Most of the existing research works can only obtain the mapping relationship between a single channel characteristic and physical channel environment information but cannot predict comprehensive channel characteristics. At the same time, the channel characteristics of any subchannels in a specified scenario cannot be predicted until now, while they play an important role on channel estimation and communication quality. CNN can compress and process

redundant channel information well, but it has not been applied to channel characteristics prediction.

In this paper, we propose an AI enabled procedure to predict channel statistical characteristics based on CNN to obtain the mapping relationship between the location information of transmitter (Tx) and receiver (Rx) antennas and almost all the characteristics of amplitude, delay, and angles. The main **contributions** of this paper are summarized as follows:

- (1) A procedure of predicting channel statistical characteristics based on a specified CNN for 3D mmWave MIMO indoor channels is proposed in this paper. With the location information of Tx and Rx antennas, the CNN can predict eleven comprehensive channel statistical characteristics, including PL, delay spread (DS), delay mean (DM), azimuth angle mean of departure (AAMD), azimuth angle mean of arrival (AAMA), azimuth angle spread of departure (AASD), azimuth angle spread of arrival (AASA), elevation angle mean of departure (EAMD), elevation angle mean of arrival (EAMA), elevation angle spread of departure (EASD), and elevation angle spread of arrival (EASA).
- (2) This is the first time to compare five different wireless channel characteristic datasets, which are collected by different ways. By comparing their training results, we can obtain better rules of data generation and collection. Therefore, it has a profound guiding significance for data generation and collection.

We have organized the rest of the paper as follows. The AI enabled procedure to predict channel statistical characteristics is shown in Section 2. In Section 3, we describe the two indoor scenarios of data collection and the principle of data preprocessing. The five datasets are also given in this section. The proposed CNN is shown in Section 4. In Section 5, we discuss and analyze the results. Conclusions and future work are given in Section 6.

## 2. System Model

The flowchart of AI enabled procedure to predict channel statistical characteristics is shown in Figure 1. Firstly, we set up the indoor scenario and obtain simulated channel information. At this time, we construct two 3D indoor scenarios by setting the sizes and materials of rooms and objects in a ray tracing software. Then we can obtain the multipath component parameters (amplitude, delay, azimuth angle of departure (AAoD), elevation angle of departure (EAoD), azimuth angle of arrival (AAoA), and elevation angle of arrival (EAoA)). We do the data preprocessing to obtain the channel statistical characteristics (PL, DM, DS, AAMA, AASA, AAMD, AASD, EAMA, EASA, EAMD, and EASD) to build the dataset. The dataset of the specified indoor scenario is built to be separated as two sets by the proportion of 7:3 randomly. One is the train set, the other is the validation set. Samples in both the train set and the validation set have 3D coordinates of Tx and Rx as

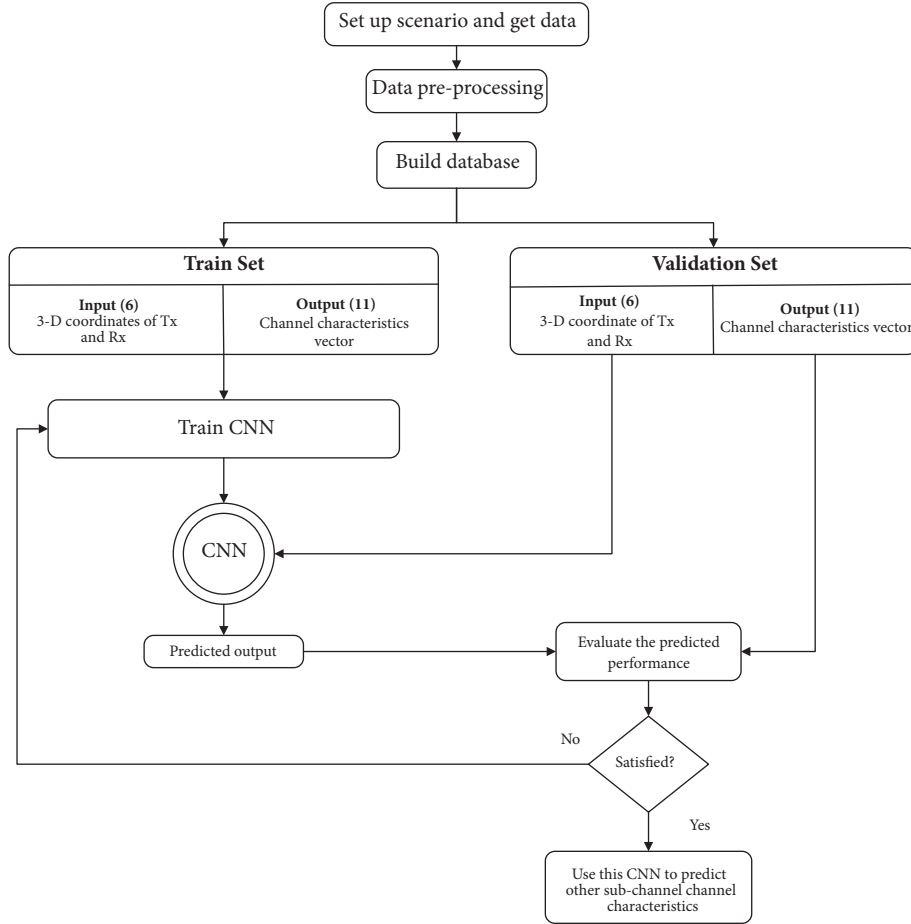


FIGURE 1: The flowchart of channel characteristic predicting procedure.

input vectors and channel statistical characteristics as output vectors. The train set is used to train the CNN. The input vectors of the validation set are put into the CNN to obtain the predicted output vectors. Whether train the CNN again or not is determined by comparing and analyzing the root mean square errors (RMSEs) and probability density functions (PDFs) between the predicted output vectors and output vectors of the validation set. More detailed information will be shown in the following sections.

### 3. Database Generation

The ray tracing software, Wireless InSite [43], is used to build the simulation datasets. Ray tracing is a classical deterministic method used for modeling radio propagations. It is based on the geometrical optic (GO) and uniform theory of diffraction (UTD). The interactions between rays and objects can be classified as reflection, transmission, scattering, and diffraction. By tracing paths in a specified scenario we build in the simulator, all the possible rays can be obtained and we can get the parameter vector  $\omega_{ij,s}$  of the  $s$ -th ( $s = 1, 2, \dots, S$ ,  $S = 250$ ) multipath between the  $i$ -th Tx antenna and  $j$ -th Rx antenna; i.e.,

$$\omega_{ij,s} = [\alpha_{ij,s}, \tau_{ij,s}, \theta_{ij,s}^T, \phi_{ij,s}^T, \theta_{ij,s}^R, \phi_{ij,s}^R] \quad (1)$$

where  $\alpha_{ij,s}$ ,  $\tau_{ij,s}$ ,  $\theta_{ij,s}^T$ ,  $\phi_{ij,s}^T$ ,  $\theta_{ij,s}^R$ , and  $\phi_{ij,s}^R$  are the amplitude, delay, AAoD, EAoD, AAoA, and EAoA of the  $s$ -th multipath between the  $i$ -th Tx antenna and  $j$ -th Rx antenna, respectively.

**3.1. The Descriptions of Data Generation.** To verify the general predicted capability of the CNN, we construct two indoor scenarios in Wireless InSite to collect multipath component parameters. One is a virtual classroom scenario shown in Figure 2; the other is a real lab scenario shown in Figure 3.

**3.1.1. The Virtual Classroom Scenario.** The virtual classroom environment is about  $8 \times 6 \times 3 \text{ m}^3$  with 12 desks whose size is about  $1.6 \times 0.4 \times 1.2 \text{ m}^3$ . The intervals of desks are 0.6 m along x-axis and 0.8 m along y-axis. Desks are made of wood. The ceiling is made of concrete. Both floor and walls are made of 3-layered dielectric in [44]. The layout of the classroom is shown in Figure 2(a). Figure 2(b) shows the 3D ray tracing scenario of the classroom constructed in Wireless InSite.

In this virtual scenario, at most 3 orders of reflection and 1 order of diffraction are simulated. The maximum number

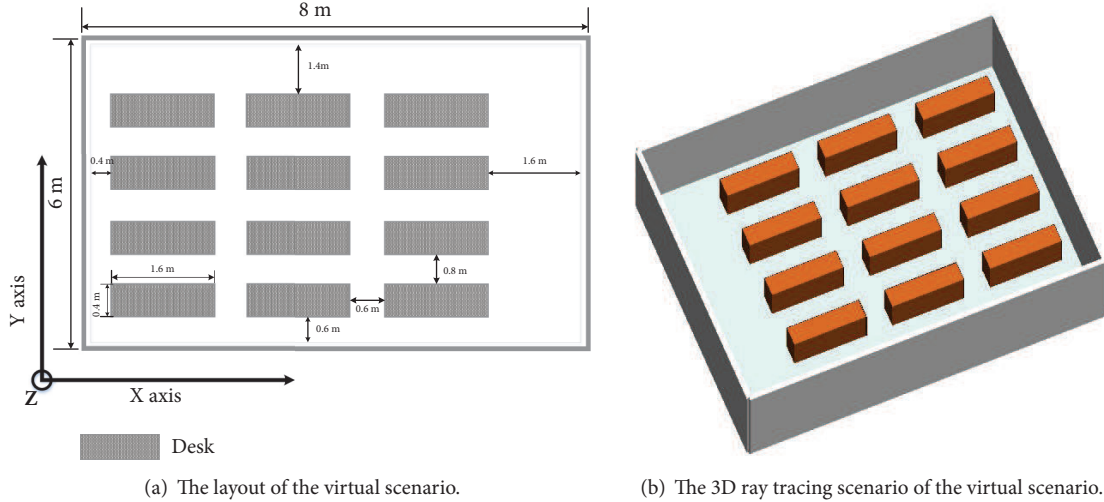


FIGURE 2: The environment information of the virtual scenario.

of paths allowed in simulation is 250. For the complexity of simulation, we do not take scattering which is caused by surface roughness into consideration while the power loss caused by rough surface of objects is calculated by the reflection coefficient multiplied with roughness coefficient. The carrier frequency and bandwidth are set to 60 GHz and 2 GHz, respectively. To evaluate the performance of data collection, two datasets are built. In the  $10 \times 100$  random dataset (10100R, R stands for random) of the virtual classroom scenario, we set up 10 Tx isotropic antennas with 0 dBi antenna gain in all directions and 100 Rx isotropic antennas at random location information to obtain multipath parameters of 1000 subchannels. 32 Tx isotropic antennas and 32 Rx isotropic antennas are randomly set up in  $32 \times 32$  random dataset (3232R, R stands for random) of the virtual classroom scenario to obtain channel parameters of 1024 subchannels. In indoor uplink communication scenarios, the Txes can be mobile phones, laptops, iPads, etc., while the Rxes are normally access points (APs) located on the ceilings. In a virtual classroom environment, we assume that the height of Txes is 1.5 m and the height of Rxes is 3 m.

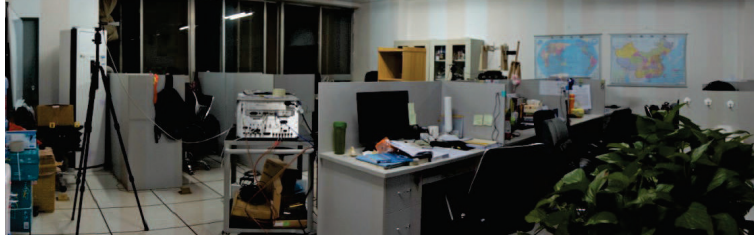
**3.1.2. The Lab Scenario.** The size of the lab scenario is approximately  $7.2 \times 7.2 \times 2.7 \text{ m}^3$ . Its floor and ceiling are made of concrete but decorated with antistatic-electricity board. It has four sides of walls. One is a partition wall which is made of plaster board, and other sides of wall are made of concrete. Two high built small windows are on the one side wall while a large window almost cover the other side of wall. The lab is furnished with multiple desks and chairs, and other office furniture such as computers, bookshelf, and electronic devices. Desks made of frosted surface chipboard are about 0.75 m height at desktop level but with two or three additional 0.45 m clapboards. Chairs are made of fabric cover and metal and plastic support. The electromagnetic properties of above-mentioned building and furniture materials were characterized by a lot of material measurements [45–47]. There is a small storage room in the corner of the lab. The

photo and layout of the lab are shown in Figures 3(a) and 3(b), respectively. Figure 3(c) shows the 3D ray tracing scenario of the lab constructed in Wireless InSite.

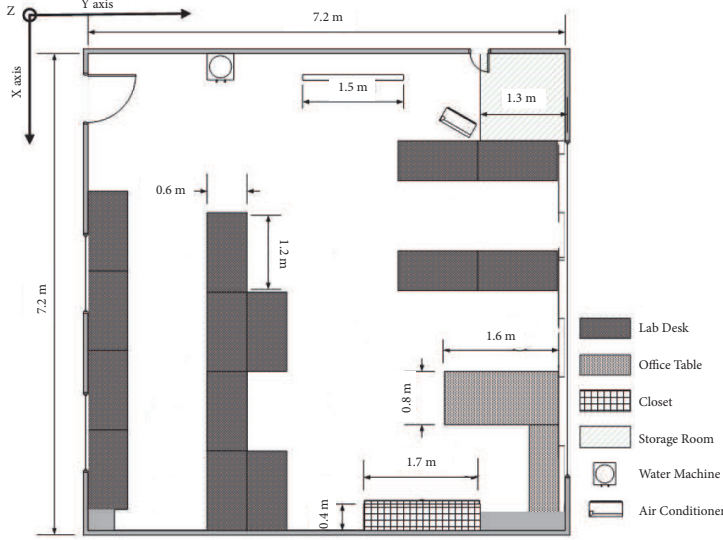
The lab supplies such as books and computer monitors on the desk are not modeled in the 3D ray tracing scenario, because their irregular shapes lead to significant increase of computational complexity and they are shadowed by higher clapboards on the desktop usually. Similarly, we also neglect chairs in the scenario modeling because they are about 0.8 m high which is lower than antennas and chairs were positioned nearby the desks. Since the lab scenario is more complex than the virtual classroom, at most 5 orders of reflection, 3 orders of transmission, and 1 order of diffraction are simulated in the ray tracing setup of this lab scenario. The maximum number of propagation paths is 250. The carrier frequency and bandwidth are set to 60 GHz and 2 GHz, respectively. To evaluate the performance of data collection, three datasets are built. In the  $30 \times 30$  random dataset (3030R, R stands for random) of the lab scenario, we set up 30 Tx isotropic antennas and 30 Rx isotropic antennas at random location information to obtain channel parameters of 900 subchannels. 30 Tx isotropic antennas and 30 Rx isotropic antennas are set up by grid at 1 m intervals in  $30 \times 30$  grid dataset (3030G, G stands for grid) of the lab scenario to obtain multipath component parameters of 900 subchannels. Similarly, 211 Tx isotropic antennas and 211 Rx isotropic antennas are set up by grid at 0.4 m intervals in the  $211 \times 211$  grid dataset (211211G, G stands for grid) of the lab scenario to get channel parameters of 44521 subchannels. In the lab environment, we assume that the height of Txes is 1.5 m and the height of Rxes is 2.7 m.

**3.2. Data Processing.** For conciseness, we use channel characteristic vector  $\hat{\omega}_{ij}$  instead of the parameter vectors of 250 multipaths; i.e.,

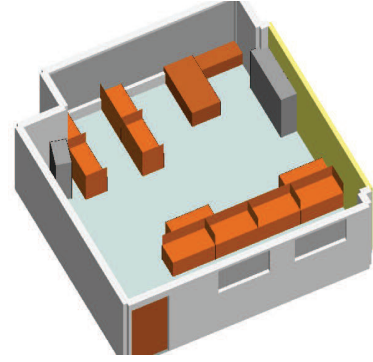
$$\hat{\omega}_{ij} = [\rho_{ij}, \mu_{ij}, \mu_{M,ij}, \vartheta_{ij}^T, \varphi_{ij}^T, \vartheta_{ij}^R, \varphi_{ij}^R, \vartheta_{M,ij}^T, \varphi_{M,ij}^T, \vartheta_{M,ij}^R, \varphi_{M,ij}^R] \quad (2)$$



(a) The photo of the lab.



(b) The layout of the lab.



(c) The 3D ray tracing scenario of the lab.

FIGURE 3: The environment information of the lab scenario.

where  $\rho_{ij}$  is PL of the subchannel from  $i$ -th Tx antenna to  $j$ -th Rx antenna,  $\mu_{ij}(\mu_{M,ij})$  is DS (DM) of the subchannel from  $i$ -th Tx antenna to  $j$ -th Rx antenna,  $\vartheta_{M,ij}^T, \vartheta_{M,ij}^R, \vartheta_{ij}^T$ , and  $\vartheta_{ij}^R$  are AAMD, AAMA, AASD, and AASA, respectively, and  $\varphi_{M,ij}^T, \varphi_{M,ij}^R, \varphi_{ij}^T$ , and  $\varphi_{ij}^R$  are EAMD, EAMA, EASD, and EASA, respectively. They can be further expressed as

$$\rho_{ij} = 10 \times \log_{10} \sum_{s=1}^S \alpha_{ij,s}^2 \quad (3)$$

$$\mu_{ij} = \sqrt{\frac{\sum_{s=1}^S \alpha_{ij,s}^2 \tau_{ij,s}^2}{\sum_{s=1}^S \alpha_{ij,s}^2} - \left( \frac{\sum_{s=1}^S \alpha_{ij,s}^2 \tau_{ij,s}}{\sum_{s=1}^S \alpha_{ij,s}^2} \right)^2} \quad (4)$$

$$\vartheta_{ij}^T = \sqrt{\frac{\sum_{s=1}^S \alpha_{ij,s}^2 \theta_{ij,s}^T{}^2}{\sum_{s=1}^S \alpha_{ij,s}^2} - \left( \frac{\sum_{s=1}^S \alpha_{ij,s}^2 \theta_{ij,s}^T}{\sum_{s=1}^S \alpha_{ij,s}^2} \right)^2} \quad (5)$$

$$\vartheta_{ij}^R = \sqrt{\frac{\sum_{s=1}^S \alpha_{ij,s}^2 \theta_{ij,s}^R{}^2}{\sum_{s=1}^S \alpha_{ij,s}^2} - \left( \frac{\sum_{s=1}^S \alpha_{ij,s}^2 \theta_{ij,s}^R}{\sum_{s=1}^S \alpha_{ij,s}^2} \right)^2} \quad (6)$$

$$\varphi_{ij}^T = \sqrt{\frac{\sum_{s=1}^S \alpha_{ij,s}^2 \phi_{ij,s}^T{}^2}{\sum_{s=1}^S \alpha_{ij,s}^2} - \left( \frac{\sum_{s=1}^S \alpha_{ij,s}^2 \phi_{ij,s}^T}{\sum_{s=1}^S \alpha_{ij,s}^2} \right)^2} \quad (7)$$

$$\varphi_{ij}^R = \sqrt{\frac{\sum_{s=1}^S \alpha_{ij,s}^2 \phi_{ij,s}^R{}^2}{\sum_{s=1}^S \alpha_{ij,s}^2} - \left( \frac{\sum_{s=1}^S \alpha_{ij,s}^2 \phi_{ij,s}^R}{\sum_{s=1}^S \alpha_{ij,s}^2} \right)^2} \quad (8)$$

$$\mu_{M,ij} = \frac{\sum_{s=1}^S \tau_{ij,s}}{S} \quad (9)$$

$$\vartheta_{M,ij}^T = \frac{\sum_{s=1}^S \theta_{ij,s}^T}{S} \quad (10)$$

$$\vartheta_{M,ij}^R = \frac{\sum_{s=1}^S \theta_{ij,s}^R}{S} \quad (11)$$

$$\varphi_{M,ij}^T = \frac{\sum_{s=1}^S \phi_{ij,s}^T}{S} \quad (12)$$

$$\varphi_{M,ij}^R = \frac{\sum_{s=1}^S \phi_{ij,s}^R}{S} \quad (13)$$

After data preprocessing, the details of datasets are shown in Table 1. We separated the total samples into train sets and validation sets by the proportion of 7:3 randomly.

#### 4. Architecture of the Proposed CNN for Channel Characteristics Prediction

The architecture of the CNN is presented in Figure 4. It includes two main stages: first stage configuring two convolutional layers and second stage configuring four dense layers which are also called fully connected layers. It requires a large number of iterations to obtain the neural network convergence to fit the thresholds nodes and the weights of connections for the least loss. The input vector  $X_{ij}$  is 3D coordinates of the  $i$ -th Tx antenna and the  $j$ -th Rx antenna. The output vector  $Y_{ij}$  is the channel characteristic vector of the subchannel between the  $i$ -th Tx and the  $j$ -th Rx  $\omega_{ij}$ . They can be expressed as

$$X_{ij} = [x_{ij}^T, y_{ij}^T, z_{ij}^T, x_{ij}^R, y_{ij}^R, z_{ij}^R] \quad (14)$$

$$Y_{ij} = \omega_{ij} = [\rho_{ij}, \mu_{ij}, \mu_{M,ij}, \vartheta_{ij}^T, \varphi_{ij}^T, \vartheta_{ij}^R, \varphi_{ij}^R, \vartheta_{M,ij}^T, \varphi_{M,ij}^T, \vartheta_{M,ij}^R, \varphi_{M,ij}^R] \quad (15)$$

The first convolutional layer filters the  $1 \times 6$  input vector with 16 kernels of size  $1 \times 3$ . The second convolutional layer takes the output of the first convolutional layer as input and filters it with 32 kernels of size  $16 \times 3$ . Both of the two convolutional layers take with a stride of one node. We zero pad the activation to match the number of features. After each convolutional layer, batch normalization in [48] and rectified linear unit (ReLU) are placed to speed up the model convergence. The output of the second convolutional layer is then fully connected to 16 neurons. The following dense layers have 16, 32, 64, and 1 neurons, respectively. In order to obtain the optimized training result, we train the 11 channel characteristics individually by the CNN. Each time we input all 6 elements of  $X_{ij}$  into CNN and output 1 element of  $Y_{ij}$ . ReLU is placed after each dense layer except the last layer. Unlike in computer vision, we do not place pooling layer between the convolutional layers, because the input of our network which is only 6 nodes is relatively sparser than the image which commonly contains millions of pixels. Pooling layer will lose useful information and make the model convergence at a high loss.

As shown in Table 2, this model has 7280 parameters in total. Most parameters are between the second convolutional layer and the first dense layer. The number of these parameters accounts for 42.20% of the total number of model parameters.

The CNN of one output node was designed. The 11 different labels (PL, DM, DS, AAMA, AASA, AAMD, AASD, EAMA, EASA, EAMD, and EASD) are individually used to train the CNN to obtain the different weights in terms of the least loss. Once the label is determined, the loss function and back propagation are applied end-to-end. The mean square error (MSE) function is used as the loss function in all CNNs of 11 labels. The learning rate is fixed throughout once

training. We used an equal learning rate which was initialized at 0.0001 for all layers. The root mean square propagation (RMSProp) in [49] with momentum of 0.9 and smooth factor of  $10^{-6}$  is used to optimize the weights of the model. The update rule for weight  $\beta$  is

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2 \quad (16)$$

$$\beta_{t+1} = \beta_t - \frac{\eta}{\sqrt{E[g^2]_t + \kappa}} g_t \quad (17)$$

where  $t$  is the iteration index,  $\eta$  is the learning rate,  $\kappa$  is smooth factor, and  $g_t$  is the gradient of the current iteration  $t$ .

Glorot uniform initializer in [50], which is also called Xavier uniform initializer, was used to initialize the weights in each layer. The weight was randomly created from a uniform distribution within  $[-\varepsilon, \varepsilon]$  with

$$\varepsilon = \sqrt{\frac{6}{l_{in} + l_{out}}} \quad (18)$$

where  $l_{in}$  is the number of input units and  $l_{out}$  is the number of output units in the weight tensor. We initialized the neuron biases in both convolutional layers and dense layers with the constant 0. This initialization accelerates the early stages of learning by providing the ReLUs with positive inputs.

## 5. Results and Analysis

The target of this section is twofold. The first intention is to verify the CNN in two indoor scenarios. Second, we will carry out comparisons between five different datasets to analyze the influence of dataset in CNN.

*5.1. Fittings between Predicted and Real Channel Characteristics.* In both the virtual classroom scenario and the lab scenario, all the predicted channel statistical characteristics generated by the CNN are in fairly good agreements with the channel statistical characteristics generated by the ray tracing software. In Figures 5–7, we show the fittings of PL, DM, and AAMA between predicted results and virtual simulation data in the two scenarios, respectively. As we can see, the predicted capability of the CNN is very good, and we can use this method to predict the channel statistical characteristics with limited simulation data in specified indoor scenarios. This shows that AI is meaningful for channel modeling. The massive data in wireless communication should be fully used and explored to make the performance of wireless communication networks better.

*5.2. RMSE.* To evaluate and compare the performances of the CNN with different datasets, we calculate the RMSE between predicted channel statistical characteristics and virtual simulation channel statistical characteristics; i.e.,

$$R(l) = \sqrt{E[(l_p - l_r)^2]} \quad (19)$$

TABLE 1: The Details of Databases.

Scenario	Dataset	Tx	Rx	No. of sample	No. of train set	No. of validation set
Virtual scenario	10100R	10	100	1000	700	300
	3232R	32	32	1024	717	307
Lab scenario	3030R	30	30	900	630	270
	3030G	30	30	900	630	270
	211211G	211	211	44521	31165	13356

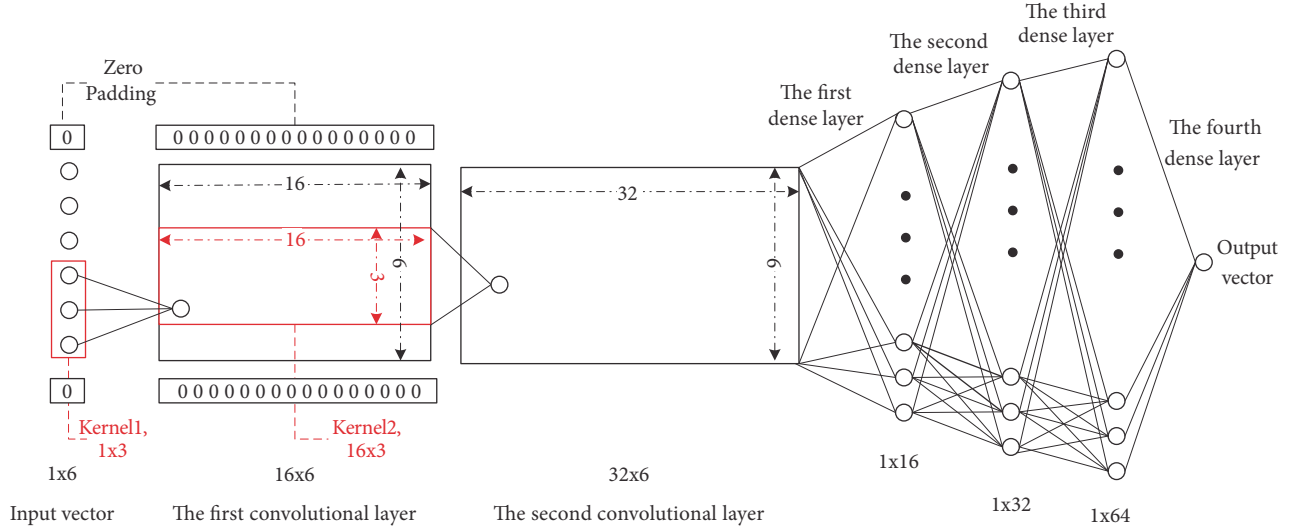


FIGURE 4: Architecture of the proposed CNN for channel statistical characteristics prediction.

TABLE 2: The parameter numbers of CNN layers.

Layers	No. of parameters
First convolutional layer	48
Second convolutional layer	1536
First dense layer	3072
Second dense layer	512
Third dense layer	2048
Fourth dense layer	64
Total	7280

where  $R(l)$  is the RMSE of the channel characteristic  $l$ , such as PL, DM, DS, AAMA, AASA, AAMD, AASD, EAMA, EASA, EAMD, and EASD.  $l_p$  and  $l_r$  denote the predicted result and virtual simulation result of the channel characteristic, respectively.

The RMSEs of channel statistical characteristics with two datasets in the virtual scenario are listed in Table 3. Train loss (TL) is the RMSE between the channel characteristic generated by the CNN and the virtual simulation channel characteristic in train data. Validation loss (VL) is the RMSE between the channel characteristic predicted by the CNN and the virtual measurement channel characteristic in the test data. In Table 3, VL of channel statistical characteristics of the 10100R is always larger than TL. Similar result is shown in the 3232R. The parameters of the CNN are trained based on MSE optimizer in the train data, and the test data is different for the

train data absolutely. So the results in the test data cannot be optimized as good as those in the train data. The performance of the CNN in the 10100R is better than in the 3232R, which is most obvious in the PL. The TL of the PL in 10100R (0.6408) is only 14.26% of that in 3232R (4.4932). The VL of the PL in 10100R (0.9586) is only 20.04% of that in 3232R (4.7832).

The RMSEs of channel statistical characteristics with two datasets in the lab scenario are listed in Table 4. As we can see, the performance of the CNN in the 3030G is better than that in the 3030R, which is most obvious in the PL. The TL of the PL in 3030G (1.0616) is only 34.70% of that in 3030R (3.0590). The VL of the PL in 3030G (1.3186) is only 41.27% of that in 3030R (3.1949). The performance of the CNN in the 211211G is better than that in the 3030G, which is most obvious in the AAMD. The TL of the AAMD in 211211G (6.7187) is only 47.27% of that in 3030G (14.2148). The VL of the AAMD in 211211G (7.1652) is only 35.24% of that in 3030G (20.3331).

There are 1024 samples and 1000 samples in the 3232R and the 10100R, respectively. The sample numbers of the two datasets belong to the same order of magnitudes, and both of them are generated when Tx and Rx are randomly located. According to the specified Tx antenna locations, there are 100 samples with different Rx antenna locations in the 10100R, while there are only 32 samples with different Rx antenna locations in the 3232R. The former is more various and more robust, which explains that the performance of 10100R is better than 3232R. The performance is determined by the robustness of data even they are in the same order



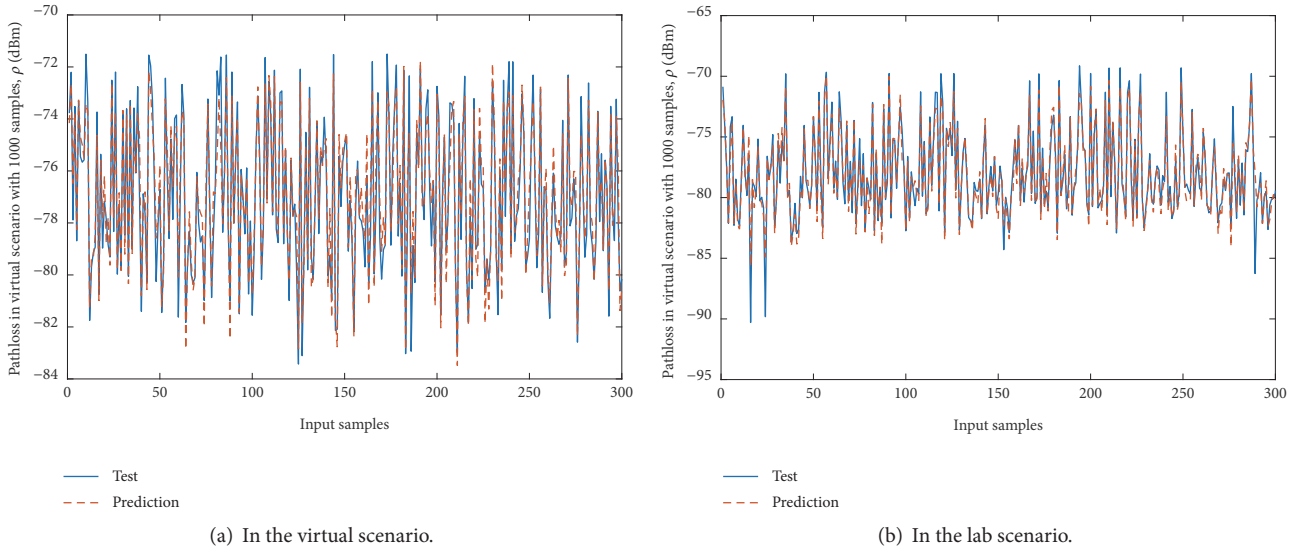


FIGURE 5: The predicted fitting of PL in two scenarios.

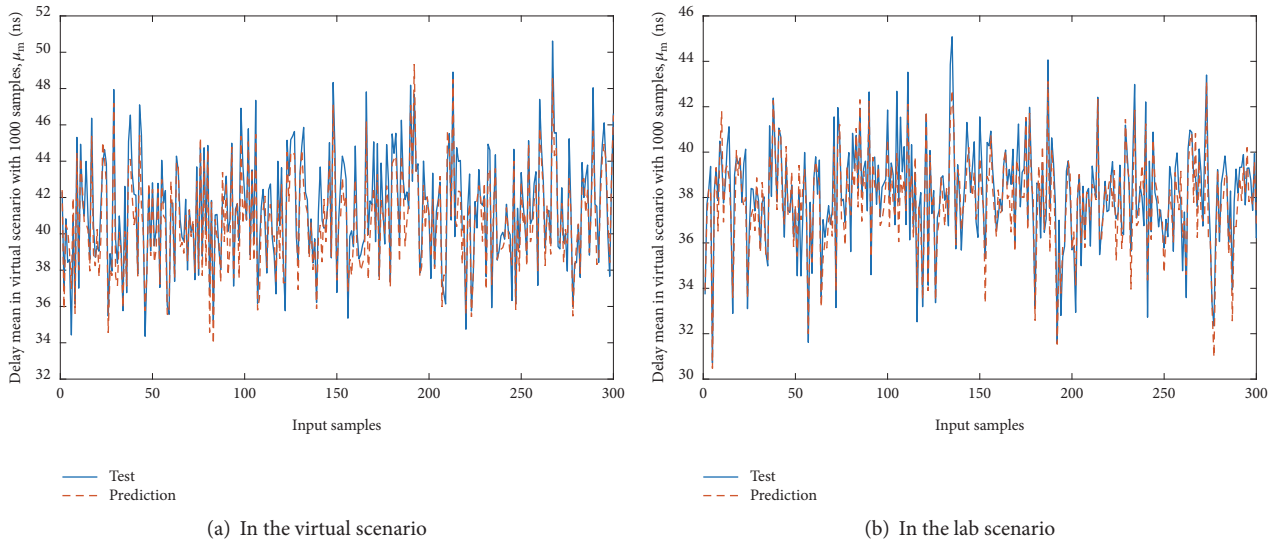


FIGURE 6: The predicted fitting of DM in two scenarios.

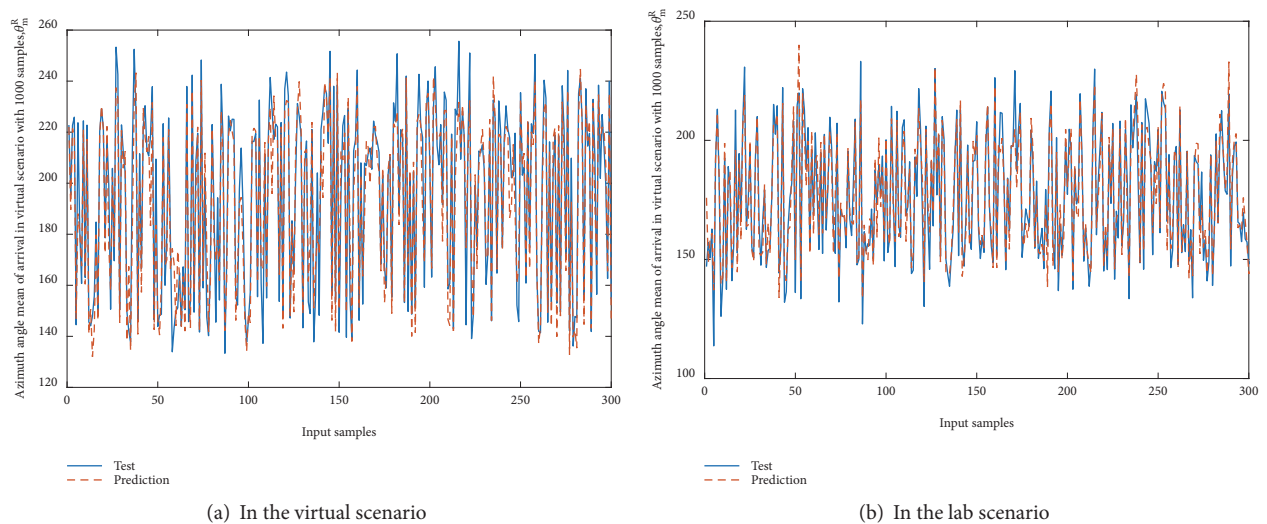


FIGURE 7: The predicted fitting of AAM in two scenarios.

TABLE 3: RMSE Loss with Different Dataset in the Virtual Scenarios.

Dataset	10100R		3232R		10100R and 3232R	
	TL	VL	TL	VL	TL in 10100R	VL in 10100R
					TL in 3232R	VL in 3232R
PL	0.6408	0.9586	4.4932	4.7832	0.1426	0.2004
DM	0.4232	0.4762	1.2986	2.3647	0.3259	0.2014
DS	0.2895	0.3708	1.7822	1.9887	0.1624	0.1865
AAMA	11.0096	13.3897	14.3276	14.4829	0.7684	0.9245
AASA	4.6625	5.6897	17.2259	18.7624	0.2707	0.3033
AAMD	11.8287	13.4187	12.8583	14.6427	0.9199	0.9164
AASD	5.3249	6.3327	21.0490	21.3898	0.2529	0.2961
EAMA	1.3355	1.4970	3.8356	4.8577	0.3482	0.3082
EASA	1.5680	1.8255	3.5929	3.8308	0.4364	0.4765
EAMD	0.8228	0.9298	2.8473	4.3977	0.2890	0.2114
EASD	2.3911	3.0512	2.7861	3.2362	0.8582	0.9428

TABLE 4: RMSE Loss with Different Dataset in the Lab Scenarios.

Dataset	3030R		3030G		211211G		3030G and 3030R		211211G and 3030G	
	TL	VL	TL	VL	TL	VL	TL in 3030G	VL in 3030G	TL in 211211G	VL in 211211G
							TL in 3030R	VL in 3030R	TL in 3030G	VL in 3030G
PL	3.0596	3.1949	1.0616	1.3186	1.0439	1.1459	0.3470	0.4127	0.9833	0.8690
DM	1.8142	2.0642	1.4970	1.8566	1.0196	1.0645	0.8252	0.8994	0.6811	0.5734
DS	0.4866	0.5037	0.2698	0.3528	0.2168	0.3038	0.5545	0.7004	0.8036	0.8611
AAMA	19.1869	19.8408	8.9489	9.0166	7.0688	7.5453	0.4664	0.4544	0.7899	0.8368
AASA	12.6992	12.2539	8.0887	8.4545	5.1746	5.8588	0.6369	0.6899	0.6397	0.6930
AAMD	17.7100	20.5364	14.2148	20.3331	6.7187	7.1652	0.8026	0.9901	0.4727	0.3524
AASD	15.1566	14.7420	8.3635	12.6681	6.2428	6.5880	0.5518	0.8593	0.7464	0.5200
EAMA	3.3544	3.4216	1.4747	1.7197	1.2807	1.4928	0.4396	0.5026	0.8684	0.8681
EASA	1.5093	1.8472	0.8160	0.6658	0.8084	0.4824	0.5356	0.3604	0.9907	0.7245
EAMD	2.2215	2.4836	1.9990	1.9906	1.6045	1.7258	0.8998	0.8015	0.8027	0.8670
EASD	0.3008	0.3739	0.2989	0.3697	0.2287	0.2360	0.9937	0.9888	0.7603	0.6384

of magnitudes. It is determined by the robustness of data. The comparison between the performance of 3030G and 3030R shows that the data collection in grid is better than that in random. The comparison between the performance of 211211G and 3030G shows that more robust data generated by the specified collection way results in a better predicted performance. The above-mentioned conclusions have significant meaning on data collection.

**5.3. PDF of Channel Characteristics Error Square.** For the further analysis of the performance of five different datasets, the PDF of channel statistical characteristics error square which can show the distribution of the channel statistical characteristics error square are given in Figure 8. The PDFs of error square of DM and AAMA are shown in Figures 8(a) and 8(b), respectively. In view of that the train loss and validation loss in the 211211G of lab scenario are only slightly lower than those of the 3030G in lab scenario in Table 4, and the advantage of large dataset is not obvious if we take the time and energy consuming of data collection into account. However, Figures 8(a) and 8(b) show that the core superiority of the 211211G in lab scenario is the PDFs of

error square in which the proportion of accurate predicted channel characteristic (error square = 0) is very large. It is better that the lower channel characteristic error square has higher probability and vice versa. In Figure 8(b), the proportion of accurate predicted AAMA of the 3232R in the virtual scenario (error square = 0) is larger than that of the 10100R in the virtual scenario, but the proportion of predicted AAMA with a high error square in the 3232R in the virtual scenario is also larger than that of the 10100R in the virtual classroom scenario. The fleet decline tendency of PDF of channel characteristic error square is what we expected.

## 6. Conclusions and Future Work

The AI enabled procedure to predict channel statistical characteristics has been proposed in this paper. The channel parameters of massive MIMO and mmWave indoor channel have been generated by a ray tracing software Wireless InSite. The channel statistical characteristics after data preprocessing, such as PL, DM, DS, AAMA, AASA, AAMD, AASD, EAMA, EASA, EAMD, and EASD, can be predicted by CNN. A virtual classroom scenario and a real lab scenario

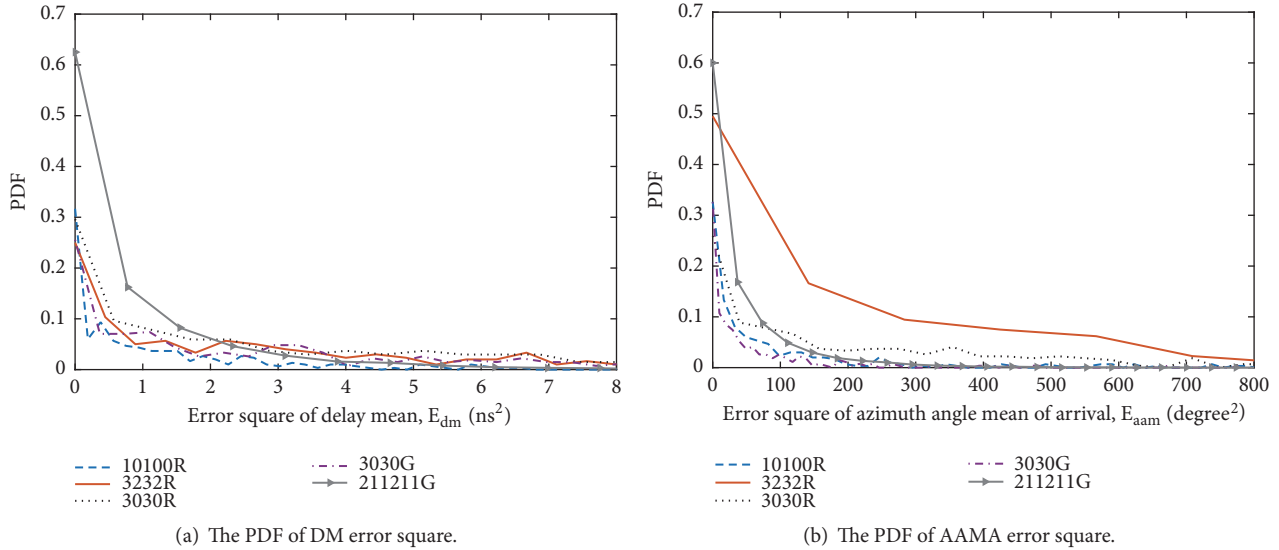


FIGURE 8: The PDF of channel statistical characteristics error square with different datasets.

have been set up to verify this algorithm. The good fittings between the predicted channel statistical characteristics and the real channel statistical characteristics have been shown in this paper. By comparing between the performance of different datasets, the better data collection rule has also been proposed. The generalization of AI enabled procedure to predict channel statistical characteristics for more scenarios is an important task to be solved in the future.

### Data Availability

The data can be made available if requested.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

The authors gratefully acknowledge the support from the National Natural Science Foundation of China (no. 61771293), the Science and Technology Project of Guangzhou (no. 201704030105), the EPSRC TOUCAN project (no. EP/L020009/1), the EU H2020 5G Wireless project (no. 641985), the EU H2020 RISE TESTBED project (no. 734325), the Key R&D Program of Shandong Province (no. 2016GGX101014), the Fundamental Research Funds of Shandong University (no. 2017JC029), and the Taishan Scholar Program of Shandong Province.

### References

- [1] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [2] Huawei, "5G a technology vision," White Paper, [https://www.huawei.com/ilink/en/download/HW\\_314849](https://www.huawei.com/ilink/en/download/HW_314849).
- [3] Samsung, "5G vision," White Paper, <https://www.samsung.com/global/business/networks/insights/5g-vision/>.
- [4] I. Gaspar and G. Wunder, "5G cellular communications scenarios and system requirements," Tech. Rep., 2013, <https://www.5gnow.eu/>.
- [5] C.-X. Wang, F. Haider, X. Gao et al., "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.
- [6] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Communications Magazine*, vol. 23, no. 1, pp. 72–79, 2016.
- [7] A. Molish, *Wireless Communications*, John Wiley & Sons, London, UK, 2011.
- [8] V. Nurmela, A. Karttunen, A. Roivainen et al., "METIS channel models," METIS ICT-317669/D1.4, 2015.
- [9] A. Maltsev, A. Pudueyev, I. Bolotin et al., "Channel modeling and characterization," MiWEBA FP7-ICT-608637/D5.1, V1.0, 2014.
- [10] ITU-R, "Preliminary draft new report ITU-R M. [IMT-2020.EVAL]," Tech. Rep. R15-WP5D-170613-TD-0332, ITU-R, Niagara Falls, Canada, 2017.
- [11] L. Liu, C. Oestges, J. Poutanen et al., "The COST 2100 MIMO channel model," *IEEE Wireless Communications Magazine*, vol. 19, no. 6, pp. 92–98, 2012.
- [12] R. Verdone and A. Zanella, *Pervasive Mobile and Ambient Wireless Communications*, Springer, London, UK, 2012.
- [13] A. Maltsev, A. Pudueyev, A. Lomayev, and I. Bolotin, "Channel models for IEEE 802.11ay," IEEE doc 802.11-15/1150r9, 2016.
- [14] mmMAGIC, "Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications," mmMAGIC H2020-ICT-671650-mmMAGIC/D2.1, v1.0, 2016.
- [15] A. Maltsev, A. Pudueyev, I. Karls et al., "Quasi-deterministic approach to mmWave channel modeling in a non-stationary environment," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps '14)*, pp. 966–971, Austin, Tex, USA, December 2014.

- [16] S. Jaeckel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, 2014.
- [17] S. Wu, C.-X. Wang, H. Aggoune, M. M. Alwakeel, and X. You, "A general 3-D non-stationary 5G wireless channel model," *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 3065–3078, 2018.
- [18] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-Frequency mmWave Massive MIMO Channel Measurements and Characterization for 5G Wireless Communication Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1591–1605, 2017.
- [19] R. Feng, J. Huang, J. Sun, and C. Wang, "A novel 3D frequency domain SAGE algorithm with applications to parameter estimation in mmWave massive MIMO indoor channels," *Science China Information Sciences*, vol. 60, no. 8, 2017.
- [20] C.-X. Wang, S. Wu, L. Bai, X. You, J. Wang, and C.-L. I, "Recent advances and future challenges for massive MIMO channel measurements and models," *Science China Information Sciences*, vol. 59, no. 2, pp. 1–16, 2016.
- [21] P. Ferrand, M. Amara, S. Valentin, and M. Guillaud, "Trends and challenges in wireless channel modeling for evolving radio access," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 93–99, 2016.
- [22] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 190–199, 2015.
- [23] L. O. Chua and T. Roska, "CNN paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 147–156, 1993.
- [24] R. He, Q. Li, B. Ai et al., "An automatic clustering algorithm for multipath components based on kernel-power-density," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '17)*, pp. 1–6, San Francisco, Calif, USA, March 2017.
- [25] R. He, Q. Li, B. Ai et al., "A kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7138–7151, 2017.
- [26] C. Huang, R. He, Z. Zhong, Y.-A. Geng, Q. Li, and Z. Zhong, "A Novel Tracking-Based Multipath Component Clustering Algorithm," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 2679–2683, 2017.
- [27] N. Czink, R. Tian, S. Wyne et al., "Tracking Time-Variant Cluster Parameters in MIMO Channel Measurements," in *Proceedings of the 2nd IEEE International Conference on Communications and Networking in China (CHINA-COM '07)*, pp. 1147–1151, Shanghai, China, August 2007.
- [28] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo, "A framework for automatic clustering of parametric MIMO channel data including path powers," in *Proceedings of the IEEE 64th Vehicular Technology Conference (VTC-Fall '06)*, pp. 1–5, Montreal, Canada, September 2006.
- [29] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo, "Automatic clustering of MIMO channel parameters using the multipath component distance measure," in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications (WPMC '05)*, 2005.
- [30] N. Zaarour, N. Kandil, and N. Hakem, "An accurate neural network approach in modeling an UWB channel in an underground mine," in *Proceedings of the IEEE Antennas and Propagation Society International Symposium (APSURSI '13)*, pp. 1608–1609, Orlando, Fla, USA, July 2013.
- [31] N. Zaarour, N. Kandil, N. Hakem, and C. Despins, "Comparative experimental study on modeling the path loss of an UWB channel in a mine environment using MLP and RBF neural networks," in *Proceedings of the 3rd International Conference on Wireless Communications in Underground and Confined Areas (ICWCUCA '12)*, Clermont Ferrand, France, August 2012.
- [32] N. Zaarour, S. Affes, N. Kandil, and N. Hakem, "Comparative study on a 60 GHz path loss channel modeling in a mine environment using neural networks," in *Proceedings of the IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB '15)*, Montreal, Canada, October 2015.
- [33] M. Kalakh, N. Kandil, and N. Hakem, "Neural networks model of an UWB channel path loss in a mine environment," in *Proceedings of the IEEE 75th Vehicular Technology Conference (VTC Spring '12)*, Yokohama, Japan, June 2012.
- [34] J. Huang, C.-X. Wang, L. Bai et al., "A big data enabled channel model for 5G wireless communication systems," *IEEE Transactions on Big Data*, In press.
- [35] T. Moazzeni, "A wireless propagation channel model with meteorological quantities using neural networks," in *Proceedings of the IEEE GCC Conference (GCC '06)*, pp. 1–4, Manama, Bahrain, March 2006.
- [36] Y. Ma, K. Liu, and Y. Guo, "Artificial neural network modeling approach to power-line communication multipath channel," in *Proceedings of the IEEE International Conference Neural Networks & Signal Processing*, Zhenjiang, China, June 2008.
- [37] L. Lv, "A novel wireless channel model with multiply feed-forward neural network," in *Proceedings of the 3rd International Conference on Natural Computation (ICNC '07)*, pp. 730–734, Haikou, China, August 2007.
- [38] X. Ye, X. Yin, X. Cai, A. Perez Yuste, and H. Xu, "Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks," *IEEE Access*, vol. 5, no. 5, pp. 12071–12087, 2017.
- [39] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: deep learning for indoor fingerprinting using channel state information," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '15)*, pp. 1666–1671, New Orleans, La, USA, March 2015.
- [40] M. Kotol and Z. Raida, "Comparison of neural models of UWB and 60 GHz in-car transmission channels," in *Proceedings of the International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom '16)*, pp. 1–4, Graz, Austria, September 2016.
- [41] M. Kotol, Z. Raida, and J. Velim, "Neural modeling of in-vehicle wireless channels: Wave propagation along the vehicle body at 60 GHz," in *Proceedings of the IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC '15)*, pp. 279–282, Turin, Italy, September 2015.
- [42] J. Zhang, "The interdisciplinary research of big data and wireless channel: a cluster-nuclei based channel model," *China Communications*, vol. 13, Article ID 7833457, pp. 14–26, 2016.
- [43] Wireless InSite Propagation Software, <https://www.remcom.com/wireless-insite-em-propagation-software>.
- [44] ITU-R, "Effects of building materials and structures on radiowave propagation above about 100 MHz," ITU-R P.2040-1, 2015.
- [45] J. Lu, D. Steinbach, P. Cabrol, P. Pietraski, and R. V. Pragada, "Propagation characterization of an office building in the 60

- GHz band,” in *Proceedings of the 8th European Conference on Antennas and Propagation (EuCAP '14)*, pp. 809–813, The Hague, Netherlands, April 2014.
- [46] D. Ferreira, I. Cuinas, R. F. Caldeirinha, and T. R. Fernandes, “A review on the electromagnetic characterisation of building materials at micro- and millimetre wave frequencies,” in *Proceedings of the 8th European Conference on Antennas and Propagation (EuCAP '14)*, pp. 145–149, The Hague, Netherlands, April 2014.
- [47] K. Korolev and M. Afsar, “Complex dielectric permittivity measurements of materials in millimeter waves,” in *Proceedings of the Joint 30th International Conference on Infrared and Millimeter Waves (IRMMW-THz '05)*, pp. 594–595, Williamsburg, Va, USA, September 2005.
- [48] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” <https://arxiv.org/abs/1502.03167>.
- [49] Keras Document, <https://keras.io/>.
- [50] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

## Research Article

# A Request-Based Handover Strategy Using NDN for 5G

Fan Jia <sup>1</sup> and Xiaolin Zheng<sup>1,2</sup>

<sup>1</sup>Beijing Jiaotong University, China

<sup>2</sup>Huawei Technologies Co. Ltd., China

Correspondence should be addressed to Fan Jia; [fjia@bjtu.edu.cn](mailto:fjia@bjtu.edu.cn)

Received 15 March 2018; Accepted 29 July 2018; Published 8 August 2018

Academic Editor: Sana Salous

Copyright © 2018 Fan Jia and Xiaolin Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the developing of 5G, it is widely accepted that 5G will use a system architecture that supports the ultradense networks (UDN) deployments. In this architecture, a user will be covered by a large amount of small cell base stations (SBS) in 5G. However, selecting an SBS for handover is a great challenge. To address the challenge, the emerging content-oriented Named Data Networking (NDN) has attractive advantages, such as providing name-based routing. In this paper, a request-based handover strategy (RBHS) is presented to improve the user experience in performance and obtain the optimal allocation of resources, and a caching mechanism based on the users' requests is introduced for it. The proposed caching mechanism and access network selection mechanism were validated utilizing ndnSIM. Simulation results demonstrate that our proposed strategy achieves around 30% higher cache hit rate and 20% more traffic reduction, compared with the access network selection based on SINR.

## 1. Introduction

Mobile data traffic and mobile devices have been exponentially growing, and monthly global mobile data traffic maybe surpasses 15 exabytes in 2018, which poses a significant challenge to the mobile communication system [1–6]. The current deploying fourth-generation mobile communication system (4G) has been unable to meet the new challenge. Then the fifth-generation mobile communication system (5G) has been proposed to address the challenge, which aims to achieve 1000 times higher mobile data volumes, 10 times higher number of connected devices, 10 times higher typical end-user data rates, 10 times the spectral efficiency, 5 times lower latency, and 25 times the average cell throughput compared with 4G [3]. In order to achieve the 5G system requirements, the 5G cellular architecture should use a system architecture that supports ultradense networks (UDN) deployments [4–6]. The UDN means that, in the coverage of macro base station (Macro), the density of SBS with low-power radio transmission technology will reach 10 times more than the existing density of SBS deployment, the distance between the SBS will be 10 meters or less [7, 8], the users per square kilometer will reach 25000 [9], and the number of active users and the number of SBS reach 1 to 1

ratio in the future [10]. Nevertheless, it is a great challenge to choose the optimal small cell base station (SBS) to connect in the environment of ultradense small cell base stations to relieve the burden of the links between Macro and backbone.

In this paper, we propose a request-based handover strategy (RBHS). The goals of this strategy are to make the optimal allocation of network resources, reduce the data traffic and decrease the latency, and let the user obtain the optimal allocation of resources, in order to improve the user experience. In RBHS, we will focus on the analysis of user requests. For the purposes of analysis and calculation of user requests, we introduce the Named Data Networking (NDN) into 5G cellular architecture. NDN is a new Internet architecture and is initiated by National Science Foundation (National Science Foundation, NSF) in 2010 [7, 8]. In NDN, each router is equipped with a fixed amount of memory to cache content, which results in the difference between NDN and the traditional IP networks. Basically, NDN runs requester-driven communication model; i.e., a client will first send out an interest packet for the desired content and then a router that has the same content in local cache memory will return the content within a data packet. Taking the exponentially growing mobile data traffic into account, we also introduce cache module into SBS as well as Macro.

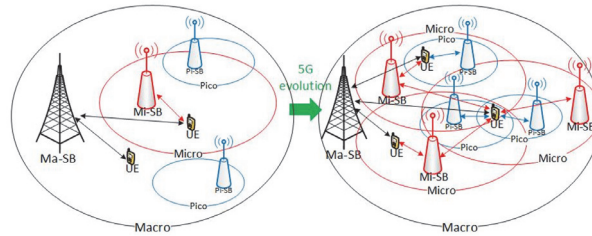


FIGURE 1: The evolution of 5G cellular architecture.

By caching the popular contents in SBS and Macro, the corresponding requests will no longer need to trace back the server which will save a considerable amount of redundant traffic. In summary, the strategy of this paper is divided into two parts; one is caching mechanism and the other one is handover mechanism.

The rest of the paper is organized as follows: Section 2 summarizes the related work. Section 3 introduces our system model which includes the caching mechanism and the handover mechanism. The RBHS is presented in Section 4. The evaluation setting, metrics, impact factors, and simulation results of RBHS are discussed in Section 5. Finally, Section 6 concludes the paper.

## 2. Related Work

As the amount of mobile data traffic continued to rise as well as the explosive growth of mobile devices [1–6], 5G has become a quite hot topic nowadays. More and more researchers pay their attention to the horizontal topic [1, 2]. METIS (mobile and wireless communications enablers for the 2020 information society) is an integrated project partly funded by the European Commission under the FP7 research framework and is considered as the 5G flagship project [2, 3]. What is more, 863 plan in China launched a 5G major project phase I and phase II separately in June 2013 and March 2014 [3]. At present, countries around the world are having a wide range of discussions for 5G development vision, application requirements, candidate of frequencies, and key technical indexes. Under the joint efforts of the countries around the world, 5G vision and capability requirements have been basically clear. The standardization for 5G has been in full gear since early 2016 [2] and would be set in 2018. According to [1–6], 5G will use a system architecture that supports the ultradense networks (UDN) deployment, which may consist of different types of infrastructure elements (BSs), such as macro-, micro-, pico-BSs. Low-power BSs like pico-BS will be used to enhance coverage and capacity by covering areas that are much smaller than a macro-BS coverage area. The UDN offers multiple options for satisfying application requirements [1, 3].

Therefore, in such complex environment, it is wise to introduce the NDN into 5G cellular architecture. NDN is a content-centric architecture, which provides name-based routing [11, 12]. Based on this characteristic, we can obtain the user request information easily which cannot be done in the IP networks. NDN has several attractive advantages,

such as network load reduction, low dissemination latency, and energy efficiency. To achieve these benefits from NDN paradigm, the content caching mechanism plays the most important role. The solution [13] proposed, called Hamlet, differs from previous work by reason that it helps users to make the decision about what information to keep, and for how long, based on a probabilistic estimate of what is cached in the neighborhood. The work [14] proposes a collaborative caching scheme guided by traffic engineering (TECC) for the emerging content-centric networks. The work [15] presents a collaborative caching scheme guided by traffic engineering (TECC) for the emerging content-centric networks. The work [16] develops a popularity-based coordinated caching strategy named the Effective Multipath Caching (EMC) scheme. The works [17–20] introduce cache into small cell base station.

Considering that the users are more likely to take a more active role in 5G (e.g., selecting the set of serving base stations, performing advanced interference rejection, or exploiting local cooperation) [2], we suggest taking the user request into the caching policy as well as handover policy. In this work, we develop a request-based handover strategy, which adopts the NDN network architecture. To our best knowledge, very few studies have tried to do that.

## 3. System Model and Problem Statement

Since NDN is newly proposed in 5G, we first sketch out the 5G cellular architecture using NDN in this section. Then, combined with our optimized objective, we illustrate the key issue of caching mechanism and handover mechanism.

### 3.1. An Overview of the 5G Cellular Architecture Using NDN.

UDN is a promising network densification cellular architecture in the 5G era, which aims at spectrum-efficient and energy-efficient solution that copes with a large number of devices and the huge mobile data traffic in future wireless applications. As illustrated in Figure 1, the 5G network will further enable the existing small cell miniaturization and distribution. And the distribution of small cell in the future will be further intensive; the density of base station deployment will increase by more than 10 times. That is, users will be more likely to be repeatedly covered by SBS (small cell base stations) and UE (user equipment) can reselect SBS. We suggest taking the users requests into consideration through introducing NDN architecture.

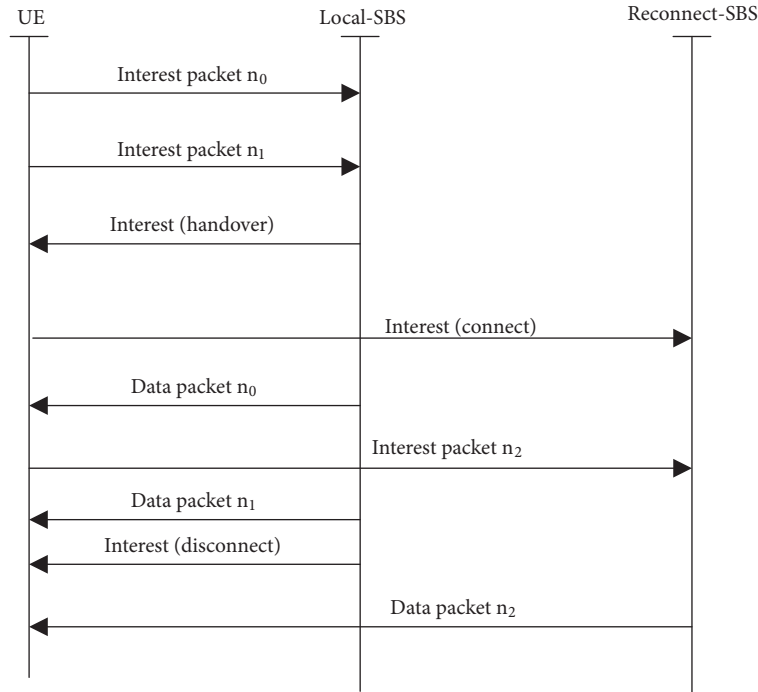


FIGURE 2: The process of handover.

As well known, NDN is a protocol stack [7, 8], which is easier to be managed and achieves better performance than IP protocol stack. There are two types of packets in NDN, interest and data. When a user requests for content, it will send out an interest packet which contains the name of the interested content. And data packets are the reply messages issued by the nodes that have the data satisfying the name of the interested content. Data is transmitted only in response to an interest and consumes that interest. NDN has three main data structures: Forwarding Information Base (FIB), Content Store (CS), and Pending Interest Table (PIT). The FIB is used to forward interest packets toward potential sources of matching data. The CS is the same as the buffer memory of an IP router but has a different replacement policy. The PIT keeps track of Interests forwarded upstream toward content sources so that returned data can be sent downstream to its requesters.

In our strategy, we propose to install FIB, CS, and PIT into SBS and Macro to fulfill the architecture of NDN. It is obvious that handover may cause data packet to be returned to an inaccessible location (the previous connected SBS). Furthermore, to receive these unreceived data packets, UE needs to initiate the recovery mechanism by retransmitting its interest packets. As a result, the handover in NDN architecture may increase the retransmission probability and introduce significant latency. How to solve this problem is the main difficulty of introducing the NDN architecture into 5G. We noticed that the user can simultaneously be connected to several wireless access technologies and seamlessly move between them (see media independent handover or vertical handover, IEEE 802.21, also expected to be provided by future 4G releases). We suggest, in 5G, UE may simultaneously

connect to the previous SBS and the new selected SBS. And the timing of disconnection to the previous SBS is the time finish of the last request. The handover process is shown in Figure 2.

What is more, we suggest adding a reconnection list to the interest packet which involves SBS that can be reconnected. And, besides the interest packet and the data packet, we add the confirmation packet which is used to request the data from SBS after reconnection. The details of the three types of packet are described in Figure 3. What is more, in order to facilitate the decision making, we add two tables: neighbor cache table (NCT) and neighbor state table (NST), which are shown in Figure 4. The NCT is used to record the content data cached by the nearby SBS. The NST is used to record the state of the nearby SBS.

In the next subsection, we will discuss mechanism about caching and handover.

**3.2. Caching Mechanism.** Based on the characteristics of 5G cellular architecture and NDN architecture, we have observed that while the popular contents are cached in the SBS (or Macro) and the user request for the same content, the SBS can directly deliver the content to the user without asking the server. For example, as shown in Figure 5, when UE1 requests for content data D1, UE1 can directly get D1 from SBS1 or Macro with no need to send the request to the server.

Since every base SBS are equipped with a limited storage space to cache content data, how to improve the cache hit rate effectively attracted our attention. It has become apparent that the technical key issues of caching mechanism fall into the following two questions [4]:



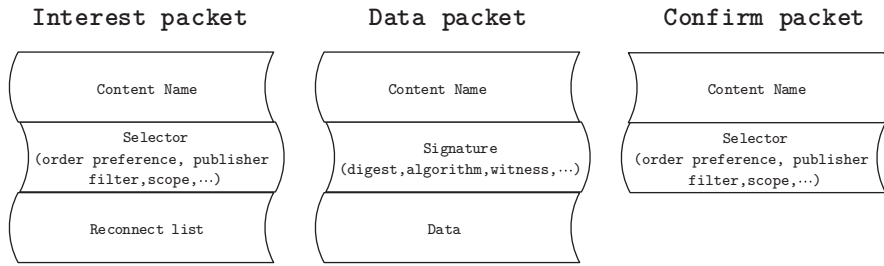


FIGURE 3: Packet types.

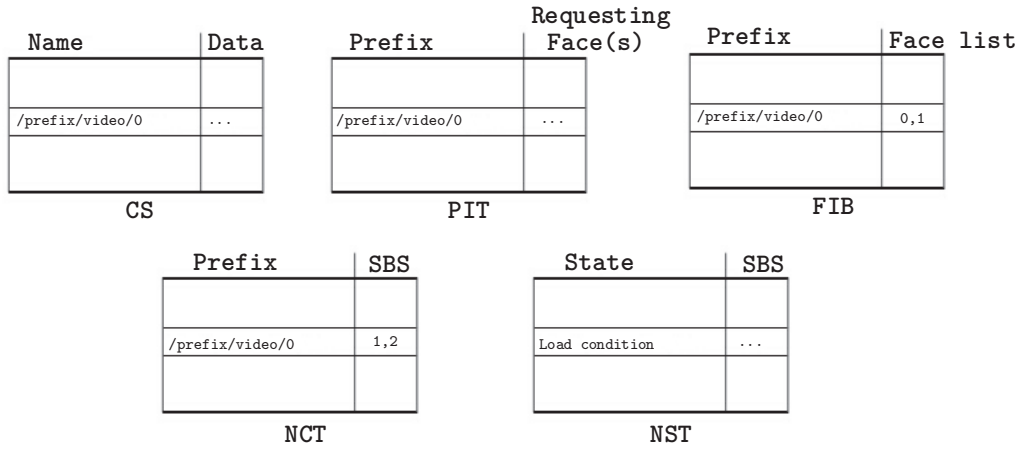


FIGURE 4: Main data structures used in NDN for SBS.

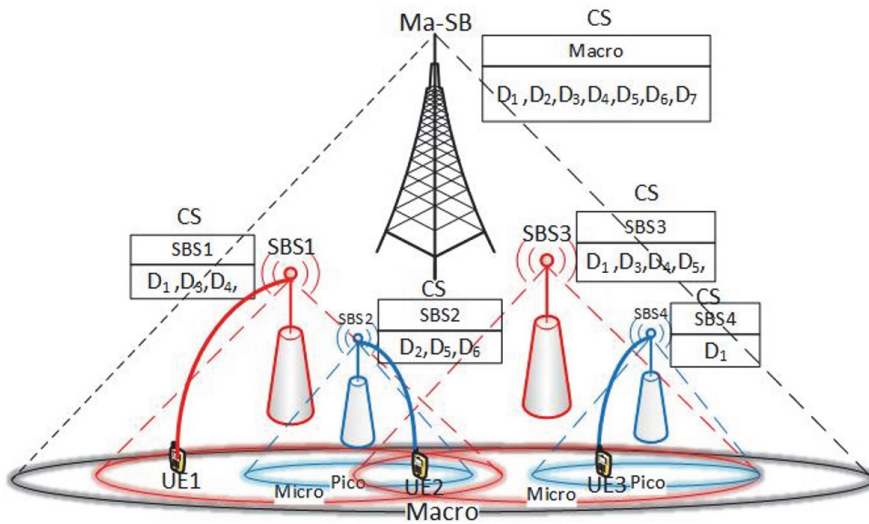


FIGURE 5: Introducing NDN in 5G.

(1) What to cache?

There are various contents in the Internet, and the cache space of SBS is limited. It is hence important to decide what content to cache taking into account content popularity. And SBS do not necessarily have to cache similar contents since the users they serve are different and they can share and exchange contents. Obviously, it is of vital importance to improve the

diversity of the cached content to augment the hit ratio of the cache content.

(2) How to cache?

Caching policies, deciding what to cache, and when to release caches are crucial for overall caching performance. And the goal of the caching policy is to augment the hit ratio. Under the consideration, the current popularity, the trend of

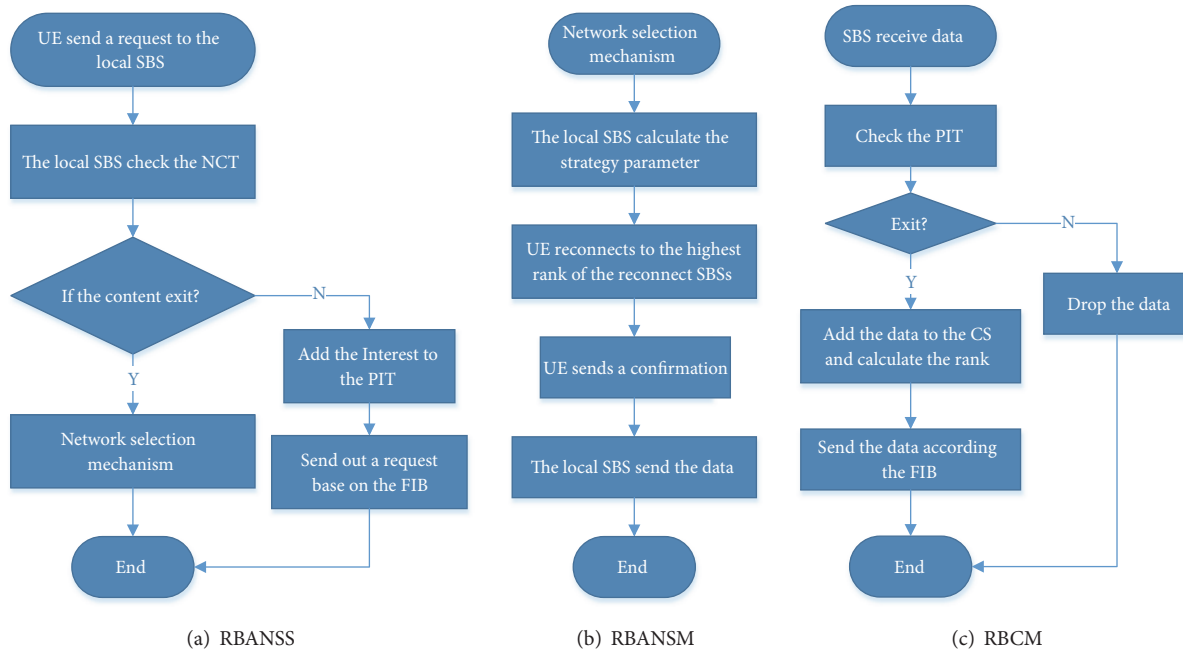


FIGURE 6: Introducing NDN in 5G.

popularity, storage size, and the location of replicas should be involved in the strategy.

**3.3. Access Network Selection Mechanism.** In wireless communication system, the density of the distribution of small cell base station led to the network repetitive coverage. And, in 5G era, this feature will become more and more obvious. For the repetitive coverage, there are diverse SBS to connect. That is, every user can reselect other SBS or Macro. Therefore, the main topic of handover mechanism can be divided into two parts:

(1) When to reselect the SBS?

As shown in Figure 5, when UE2 requests content data D6, UE2 will connect to SBS2 and obtain content data D6 from the CS of SBS2. After that, if UE2 requests content data D1, then UE2 will first consider reconnecting to SBS1, SBS3, or Macro.

(2) How to reselect SBS?

In the above case, when UE2 decides to reselect SBS, it is obvious that it need choose one from SBS1, SBS3, and Macro. We suggest taking the condition of UE and the condition of base station into account.

#### 4. RBHS: The Effective Handover Strategy

With the object of reducing the traffic between the macro and the server, we need to gain a high cache hit rate. We suggest that the timing of handover and the cache update mechanism are very important.

Our strategy process is shown in Figure 6.

When UE sends an interest packet to the local SBS, the local SBS will first check the NCT according to the

reconnection list in the interest packet. If not found, the local SBS will add the interest to the PIT and send the request to the next hop according to the FIB. Else, it will trigger the handover mechanism.

When the handover mechanism is active, the local SBS will compute the rank of the SBS in the reconnection list. Then the local SBS will let UE reconnect to the best SBS. After reconnection, UE will send a confirmation packet to the SBS to get the request data.

Considering the above content, we consider the receiving contents as a trigger of the caching mechanism. When the SBS receives data, it will first check the PIT. If not matching, it will drop the data. Or else it will add the data to the CS. If the CS is full, it will replace the data which is of the lowest rank and then send the data to UE according to PIT and FIB.

**4.1. Request-Based Caching Mechanism (RBCM).** Following the discussion in Section 3.2, we need to decide which data should be cache in the limit cache space and the update strategy which will be active when the cache space is full.

As illustrated in Figure 5, the local SBS will send the UE's request to the next hop only if there is no cache in the local SBS and the neighbors. Hence, if the incoming data match the PIT entry, the data will be stored in the CS and also will be sent to the requested UE according to the PIT and FIB. The question is, when the CS is full, how could the new incoming data be stored in the CS? It is obvious that we need to replace a content in the CS with the new incoming data. Therefore, it is meaningful to rank the data in the CS by setting a factor "value" to make sure the most valuable contents are stored in the caching space.

As well known, the goal of increasing the traffic saving is equivalent to the goal of storing the much more popular

data in the CS. Obviously, different data has different popularity, and the probability of UE requesting different data is different. But we cannot get the popularity of data. Because the more popular the data is, the higher cache hit it will be, we believe that the cache hit would affect the factor of "value", and we will take the cache hit count into consideration. We define the "hitcount" for each data stored in the CS, which indicates the popularity of the data. For  $data_i$  in the CS, when the UE's request matches the  $data_i$ , then  $hitcount_i = hitcount_i + 1$ . As we noticed, the popularity of data changes over time. Although the popularity of the data is high in this period, it may decline in the next period. Hence, the time of the data to be stored also should be taken into account.

As the UE can reselect the access network based on the request, it can reach all the data that are cache in the local SBS or in the SBS in the reconnection list. In other words, from the UE's viewpoint, all the CS in the local SBS and in the SBS in the reconnection list can be seen as a whole, therefore increasing the cache hit rate which means that we need to improve the diversity in the adjacent SBS. If the data in the CS is also stored in the CS of the adjacent SBS, the value of the data will decrease. The more the replicas exist in the CS of the adjacent SBS, the less the value of data will be.

Considering the above, the mathematical expression of the value is as follows:

$$val_{j,i}(t) = \frac{hitcount_i}{t - t_i} * \frac{1}{n_i} \quad (1)$$

As shown in formula (1),  $i$  means the  $data_i$ ;  $j$  means the SBS  $j$ .  $n_i$  is the number of replicas existing in the adjacent SBS.  $t$  is the current time,  $t_i$  is the time when  $data_i$  is stored, and  $hitcount_i$  represents the total number of the UE's request from  $t_i$  to  $t$ . So  $hitcount_i/(t - t_i)$  indicates the popularity of  $data_i$  in  $(t_i, t)$ .

After calculating all the values of data stored in the SBS, we could rank the data by value. And the value will be periodically updated. If new incoming data matches the PIT and will be stored in the CS while the CS is full, the data with min value will be deleted to cache the new incoming data.

**4.2. Request-Based Handover Mechanism (RBHM).** Following the discussion in Section 3.3, we need to decide the best timing of reselection and choose SBS. Based on the RBHM, we assume that the data in the adjacent SBS are sufficiently diverse. From the UE's viewpoint, the CS of all the adjacent SBS can be seen as a whole. But, in fact, each SBS is present individually. As a result, if we want to get the data from the SBS directly, we need to select the SBS which has already stored the data. According to the above analysis, the best timing of handover is the time after the UE determines to request some new content, which makes the currently connected network no longer the best access network. What is more, the request should be the most important part of handover. Channel capacity for a device may be determined by its RSS. In general, RSS depends on the distance between the UE and its attached BS. Also, to some extent, the RSS represents the mobility. Hence, the RSS should be taken into consideration. As the SINR and the load condition of the SBS

will affect the performance of access SBS, we also take SINR and load condition into consideration.

UE will periodically measure the RSS and SINR and choose the SBS with the RSS and SINR over the threshold. First order the SBS by RSS. And select the top  $n$  SBS by RSS and record sorting number as  $rank_{RSS,i}$ . Then order the top  $n$  SBS by SINR and record sorting number as  $rank_{SINR,i}$ . Finally add the top  $n$  SBS into the reconnection list in interest packet ( $n$  is the length of reconnection list). From the interest packet, the local SBS can obtain the reconnection SBS as candidate. We define the Weighted RSS ( $W_{RSS}$ ) and Weighted SINR ( $W_{SINR}$ ) for each candidate. So  $W_{RSS,i}$  and  $W_{SINR,i}$  are expressed as

$$W_{RSS,i} = \frac{rank_{RSS,i}}{\sum_i rank_{RSS,i}} \quad (2)$$

$$W_{SINR,i} = \frac{rank_{SINR,i}}{\sum_i rank_{SINR,i}} \quad (3)$$

The range of  $W_{RSS,i}$  and  $W_{SINR,i}$  is  $[0, 1]$ .

After that we need to define the weighted data ( $W_{data}$ ) for each candidate. From the predefined neighbor cache table (NCT) which records the content data stored in neighbor, we could find out whether the candidate has the data or not. So  $W_{data,i}$  is expressed as

$$W_{data,i} = \begin{cases} 0, & \text{the data do not exist in the CS of } C_i \\ 1, & \text{the data exist in the CS of } C_i \end{cases} \quad (4)$$

The range of  $W_{data,i}$  is  $[0, 1]$ .

We also define the weighted load ( $W_{load}$ ) for each candidate. From the predefined neighbor state table (NST) which records the amount of UE the SBS connect to. For the candidate  $C_i$ , set its load as  $load_i$ ; set max connection number as  $max_i$ . So  $W_{load,i}$  is expressed as

$$W_{load,i} = 1 - \frac{load_i}{max_i} \quad (5)$$

The range of  $W_{load,i}$  is  $[0, 1]$ .

After that, we use the parameter  $S_{u,i}$ , to represent the cost of the UE  $u$  connecting to the candidate  $C_i$ . Based on the aforementioned analysis,  $S_{u,i}$  is expressed as

$$S_{u,i} = w_1 * W_{data,i} + w_2 * W_{RSS,i} + w_3 * W_{SINR,i} + w_4 * W_{load,i} \quad (6)$$

We utilize the Analytic Hierarchy Process (AHP) [21] to calculate  $w_1, w_2, w_3, w_4$ . The AHP is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology. Rather than prescribing a "correct" decision, the AHP helps decision makers find one that best suits their goal and their understanding of the problem. In our consideration, compared with  $W_{RSS,i}, W_{SINR,i}$  and  $W_{load,i}, W_{data,i}$  is strongly preferred. And compared with

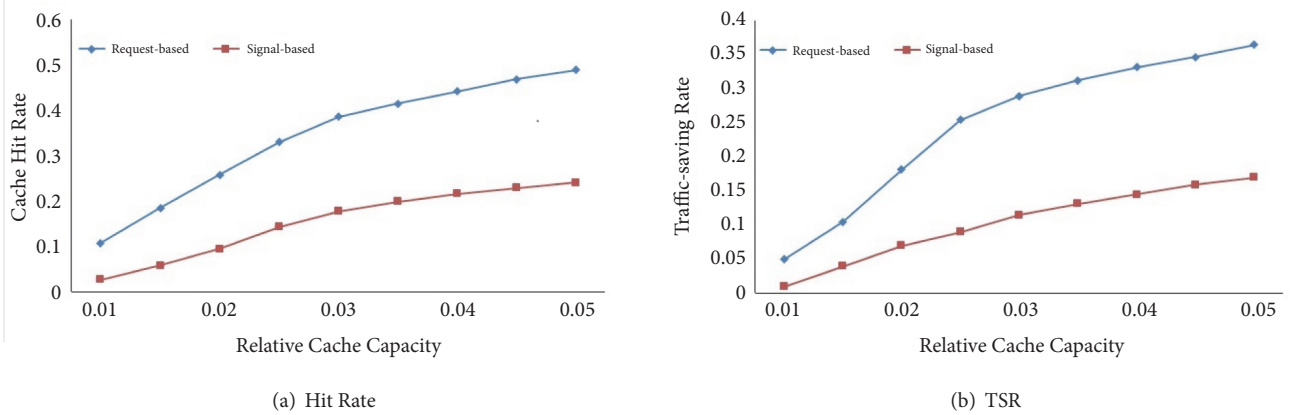


FIGURE 7: Strategy performance versus relative cache capacity.

$W_{load,i}$  and  $W_{SINR,i}$ ,  $W_{RSS,i}$  is moderately preferred. And  $W_{load,i}$  is as important as  $W_{SINR,i}$ .

From the above, we can obtain  $S_{u,i}$  which represent the benefit of UE  $u$  reconnection to the SBS candidate  $C_i$ . After calculate all  $S_{u,i}$  of UE  $u$ , we could acquire the maximum of all  $S_{u,i}$ . Then UE  $u$  will reconnects to  $C_{max}$ . And as we known, the requested content has already been stored in the reconnection SBS; then we can get the requested content by the reconnection SBS instead of sending a request to the server.

## 5. Simulation Results

In this section, we implement the handover strategy of RBHS in the ndnSIM simulator [22].

**5.1. Simulation Setting.** Considering the peculiar character which we add into the NDN architecture, we have modified the source code of ndnSIM. The basic configurations are explained as follows.

(1) Simulation environment: We set the simulation environment to the population density area. And we assume that all the SBS are evenly distributed around the Macro and all the UE pieces are randomly distributed in this area. For simplicity, the mobility model of UE is set to random walk with low speed.

(2) Performance metric: We take the traffic-saving rate (TSR) as the dominant metric to show the importance of saving the overall traffic between Macro and server. The TSR is the ratio of the average amount of traffic reduced by adopting RBHS to the amount of traffic which reselects SBS according to SINR. In our context, the traffic is equivalent to the number of the incoming packets. The cache hit rate indicates that the SBS has the data in the CS and could directly deliver the data to the UE without sending the request to server, which also represents saving the traffic. So we also involve it into our work.

(3) The compared handover strategy: The main idea of our strategy is request-based. The cache hit rate and the traffic-saving rate are increased by introducing the request-based mechanism to relieve the burden of the link between Macro and backbone caused by the growing mobile traffic

data and the UDN. So we take the handover strategy based on the SINR and load condition.

(4) Input data: We generate the synthetic input data as the following descriptions. Let  $D = \{data_1, data_2, data_3, \dots, data_i, \dots, data_n\}$  donate the set of content items. All the requests are identical and independently distributed within the set  $D$ .

The requests of each UE follow the Zipf-Mandelbrot law also known as the Pareto-Zipf law. The probability mass function is given by

$$f(k; N, q, s) = \frac{1/(k+q)^s}{H_{N,q,s}} \quad (7)$$

where  $H_{N,q,s}$  is given by

$$H_{N,q,s} = \sum_{i=1}^N \frac{1}{(i+q)^s} \quad (8)$$

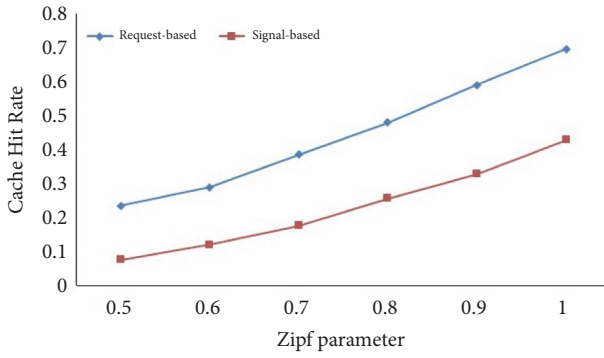
In the formula,  $N$  is the total number of data,  $k$  is the rank of the data, and  $q$  and  $s$  are the parameters of distribution. And  $s$  is the skewness factor indicating the consideration degree of the arrival of the requests.

(5) Impact factor and default setting: To explore the effectiveness and the scalability of RBHS, we take impact factors into consideration, including the cache size, request pattern, content population, and the size of reconnection list attaching in the interest packet.

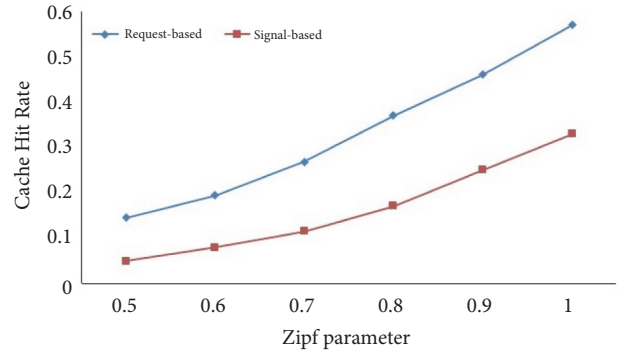
We set a default setting as follows. Associating with the size of chunks, we describe the cache size as relative cache size, which is the proportion of the cache size per SBS in the total size of all data. The relative cache size at each SBS is set to 3%. And the total number of data is  $N=1000$ . And the skewness factor is  $s=0.7$ . And the size of reconnecting list is list size=3.

**5.2. Experiment Results.** (1) Impact of cache size: We conduct the experiment in the range of relative cache size from 1% to 5%, while other parameters follow the default setting.

Figure 7 compares the cache hit rate and the traffic-saving rate (TSR) gained by the two strategies. Obviously, our



(a) Hit rate



(b) TSR

FIGURE 8: Strategy performance versus Zipf parameter.

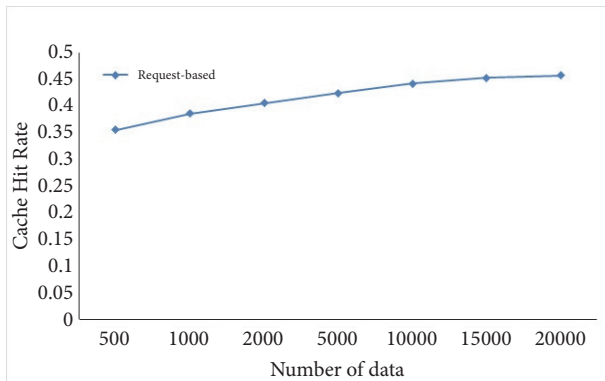


FIGURE 9: Cache hit rate versus number of data.

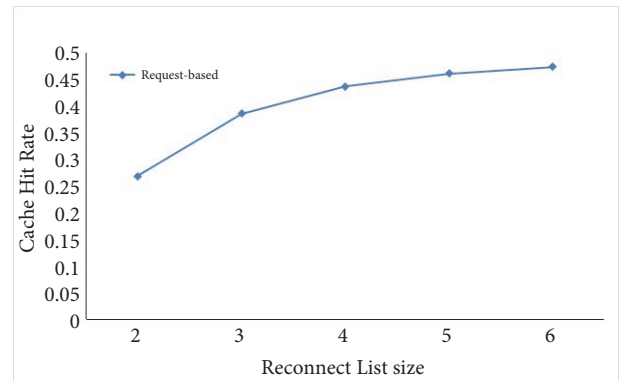


FIGURE 10: Cache hit rate versus reconnection list size.

proposed RBHS significantly outperforms the baseline of the strategy based on the SINR.

(2) Impact of the request pattern: As explained before, the requests follow the Zipf–Mandelbrot law. The parameter  $s$  is the key factor of Zipf–Mandelbrot law indicating the degree of the concentration of requests. In short, the larger  $s$  is, the fewer the data covering the major requests are.

In Figure 8, we test the impact of request patterns on effectiveness of the two strategy. Parameter  $s$  varies from 0.5 to 1.0 under the default setting. As shown in Figure 7, obviously, the more concentrated the requests are, the more effective the strategy is.

(3) Impact of the data population: To examine the scalability of the RBHS, we collect test data covering a large range of data scales, whose number of items varies from 500 up to 20,000.

Given that the relative cache size is fixed to 3%, we get the result in Figure 9. We are excited to see the saved traffic is increasing smoothly as the scale enlarges, which means RBHS will achieve even better performance if being deployed in the large network, considering the UE requests of data are increasing exponentially nowadays.

(4) Impact of the reconnection list: As mentioned in Section 4.2, by setting the size of reconnection list, we can

pretend that the cache size of the SBS is extended to (list size\*cache size) or smaller than (list size\*cache size) because there are copies existing in the adjacent SBS. As shown in Figure 10, the cache hit rate is increasing by the size of reconnection list and tends to be stable. We think that the reason why cache hit rate tends to be stable is that the effect of the reconnection list size is limited to the variety of the CS in the SBS.

## 6. Conclusion and Future Work

In this paper, we propose a request-base handover strategy, which can be divided into caching mechanism and handover mechanism, in order to deal with the UND in 5G. For the sake of the analysis of the user requests, we introduce NDN into 5G cellular architecture. And the simulation results demonstrate our proposed RBHS is effective and scalable.

In the future, since there are some horizontal topics of 5G such as D2D, we are to develop a handover strategy that takes the D2D communication into account. That means the conditions we need to consider become more complicated.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The work described in this paper was fully supported by “the Fundamental Research Funds for the Central Universities” (no. 2017JBM005).

## References

- [1] P. Demestichas, A. Georgakopoulos, D. Karvounas et al., “5G on the Horizon: key challenges for the radio-access network,” *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 47–53, 2013.
- [2] T. Inoue, “5G standards progress and challenges,” in *Proceedings of the 2017 IEEE Radio and Wireless Symposium (RWS)*, pp. 1–4, IEEE, Phoenix, Ariz, USA, January 2017.
- [3] C.-X. Wang, F. Haider, X. Gao et al., “Cellular architecture and key technologies for 5G wireless communication networks,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.
- [4] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, “Cache in the air: exploiting content caching and delivery techniques for 5G systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131–139, 2014.
- [5] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, “Networks and devices for the 5G era,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 90–96, 2014.
- [6] N. Bhushan, J. Li, D. Malladi et al., “Network densification: the dominant theme for wireless evolution into 5G,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, 2014.
- [7] R. Arshad, H. Elsayy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, “Handover Management in 5G and Beyond: A Topology Aware Skipping Approach,” *IEEE Access*, vol. 4, pp. 9073–9081, 2016.
- [8] I. Hwang, B. Song, and S. Soliman, “A holistic view on hyperdense heterogeneous and small cell networks,” *IEEE Communications Magazine*, vol. 51, no. 6, pp. 20–27, 2013.
- [9] S. Liu, J. Wu, C. H. Koh, and V. K. N. Lau, “A 25 Gb/s/(km<sup>2</sup>) Urban wireless network beyond IMT-advanced,” *IEEE Communications Magazine*, vol. 49, no. 2, pp. 122–129, 2011.
- [10] Qualcomm Research, *LTE Rel-12 & Beyond*, 2012, <http://www.qualcomm.com/1000x/>.
- [11] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *Proceedings of the 5th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '09)*, pp. 1–12, ACM, December 2009.
- [12] L. Zhang, D. Estrin, J. Burke et al., “Named data networking (ndn) project,” Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC, 2010.
- [13] H. Wu, J. Li, T. Pan, and B. Liu, “A novel caching scheme for the backbone of Named data networking,” in *Proceedings of the 2013 IEEE International Conference on Communications, ICC 2013*, pp. 3634–3638, IEEE, June 2013.
- [14] M. Fiore, F. Mininni, C. Casetti, and C.-F. Chiasserini, “To cache or not to cache?” in *Proceedings of the IEEE INFOCOM*, pp. 235–243, IEEE, Rio de Janeiro, Brazil, April 2009.
- [15] H. Xie, G. Shi, and P. Wang, “TECC: towards collaborative in-network caching guided by traffic engineering,” in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 2546–2550, Orlando, Fla, USA, March 2012.
- [16] Y. Zeng and X. Hong, “A caching strategy in mobile ad hoc named data network,” in *Proceedings of the 2011 6th International ICST Conference on Communications and Networking in China, CHINACOM 2011*, pp. 805–809, China, August 2011.
- [17] H. Ahlehagh and S. Dey, “Video caching in radio access network: impact on delay and capacity,” in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '12)*, pp. 2276–2281, IEEE, Shanghai, China, April 2012.
- [18] E. Bastug, M. Bennis, and M. Debbah, “Cache-enabled small cell networks: Modeling and tradeoffs,” in *Proceedings of the 2014 11th International Symposium on Wireless Communications Systems, ISWCS 2014*, pp. 649–653, Barcelona, Spain, August 2014.
- [19] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, “FemtoCaching: Wireless video content delivery through distributed caching helpers,” in *Proceedings of the IEEE Conference on Computer Communications, INFOCOM 2012*, pp. 1107–1115, IEEE, March 2012.
- [20] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, “Cache-aware user association in backhaul-constrained small cell networks,” in *Proceedings of the 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt 2014*, pp. 37–42, May 2014.
- [21] T. L. Saaty, “The analytic hierarchy process,” in *Proceedings of the Second International Seminar on Operational Research in the Basque Provinces*, vol. 4, 29, pp. 189–234, 1996.
- [22] ndnSIM, <http://ndnsim.net/1.0/index.html>.

## Research Article

# Channel Characteristics of Rail Traffic Tunnel Scenarios Based on Ray-Tracing Simulator

Jinmeng Zhao <sup>1</sup>, Lei Xiong <sup>1</sup>, Danping He,<sup>1</sup> and Jiadong Du<sup>2</sup>

<sup>1</sup>State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China

<sup>2</sup>The China Academy of Information and Communications Technology (CAICT), China

Correspondence should be addressed to Lei Xiong; [lxiong@bjtu.edu.cn](mailto:lxiong@bjtu.edu.cn)

Received 30 March 2018; Accepted 27 June 2018; Published 1 August 2018

Academic Editor: Sana Salous

Copyright © 2018 Jinmeng Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The tunnel scenario is a major rail communication scenario. In this paper, the radio channel characteristics of tunnel scenarios with different carrier frequencies, different distances between the transmitter (Tx) and receiver (Rx), and cross sections are simulated with a ray-tracing tool. Key parameters such as path loss, Rician K-factor, root mean square (RMS) delay spread, and angular spread are studied. According to the results, higher frequencies introduce larger path loss and the presence of the vehicle body increases the path loss by about 35 dB in the scenario; at the same time it will also cause the fluctuation and instability of the path loss. Besides, the influence of reflections from the side walls is significant on radio propagation. The channel experiences more severe fading in a narrow tunnel compared with others.

## 1. Introduction

The rail traffic communications have experienced a rapid development recently, such as high-speed railway, municipal railway, and urban railway system. The rail traffic scenario is indispensable for both private and public mobile communications. It is widely agreed that the wireless channel model is significant to carry out mobile communications research, system development, and network deployment, and so on [1]. Currently, the Long-Term Evolution Railway (LTE-R) system has been recommended to replace the global system for mobile communication railway (GSM-R) system for high-speed train (HST) communication system as a part of intelligent transportation systems (ITS) [2, 3]. In addition, the research on 5th-Generation wireless systems (5G) based on high-speed railway (HSR) has become a trend to meet the need of transmission capabilities [4–6]. Thus the study on channel models in rail traffic system for new communication system is indispensable.

The wireless channel research needs to be carried out for various typical rail traffic scenarios (viaducts, tunnels, cuttings, etc.), considering the significant differences between the rail traffic scenarios and the public network scenarios [7].

Field test and ray tracer (RT) are two well-known methods for implementing channel characterization. Field test in the field of rail traffic is hard, costly, and long-term. RT provides a means of accurately prediction of the wave propagation which is timesaving and convenient. Therefore, RT is widely used in channel modeling for confined environments.

Tunnel scenario is one of the most common scenarios, especially in the mountainous and hilly areas. Considering the unique construction, the wireless propagation in tunnel scenarios is different from other HST scenarios which attracts a lot of research interests [8–10]. Leaky cables and Distributed Antenna System (DAS) [11, 12] are mainly two promising methods to provide radio coverage in tunnels. Several tunnel channel models were proposed in recent years, such as ray-tracing model, multimode model, and propagation-graph theory based model [13, 14]. A real multipath propagation model for radio transmission in typical rectangular subway tunnel was presented in [15] using RT to analyze Doppler spread. The three-dimensional (3D) models of six scenario modules for mmWave and THz train-to-infrastructure channels were defined and constructed for the first time in [16], which considered reality obstacle objects. Based on the wideband measurements conducted in the tunnel scenario

by using the mobile hotspot network system, the authors in [17] explored key channel characteristics in different HSR scenarios by 3D RT. Besides, suggestions on symbol rate, subframe bandwidth, and polarization configuration were provided to guide the 5G mmWave communication system design in typical HSR scenarios.

The aforementioned research summarized that the majority of existing works contribute to modeling methods based on large-scale parameters. However, previous models were analyzed and compared based on field test. The measurements results such as channel impulse response (CIR) are mixed with a variety of factors. In addition, measurements focusing on narrowband single-input single-output (SISO) systems at low frequency band result in missing multipath parameters such as angular spread and Doppler shift. There is still an urgent demand for an accurate complete channel model considering large-scale parameters, small-scale parameters, and spatial parameters [18, 19] for multitype tunnel scenarios in SISO systems or multiple-input multiple-output (MIMO) systems. Based on RT, channel parameters such as the path loss, multipath delay, and angular spread can be obtained exactly and analyzed. This paper provides the simulation results in different type of tunnels at 1.8 GHz. Optimized deployment recommendations are provided for MIMO systems and antenna configuration.

The rest of the paper is organized as follows. The existing channel models for the tunnel scenario of 3GPP are introduced in Section 2. The defined tunnel scenarios and simulation setup are presented in Section 3. Key parameters such as path loss, Rician K-factor, root mean square (RMS) delay spread, and angular spread in tunnels with different cross sections at different frequencies are analyzed in Section 4. Conclusions are drawn in Section 5.

## 2. Channel Models

3GPP TSG-RAN WG4 in RAN#66 meeting defined four typical high-speed railway scenarios. A variety of standard channel models in tunnel scenarios were adopted by 3GPP [20–23], like the single tap channel, time delay line (TDL) model, etc.

Scenarios 2c and 2g are corresponding to tunnel scenarios covered by leaky cables. The coverage of the Evolved Node B (eNB) is extended by deploying leaky cables in the 2g and 2c scenarios, which is shown in Table 1. In scenario 2g, the User Equipment (UE) communicates directly with the leaky cable in the tunnel. The signal is severely attenuated due to the shielding of the train body in this case. In scenarios 2c, relays are set up on the train. The UE achieves two-hop communication with the leaky cable through relay as shown in Figure 1. Since the transmit power gradually attenuates along the leaky cable, RF amplifiers will typically be applied to ensure adequate signal strength. The propagation channel between the in-car relay and the UE does not involve high-speed movement.

3GPP R4-154106 proposes a one-tap channel model for leaky cables deployment in the tunnel scenario as shown in Figure 2, where  $D_0$  is the distance between two neighbor

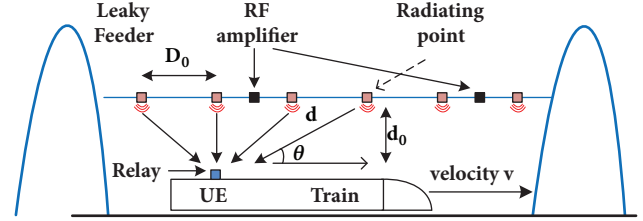


FIGURE 1: Leaky cable coverage with relay.

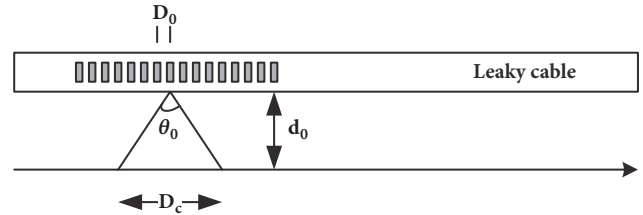


FIGURE 2: Detail illustration of leaky cable.

radiating points; the radiation angle of radiating point is denoted as  $\theta_0$ . It can be assumed that the receiver (Rx) is covered by 8-9 radiation spots. The RF receive signal is

$$\begin{aligned} r(t) &= \sum_k x(t) \sqrt{P_k} e^{j2\pi(f_c - f_{d,k})t} + z(t) \\ &= x(t) e^{j2\pi f_c t} \sum_k \sqrt{P_k} e^{-j2\pi f_{d,k} t} + z(t) \end{aligned} \quad (1)$$

where  $P_k$  denotes the normalized received power of the  $k_{th}$  radiation spot,  $f_{d,k}$  is the respective Doppler shift,  $z(t)$  is the received noise, and  $x(t)$  is the transmitted signal. The power of the fading gain can be modeled as a Rician distribution based on the central limit theorem.

However, this model is only a simplified model. It ignores the reflection component and the scattering component which should be considered in a limited space. Besides, the parameters, such as tunnel size, cross section, internal electromagnetic (EM) properties of tunnel walls, surface roughness, and antenna polarization mode, affect the channel characteristics. RT views radio waves reflected by surfaces of the tunnel. It solves the problem of large workload and poor applicability in field test and also compensates for the inability of the traditional model to provide specific wireless channel parameters, such as amplitude, delay, Doppler shift, and angular spread, which is of great significance for wireless communication systems. It is an effective modeling method to establish a 3D channel model of the rail communication scenarios and analyze the radio signal propagation under different coverage patterns and multiple fading characteristics.

## 3. Scenario Definition and Simulation Setup

**3.1. Tunnel Scenarios Definition.** Given the different geological conditions, the cross section of the tunnel varied in reality. Four typical shapes of cross sections: rectangular, arched, long arched (combined rectangular and semicircular), and



TABLE 1: Tunnel scenarios by leaky cables.

Scenarios	Description	Notation
2c	The tunnel is equipped with a leaky cable that communicates with the train's roof RP and with the UE through the RP	Two hop
2g	Laying leaky cables in the tunnel to achieve direct coverage of the UE	One hop

semicircular, are shown in Figure 3. The side length of rectangular is 6 m in *tunnel a*; the height of arched is 6 m, and the center of circle is 2 m high above the ground in *tunnel b*; *tunnel c* contains a rectangular whose side length is 4 m and a semicircular whose radius is 2 m; the radius of semicircular is 6 m in *tunnel d*. The number of tunnel surface in the simulation is 4, 10, 9, and 13, respectively. Figure 4 shows the details of the simplified tunnel model referring to the metro tunnel project boundary map provided by the China Railway Fourth Survey and Design Institute Group Co., Ltd. The train model is shown in Figure 5. The 3D tunnel models in this work are built by Google SketchUp. These models are considered in the following simulation analysis.

**3.2. Parameters Setup.** The simulation parameters should be carefully determined to ensure that the simulation results are accurate and effective and make simulation energy-friendly and timesaving. RT supports a variety of radio wave propagation mechanisms, such as direction, reflection, scattering, diffraction, and transmission. Reflection order means that rays experience up to several times specular reflection from transmitter (Tx) to Rx. The appropriate reflection order is necessary not only to reflect the actual reflectance, but also to avoid a longer simulation time.

Based on the above ideas, the channel characteristics under different reflection orders are simulated and analyzed in tunnel scenario in Figure 4 with all radio wave propagation mechanisms. The key channel characteristics, such as Rician K-factor, delay spread, angle spread of angle of arrival (AOA), and angle of departure (AOD), are compared, respectively, under different reflection orders range from 6<sup>th</sup>-9<sup>th</sup>. When the reflection order is set to higher than 6<sup>th</sup>, there is rarely a change. In addition, we define

$$PL_{error}(i, j) = \frac{|PL_{i,f} - PL_{j,f}|}{PL_{j,f}} \quad (2)$$

where the  $PL_{i,f}$  is the path loss when the frequency is  $f$  and the reflection order is  $i$ . The  $PL_{error}(i, j)$  shows the difference rate between  $i^{th}$  and  $j^{th}$  order in a series of frequencies. The CDF of  $PL_{error}$  is shown in Figure 6. The means of  $PL_{error}(8, 9)$ ,  $PL_{error}(7, 8)$ , and  $PL_{error}(6, 7)$  are  $4.26 \times 10^{-12}$ ,  $1.9 \times 10^{-2}$ , and 0.18, respectively. The  $PL_{error}(8, 9)$  is the smallest, ranging from  $6.41 \times 10^{-13}$  to  $1.72 \times 10^{-11}$ . The  $PL_{error}(7, 8)$  ranges from  $4.97 \times 10^{-5}$  to 0.12. The  $PL_{error}(6, 7)$  ranges from  $1.73 \times 10^{-4}$  to 0.53, which is larger than others. As a result, the path loss up to higher than 8<sup>th</sup> order is hardly change. Thus the 8<sup>th</sup> order is adopted in simulation due to a tradeoff

TABLE 2: Parameters setup.

Parameter	Value
Frequency	1.8 GHz / 5.8 GHz
Transmitting power	25 dBm
Antenna(Tx, Rx)	Omnidirectional antenna (vertically polarized)
Tx location	(0, 0, 3)
Rx location	(x, 0, 5)
Reflection order	8
Bandwidth	10 MHz
Resolution	1 MHz

between computational complexity and precision as shown in Table 2. Considering the fact that the tunnel scenarios are long straight tunnels covered by concrete with no vents and pipes and other obstacles, diffraction is irrelevant here as well as transmission because the Tx is not fixed in the wall.

Detailed parameters setup is shown in Table 2. The frequency is set to 1.8 GHz and 5.8 GHz in order to compare channel characteristics at different frequencies. Signal propagates with 25 dBm of transmitting power. Tx is fixed at one end of the tunnel with the height of 3 m. Rx moves along the x-axis with the height of 5 m. Both of them use vertically polarized omnidirectional antennas. The channel bandwidth  $B = 10$  MHz, and the system time domain resolution  $\Delta t = 1 / B$ . In order to better observe the path with the maximum excess delay max, the number of simulated frequency points  $N_f$  should be determined enough [24]:

$$\tau_{max} \leq N_f \times \Delta t \quad (3)$$

in which, in other words, the frequency domain resolution  $\Delta f = B / N_f$  should be set sufficiently.

## 4. Simulation Results and Analyses

In this section, the simulation results of tunnel channel characteristics in different cross sections and different frequencies are presented and analyzed. Furthermore, suggestions on optimized deployment are provided.

**4.1. Large-Scale Parameters.** The reliable large-scale channel models are essential to network deployment and optimization. Typically, the path loss is expressed as

$$PL(d) = A + 10n \lg(d) + X \quad (4)$$

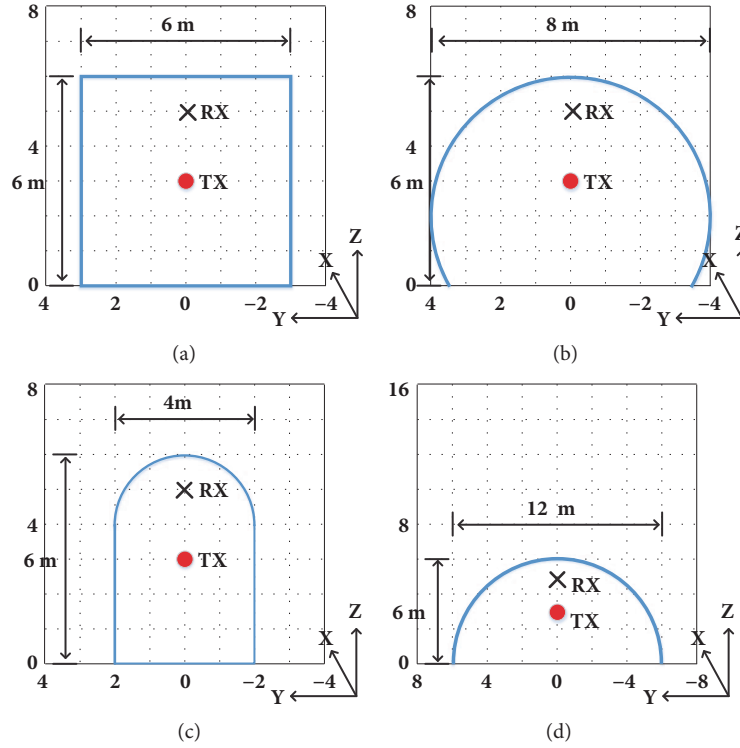


FIGURE 3: Regular cross sections for tunnels. (a) Rectangular. (b) Arched. (c) Long arched. (d) Semicircular.

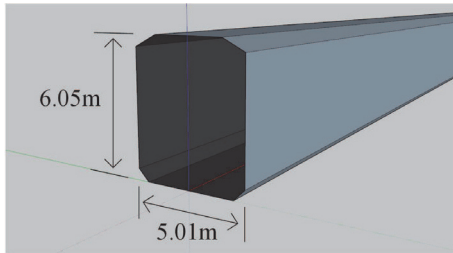


FIGURE 4: Simplified tunnel model (actual scenario).

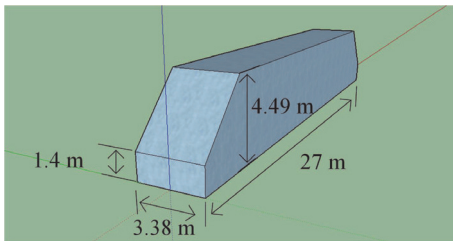


FIGURE 5: Train model for HST.

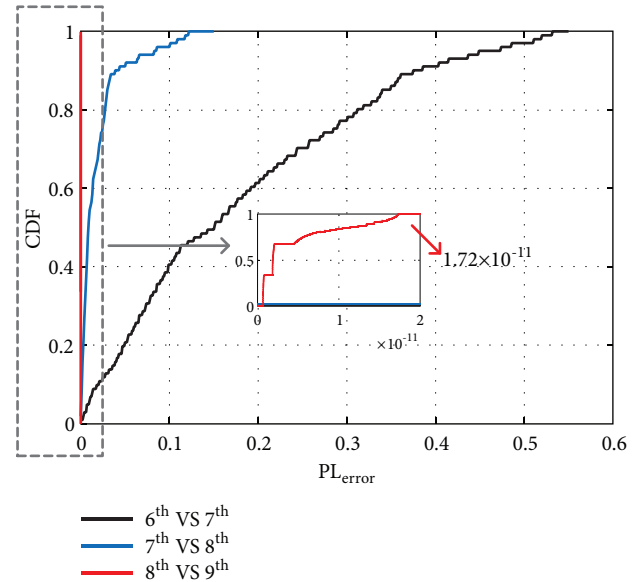


FIGURE 6: CDF of  $PL_{error}$ .

where the  $PL(d)$  is path loss without small-scale fading and it is a function of the distance between Tx and Rx [25].  $n$  is path loss exponent, and  $X$  (shadow fading) is the zero-mean Gaussian random variable with standard deviation  $\sigma$ . Figure 7 shows the path loss at 1.8 GHz in different cross sections. In the 0-190 m section, the path loss is nearly the same. However

in the 190-500 m section, the path loss in *tunnel c* has a higher path loss than other tunnels. The path loss in *tunnel c* can be divided into two phases, and path loss exponents are 1.96 and 3.62, respectively, which means that there is a breakpoint ( $d_0 = 190.10$  m) in this scenario. These tunnel cross sections have the same height, and the width varies in shape and size. The path loss fitting result given in Table 3

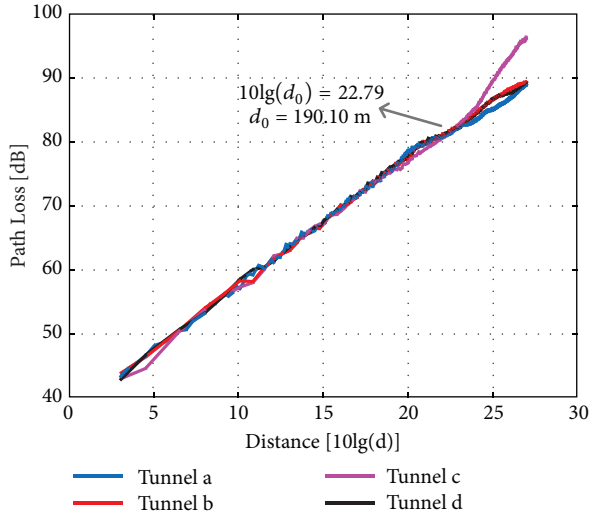


FIGURE 7: Path loss in different tunnel cross sections at 1.8 GHz.

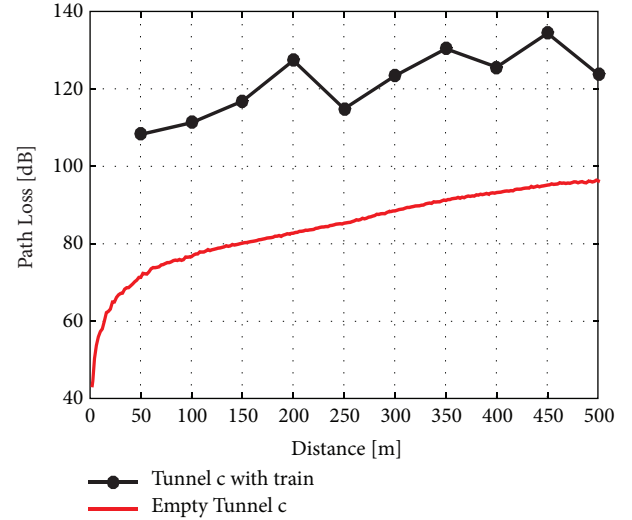
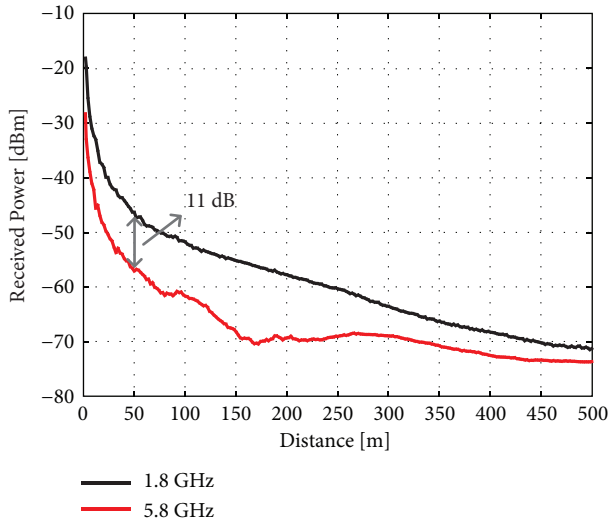


FIGURE 9: Path loss in empty tunnel and tunnel with train.


 FIGURE 8: Received power at different frequencies in *tunnel c*.

indicates that the lateral shape has an important impact on the path loss in the empty tunnel with strong line of sight (LOS) path. The path loss exponent in *tunnel c* is greater than 2 because of the existing breakpoint. The path loss exponents are 1.77, 1.84, and 1.88 in *tunnel a*, *d*, and *b*, respectively. Compared with the measurement results in straight arched tunnel [26, 27] and rectangular tunnel [28–30], the value of path loss exponent 1.96 in *tunnel c* is similar to 1.40–2.03 at 954–2000 MHz in straight arched tunnel. Meanwhile, path loss exponent 1.77 in *tunnel a* is similar to 1.65–1.94 at 945–2650 MHz in rectangular tunnel. Simulation results in this paper are basically consistent with the measurement results in previous work.

Figure 8 shows the received power at different frequencies in *tunnel c*. As the frequency increases, the received power decreases. When the distance between Tx and Rx equals 50 m, the received power at 1.8 GHz is 11 dB higher than that at 5.8 GHz. Received power has larger fluctuation, and deeper

TABLE 3: Path loss fitting result at 1.8 GHz.

Parameter	A	n	$\sigma$
Tunnel a	41.29	1.77	0.74
Tunnel b	39.46	1.88	0.45
Tunnel c	31.82	2.30	1.60
Tunnel d	41.29	1.84	0.60

TABLE 4: Path loss fitting result in *tunnel c*.

Parameter	A	n	$\sigma$
1.8 GHz	31.82	2.31	1.60
5.8 GHz	51.51	1.78	1.45

fading is prone to occur in higher frequency. Based on the similar simulation scenario and parameters setup, the same conclusion is also reflected on the simulation results at three different typical carrier frequencies, i.e., 900 MHz, 2.45 GHz, and 5.75 GHz in [31]. Table 4 shows the fitting parameters at different frequencies. The path loss exponent is 2.31 at 1.8 GHz, while the path loss exponent is less than 2 at higher frequency. Results confirm that the frequency band has a strong impact on the wireless propagation.

For the tunnel with train, the path loss result in selected locations is shown in Figure 9. Rx is fixed on the head of train when the train gradually moves away from the Tx. The obstruction and reflection of the train make the received signal experience longer distance and greater attenuation. The corresponding number of arriving rays reduces sharply. The path loss increases by 35.58 dB with the presence of the vehicle body. The result shows that the path loss in the empty tunnel scenario rises more steadily. Suggestions are given that the antenna deployment in reality scenarios should make sure of the existence of LOS path between Tx and Rx. Moreover, antenna placement is selected carefully to make the blocking effect of the train as small as possible for which further researches are needed.

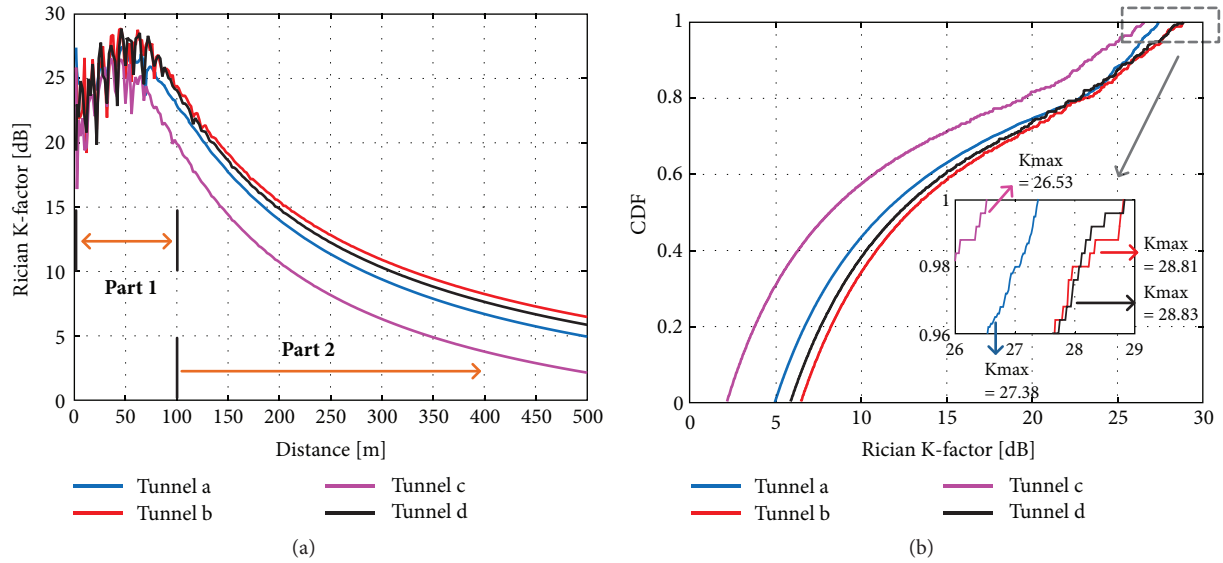


FIGURE 10: (a) Rician K-factor at 1.8 GHz. (b) CDF of Rician K-factor at all the snapshots.

TABLE 5: Small scale parameters.

Parameter	Mean K-factor [dB]	Percentage of $K < 25$ dB	Mean delay spread [ns]	95 % of delay spread range [ns]
Tunnel a	13.7	88%	0.47	0.31-0.76
Tunnel b	14.98	86%	0.37	0.09-0.75
Tunnel c	10.82	94.8%	0.33	0.17-0.66
Tunnel d	14.46	87%	0.40	0.16-0.80

**4.2. Small-Scale Parameters.** Small-scale parameters have an important impact in wireless communication system design and analysis. Rician K-factor, RMS delay spread, and angular spread are key parameters to describe the small-scale characteristics. Table 5 shows the K-factor and RMS delay spread in different tunnel scenarios.

**4.2.1. Rician K-Factor.** Rician K-factor is defined as the ratio of the power of LOS path to the power of non-line-of-sight (NLOS) paths [32]. Figure 10 shows K-factor at 1.8 GHz and CDF of K-factor at all the snapshots. Figure 10(a) depicts that the simulation section can be divided into two parts; K-factor fluctuates severely in part one (0-100 m) and then decreases smoothly in part two (100-500 m). In the first 50 m, the fading is weak, but, after that, the fading is increasingly serious whatever the tunnel scenario is. Table 5 demonstrates that 94.8% of K-factor in all the snapshots in *tunnel c* is lower than 25 dB, but there is about 88% of K-factor lower than 25 dB in other tunnel scenarios as shown in Figure 10(b). The mean K-factor of tunnel scenarios is around 13.49 dB larger than the result 11.5 dB of empirical cluster characteristics extracted for tunnel scenarios at 2.14 GHz [33]. Since the path loss increases with frequency, LOS path experiences a deeper fading at 2.14 GHz. In conclusion, channel characteristics in *tunnel c* experience a severe fading than others.

**4.2.2. RMS Delay Spread.** Figure 11 shows that RMS delay spread at 1.8 GHz and CDF of RMS delay spread at all the snapshots. *Tunnel a* is the rectangular tunnel ( $W6\text{ m} \times H6\text{ m}$ ), and *tunnel c* is the long arched tunnel ( $W4\text{ m} \times H6\text{ m}$ ). The most apparent difference between *tunnel a* and *tunnel c* is their widths. RMS delay spread in *tunnel a* is larger than that in *tunnel c* along the simulation section. In Table 5, we notice that RMS delay spread is lower than 0.8 ns in at least 95% at all the snapshots, and the maximum RMS delay spread is 1.18 ns. The results in Table 5 illustrate that 95% of the RMS delay spreads at all the snapshots in *tunnel a* and *tunnel c* are 0.45 ns and 0.49 ns, respectively, which is narrow compared with other scenarios. On the other hand, mean RMS delay spread is the largest in *tunnel a*, and the weighted average of the delay spread in *tunnel c* is the smallest. As a result, the channel in *tunnel c* experiences a more stable fading. The RMS delay spread varies in ns because of fewer obstacles in empty tunnel, which is similar to the measurement results of delay spread 2-27 ns in mine tunnels [34].

**4.3. Spatial Parameters.** The RMS angular spreads in *tunnel c* and CDF of angular spreads are shown in Figure 12. ASA, ESA, ASD, and ESD are angular spreads of the azimuth angle of arrival, the elevation angle of arrival, the azimuth angle of departure, and elevation angle of departure, respectively.

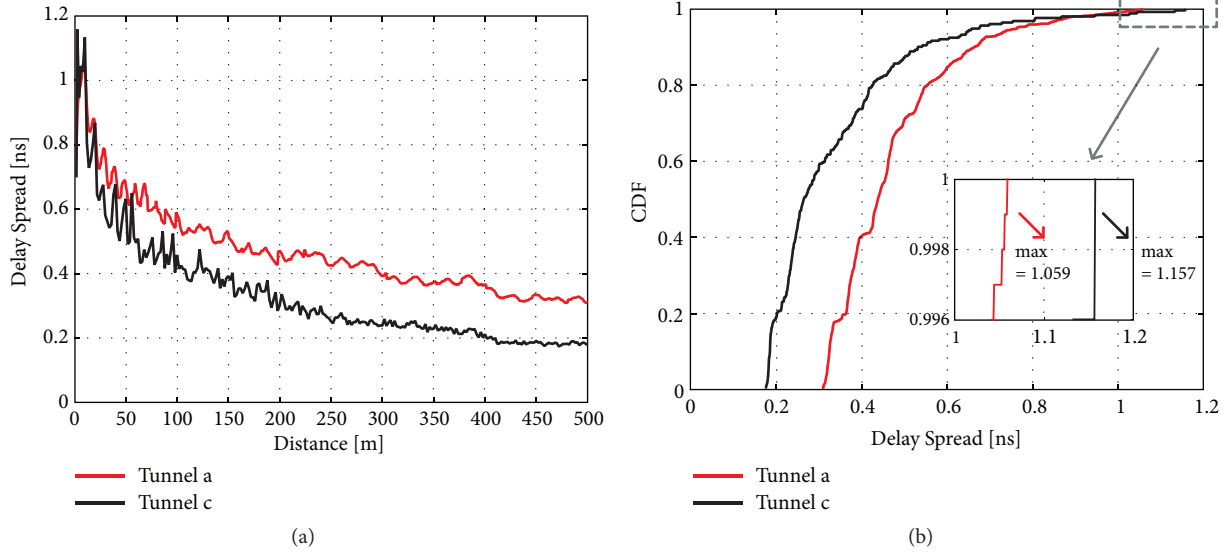


FIGURE 11: (a) RMS delay spread at 1.8 GHz. (b) CDF of RMS delay spread at all the snapshots.

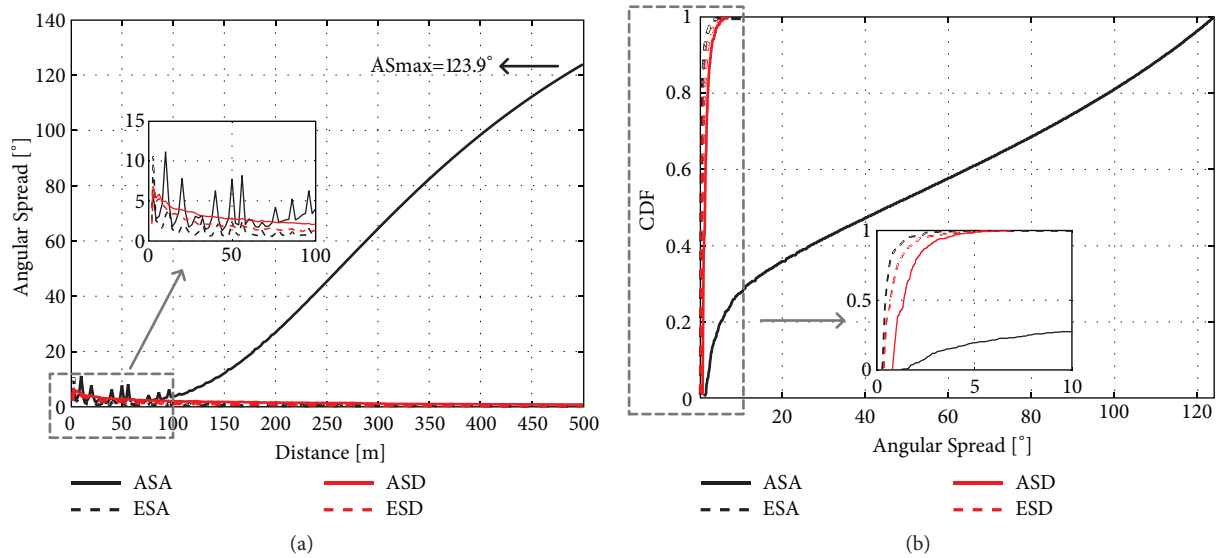


FIGURE 12: Tunnel c scenario. (a) RMS angular spread at 1.8 GHz. (b) CDF of RMS angular spread at all the snapshots.

Figure 11 illustrates that ESA, ASD, and ESD decrease slowly as the Rx moves far away from Tx, and they are lower than  $10.20^\circ$  at all the snapshots. The means of ASA, ESA, ASD, and ESD are  $51.25^\circ$ ,  $0.67^\circ$ ,  $1.63^\circ$ , and  $0.95^\circ$ , respectively. 99% of ESA, ASD, and ESD are less than  $5^\circ$ . However, ASA varies differently compared with others. It gradually increases and the max ASA is  $123.90^\circ$  when the distance between Tx and Rx reaches 500m. Angular spread in straight tunnel scenario in work [17] varies similarly, in which ASD becomes larger as the distance of Tx and Rx increases and others decrease at the same time. ASA is greater than others in the majority of the snapshots, which indicates that the impact of reflections from the side walls is significant on radio propagation. Multiantennas can be considered to implement

a MIMO system so that the system capacity can be increased by achieving diversity and multiplexing. Besides, directional antennas can be deployed to drop the ASA and reduce the impact of Doppler spread on the channel.

Simulation results reveal that ASA, ESA, ASD, and ESD vary similarly in tunnels b, c, and d. The ASA is compared in tunnels b, c, and d as shown in Figure 13. The maximum of ASA in tunnels b, c, and d is  $65.8^\circ$ ,  $123.9^\circ$ , and  $73.8^\circ$ , respectively. It is obvious that ASA in tunnel c is larger than that in tunnel d, whereas ASA in tunnel b is the lowest at the majority of snapshots. This observation indicates that spatial fading is affected by the cross section. Furthermore, the side walls have a more significant impact on the channel in a narrow tunnel than a wider one.

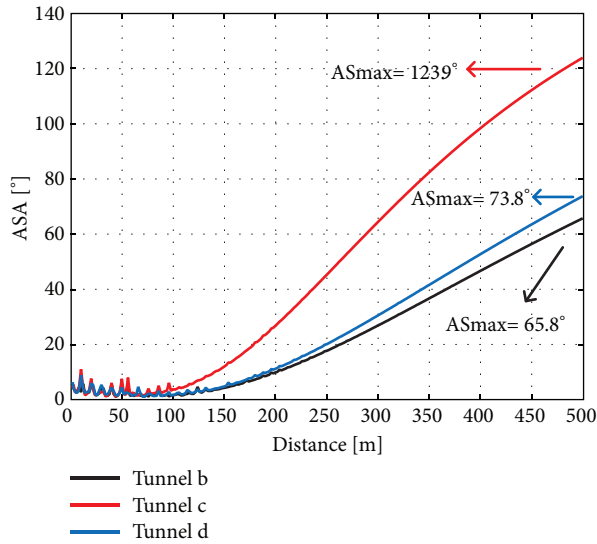


FIGURE 13: ASA in tunnels b, c, and d.

## 5. Conclusion

In this paper, the radio channel characteristics of several tunnel scenarios were simulated with RT. The channel characteristics in different carrier frequencies and tunnel cross sections were analyzed. The key parameters, such as path loss, Rician K-factor, RMS delay spread, and angular spread were simulated and analyzed. The presence of the vehicle body introduces additional 35 dB of the path loss, which leads to the fluctuation and instability of the channel. K-factor changes severely when the distance between Tx and Rx is smaller than 100 m and then decreases smoothly in far region. The channel has short delay spread (1.18 ns) due to strong LOS and the limited space. The side walls have a significant influence on radio propagation especially in a narrow tunnel. Therefore, the conclusion can be drawn that the channel experiences a severe and stable fading in long arched tunnels compared to other tunnel scenarios. The other parameters (i.e., Doppler spread, coherence time, antenna polarization, etc.) and leaky cable for coverage will be studied in the future work.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities (2018JBM079), the National Key Research and Development Program under

Grant 2016YFE0200900, the National Science and Technology Major Project (2015ZX03001027-003), the National Key R&D Program of China under Grant no. 2016YFB1200100, the National Natural Science Foundation of China under Grant 61471030, the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2018ZZ006, and the Fundamental Research Funds for the Central Universities under Grant no. 2017YJS014.

## References

- [1] C.-X. Wang, A. Ghazal, B. Ai, Y. Liu, and P. Fan, "Channel measurements and models for high-speed train communication systems: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 974–987, 2015.
- [2] D. W. Matolak, M. Berbineau, D. G. Michelson, and C. Chen, "Future railway communications [Guest Editorial]," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 60–61, 2015.
- [3] J. Marais, J. Beugin, and M. Berbineau, "A Survey of GNSS-Based Research and Developments for the European Railway Signaling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2602–2618, 2017.
- [4] J. Zhang, Z. Zheng, Y. Zhang, J. Xi, X. Zhao, and G. Gui, "3D MIMO for 5G NR: Several Observations from 32 to Massive 256 Antennas Based on Channel Measurement," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 62–70, 2018.
- [5] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, no. 99, article 1, 2017.
- [6] C. T. Neil, M. Shafi, P. J. Smith, P. A. Dmochowski, and J. Zhang, "Impact of Microwave and mmWave Channel Models on 5G Systems Performance," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6505–6520, 2017.
- [7] A. Hrovat, G. Kandus, and T. Javornik, "A survey of radio propagation modeling for tunnels," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 658–669, 2014.
- [8] Y. Liu, A. Ghazal, C. Wang, X. Ge, Y. Yang, and Y. Zhang, "Channel measurements and models for high-speed train wireless communication systems in tunnel scenarios: a survey," *Science China Information Sciences*, vol. 60, no. 10, 2017.
- [9] C. Zhou, "Ray tracing and modal methods for modeling radio propagation in tunnels with rough walls," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 5, pp. 2624–2634, 2017.
- [10] K. Guan, B. Ai, Z. Zhong et al., "Measurements and Analysis of Large-Scale Fading Characteristics in Curved Subway Tunnels at 920 MHz, 2400 MHz, and 5705 MHz," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2393–2405, 2015.
- [11] K. Guan, Z. Zhong, B. Ai, and C. Briso-Rodriguez, "Statistic modeling for propagation in tunnels based on distributed antenna systems," in *Proceedings of the 2013 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, pp. 1920–1921, Orlando, FL, USA, July 2013.
- [12] C. Briso-Rodriguez, J. M. Cruz, and J. I. Alonso, "Measurements and modeling of distributed antenna systems in railway tunnels," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2870–2879, 2007.

- [13] J. A. Castiblanco, D. Seetharamdoo, M. Berbineau et al., "Surface impedance boundary conditions in time domain for guided structures of arbitrary cross section with lossy dielectric walls," *Institute of Electrical and Electronics Engineers. Transactions on Antennas and Propagation*, vol. 63, no. 3, pp. 1086–1097, 2015.
- [14] A. Hrovat, G. Kandus, and T. Javornik, "Four-slope channel model for path loss prediction in tunnels at 400 MHz," *IET Microwaves, Antennas & Propagation*, vol. 4, no. 5, pp. 571–582, 2010.
- [15] K. Guan, Z. D. Zhong, J. I. Alonso, and C. Briso-Rodríguez, "Measurement of distributed antenna systems at 2.4 GHz in a realistic subway tunnel environment," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 834–837, 2012.
- [16] K. Guan, X. Lin, D. He et al., "Scenario modules and ray-tracing simulations of millimeter wave and terahertz channels for smart rail mobility," in *Proceedings of the 2017 11th European Conference on Antennas and Propagation (EUCAP)*, pp. 113–117, Paris, France, March 2017.
- [17] D. He, B. Ai, K. Guan et al., "Channel Measurement, Simulation, and Analysis for High-Speed Railway Communications in 5G Millimeter-Wave Band," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–15, 2017.
- [18] L. Zhang, C. Briso, J. R. O. Fernandez et al., "Delay spread and electromagnetic reverberation in subway tunnels and stations," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 585–588, 2016.
- [19] J. Zhang, C. Pan, F. Pei, G. Liu, and X. Cheng, "Three-dimensional fading channel models: a survey of elevation angle research," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 218–226, 2014.
- [20] Nokia Networks, "Channel modelling for high speed leaky cable scenarios," in *Proceedings of the 3GPP TSG-RAN WG4 Meeting #76*, R4-154106, Beijing, China, 2015.
- [21] Huawei, "Channel models for the leaky cable," in *Proceedings of the 3GPP TSG-RAN WG4 Meeting #76*, HiSilicon, R4-154241, Beijing, China, 2015.
- [22] Huawei, HiSilicon, "Channel models for the leaky cable," in *3GPP TSG-RAN WG4 Meeting #75*, Japan, Fukuoka, 2015.
- [23] Huawei, "TP: Channel model for leaky cable in tunnel," in *Proceedings of the 3GPP TSG-RAN WG4 Meeting #76*, HiSilicon, R4-154242, Beijing, China, 2015.
- [24] S. Priebe, *towards THz communications: propagation studies, indoor channel modeling and interference investigations*, Shaker Verlag, 2013.
- [25] D. He, J. Yang, K. Guan et al., "Ray-tracing simulation and analysis of propagation for 3GPP high speed scenarios," in *Proceedings of the 2017 11th European Conference on Antennas and Propagation (EUCAP)*, pp. 2890–2894, Paris, France, March 2017.
- [26] *Ningde high-speed railway tunnel covering research report*, GCI Science Technology Co., Ltd, Guangzhou, China, Tech. Rep, 2008.
- [27] N. Wang, "A uniform path-loss model by power balance theory (PBT) and its application on tunnels," in *Proceedings of the 2010 9th International Symposium on Antennas Propagation and EM Theory, ISAPE 2010*, pp. 481–484, China, December 2010.
- [28] D. Didascalou, J. Maurer, and W. Wiesbeck, "Subway tunnel guided electromagnetic wave propagation at mobile communications frequencies," *IEEE Transactions on Antennas and Propagation*, vol. 49, no. 11, pp. 1590–1596, 2001.
- [29] S. Shinozaki, M. Wada, A. Teranishi, H. Furukawa, and Y. Akaiwa, "Radio propagation characteristics in subway platform and tunnel in 2.5 GHz band," in *Proceedings of the 1995 6th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'95. Part 3 (of 3)*, pp. 1175–1179, September 1995.
- [30] M. Choi, D. Kim, H. Jo, J. Yook, and H. Park, "Path-loss characteristics in subway tunnels at 2.65 GHz," *Microwave and Optical Technology Letters*, vol. 48, no. 2, pp. 383–386, 2006.
- [31] Y. Zhang, Y. Liu, J. Sun, C. Wang, and X. Ge, "Impact of Different Parameters on Channel Characteristics in a High-Speed Train Ray Tracing Tunnel Channel Model," in *Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Sydney, NSW, June 2017.
- [32] A. F. Molisch, *Wireless Communications*, IEEE-Wiley, 2nd edition, 2011.
- [33] X. Cai, X. Yin, X. Cheng, and A. Perez Yuste, "An Empirical Random-Cluster Model for Subway Channels Based on Passive Measurements in UMTS," *IEEE Transactions on Communications*, vol. 64, no. 8, pp. 3563–3575, 2016.
- [34] R. He, Z. Zhong, B. Ai et al., "Propagation channel measurements and analysis at 2.4 GHz in subway tunnels," *IET Microwaves, Antennas & Propagation*, vol. 7, no. 11, pp. 934–941, 2013.

## Review Article

# A Survey on Machine Learning-Based Mobile Big Data Analysis: Challenges and Applications

Jiyang Xie <sup>1</sup>, Zeyu Song,<sup>1</sup> Yupeng Li,<sup>2</sup> Yanting Zhang,<sup>3</sup> Hong Yu,<sup>1</sup> Jinnan Zhan,<sup>2</sup> Zhanyu Ma <sup>1</sup>, Yuanyuan Qiao,<sup>3</sup> Jianhua Zhang <sup>2</sup>, and Jun Guo<sup>1</sup>

<sup>1</sup>Pattern Recognition and Intelligent Systems Lab., Beijing University of Posts and Telecommunications, Beijing, China

<sup>2</sup>State Key Lab. of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

<sup>3</sup>Center for Data Science, Beijing University of Posts and Telecommunications, Beijing, China

Correspondence should be addressed to Zhanyu Ma; mazhanyu@bupt.edu.cn and Jianhua Zhang; jhzhang@bupt.edu.cn

Received 9 April 2018; Accepted 7 June 2018; Published 1 August 2018

Academic Editor: Liu Liu

Copyright © 2018 Jiyang Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper attempts to identify the requirement and the development of machine learning-based mobile big data (MBD) analysis through discussing the insights of challenges in the mobile big data. Furthermore, it reviews the state-of-the-art applications of data analysis in the area of MBD. Firstly, we introduce the development of MBD. Secondly, the frequently applied data analysis methods are reviewed. Three typical applications of MBD analysis, namely, wireless channel modeling, human online and offline behavior analysis, and speech recognition in the Internet of Vehicles, are introduced, respectively. Finally, we summarize the main challenges and future development directions of mobile big data analysis.

## 1. Introduction

With the success of wireless local access network (WLAN) technology (a.k.a. Wi-Fi) and the second/third/fourth generation (2G/3G/4G) mobile network, the number of mobile phones, which is 7.74 billion, 103.5 per 100 inhabitants all over the world in 2017, is rising dramatically [1]. Nowadays, mobile phone can not only send voice and text messages, but also easily and conveniently access the Internet which has been recognized as the most revolutionary development of mobile Internet (M-Internet). Meanwhile, worldwide active mobile-broadband subscriptions in 2017 have increased to 4.22 billion, which is 9.21% higher than that in 2016 [1]. Figure 1 shows the numbers of mobile-cellular telephone and active mobile-broadband subscriptions of the world and main districts from 2010 to 2017. The numbers which are up to the bars are the mobile-cellular telephone or active mobile-broadband subscriptions (million) in the world of the year which increase each year. Under the M-Internet, various kinds of content (image, voice, video, etc.) can be sent and received everywhere and the related applications emerge to satisfy people's requirements, including working, study,

daily life, entertainment, education, and healthcare. In China, mobile applications giants, *i.e.*, Baidu, Alibaba, and Tencent, held 78% of M-Internet online time per day in apps which was about 2,412 minutes in 2017 [2]. This figure indicates that M-Internet has entered a rapid growth stage.

Nowadays, more than 1 billion smartphones are in use and producing a great quantity of data every day. This situation brings far-reaching impacts on society and social interaction and increases great opportunities for business. Meanwhile, with the rapid development of the Internet-of-Things (IoT), much more data is automatically generated by millions of machine nodes with growing mobility, for example, sensors carried by moving objects or vehicles. The volume, velocity, and variety of these data are increasing extremely fast, and soon they will become the new criterion for data analytics of enterprises and researchers. Therefore, mobile big data (MBD) has been already in our lives and is being enriched rapidly. The trend for explosively increased data volume with the increasing bandwidth and data rate in the M-Internet has followed the same exponential increase as Moore's Law for semiconductors [3]. The prediction [2] about the global data volume will grow up to 47 zettabytes



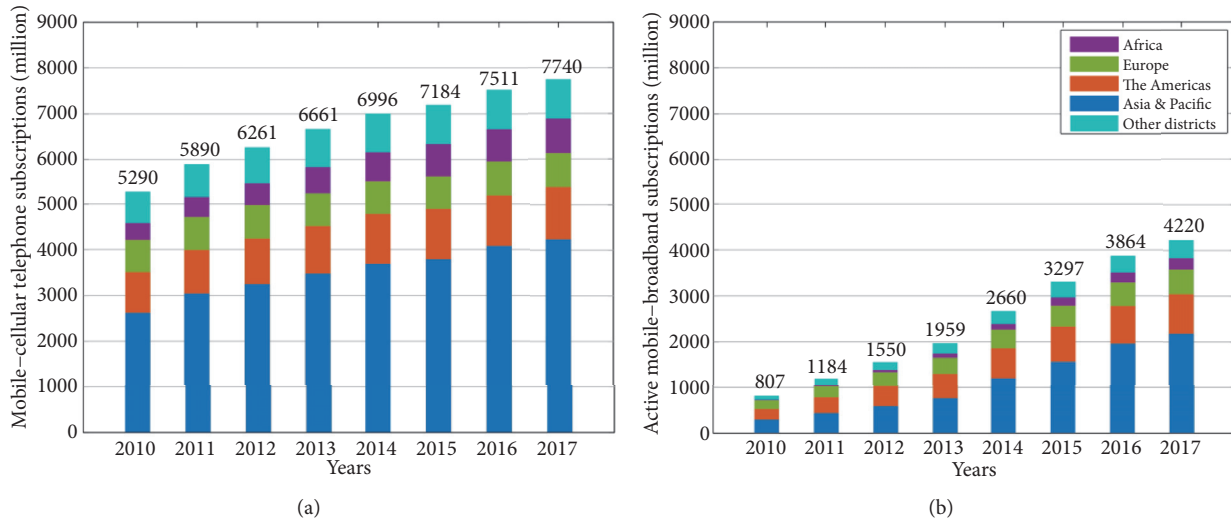


FIGURE 1: Mobile-cellular telephone subscriptions (million) in (a) and active mobile-broadband subscriptions (million) in (b) of the world and main districts [1].

(1 zettabyte =  $1 \times 10^{21}$  bytes) by 2020 and 163 zettabytes by 2025. For M-Internet, 3.7 exabytes (1 exabyte =  $1 \times 10^{18}$  bytes) data have been generated per month from the mobile data traffic in 2015 [4], 7.2 exabytes in 2016 [5], 24 exabytes by 2019 on forecasting [5], and 49 exabytes by 2021 on forecasting [5]. According to the statistical and prediction results, a concept called MBD has appeared.

The MBD can be considered as a huge quantity of mobile data which are generated from a massive number of mobile devices and cannot be processed and analyzed by a single machine [6, 7]. MBD is playing and will play a more important role than ever before by the popularization of mobile devices including smartphones and IoT gadgets especially in the era of 4G and the forthcoming the fifth generation (5G) [4, 8].

With the rapid development of information technologies, various data generated from different technical fields are showing explosive growth trends [9]. Big data has broad application prospects in many fields and has become important national strategic resources [10]. In the era of big data, many data analysis systems are facing big challenges as the volume of data increases. Therefore, analysis for MBD is currently a highly focused topic. The importance of MBD analysis is determined by its role in developing complex mobile systems which supports a variety of intelligently interactive services, for example, healthcare, intelligent energy networks, smart buildings, and online entertainments [4]. MBD analysis can be defined as mining terabyte-level or petabyte-level data collected from mobile users and wireless devices at the network-level or the app-level to discover unknown, latent, and meaningful patterns and knowledge with large-scale machine learning methods [11].

Present requirements of MBD are based on software-defined in order to be more scalable and flexible. M-Internet environment in the future will be even more complex and interconnected [12]. For this purpose, data centers of MBD need to collect user statistics information of millions of users

and obtain meaningful results by proper MBD analysis methods. For the decreasing price of data storage and widely accessible high performance computers, an expansion of machine learning has come into not only theoretical researches, but also various application areas of big data. Even though, there is a long way to go for the machine learning-based MBD analysis.

Machine learning technology has been used by many Internet companies in their services: from web searches [13, 14] to content filtering [15] and recommendation [16, 17] on online social communities, shopping websites, or content distribution platforms. Furthermore, it is also frequently appearing in products like smart cellphones, laptop computers, and smart furniture. Machine learning systems are used to detect and classify objects, return most relevant searching results, understand voice commands, and analyze using habits. In recent years, big data machine learning has become a hot spot [18]. Some conventional machine learning methods based on Bayesian framework [19–22], distributed optimization [23–26], and matrix factorization [27] can be applied into the aforementioned applications and have obtained good performances in small data sets. On this foundation, researchers have always been trying to fill their machine learning model with more and more data [28]. Furthermore, the data we got is not only big but also has features such as multisource, dynamic and sparse value; these features make it harder to analyze MBD with conventional machine learning methods. Therefore, the aforementioned applications implemented with conventional machine learning methods have fallen in a bottleneck period for low accuracy and generalization. Recently, a class of novel techniques, called deep learning, is applied in order to make the effort to solve the problems and has obtained good performances [29]. Machine learning, especially deep learning, has been an essential technique in order to use big data effectively.

Most conventional machine learning methods are shallow learning structures with one or none hidden layers.

These methods performed well in practical use and were precisely analyzed theoretically. But when dealing with high-dimensional or complicated data, shallow machine learning methods show their weakness. Deep learning methods are developed to learn better representations automatically with deep structure by using supervised or unsupervised strategies [30, 31]. The features extracted by deep hidden layers are used for regression, classification, or visualization. Deep learning uses more hidden layers and parameters to fit functions which could extract high level features from complex data; the parameters will be set automatically using large amount of unsupervised data [32, 33]. The hidden layers of deep learning algorithms help the model learn better representation of data; the higher layers learn specific and abstract features from global features learned by lower layers. Many surveys show that nonlinear feature extractors that are linked up as stacks such as deep learning methods always perform better in machine learning tasks, for example, a more accurate classification method [34], better learning of data probabilistic models [35], and the extraction of robust features [36]. Deep learning methods have proved useful in data mining, natural language processing, and computer vision applications. A more detailed introduction of deep learning is presented in Section 3.1.4.

Artificial Intelligence (AI) is a technology that develops theories, methods, techniques, and applications that simulate or extend human brain abilities. The research of observing, learning, and decision-making process in human brain motivates the development of deep learning, which was first designed aiming to emulate the human brain's neural structures. Further observation on neural signals processing and the effect on brain mechanisms [37–39] inspired the architecture design of deep learning network, using layers and neuron connections to generalize globally. Conventional methods such as support vector machines, decision trees, and case-based reasoning which are based on statistics or logic knowledge of human may fall short when facing complex structure or relationships of data. Deep learning methods can learn patterns and relationships from hidden layers and may benefit the signal processing study in human brain with visualization methods of neural network. Deep learning has attracted much attention from AI researchers recently because of its state-of-the-art performance in machine learning domains including not only the aforementioned natural language processing (NLP), but also speech recognition [40, 41], collaborative filtering [42], and computer vision [43, 44].

Deep learning has been successfully used in industry products which have access to big data from users. Companies in United States such as Google, Apple, Facebook, and Chinese companies like Baidu, Alibaba, and Tencent have been collecting and analyzing data from millions of users and pushing forward deep learning based applications. For example, Tencent YouTu Lab has developed identification (ID) card identification and bank card identification systems. These systems can read information from card images to check user information while registering and bank information while purchasing. The identification systems are based on deep learning model and large volume of user data provided by Tencent. Apple develops Siri, a virtual

intelligent assistant in iPhones, to answer questions about weather, location, news according to voice commands and dial numbers or send text messages. Siri also utilizes deep learning methods and uses data from apple services [45]. Google uses deep learning on Google translation service with massive data collected by Google search engine.

MBD contains a large variety of information of offline data and online real-time data stream generated from smart mobile terminals, sensors, and services and hastens various applications based on the advancement of data analysis technologies, such as collaborative filtering-based recommendation [46, 47], user social behavior characteristics analysis [48–51], vehicle communications in the Internet of Vehicles (IoV) [52], online smart healthcare [53], and city residents' activity analysis [6]. Although the machine learning-based methods are widely applied in the MBD fields and obtain good performances in real data test, the present methods still need to be further developed. Therefore, five main challenges facing MBD analysis regarding the machine learning-based methods include large-scale and high-speed M-Internet, overfitting and underfitting problems, generalization problem, cross-modal learning, and extended channel dimensions and should be considered.

This paper attempts to identify the requirement and the development of machine learning-based mobile big data analysis through discussing the insights of challenges in the MBD and reviewing state-of-the-art applications of data analysis in the area of MBD. The remainder of the paper is organized as follows. Section 2 introduces the development of data collection and properties of MBD. The frequently adopted methods of data analysis and typical applications are reviewed in Section 3. Section 4 summarizes the future challenges of MBD analysis and provides suggestions.

## 2. Development and Collection of the Mobile Big Data

*2.1. Data Collection.* Data collection is the foundation of a data processing and analysis system. Data are collected from mobile smart terminals and Internet services, or called mobile Internet devices (MIDs) generally, which are multimedia-capable mobile devices providing wireless Internet access and contain smartphones, wearable computers, laptop computers, wireless sensors, etc. [54].

MBD can be divided into two hierarchical data form: transmission and application data, from bottom to top. The transmission data focus on solving channel modeling [55, 56] and user access problems corresponding to the physical transmission system of M-Internet. On this foundation, application data focus on the applications based on the MBD including social networks analysis [57–59], user behavior analysis [48, 50, 60], speech analysis and decision in IoV [61–66], smart grid [67, 68], networked healthcare [53, 69, 70], finance services [46, 71], etc.

Due to the heterogeneity of the M-Internet and the variety of the access devices, the collected data are unstructured and usually in many categories and formats, which make data preprocessing become an essential part of a data processing and analysis system in order to ensure the input data complete

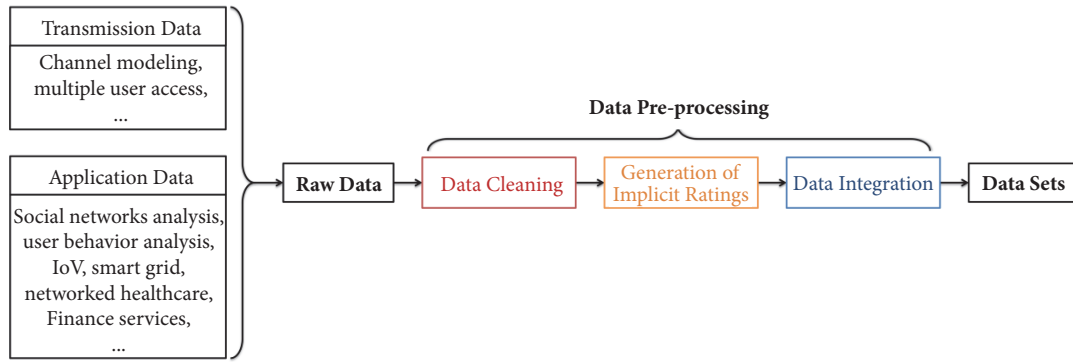


FIGURE 2: The procedures of data collection and preprocessing.

and reliable [72]. Data preprocessing can be divided into three steps which are data cleaning, generation of implicit ratings, and data integration [46].

(1) *Data Cleaning*. Due to possible equipment failures, transmission errors, or human factor, raw data are “dirty data” which cannot be directly used, generally [46]. Therefore, data cleaning methods including outlier detection and denoising are applied in the data preprocessing to obtain the data meet required quality. Manual removal of error data is difficult and impossible to accomplish in MBD due to the massive volume. Common data cleaning methods can alleviate the dirty data problem to some extent by training support vector regression (SVR) classifiers [73], multiple linear regression models [74], autoencoder [75], Bayesian methods [76–78], unsupervised methods [79], or information-theoretic models [79].

(2) *Generation of Implicit Ratings*. Generation of implicit ratings is mainly applied in recommend systems. The volume of rating data increases rapidly by analyzing specific user behaviors to solve data sparsity problem with machine learning algorithms, for example, neural networks and decision trees [46].

(3) *Data Integration*. Data integration is a step to integrate data from different resources with different formats and categories and to handle missing data fields [7].

Figure 2 represents the procedures of data collection and preprocessing.

**2.2. Properties of Mobile Big Data.** The MBD brings a massive amount of new challenges to conventional data analysis methods for its high dimensionality, heterogeneity, and other complex features from applications, such as planning, operation and maintenance, optimization, and marketing [57]. This section discusses the five Vs (short for volume, velocity, variety, value, and veracity) features [80] deriving from big data towards the MBD. The five Vs features have been improved in M-Internet, while it makes users access Internet anytime and anywhere [81].

(1) *Volume: Large Number of MIDs, Exabyte-Level Data, and High-Dimensional Data Space.* Volume is the most obvious

feature of MBD. In the forthcoming 5G network and the era of MBD, conventional store and analysis methods are incapable of processing the 1000x or more wireless traffic volume [7, 82]. It is of great urgency to improve present MBD analysis methods and propose new ones. The methods should be simple and cost-effective to be implemented for MBD processing and analysis. Moreover, they should also be effective enough without requiring a massive amount of data for model training. Finally, they are precise to be applied in various fields [81].

(2) *Velocity: Real-Time Data Streams and Efficiency Requirement.* Velocity can be considered as the speed at which data are transmitted and analyzed [83]. The data is now continuously streaming into the servers in real-time and makes the original batch process break down [84]. Due to the high generating rate of MBD, velocity is the efficiency requirement of MBD analysis since real-time data processing and analysis are extremely important in order to maximize the value of MBD streams [7].

(3) *Variety: Heterogeneous and Nonstructured Mobile Multimedia Contents.* Due to the heterogeneity of MBD which means that mobile data traffic comes from spatially distributed data resources (*i.e.*, MIDs), the variety of MBD arises and makes the MBD more complex [4]. Meanwhile, the nonstructured MBD also causes the variety. The MBD can be divided into structured data, semistructured data, and unstructured data. Here, unstructured data are usually collected in new applications and have random data fields and contents [7]; therefore, they are difficult to analyze before data cleaning and integration.

(4) *Value: Mining Hidden Knowledge and Patterns from Low Density Value Data.* Value, or low density value of MBD, is caused by a large amount of useless or repeated information in the MBD. Therefore, we need to mine the big value by MBD analyzing which is hidden knowledge and patterns extraction. The purified data can provide comprehensive information to conduct more effectively analysis results about user demands, user behaviors, and user habits [85] and to achieve better system management and more accurate demand prediction and decision-making [86].

(5) *Veracity: Consistency, Trustworthiness, and Security of MBD*. The veracity of MBD includes two parts: data consistency and trustworthiness [80]. It can also be summarized as data quality. MBD quality is not guaranteed due to the noise of transmission channel, the equipment malfunctioning, and the uncalibrated sensors of MIDs or the human factor (for instance, malicious invasion) resulting in low-quality data points [4]. Veracity of MBD ensures that the data used in analysis process are authentic and protected from unauthorized access and modification [80].

### 3. Applications of Machine Learning Methods in the Mobile Big Data Analysis

*3.1. Development of Data Analysis Methods.* In this section, we present some recent achievements in data analysis from four different perspectives.

*3.1.1. Divide-and-Conquer Strategy and Sampling of Big Data.* The strategies dividing and conquering big data is a computing paradigm dealing with big data problems. The development of distributed and parallel computing makes divide-and-conquer strategy particularly important.

Generally speaking, whether the diversity of samples in learning data benefits the training results varies. Some redundant and noisy data can cause a large amount of storage cost as well as reducing the efficiency of the learning algorithm and affecting the learning accuracy. Therefore, it is more preferable to select representative samples to form a subset of original sample space according to a certain performance standard, such as maintaining the distribution of samples, topological structure, and keeping classification accuracy. Then learning method will be constructed on previous formed subset to finish the learning task. In this way, we can maintain or even improve the performance of big data analyzing algorithm with minimum computing and stock resources. The need to learn with big data demands on sample selection methods. But most of the sample selection method is only suitable for smaller data sets, such as the traditional condensed nearest neighbor [93], the reduced nearest neighbor [94], and the edited nearest neighbor [95]; the core concept of these methods is to find the minimum consistent subset. To find the minimum consistent subset, we need to test every sample and the result is very sensitive to the initialization of the subset and samples setting order. Li et al. [96] proposed a method to select the classification and edge boundary samples based on local geometry and probability distribution. They keep the space information of the original data but need to calculate k-means for each sample. Angiulli et al. [97, 98] proposed a fast condensation nearest neighbor (FCNN) algorithm based on condensed nearest neighbor, which tends to choose the classification boundary samples.

Jordan [99] proposed statistical inference method for big data. When dealing with statistical inference with divide-and-conquer algorithm, we need to get confidence intervals from huge data sets. By data resampling and then calculating confidence interval, the Bootstrap theory aims to obtain the fluctuation of the evaluation value. But it does not fit big data. The incomplete sampling of data can lead to erroneous

range fluctuations. Data sampling should be correct in order to provide statistical inference calibration. An algorithm named Bag of Little Bootstraps was proposed, which can not only avoid this problem, but also has many advantages on computation. Another problem discussed in [99] is massive matrix calculation. The divide-and-conquer strategy is heuristic, which has a good effect in practical application. However, new theoretical problems arise when trying to describe the statistical properties of partition algorithm. To this end, the support concentration theorem based on the theory of random matrices has been proposed.

In conclusion, data partition and parallel processing strategy is the basic strategy to deal with big data. But the current partition and parallel processing strategy uses little data distribution knowledge, which has influence on the load balancing and the calculation efficiency of big data processing. Hence, there exists an urgent requirement to solve the problem about how to learn the distribution of big data for the optimization of load balancing.

*3.1.2. Feature Selection of Big Data.* In the field of data mining, such as document classification and indexing, the dataset is always large, which contains a large number of records and features. This leads to the low efficiency of algorithm. By feature selection, we can eliminate the irrelevant features and increase the speed of task analysis. Thus, we can get a better preformed model with less running time.

Big data processing faces a huge challenge on how to deal with high-dimensional and sparse data. Traffic network, smartphone communication records, and information shared on Internet provide a large number of high-dimensional data, using tensor (such as a multidimensional array) as natural representation. Tensor decomposition, in this condition, becomes an important tool for summary and analysis. Kolda [100] proposed an efficient use of the memory of the Tucker decomposition method named as memory-efficient Tucker (MET) decomposition decreasing time and space cost which traditional tensor decomposition algorithm cannot do. MET adaptively selects execution strategy based on available memory in the process of decomposition. The algorithm maximizes the speed of computation in the premise of using the available memory. MET avoid dealing with the large number of sporadic intermediate results proceeded during the calculation process. The adaptive selections of operation sequence not only eliminate the intermediate overflow problem, but also save memory without reducing the precision. On the other hand, Wahba [101] proposed two approaches to the statistical machine learning model which involve discrete, noisy, and incomplete data. These two methods are regularized kernel estimation (RKE) and robust manifold unfolding (RMU). These methods use dissimilarity between training information to get nonnegative low rank definite matrix. The matrix will then be embedded into a low dimensional Euclidean space, which coordinate can be used as features of various learning modes. Similarly, most online learning research needs to access all features of training instances. Such classic scenario is not always suitable for practical applications when facing high-dimensional data instances or expensive feature sets. In order to break through this limit,

Hoi et al. [102] propose an efficient algorithm to predict online feature solving problem using some active features based on their study of sparse regularization and truncation technique. They also test the proposed algorithm in some public data sets for feature selection performance.

The traditional self-organizing map (SOM) can be used for feature extraction. But the low speed of SOM limits its usage on large data sets. Sagheer [103] proposed a fast self-organizing map (FSOM) to solve this problem. The goal of this method is to find a feature space where data is mainly distributed in. If there exists such area, data can be extracted in these areas instead of information extraction in overall feature spaces. In this way, we can greatly reduce extraction time.

Anaraki [104] proposed a threshold method of fuzzy rough set feature selection based on fuzzy lower approximation. This method adds a threshold to limit the QuickReduct feature selection. The results of the experiment prove that this method can also help the accuracy of feature extraction with lower running time.

Gheyas et al. [105] proposed a hybrid algorithm of simulated annealing and genetic algorithm (SAGA), combining the advantages of simulated annealing algorithm, genetic algorithm, greedy algorithm, and neural network algorithm, to solve the NP-hard problem of selecting optimal feature subset. The experiment shows that this algorithm can find better optimal feature subset, reducing the time cost sharply. Gheyas pointed in as conclusion that there is seldom a single algorithm which can solve all the problems; the combination of algorithms can effectively raise the overall affect.

To sum up, because of the complexity, high dimensionality, and uncertain characteristics of big data, it is an urgent problem to solve how to reduce the difficulty of big data processing by using dimension reduction and feature selection technology.

*3.1.3. Big Data Classification.* Supervised learning (classification) faces a new challenge of how to deal with big data. Currently, classification problems involving large-scale data are ubiquitous, but the traditional classification algorithms do not fit big data processing properly.

(1) *Support Vector Machine (SVM).* Traditional statistical machine learning method has two main problems when facing big data. (1) Traditional statistical machine learning methods are always involving intensive computing which makes it hard to apply on big data sets. (2) The prediction of model that fits the robust and nonparameter confidence interval is unknown. Lau et al. [106] proposed an online support vector machine (SVM) learning algorithm to deal with the classification problem for sequentially provided input data. The classification algorithm is faster, with less support vectors, and has better generalization ability. Laskov et al. [107] proposed a rapid, stable, and robust numerical incremental support vector machine learning method. Chang et al. [108] developed an open source package called LIBSVM as a library for SVM code implementation.

In addition, Huang et al. [109] present a large margin classifier M4. Unlike other large margin classifiers which

locally or globally constructed separation hyperplane, this model can learn both local and global decision boundary. SVM and minimax probability machine (MPM) has a close connection with the model. The model has important theoretical significance and furthermore, the optimization problem of maxi-min margin machine ( $M^4$ ) can be solved in polynomial time.

(2) *Decision Tree (DT).* Traditional decision tree (DT), as a classic classification learning algorithm, has a large memory requirement problem when processing big data. Franco-Arcega et al. [110] put forward a method of constructing DT from big data, which overcomes some weakness of algorithms in use. Furthermore, it can use all training data without saving them in memory. Experimental results showed that this method is faster than current decision tree algorithm on large-scale problems. Yang et al. [111] proposed a fast incremental optimization decision tree algorithm for large data processing with noise. Compared with former decision tree data mining algorithm, this method has a major advantage on real-time speed for data mining, which is quite suitable when dealing with continuous data from mobile devices. The most valuable feature of this model is that it can prevent explosive growth of the decision tree size and the decrease of prediction accuracy when the data packet contains noise. The model can generate compact decision tree and predict accuracy even with highly noisy data. Ben-Haim et al. [112] proposed an algorithm of building parallel decision tree classifier. The algorithm runs in distributed environment and is suitable for large amount and streaming data. Compared with serial decision tree, the algorithm can improve efficiency under the premise of accuracy error approximation.

(3) *Neural Network and Extreme Learning Machine (ELM).* Traditional feedforward neural networks usually use gradient descent algorithm to tune weight parameters. Generally speaking, slow learning speed and poor generalization performance are the bottlenecks that restrict the application of feedforward neural network. Huang et al. [113] discarded the iterative adjustment strategy of the gradient descent algorithm and proposed extreme learning machine (ELM). This method randomly assigns the input weights and the deviations of the single hidden layer neural network. It can analyze the output weights of the network by one step calculation. Compared to the traditional feedforward neural network training algorithm, the network weights can be determined by multiple iterations, and the training speed of ELM is significantly improved.

However, due to the limitation of computing resource and computational complexity, it is a difficult problem to train a single ELM on big data. There are usually two ways to solve this problem: (1) training ELM [114] based with divide-and-conquer strategy; (2) introducing parallel mechanism [115] to train a single ELM. It is shown in [116, 117] that a single ELM has strong function approximation ability. Whether it is possible to extend this approximation capability to ELM based on divide-and-conquer strategy is a key index to evaluate the possibility that ELM can be applied to big data.

Some of the related studies also include effective learning to solve such problem [118].

In summary, the traditional classification method of machine learning is difficult to apply to the analysis of big data directly. The study of parallel or improved strategies of different classification algorithms has become the new direction.

*3.1.4. Big Data Deep Learning.* With the unprecedentedly large and rapidly growing volumes of data, it is hard for us to get hidden information from big data with ordinary machine learning methods. The shallow-structured learning architectures of most conventional learning methods are not fit for the complex structures and relationships in these input data. Big data deep learning algorithm, with its deep architectures and globally feature extracting ability, can learn complex patterns and hidden connections beyond big data [37, 119]. It has had state-of-the-art performances in many benchmarks and also been applied in industry products. In this section, we will introduce some deep learning methods in big data analytics.

Big data deep learning has some problems: (1) the hidden layers of deep network make it difficult to learn from a given data vector, (2) the gradient descent method for parameters learning makes the initialization time increasing sharply as the number of parameters arises, and (3) the approximations at the deepest hidden layer may be poor. Hinton et al. [32] proposed a deep architecture: deep belief network (DBN) which can learn from both labeled and unlabeled data by using unsupervised pretraining method to learn unlabeled data distributions and a supervised fine-tune method to construct the models, and solved part of the aforementioned problems. Meanwhile, subsequent researches, for example, [120], improved the DBN trying to solve the problems.

Convolutional neural network (CNN) [121] is another popular deep learning network structure for big data analyzing. A CNN has three common features including local receptive fields, shared weights, and spatial or temporal subsampling, and two typical types of layers [122, 123]. Convolutional layers are key parts of CNN structure aiming to extract features from image. Subsampling layers, which are also called pooling layers, adjust outputs from convolutional layer to get translation invariance. CNN is mainly applied in computer vision field for big data, for example, image classification [124, 125] and image segmentation [126].

Document (or textual) representation, also part of NLP, is the basic method for information retrieval and important to understand natural language. Document representation finds specific or important information from the documents by analyzing document structure and content. The unique information could be document topic or a set of labels highly related to the document. Shallow models for document representation only focus on small part of the text and get simple connection between words and sentences. Using deep learning can get global representation of the document because of its large receptive field and hidden layers which could extract more meaningful information. The deep learning methods for document representation make it possible to obtain features from high-dimensional textual data. Hinton

et al. [127] proposed deep generative model to learn binary codes for documents which make documents easy to store up. Socher et al. [128] proposed a recursive neural network on analyzing natural language and contexts, achieving state-of-the-art results on segmentation and understanding of natural language processing. Kumer et al. [129] proposed recurrent neural networks (RNN) which construct search space from large amount of textual data.

With the rapid growth and complexity of academic and industry data sets, how to train deep learning models with large amount of parameters has been a major problem. The works in [40, 41, 43, 130–133] proposed effective and stable parameter updating methods for training deep models. Researchers focus on large-scale deep learning that can be implemented in parallel including improved optimizers [131] and new structures [121, 133–135].

In conclusion, big data deep learning methods are the key methods of data mining. They use complex structure to learn patterns from big data sets and multimodal data. The development of data storage and computing technology promotes the development of deep learning methods and makes it easier to use in practical situations.

*3.2. Wireless Channel Modeling.* As is well known, wireless communication transmits information through electromagnetic waves between a transmitting antenna and a receiving antenna, which is deemed as a wireless channel. In the past few decades, the channel dimension has been extended to space, time, and frequency, which means the channel property is comprehensively discovered. Another development is that channel characteristics can be accurately described by different methods, such as channel modeling [136].

Liang et al. [137] used machine learning to predict channel state information so as to decrease the pilot overhead. Especially for 5G, wireless big data emerges and its related technologies are employed to traditional communication research to meet the demand of 5G. However, the wireless channel is essentially a physical electromagnetic wave, and the current 5G channel model research follows the traditional way. Zhang [138] proposed an interdisciplinary study of big data and wireless channels, which is a cluster-based channel model. In the cluster-nuclei based channel model, the multipath components (MPCs) are aggregated into a traditional stochastically channel model. At the same time, the scene is discerned by the computer and the environment is rebuilt by machine learning methods. Then, by matching the real propagation objects with the clusters, the cluster-nuclei, which are the key factors in contacting deterministic environment and stochastic clusters, can be easily found. There are two main steps employing the machine learning methods in the cluster-nuclei based channel model. The recent progress is shown as follows.

*3.2.1. A Gaussian Mixture Model (GMM) Based Channel MPCs Clustering Method.* The MPCs are clustered with the Gaussian mixture model (GMM) [87, 139]. Using sufficient statistic characteristics of channel multipath, the GMM can get clusters corresponding to the multipath propagation characteristics. The GMM assumes that all the MPCs consist

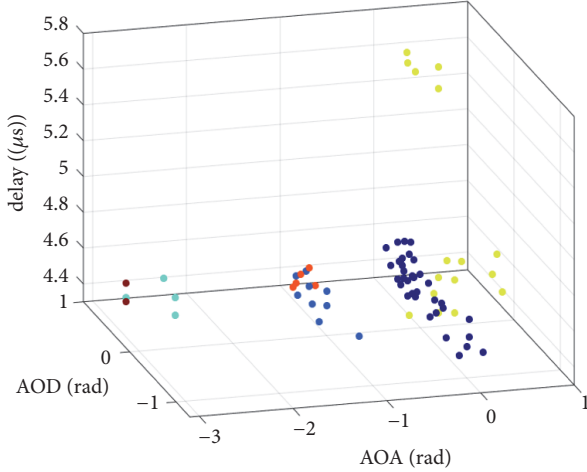


FIGURE 3: Clustering results of GMM [87].

of several Gaussian distributions in varying proportions. Given a set of  $N$  channel multipath  $X$ , the log-likelihood of the Gaussian mixture model is

$$L(X; \Theta) = \sum_{i=1}^N \log \sum_{k=1}^K \pi_k p(x_i | z_i; \mu_k, \Sigma_k), \quad (1)$$

where  $\Theta = \{\pi_k, \mu_k, \Sigma_k, k = 1, \dots, K\}$  is the set of all the parameters and  $\pi_k \in [0, 1]$  is the prior probability satisfying the constraint  $\sum_{k=1}^K \pi_k = 1$ . To estimate the GMM parameters, expectation maximization (EM) algorithm is employed to solve the log-likelihood function of GMM [87]. Figure 3 illustrates the simulation result of GMM clustering algorithm.

As seen in Figure 3, the GMM clustering obtains clearly compact clusters. As scattering property of the channel multipath obeys Gaussian distribution, the compact clusters can accord with the multipath scattering property. Moreover, corresponding to the clustering mechanism of GMM, paper [87] proposed a compact index (CI) to evaluate the clustering results shown as follows:

$$CI = \frac{\text{tr}(B) / (K - 1)}{\text{tr}(W) / (L - K)} \cdot \left( \sum_{k=1}^K S_k^2 \right), \quad (2)$$

where  $S_k^2$  is the variance of the  $k$ th cluster and  $\text{tr}(B)$  and  $\text{tr}(W)$  are given as

$$\text{tr}(B) = \sum_{k=1}^K L_k \cdot MCD(c_k, \bar{c})^2, \quad (3)$$

$$\text{tr}(W) = \sum_{k=1}^K \sum_{j \in C_n} MCD(x_j, c_k)^2, \quad (4)$$

where  $L_k$  is the number of multipaths corresponding to the  $k$ th cluster. Both the means and variances of the clusters are considered in CI. Considering sufficient statistics characteristics, CI can uncover the inherent information of multipath parameters and provide appropriate explanation to

the clustering result. Besides, considering sufficient statistics characteristics, the CI can evaluate the clustering results more reasonably.

**3.2.2. Identifying the Scatters with the Simultaneous Localization and Mapping Algorithm (SLAM).** In order to reconstruct three-dimensional (3D) propagation environment and to find the main deterministic objects, simultaneous localization and mapping (SLAM) algorithm is used to identify the texture from the measurement scenario picture [140, 141]. Figure 4 illustrates our indoor reconstruction result with SLAM algorithm.

The texture of propagation environment can be used to search for the main scatters in the propagation environment. Then, the three-dimensional propagation environment can be reconstructed with the deep learning method.

Then the mechanism to form the cluster-nuclei is clear. The channel impulse response can be produced by machine learning with a limited number of cluster-nuclei, *i.e.*, decision tree [142], neural network [143], and mixture model [144]. Based on the database from various scenarios, antenna configurations, and frequency, channel changing rules can be explored and then input into the cluster-nuclei based modeling. Finally, the predication of channel impulse response in various scenarios and configuration can be realized [138].

**3.3. Analyses of Human Online and Offline Behavior Based on Mobile Big Data.** The advances of wireless networks and increasing mobile applications bring about explosion of mobile traffic data. It is a good source of knowledge to obtain the individuals' movement regularity and acquire the mobility dynamics of populations of millions [145]. Previous researches have described how individuals visit geographical locations and employed mobile traffic data to analyze human offline mobility patterns. Representative works like [146, 147] explore the mobility of users in terms of the number of base stations they visited, which turned out to be a heavy tail distribution. Authors in [146, 148, 149] also reveal that a few important locations are frequently visited by users. In particular, these preferred locations are usually related to home and work places. Moreover, through defining a measure of entropy, Song et al. [150] believe that 93% of individual movements are potentially predictable. Thus, various models have been applied to describe the human offline mobility behavior [151]. Passively collecting human mobile traffic data while users are accessing the mobile Internet has many advantages like low energy consumption. In general, the mobile big data covers a wide range and a great number of populations with fine time granularity, which gives us an opportunity to study human mobility at a scale that other data sources are very hard to reach [152]. Novel offline user mobility models developed based on the mobile big data are expected to benefit many fields, including urban planning, road traffic engineering, telecommunication network construction, and human sociology [145].

Online browsing behavior is another important facet regarding user behavior when it comes to network resource consumption. A variety of applications are now available on smart devices, covering all aspects of our daily life and

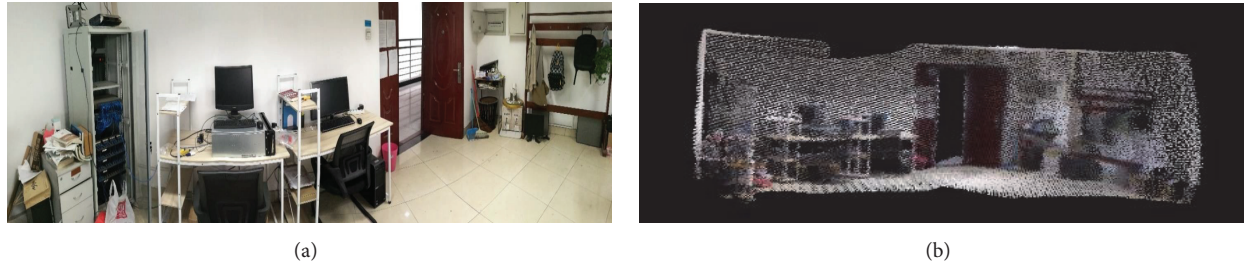
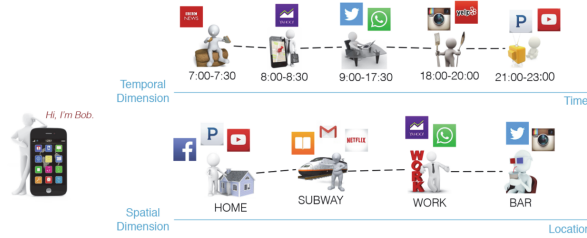
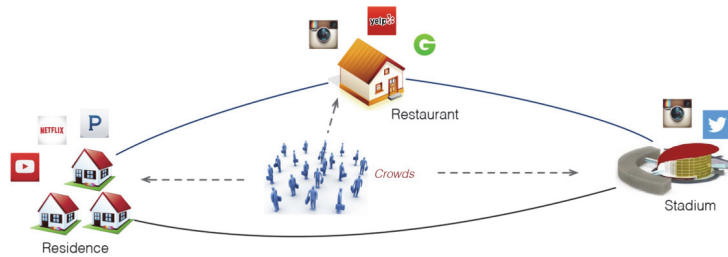


FIGURE 4: Recognition of multiobjects with SLAM algorithm: (a) real indoor scene and (b) reconstruction result with SLAM algorithm.



(a) App usage behavior of Bob in temporal and spatial dimension



(b) App usage behavior of crowds at crowd gathering place

FIGURE 5: App usage behavior in daily life: (a) the app usage behavior of an individual and (b) app usage behavior of crowds at crowd gathering places [50].

providing convenience. For example, we can order taxis, shop, and book hotels using mobile phones. Yang et al. [49] provide a comprehensive study on user behaviors in exploiting the mobile Internet. It has been found that many factors, such as data usage and mobility pattern, may impact people’s online behavior on mobile devices. It is discovered that the more the number of distinct cells a user visit, the more diverse applications user has visited. Zheng et al. [153] analyze the longitudinal impact of proximity density, personality, and location on smartphone traffic consumption. In particular, location has been proven to have strong influences on what kinds of apps users prefer to use [149, 153]. The aforementioned observations point out that there is a close relationship between online browsing behavior and offline mobility behavior.

Figure 5(a) is an example of how browsed applications and current location related to each other from the view of temporal and spatial regularity. It has been found that the mobility behaviors have strong influences on online browsing behavior [149, 153, 154]. Similar trends can also be observed for crowds at crowd gathering places, as is shown in Figure 5(b); *i.e.*, certain apps are favored at places that

group people together and provide some specific functions. The authors in [50] tried to measure the relationship between human mobility and app usage behavior. In particular, the authors proposed a rating framework which can forecast the online app usage behavior for individuals and crowds. Building the bridge between human offline mobility and online mobile Internet behavior can tell us what people really need in daily life. Content providers can leverage this knowledge to appropriately recommend content for mobile users. At the same time, Internet service providers (ISPs) can use this knowledge to optimize networks for better end-user experiences.

In order to make full use of users’ online and offline information, some researchers begin to quantize the interplay between online social network and offline social network and investigate network dynamics from the view of mobile traffic data [155–158]. Specifically, the online and offline social networks are, respectively, constructed based on online interest based and location based social network among mobile users. The two different networks are grouped into layers of a multilayer social network  $M = \{G^{on}, G^{off}\}$ , as shown in Figure 6.  $G^{off}$  and  $G^{on}$  depict offline and online



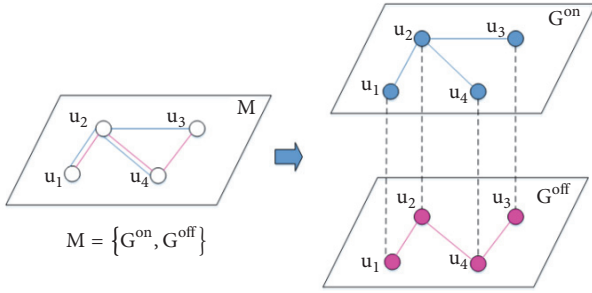


FIGURE 6: Multilayer model of a network [88].

social network separately. In each layer, the graph is described as  $G = \langle V, E \rangle$ , where  $V$  and  $E$ , respectively, represent node sets and edge sets. Nodes, such as  $u_1, \dots, u_4$ , represent users. Edges exist among users when users share similar object-based interests [88]. Combining information from manifold networks in a multilayer structure provides a new insight into user interactions between virtual and physical worlds. It sheds light on the link generation process from multiple views, which will improve social bootstrapping and friend recommendations in various valuable applications by a large margin [158].

So far, we have summarized some representative works related to human online and offline behaviors. It is meaningful to note that owing to the highly spatial-temporal and nonhomogeneous nature of mobile traffic data, a pervasive framework is challenging yet indispensable to realize the collection, processing, and analyses of massive data, reducing resource consumption and improving Quality of Experience (QoE). The seminal work by Qiao et al. [60] proposes a framework for MBD (FMBD). It provides comprehensive functions on data collection, storage, processing, analyzing, and management to monitor and analyze the massive data. Figure 7(a) displays the architecture of FMBD, while Figure 7(b) shows the considered mobile networks framework. With the interaction between user equipment and 2G/3G/4G network, real massive mobile data can be collected by traffic monitoring equipment (TME). The implementation modules are employed based on Apache software [159]. FMBD builds a security environment and easy-to-use platform both for operators and data analysts, showing good performance on energy efficiency, portability, extensibility, usability, security, and stability. In order to meet the increasing demands on traffic monitoring and analyzing, the framework provides a solution to deal with large-scale mobile big data.

In conclusion, the prosperity of continuously emerging mobile applications and users' increasing demands on accessing Internet all bring about challenges for current and future mobile networks. This section surveys the literature on analyses of human online and offline behavior based on the mobile traffic data. Moreover, a framework has also been investigated, in order to meet the higher requirement of dealing with dramatically increased mobile traffic data. The analyses based on the big data will provide valuable information for the ISPs on network deployment, resource

management, and the design of future mobile network architectures.

**3.4. Speech Recognition and Verification for the Internet of Vehicles.** With the significant development of smart vehicle produces, intelligent vehicle based Internet of Vehicle (IoV) technologies have received widespread attention of many giant Internet businesses [160–162]. The IoV technologies include the communication between different vehicles and vehicles to sensors, roads, and humans. These communications can help the IoV system sharing and the gathering information on vehicles and their surrounds.

One of the challenges in the real-life applications of smart vehicles and IoV systems is how to design a robust interactive method between drivers and the IoV system [163]. The level of focusing on driving will directly affect the danger of driver and passengers; hence, the attention of drivers should be paid on the complex road situation in order to avoid accidents during an intense driving. So, using the voices transfer information to the IoV systems is an effective solution for assistant and cooperative driving. By building a speech recognition interactive system, the driver can check traffic jams near the destination or order a lunch in the restaurant near the rest stop through the IoV system by using voice-based interaction. The speech recognition interactive system for IoV system can reduce the risk of vehicle accident, and the drivers do not need to touch the control panels or any buttons. A useful speech recognition system in IoV can simplify the life of the drivers and passengers in vehicles [164]. In the IoV system, drivers want to use their own voice commands to control the driving vehicles, and the IoV system must recognize the difference between an authorized and unauthorized user. Therefore, an automatic speaker verification system is necessary in IoV, which can protect the vehicle from the imposters.

Recently, many deep learning methods have been applied in the speech recognition and speaker verification systems [41, 165–167], and published results show that speech processing methods driven by MBD and deep learning can obviously improve the performance of the existing speech recognition and speaker verification system [40, 168, 169]. In the IoV systems, millions of sensors collect abundant vehicles and environmental noises from engines and streets will significantly reduce the accuracy of speech processing system, while the traditional speech enhancement methods, for example, Wiener filtering [170] and minimum mean-square error estimation (MMSE) [171] which focus on advancing signal noise ratio (SNR), do not take full advantage of a priori distribution of noises around vehicles. With the help of machine learning and deep learning methods, we can use a priori knowledge of the noises to improve the robustness of speech processing systems.

For speech recognition task, deep-neural-network (DNN) can be applied to train an effective monophone classifier, instead of the traditional GMM based classifier. Moreover, the deep-neural-network hidden Markov model (DNN-HMM) speech recognition model can significantly improve the performance of Gaussian mixture model hidden Markov model (GMM-HMM) models [172–174]. As shown

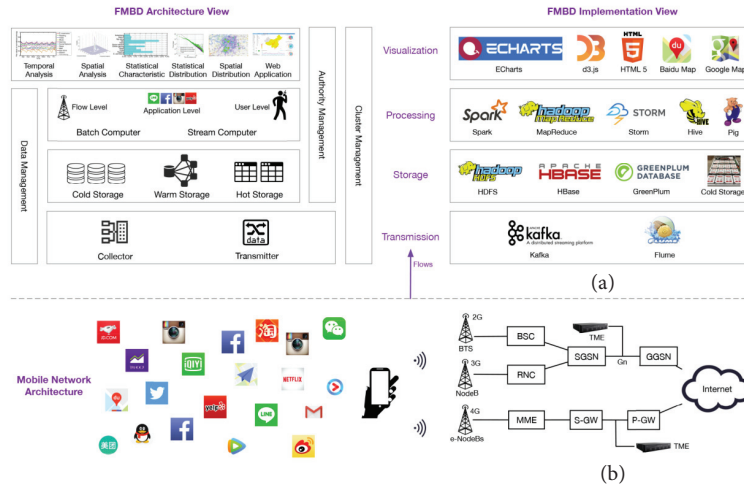


FIGURE 7: The overall architecture of framework for mobile big data (FMBD) and our considered mobile networks architecture [60].

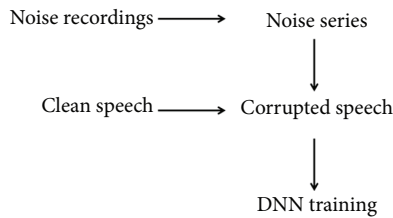


FIGURE 8: Multitraining DNN [89].

in Figure 8, making full use of the self-adaption power of DNN, we can use the multitraining methods to improve the robustness of DNN monophone classifier by adding noise into the training data [89]. The experimental results in [89, 175] show that the multitraining method can build a matched training and testing condition which can improve the accuracy of noisy speech recognition, especially for the prior knowledge of noise types that we can easily obtain in vehicles.

As shown in Figure 9, a DNN can also be used to train a feature mapping network (FMN) which uses noisy features as input and corresponding clean features as training target. Enhanced features extracted by the FMN can improve the performance of speech recognition systems. Han et al. [176] used FMN to extract one enhanced Mel-frequency cepstral coefficient (MFCC) frame from 15 noisy MFCCs frames. Xu et al. [90] built a FMN which learned the mapping from a log spectrogram to a log Mel filter bank. The enhanced feature can remarkably reduce the word error rate in speech recognition.

Besides getting the mapping feature directly, the DNN can also be used to train an ideal binary mask (IBM) which can be used to separate the clean speech from background noise as shown in Figure 10 [91, 177, 178]. With a priori knowledge of noise types and SNR, we can generate IBMs as training targets and use noisy power spectral as training data. In the test phase, we can use the learned IBMs to get enhanced features which can improve the robustness of speech recognition.

In speaker verification tasks, the classical GMM based methods, for example, Gaussian mixture model universal background model (GMM-UBM) [179] and i-vector systems [180], need to build a background GMM, firstly, using a large quantity of speaker independent speeches. Then, by computing the statistics information on each GMM component of enrollment speakers, we can get speaker models or speaker i-vectors. However, a trained monophone classification DNN can replace the function of GMM by computing the statistics information on each monophone instead of on GMM components. Many published papers [181–184] show that the DNN-i-vector based speaker verification systems work better than the GMM-i-vector method on detection accuracy and robustness.

Unlike in the speech recognition tasks where the DNNs are used to get enhanced features from noisy features, researchers more prefer to use a DNN or convolutional neural network (CNN) to generate noise robustness bottleneck feature directly in speaker verification tasks [185–187]. As shown in Figure 11, acoustic features or feature maps are used to train a DNN/CNN with a bottleneck layer which has less nodes and closes to the output layer. Speaker ID, noise types, monophone labels, or combination of these labels are used as training targets. Outputs of bottleneck layers include abundant differentiated information and can be used as speaker verification features which improve the performance of classical speaker verification methods such as the aforementioned GMM-UBM and i-vector. Similar to the multitraining method, adding noisy speeches into the training data can also improve the robustness of extracted bottleneck features [65, 92].

Recently, some adversarial training methods are introduced to extract noise invariant bottleneck features [64, 188]. As shown in Figure 12, the adversarial network includes two parts, *i.e.*, an encoding network (EN) which can extract noise invariant features and a discriminative network (DN) which can judge noise types of the noise invariant feature generated from EN. Therefore, we can get robustness noise invariant features from EN which can improve the performance of

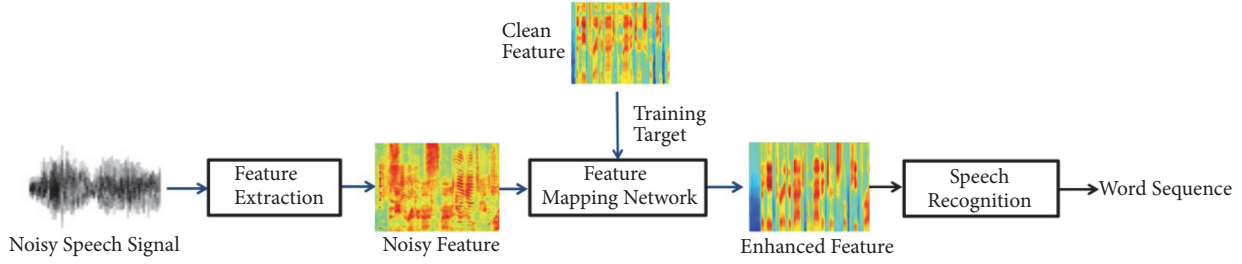


FIGURE 9: DNN used for feature mapping [90].

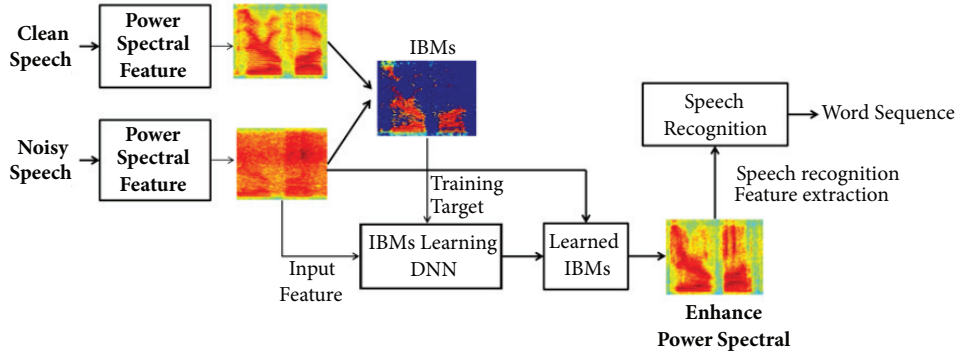


FIGURE 10: DNN used for IBMs learning [91].

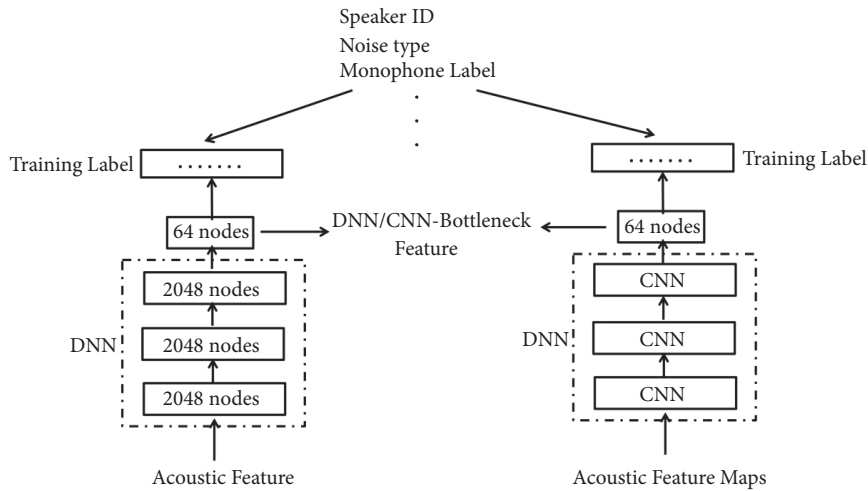


FIGURE 11: DNN/CNN used for extracting bottleneck feature [92].

speaker verification system by adversarial training these two parts in turn [64, 188].

In conclusion, using DNN and machine learning methods can make full use of the MBD collected from the IoV systems. Moreover, it improves the performance of speech recognition and speaker verification methods applied in the voice interactive systems.

#### 4. Conclusions and Future Challenges

Although the machine learning-based methods introduced in Section 3 are widely applied in the MBD fields and obtain

good performances in real data test, the present methods still need to be further developed. Therefore, five main challenges facing MBD analysis regarding the machine learning-based methods should be considered as follows.

(1) *Large-Scale and High-Speed M-Internet.* Due to the growth of MIDs and high speed of M-Internet, increasingly various mobile data traffic is introduced and results in a heavy load to the wireless transmission system, which leads us to improve wireless communication technologies including WLAN and cellular mobile communication. In addition, the requirement of real-time services and applications depends

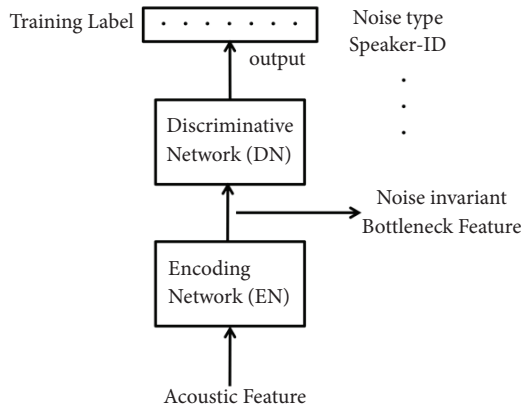


FIGURE 12: Adversarial training network for noise invariant bottleneck feature extraction [64].

on the development of machine learning-based MBD analysis methods towards high efficiency and precision.

(2) *Overfitting and Underfitting Problems.* A benefit of MBD to machine learning and deep learning lies in the fact that the risk of overfitting becomes smaller with more and more data available for training [28]. However, underfitting is another problem for the oversize data volume. In this condition, a larger model might be a better selection, while the model can express more hidden information of the data. Nevertheless, larger model which generally implies a deeper structure increases runtime of the model which affects the real-time performance. Therefore, the model size in machine learning and deep learning, which represents number of parameters, should be balanced to model performance and runtime.

(3) *Generalization Problem.* As the massive scale of MBD, it is impossible to gain entire data even if they are only in a specific field. Therefore, the generalization ability which can be defined as suitable of different data subspace, or called scalability, of a trained machine learning or deep learning model is of great importance for evaluating the performance.

(4) *Cross-Modal Learning.* The variety of MBD causes multiple modalities of data (for example, images, audios, personal location, web documents, and temperature) generated from multiple sensors (correspondingly, cameras, sound recorders, position sensor, and temperature sensor). Multimodal learning should learn from multimodal and heterogeneous input data with machine learning and deep learning [4, 189] and obtain hidden knowledge and meaningful patterns; however, it is quite difficult to discover.

(5) *Extended Channel Dimensions.* The channel dimensions have been extended to three domains, *i.e.*, space, time, and frequency, which means that the channel property is comprehensively discovered. Meanwhile, the increasing antenna number, high bandwidth, and various application scenarios bring the big data of channel measurements and estimations,

especially for 5G. The finding channel characteristics need to be precisely described by more advanced channel modeling methodologies.

In this paper, the applications and challenges of machine learning-based MBD analysis in the M-Internet have been reviewed and discussed. The development of MBD in various application scenarios requires more advanced data analysis technologies especially machine learning-based methods. Three typical applications of MBD analysis focus on wireless channel modeling, human online and offline behavior analysis, and speech recognition and verification in the Internet of Vehicles, respectively, and the machine learning-based methods used are widely applied in many other fields. In order to meet the aforementioned future challenges, three main study aims, *i.e.*, accuracy, feasibility, and scalability [28], are highlighted for present and future MBD analysis research. In future work, accuracy improving will be also the primary task on the basis of a feasible architecture for MBD analysis. In addition, as the aforementioned discussion of the generalization problem, scalability has obtained more and more attentions especially in a classification or recognition problem where scalability also includes the increase in the number of inferred classes. It is of great importance to improve the scalability of the methods with the high accuracy and feasibility in order to face the analysis requirements of MBD.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China (NSFC) [Grant no. 61773071]; in part by the Beijing Nova Program Interdisciplinary Cooperation Project [Grant no. Z181100006218137]; in part by the Beijing Nova Program [Grant no. Z171100001117049]; in part by the Beijing Natural Science Foundation (BNSF) [Grant no. 4162044]; in part by the Funds of Beijing Laboratory of Advanced Information Networks of BUPT; in part by the Funds of Beijing Key Laboratory of Network System Architecture and Convergence of BUPT; and in part by BUPT Excellent Ph.D. Students Foundation [Grant no. XTCX201804].

## References

- [1] International Telecommunication Union (ITU), "ICT Facts and Figures 2017," <https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>, 2017.
- [2] Meeker, "Internet Trend 2017," <http://www.kpcb.com/internet-trends>, 2017.
- [3] G. Fettweis and S. Alamouti, "5G: personal mobile internet beyond what cellular did to telephony," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 140–145, 2014.
- [4] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE Network*, vol. 30, no. 3, pp. 22–29, 2016.

- [5] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper," <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>, 2017.
- [6] Y. Guo, J. Zhang, and Y. Zhang, "An algorithm for analyzing the city residents' activity information through mobile big data mining," in *Proceedings of the Joint 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 10th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Symposium on Parallel and Distributed Processing with Applications, IEEE TrustCom/BigDataSE/ISPA 2016*, pp. 2133–2138, China, August 2016.
- [7] Z. Liao, Q. Yin, Y. Huang, and L. Sheng, "Management and application of mobile big data," *International Journal of Embedded Systems*, vol. 7, no. 1, pp. 63–70, 2015.
- [8] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.
- [9] W. Li and Z. Zhou, "Learning to hash for big data: current status and future trends," *Chinese Science Bulletin (Chinese Version)*, vol. 60, no. 5-6, p. 485, 2015.
- [10] V. Mayerschönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt, Boston, 2013.
- [11] D. Z. Yazti and S. Krishnaswamy, "Mobile big data analytics: research, practice, and opportunities," in *Proceedings of the 15th IEEE International Conference on Mobile Data Management, IEEE MDM 2014*, pp. 1-2, Australia, July 2014.
- [12] E. Zeydan, E. Bastug, M. Bennis et al., "Big data caching for networking: moving from cloud to edge," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 36–42, 2016.
- [13] Z. Liu, Y. Qi, Z. Ma et al., "Sentiment analysis by exploring large scale web-based Chinese short text," in *Proceedings of the International Conference on Computer Science and Application Engineering (CSAE)*, pp. 21–23, 2017.
- [14] Z. Wang, Y. Qi, J. Liu, and Z. Ma, "User intention understanding from scratch," in *Proceedings of the 1st International Workshop on Sensing, Processing and Learning for Intelligent Machines, SPLINE 2016*, Denmark, July 2016.
- [15] C. Zhang, Z. Si, Z. Ma, X. Xi, and Y. Yin, "Mining sequential update summarization with hierarchical text analysis," *Mobile Information Systems*, vol. 2016, Article ID 1340973, 10 pages, 2016.
- [16] C. Zhang, Y. Zhang, W. Xu, Z. Ma, Y. Leng, and J. Guo, "Mining activation force defined dependency patterns for relation extraction," *Knowledge-Based Systems*, vol. 86, pp. 278–287, 2015.
- [17] C. Zhang, W. Xu, Z. Ma, S. Gao, Q. Li, and J. Guo, "Construction of semantic bootstrapping models for relation extraction," *Knowledge-Based Systems*, vol. 83, pp. 128–137, 2015.
- [18] M. Jordan, "Message from the president: the era of big data," *ISBA Bull*, vol. 18, pp. 1–3, 2011.
- [19] W. Chen, D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, "Simultaneous Bayesian sparse approximation with structured sparse models," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6145–6159, 2016.
- [20] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: a tight frame based approach," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2016–2029, 2013.
- [21] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [22] Q. Xie, D. Zeng, Q. Zhao et al., "Robust low-dose CT sinogram preprocessing via exploiting noise-generating mechanism," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2487–2498, 2017.
- [23] M. O'Connor, G. Zhang, W. B. Kleijn, and T. D. Abhayapala, "Function splitting and quadratic approximation of the primal-dual method of multipliers for distributed optimization over graphs," *IEEE Transactions on Signal and Information Processing over Networks*, pp. 1-1, 2018.
- [24] G. Zhang and R. Heusdens, "Distributed optimization using the primal-dual method of multipliers," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 173–187, 2018.
- [25] G. Zhang and R. Heusdens, "Linear coordinate-descent message passing for quadratic optimization," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 20055-2008, 2012.
- [26] G. Zhang, R. Heusdens, and W. B. Kleijn, "Large scale LP decoding with low complexity," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2152–2155, 2013.
- [27] Z. Ma, A. E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational bayesian matrix factorization for bounded support data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 876–889, 2015.
- [28] Z.-H. Zhou, N. V. Chawla, Y. Jin, and G. J. Williams, "Big data opportunities and challenges: discussions from data analytics perspectives," *IEEE Computational Intelligence Magazine*, vol. 9, no. 4, pp. 62–74, 2014.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [30] Y. Bengio and S. Bengio, "Modeling high-dimensional discrete data with multi-layer neural networks," in *Proceedings of the 13th Annual Neural Information Processing Systems Conference, NIPS 1999*, pp. 400–406, USA, December 1999.
- [31] M. Ranzato, Y.-L. Boureau, and Y. Le Cun, "Sparse feature learning for deep belief networks," in *Advances in Neural Information Processing Systems*, pp. 1185–1192, 2008.
- [32] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [33] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proceedings of the 20th Annual Conference on Neural Information Processing Systems (NIPS '06)*, pp. 153–160, Cambridge, Mass, USA, December 2006.
- [34] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *Journal of Machine Learning Research*, vol. 10, pp. 1–40, 2009.
- [35] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, vol. 24, pp. 448–455, 2009.
- [36] I. Goodfellow, H. Lee, and Q. V. Le, "Measuring invariances in deep networks," *Neural Information Processing Systems*, pp. 646–654, 2009.
- [37] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards, AI," *Large Scale Kernel Machines*, vol. 34, pp. 321–360, 2007.
- [38] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on*

- Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [39] I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning—a new frontier in artificial intelligence research,” *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.
- [40] G. E. Dahl, D. Yu, L. Deng et al., “Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30–42, 2012.
- [41] G. Hinton, L. Deng, D. Yu et al., “Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [42] R. Salakhutdinov, A. Mnih, and G. Hinton, “Restricted Boltzmann machines for collaborative filtering,” in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, vol. 227, pp. 791–798, Corvallis, Oregon, June 2007.
- [43] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, “Deep, big, simple neural nets for handwritten digit recognition,” *Neural Computation*, vol. 22, no. 12, pp. 3207–3220, 2010.
- [44] M. D. Zeiler, G. W. Taylor, and R. Fergus, “Adaptive deconvolutional networks for mid and high level feature learning,” in *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV 2011*, pp. 2018–2025, Spain, November 2011.
- [45] A. Efrati, “How deep learning works at Apple, beyond,” <https://www.theinformation.com/How-Deep-Learning-Works-at-Apple-Beyond>, 2013.
- [46] Z. Yang, B. Wu, K. Zheng, X. Wang, and L. Lei, “A survey of collaborative filtering-based recommender systems for mobile internet applications,” *IEEE Access*, vol. 4, pp. 3273–3287, 2016.
- [47] K. Zhu, L. Zhang, and A. Pattavina, “Learning geographical and mobility factors for mobile application recommendation,” *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 36–44, 2017.
- [48] S. Jiang, B. Wei, T. Wang, Z. Zhao, and X. Zhang, “Big data enabled user behavior characteristics in mobile internet,” in *Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–5, Nanjing, October 2017.
- [49] J. Yang, Y. Qiao, X. Zhang, H. He, F. Liu, and G. Cheng, “Characterizing user behavior in mobile internet,” *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 95–106, 2015.
- [50] Y. Qiao, X. Zhao, J. Yang, and J. Liu, “Mobile big-data-driven rating framework: measuring the relationship between human mobility and app usage behavior,” *IEEE Network*, vol. 30, no. 3, pp. 14–21, 2016.
- [51] Y. Qiao, J. Yang, H. He, Y. Cheng, and Z. Ma, “User location prediction with energy efficiency model in the Long Term-Evolution network,” *International Journal of Communication Systems*, vol. 29, no. 14, pp. 2169–2187, 2016.
- [52] M. Gerla and L. Kleinrock, “Vehicular networks and the future of the mobile internet,” *Computer Networks*, vol. 55, no. 2, pp. 457–469, 2011.
- [53] M. M. Islam, M. A. Razzaque, M. M. Hassan, W. N. Ismail, and B. Song, “Mobile cloud-based big healthcare data processing in smart cities,” *IEEE Access*, vol. 5, pp. 11887–11899, 2017.
- [54] Texas Instruments, “Wireless Handset Solutions: Mobile Internet Device,” [http://www.ti.com/solution/handset\\_smartphone](http://www.ti.com/solution/handset_smartphone), 2008.
- [55] X. Ma, J. Zhang, Y. Zhang, and Z. Ma, “Data scheme-based wireless channel modeling method: motivation, principle and performance,” *Journal of Communications and Information Networks*, vol. 2, no. 3, pp. 41–51, 2017.
- [56] X. Ma, J. Zhang, Y. Zhang, Z. Ma, and Y. Zhang, “A PCA-based modeling method for wireless MIMO channel,” in *Proceedings of the 2017 IEEE Conference on Computer Communications: Workshops (INFOCOM WKSHPS)*, pp. 874–879, Atlanta, GA, May 2017.
- [57] X. Zhang, Z. Yi, Z. Yan et al., “Social computing for mobile big data,” *The Computer Journal*, vol. 49, no. 9, pp. 86–90, 2016.
- [58] K. Zhu, Z. Chen, L. Zhang, Y. Zhang, and S. Kim, “Geocascading and community-cascading in social networks: comparative analysis and its implications to edge caching,” *Information Sciences*, vol. 436–437, pp. 1–12, 2018.
- [59] S. Gao, H. Luo, D. Chen et al., “A cross-domain recommendation model for cyber-physical systems,” *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 2, pp. 384–393, 2013.
- [60] Y. Qiao, Z. Xing, Z. M. Fadlullah, J. Yang, and N. Kato, “Characterizing flow, application, and user behavior in mobile networks: a framework for mobile big data,” *IEEE Wireless Communications Magazine*, vol. 25, no. 1, pp. 40–49, 2018.
- [61] H. Yu, Z. Tan, Z. Ma, R. Martin, and J. Guo, “Spoofing detection in automatic speaker verification systems using DNN classifiers and dynamic acoustic features,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12.
- [62] H. Yu, Z.-H. Tan, Y. Zhang, Z. Ma, and J. Guo, “DNN filter bank cepstral coefficients for spoofing detection,” *IEEE Access*, vol. 5, pp. 4779–4787, 2017.
- [63] Z. Ma, H. Yu, Z.-H. Tan, and J. Guo, “Text-independent speaker identification using the histogram transform model,” *IEEE Access*, vol. 4, pp. 9733–9739, 2016.
- [64] H. Yu, Z.-H. Tan, Z. Ma, and J. Guo, “Adversarial network bottleneck features for noise robust speaker verification,” in *Proceedings of the 18th Annual Conference of the International Speech Communication Association, INTERSPEECH 2017*, pp. 1492–1496, Sweden, August 2017.
- [65] H. Yu, A. Sarkar, D. A. L. Thomsen, Z.-H. Tan, Z. Ma, and J. Guo, “Effect of multi-condition training and speech enhancement methods on spoofing detection,” in *Proceedings of the 1st International Workshop on Sensing, Processing and Learning for Intelligent Machines, SPLINE 2016*, Denmark, July 2016.
- [66] H. Yu, Z. Ma, and M. Li, “Histogram transform model Using MFCC features for text-independent speaker identification,” in *Proceedings of the IEEE Asilomar Conference on Signals, Systems, pp. 500–504, 2014.*
- [67] Z. Ma, J. Xie, H. Li et al., “The role of data analysis in the development of intelligent energy networks,” *IEEE Network*, vol. 31, no. 5, pp. 88–95, 2017.
- [68] Z. Ma, H. Li, Q. Sun, C. Wang, A. Yan, and F. Starfelt, “Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems,” *Energy and Buildings*, vol. 85, pp. 464–472, 2014.
- [69] D. West, “How mobile devices are transforming healthcare,” *Issues in Technology Innovation*, vol. 18, no. 1, pp. 1–11, 2012.
- [70] L. A. Tawalbeh, R. Mehmood, E. Benkhelifa, and H. Song, “Mobile cloud computing model and big data analysis for healthcare applications,” *IEEE Access*, vol. 4, pp. 6171–6180, 2016.
- [71] S. Sagiroglu and D. Sinanc, “Big data: a review,” in *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS '13)*, pp. 42–47, IEEE, San Diego, Calif, USA, May 2013.

- [72] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu, and L. Lei, "Soft-defined heterogeneous vehicular network: architecture and challenges," *IEEE Network*, vol. 30, no. 4, pp. 72–80, 2016.
- [73] H. Hsieh, V. Klyuev, Q. Zhao, and S. Wu, "SVR-based outlier detection and its application to hotel ranking," in *Proceedings of the 2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)*, pp. 1–6, Paris, France, October 2014.
- [74] S. Rahman, M. Sathik, and K. Kannan, "Multiple linear regression models in outlier detection," *International Journal of Research in Computer Science*, vol. 2, no. 2, pp. 23–28, 2012.
- [75] H. A. Dau, V. Ciesielski, and A. Song, "Anomaly Detection using replicator neural networks trained on examples of one class," in *Simulated Evolution and Learning*, vol. 8886 of *Lecture Notes in Computer Science*, pp. 311–322, Springer International Publishing, Cham, 2014.
- [76] Z. Ma, J.-H. Xue, A. Leijon, Z.-H. Tan, Z. Yang, and J. Guo, "Decorrelation of neutral vector variables: theory and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 129–143, 2018.
- [77] Z. Ma, S. Chatterjee, W. B. Kleijn, and J. Guo, "Dirichlet mixture modeling to estimate an empirical lower bound for LSF quantization," *Signal Processing*, vol. 104, pp. 291–295, 2014.
- [78] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2160–2173, 2011.
- [79] C. C. Aggarwal, "Outlier analysis," in *Data Mining*, Springer, 2015.
- [80] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *Proceedings of the IEEE International Conference on Collaboration Technologies and Systems (CTS '13)*, pp. 48–55, May 2013.
- [81] C. Zhou, H. Jiang, Y. Chen, L. Wu, and S. Yi, "User interest acquisition by adding home and work related contexts on mobile big data analysis," in *Proceedings of the IEEE INFOCOM 2016 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 201–206, San Francisco, CA, USA, April 2016.
- [82] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: challenges and research advances," *IEEE Network*, vol. 28, no. 6, pp. 6–11, 2014.
- [83] S. Landset, T. M. Khoshgoftaar, A. N. Richter, and T. Hasanin, "A survey of open source tools for machine learning with big data in the Hadoop ecosystem," *Journal of Big Data*, vol. 2, no. 1, pp. 24–59, 2015.
- [84] D. Soubra, "The 3Vs that define Big Data," <http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>, 2012.
- [85] L. Ma, F. Nie, and Q. Lu, "An analysis of supply chain restructuring based on big data and mobile internet—a case study of warehouse-type supermarkets," in *Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services, GSIS 2015*, pp. 446–451, UK, August 2015.
- [86] A. McAfee and E. Brynjolfsson, "Big data: the management revolution," *Harvard Business Review*, vol. 90, no. 10, pp. 60–128, 2012.
- [87] Y. Li, J. Zhang, and Z. Ma, "Clustering in wireless propagation channel with a statistics-based framework," in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Barcelona, April 2018.
- [88] P. Kazienko, K. Musiał, and T. Kajdanowicz, "Multidimensional social network in the social recommender system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 41, no. 4, pp. 746–759, 2011.
- [89] A. Abe, K. Yamamoto, and S. Nakagawa, "Robust speech recognition using DNN-HMM acoustic model combining noise-aware training with spectral subtraction," in *Proceedings of the 16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015*, pp. 2849–2853, Germany, September 2015.
- [90] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65–68, 2014.
- [91] A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in *Proceedings of the 2013 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, pp. 7092–7096, Canada, May 2013.
- [92] D. Serdyuk, K. Audhkhasi, and P. Brakel, "Invariant representations for noisy speech recognition," *Computation and Language*, 2016, arXiv:1612.01928.
- [93] P. E. Hart, "The condensed nearest neighbor rule," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 515–516, 1968.
- [94] G. Gates, "The reduced nearest neighbor rule," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 431–433, 1972.
- [95] H. Brighton and C. Mellish, "Advances in instance selection for instance-based learning algorithms," *Data Mining and Knowledge Discovery*, vol. 6, no. 2, pp. 153–172, 2002.
- [96] Y. Li and L. Maguire, "Selecting critical patterns based on local geometrical and statistical information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1189–1201, 2011.
- [97] F. Angiulli, "Fast nearest neighbor condensation for large data sets classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 11, pp. 1450–1464, 2007.
- [98] F. Angiulli and G. Folino, "Distributed nearest neighbor-based condensation of very large data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 12, pp. 1593–1606, 2007.
- [99] M. I. Jordan, "Divide-and-conquer and statistical inference for big data," in *Proceedings of the the 18th ACM SIGKDD international conference*, p. 4, Beijing, China, August 2012.
- [100] T. G. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," in *Proceedings of the 8th IEEE International Conference on Data Mining, ICDM 2008*, pp. 363–372, Italy, December 2008.
- [101] G. Wahba, "Dissimilarity data in statistical model building and machine learning," in *Proceedings of the 5th International Congress of Chinese Mathematicians*, pp. 785–809, 2012.
- [102] S. C. Hoi, J. Wang, P. Zhao, and R. Jin, "Online feature selection for mining big data," in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pp. 93–100, Beijing, China, August 2012.
- [103] A. Sagheer, N. Tsuruta, R.-I. Taniguchi, D. Arita, and S. Maeda, "Fast feature extraction approach for multi-dimension feature space problems," in *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, pp. 417–420, China, August 2006.
- [104] J. R. Anaraki and M. Eftekhari, "Improving fuzzy-rough quick reduct for feature selection," in *Proceedings of the 2011 19th*

- Iranian Conference on Electrical Engineering, ICEE 2011*, Iran, May 2011.
- [105] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognition*, vol. 43, no. 1, pp. 5–13, 2010.
- [106] K. W. Lau and Q. H. Wu, "Online training of support vector classifier," *Pattern Recognition*, vol. 36, no. 8, pp. 1913–1920, 2003.
- [107] P. Laskov, C. Gehl, S. Krüger, and K.-R. Müller, "Incremental support vector learning: analysis, implementation and applications," *Journal of Machine Learning Research*, vol. 7, pp. 1909–1936, 2006.
- [108] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [109] K. Huang, H. Yang, I. King, and M. R. Lyu, "Maxi-min margin machine: learning large margin classifiers locally and globally," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 19, no. 2, pp. 260–272, 2008.
- [110] A. Franco-Arcega, J. A. Carrasco-Ochoa, G. Sanchez-Daz et al., "Building fast decision trees from large training sets," *Intelligent Data Analysis*, vol. 16, no. 4, pp. 649–664, 2012.
- [111] H. Yang and S. Fong, "Incrementally optimized decision tree for noisy big data," in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '12)*, pp. 36–44, Beijing, China, August 2012.
- [112] Y. Ben-Haim and E. Tom-Tov, "A streaming parallel decision tree algorithm," *Journal of Machine Learning Research (JMLR)*, vol. 11, pp. 849–872, 2010.
- [113] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [114] N. Liu and H. Wang, "Ensemble based extreme learning machine," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 754–757, 2010.
- [115] Q. He, T. Shang, F. Zhuang, and Z. Shi, "Parallel extreme learning machine for regression based on mapReduce," *Neurocomputing*, vol. 102, pp. 52–58, 2013.
- [116] R. Zhang, Y. Lan, G.-B. Huang, and Z.-B. Xu, "Universal approximation of extreme learning machine with adaptive growth of hidden nodes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 365–371, 2012.
- [117] H.-J. Rong, G.-B. Huang, N. Sundararajan, P. Saratchandran, and H.-J. Rong, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 4, pp. 1067–1072, 2009.
- [118] Y. Yang, Y. Wang, and X. Yuan, "Bidirectional extreme learning machine for regression problem and its learning effectiveness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1498–1505, 2012.
- [119] W. X. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.
- [120] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *The American Association for the Advancement of Science: Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [121] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [122] D. C. Ciresan, U. Meier, and J. Masci, "Flexible, high performance convolutional neural networks for image classification," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1237–1242, 2011.
- [123] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 92–101, 2010.
- [124] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12)*, pp. 1097–1105, Lake Tahoe, Nev, USA, December 2012.
- [125] J. Dean, G. Corrado, and R. Monga, "Large scale distributed deep networks," in *Neural Information Processing Systems*, pp. 1223–1231, 2012.
- [126] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015*, pp. 1742–1750, Chile, December 2015.
- [127] G. Hinton and R. Salakhutdinov, "Discovering binary codes for documents by learning deep generative models," *Topics in Cognitive Science*, vol. 3, no. 1, pp. 74–91, 2011.
- [128] R. Socher, C. C.-Y. Lin, C. D. Manning, and A. Y. Ng, "Parsing natural scenes and natural language with recursive neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*, pp. 129–136, Bellevue, Wash, USA, June 2011.
- [129] R. Kumar, J. O. Talton, and S. Ahmad, "Data-driven web design," in *Proceedings of the International Conference on Machine Learning*, pp. 3–4, 2012.
- [130] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*, pp. 873–880, Canada, June 2009.
- [131] J. Martens, "Deep learning via Hessian-free optimization," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 735–742, June 2010.
- [132] K. Zhang and X.-W. Chen, "Large-scale deep belief nets with mapreduce," *IEEE Access*, vol. 2, pp. 395–403, 2014.
- [133] L. Deng, D. Yu, and J. Platt, "Scalable stacking and learning for building deep architectures," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 2133–2136, Kyoto, Japan, March 2012.
- [134] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. LeCun, "Learning invariant features through topographic filter maps," in *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 1605–1612, USA, June 2009.
- [135] B. Hutchinson, L. Deng, and D. Yu, "Tensor deep stacking networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1944–1957, 2013.
- [136] J. Zhang, "Review of wideband MIMO channel measurement and modeling for IMT-Advanced systems," *Chinese Science Bulletin*, vol. 57, no. 19, pp. 2387–2400, 2012.
- [137] C. Liang, H. Li, Y. Li, S. Zhou, and J. Wang, "A learning-based channel model for synergetic transmission technology," *China Communications*, vol. 12, no. 9, pp. 83–92, 2015.



- [138] J. Zhang, "The interdisciplinary research of big data and wireless channel: a cluster-nuclei based channel model," *China Communications*, vol. 13, no. supplement 2, Article ID 7833457, pp. 14–26, 2016.
- [139] Y. Li, J. Zhang, Z. Ma, and Y. Zhang, "Clustering analysis in the wireless propagation channel with a variational gaussian mixture model," *IEEE Transactions on Big Data*, pp. 1-1, 2018.
- [140] F. Bai, T. Vidal-Calleja, and S. Huang, "Robust incremental SLAM under constrained optimization formulation," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1–8, 2018.
- [141] I. Z. Ibragimov and I. M. Afanasyev, "Comparison of ROS-based visual SLAM methods in homogeneous indoor environment," in *Proceedings of the 2017 14th Workshop on Positioning, Navigation and Communications (WPNC)*, pp. 1–6, Bremen, October 2017.
- [142] U. M. Fayyad, *On the Induction of Decision Trees for Multiple Concept Learning*, University of Michigan, 1992.
- [143] G. Cybenko, "Approximation by Superpositions of a sigmoidal function," *Mathematics of Control Signals & Systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [144] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of dirichlet mixture model with variational inference," *Pattern Recognition*, vol. 47, no. 9, pp. 3143–3157, 2014.
- [145] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 124–161, 2016.
- [146] E. Halepovic and C. Williamson, "Characterizing and modeling user mobility in a cellular data network," in *Proceedings of the PE-WASUN'05 - Second ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks*, pp. 71–78, Canada, October 2005.
- [147] S. Scepanovic, P. Hui, and A. Yla-Jaaski, "Revealing the pulse of human dynamics in a country from mobile phone data," *NetMob D4D Challenge*, pp. 1–15, 2013.
- [148] S. Isaacman, R. A. Becker, and R. Caceres, "Identifying important places in people's lives from cellular network data," in *Proceedings of the International Conference on Pervasive Computing*, pp. 133–151, 2011.
- [149] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: connecting people, locations and interests in a mobile 3G network," in *Proceedings of the 2009 9th ACM SIGCOMM Internet Measurement Conference, IMC 2009*, pp. 267–279, USA, November 2009.
- [150] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [151] Q. Lv, Y. Qiao, N. Ansari, J. Liu, and J. Yang, "Big data driven hidden markov model based individual mobility prediction at points of interest," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5204–5216, 2017.
- [152] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1443–1455, 2017.
- [153] L. Meng, S. Liu, and A. Striegel, "Analyzing the longitudinal impact of proximity, location, and personality on smartphone usage," *Computational Social Networks*, vol. 1, no. 1, 2014.
- [154] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage," in *Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '11)*, pp. 47–56, September 2011.
- [155] D. Hristova, M. Musolesi, and C. Mascolo, "Keep your friends close and your facebook friends closer: a multiplex network approach to the analysis of offline and online social Ties," in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, pp. 206–215, USA, June 2014.
- [156] R. I. M. Dunbar, V. Arnaboldi, M. Conti, and A. Passarella, "The structure of online social networks mirrors those in the offline world," *Social Networks*, vol. 43, pp. 39–47, 2015.
- [157] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo, "Measuring urban social diversity using interconnected geo-social networks," in *Proceedings of the the 25th International Conference*, pp. 21–30, Canada, April 2016.
- [158] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, "A multilayer approach to multiplexity and link prediction in online geo-social networks," *EPJ Data Science*, vol. 5, no. 1, 2016.
- [159] Apache, "Apache software foundation," <http://apache.org>, 2017.
- [160] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds," in *Proceedings of the IEEE World Forum on Internet of Things (WF-IoT '14)*, pp. 241–246, March 2014.
- [161] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of internet of vehicles," *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.
- [162] K. M. Alam, M. Saini, and A. El Saddik, "Toward social internet of vehicles: concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, 2015.
- [163] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 4, pp. 631–640, 2001.
- [164] C. Y. Loh, K. L. Boey, and K. S. Hong, "Speech recognition interactive system for vehicle," in *Proceedings of the 13th IEEE International Colloquium on Signal Processing and its Applications, CSPA 2017*, pp. 85–88, Malaysia, March 2017.
- [165] D. Amodei, S. Ananthanarayanan, and R. Anubhai, "Deep speech 2: end-to-end speech recognition in English and mandarin," in *Proceedings of the International Conference on Machine Learning*, pp. 173–182, 2016.
- [166] E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. Gonzalez-Dominguez, "Deep neural networks for small footprint text-dependent speaker verification," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014*, pp. 4052–4056, Italy, May 2014.
- [167] K. Chen and A. Salman, "Learning speaker-specific characteristics with a deep neural architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 11, pp. 1744–1756, 2011.
- [168] L. Deng, G. E. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8599–8603, IEEE, Vancouver, Canada, May 2013.
- [169] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 6645–6649, May 2013.

- [170] J. Meyer and K. U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP. Part 1 (of 5)*, pp. 1167–1170, April 1997.
- [171] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, pp. 4266–4269, USA, March 2010.
- [172] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH '11)*, vol. 33, pp. 437–440, August 2011.
- [173] H. Ze, A. Senior, and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '13)*, pp. 7962–7966, IEEE, Vancouver, Canada, May 2013.
- [174] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8609–8613, May 2013.
- [175] Y. Qian, M. Bi, T. Tan, and K. Yu, "Very deep convolutional neural networks for noise robust speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 12, pp. 2263–2276, 2016.
- [176] K. Han, Y. He, D. Bagchi et al., "Deep neural network based spectral feature mapping for robust speech recognition," *INTERSPEECH*, pp. 2484–2488, 2015.
- [177] B. Li and K. C. Sim, "Improving robustness of deep neural networks via spectral masking for automatic speech recognition," in *Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013*, pp. 279–284, Czech Republic, December 2013.
- [178] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.
- [179] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1, pp. 19–41, 2000.
- [180] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.
- [181] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 4277–4280, IEEE, March 2012.
- [182] M. McLaren, Y. Lei, and L. Ferrer, "Advances in deep neural network approaches to speaker recognition," in *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2015*, pp. 4814–4818, Australia, April 2014.
- [183] C. Yu, A. Ogawa, M. Delcroix, T. Yoshioka, T. Nakatani, and J. H. L. Hansen, "Robust i-vector extraction for neural network adaptation in noisy environment," in *Proceedings of the 16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015*, pp. 2854–2857, Germany, September 2015.
- [184] N. Li, M.-W. Mak, and J.-T. Chien, "Deep neural network driven mixture of PLDA for robust i-vector speaker verification," in *Proceedings of the 2016 IEEE Workshop on Spoken Language Technology, SLT 2016*, pp. 186–191, USA, December 2016.
- [185] Z. Zhang, L. Wang, A. Kai, T. Yamada, W. Li, and M. Iwahashi, "Deep neural network-based bottleneck feature and denoising autoencoder-based dereverberation for distant-talking speaker identification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, p. 12, 2015.
- [186] M. McLaren, Y. Lei, and N. Scheffer, "Application of convolutional neural networks to speaker recognition in noisy conditions," in *Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [187] T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Auto-encoder bottleneck features using deep belief networks," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012*, pp. 4153–4156, Japan, March 2012.
- [188] Y. Shinohara, "Adversarial multi-task learning of deep neural networks for robust speech recognition," *INTERSPEECH*, pp. 2369–2372, 2016.
- [189] N. D. Lane and P. Georgiev, "Can deep learning revolutionize mobile sensing?" in *Proceedings of the the 16th International Workshop*, pp. 117–122, Santa Fe, NM, USA, February 2015.

## Research Article

# Analysis of Nonstationary Characteristics for High-Speed Railway Scenarios

Tao Zhou <sup>1,2</sup>, Cheng Tao <sup>1</sup> and Kai Liu<sup>1</sup>

<sup>1</sup>*Institute of Broadband Wireless Mobile Communications, Beijing Jiaotong University, Beijing 100044, China*

<sup>2</sup>*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*

Correspondence should be addressed to Tao Zhou; [taozhou.china@gmail.com](mailto:taozhou.china@gmail.com)

Received 21 March 2018; Accepted 29 May 2018; Published 21 June 2018

Academic Editor: Tommi Jamsa

Copyright © 2018 Tao Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents the analysis of nonstationary characteristics for high-speed railway (HSR) scenarios, according to passive long-term evolution- (LTE-) based channel measurements. The measurement data collected in three typical scenarios, rural, station, and suburban, are processed to obtain the channel impulse responses (CIRs). Based on the CIRs, the nonstationarity of the HSR channel is studied focusing on the stationarity interval, and a four-state Markov chain model is generated to describe the birth-death process of multipath components. The presented results will be useful in dynamic channel modeling for future HSR mobile communication systems.

## 1. Introduction

With the rapid development of high-speed railways (HSRs), there appears a growth in demand for new railway communication services, for example, real-time monitoring, train multimedia dispatching, railway emergency communications, railway Internet of Things (IoT), and broadband wireless access for train passengers [1]. To satisfy such ever-increasing requirements, broadband wireless communication systems for HSR have recently attracted much attention in the world. Since 2014, International Union of Railways (UIC) has considered replacing the current global system for mobile communications for railway (GSM-R) with the long-term evolution for railway (LTE-R) [2]. In China, fourth-generation (4G) networks have been deployed along most of HSRs, for a total of 15,000 km by 2014, and as HSRs continue to grow, the dedicated 4G networks that grow with them will exceed 30,000 km in 2020. For future fifth-generation (5G) mobile communication system, it is reported that one of its aims is to provide high-data-rate access under high mobility scenarios [3].

Since the radio channel determines the performance of broadband wireless mobile communication systems, detailed knowledge and accurate characterization of its parameters in diverse scenarios are vital. The propagation characteristics are

in disparity under various HSR environments. In this case, the characterization of HSR channels should consider the influence of scenarios.

So far, a wide variety of the studies have concentrated on the long-term fading behavior, involving path loss and shadowing, in multiple HSR scenarios [4–8]. There are a few research works on the short-term fading behavior, based on wideband channel measurements conducted on HSR. By contrast, the nonstationary behavior, which has been widely studied for vehicle-to-vehicle (V2V) channels [9, 10], is rarely investigated in HSR environments. To fill this research gap, we present the analysis of the nonstationary characteristics for multiple HSR scenarios. Passive channel measurements are conducted for three typical scenarios, rural, station, and suburban, in an HSR LTE network. Considering the interval of stationarity and the dynamic evolution of multipath components (MPCs), the nonstationary behavior is analyzed and compared in different scenarios.

The remainder of this paper is outlined as follows. Section 2 reviews the related work focusing on short-term fading and nonstationary behaviors. In Section 3, our passive channel measurements based on LTE are introduced. Then, the nonstationary characteristics are analyzed in Section 4, respectively. Finally, conclusions are drawn in Section 5.

TABLE 1: Summary of investigation on short-term fading and nonstationary behaviors in HSR environments.

Scenario	Short-term fading behavior			Nonstationary behavior		
	Fading severity	Time dispersion	Frequency dispersion	Space dispersion	Stationarity interval	Birth-death process
Viaduct	[11, 12, 16]	[11]	[11, 12]	[15]	[25]	[26]
Cutting	[13, 14, 17]	[13, 14]	[13, 14]	[15]	-	-
Rural	[18]	[18, 19]	[19]	[18, 19]	-	-
Hilly terrain	[20]	[21]	[21]	-	-	-
Tunnel	[22]	[22]	[22]	-	-	-
Station	[23, 24]	[24]	-	-	[24]	-
Suburban/urban	-	-	-	-	-	-

## 2. Related Work

Unique HSR scenarios, such as viaduct, cutting, tunnel, station, hilly terrain, rural, and suburban, have a significant impact on propagation characteristics. In our previous work, fading severity and time-frequency-space dispersion of HSR channels in the viaduct and cutting scenarios were deeply characterized based on measurements using Propsound [11–15]. In [15], spatial characteristics, involving angle of arrival (AOA), root-mean-square (RMS) angle spread (AS), and spatial correlation (SC), were analyzed according to a so-called moving virtual antenna array (VAA) scheme. Authors in [16, 17] proposed a statistic model for Ricean K-factor, which investigated the impact of the viaduct height and the cutting width. The WINNER II model [18] and COST 2100 TD [19] provided some measurement results of short-term fading behavior for rural scenarios. There were also a few results of K-factor, RMS delay spread (DS), and Doppler power spectral density (DPSD) measured in the hilly terrain and tunnel scenarios [20–22]. Authors in [23, 24] presented detailed analysis of fading severity and time dispersion in open-type and semiclosed station scenarios. For the nonstationary behavior, stationarity interval (SI) in the viaduct scenario was investigated based on GSM-R measurements, which showed that conventional channel models offered SI much larger than the actual measured ones [25]. We also tried to adopt a RUN test method to obtain the SI for the open-type station scenario in [24]. Additionally, a four-state Markov chain was used to model the birth-death (B-D) process of MPCs in the viaduct scenario [26].

Table 1 summarizes the existing measurement campaigns about short-term fading and nonstationary behaviors in different HSR environments. It can be found that there are a few results of time-frequency-space characteristics in some HSR scenarios. However, the characterization of the nonstationary behavior is largely neglected in most scenarios. Therefore, this paper aims to investigate the nonstationary behavior in the rural, station, and suburban scenarios.

## 3. Channel Measurements

*3.1. Measurement Scenarios.* An LTE network deployed on Beijing to Tianjin (BT) HSR in China was chosen in our measurements [27]. It is a hybrid network composed of a dedicated network and a common network. The architecture of the dedicated network is completely different from the

TABLE 2: Measurement parameters in the BT HSR LTE network.

Scenario	Rural	Station	Suburban
	eNB side		
Frequency	1.89 GHz		2.605 GHz
Bandwidth	18 MHz		18 MHz
CRS power	12.2 dBm		12.2 dBm
Antenna type	Directional		Directional
Antenna gain	17.4 dBi		18.6 dBi
Horizontal beamwidth	67 deg		60 deg
Vertical beamwidth	6.6 deg		4.9 deg
Electric tilted angle	3 deg		3 deg
	Sounder side		
Antenna type	Omnidirectional		Omnidirectional
Antenna gain	8.5 dBi		8.5 dBi
Speed of HST	285 km/h		185 km/h
	Network side		
Distance between PSs		1.2 km	
Height of PS		30 m	
Height of rail track		10 m	
Distance between PS and rail track		30 m	

common network, which adopts building baseband unit (BBU) plus remote radio unit (RRU) to achieve the special narrow-strip-shaped coverage instead of the cellular coverage. In this architecture, one physical site (PS) deploys two RRUs, which transmit signals by directional antennas in opposite directions along the railway track. The RRUs are connected together via optical fiber and then to a BBU that is in charge of radio frequency (RF) signal processing.

The LTE sounder is placed on a high-speed train (HST) to collect the channel data, which experiences multiple scenarios along the BT railway line, such as rural, station, and suburban. In our measurement, the rural and station scenarios are within the coverage of dedicated network, while the suburban scenario is covered by the common network. The detailed measurement parameters for different scenarios in the network are listed in Table 2. The carrier frequency is 1.89 GHz for the rural and station scenarios and 2.605 GHz for the suburban scenario. When the HST moves into

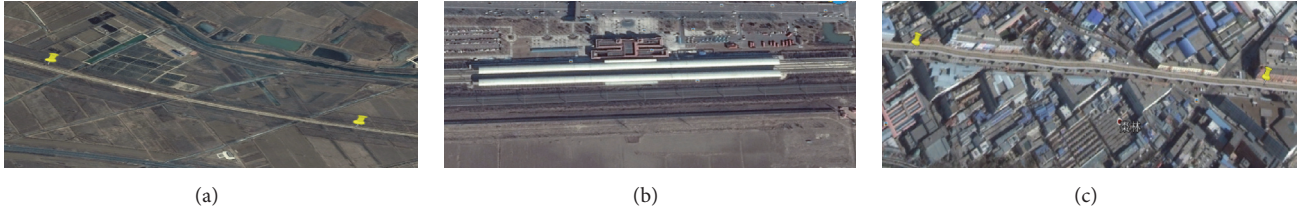


FIGURE 1: Measurement scenarios. (a) Rural. (b) Station. (c) Suburban.

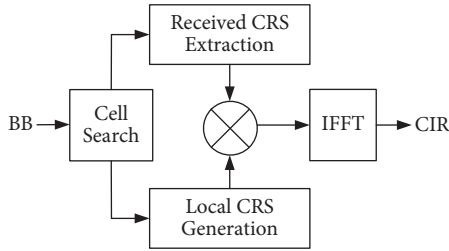


FIGURE 2: The procedure of data processing.

the suburban area, the speed is decreased from 285 km/h to 185 km/h. The specification of the directional eNB antenna such as gain and beamwidth has slight difference in different frequencies. At Rx side, the LTE sounder is connected to a train-mounted omnidirectional antenna. The average spacing between neighboring PSs is around 1.2 km. The height and distance difference between the PS and the rail track are about 20 m and 30 m, respectively.

The measured scenarios are shown in Figure 1. As for the rural scenario, the transmit antenna is much higher than the surroundings, which are light forests and a few buildings with an average height of less than 10 m. The link between the transmitter (Tx) and Rx generally has a strong line-of-sight (LoS) component. However, after a certain distance, the impact of the sparse scatterers will be noticed at the Rx represented by non-LoS (NLoS) components. With regard to the station scenario, in the measurement, the HST runs through the station without stopping. The measured station can be regarded as an open-type station with two awnings that only cover the platform supporting a clear free space over the rail. However, the awnings can still produce lots of NLoS components to complicate the fading behavior. The length of the station is 440 m, the width of the awning is 14 m, and the width of the gap between the two awnings is 9 m. Suburban is a transition zone between the rural and urban. The NLoS components in the suburban environment will be much richer than those in the rural environment. The density of the buildings in the suburban scenario is similar to that in the urban scenario, but the height of the buildings is lower. Since the measured suburban is close to the urban area, some remote high buildings could affect the results.

**3.2. Data Processing.** Baseband data (BB) collected by the LTE sounder in the multilink regions are used for offline processing. The procedure of data processing is shown in Figure 2. Firstly, cell search is implemented to determine the

cell identity and obtain synchronized frames for extracting received CRSs and generating local CRSs. Then, frequency-domain correlation is used to estimate channel frequency responses which can be subsequently transformed to the raw CIRs by inverse fast Fourier transform (IFFT) operation [28].

## 4. Results and Analysis

**4.1. Stationarity Interval.** High mobility leads to the violation of wide sense stationary (WSS) condition for wireless channels under HSR scenarios. The stationarity interval (SI) is defined as the maximum time or distance duration, over which the channel satisfies the WSS condition. It is summarized by [29] that there are several metrics that can be used for measuring the SI, involving local region of stationarity (LRS), correlation matrix distance (CMD), and spectral divergence (SD). Besides, some statistical tests for the WSS of a random process can also be applied to the determination of SI, such as RUN test and reverse arrangement test. The LRS method has been used to estimate the time interval of HSR channels in [25]. The RUN test was applied to the RMS DS data to identify the stationary distance in the station scenario [24]. In this paper, we choose the classical LRS approach to characterize the SI in the measured scenarios.

The aim of the LRS method is to find the maximum interval within which the correlation coefficient between two consecutive PDPs exceeds a predefined threshold  $c_{th}$ . The correlation coefficient between PDPs is defined as

$$c(x_k, \Delta x) = \frac{\overline{P(x_k)P(x_k + \Delta x)}}{\max\left\{\overline{P(x_k)^2}, \overline{P(x_k + \Delta x)^2}\right\}}, \quad (1)$$

where

$$\overline{P(x_k)} = \frac{1}{N} \sum_{m=k}^{k+N-1} |h(x_m)|^2, \quad (2)$$

where  $N$  is the window size and  $h(x_m)$  are the samples of channel impulse response.

Then, the SI can be estimated as

$$I_k = (k_{\max} - k_{\min}) \Delta x \quad (3)$$

where

$$k_{\max} = \arg \max_{k+1 \leq m \leq L-N} c(x_k, m\Delta x) < c_{th} \quad (4)$$

$$k_{\min} = \arg \min_{1 \leq m \leq k-1} c(x_k, m\Delta x) < c_{th} \quad (5)$$

TABLE 3: Analysis results of the SI in different scenarios.

Scenario	Method	Statistics	SI (m)		
			$c_{th} = 0.7$	$c_{th} = 0.8$	$c_{th} = 0.9$
Rural	LRS	Mean value	14.3	6.46	2.84
		60% of CCDF	8.22	4.29	1.65
		80% of CCDF	4.74	2.64	0.99
Station	LRS	Mean value	7.27	3.84	1.87
		60% of CCDF	5.02	2.55	1.17
		80% of CCDF	3.07	1.25	0.37
Suburban	LRS	Mean value	7.62	3.74	1.37
		60% of CCDF	5.43	2.33	0.86
		80% of CCDF	3.54	1.25	0.35
Viaduct [26]	LRS	60% of CCDF	-	1.8	-
		80% of CCDF	-	0.81	-
Standard models [26]	LRS	60% of CCDF	-	3.4	-
Station [25]	RUN test	80% of CCDF		4	

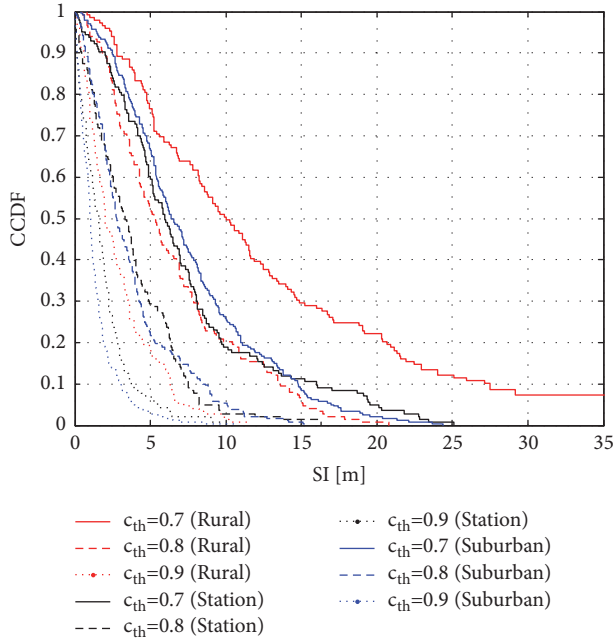


FIGURE 3: CCDFs of SI in rural, station, and suburban scenarios.

where  $L$  is the length of used data. Here, we consider three typical correlation threshold values:  $c_{th} = 0.7$ ,  $c_{th} = 0.8$ , and  $c_{th} = 0.9$ .

Figure 3 compares the derived SI results for different  $c_{th}$  and scenarios in terms of complementary CDF (CCDF). With the increase of  $c_{th}$ , the SI is gradually decreasing. For  $c_{th} = 0.9$ , only slight difference can be observed in different environments, while there is an obvious deviation for  $c_{th} = 0.7$  and  $c_{th} = 0.8$ . Focusing on the cases of  $c_{th} = 0.7$  and  $c_{th} = 0.8$ , we find that the SI values in the rural scenario are larger than those in the station or suburban scenarios. This is understandable from the completely different scattering environment. Since LoS is dominant in the rural scenario,

the MPCs have smaller dynamic changes over time, and thus the stationary distance is longer. When it comes to the station and suburban scenarios, due to the rich reflection and scattering components from the awnings or buildings, the nonstationarity is more serious, and thus the SI decreases. This nonstationarity could be originated from a special physical phenomena, for example, “appearance and disappearance” or “birth and death” of MPCs [26], which will be further investigated in the following subsection.

The detailed statistical SI results including mean value and 60% and 80% of CCDF for different scenarios are listed in Table 3. For  $c_{th} = 0.8$ , the mean value of SI is 6.46 m in the rural scenario, while those are 3.84 m and 3.74 m in the station and suburban scenarios. In 60% and 80% of cases, the channel could be stationary over a distance of 2.33-4.29 m and 1.25-2.64 m for the measured scenarios, respectively. These values are higher than the results of 1.8 m for 60% and 0.81 m for 80% reported in [25]. From [25], the calculated stationary interval for standard channel models is equal to 3.4 m for 60%, which is shorter than the one of 4.29 m for the rural scenario but is longer than the ones of 2.55 m and 2.33 m for the station and suburban scenarios. It is also observed that the value of SI in 80% of the cases in the measured station scenario is smaller than that of around 4 m in the station scenario reported in [24]. This variance could be due to the use of different calculation methods.

**4.2. Birth-Death Process.** The nonstationarity of the channel is basically due to the dynamic evolution of MPCs when the Rx is in motion, for example, appearance to disappearance or B-D. To describe this B-D process, a four-state Markov chain model (MCM) is used, where each state is defined as follows [30]:

- (i)  $S_0$ : no “births” or “deaths”
- (ii)  $S_1$ : “births” only
- (iii)  $S_2$ : “deaths” only
- (iv)  $S_3$ : both “births” and “deaths”

TABLE 4: Analysis results of the B-D process in different scenarios.

Scenario	State transition probability matrix	Steady-state probability
Rural	$\begin{bmatrix} 0.3361 & 0.4025 & 0.1203 & 0.1411 \\ 0.0497 & 0.0315 & 0.5348 & 0.3841 \\ 0.1658 & 0.7148 & 0.0282 & 0.0912 \\ 0.0332 & 0.0600 & 0.2505 & 0.6563 \end{bmatrix}$	$[0.1017 \quad 0.2533 \quad 0.2530 \quad 0.3920]$
Station	$\begin{bmatrix} 0.2791 & 0.2093 & 0.1628 & 0.3488 \\ 0.0224 & 0.0096 & 0.4936 & 0.4744 \\ 0.0354 & 0.7235 & 0.0129 & 0.2283 \\ 0.0081 & 0.0471 & 0.0905 & 0.8544 \end{bmatrix}$	$[0.0189 \quad 0.1374 \quad 0.1366 \quad 0.7071]$
Suburban	$\begin{bmatrix} 0.1870 & 0.2439 & 0.1463 & 0.4228 \\ 0.0557 & 0.0557 & 0.2256 & 0.6630 \\ 0.0780 & 0.4150 & 0.0613 & 0.4457 \\ 0.0151 & 0.0464 & 0.0690 & 0.8695 \end{bmatrix}$	$[0.0287 \quad 0.0837 \quad 0.0837 \quad 0.8039]$
Viaduct [17]	$\begin{bmatrix} 0.2917 & 0.5208 & 0.1667 & 0.0208 \\ 0.0200 & 0.1000 & 0.5200 & 0.3600 \\ 0.2577 & 0.4330 & 0.0928 & 0.2165 \\ 0.0673 & 0.2115 & 0.2788 & 0.4423 \end{bmatrix}$	$[0.1371 \quad 0.2800 \quad 0.2857 \quad 0.2971]$

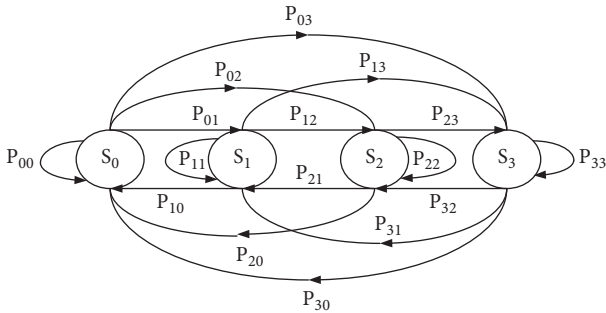


FIGURE 4: State transition diagram of the four-state MCM.

Note that the state in the MCM only considers the variation of MPCs from the current moment to the next moment. With the motion of the Rx, the states can be transformed to each other. Figure 4 illustrates the state transition diagram of the four-state MCM [30]. The probabilistic switching process between states in the MCM is controlled by the state transition probability matrix  $\mathbf{P}$  given by

$$\mathbf{P} = \{p_{ij}\} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{bmatrix} \quad (6)$$

where  $i$  and  $j$  represent the state index, while  $p_{ij}$  is the transition probability from state  $S_i$  to state  $S_j$ . Note that  $p_{ij}$  must satisfy the following requirement:

$$0 \leq p_{ij} \leq 1, \quad i, j = 0, 1, \dots, N-1 \quad (7)$$

$$\sum_{j=0}^{N-1} p_{ij} = 1, \quad i = 0, 1, \dots, N-1 \quad (8)$$

where  $N$  is the number of states; that is,  $N = 4$  in our case.

The steady-state probability can be expressed as

$$\mathbf{P}_S = [P_{S_0} \quad P_{S_1} \quad P_{S_2} \quad P_{S_3}] \quad (9)$$

which satisfies  $P_{S_0} + P_{S_1} + P_{S_2} + P_{S_3} = 1$ . Each element in  $\mathbf{P}_S$  indicates the overall state occupancy probability.

The obtained results of the state transition probability matrix in different scenarios are listed in Table 4. Here, the state transition matrix is derived from each CIR. Since the sample rate of CIR is 2000 Hz and the velocity of train is 79 m/s, the reference value for state transition matrix is 0.04 m. It is observed that in the suburban scenario the next state is more likely to transit into the state  $S_3$  no matter what the current state is  $S_0, S_1$ , or  $S_2$ . This means some new MPCs are born and older MPCs die most of the time because of the rich reflection and scattering components in the suburban scenario. For other scenarios, in the case of  $S_0$ , the next state could be any one of the four states; in the case of  $S_1, S_2$  or  $S_3$  have the maximum probability to be the next state; in the case of  $S_2$ , the next state is more likely to transit into  $S_1$ ; in the case of  $S_3$ , the next state is still likely to be  $S_3$ . The results show the nonsymmetric transition matrix, which means that the transition probability from state A to state B has no relationship with that from state B to state A.

According to the state transition probability matrix, the steady-state probability can be derived, as listed in Table 4. It is found that  $S_3$  is the most likely state in the station and suburban scenarios. For the rural scenario, either  $S_1, S_2$ ,

or  $S_3$  could be the steady state. It is worth noting that  $S_1$  and  $S_2$  have approximately similar steady-state probability in any scenario. This confirms that the appearance and disappearance of MPCs are equivalent. The above results in the rural scenario are similar to those in the open viaduct scenario reported in [26]. The obtained results can be applied for ON/OFF tapped delay line models that use the Markov chain to model the ON/OFF process of MPCs [31].

## 5. Conclusion

This paper analyzes the nonstationary characteristics in typical HSR scenarios, rural, station, and suburban, depending on the passive LTE-based channel measurements. With regard to the nonstationary characteristics, it is found that the SI is longest in the rural scenario. Additionally, a four-state MCM is established to characterize the B-D process of MPCs, and the corresponding state transition probability matrix and steady-state probability are provided. These results show the realistic channel characteristics in the HSR communication network, which will provide helpful information for nonstationary channel modeling of HSR communication systems.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant 2018JBZ102), the National Natural Science Foundation of China (Grant 61701017), the Beijing Municipal Natural Science Foundation (Grant 4174102), and the Open Research Fund through the National Mobile Communications Research Laboratory, Southeast University (Grant 2018D11).

## References

- [1] R. He, B. Ai, G. Wang et al., "High-Speed railway communications: from GSM-R to LTE-R," *IEEE Vehicular Technology Magazine*, vol. 11, no. 3, pp. 49–58, 2016.
- [2] B. Ai, X. Cheng, T. Kurner et al., "Challenges toward wireless communications for high-speed railway," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2143–2158, 2014.
- [3] *White Paper for 5G High Mobility*, [Online], Available, <http://www.future-forum.org>.
- [4] H. Wei, Z. Zhong, K. Guan, and B. Ai, "Path loss models in viaduct and plain scenarios of the High-speed Railway," in *Proceedings of the 5th International ICST Conference on Communications and Networking in China*, Beijing, China, August 2010.
- [5] R. He, Z. Zhong, B. Ai, and J. Ding, "An empirical path loss model and fading analysis for high-speed railway viaduct scenarios," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 808–812, 2011.
- [6] K. Guan, Z. Zhong, B. Ai, and T. Kurner, "Empirical models for extra propagation loss of train stations on high-speed railway," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1395–1408, 2014.
- [7] R. He, Z. Zhong, B. Ai, and C. Oestges, "Shadow fading correlation in high-speed railway environments," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 2762–2772, 2015.
- [8] L. Liu, C. Tao, D. W. Matolak, T. Zhou, and H. Chen, "Investigation of Shadowing Effects in Typical Propagation Scenarios for High-Speed Railway at 2350 MHz," *International Journal of Antennas and Propagation*, vol. 2016, pp. 1–8, 2016.
- [9] O. Renaudin, V.-M. Kolmonen, P. Vainikainen, and C. Oestges, "Non-stationary narrowband MIMO inter-vehicle channel characterization in the 5-GHz band," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 2007–2015, 2010.
- [10] D. W. Matolak, "Modeling the vehicle-to-vehicle propagation channel: A review," *Radio Science*, vol. 49, no. 9, pp. 721–736, 2014.
- [11] L. Liu, C. Tao, J. Qiu et al., "Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 834–845, 2012.
- [12] T. Zhou, C. Tao, L. Liu, and Z. Tan, "A semiempirical MIMO channel model in obstructed viaduct scenarios on high-speed railway," *International Journal of Antennas and Propagation*, vol. 2014, Article ID 287159, 10 pages, 2014.
- [13] T. Zhou, C. Tao, S. Salous, Z. Tan, L. Liu, and L. Tian, "Graph-based stochastic model for high-speed railway cutting scenarios," *IET Microwaves, Antennas & Propagation*, vol. 9, no. 15, pp. 1691–1697, 2015.
- [14] R. Sun, C. Tao, L. Liu, and Z. Tan, "Channel Measurement and Characterization for HSR U-Shape Groove Scenarios at 2.35 GHz," in *Proceedings of the 2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Las Vegas, NV, USA, September 2013.
- [15] T. Zhou, C. Tao, S. Salous, and L. Liu, "Spatial Characterization for High-Speed Railway Channels Based on Moving Virtual Array Measurement Scheme," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 1423–1426, 2017.
- [16] R. He, Z. Zhong, B. Ai, G. Wang, J. Ding, and A. F. Molisch, "Measurements and analysis of propagation channels in high-speed railway viaducts," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 794–805, 2013.
- [17] R. He, Z. Zhong, B. Ai, J. Ding, Y. Yang, and A. F. Molisch, "Short-term fading behavior in high-speed railway cutting scenario: Measurements, analysis, and statistical models," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 2209–2222, 2013.
- [18] P. Kyösti et al., *WINNER II channel models part II radio channel measurement and analysis results*, vol. ST-4-027756, 2007, WINNER II D1.1.2, v1.0.
- [19] R. Parviainen, P. Ky, Y. Hsieh, P. Ting, and J. Chiou, "Results of high speed train channel measurements," in *Proc. COST 2100 TD*, France, Lille, 2008.
- [20] F. Luan, Y. Zhang, L. Xiao, C. Zhou, and S. Zhou, "Fading characteristics of wireless channel on high-speed railway in hilly terrain scenario," *International Journal of Antennas and Propagation*, vol. 2013, Article ID 378407, 9 pages, 2013.



- [21] Y. Zhang, Z. He, W. Zhang, L. Xiao, and S. Zhou, "Measurement-based delay and doppler characterizations for high-speed railway hilly scenario," *International Journal of Antennas and Propagation*, vol. 2014, Article ID 875345, 8 pages, 2014.
- [22] P. Aikio, R. Gruber, and P. Vainikainen, "Wideband radio channel measurements for train tunnels," in *Proceedings of the VTC '98. 48th IEEE Vehicular Technology Conference. Pathway to a Global Wireless Revolution*, pp. 460–464, Ottawa, Ont., Canada.
- [23] K. Guan, Z. Zhong, B. Ai, and T. Kürner, "Propagation measurements and modeling of crossing bridges on high-speed railway at 930 MHz," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3349–3516, 2014.
- [24] T. Zhou, C. Tao, S. Salous, L. Liu, and Z. Tan, "Channel characterization in high-speed railway station environments at 1.89 GHz," *Radio Science*, vol. 50, no. 11, pp. 1176–1186, 2015.
- [25] B. Chen, Z. Zhong, and B. Ai, "Stationarity intervals of time-variant channel in high speed railway scenario," *China Communications*, vol. 9, no. 8, pp. 64–70, 2012.
- [26] L. Liu, C. Tao, J. Qiu, T. Zhou, R. Sun, and H. Chen, "The dynamic evolution of multipath components in High-Speed Railway in viaduct scenarios: From the birth-death process point of view," in *Proceedings of the 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC 2012)*, pp. 1774–1778, Sydney, Australia, September 2012.
- [27] T. Zhou, C. Tao, S. Salous, L. Liu, and Z. Tan, "Channel sounding for high-speed railway communication systems," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 70–77, 2015.
- [28] T. Zhou, C. Tao, S. Salous, L. Liu, and Z. Tan, "Implementation of an LTE-based channel measurement method for high-speed railway scenarios," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 25–36, 2016.
- [29] R. He, O. Renaudin, V.-M. Kolmonen et al., "Characterization of quasi-stationarity regions for vehicle-to-vehicle radio channels," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 5, pp. 2237–2251, 2015.
- [30] C.-C. Chong, C.-M. Tan, D. I. Laurenson, S. McLaughlin, M. A. Beach, and A. R. Nix, "A novel wideband dynamic directional indoor channel model based on a Markov process," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1539–1552, 2005.
- [31] I. Sen and D. Matolak, "Vehicle channel models for the 5-ghz band," *IEEE T. Intell. Transp.*, vol. 9, no. 2, pp. 235–245, June 2008.

## Research Article

# Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments

Yan Zhang , Jinxiao Wen , Guanshu Yang , Zunwen He , and Xinran Luo

*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Zunwen He; [hezunwen@bit.edu.cn](mailto:hezunwen@bit.edu.cn)

Received 6 April 2018; Accepted 14 May 2018; Published 13 June 2018

Academic Editor: Liu Liu

Copyright © 2018 Yan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, unmanned aerial vehicle (UAV) plays an important role in many applications because of its high flexibility and low cost. To realize reliable UAV communications, a fundamental work is to investigate the propagation characteristics of the channels. In this paper, we propose path loss models for the UAV air-to-air (AA) scenario based on machine learning. A ray-tracing software is employed to generate samples for multiple routes in a typical urban environment, and different altitudes of Tx and Rx UAVs are taken into consideration. Two machine-learning algorithms, Random Forest and KNN, are exploited to build prediction models on the basis of the training data. The prediction performance of trained models is assessed on the test set according to the metrics including the mean absolute error (MAE) and root mean square error (RMSE). Meanwhile, two empirical models are presented for comparison. It is shown that the machine-learning-based models are able to provide high prediction accuracy and acceptable computational efficiency in the AA scenario. Moreover, Random Forest outperforms other models and has the smallest prediction errors. Further investigation is made to evaluate the impacts of five different parameters on the path loss. It is demonstrated that the path visibility is crucial for the path loss.

## 1. Introduction

In recent years, unmanned aerial vehicles (UAVs), as aircraft without pilots on board, have shown great promise due to their high mobility and deployment flexibility. With the development of UAV manufacturing, its cost is reduced while its performance continuously increases. As a result, there are more and more attractive applications for UAV, such as traffic monitoring, emergency rescue, forest fire detection, cargo transport, and so on [1, 2]. Stable and efficient wireless communication links are indispensable in most UAV applications. Therefore, the UAV communications play an important role in the future fifth-generation wireless networks (5G), providing vast coverage and reliable relaying.

Meanwhile, the propagation environment of UAV-aided communication systems differs from that of traditional ones, which brings enormous challenges. An accurate understanding of the UAV wireless channels is crucial for the design and deployment of these communication systems. The wireless signals from/to UAVs may be obstructed and

may encounter different propagation conditions along the path. The attenuation of the electromagnetic wave, which is usually described by the path loss, is of great significance for the link budget analysis and network planning for UAV communications. Therefore, many works have been finished to develop flexible and precise models for the path loss in the UAV communication scenarios.

The high altitude range of 500 m to 2000 m was considered in [3], and the air-to-ground (AG) channel based on the curved-earth two-ray model was investigated in various scenarios. In [4], a statistical propagation model was proposed for the UAV channel at low altitude in the urban environment, and it was shown that the prediction results were dependent on the elevation angle between the airborne transmitter (Tx) and the receiver (Rx) on the ground. In [5], the impact of the UAV altitude on path loss exponent and shadow fading in the rural scenario was studied. In [6], statistical path loss models were established by modifying the current 3GPP terrestrial channel models for urban macrocell and rural macrocell scenarios. In [7, 8], measurements were

carried out in suburban scenarios, and large-scale parameters and multipath components were extracted and analyzed. In [9], measurement campaigns were conducted in an urban scenario and a distance-dependent model was proposed for the UAV path loss prediction.

Most of these aforementioned works are focused on the AG communication. An alternative application approach of UAVs is to use them as both sides of the communication, i.e., air-to-air (AA) communication. Until now, only a few papers have investigated the channel models in the AA scenario. In [10], Rice model was extended to derive the AA channel parameters and it was reported that the attenuation caused by the distance effect followed a free-space model. In [11], with data generated by a ray-tracing software, the close-in free-space model and excess fading loss model were adopted to characterize the path loss in the AA channel.

Furthermore, the existing works are mainly based on empirical models, such as free-space model and log-distance model, which rely on data collected in specific propagation scenarios. Statistical analysis is performed to build the mapping relationship between path loss and parameters such as propagation distance and flight altitude. The empirical models are computationally efficient and easy to implement. They can describe the statistical characteristics of the path loss at a given distance in the measured scenario. However, the actual path loss at a specific location cannot be obtained. Besides, the accuracy of these models decreases when they are applied to more general environments [12].

Another candidate solution is to utilize deterministic approaches, such as ray tracing and finite-difference time-domain (FDTD), within which the path loss values are calculated by applying radio wave propagation mechanisms and numerical analysis techniques to model computational electromagnetics. With detailed geographic information and dielectric properties of materials, these methods are very accurate and reliable for predicting the spatial distribution of electromagnetic fields. Due to the high cost of carrying out measurement campaigns, the deterministic approaches have been widely used for wireless network planning. The only disadvantage is that the computation procedure consumes huge time and memory resources and thus it is inappropriate to use these approaches for real-time applications. Moreover, the complicated calculation has to be run again once the propagation environment changes.

Actually, path loss modeling is a supervised regression problem and can be solved by machine learning [13]. It has been proved that machine-learning-based models are able to provide more accurate path loss prediction results than the empirical ones and are more computationally efficient than the deterministic approaches [12]. Different algorithms have been adopted to train prediction models in traditional terrestrial communication scenarios. For example, artificial neural networks (ANNs) were used for path loss prediction in urban [14], suburban [15], rural [16], and railway [17] scenarios. Support vector regression (SVR) was applied for the prediction of path loss in suburban environment in [18]. In order to build a connectivity model for an environmental wireless sensor network, several methods, including Random Forest, Adaboost, ANNs, and K-Nearest-Neighbors (KNN),

were analyzed and compared in [13]. It was reported that Random Forest performed better than others for the considered complex terrain environments.

In this paper, we build the prediction models for path loss in the AA scenario based on machine learning. Two algorithms, Random Forest and KNN, are taken into consideration. To evaluate the feasibility of the proposed models, the ray-tracing approach is used to generate data for training and testing purposes. In addition, the prediction accuracies of machine-learning-based models are compared with those of the empirical ones, such as the Stanford University Interim (SUI) model [19] and the COST231-W-I model [20]. It is shown that the proposed models outperform the empirical ones. Furthermore, we analyze the commonly used parameters related to the path loss in the AA scenario, including the propagation distance, Tx UAV altitude, Rx UAV altitude, path visibility, and elevation angle. Meanwhile, the importance of these parameters is discussed.

We summarize the major contributions and novelties of this paper as follows.

- (1) The path loss for the UAV communication in the AA scenario is modeled based on machine learning methods, including Random Forest and KNN algorithms.
- (2) The prediction results are evaluated with the data generated by a ray-tracing software. It is proved that the machine-learning-based models are able to provide better accuracy than the empirical ones.
- (3) We analyze the impacts of different parameters on the AA path loss and sort these parameters by their importance.

The remainder of this paper is organized as follows. The considered AA propagation environment and the ray-tracing-based data generation are described in Section 2. Section 3 presents the machine-learning-based methods for path loss prediction. The model training procedure is introduced in Section 4. In Section 5, the performance of machine-learning-based models is evaluated and the importance ranking of different parameters is discussed. At last, conclusions are drawn in Section 6.

## 2. Propagation Environment Description

In order to investigate the path loss model in the AA scenario, we consider a typical urban environment in which two UAVs are employed as Tx and Rx. As a new emerging scenario, measurements for AA communications are still in a very preliminary stage. Since the machine-learning-based models require a large amount of data for training purpose, a ray-tracing software is employed to generate data for model building and performance evaluation. It has been proved that the channel data obtained by the ray-tracing software are in good agreement with the actual measured values [21]. As illustrated in Figure 1, the considered environment is a region in Helsinki, with dimensions of 1000 m by 600 m. Gray areas and green areas indicate buildings and grounds, respectively. The maximum height of the buildings is 50 m. All the buildings are assumed to be made of concrete with the following dielectric half-space properties: permittivity 6, conductivity 0.02, and thickness 0.3 m. The material of the



FIGURE 1: Urban environment for AA communications.

ground surface is assigned as asphalt whose permittivity is 10 and conductivity is 0.01.

The simulations were performed at the central frequency of 2.4 GHz, with a bandwidth of 100 MHz. The red square in Figure 1 represented the position of the Tx UAV, which was equipped with a directional antenna. Rx UAV was moved at a spacing of 2 m along six different routes. Different flight altitudes were taken into account for both Tx and Rx UAVs. The altitudes of the Tx UAV included 60 m, 70 m, and 80 m. Meanwhile, Rx UAV were assumed to fly at heights of 10 m, 20 m, 30 m, and 40 m, lower than the maximum height of the buildings. It should be noted that the direct, reflected, and diffracted paths were considered, whereas the penetration paths were neglected because of the high attenuation through building. Details of the parameter setting can be found in Table 1.

Through calculations, we obtain the spatial distribution of received powers. Then, path loss values at different locations can be extracted. In practice, the path loss in the AA scenario is related to many environmental parameters. The goal of the machine learning method is to find the optimal function describing the relationship between these parameters and the path loss. In the following analysis, five parameters which have impacts on the path loss are selected as the input features of the machine-learning-based models and they are listed as follows.

- (1) **Propagation distance** ( $d$ , in meter): the distance between the Tx and Rx UAVs calculated from their coordinates.
- (2) **Tx altitude** ( $h_t$ , in meter): the height of Tx UAV from the ground, with three values of 50 m, 60 m, and 70 m.

TABLE 1: Parameter configuration.

Parameter	Value
Environment	Helsinki urban scenario
Area	1000 m $\times$ 600 m
Max. building height	50 m
Carrier frequency	2.4 GHz
Bandwidth	100 MHz
Transmit power	15 dBm
Tx altitude	60, 70, 80 m
Rx altitude	10, 20, 30, 40 m
Distance between adjacent Rx positions	2 m
Number of Tx locations	1 $\times$ 3
Number of Rx locations	405 $\times$ 4
Max. number of reflection	10
Max. number of diffraction	1
Max. penetration	Not simulated

- (3) **Rx altitude** ( $h_r$ , in meter): the height of Rx UAV from the ground, with four values of 10 m, 20 m, 30 m, and 40 m.
- (4) **Path visibility** ( $I_v$ , 0 or 1): parameter indicating whether there exists line-of-sight (LOS) path between the Tx and Rx UAVs.  $I_v = 1$  for the LOS case and  $I_v = 0$  for the non-line-of-sight (NLOS) case.
- (5) **Elevation angle** ( $\theta$ ,  $-\pi/2$  to  $\pi/2$ ): the angle between the LOS path and the horizontal.

We collected all the samples when the Tx UAV was in different altitudes and the Rx UAV flew along six routes at

**Input:**

Training set  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  with responses  $Y = \{PL_1, PL_2, \dots, PL_N\}$ , where  $\mathbf{x}_i = (d_i, h_{ii}, h_{ri}, I_{vi}, \theta_i)$ ,  $i = 1, \dots, N$ .  
Number of ensemble members  $T$ .

**Training Process:**

For  $t = 1$  to  $T$ :

- (1) Take a bootstrap sample  $\{X_t, Y_t\}$  of size  $N$  from  $\{X, Y\}$ .
- (2) Use  $\{X_t, Y_t\}$  as the training data to train the  $t$ th ensemble member by using binary recursive partitioning.
- (3) Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
  - (i) Select  $m$  features randomly from the  $f$  available features ( $f = 5$  in this study).
  - (ii) Calculate the square error for each possible splitting point of each feature, and find the best binary split among all binary splits on the  $m$  features.
  - (iii) Split the node into two descendant nodes using the best split.

**Prediction:**

Given a new  $\mathbf{x} = (d, h, h_r, I_r, \theta)$ , the predicted path loss value is obtained by  $PL' = (1/T) \sum_{t=1}^T \hat{h}_t(\mathbf{x})$ , where  $\hat{h}_t(\mathbf{x})$  is the prediction of the  $t$ th ensemble member.

ALGORITHM 1: Random Forest algorithm for path loss prediction in the AA scenario.

different heights. Each sample was with an output (path loss value) and five input features. Removing the locations where the received signals are too weak to detect, we obtained 5508 samples in total. Then, these samples were separated into two set, training set and test set. The former were utilized to train the models while the latter were employed to evaluate the performance of the trained models.

### 3. Machine-Learning-Based Models for AA Path Loss Prediction

Machine learning is a method to improve performance on a specific task based on extensive data and a flexible model architecture. In recent years, it has been widely used in many fields like computer vision, speech recognition, autonomous driving, and so on. Machine learning tasks can be broadly classified into supervised learning and unsupervised learning, depending on whether data samples have labels or not. For supervised learning, tasks can be further divided into classification problems and regression problems based on whether the predicted values are discrete or continuous. The AA path loss prediction is a typical regression task, which can be solved by many algorithms, such as Random Forest, ANN, and SVR. We aim to build the path loss prediction model in the AA scenario based on machine learning. With given path loss values and corresponding input features, the model can be trained and then the path loss values in new conditions can be predicted with the various inputs.

In this study, two typical supervised learning algorithms, Random Forest and KNN, are chosen to build prediction models for the AA path loss. Their performance evaluation results will be compared in Section 5 and Random Forest will be proved to have a better agreement with the test data compared with KNN. The major principles of these two algorithms are introduced as follows.

**3.1. Random Forest.** Ensemble learning, which uses multiple individual learners to solve classification and regression problems, can achieve a significantly superior generalization performance [22]. Random Forest is a commonly used ensemble

learning algorithm and employs decision tree as ensemble member. It applies bootstrap aggregating to select training samples for each ensemble member. Ensemble members are trained based on these samples and then the final result is obtained by averaging the results of all the ensemble members.

Besides, Random Forest further introduces random selection of features in the decision tree training process. Usually, the traditional decision tree selects an optimal feature from the feature set of current node for split. For Random Forest, a subset is randomly chosen from the feature set of each node, and then the optimal feature is selected from this subset.

In ensemble learning, the greater the diversity of ensemble members, the better the prediction performance. By introducing sample perturbations and feature perturbations, the diversity of ensemble members in Random Forest is increased. It can improve the generalization performance of the model. It is worth noting that the decision trees grow without pruning, due to the randomness of samples and features. Algorithm 1 shows the method we used for path loss prediction in the AA scenario. The detailed descriptions of Random Forest can be found in [23].

Random Forest is easy to implement and can realize parallel computing. It is also insensitive to input data and can handle thousands of input features. In addition, an important advantage of Random Forest is that it can sort the importance of features. In the following, we will use this property to analyze the significance of different input features for the AA path loss.

**3.2. KNN.** KNN is a classical machine learning algorithm that is often used to solve classification problems. It has no explicit training process and its implementation is simple. The mechanism of KNN is to find the  $k$  training samples closest to the sample to be predicted based on a distance metric and then to perform prediction based on the information of these  $k$  neighbors. It is also suitable for regression tasks by averaging the values of  $k$  neighbors to get the final prediction result.

The distance metric plays a very important role in KNN. The distance reflects the difference between two samples. Commonly used distance metrics include Manhattan distance, Euclidean distance, and so on. In this study, Euclidean distance is chosen for analysis. In general, features have different ranges of values and their influences on the distance calculation are not the same. As a result, KNN algorithm is more sensitive to the input data compared with Random Forest. For the sake of fairness, the samples need to be normalized before model training. In this study, Z-score normalization method is adopted and it can be expressed as

$$x_N = \frac{x_i - \mu}{\sigma} \quad (1)$$

where  $x_i$  is the input value of the feature,  $x_N$  is the normalized value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

#### 4. Model Training and Accuracy Metrics

The procedure of machine-learning-based path loss predictors for AA channel is introduced as follows. Firstly, we collect enough data samples for analysis, each with path loss record and corresponding input features. As mentioned above, for KNN the features need to be scaled by the normalization process, while for Random Forest it is not necessary. Secondly, these samples can be divided into two categories, training set and test set, which are used for model training and evaluation purposes, respectively. Thirdly, based on the training data and selected algorithms, we train the model and tune its parameters. Finally, some metrics are employed to assess the prediction accuracy of the trained model, and then in view of the evaluated results we can further improve the machine-learning-based predictor for the path loss in the AA scenario.

In this section, we will introduce the division of training set and test set. Then, the model training process is explained in detail. In addition, accuracy metrics for model validation are presented.

**4.1. Data Division.** The performance of the machine-learning-based models strongly depends on the amount and quality of training data. In general, more training samples lead to more accurate reflections of the inherent laws. Thus, we must try to obtain enough samples in order to get accurate models for path loss prediction. In addition, the rules extracted from the model training are hidden in the samples, so the training samples must be representative. Different from the training data, the test set is used to assess and further improve the trained models.

As aforementioned, 5508 samples were collected in the considered AA scenario, including samples from six routes at all different Tx/Rx altitudes. In this study, the samples from the third route with the Rx altitudes of 20 m and 30 m were used for test purpose, and they did not participate in the training process. The remaining samples were included in the training set. Then, the proportions of the training samples and test ones were 84% and 16%.

**4.2. Model Training.** Model training aims at acquiring parameters for the model to optimize the performance and effectiveness of the path loss prediction. Some machine learning algorithms, such as ANN and SVR, have many parameters whose values need to be set before the learning process begins. In contrast, Random Forest and KNN have only a few parameters that need to be tuned and thus they are both efficient for implementation.

For Random Forest, the model accuracy is affected by the parameters including maximum tree depth and the number of ensemble members. The former controls the maximum split number of the decision tree and the latter determines the size of the ensemble. Generally, a small ensemble with deep decision trees has a greater tendency to overfit than a shallow ensemble of many decision trees [13].

For KNN, the number of neighbors,  $k$ , is deterministic for the prediction performance. If  $k$  is too small, the model becomes complicated and may overlearn when the neighboring points are noises. Meanwhile, large  $k$  makes the model structure simple but neighboring samples with large differences will affect the prediction result.

These aforementioned parameters cannot be learned directly from the data. The optimization methods for tuning parameters mainly include grid search, random search, and Bayesian optimization. In this study, grid search is used to find the optimal combination of parameters by searching all possible points in the given range. For Random Forest, we evaluate the following parameters: the number of ensemble members between 10 and 200 (at 10-unit intervals) and the maximum tree depths between 5 and 50 (at 5-unit intervals). For KNN, the value of  $k$  is set between 2 to 10 with the interval of 1. The obtained parameters are as follows. For Random Forest, the depth of trees and the number of ensemble members are finally set as 30 and 140, respectively. The number of neighbors is equal to 5 in KNN.

**4.3. Metrics for Evaluating Prediction Accuracy.** To evaluate the performance of different models, two statistical properties, mean absolute error (MAE) and root mean square error (RMSE) [24], are chosen as metrics. They can be calculated by comparing the predicted path loss with the data in the test set as

$$\text{MAE} = \frac{1}{I} \sum_{i=1}^I |PL_i - PL_i'|, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{I} \sum_{i=1}^I (PL_i - PL_i')^2} \quad (3)$$

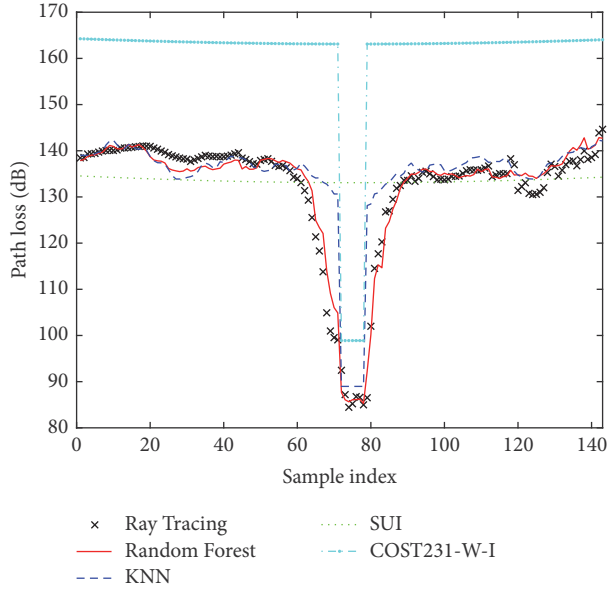
where  $I$  is the total number of test samples,  $PL_i$  is the path loss value of the  $i$ th sample in the test set, and  $PL_i'$  is the predicted value.

### 5. Model Validation and Results

In this section, we will evaluate the performance of these machine-learning-based models in the AA scenario. Two empirical models are also considered for comparison. In

TABLE 2: Statistics analysis for different predictors.

Evaluation indicators	Random Forest	KNN	SUI	COST231-W-I
MAE (dB)	2.27	4.56	7.54	26.67
RMSE (dB)	3.06	8.90	13.40	28.53

FIGURE 2: Prediction performance of different models when Rx UAV moves in the third route ( $h_t = 60$  m and  $h_r = 30$  m).

addition, the impacts of different features on the path loss are analyzed.

*5.1. Comparisons between Empirical Models and Machine-Learning-Based Models.* As an example, we consider the samples in the test set gathered when the Tx and Rx altitudes are 60 m and 30 m, respectively. The predicted path loss results from different models are shown in Figure 2. Sample indexes are corresponding to different positions of Rx UAV in the third route from up to down in Figure 1. As mentioned in Section 2, the distance between two adjacent Rx positions is 2 m. As shown in Figure 1, LOS path exists when the Rx UAV moves in the middle, corresponding to the sample index from 72 to 79. It can be found that the path loss values are quite small in this area. In Figure 2, it is illustrated that the machine-learning-based models can accurately approximate the realistic path loss values generated by the ray-tracing software. Two empirical models, SUI model and COST231-W-I model, are chosen for comparison. The description of whether the LOS path exists is not included in the SUI model. Thus, there are large gaps from the path loss results predicted by the SUI model to the true values under the LOS condition. The COST231-W-I model can describe the path loss variations in both LOS and NLOS conditions. Apart from the path visibility, this model also involves the distance, Tx altitude, Rx altitude, and elevation angle into the model parameters. However, its predicted results are much larger than the true values. The major reason may be

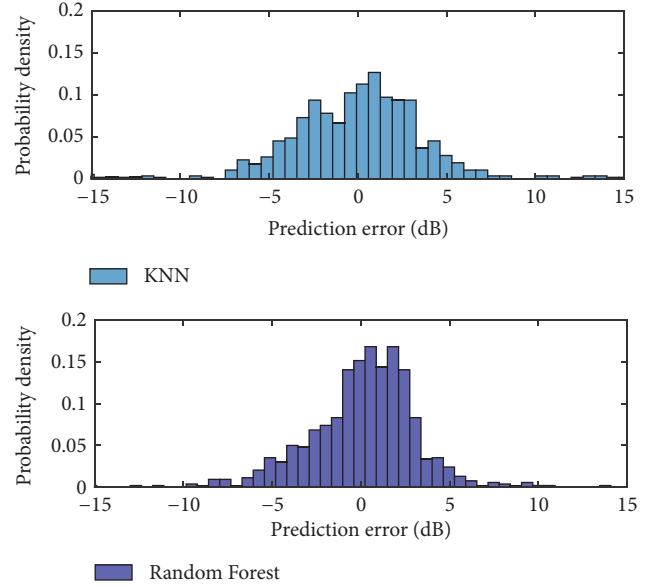


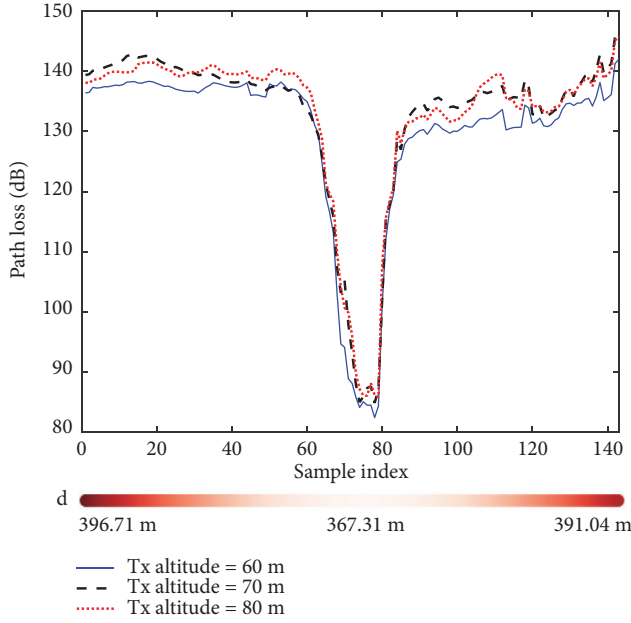
FIGURE 3: Prediction error distributions for the machine-learning-based models.

the fact that the application scenarios of the COST231-W-I model differ from what we used during the analysis. It reflects the poor generalization performance of the empirical model; i.e., its accuracy decreases when it is applied to a different environment. Furthermore, as shown in Figure 2, these empirical models neglect some details of the path loss fluctuations and it is difficult to use them for characterizing the path loss value at a specific location.

Considering all samples in the test set, we can get the statistical assessment of these different models. The MAEs and RMSEs of prediction results are illustrated in Table 2. It is shown that both Random Forest and KNN outperform the empirical models. These machine-learning-based models can also depict the fluctuations of path loss in detail.

*5.2. Comparisons between Random Forest and KNN.* As listed in Table 2, Random Forest provides the best fit to the test data, with 2.27 dB MAE and 3.06 dB RMSE. KNN also offers acceptable results whereas its predicted values are almost unchanged under the LOS condition. The reason is that, within the KNN model, the path loss is predicted by averaging the values of the nearest  $k$  neighbors. Due to the limited number of collected LOS samples, a similar path loss value is probably estimated.

Figure 3 shows the distributions of prediction errors for the two machine-learning-based models. It is shown that most errors concentrate in the range of  $-5$  dB to  $5$  dB and

FIGURE 4: Path loss values with different Tx altitudes ( $h_t = 10$  m).

Random Forest shows a higher prediction accuracy than KNN.

**5.3. Computational Efficiency.** Another aspect to be evaluated is the computational efficiency. The path loss values should be generated in a short time so that the spatial distribution of electromagnetic fields can be quickly updated when the propagation environment changes. The generation durations of our machine-learning-based models are recorded. The computer we used to run the programs has an AMD A8-4500M processor and 4 GB of memory. The required times of Random Forest and KNN predictors are 8.71 s and 5.95 s, respectively. In contrast, running the ray-tracing software would take more than 10 minutes to generate all the samples in the test set. This comparison result is preliminary but it still reflects that the machine-learning-based model can provide higher computational efficiency to the network planning than the deterministic approaches.

**5.4. Analysis of Feature Importance.** As mentioned, there are many parameters related to the path loss in the AA scenario and they serve as the input features in our machine-learning-based models. For example, the ray-tracing-based path loss values in the third route at different Tx altitudes are shown in Figure 4. The altitude of the Rx UAV is 10 m and three different Tx altitudes are taken into account, including 60 m, 70 m, and 80 m. It is shown that in the selected low-altitude UAV AA scenario, the path loss values at different Tx altitudes are very close. Besides, when the altitude of the Tx UAV is fixed at 70 m. The path loss values at different Rx altitudes are illustrated in Figure 5. According to Figures 4 and 5, the path visibility is vital for the path loss in the considered AA scenario. The propagation distances corresponding to sample indexes are also shown in these two figures.

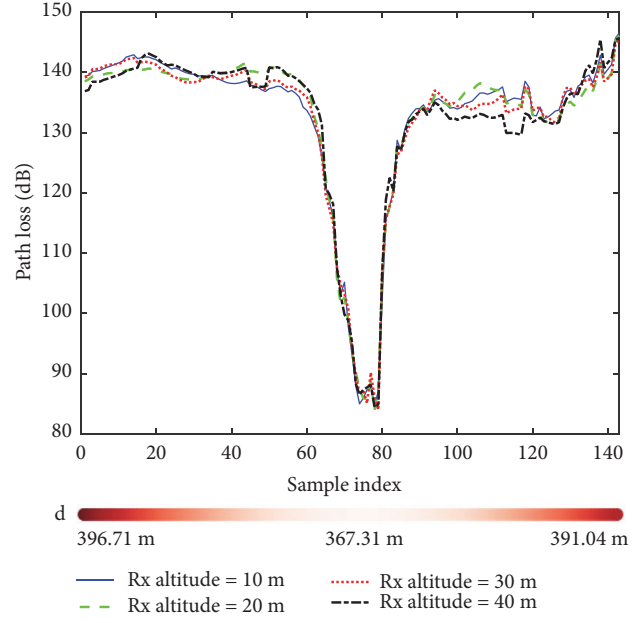
FIGURE 5: Path loss values with different Rx altitudes ( $h_t = 70$  m).

TABLE 3: Normalized importance of different features.

Feature	Importance
Path visibility	0.7438
Propagation distance	0.1137
Elevation angle	0.1045
Tx altitude	0.0317
Rx altitude	0.0064

The following task is to investigate the significance of different input parameters. Fortunately, Random Forest can give a natural ranking of the features in the model. In this study, the mean decrease impurity method [23] is employed to analyze the importance of features. As introduced above, Random Forest is composed of multiple decision trees. Each node in the decision tree is a condition about a feature, in order to divide the data into two sets according to different response variables. For regression problems, variance or least-squares fitting is often used as impurity. During the training process, it can be calculated how much impurity of the tree is reduced by a feature. For Random Forest, it is possible to calculate the average reduced impurity of each feature and to use it as the importance of the feature.

Table 3 shows the normalized contribution of each parameter used in the model based on Random Forest. The path visibility has the greatest impact, followed by the propagation distance, elevation angle, Tx altitude, and Rx altitude. Similar to results shown in Figures 4 and 5, the flight altitudes of UAVs have small influences on the path loss. A possible reason is that the selected scenario is a low-altitude UAV AA scenario and the heights of buildings are close to the flight altitudes of UAVs.



## 6. Conclusions

In this paper, we have proposed a modeling mechanism for AA path loss based on machine learning. A ray-tracing software has been utilized to generate the data for an urban AA scenario, which was subsequently divided into a training set and a test set to be used by the models. The models have been learned by two machine learning algorithms, Random Forest and KNN. The test data have been used to evaluate the accuracy performance of these machine-learning-based models and two empirical models, SUI model and COST231-W-I model. It has been demonstrated that machine learning provides a flexible modeling approach based on the training data for such complex environment and Random Forest has the best prediction performance. In addition, we have analyzed the importance of five input features for the path loss in the AA scenario. Results have confirmed that the path visibility is the dominant factor. Propagation distance and elevation angle have also shown great influences.

Since the UAV AA communication is a newly emerging scenario, the channel modeling and path loss prediction in such a scenario are still in a very preliminary stage. Future work should incorporate introduction of more machine-learning-based models like ANN and SVR. Different scenarios should also be taken into account to verify the generalization property of these models. Last but not least, measurement campaigns should be carried out in the AA scenario. More measured data are expected to further improve the performance and feasibility of the machine-learning-based path loss predictors.

## Data Availability

The data that support the findings of this study are available on reasonable request from the first author, Yan Zhang (zhangy@bit.edu.cn), or the corresponding author, Zunwen He (hezunwen@bit.edu.cn). The data are not publicly available since they are generated based on the mentioned parameters by commercial software.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by National Nature Science Foundation of China under Grant no. 61201192 and the National High Technology Research and Development Program of China (863 Program) under Grant no. 2015AA01A706. It was also supported by Ericsson Company.

## References

- [1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.
- [2] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeter-wave communication: potentials and approaches," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 66–73, 2016.
- [3] D. W. Matolak and R. Sun, "Air-ground channels for UAS: Summary of measurements and models for L- and C-bands," in *Proceedings of the 2016 Integrated Communications Navigation and Surveillance (ICNS)*, pp. 8B2-1–8B2-11, Herndon, VA, USA, April 2016.
- [4] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Proceedings of the 2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 2898–2904, Austin, TX, USA, December 2014.
- [5] R. Amorim, H. Nguyen, P. Mogensen, I. Z. Kovács, J. Wigard, and T. B. Sørensen, "Radio channel modeling for UAV communication over cellular networks," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 514–517, 2017.
- [6] K. Wang, R. Zhang, L. Wu et al., "Path loss measurement and modeling for low-altitude UAV access channels," in *Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, Toronto, Canada, September 2017.
- [7] X. Cai, A. Gonzalez-Plaza, D. Alonso et al., "Low altitude UAV propagation channel modelling," in *Proceedings of the 2017 11th European Conference on Antennas and Propagation (EuCAP)*, pp. 1443–1447, Paris, France, March 2017.
- [8] W. Khawaja, I. Guvenc, and D. Matolak, "UWB channel sounding and modeling for UAV air-to-ground propagation channels," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, pp. 1–7, Washington, DC, USA, December 2016.
- [9] Z. Yang, L. Zhou, G. Zhao, and S. Zhou, "Channel model in the urban environment for unmanned aerial vehicle communications," *Accepted by EuCAP'18*, 2018.
- [10] N. Goddemeier and C. Wietfeld, "Investigation of air-to-air channel characteristics and a UAV specific extension to the rice model," in *Proceedings of the IEEE Globecom Workshops, GC Wkshps 2015*, pp. 1–5, San Diego, Cal, USA, December 2015.
- [11] L. Zhou, Z. Yang, G. Zhao, L. Xiao, and S. Zhou, "Modeling air-to-air path loss for dense urban environments," in *submitted to IEEE Globecom'18*.
- [12] M. Ayadi, A. Ben Zineb, and S. Tabbane, "A UHF path loss model using learning machine for heterogeneous networks," *Institute of Electrical and Electronics Engineers. Transactions on Antennas and Propagation*, vol. 65, no. 7, pp. 3675–3683, 2017.
- [13] C. A. Oroza, Z. Zhang, T. Watteyne, and S. D. Glaser, "A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 576–584, 2017.
- [14] I. Popescu, D. Nikitopoulos, P. Constantinou, and I. Nafornita, "ANN prediction models for outdoor environment," in *Proceedings of the 2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, Helsinki, Finland.
- [15] I. Popescu, I. Nafornita, and P. Constantinou, "Comparison of neural network models for path loss prediction," in *Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2005. (WiMob '2005)*, pp. 44–49, Montreal, Canada.
- [16] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2735–2747, 2010.
- [17] D. Wu, G. Zhu, and B. Ai, "Application of artificial neural networks for path loss prediction in railway environments,"

- in *Proceedings of the 5th International ICST Conference on Communications and Networking in China*, pp. 1–5, Beijing, China, August 2010.
- [18] K. Lin, K. Hung, J. Lin, C. Wang, and P. Pai, “Applying least squares support vector regression with genetic algorithms for radio-wave path-loss prediction in suburban environment,” in *Advances in Neural Network Research and Applications*, vol. 67 of *Lecture Notes in Electrical Engineering*, pp. 861–868, Springer, Berlin, Germany, 2010.
- [19] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown, “Comparison of empirical propagation path loss models for fixed wireless access systems,” in *Proceedings of the 2005 IEEE 61st Vehicular Technology Conference VTC’05-Spring*, pp. 73–77, Stockholm, Sweden, May-June 2005.
- [20] V. Erceg, “Channel models for fixed wireless applications, tech. rep,” in *IEEE 802, 16 Broadband Wireless Access Working Group*, 2001.
- [21] P. Mededovic, M. Veletic, and Z. Blagojevic, “Wireless insite software verification via analysis and comparison of simulation and measurement results,” in *Proceedings of the 35th International Convention MIPRO*, pp. 776–781, Opatija, Croatia.
- [22] W. Hou, D. Shi, Y. Gao, and C. Yao, “A new method for radio wave propagation prediction based on finite integral method and machine learning,” in *Proceedings of the 2017 IEEE 5th International Symposium on Electromagnetic Compatibility (EMC-Beijing)*, pp. 1–4, Beijing, China, October 2017.
- [23] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [24] J. Isabona and V. M. Srivastava, “Hybrid neural network approach for predicting signal propagation loss in urban micro-cells,” in *Proceedings of the 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 1–5, Agra, India, December 2016.

## Research Article

# A Full Duplex D2D Clustering Resource Allocation Scheme Based on a $K$ -Means Algorithm

Xu Huang <sup>1,2</sup>, Mengjia Zeng,<sup>1</sup> Jing Fan,<sup>2</sup> Xiangxiang Fan,<sup>1</sup> and Xuefeng Tang<sup>1</sup>

<sup>1</sup>School of Information Engineering, Huzhou University, Huzhou, Zhejiang 313000, China

<sup>2</sup>College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310058, China

Correspondence should be addressed to Xu Huang; hx@zjhu.edu.cn

Received 5 March 2018; Accepted 22 April 2018; Published 23 May 2018

Academic Editor: Liu Liu

Copyright © 2018 Xu Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although the Device-to-Device (D2D) technology in cellular networks can improve the performance of cellular systems, it creates a large amount of interference in traditional communications. In this paper, the problem of resource allocation and control in a single-cell scene is studied. First, the concept of a restricted D2D communication area and a restricted D2D user-reusage area is put forward to reduce the complexity and interference intensity of resource allocation. Second, under the premise of satisfying the QoS (Quality of Service) demands of every system user, the resource allocation algorithm is improved, the optimal allocation of resources is carried out, and the algorithm's processes are given in detail. Our simulated experiments show that the proposed method greatly improves the spectrum efficiency and the system fairness.

## 1. Introduction

In recent years, the rise and rapid development of smart phones have profoundly changed the way we communicate and enjoy digital entertainment. The wide use of wireless applications, such as cloud computing, surfing the Internet, and downloading and watching digital multimedia, has created a large demand for high-speed and efficient wireless communication technology. 5G is the next-generation mobile communication system that is being developed for the expected demand of information and communication after 2020. It will have higher spectrum utilization and transmission rate, significantly improved transmission delay and QoS (Quality of Service) perception, and an increased number of access links and security [1].

Device-to-Device (D2D) technology [2, 3] is a hot topic in the field of mobile communication. Using D2D technology, adjacent terminals can transmit data within a close range through a direct link without a central node. In this case, the base station sends only some control information, which greatly reduces the load of the base station [2, 4, 5]. D2D users can effectively reuse the wireless resources authorized by the network to enhance the reusage rate of the wireless spectrum and expand the throughput and coverage of heterogeneous

cellular networks [6, 7]. In full duplex mode, the spectrum benefit is doubled as a result of allowing users to synchronize and send and receive signals simultaneously [8, 9]. This technology can greatly improve the wireless transmission rate, and it has great technical advantages and application prospects. As such, it is a promising option for future 5G communication.

However, under the same cell, channel resources are reused by D2D communication and cellular users simultaneously. So, while D2D communication technology brings convenience, it also results in frequency interference [10]. As the number of users increases, the interference between them also rises. The problem of network capacity optimization and power allocation also arises, which leads to an increase in the power consumption of the entire system. In heterogeneous networks, there are two resource allocation schemes: one is local resource allocation (fixed-cell user resource and adaptive resource allocation for a D2D pair) and the other is global resource allocation (the cell service user and the D2D pair are allocated resources jointly) [11]. Because D2D communications can compete and cooperate with each other to share resources, their individual or group behaviors conform to the inherent nature of game theory, which allows them to be effectively modeled and analyzed [12].

The combination of D2D communication technology and a cognitive radio (CR) can effectively reduce interference [13]. CR technology, through interactions with the external environment in terms of multidimensional spectrum detection, as well as real-time and interactive environments, is able to perceive any interference and make subsequent judgments so that cognitive users can choose the most appropriate communication frequency to avoid interference to primary users under the condition of the spectrum with the primary users sharing [14]. How to manage the spectrum resources of the community as a whole, reasonably determine the communication power of each device, and minimize the interference between devices have become the main bottleneck for D2D communication to enter the practical stage. It is worthwhile to try to achieve power stability [15, 16] or learn from software reliability prediction through context sensitive rate Boolean control network [17].

Considering the strengths and weaknesses of current D2D communication research, we have studied resource allocation and power control in the single-cell scenario based on the work of [18]. First, the concept of a restricted D2D communication area and a restricted D2D usage area is put forward to reduce the complexity and interference intensity of resource allocation. Second, under the premise of satisfying the QoS of all users in the system, the resource allocation algorithm is improved, the optimal allocation of resources is carried out, and a detailed description of the algorithm's processes is given. Our simulated experiments show that the proposed method greatly improves the spectrum efficiency and system fairness.

## 2. Problem Description

For a single-cell model, it is assumed that the cell includes  $M$  cellular users and  $N$  D2D user pairs, and the number of subchannels in this cell is  $L$ . The set of cellular users is  $C = \{C_1, C_2, \dots, C_M\}$ , while the set of D2D user pairs is  $D = \{D_1, D_2, \dots, D_N\}$ . D2D user pairs can reuse the cellular users' uplink resources. Compared with a user device, the base station has stronger anti-interference ability and processing power, and the amount of data in the cellular network is asymmetric, where the uplink resources are not fully utilized [19, 20]. Therefore, the D2D communication discussed in this paper chooses to reuse the uplink resources of the cellular system.

Assuming that the location coordinates of each user device are known to the base station, the channel gain between any users and between the users and the base stations can be calculated. In a cellular system model, BS is the base station, and the  $i$ th D2D user pair  $D2D_i$  reuses the uplink resource of the  $j$ th cellular user. See Figure 1.

As shown in Figure 1,  $CU_j$  send signals to the BS, where the channel gain is  $g_{j,B}$ . The transmitter  $D2D_{i,tx}$  of the  $i$ th D2D pairs  $D2D_i$  sends signals to the receiver  $D2D_{i,rx}$ , where the channel gain is  $g_i$ . When  $CU_j$  transmits a signal, there is interference to  $D2D_{i,rx}$ , where the channel gain is  $h_{i,j}$ . Additionally, there are interferences to the base station when  $D2D_{i,tx}$  transmits a signal, where the channel gain is  $h_{i,B}$  and the Gaussian white noise is  $N_0$ .

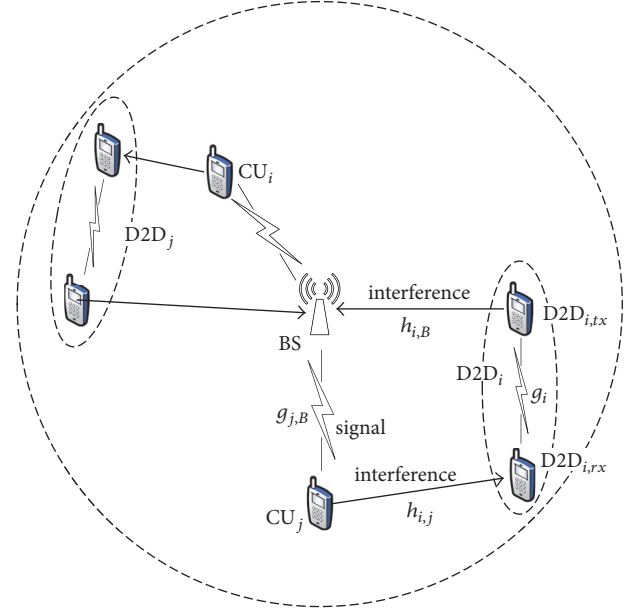


FIGURE 1: System model.

Regarding the road-loss model, the slow fading caused by the multipath effect and the fast fading caused by the shadow effect should also be considered. As such, the channel gains of the D2D pairs' transmitter  $D2D_{i,tx}$  to the base station and  $CU_j$  to  $D2D_{i,rx}$  can be expressed as

$$h_{i,B} = K \cdot \delta_{i,B} \cdot \xi_{i,B} \cdot d_{i,B}^{-\alpha}, \quad (1)$$

$$h_{i,j} = K \cdot \delta_{i,j} \cdot \xi_{i,j} \cdot d_{i,j}^{-\alpha}. \quad (2)$$

Among them,  $K$  is the road-loss constant and  $\alpha$  is the road loss index, which are both determined by the cellular system environment.  $d_{i,B}$  is the distance from  $D2D_{i,tx}$  to the base station, and  $d_{i,j}$  is the distance from  $CU_j$  to  $D2D_{i,rx}$ .  $\delta_{i,B}$  is the fast fading gain from  $D2D_{i,tx}$  to the base station, which has an exponential distribution,  $\delta_{i,j}$  is the fast fading gain from  $CU_j$  to  $D2D_{i,rx}$ , which also has an exponential distribution,  $\xi_{i,B}$  is the slow fading gain from  $D2D_{i,tx}$  to the base station, which has a logarithmic distribution, and  $\xi_{i,j}$  is the slow fading gain from  $CU_j$  to  $D2D_{i,rx}$ , which also has a logarithmic distribution.

The foregoing description shows that there are three kinds of interference in the considered scenario: (1) interference from the D2D transmitter to the cellular system; (2) interference from the cellular users to the D2D receiver; and (3) interference between D2D pairs sharing the same spectrum resources. The problem to be solved in this paper is to establish a cellular resource reuse state system model under the consideration of all three kinds of interference and use it to determine the best resource allocation plan.

## 3. Resource Allocation Scheme Satisfying System QoS

3.1. D2D User Pairs Delimit Communication Restricted Areas. As a D2D user reuses the upstream resource of a cellular user,

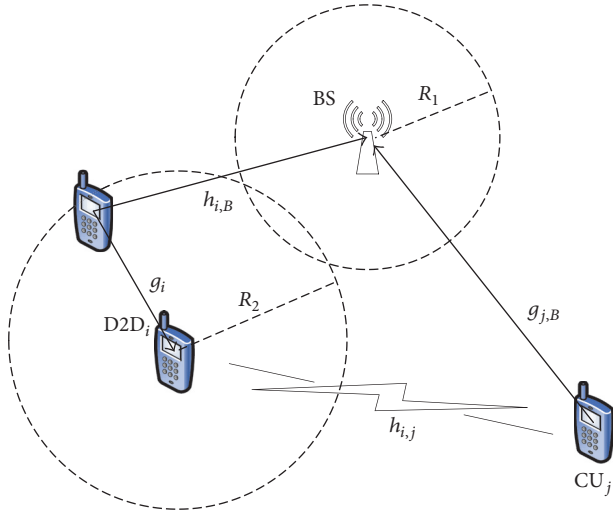


FIGURE 2: Cellular users in relation to the limited D2D communication area and the reuse area.

the closer the D2D user is to the base station, the greater the amount of interference is imparted to the BS receiving the CU (cellular user) signals. Therefore, to reduce the interference, we propose an improvement to the total throughput of the system while also reducing the computation complexity; we suggest restricting the size of the D2D communication area. In this area, D2D is considered to contribute a large amount of interference to the base station. Out of this area, any interference from D2D to the base station can be ignored. Assuming that  $P_{\max}^d$  is the maximum transmission power of all D2D users,  $I_{d,B}$  indicates the interference threshold value of all D2D users to the base station, which must satisfy

$$I_{d,B} \geq P_{\max}^d \cdot h_{i,B}. \quad (3)$$

The interference of all D2D users to the base station must be less than this threshold value; otherwise, it will cause more serious interference to the base station, so that cellular users cannot communicate effectively. Among them,  $I_{d,B}$  is obtained via long-term observations of a cellular cell. By combining (1) and (3), we can solve for the radius  $R_1$ , which is the radius of D2D communication limited area:

$$d_{i,B} \geq \left[ \frac{P_{\max}^d \cdot K \cdot \delta_{i,B} \cdot \xi_{i,B}}{I_{d,B}} \right]^{\alpha^{-1}} = R_1. \quad (4)$$

Therefore, the limited D2D communication area is a circular area of radius  $R_1$  with the BS at the center (see Figure 2). When the base station receives a request to establish a D2D link, the base station first determines whether the D2D transmitter is in the D2D communication area, and if it is, then this prohibits the establishment of D2D link.

**3.2. Cellular Users Are Restricted to the Reusage Area.** We can see from Figure 1 that when D2D users and cellular users use the same uplink resources, there will be interference from  $CU_j$  to  $D2D_{i,rx}$ , where the closer the distance between them

is, the greater the interference will become. Therefore, we propose to delineate a restricted reuse area for the cellular users to reduce this kind of interference. Assuming that  $P_{\max}^c$  represents the maximum transmit power of the cellular users,  $I_{c,d}$  represents the threshold value of the cellular users' interference to the D2D users. Hence, the total imparted interference from all cellular users to the D2D users must be less than this threshold value; otherwise the D2D users cannot communicate normally. That is to say, the following formula should be satisfied:

$$I_{c,d} \geq P_{\max}^c \cdot h_{i,j}. \quad (5)$$

Among them,  $I_{c,d}$  is determined by long-term observations of cellular cells. If we combine (2) and (5), we can calculate the restricted area radius  $R_2$  of the cellular users:

$$d_{i,j} \geq \left[ \frac{P_{\max}^c \cdot K \cdot \delta_{i,j} \cdot \xi_{i,j}}{I_{c,d}} \right]^{\alpha^{-1}} = R_2. \quad (6)$$

So, the limited, reuse area for a cellular user is a circle with radius  $R_2$ , where  $D2D_{i,rx}$  is located at the center (Figure 2). Any cellular users in this area are not selected as potential reuse objects by any D2D pairs.

**3.3. System Model Establishment.** To satisfy all users QoS requests, we must ensure that the D2D users and cellular users reach their respective minimum SINR (signal-to-interference-plus-noise ratio). For device  $K$  in the cellular  $m$ , the SINR of the device when communicating is

$$\text{SINR}(m, k) = \frac{P_R(m, k)}{I(m, k) + P_n}, \quad \forall m, 1 \leq m \leq M. \quad (7)$$

Among them,  $P_R(m, k) = P_T(m, k) \cdot G_T(m, k)$  indicates the valid signal strength that device  $K$  receives from cellular  $m$ ,  $P_T(m, k)$  is the transmission power of each base station RB (resource block),  $G_T(m, k)$  is the channel gain between the BS and the device, and  $I(m, k)$  is the summed interference from all other devices.

The goal is to maximize the system's throughput:

$$\max \left( \sum_{i=1}^N \log_2(1 + \gamma_{CU_j}) + \sum_{j=1}^K \log_2(1 + \gamma_{DU_i}) \right) \quad (8)$$

$$\text{s.t. } \gamma_{CU_j} = \frac{P_j^c h_{c_j,B}}{\sum_{i=1}^k P_i^d h_{d_i,c_j} x_{i,j} + N_0} \geq \text{SINR}_{\min}^c, \quad (8a)$$

$$\forall j \in C,$$

$$\begin{aligned} \gamma_{DU_i} &= \frac{P_i^d h_{d_i,d_i}}{\sum_{j=1, j \neq i}^k P_j^d h_{d_l,d_i} x_{l,i} + \sum_{j=1}^N P_j^c h_{c_j,d_i} x_{i,j} + N_0} \quad (8b) \end{aligned}$$

$$\geq \text{SINR}_{\min}^d, \quad \forall i \in D,$$

$$0 \leq P_j^c \leq P_{\max}^c, \quad \forall i \in C, \quad (8c)$$

$$0 \leq P_i^d \leq P_{\max}^d, \quad \forall j \in D. \quad (8d)$$

Among them,  $\gamma_{CU_j}$  represents the actual SINR of the cellular users, and  $\gamma_{DU_i}$  represents the actual SINR of the D2D users.  $P_j^c$  and  $P_i^d$ , respectively, represent the actual transmission power of the cellular users  $j$  and the D2D users  $i$ , while  $h_{cj,B}$  indicates the channel gain between the cellular users  $j$  and the base station. The model is  $128.1 + 37.6 \cdot \lg D$ , where  $D$  is the distance between the cellular users and the BS in units of Km.  $h_{di,di}$ ,  $h_{dl,dl}$ ,  $h_{cj,di}$ , and  $h_{di,cj}$ , respectively, represent the channel gain between users, where the model is  $10 \cdot \lg D4$  and  $D$  is the distance between two users in units of m.  $x_{i,j}$  is a binary value: if users  $i$  and  $j$  share the same resources,  $x_{i,j} = 1$ ; otherwise  $x_{i,j} = 0$ . Finally,  $P_{\max}^c$  and  $P_{\max}^d$ , respectively, represent the minimum SINR of a cellular user and a D2D pair when they communicate normally.

The foregoing objective function represents the maximum throughput of the complete system. The first two restricted conditions (see (8a) and (8b)) ensure that cellular users and D2D users meet their QoS needs. Only those D2D users that satisfy these two conditions can be connected to the network at the same time. The last two restricted conditions (see (8c) and (8d)) indicate that the transmission power of the cellular and D2D users should not exceed the maximum transmission power.

**3.4. D2D Pair Clustering to Solve the Problem of Interference between D2D Pairs and Each Other.** In cellular communication systems, there is interference between D2D pairs and each other. To effectively use the available system resources, it is necessary to reduce this kind of interference by grouping D2D pairs into clusters. It is considered that, in the same cluster, the interference between D2D pairs and each other can be ignored, and it is possible to reuse resources for the same CU. As their location coordinates are known, the distance between the D2D pairs can be calculated. The closer the distance between D2D pairs is, the greater the interference becomes; that is, the distance between D2D pairs is inversely proportional to the interference value. The reciprocal of the distance value is used to represent of the interference, where a matrix of interference values between the D2D pairs can be formed. Then, the  $K$ -means clustering method is used to divide the D2D clusters according to their interference value in the matrix. Assuming that the total number of D2D pairs is  $N$ , all D2D devices are grouped as  $X = \{x(i), i = 1, 2, \dots, N\}$ . After clustering,  $K_C$  clusters are formed and recorded as  $C = \{c(j), j = 1, 2, \dots, K_C\}$ . The specific algorithm process is as follows:

(1) Randomly select the initial cluster centers of  $K_C$  as  $U = \{u(j), j = 1, 2, \dots, K_C\}$ .

(2) The distance between samples is calculated, and their interference value in the matrix is formed from the reciprocal of the distance value, and the sample is added to the cluster with the smallest interference value.

(3) Calculate the summed square distance between points in cluster  $c(j)$  and the cluster center  $u(j)$ , as well as the total summed square distance:

$$J[c(j)] = \sum_{x(i) \in c(j)} \|x(i) - u(j)\|^2,$$

$$J(C) = \sum_{j=1}^{K_C} J[c(j)] = \sum_{j=1}^{K_C} \sum_{i=1}^{N_{UE}} d_{ji} \|x(i) - u(j)\|^2,$$

$$d_{ji} = \begin{cases} 1, & x(i) \in c(j) \\ 0, & x(i) \notin c(j). \end{cases} \quad (9)$$

According to the least-square method and the Lagrange principle, the cluster center  $u(j)$  takes the average of each sample point within the corresponding cluster  $c(j)$ .

(4) According to these steps, we iterate and update the cluster until  $J(C)$  converges to a minimum value, whereby the iteration is complete.

After iteration, we output the centers of the  $K_C$  clusters  $\{u_1, u_2, \dots, u_{K_C}\}$  and decide which D2D devices are included in each cluster.

**3.5. Resource Allocation of the D2D User Cluster under a QoS Guarantee.** Through the aforementioned steps, all D2D pairs are divided into  $K$  clusters, each of which is called  $g_k$ . Because the interference arising from the D2D network is different for each TTI (Transmission Time Interval), the number and the size of the clusters are not fixed, which makes full use of the instantaneous channel state information of the D2D network. The main task of this section is to determine which subchannel is assigned to which cluster.

Let us consider the following model for resource sharing: the cellular users are allocated resources as a prior, and each user occupies a cellular RB. A single D2D cluster can reuse most of the cellular users' spectrum resources; otherwise, a cellular spectrum resource is only reused by a single D2D cluster. The channel allocation is represented by the matrix  $Y = [y_{kn}]$ . Element  $y_{kn} = 1$  indicates that channel  $n$  is assigned to the D2D cluster  $K$ ; otherwise, it is equal to 0. The goal of this article is to find an allocation method  $Y_{\text{opt}}$  to maximize the throughput of the whole system; that is,

$$Y_{\text{opt}} = \arg \max_{Y \in J} \sum_K \sum_n R_{kn} \cdot y_{kn}, \quad (10)$$

among which  $R_{kn}$  represents the data rate of the cluster  $K$  to the RB $_n$ . To ensure normal communication throughout the original cellular network, in the actual resource allocation stage, it is necessary to screen any D2D users not only meeting the D2D QoS request from each cluster and to also ensure the normal communication of all cellular users and grant them access to the network. Therefore, the D2D pairs in the same cluster  $g_k$  are different, which shared different resource blocks of the same cellular user. Cluster  $g_{kj\text{Share}}$  is used to represent a D2D set that can share RB $_j$  in cluster  $g_k$ .

For a particular RB, the screening process for the D2D users that can eventually access the network is as follows:

- (1) **Initialization:** consider the case of a particular cluster  $g_k$  sharing RB $_j$ , and create two new clusters of  $g_{kj}$  and  $g_{kj\text{Share}}$
- (2) **for**  $i = 1 : M$  ( $I, GK$ ) **do**

calculate the distance  $d_{i,j}$  between  $CU_j$  and  $D2D_i$  on  $RB_j$

calculate the distance  $d_{i,B}$  between  $D2D_i$  and the base station BS

calculating the  $SINR_{di}$  of  $D2D_i$

**if**  $D2D_i$  satisfies  $SINR_{di} \geq SINR_{min}^d$  and  $d_{i,B} \geq R_1$  and  $d_{i,j} \geq R_2$

    put  $D2D_i$  into cluster  $g_{kj}$

**end if**

**end for**

(3) **for**  $i = 1 : M$  ( $i \in g_{kj}$ ) **do**

    calculate the interference value of  $D2D_i$  to the cellular user in  $RB_j$

**end for**

Ascendingly sort the interference values of all  $D2D$  users in  $g_{kj}$

(4) **for**  $D2D$  users after sorting **do**

    calculate the SINR of the cellular user after added the  $D2D$  pairs

**if** the SINR of the cellular user satisfies  $SINR \geq SINR_{min}^c$

**break**

**end if**

**else**

        put this  $D2D$  user into cluster  $g_{kjShare}$

**end else**

**end for**

Step (2) filters the  $D2D$  pairs in cluster  $g_k$  by considering the interference between the  $D2D$  pairs and the base station and from the cellular users to the  $D2D$  users. If a QoS request cannot be fulfilled, we then obtain cluster  $g_{kj}$ . Step (3) calculates the interference value of each  $D2D$  pair in cluster  $g_{kj}$  and, according to the interference value sorts the corresponding  $D2D$  users. Step (4) starts with the  $D2D$  pair that has the minimum value of interference from cellular users in cluster  $g_{kj}$  and then calculates the SINR of the cellular user when this  $D2D$  pair is shared  $RB_j$ . If the SINR is less than the minimum SINR of the cellular user, the system does not allow the  $D2D$  pair to share  $RB_j$ ; otherwise, it puts the  $D2D$  pair into cluster  $g_{kjShare}$ . The system then considers the  $D2D$  user with the second smallest interference value and calculates these cellular users' SINR. The entire process is repeated until every  $D2D$  pair in  $g_{kj}$  has been calculated. Finally, all  $D2D$  users in cluster  $g_{kj}$  which can share  $RB_j$  are obtained and are represented by the cluster  $g_{kjShare}$ .

In this paper, we use  $throughput_{R_j}$  to characterize the throughput achieved by each cluster to each RB. Finally, we construct a two-dimensional  $throughput_{R_j}$  matrix based on the cluster number and the RB number and allocate resources for each cluster according to the constructed matrix.

The  $throughput_{R_j}$  value of the cluster  $g_k$  sharing  $RB_j$  is defined as

$$throughput_{R_j} = \left[ \log_2(1 + \gamma_{CU}) + \sum_i \log_2(1 + \gamma_{DU_i}) \right], \quad (11)$$

$$\gamma_{DU_i} = \frac{P_i^d h_{di,di}}{\sum_{j,l \neq i} P_l^d h_{dl,dj} + P^c h_{c,di} + N_0}, \quad \forall i, l \in g_{kjShare}, \quad (12)$$

$$\gamma_{CU} = \frac{P^c h_{c,B}}{\sum_i P_{max}^d h_{di,c} + N_0}, \quad \forall i \in g_{kjShare}. \quad (13)$$

Among them,  $\gamma_{DU_i}$  is the  $D2D$  users' SINR in cluster  $g_{kjShare}$  in  $RB_j$  and  $\gamma_{CU}$  indicates the cellular users' SINR in  $RB_j$  when there is interference from the  $D2D$  users in cluster  $g_{kjShare}$ .

The specific resource allocation process is as follows:

(1) **Initialize:** according to (11), calculate the  $throughput_{R_j}$  of each cluster in each RB, and then construct the  $throughput_{R_j}$  matrix. Assuming  $S$  is the set of RBs that have not yet been allocated,  $j$  shows a single RB,  $j \in S$ .

(2) **while** ( $S \neq \emptyset$ ):

In the  $throughput_{R_j}$  matrix, find the maximum  $throughput_{R_j}$  value, and its corresponding  $RB_j$  and cluster  $g_{kjShare}$ ;

Allocate  $RB_j$  to cluster  $g_{kjShare}$ ;

Set all  $RB_j$  corresponding to  $throughput_{R_j}$  to  $-1,000$ .

$S = S - \{j\}$ ;

**End while**

(3) **Power distribution:** the final transmission power of each  $D2D$  pair  $i$  is

$$P_i^d = \frac{P_{max}^d}{(\text{the number of RB that been allocated to } D2D_i)}, \quad (14)$$

where the transmission power of cellular users is  $P^c = P_{max}^c$ .

## 4. Simulation Results and Analysis

**4.1. Simulation Parameters.** A single-cell scenario with a radius of 500 m is considered, where all CUs and DUs are randomly distributed uniformly in the cell. The  $D2D$  pairs and CU are static, and each receiver and each transmitter have a single antenna. In the simulation, it is assumed that the amount of self-interference cancellation of the full duplex  $D2D$  node is 110 dB [24], the antenna gain is 14.0 dBi [25], the uplink bandwidth is set to 1.4 MHz, and the resource block RB is six. All parameters are shown in Table 1. The

TABLE 1: Simulation parameters.

Parameter	Parameter value
Radius of cellular users (Km)	0.5
Average distance of D2D pair (m)	25
System bandwidth (MHz)	1.4
RB bandwidth (KHz)	240
Number of cellular users	6
Number of D2D pairs	0, 10, 20, 30, 40, 50, 60, 70, 80
The maximum transmit power of cellular user (dBm)	23
The maximum transmit power of D2D pair (dBm)	10
Noise spectrum density (dBm/Hz)	-174
Cellular link road-loss model (Km) [21-23]	$128.1 + 37.6 \lg D$
D2D link road-loss model (m) [21-23]	$127 + 30 \lg D$
Cellular user SINR threshold (dB)	[0, 25] uniform distribution
D2D pair SINR threshold (dB)	[0, 25] uniform distribution

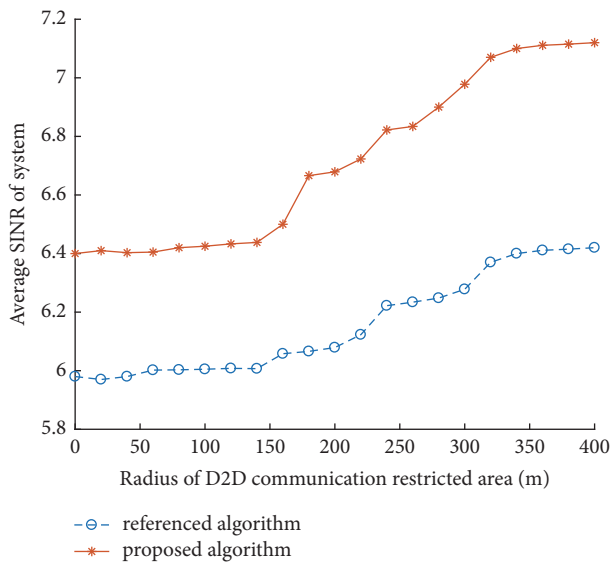


FIGURE 3: Influence of the radius of the D2D communication area versus the average SINR of the system.

simulation was performed using Matlab, where the goal of our proposed scheme is to determine the optimized total system throughput and average SINR, which are then compared with the full duplex random resource allocation scheme of [18, 26] to verify its validity.

#### 4.2. Verification Analysis

**4.2.1. D2D Communication Restricted Area Verification.** Figure 3 is a chart of the average SINR change of the system for increasing radii of the D2D communication area. From Figure 3, we can see that as the D2D communication area radius increases from 0 to 150 m, the average SINR increase is not very obvious. However, when the radius exceeds 150 m, the average SINR of the system increases with an increase of the radius of the D2D communication area. It can be seen that, to reduce the interference from the D2D users to base

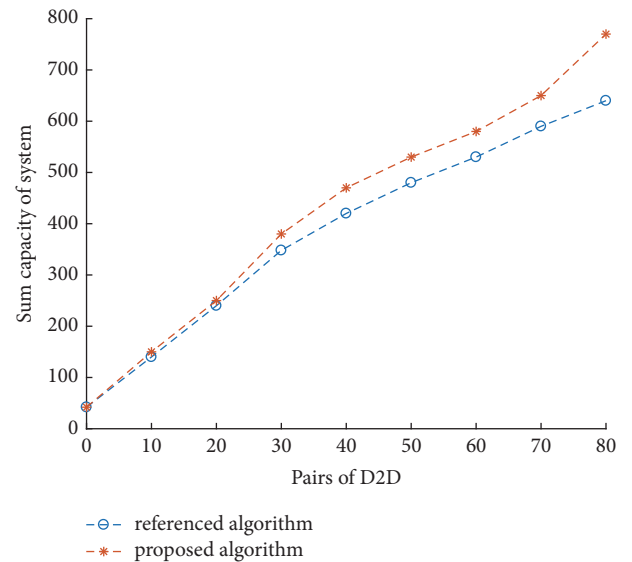


FIGURE 4: Influence of the number of D2D pairs on the total capacity of the system.

station, it is necessary for the cell to delimit a restricted D2D communication area. Under the experimental conditions set in this paper, the radius of the restricted D2D communication area is ~150 m. At the same time, it can be seen from Figure 4 that the proposed algorithm is obviously better than the resource block random allocation algorithm used by other authors [18, 26].

**4.2.2. Influence of the Number of D2D Pairs to the Total System Throughput Verification.** Figure 4 shows the total system throughput of the two resource allocation schemes as the number of D2D pairs increases. From this figure, we can see that when the number of D2D pairs is <20, the throughput of the proposed resource allocation scheme is similar to that of the random allocation scheme. As the number of D2D pairs increases, the total throughput of the two schemes gradually increases, but the throughput of our



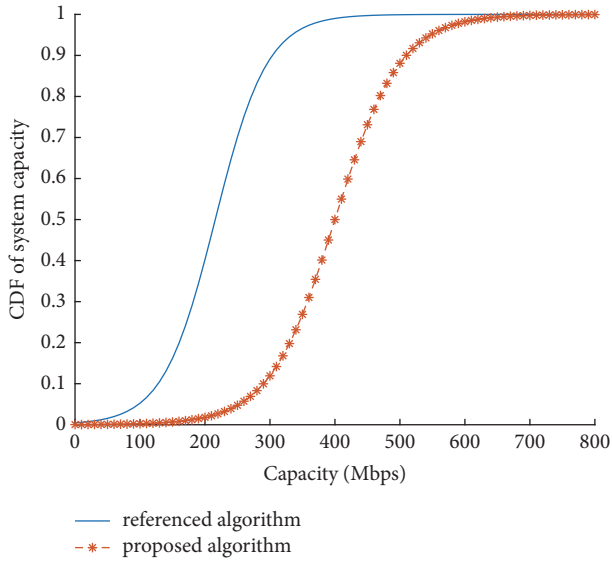


FIGURE 5: CDFs of the system capacity for the two considered schemes.

proposed scheme is significantly higher than that of the random resource allocation scheme.

**4.2.3. Comparison of the Cumulative Distribution of the Total System Throughput.** Figure 5 shows the cumulative distribution functions (CDFs) of the total throughput of the system for the two schemes. Since our method uses a  $K$ -means algorithm to cluster the D2D pairs, where interference can be ignored for each pair, and then allocates the optimal RB to the D2D cluster, we are able to fully utilize the system's available and reusable RBs, and the number of D2D links established in the cellular system is maximized. Therefore, compared with the reference algorithm, the total throughput of the system has been significantly improved.

## 5. Conclusions

We analyzed the status of D2D pairs that reuse the resources of cellular users in the same 5G channel in a cellular communication system to eliminate the interference of D2D pairs to the base station, CU users to D2D receivers, and D2D pairs with each other. We proposed several countermeasures and solutions, where a function model was constructed to maximize the system's efficiency. To eliminate the interference of D2D users to the base station, the concept of a limited D2D communication area is proposed. The interference of D2D users to the base station outside of the restricted area is negligible. The concept of a limited reuse area to combat the interference of CU users to D2D users is proposed. It is possible for the D2D users to reuse the CU resource outside of the restricted area. To address the interference between D2D pairs, we suggest using the interference intensity value to the cluster D2D pairs based on a  $K$ -means algorithm, where the larger the interference value is between D2D pairs, the smaller the probability of being placed into a cluster is. In the same cluster, the interference can be neglected between D2D

pairs; that is to say, they can reuse the same CU resources. Then, we described the resource allocation algorithm, where we allocated CU resource blocks to the clustered D2D pairs and allocated the best RBs to D2D pairs. As demonstrated with a simulation performed with Matlab, our method is able to eliminate interference and improve the system's overall performance.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (61202290, 61370173, and 61772198).

## References

- [1] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ö. Bulakci, "METIS research advances towards the 5G mobile and wireless system definition," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, article 53, 2015.
- [2] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hug, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, 2009.
- [3] J. Du, W. Zhu, J. Xu, Z. Li, and H. Wang, "A compressed HARQ feedback for device-to-device multicast communications," in *Proceedings of the IEEE Vehicular Technology Conference (VTC-Fall '12)*, pp. 1–5, IEEE, Quebec City, Canada, September 2012.
- [4] Z. Li, M. Moisio, M. Uusitalo et al., "Overview on initial METIS D2D Concept," in *Proceedings of the 1st International Conference on 5G for Ubiquitous Connectivity*, pp. 203–208, IEEE Press, Levi, Finland, November 2014.
- [5] R. Tang, J. Zhao, H. Qu, Z. Zhu, and Y. Zhang, "Joint mode selection and resource allocation for mobile relay-aided device-to-device communication," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 3, pp. 950–975, 2016.
- [6] S. Ali, N. Rajatheva, and M. Latva-Aho, "Full duplex device-to-device communication in cellular networks," in *Proceedings of the 2014 European Conference on Networks and Communications, EuCNC 2014*, pp. 1–5, Bologna, Italy, June 2014.
- [7] X. Yao, X. Xie, and Y. Tian, "A new interference alignment algorithm for device-to-device communication underlying multi-cell MIMO networks," *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, vol. 28, no. 5, pp. 641–647, 2016.
- [8] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: challenges, solutions, and future research directions," *Proceedings of the IEEE*, vol. 104, no. 7, pp. 1369–1409, 2016.

- [9] R. Tang, J. Zhao, H. Qu, and Z. Zhang, "Energy-efficient resource allocation for 5G full-duplex enabled device-to-device communication," in *Proceedings of the 1st IEEE Global Communications Conference on Full Duplex Wireless Communications International Workshop*, pp. 1-7, IEEE Press, Washington, DC, USA, December 2016.
- [10] G. Fodor and N. Reider, "A distributed power control scheme for cellular network assisted D2D communications," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1-6, Houston, Tex, USA, December 2011.
- [11] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Communications Magazine*, vol. 21, no. 3, pp. 136-144, 2014.
- [12] L. Yang, S. He, Y. Wang et al., "Key technologies for 5G wireless communication system," *Journal of Data Acquisition & Processing*, vol. 30, no. 3, pp. 469-485, 2015.
- [13] A. H. Sakr, H. Tabassum, E. Hossain, and D. I. Kim, "Cognitive spectrum access in device-to-device-enabled cellular networks," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 126-133, 2015.
- [14] T. E. Bogale and L. Vandendorpe, "Linearly combined signal energy based spectrum sensing algorithm for cognitive radio networks with noise variance uncertainty," in *Proceedings of the 2013 8th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM 2013*, pp. 80-86, IEEE Press, Washington, DC, USA, July 2013.
- [15] L. Tong, Y. Liu, J. Lou, J. Lu, and F. E. Alsaadi, "Static output feedback set stabilization for context-sensitive probabilistic Boolean control networks," *Applied Mathematics and Computation*, vol. 332, pp. 263-275, 2018.
- [16] Y. Liu, L. Y. Tong, J. G. Lou et al., "Sampled-data control for the synchronization of boolean control networks," *IEEE Transactions on Cybernetics*, pp. 1-7, 2018.
- [17] J. Lou, Y. Jiang, Q. Shen, Z. Shen, Z. Wang, and R. Wang, "Software reliability prediction via relevance vector regression," *Neurocomputing*, vol. 186, pp. 66-73, 2016.
- [18] N. Chen, H. Tian, and Z. Wang, "Resource allocation for intra-cluster D2D communications based on Kuhn-Munkres algorithm," in *Proceedings of the 80th IEEE Vehicular Technology Conference, VTC 2014-Fall*, pp. 1-5, September 2014.
- [19] Y. Liu, Y. Xu, D. Li, and W. Wang, "Device-to-Device communication in LTE-A cellular networks: Standardization, architecture, and challenge," in *Proceedings of the 2014 IEEE Vehicular Technology Conference (VTC 2014-Spring)*, pp. 1-5, IEEE, Seoul, South Korea, May 2014.
- [20] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40-48, 2014.
- [21] C. Lee, S. Oh, and A. Park, "Interference avoidance resource allocation for D2D communication based on graph-coloring," in *Proceedings of the 2014 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 895-896, Busan, South Korea, October 2014.
- [22] J. Han, Q. Cui, C. Yang, and X. Tao, "Bipartite matching approach to optimal resource allocation in device to device underlying cellular network," *IEEE Electronics Letters*, vol. 50, no. 3, pp. 212-214, 2014.
- [23] D. Zhu, J. Wang, A. L. Swindlehurst, and C. Zhao, "Downlink resource reuse for device-to-device communications underlying cellular networks," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 531-534, 2014.
- [24] G. Zhu, T. Liu, and J. Yang, "Optimal power control for 5G full-duplex enabled D2D communication," *Application Research of Computers*, vol. 34, no. 12, pp. 1-5, 2017.
- [25] R. Tang, J. Dang, X. Zhuang et al., "Throughput-capacity trade-off for device-to-device communication underlying cellular network via joint resource allocation," in *Proceedings of the International Conference on Communication Technology*, pp. 181-190, IEEE Press, 2015.
- [26] D. J. Son, C. H. Yu, and D. I. Kim, "Resource allocation based on clustering for D2D communications in underlying cellular networks," in *Proceedings of the 5th International Conference on Information and Communication Technology Convergence, ICTC 2014*, pp. 232-237, October 2014.

## Research Article

# MU-MIMO Downlink Capacity Analysis and Optimum Code Weight Vector Design for 5G Big Data Massive Antenna Millimeter Wave Communication

Adam Mohamed Ahmed Abdo,<sup>1</sup> Xiongwen Zhao ,<sup>1</sup> Rui Zhang,<sup>2</sup> Zhenyu Zhou ,<sup>1</sup> Jianhua Zhang ,<sup>3</sup> Yu Zhang,<sup>1</sup> and Imran Memon <sup>4</sup>

<sup>1</sup>School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

<sup>2</sup>National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation, Qingdao 266107, China

<sup>3</sup>School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>4</sup>College of Computer Science, Zhejiang University, Hangzhou, Zhejiang 310027, China

Correspondence should be addressed to Xiongwen Zhao; zhaowx@ncepu.edu.cn

Received 8 January 2018; Revised 1 April 2018; Accepted 18 April 2018; Published 22 May 2018

Academic Editor: Javier Prieto

Copyright © 2018 Adam Mohamed Ahmed Abdo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiuser multiple input multiple output (MU-MIMO) wireless communication system provides substantial downlink throughput in millimeter wave (mmWave) communication by allowing multiple users to communicate at the same frequency and time slots. However, the design of the optimum beam-vector for each user to minimise interference from other users is challenging. In this paper, based on the concept of signal-to-leakage plus noise ratio (SLNR), we analyze the ergodic sum-rate capacity using statistical Eigen-mode (SE) and zero-forcing (ZF) models with Ricean fading channel. In the analysis, the orthogonality of channel vectors between users is assumed to guarantee interference cancellation from other cochannel users. The impact of the number of antenna elements on the achievable sum-rate capacity obtained by dirty paper coding (DPC) method considered as a nonlinear scheme for approximating average system capacity is studied. A power iterative precoding scheme that iteratively finds the most dominant eigenvector (optimum weight vector) for minimising cochannel interference (CCI), that is, maximising the SLNR for all users simultaneously, is designed resulting in enhancement of average system capacity. The average system capacities achieved by the proposed power iterative technique in this study compared with the singular value decomposition (SVD) method are in the ranges of 5–11 bps/Hz and 1–6 bps/Hz, respectively. Therefore, the proposed power iterative method achieves higher performance than the SVD regarding achievable sum-rate capacity.

## 1. Introduction

Millimeter wave (mmWave) communication which explores shorter propagation distance in frequency band of 20–40 GHz is a key enabler for the fifth generation (5G) mobile communication systems [1]. The mmWave communication also provides significant benefits to a variety of applications such as vehicular communication, wire-able networks, and autonomous robots [2]. In downlink transmission for a MU-MIMO system, a base station (BS) serves

multiple users simultaneously in the same frequency and time slot. Thus the throughput can be enhanced by spatial multiplexing. However, the cochannel interference (CCI) becomes a dominant factor in capacity due to nonorthogonal signalling [3]. Hence, minimising CCI as much as possible at the end users is of importance. CCI can be suppressed by using linear precoders and decoders at transmitter and receiver sides [4]. Additionally, the channel state information (CSI) as well as partial information on the transmitter side can be used to improve the system's performance [5]. For all

users served by the BS, the CSI should be known in advance at the BS to support the CCI minimisation which is not available at user's end. Explicitly, the overhead on the system is reduced in this approach because the channel information feedback is not required [6, 7]. To perfectly cancel CCI at each end user, the restriction on the system configuration is necessary. For instance, the number of antenna elements at the BS should be larger than the total number of antennas at the end users [7]. This assumption is usually valid for single antenna per user which is generally adopted in the MU-MIMO system.

There are several studies that have been undertaken in MU-MIMO downlink systems on how to minimise the CCI problem related to CSI and analyze the achievable sum-rate per user as well as the average system capacity. In [8], an in-depth capacity analysis for nonorthogonal multiple access (NOMA) mmWave massive MIMO systems was provided. A simplified mmWave channel model was also explored by extending the uniform random single-path (UR-SP) model with the angle-of-arrival (AOA). Furthermore, the capacity analysis was divided into high and low signal to noise ratio (SNR) regimes, where the dominant factors for the signal to interference plus noise ratio (SINR) were determined as interference and noise. A multicast beamforming approach was proposed in [9], where users located within the proximity of the BS can receive two different data streams simultaneously, while those away can receive only one data stream. In [7], a model was designed to transmit beamforming vectors to maximise the signal to leakage ratio (SLR), minimising transmission power which may cause interference to other users. The model is not restricted to a certain number of transmission antenna elements and can be extended to more general scenarios. An asymptotic deterministic SLNR optimisation approach for regularized zero-forcing (RZF) considering perfect CSI and antenna correlation was proposed [10]. It was found that when the users are homogeneously distributed and the number of antenna elements is large enough ( $M \rightarrow \infty$ ), the SLNR is asymptotically equal to SINR. In [11], an efficient statistical Eigen-mode space division multiple access (SE-SDMA) scheme for downlink ergodic sum-rate analysis based on SLNR was proposed. To maximise the approximate ergodic sum-rate capacity, authors in [12] designed an optimal beamforming vector for each user based on a three-dimensional beamforming algorithm. The authors in [3] focused on imperfect CSI scenario and developed a robust SLNR approach for compensating performance degradation caused by random CSI errors. In [13], a precoding scheme for heterogeneous networks (Het-Nets) was proposed based on SLNR under imperfect CSI scenario. In this case, regularization parameter was used to "weight" the precoding information for other cells (BSs).

In this paper based on the SLNR concept and Ricean fading channel model with perfect CSI in MU-MIMO systems, the statistical Eigen-mode (SE) and zero-forcing (ZF) models are derived. In addition, the average achievable sum-rate by each user and the overall average system capacity with different number of users are analyzed. Besides, the expression of dirty paper coding (DPC) method as a function of number of users ( $U$ ) and number of antenna elements ( $M$ ) is addressed. Moreover, the impact of  $U$  on

the average achievable sum-rate is demonstrated. The Ricean fading channel with the line-of-sight (LOS) and scattering components as channel model are adopted. We find that the ergodic capacity of DPC is approximately free from channel matrix which mainly depends on the number of BS antenna elements and average transmission power.

Furthermore, by exploring the leakage signal criteria, we propose a new design solution for precoding based on the proposed power iteration technique. The technique finds the optimum weight vector which maximises the SLNR, that is, enhances the system capacity by obtaining the dominant eigenvector for minimising the CCI. Finally, we compare the proposed method with the conventional solutions such as SVD. The numerical results demonstrate that significant throughput performance can be achieved with the proposed technique.

A comparative study of the proposed and SVD methods regarding computation complexity and storage is also carried out. The simplicity of the proposed algorithm is observed in approximating only one eigenvalue of a matrix in a sequence which is considerably more efficient as the number of iterations increase. In comparison, the SVD has to calculate all eigenvectors and pick the one with the most significant eigenvalue (the maximum eigenvalue), which requires larger computation capability and storage capacity.

The contributions of this paper are in twofold:

(1) The SE and ZF beam former models in Ricean fading channel are examined based on SLNR instead of the commonly used SINR and orthogonal condition of the beam weight vector in MU-MIMO systems. The linear achievable sum-rate capacity is investigated using these models to validate the proposed method. Moreover, in nonlinear achievable sum-rate capacity, the DPC technique regarding the impact of the number of users and antenna elements on the ergodic capacity is addressed. The results show that the number of antenna elements has a significant contribution to the ergodic capacity, and the DPC sum-rate capacity logarithmically increases with the number of users.

(2) A new optimum weight vector is developed based on the proposed power iteration method, which allows each user to maximise the SLNR and minimise the CCI from other users. The proposed method compared with the conventional SVD method shows that the proposed method can achieve better performance with relatively low cost.

This paper is organized as follows. In Section 2, the MU-MIMO downlink system is introduced. The beamforming and leakage signal approaches are presented in Section 3. In Section 4, the results and discussion of the study are presented. Finally, the conclusion is drawn in Section 5. The significant proofs of the algorithm are addressed in the appendix.

In this paper, the superscript  $(\bullet)^H$ ,  $(\bullet)^T$  and  $(\bullet)^*$  denote the conjugate-transpose (Hermitian), transpose, and conjugate, respectively.  $E\{\bullet\}$ ,  $\mathbf{I}_M$ , and  $CN$  are the expectation operator, square identity matrix of size  $M \times M$ , and Gaussian random complex numbers, respectively.

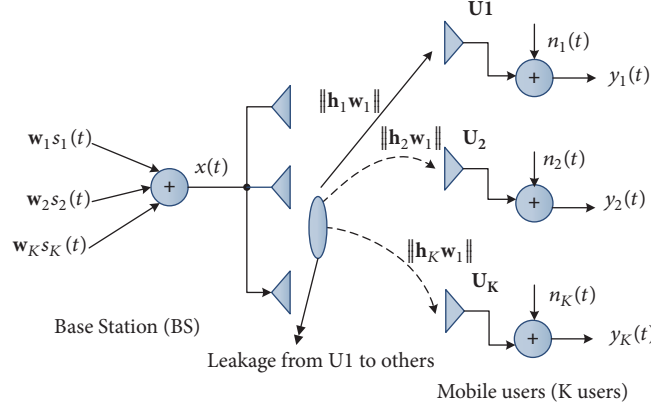


FIGURE 1: Schematic diagram for leakage signals in MU-MIMO system model from the first user to all other users.

## 2. System Model

Consider the MU-MIMO system model shown in Figure 1. The BS is equipped with  $M$  uniform linear transmission antennas that serve  $U$  users simultaneously. Assume that each user is equipped with a single antenna element [3]. Additionally, the number of the BS antenna elements is assumed to be larger than that of users ( $M \geq U$ ); and the power is equally allocated among all users [11, 12].

*2.1. Signal Model.* From the previous assumptions, the received signal  $y_k$  at  $k$ th user is given by

$$y_k = \sqrt{p_k} \mathbf{h}_k^H \mathbf{w}_k x_k + \sqrt{p_j} \sum_{j=1, j \neq k}^U \mathbf{h}_k^H \mathbf{w}_j s_j + n_k. \quad (1)$$

The expression in (1) can be rewritten in a more simplified form as

$$y_k = \sqrt{p_k} \sum_{j=1}^U \mathbf{h}_k^H \mathbf{w}_j s_j + n_k, \quad (2)$$

where  $\mathbf{h}_k \in \mathbb{C}^{1 \times M}$  is the flat Ricean downlink channel vector between the BS and the  $k$ th user. The weight vector  $\mathbf{w}_j \in \mathbb{C}^{M \times 1}$  is unit normalised beamforming vector of user  $j$  and satisfies  $\|\mathbf{w}_j\| = 1$ . Furthermore, the vector  $s_j$  is the transmitted data symbol of user  $j$  with  $E[|s_j|^2] = 1$ .  $n_k \rightarrow CN(0, \sigma_k^2)$  is unit normalised ( $\sigma_k^2 = 1$ ) complex additive white Gaussian noise (AWGN). Moreover,  $p_k$  is the transmission power of the  $k$ th user under the constraint  $\sum_{k=1}^U p_k \leq P$ . Since the total transmission power ( $P$ ) is divided among all the users equally, the average power can be written as  $p_k = P/U$  for each user [3, 11, 12]. Consequently, the signal magnitude of each user can be written as  $p |\mathbf{h}_k^H \mathbf{w}_k|^2$  while the interference from all other users is determined by the sum of interference from all other users. Hence, the interference can be denoted as  $\sum_{j=1, j \neq k}^U p |\mathbf{h}_k^H \mathbf{w}_j|^2$ . Thus the SINR of the  $k$ th user can be expressed as in (3) [12, 14]:

(Note: for simplicity we drop the “subscript  $k$ ” from  $p_k$  in the rest of the paper).

$$\text{SINR}_k = \frac{p |\mathbf{h}_k^H \mathbf{w}_k|^2}{1 + \sum_{j=1, j \neq k}^U p |\mathbf{h}_k^H \mathbf{w}_j|^2}. \quad (3)$$

The corresponding average achievable data rate for user  $k$  can be written as

$$R_k = E [\log_2 (1 + \text{SINR}_k)]. \quad (4)$$

By substituting (3) into (4), the average achievable sum rate per user is given as in (5) [15, 16].

$$R_k = E \left\{ \log_2 \left( 1 + \frac{p |\mathbf{h}_k^H \mathbf{w}_k|^2}{1 + \sum_{j=1, j \neq k}^U p |\mathbf{h}_k^H \mathbf{w}_j|^2} \right) \right\}. \quad (5)$$

Subsequently, the overall system ergodic sum-rate capacity can be written in bps/Hz as

$$R_{\text{system}} = \sum_{k=1}^U R_k. \quad (6)$$

*2.2. Channel Model.* In this subsection, we describe the channel model as in [11, 12]. In the most recent studies on mmWave communication, the Rayleigh fading channel model is employed which faces difficulty in capturing the fading variation in the presence of the LOS component. Moreover, mmWave MU massive antenna system can be used mainly in 5G hotspot scenarios in the LOS case. To overcome the capturing problem in Rayleigh fading channel, the Ricean fading channel is applied in this paper. Considering the Ricean fading channel model, the channel vector has two components: LOS and Rayleigh random distribution components. Thus, the channel vector can be expressed as [11]

$$\mathbf{h}_k = \sqrt{\frac{K_k}{K_k + 1}} \bar{\mathbf{h}}_k + \sqrt{\frac{1}{K_k + 1}} \mathbf{h}_{w,k}, \quad (7)$$

where  $K_k$  is the power ratio between the LOS component and scattering components, also called Ricean  $K$ -factor of the

$k$ th user. Moreover, the channel vector  $\mathbf{h}_{w,k}$  is the random component, and the entries are independent and identical distribution complex Gaussian random variables. Vector  $\bar{\mathbf{h}}_k$  ( $k = 1, 2, \dots, U$ ) is the deterministic component of the channel mean vector of the  $k$ th user. Based on the assumption that the BS is equipped with  $M$  uniform linear array antennas, the deterministic channel vector for each user can be written as [17, 18]

$$\bar{\mathbf{h}}_k = \left[ 1, e^{j2\pi\lambda^{-1}d \sin(\theta_k)} \dots e^{j(M-1)2\pi\lambda^{-1}d \sin(\theta_k)} \right], \quad (8)$$

where  $d$  is the space between two antenna elements,  $\lambda$  is the wavelength of the signal, and  $\theta_k$  is the angle of departure (AoD) for the  $k$ th user,  $\theta_k \in [-\pi, \pi]$ .

### 3. Beamforming and Leakage Signals

It is assumed that all users have a perfect instantaneous knowledge of their channel vectors while the BS knows all users channel vector including the channel mean vector. Additionally, assuming  $U$  users, it is difficult to maximise the average system capacity due to a couple of  $U$  calculations (capacity of each user) required. Therefore, it is hard to use SINR directly to obtain the optimum beamforming vector ( $\mathbf{w}_k$ ). To overcome this problem, the SLNR can be easily implemented to control the leakage signals from a specific user to other users as shown in Figure 1. The SLNR of the  $k$ th user can be written as in (9) by assuming unit normalised Gaussian noise [4, 16]:

$$\text{SLNR}_k = \frac{p |\mathbf{h}_k^H \mathbf{w}_k|^2}{1 + p \sum_{j=1, j \neq k}^U |\mathbf{h}_j^H \mathbf{w}_k|^2}. \quad (9)$$

$$\mathbf{R}_k = \begin{pmatrix} \frac{K_k M}{K_k + 1} + \frac{1}{K_k + 1} & 0 & \dots & 0 \\ 0 & \frac{1}{K_k + 1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{K_k + 1} \end{pmatrix} \begin{pmatrix} \frac{\bar{\mathbf{h}}_k^H}{\sqrt{M}} \\ \bar{\mathbf{U}}_k^H \end{pmatrix}, \quad (12)$$

where  $\bar{\mathbf{U}}_k^H$  is the orthogonal subspace of  $\bar{\mathbf{h}}_k^H / \sqrt{M}$ . Then we can obtain the optimum beamforming vector which maximises the lower bound of the SLNR as follows:

$$\mathbf{w}_k^{\text{opt}} = \frac{1}{\sqrt{M}} \bar{\mathbf{h}}_k^H \quad \text{for } k = 1, 2, \dots, U. \quad (13)$$

On the other hand, to minimise the denominator of (9) we have

$$\bar{\mathbf{h}}_j \mathbf{w}_k^{\text{opt}} = \frac{1}{\sqrt{M}} \bar{\mathbf{h}}_j \bar{\mathbf{h}}_k^H = 0, \quad j \neq k, \quad j = 1, 2, \dots, U, \quad (14)$$

where vector  $\mathbf{w}_k^{\text{opt}}$  is orthogonal to  $(1/\sqrt{M})\bar{\mathbf{h}}_j^H$ . Accordingly, from (12) we have

Equations (3) and (9) have different denominators to define SINR and SLNR. In (3), the user channel vector  $\mathbf{h}_k$  is used with other users' beamforming vectors  $\mathbf{w}_j|_{j=1}^U$ ,  $j \neq k$  to calculate SLNR of the  $k$ th user. While in (9) the  $k$ th user beamforming vector  $\mathbf{w}_k$  is used with other users' channel vectors  $\mathbf{h}_j|_{j=1}^U$ ,  $j \neq k$  to calculate SLNR of the  $k$ th user. Besides, the lower bound (LB) on the average SLNR of the  $k$ th user can be represented as  $E\{\text{SLNR}_k\} \geq \{\text{SLNR}_k\}_{\text{LB}}$  demonstrated in [11, 16]:

$$\begin{aligned} \text{SLNR}_k &= \{\text{SLNR}_k\}_{\text{LB}} \\ &= \frac{p \mathbf{w}_k^H \mathbf{h}_k^H \mathbf{h}_k \mathbf{w}_k}{1 + p \sum_{j=1, j \neq k}^U \mathbf{w}_k^H \mathbf{h}_j^H \mathbf{h}_j \mathbf{w}_k} \quad (10) \\ \{\text{SLNR}_k\}_{\text{LB}} &\triangleq \frac{p \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k}{1 + p \mathbf{w}_k^H \sum_{j=1, j \neq k}^U \mathbf{R}_j \mathbf{w}_k}. \end{aligned}$$

In (10), we can have  $\mathbf{R}_k \triangleq E\{\mathbf{h}_k \mathbf{h}_k^H\}$ :

$$\mathbf{R}_j \triangleq E\{\mathbf{h}_j \mathbf{h}_j^H\} = \frac{K_j}{K_j + 1} \bar{\mathbf{R}}_j + \frac{1}{K_j + 1} \mathbf{I}_M, \quad (11)$$

where  $\mathbf{R}_k$  is a Hermitian matrix (channel correlation matrix) which can be constructed as in (12), where  $\bar{\mathbf{R}}_j \triangleq \bar{\mathbf{h}}_j^H \bar{\mathbf{h}}_j$ , and  $\bar{\mathbf{h}}_j$  is calculated in (8). Hence,  $\mathbf{R}_k$  of the  $k$ th user can be decomposed regarding the deterministic component of channel vector  $\bar{\mathbf{h}}_k$  after normalisation by a factor  $\sqrt{M}$  [6]. Therefore, we can use the normalised version  $\bar{\mathbf{h}}_k / \sqrt{M}$  in the decomposition of the channel correlation matrix as in

$$\begin{aligned} \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k &= \frac{K_k}{K_k + 1} M + \frac{1}{K_k + 1} \\ \mathbf{w}_k^H \mathbf{R}_j \mathbf{w}_k &= \frac{1}{K_k + 1}, \quad \text{for } j \neq k. \end{aligned} \quad (15)$$

The maximum value of the lower bound can be achieved if  $\mathbf{w}_k^{\text{opt}}$  is orthogonal to  $(1/\sqrt{M})\bar{\mathbf{h}}_j^H$ . By substituting (15) in (10), we have

$$\{\text{SLNR}_k\}_{\text{LB}}^{\text{max}} = \frac{p ((K_k / (K_k + 1)) M + 1 / (K_k + 1))}{1 + p \sum_{j=1, j \neq k}^U (1 / (K_j + 1))}. \quad (16)$$

In (16), the lower bound is directly affected by the number of antenna elements  $M$ , the Ricean factor  $K$ , and the number of users  $U$ . Assuming that the number of users is increased as  $k = 1 \rightarrow U$ , taking  $K = 0$  and  $K \rightarrow \infty$ , the lower bounds of SLNR are expressed approximately as  $(p/(1 + p(U - 1)))|_{K=0}$  and  $M * p|_{K \rightarrow \infty}$ , respectively. Therefore, it is seen that the lower bound is independent of  $M$  when  $K = 0$ .

**3.1. Analysis of Linear Achievable Rates by SE and ZF.** In this section, we evaluate the Ergodic sum-rate capacity of SE obtained by (13) and (14) and assume that ZF capacity has perfect CSI when calculating the mean gap loss between  $R^{\text{SE}}$  and  $R^{\text{ZF}}$ . Based on the orthogonal beamforming condition in (13) and (14), the achievable sum-rate capacity obtained by the  $k$ th user is given by

$$\begin{aligned} R_k^{\text{SE}} &= E \left\{ \log_2 \left( 1 + \frac{p \left| (1/\sqrt{M}) \mathbf{h}_k \bar{\mathbf{h}}_k^{-H} \right|^2}{1 + p \sum_{j=1, j \neq k}^U \left| (1/\sqrt{M}) \mathbf{h}_k \bar{\mathbf{h}}_j^{-H} \right|^2} \right) \right\} \\ &\approx E \left\{ \log_2 \left( 1 + \frac{p \left| \mathbf{h}_k \mathbf{w}_k^{\text{opt}} \right|^2}{1 + p \sum_{j=1, j \neq k}^U \left| \mathbf{h}_k \mathbf{w}_j^{\text{opt}} \right|^2} \right) \right\}. \end{aligned} \quad (17)$$

The achievable sum-rate capacity by ZF is expressed as

$$R_k^{\text{ZF}} = \log_2 \left( 1 + p \left| \mathbf{h}_k \mathbf{w}_k^{\text{ZF}} \right| \right), \quad (18)$$

where  $\mathbf{w}_k^{\text{ZF}}$  is the unit normalised beamforming vector, which is selected as the  $k$ th column of the normalised matrix  $\mathbf{w} = \mathbf{H}(\mathbf{H}\mathbf{H}^H)^{-1}$  with  $\mathbf{H} = [\mathbf{h}_1^H \ \mathbf{h}_2^H \ \dots \ \mathbf{h}_U^H]$  to suppress the interference from all other users for  $|\mathbf{h}_j \mathbf{w}_k^{\text{ZF}}| = 0$ , if  $k \neq j$ . Accordingly, the mean gap loss  $\Delta R^{\text{ZF-SE}}$  with perfect instantaneous CSI is given by

$$\Delta R^{\text{ZF-SE}} = R^{\text{ZF}} - R^{\text{SE}}. \quad (19)$$

From the orthogonal condition  $(1/\sqrt{M}) \bar{\mathbf{h}}_j \bar{\mathbf{h}}_k^{-H} = 0|_{j \neq k}$  which is addressed in (14), we neglect the interference from the other users to the  $k$ th user regarding the signal component. In this case the sum-rate capacity  $R^{\text{SE}}$  used in (14) changes to the loose bound sum-rate capacity of the  $k$ th user expressed as [11]

$$\begin{aligned} R_k^{\text{SE}} &\geq E \left\{ \log_2 \left( 1 + p \left| \frac{1}{\sqrt{M}} \mathbf{h}_k \bar{\mathbf{h}}_k^{-H} \right|^2 \right) \right\} \\ &\quad - E \left\{ \log_2 \left( 1 + p \sum_{j=1, j \neq k}^U \left| \frac{1}{\sqrt{M}} \mathbf{h}_k \bar{\mathbf{h}}_j^{-H} \right|^2 \right) \right\}. \end{aligned} \quad (20)$$

Using (18) and (20), we rewrite (19) as the mean gap loss in (21) as

$$\begin{aligned} \Delta R^{\text{ZF-SE}} &\leq \sum_{k=1}^U E \left\{ \log_2 \left( 1 + p \left| \mathbf{h}_k \mathbf{w}_k^{\text{ZF}} \right|^2 \right) \right\} \\ &\quad - E \left\{ \log_2 \left( 1 + p \left| \frac{1}{\sqrt{M}} \mathbf{h}_k \bar{\mathbf{h}}_k^{-H} \right|^2 \right) \right\} \\ &\quad + E \left\{ \log_2 \left( 1 + p \sum_{j=1, j \neq k}^U \left| \frac{1}{\sqrt{M}} \mathbf{h}_k \bar{\mathbf{h}}_j^{-H} \right|^2 \right) \right\}. \end{aligned} \quad (21)$$

**3.2. Nonlinear Achievable Rate Analysis by DPC.** The BS can serve all users simultaneously and achieve maximum system capacity as much as possible as the CSI is known at the BS. In MIMO systems, the system capacity obtained by downlink strategy is called DPC. Practically, it is challenging to implement DPC because the encoding and decoding have high computations which is ineffective for large number of users. The achievable sum-rate capacity of the DPC can be written as [19]

$$\begin{aligned} R_{\text{DPC}} &= \max \log \left( \left| 1 + \sum_{k=1}^U p \mathbf{h}_k^* \mathbf{h}_k \right| \right), \\ &\quad p \geq 0, \quad \sum_{k=1}^U p \leq P. \end{aligned} \quad (22)$$

When the number of users is large, the DPC achievable sum-rate capacity is approximately expressed as [19, 20]

$$R_{\text{DPC}} = M \log \left( 1 + \frac{P}{M} \log U \right). \quad (23)$$

From (23), we can observe that the sum-rate capacity increases linearly with the number of antenna elements. However, the approximate achievable sum-rate capacity is a nonlinear function of  $U$ . Therefore, the DPC is regarded as a nonlinear model.

**3.3. Optimum Weight Vector Formulation.** In this subsection, we find the solution for optimum weight vector  $\mathbf{w}^{\text{opt}}$ , in which the maximum SLNR can be achieved to guarantee that all users have the ability to access the limited resources. For equal transmission of power to all users in a given case, we need to design  $\mathbf{w}_k$ ,  $k = 1, 2, \dots, U$  to maximise the SLNR for each user at the same time minimise interference. Hence, the optimum weight vector  $\mathbf{w}_k^{\text{opt}} = \arg \max(\text{SLR}_k)$  can be written as follows:

$$\mathbf{w}_k^{\text{opt}} = \arg \max \left( \frac{\left\| \mathbf{h}_k \mathbf{w}_k \right\|^2}{\sum_{j=1, j \neq k}^U \left\| \mathbf{h}_j \mathbf{w}_k \right\|^2} \right) \quad (24)$$

$$\text{SLR}_k = \frac{\left\| \mathbf{h}_k \mathbf{w}_k \right\|^2}{\sum_{j=1, j \neq k}^U \left\| \mathbf{h}_j \mathbf{w}_k \right\|^2}. \quad (25)$$

```

Step 1: Initial Inputs. number of users  $U$ ;
                        number of iteration  $N$ ;
                        initial non-zero vector  $\mathbf{X}_0$ ;
                        Tolerance Tol.
Step 2: For  $k = 1: U$  //to construct the CICTM for each user;
    Calculate  $\mathbf{h}_{w,k}$  and  $\tilde{\mathbf{h}}_k$  as defined in Section 2.2;
    Calculate for  $\mathbf{h}_k$  as in (7);
    Construct the CICTM  $\tilde{\mathbf{h}}_k$  for each user as in (29);
End
Step 3: For  $k = 1: U$  //to compute  $\mathbf{w}_k^{\text{opt}}$  for each user.
    Calculate matrix  $\mathbf{A}$  for  $k$ th user as in (31)
Step 4: While  $i \leq N$ ;
    set:  $\mathbf{y}_i = \mathbf{A}\mathbf{x}_{i-1}$ , and  $\mathbf{x}_i = \mathbf{y}_i/\alpha_i$ ; (normalisation)
    “the value of  $\mathbf{x}_i$  is the scaled version of  $\mathbf{y}_i$ ”
    if  $|\alpha_i - \alpha_{i-1}| \leq \text{Tol}$ , then  $i = N + 1$ ;
    else, set  $i = i + 1$ ;
    end
Step 5: calculate the dominant eigenvector  $\mathbf{v}_1 \approx \mathbf{x}_i$ ;
    the dominant eigenvalue  $\lambda_1 = \mathbf{x}_i^* \mathbf{A} \mathbf{x}_i / \mathbf{x}_i^* \mathbf{x}_i$ ;
End;
Step 6: SLNR maximise weight vector  $\tilde{\mathbf{w}}_k = \mathbf{x}_i$ ;
End;
Step 7: Output: the optimum weight vector  $\tilde{\mathbf{w}}_k$  for each user.

```

ALGORITHM 1: Steps for the proposed power iteration method.

Equation (24) is subjected to

$$\text{Constraint 1: } \|\mathbf{w}_k\|^2 = 1, \quad k = 1, 2, \dots, U \quad (26)$$

$$\text{Constraint 2: } \|\mathbf{h}_j \mathbf{w}_k\|^2 = 0, \quad j, k = 1, 2, \dots, U, \quad j \neq k.$$

Consequently, the SLNR is written as

$$\text{SLNR}_k = \frac{\|\mathbf{h}_k \mathbf{w}_k\|^2}{1 + \sum_{j=1, j \neq k}^U \|\mathbf{h}_j \mathbf{w}_k\|^2}. \quad (27)$$

Equation (27) is further written as

$$\text{SLNR}_k = \frac{\|\mathbf{h}_k \mathbf{w}_k\|^2}{1 + \|\tilde{\mathbf{h}}_k \mathbf{w}_k\|^2}, \quad (28)$$

where  $\tilde{\mathbf{h}}_k$  is congestate interfering channel transfer matrix (CICTM) of the  $k$ th user, which is an extended channel matrix that excludes  $\mathbf{h}_k$  only for the  $k$ th user. Then  $\mathbf{h}_k$  is written as

$$\tilde{\mathbf{h}}_k = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \dots \quad \mathbf{h}_{k-1} \quad \mathbf{h}_{k+1} \quad \dots \quad \mathbf{h}_{U-1} \quad \mathbf{h}_U]^T \cdot \left( \sum_{\substack{j=1 \\ j \neq k}}^U 1 \times M \right), \quad (29)$$

where  $\tilde{\mathbf{h}}_k \in \mathbb{C}^{M \times (U-1)}$  and  $\mathbf{h}_k \in \mathbb{C}^{1 \times M}$ . The optimum weight code vector can be derived from (28) as follows [21]:

$$\mathbf{w}_k^{\text{opt}} \propto \text{max. eigenvector} \left( (\mathbf{I}_M + \tilde{\mathbf{h}}_k^* \tilde{\mathbf{h}}_k)^{-1} \mathbf{h}_k^* \mathbf{h}_k \right), \quad (30)$$

where  $\mathbf{I}_M$  is  $M \times M$  square identity matrix.

The optimum weight code vector in (30) is called SVD solution, which is determined from the matrix in (29). Sometimes, when calculating the maximum eigenvalues for the given matrix, the dominant eigenvector does not correspond to the optimum code weight vector. Hence, the eigenvalues are put in ascending or descending order; then an optimum vector corresponding to the most significant eigenvalue is easily selected but at the expense of extra overhead to the system [4, 7]. To enhance the system performance and minimise the cost, a new approximation method expected to improve the average user achievable sum-rate capacity is proposed.

**3.4. Proposed Optimum Weight Vector Design.** A new method is developed based on power iteration method to find the optimum weight vector. The method requires only one weight vector during implementation. Equation (30) is modified and adopted as the main matrix in the proposed power approximation method as follows:

$$\mathbf{A} = (\mathbf{I}_M + \tilde{\mathbf{h}}_k^* \tilde{\mathbf{h}}_k)^{-1} \mathbf{h}_k^* \mathbf{h}_k. \quad (31)$$

**3.4.1. Power Approximation Method.** The power approximation method generates a sequence of vectors  $\mathbf{A}^i \mathbf{x}_0$ , where  $\mathbf{x}_0$  is a nonzero initial selected vector. By normalising these sequence vectors under conditions stated in Section 3.4.3, the vector converges to the dominant eigenvector corresponding to the most significant eigenvalue. The normalisation is used to ensure that the most significant component of the given iteration is equal to one [22]. The steps in the proposed power iteration method are given in Algorithm 1.



This algorithm is summarised in a sequence of iterations as follows:

$$\mathbf{x}_i = \frac{1}{\alpha_i} \mathbf{A} \mathbf{x}_{i-1}, \quad (32)$$

where  $\alpha_i$  is the component (magnitude) of vector  $\mathbf{A} \mathbf{x}_{i-1}$ . Before proceeding to the next iteration, it is necessary to scale down (normalised) the sequence vectors in each approximation so as to keep the largest component at unity.

**3.4.2. Convergence of the Algorithm.** The matrix  $\mathbf{A}$  is  $n \times n$  square matrix with  $n$  eigenvalues  $[\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n]$ , which are in descending order as  $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \dots \geq |\lambda_n|$ . With the initial vector  $\mathbf{X}_0$  chosen, the sequences  $\{\mathbf{X}_k = [x_1^{(k)} \ x_2^{(k)} \ \dots \ x_n^{(k)}]^T\}$  and  $\{c_k\}$  are recursively generated by the relations  $\mathbf{Y}_k = \mathbf{A} \mathbf{X}_k$  and  $\mathbf{X}_{k+1} = (1/c_{k+1}) \mathbf{Y}_k$ , where  $c_{k+1} = x_j^{(k)}$  and  $x_j^{(k)} = \max_{1 \leq i \leq n} \{|x_i^{(k)}|\}$ . These sequences converge to the dominant eigenvector  $\mathbf{V}_1$  and eigenvalue  $\lambda_1$ , respectively, as  $\lim_{k \rightarrow \infty} \mathbf{X}_k = \mathbf{V}_1$  and  $\lim_{k \rightarrow \infty} c_k = \lambda_1$ . The proof of convergence is referred to in Appendix A.

**3.4.3. The Speed of Convergence.** Referring to (A.6) in Appendix A, we observe that the coefficient of  $\mathbf{V}_j$  in the sequence  $\mathbf{X}_k$  goes to zero, which is proportional to  $(\lambda_j/\lambda_1)^k$ ; the convergence speed of sequence  $\{\mathbf{X}_k\}$  to  $\mathbf{V}_1$  is governed by  $(\lambda_2/\lambda_1)^k$ . Therefore, the rate of convergence and the convergence of the constants  $\{c_k\}$  to  $\lambda_1$  are linear. For any linear convergent sequence  $\{p_k\}$ , we can use the Aitken  $\Delta^2$  technique which is used to make the linearly convergent sequences fast. The new convergence of the sequence  $\{p_k\}$  can be written as

$$\hat{p}_k = \frac{(p_{k+1} - p_k)^2}{p_{k+2} - 2p_{k+1} + p_k}. \quad (33)$$

The convergence property of the algorithm is given by the following.

**Theorem 1.** Assume that there is one and only one eigenvalue  $\lambda_1$  of  $\mathbf{A}$ , and  $\lambda_1$  is semisimple; then either the initial vector  $\mathbf{x}_0$  which has no component in the invariant subspace associated with  $\lambda_1$  or the sequence of vectors generated by the algorithm converges to the eigenvector associated with  $\lambda_1$  and  $i$  and converges to  $\lambda_1$ . The proof is shown in Appendix B.

**Definition 2.** Eigenvalue  $\lambda_1$  of matrix  $\mathbf{A}$  is semisimple with the degree of  $n$  if it has a geometric multiplicity one and algebraic multiplicity  $n$ . The vector  $\mathbf{A}^i \mathbf{x}_0$  is normalised by a specific scalar  $\alpha_i$  to make the most significant component of the vector unity. The initial vector  $\mathbf{x}_0$  is decomposed as follows:

$$\mathbf{x}_0 = \sum_{j=1}^p P_j \mathbf{x}_0, \quad (34)$$

where  $P_j$ ,  $j = 1, 2, \dots, p$ , are the spectral projectors associated with the eigenvalues  $\lambda_j$ ,  $j = 1, 2, \dots, p$ .

From the formula  $\mathbf{A} P_j = P_j (\lambda_j \mathbf{I} + \mathbf{D}_j)$ , the power can be written as  $\mathbf{A}^i P_j = P_j (\lambda_j \mathbf{I} + \mathbf{D}_j)^i$ , where  $\mathbf{I}$  is the identity matrix and  $\mathbf{D}_j$  is the diagonal matrix. Consequently, we have

$$\begin{aligned} \mathbf{x}_i &= \frac{1}{\alpha_i} \mathbf{A}^i \sum_{j=1}^p P_j \mathbf{x}_0 = \frac{1}{\alpha_i} \sum_{j=1}^p \mathbf{A}^i P_j \mathbf{x}_0 \\ &= \frac{1}{\alpha_i} \sum_{j=1}^p P_j (\lambda_j \mathbf{I} + \mathbf{D}_j)^i \mathbf{x}_0. \end{aligned} \quad (35)$$

Referring to the Definition 2,  $\mathbf{D}_1 = 0$  because  $\lambda_1$  is a semisimple eigenvalue. Thus, we obtain

$$\begin{aligned} \mathbf{x}_i &= \frac{1}{\alpha_i} \sum_{j=1}^p P_j (\lambda_j P_j + \mathbf{D}_j)^i \mathbf{x}_0 \\ &= \frac{1}{\alpha_i} \left( \lambda_1^i P_1 \mathbf{x}_0 + \sum_{j=2}^p P_j (\lambda_j P_j + \mathbf{D}_j)^i \mathbf{x}_0 \right) \\ &= \frac{\lambda_1^i}{\alpha_i} \left( P_1 \mathbf{x}_0 + \sum_{j=2}^p \frac{1}{\lambda_1^i} (\lambda_j P_j + \mathbf{D}_j)^i P_j \mathbf{x}_0 \right). \end{aligned} \quad (36)$$

The spectral radius of each operator  $(\lambda_j P_j + \mathbf{D}_j) \lambda_1^{-1} < 1$  and  $|\lambda_j \lambda_1^{-1}| < 1$  means that the  $i$ th power will converge to zero. Theorem 1 is true when  $P_1 \mathbf{x}_0 = 0$ . However, if  $P_1 \mathbf{x}_0 \neq 0$ ,  $\mathbf{x}_i$  converge to normalise  $P_1 \mathbf{x}_0$  so that the most significant component is one. Meanwhile, the scalar  $\alpha_i$  converges to the eigenvalue  $\lambda_1$ , which is an immediate consequence in the form

$$\mathbf{A} \mathbf{x}_{i-1} = \alpha_i \mathbf{x}_i. \quad (37)$$

Thus, the sequence of vectors  $\mathbf{x}_i$  is proven to converge to the optimum weight vector which maximises the SLNR in (28). The optimum weight vector is equal to the vector  $\mathbf{x}_i$ :

$$\widehat{\mathbf{w}} = \mathbf{x}_i. \quad (38)$$

The result in (38), which is the main objective of this paper, maximises the  $k$ th user SLNR when substituted in (28) to find  $\text{SLNR}_k$ .

**3.5. Cost Analysis.** In this subsection, the cost function is analyzed regarding storage space required for the proposed power iteration technique and SVD method. Referring to (30), the value  $(\mathbf{I}_M + \tilde{\mathbf{h}}_k^* \tilde{\mathbf{h}}_k)^{-1} \tilde{\mathbf{h}}_k^* \tilde{\mathbf{h}}_k$  results in a new square matrix with dimension  $M \times M$ , where  $M$  denotes the number of antenna elements. Therefore, finding  $\mathbf{w}_k^{\text{opt}}$  for each user by using the solution of SVD in (30), we need to find at least two matrices,  $M \times M$  left eigenvectors matrix,  $M \times M$  right eigenvectors matrix, and  $M$  diagonal eigenvalues from the original  $M \times M$  matrix. On the other hand, the required storage memory for the  $k$ th user is equal to the size of three matrices in addition to the original square matrix. Hence, the total storage becomes  $(3M^2 + M) * U$  (full SVD), where  $U$  is the number of users. Then, the determination of the

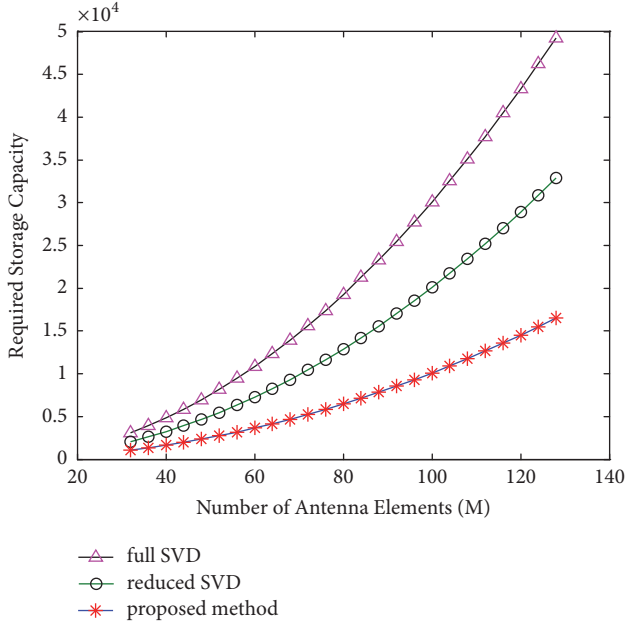


FIGURE 2: Storage capacity required by the proposed technique and the SVD (full SVD and reduced SVD) method regarding the number of antenna elements.

optimum weight vector can be from the right eigenvector or left eigenvector based on the corresponding maximum eigenvalue. Furthermore, we can reduce the size by taking only one side eigenvectors matrix (left or right) to save  $M^2$  redundant storage size. Thus the reduced new storage capacity can be written as  $(2M^2 + M) * U$  (reduced SVD).

Similarly, only the original square matrix with dimension  $M \times M$  and a vector with size  $M \times 1$  is required in the proposed power approximation method. In this case, there is no need for storing the previous iteration result as it can be overwritten up to the last iteration or reaches the determined tolerance. Thus, the required storage for the power iteration technique is  $(M^2 + M) * U$  which is much less than full and reduced SVDs.

Figure 2 shows the storage capacity required in the proposed power iteration technique and SVD method. From Figure 2, it is observed that the number of antenna elements has a significant effect on the cost function. However, the proposed method has much less storage space compared to SVD hence recommended for a massive MIMO with large antennas.

#### 4. Simulation Results and Discussion

In this section, the numerical results are presented with the following assumptions: the equal power allocation strategy ( $p_k = P/U$ ) and noise effect ( $\sigma_k = \sigma$ ) for all users are the same [23]. Additionally, the AoD for each user is in a horizontal direction with the uniform distribution in the range  $[-\pi, \pi]$ . Furthermore, for each user the Ricean  $K$ -factor has uniform distribution in the range of  $[K_{\min}, K_{\max}]$ , where  $K_{\min} = 10$  dB and  $K_{\max} = 30$  dB. Moreover, the number of users is increased

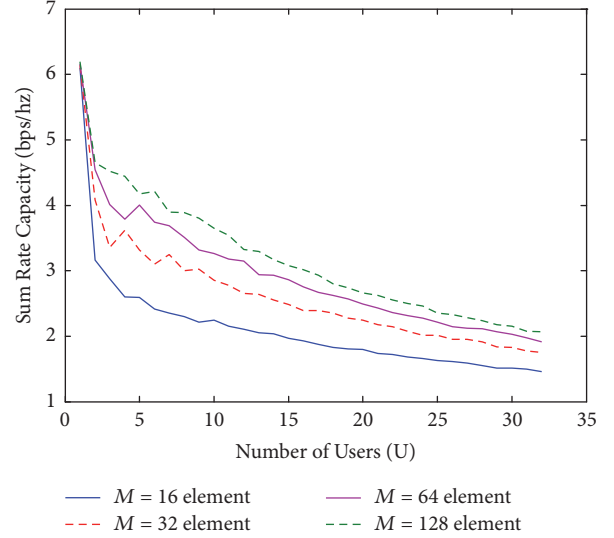


FIGURE 3: Sum-rate capacities for SE on the number of antenna elements.

up to  $U = 32$ . For SE and ZF analysis  $M = 16, 32, 64,$  and  $128$  are selected as the numbers of transmission antenna elements. The ergodic sum-rate capacities of SE and ZF are obtained in Figures 3 and 4. In Figure 3. It is observed that the average system capacity of SE is, respectively, directly and inversely proportional to the number of antenna elements and number of users. Moreover, when the number of users is small, the impact of  $M$  on the SE capacity is more significant and conversely when the numbers are large, the impact is less significant. The SE achieves average capacities approximately in the range of 1.6–6.2 bps/Hz. At maximum number of users ( $U = 32$ ) with  $M = 16$ , the corresponding SE sum-rate capacity is less than 1.6 bps/Hz. This is because each user faces CCI interference which is considered as sum of leakage signals from all other cochannel users. Referring to (17), it is validated that as the number of users increases, the overall contributed interference to the  $k$ th user also increases. Therefore, the increase in the denominator in (17) causes a reduction in the sum-rate capacity per user resulting in an average lower system capacity.

In Figure 4, it is shown that the ergodic sum-rate capacities obtained by ZF are directly proportional to the number of antenna elements. By referring to (18) the interference from the other users is perfectly suppressed by the orthogonal beamforming for user channel and weight vectors. These vectors are orthogonal to each other based on the condition  $|\mathbf{h}_j \mathbf{w}_k^{\text{ZF}}| = 0$  in which  $k \neq j$ .

Moreover, the number of antenna elements  $M$  has much contribution on the average system capacity with ZF than SE as comparatively shown in Figures 4 and 3. The maximum achievable capacities are 105 bps/Hz and 6.2 bps/Hz for ZF and SE, respectively, because of the interference effects on SE.

Figure 5 shows the upper bound of the mean-rate capacity gap loss ( $R^{\text{ZF}} - R^{\text{SE}}$ ) between ZF and SE. Since the number of users tends to be large, the difference in the mean rate capacity is most likely the ZF capacity, because the capacity obtained

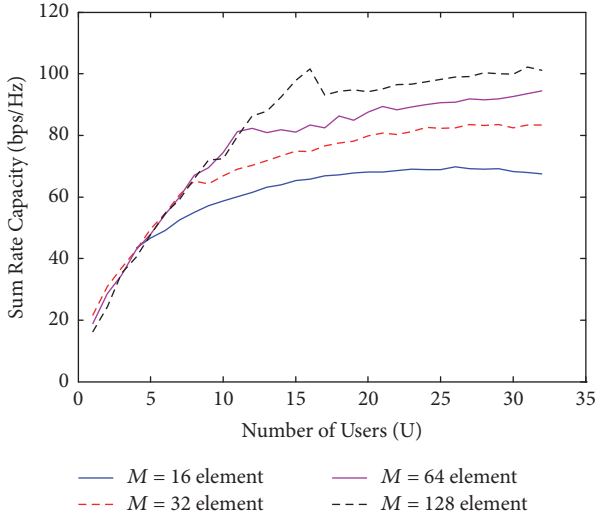


FIGURE 4: Achievable sum-rate capacities for ZF beamforming on the number of antenna elements.

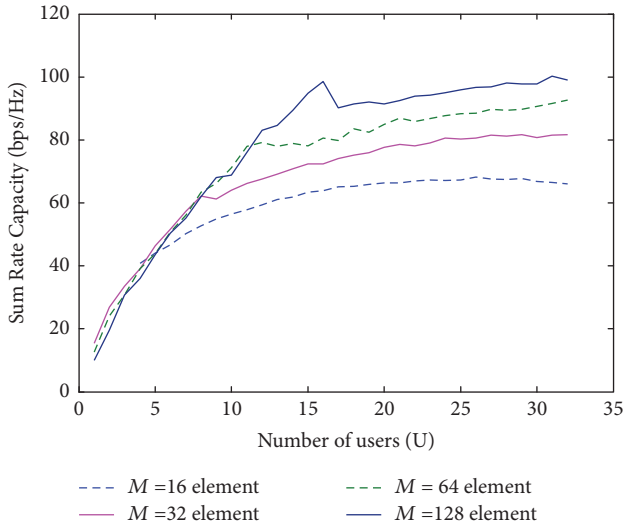


FIGURE 5: Mean gap loss capacity ( $R^{ZF} - R^{SE}$ ) which is the difference between ZF and SE capacities.

by ZF is close to 15 times that obtained by SE, which precisely affects the mean-gap loss. Nevertheless, when the number of users increases, the difference tends to be very big. It is explicit that when the number of users is large under any number of antenna elements, the achievable sum-rate capacity obtained by SE almost tends to be zero. However, the ergodic sum-rate capacity achieved by ZF positively increases with increase in the number of antenna elements and tends to a fixed level with large number of users. Meanwhile, the achievable sum-rate capacity is not profoundly affected by the number of users.

As shown in Figures 3–5, the simulation results depict that the ZF seems ideal due to a perfect cancellation of CCI. However, in practice SE is dominant and realisable because it is challenging to ensure that the interference is perfectly cancelled by the orthogonal condition between the channel

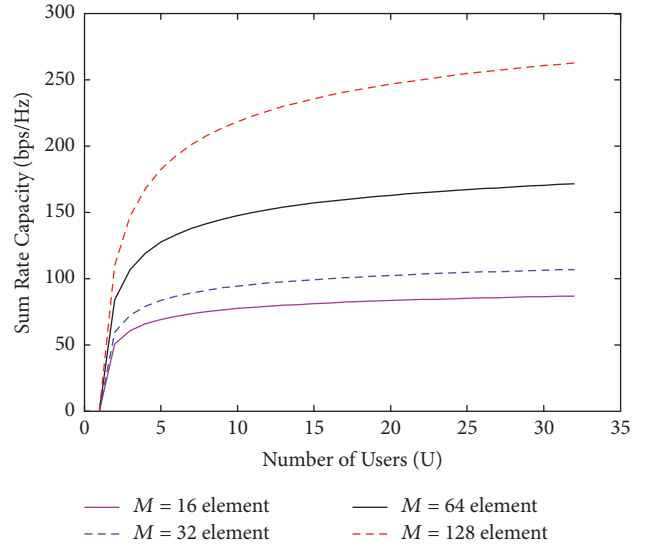


FIGURE 6: Achievable sum-rate capacities for DPC on the number of antenna elements.

vector and optimum weight vector. Thus, carefully focusing the beams alignment between the BS and the users a high system capacity can be achieved.

In Figure 6, the antenna elements  $M = 16, 32, 64,$  and  $128$  are considered with number of users up to 32 to satisfy  $M \geq U$  and show the effects of  $M < U$  ( $M = 16$ ). The results show that the ergodic sum-rate capacity achieved by DPC is nonlinearly increased with the number of users. Moreover, if the number of users is fixed to 25, at the number of antenna elements 16, 32, 64, and 128, respectively, the corresponding ergodic sum-rate capacities are 80, 105, 175, and 250 bps/Hz, respectively. As the number of users is increased, the system capacity also increases. Thus, comparing the DPC with ZF and SE achieves much higher capacity with large number of antenna elements and users; hence it can provide multiuser diversity gain.

The results in Figures 7–9 are used to comparatively evaluate the proposed power iteration method in (38) and the SVD method in (30). In Figure 7, it is observed that the SLNR in the range  $-10.5$  dB to 37 dB achieves the system capacities in the range 5 bps/Hz to 11 bps/Hz for the proposed method. On the other hand, the SVD method has the SLNR in the range  $-15$  dB to 31 dB that achieves system capacities in the range 1 bps/Hz to 6 bps/Hz. It is clear that the proposed method has the average throughput nearly two times that of SVD. The main reason is due to the efficiency of the weight vector obtained in the proposed method that is capable of minimising the CCI much more than in the SVD. It is expected that the system capacity can be enhanced in case the BS has more antenna elements but at a higher simulation time. Additionally, the SVD method has a drawback in finding the exact eigenvector corresponding to the dominant eigenvalue in which the eigenvalues are not always in an orderly way (ascending or descending), which limits the SVD performance. Figures 8 and 9 are the Cumulative Distribution Function (CDF) and Probability Density Function (PDF) of

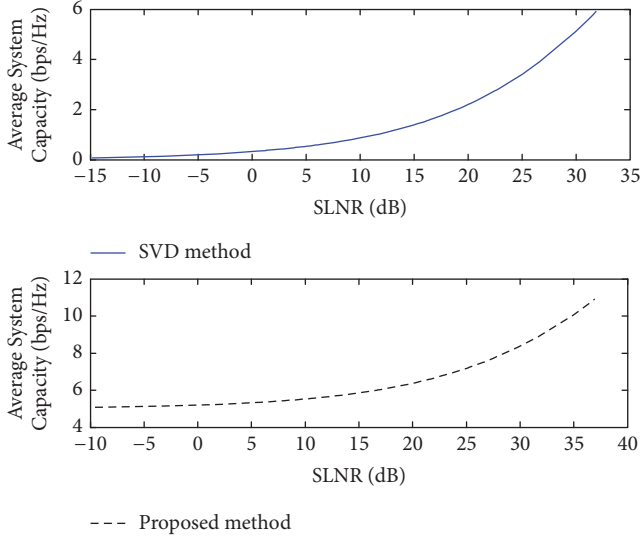


FIGURE 7: Sum-rate capacities for the proposed power iteration technique and SVD method.

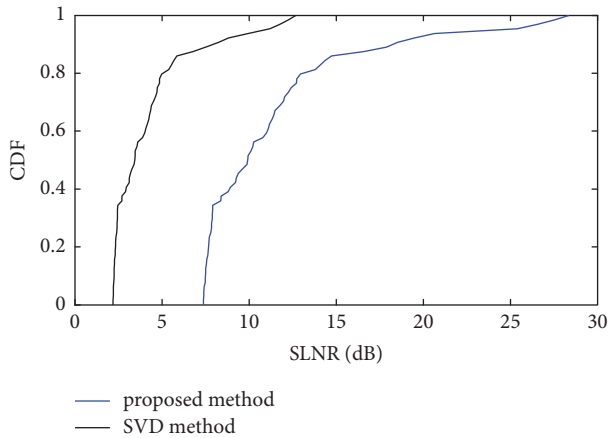


FIGURE 8: CDFs of the SLNRs for the proposed power iteration and SVD methods.

SLNR of the proposed and SVD methods, respectively. It is observed that the proposed method achieves a much better SLNRs than the SVD method.

## 5. Conclusion

In this paper, based on MU-MIMO with massive antennas, the ergodic capacity for linear sum-rate analysis including SE and ZF beamforming based on SLNR technique is investigated. In addition, the nonlinear sum-rate analysis using DPC in Ricean fading channels based on SLNR was undertaken. A new method is proposed to find an optimum beam weight vector by exploring the power iteration method using eigenvector approximation. The number of antenna elements is selected up to 128 dramatically increasing the system capacity. By way of simulation, the most significant dominant eigenvector to maximise SLNR as well as minimise the CCI is obtained. By comparing with the SVD method,

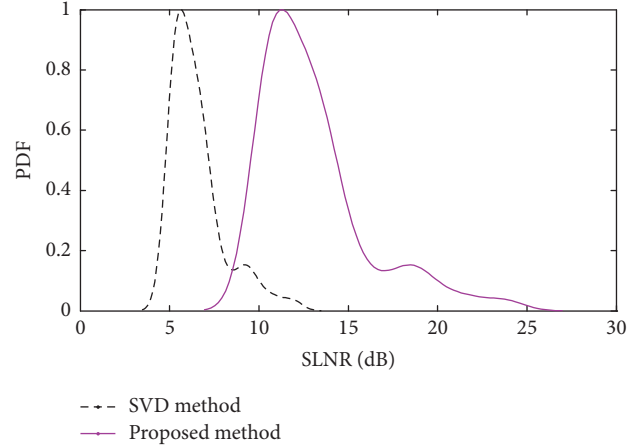


FIGURE 9: PDFs of the SLNRs for the proposed power iteration and SVD methods.

the proposed method achieves higher performance regarding mean achievable sum-rate capacity per user. The throughput of the proposed method is in the range of 5 bps/Hz–11 bps/Hz, while that of SVD in the range of 1 bps/Hz–6 bps/Hz. Therefore, the proposed method can provide significant system capacity enhancement. In future work, the proposed power iteration technique is recommended for a 5G MU massive antenna system.

## Appendix

### A. Proof of the Algorithm

As we know, matrix  $\mathbf{A}$  has  $n$  eigenvalues and  $n$  corresponding eigenvectors  $\mathbf{V}_j$  ( $j = 1, 2, \dots, n$ ) which are linearly independent and normalised and form a basis for  $n$  dimensional space. Thus, let the initial vector  $\mathbf{X}_0$  be written as the linear combination form as

$$\mathbf{X}_0 = b_1 \mathbf{V}_1 + b_2 \mathbf{V}_2 + \dots + b_n \mathbf{V}_n. \quad (\text{A.1})$$

Assume vector  $\mathbf{X}_0 = [x_1 \ x_2 \ \dots \ x_n]^T$  is chosen in such a way that  $b_1 \neq 0$ , and the elements of  $\mathbf{X}_0$  are scaled so that  $\max_{1 \leq j \leq n} \{|x_j|\} = 1$ . As the vectors  $\{\mathbf{V}_j\}_{j=1}^n$  represent the eigenvectors of the matrix  $\mathbf{A}$ , the multiplication  $\mathbf{A}\mathbf{X}_0$  is followed by normalisation as follows:

$$\begin{aligned} \mathbf{Y}_0 &= \mathbf{A}\mathbf{X}_0 = \mathbf{A}(b_1 \mathbf{V}_1 + b_2 \mathbf{V}_2 + \dots + b_n \mathbf{V}_n) \\ &= b_1 \mathbf{A}\mathbf{V}_1 + b_2 \mathbf{A}\mathbf{V}_2 + \dots + b_n \mathbf{A}\mathbf{V}_n \\ &= b_1 \lambda_1 \mathbf{V}_1 + b_2 \lambda_2 \mathbf{V}_2 + \dots + b_n \lambda_n \mathbf{V}_n \\ &= \lambda_1 \left( b_1 \mathbf{V}_1 + b_2 \left( \frac{\lambda_2}{\lambda_1} \right) \mathbf{V}_2 + \dots + b_n \left( \frac{\lambda_n}{\lambda_1} \right) \mathbf{V}_n \right). \end{aligned} \quad (\text{A.2})$$

Moreover, also we have

$$\begin{aligned} \mathbf{X}_1 &= \frac{\lambda_1}{c_1} \left( b_1 \mathbf{V}_1 + b_2 \left( \frac{\lambda_2}{\lambda_1} \right) \mathbf{V}_2 + \dots + b_n \left( \frac{\lambda_n}{\lambda_1} \right) \mathbf{V}_n \right). \end{aligned} \quad (\text{A.3})$$

After  $k$  iterations we arrive at

$$\begin{aligned}
\mathbf{Y}_{k-1} &= \mathbf{A}\mathbf{X}_{k-1} = \mathbf{A} \frac{\lambda_1^{k-1}}{c_1 c_2 \cdots c_{k-1}} \left( b_1 \mathbf{V}_1 \right. \\
&\quad \left. + b_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{V}_2 + \cdots + b_n \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{V}_n \right) \\
&= \frac{\lambda_1^{k-1}}{c_1 c_2 \cdots c_{k-1}} \left( b_1 \mathbf{A}\mathbf{V}_1 + b_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{A}\mathbf{V}_2 + \cdots \right. \\
&\quad \left. + b_n \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{A}\mathbf{V}_n \right) = \frac{\lambda_1^{k-1}}{c_1 c_2 \cdots c_{k-1}} \left( b_1 \lambda_1 \mathbf{V}_1 \right. \\
&\quad \left. + b_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \lambda_2 \mathbf{V}_2 + \cdots + b_n \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \lambda_n \mathbf{V}_n \right) \\
&= \frac{\lambda_1^k}{c_1 c_2 \cdots c_{k-1}} \left( b_1 \mathbf{V}_1 + b_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \mathbf{V}_2 + \cdots \right. \\
&\quad \left. + b_n \left( \frac{\lambda_2}{\lambda_1} \right)^k \mathbf{V}_n \right).
\end{aligned} \tag{A.4}$$

Furthermore, we have

$$\begin{aligned}
\mathbf{X}_k &= \frac{\lambda_1^k}{c_1 c_2 \cdots c_k} \left( b_1 \mathbf{V}_1 + b_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{V}_2 + \cdots \right. \\
&\quad \left. + b_n \left( \frac{\lambda_2}{\lambda_1} \right)^{k-1} \mathbf{V}_n \right).
\end{aligned} \tag{A.5}$$

Since we assumed that  $|\lambda_j|/|\lambda_1| < 1$  for  $j = 2, 3, \dots, n$ , then we have

$$\lim_{k \rightarrow \infty} b_j \left( \frac{\lambda_j}{\lambda_1} \right)^{k-1} \mathbf{V}_j = 0, \quad \text{for } j = 2, 3, \dots, n. \tag{A.6}$$

Hence it follows that

$$\lim_{k \rightarrow \infty} \mathbf{X}_k = \lim_{k \rightarrow \infty} \frac{b_1 \lambda_1^k}{c_1 c_2 \cdots c_k} \mathbf{V}_1. \tag{A.7}$$

We need both vectors  $\mathbf{X}_k$  and  $\mathbf{V}_1$  to be normalised and their most significant component is 1.

$$\lim_{k \rightarrow \infty} \frac{b_1 \lambda_1^k}{c_1 c_2 \cdots c_k} = 1. \tag{A.8}$$

## B. Proof of the Convergence

For the sequence of vectors  $\{\mathbf{X}_k\}$  which converges to the dominant eigenvector:

$$\lim_{x \rightarrow \infty} \mathbf{X}_k = \mathbf{V}_1. \tag{B.1}$$

By substituting  $k$  with  $k-1$ , it yields

$$\lim_{k \rightarrow \infty} \frac{b_1 \lambda_1^{k-1}}{c_1 c_2 \cdots c_{k-1}} = 1. \tag{B.2}$$

By merging both into (A.8), we get

$$\lim_{x \rightarrow \infty} \frac{\lambda_1}{c_k} = \lim_{x \rightarrow \infty} \frac{b_1 \lambda_1^k / c_1 c_2 \cdots c_k}{b_1 \lambda_1^{k-1} / c_1 c_2 \cdots c_{k-1}} = \frac{1}{1} = 1. \tag{B.3}$$

Therefore, the sequences of the constants  $\{c_k\}$  converge to the dominant eigenvalue as

$$\lim_{x \rightarrow \infty} c_k = \lambda_1. \tag{B.4}$$

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work is supported by Key program of Beijing Municipal Natural Science Foundation with no. 17L20052 and also supported by National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation under Grant no. 201600012, and the National Nature Science Foundation of China (NSFC) under Grant no. 61771194.

## References

- [1] M. S. Islam, M. Kamruzzaman, T. Jessy, M. S. Zahan, and M. S. Hassan, "Performance analysis of massive MIMO for 5G wireless communication systems," in *Proceedings of the 2016 IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, pp. 1579–1583, April 2016.
- [2] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "Introduction to the special issue on signal processing for millimeter wave wireless communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 433–435, 2016.
- [3] H. Shen, W. Xu, A. L. Swindlehurst, and C. Zhao, "Transmitter optimization for per-antenna power constrained multi-antenna downlinks: an SLNR maximization methodology," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2712–2725, 2016.
- [4] M. Sadek, A. Tarighat, and A. H. Sayed, "A lalage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, 2007.
- [5] M. Vu and A. Paulraj, "Linear precoding for MIMO wireless correlated channels with non-zero means: K factor analysis, extension to non-orthogonal STBC," in *Proceedings of the (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, pp. 1113–1116, Philadelphia, PA, USA, 2005.
- [6] C. Zhang, Z. Lu, Y. Huang, J. Zhang, and L. Yang, "Statistical beamforming for FDD massive MIMO downlink systems," in *Proceedings of the IEEE/CIC International Conference on Communications in China, ICC 2015*, November 2015.
- [7] A. Tarighat, M. Sadek, and A. H. Sayed, "A multi user beamforming scheme for downlink mimo channels based on

- maximizing signal-to-leakage ratios,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 3, pp. 1129–1132, IEEE, March 2005.
- [8] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, “Capacity analysis of NOMA with mmWave massive MIMO systems,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1606–1618, 2017.
- [9] J. Choi, “Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems,” *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 791–800, 2015.
- [10] S. Park, J. Park, A. Yazdan, and R. W. Heath, “Optimal user loading in massive MIMO systems with regularized zero forcing precoding,” *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 118–121, 2017.
- [11] S. Jin, W. Tan, M. Matthaiou, J. Wang, and K.-K. Wong, “Statistical eigenmode transmission for the MU-MIMO downlink in rician fading,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6650–6663, 2015.
- [12] X. Li, S. Jin, H. A. Suraweera, J. Hou, and X. Gao, “Statistical 3-D beamforming for large-scale MIMO downlink systems over rician fading channels,” *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1529–1543, 2016.
- [13] I. Boukhedimi, A. Kammoun, and M.-S. Alouini, “Coordinated SLNR based precoding in large-scale heterogeneous networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 3, pp. 534–548, 2017.
- [14] X. Wu, W. Zhao, X. Sha, F. Labeau, and Y. Wang, “SLNR beamforming based iterative power allocation in TD-LTE-A downlink,” in *Proceedings of the 11th International Wireless Communications and Mobile Computing Conference, IWCMC 2015*, pp. 84–89, August 2015.
- [15] L. Liu, D. W. Matolak, C. Tao, and Y. Li, “Analysis of an upper bound on the effects of large scale attenuation on uplink transmission performance for massive MIMO systems,” *IEEE Access*, vol. 5, pp. 4285–4297, 2017.
- [16] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, “Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2181–2194, 2017.
- [17] X. Zhao, A. M. Abdo, C. Xu, S. Geng, J. Zhang, and I. Memon, “Dimension reduction of channel correlation matrix using CUR-decomposition technique for 3-D massive antenna system,” *IEEE Access*, vol. 6, pp. 3031–3039, 2018.
- [18] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [19] T. Yoo and A. Goldsmith, “Optimality of zero-forcing beamforming with multiuser diversity,” in *Proceedings of the 2005 IEEE International Conference on Communications, ICC 2005*, pp. 542–546, May 2005.
- [20] M. Sharif and B. Hassibi, “A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users,” *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 11–15, 2007.
- [21] J. Wang, X. Xie, and Q. Zhang, “A way to reduce ICI of multi-user MIMO-OFDM system with precoding,” in *Proceedings of the International Conference on Advanced Computer Control, ICACC 2009*, pp. 134–137, January 2009.
- [22] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 2nd edition, 2011.
- [23] A. M. Abdo, X. Zhao, R. Zhang, Y. Zhang, K. Eguchi, and T. Chen, “Codebook metrics evaluation for millimeter wave communications by antenna array response and signal to noise ratio,” *ITM Web of Conferences*, vol. 17, p. 03011, 2018.