An International ICT R&D Journal Sponsored by ZTE Corporation

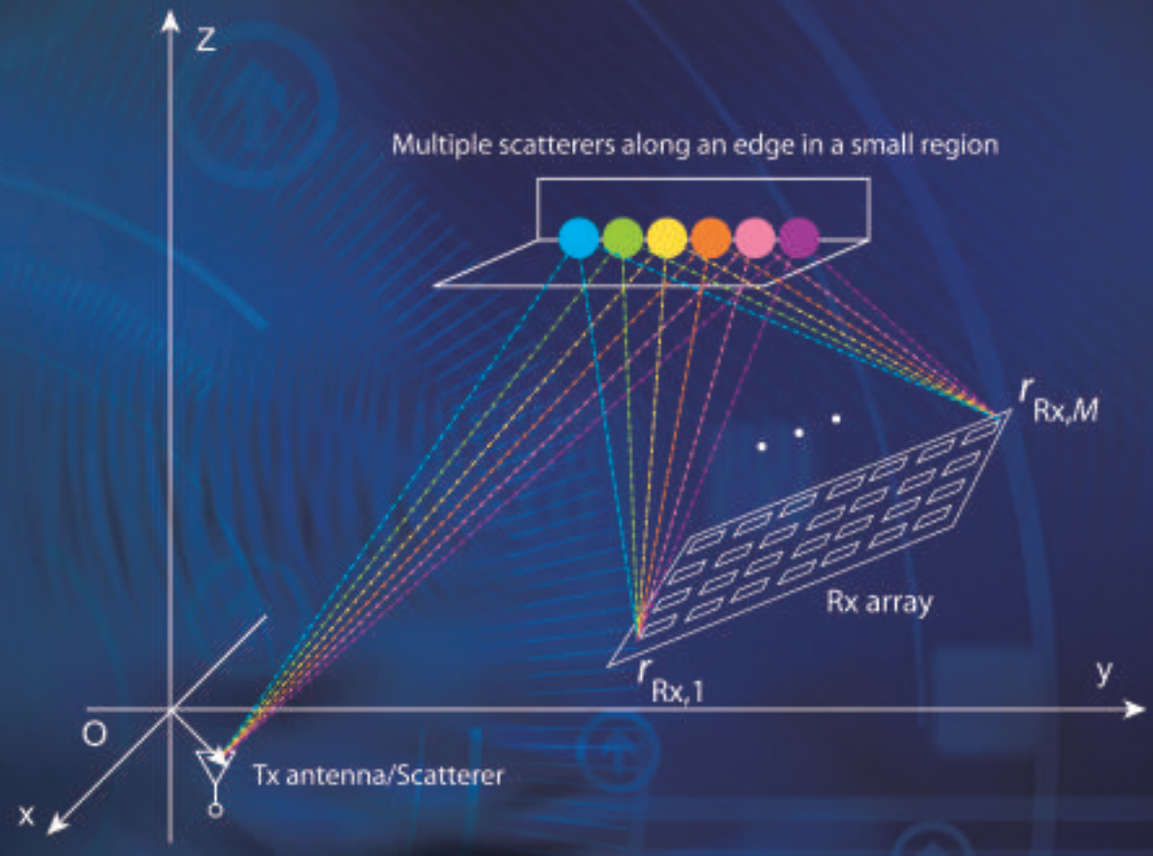
ISSN 1673-5188
CN 34-1294/ TN
CODEN ZCTOAK

ZTE COMMUNICATIONS

tech.zte.com.cn

February 2017, Vol. 15 No. 1

SPECIAL TOPIC:
Channel Measurement and Modeling for Heterogeneous 5G



ZTE中兴 ZTE COMMUNICATIONS

VOLUME 15 NUMBER 1 FEBRUARY 2017

ZTE Communications Editorial Board

CONTENTS

Special Topic: Channel Measurement and Modeling for Heterogeneous 5G

CONTENTS

ZTE COMMUNICATIONS
Vol. 15 No. 1 (Issue 55)
Quarterly
First English Issue Published in 2003

Supervised by:
Anhui Science and Technology Department

Sponsored by:
Anhui Science and Technology Information Research Institute and ZTE Corporation

Staff Members:
Editor-in-Chief: CHEN Jie
Executive Associate Editor-in-Chief: HUANG Xinming
Editor-in-Charge: ZHU Li
Editors: XU Ye, LU Dan, ZHAO Lu
Producer: YU Gang
Circulation Executive: WANG Pingping
Assistant: WANG Kun

Editorial Correspondence:
Add: 12F Kaixuan Building,
329 Jinzhai Road,
Hefei 230061, P. R. China
Tel: +86-551-65533356
Fax: +86-551-65850139
Email: magazine@zte.com.cn

Published and Circulated (Home and Abroad) by:
Editorial Office of
ZTE Communications

Printed by:
Hefei Tiancai Color Printing Company

Publication Date:
February 25, 2017

Publication Licenses:

ISSN 1673-5188
CN 34-1294/ TN

Advertising License:
皖合工商广字0058号

Annual Subscription:
RMB 80

Review

Research Paper

Roundup

Guest Editorial
CUI Shuguang and CHENG Xiang

Channel Measurement and Modeling for Heterogeneous 5G

▶ CUI Shuguang

CUI Shuguang received his Ph.D. in electrical engineering from Stanford University, USA in 2005. He is currently a Childs Family Endowed Professor in electrical and computer engineering at the University of California-Davis. His current research interests focus on data driven large-scale information analysis and system design. He was selected as the Thomson Reuters Highly Cited Researcher and listed in the Worlds' Most Influential Scientific Minds by ScienceWatch in 2014. He was the recipient of the IEEE Signal Processing Society 2012 Best Paper Award. He has served as the general co-chair and TPC co-chairs for many IEEE conferences. He has also been serving as the area editor for *IEEE Signal Processing Magazine*, and associate editors for *IEEE Transactions on Big Data*, *IEEE Transactions on Signal Processing*, *IEEE JSAC Series on Green Communications and Networking*, and *IEEE Transactions on Wireless Communications*. He was the elected member for IEEE Signal Processing Society SPCOM Technical Committee (2009–2014) and the elected Chair for IEEE ComSoc Wireless Technical Committee (2017–2018). He is a member of the Steering Committee for both *IEEE Transactions on Big Data* and *IEEE Transactions on Cognitive Communications and Networking*. He is also a member of the IEEE ComSoc Emerging Technology Committee. He was elected as an IEEE Fellow in 2013 and an IEEE ComSoc Distinguished Lecturer in 2014.

▶ CHENG Xiang

CHENG Xiang received the Ph.D. degree from Heriot-Watt University, UK, and the University of Edinburgh, UK in 2009, where he received the Postgraduate Research Thesis Prize. He is currently an associate professor at Peking University, China. He has published more than 120 research papers in journals and conference proceedings. His research interests include mobile propagation channel modeling and simulation, next generation mobile cellular systems, intelligent transportation systems, and hardware prototype development. Dr. Cheng was the recipient of the IEEE Leonard G. Abraham Prize (IEEE JSAC best paper award) in 2016, the IEEE Asia Pacific Outstanding Young Researcher Award in 2015, and Best Paper Awards at IEEE ITST'12, ICCC'13, ITSC'14, ICC'16, ICNC'17. He has served as Symposium Lead Chair, Co-Chair, and a member of the Technical Program Committees for several international conferences. He is an associate editor for *IEEE Transactions on Intelligent Transportation Systems*.

While cellular networks have continuously evolved in recent years, the industry has clearly seen unprecedented challenges to meet the exponentially growing expectations in the near future. The 5G system is facing grand challenges such as the ever-increasing traffic volumes and remarkably diversified services connecting humans and machines alike. As a result, the future network has to deliver massively increased capacity, greater flexibility, incorporated computing capability, support of significantly extended battery lifetime, and accommodation of varying payloads with fast setup and low latency, etc. In particular, as 5G requires more spectrum resource, higher frequency bands are desirable. Nowadays, millimeter wave has been widely accepted as one of the main communication bands for 5G. As a result, envisioned 5G research and development are inclined to be heterogeneous, with possibly ultra dense network layouts due to their capability to support high speed connections, flexibility of resource management, and integration of distinct access technologies. In such a heterogeneous 5G structure, a large number of communication scenarios should be fully supported, including special ones involving high mobility (such as vehicular and high speed train communications and networks).

Towards the heterogeneous 5G, the first and foremost hurdle lies in the channel measurement and modeling in the broad and diversified 5G scenarios. This special issue is dedicated to providing a platform to share and present the latest views and developments on 5G channel measurement and modeling issues.

This special issue includes five technical contributions from leading researchers in channel measurements and modeling. The first paper entitled "An Overview of Non-Stationary Property for Massive MIMO Channel Modeling" by ZHANG, CHEN, and TANG presents an overview of methods of modeling non-stationary properties on both the array and time axes, which are mainly divided into two major categories: birth-death (BD) process and cluster visibility region (VR) method. The main concepts and theories are described, together with useful implementation guidelines. In conclusion, a comparison between these two methods is made. The second paper is entitled "Measurement-Based Channel Characterization for 5G Wireless Communications on Campus Scenario" by YANG, HE, AI, XIONG, DONG, LI, WANG, FAN, and QIN. It investigates the radio channels of 5G communications below 6 GHz according to the requirements and scenarios of 5G communications. Channel measurements were conducted on campus of Beijing Jiaotong University, China at two key optional frequency bands below 6 GHz. By using the measured data, the key channel parameters at 460 MHz and 3.5 GHz are analyzed, such as power delay profile, path loss exponent, shadow fading, and delay spread. The results are helpful for the 5G communication system design. The third paper, co-authored by ZHANG, WANG, WU, and ZHANG, is entitled "A Survey of Massive MIMO Channel Measurements and Models". In this paper, the channel measurements and models of massive MIMO in recent years are summarized globally. Besides, their

Guest Editorial
CUI Shuguang and CHENG Xiang

work on related 256 antenna elements with 200 MHz bandwidth at 3.5 GHz, the verification of rationality of measurement method, and the spatial evolution of clusters in mobile scenario are provided. The next paper, co-authored by WANG, GENG, ZHAO, HONG, and Haneda, is entitled "Feasibility Study of 60 GHz UWB System for Gigabit M2M Communications". In this paper, the feasibility and performance of mm-wave 60 GHz ultra-wide band (UWB) systems for gigabit machine-to-machine (M2M) communications are analyzed. Specifically, based on specifications and channel measurements and models for both line-of-sight LOS and non-LOS (NLOS) scenarios, 60 GHz propagation mechanisms are summarized and 60 GHz UWB link budget and performance are analyzed. The goal of this work is to provide useful information for standardizations and design of 60 GHz UWB systems. The last (but not least) paper "Measurement-Based Spatial-Consistent Channel Modeling Involving Clusters of Scatterers" is co-authored by YIN, ZHANG, WANG, and CHENG. In this paper, the conventional method of establishing spatial channel models (SCMs) based on measurements is extended by including clusters-of-scatterers (CoSs) that exist along propagation paths. Channel models resulted by utilizing this new method are applicable for generating channel realizations of reasonable spatial consistency which is required for designing techniques and systems of the 5G wireless communications.

We would like to thank all the authors for choosing this special issue to publish their new research results and all the reviewers for their meticulous review comments and suggestions that help to improve the technical quality and presentation of this special issue. We hope that our readers will enjoy reading the articles and find this special issue helpful to their own research work.

Call for Papers

ZTE Communications Special Issue on

Cloud Computing, Fog Computing, and Dew Computing

Cloud Computing has drastically changed the landscape of IT industry by providing some major benefits to IT customers: eliminating upfront IT investment, scalability, proportional costs, and so on. However, the delay-sensitive applications face the problem of large latency, especially when several smart devices are getting involved. Therefore, cloud computing is unable to meet the requirements of low latency, location awareness, and mobility support.

To overcome this problem, Cisco has first introduced a trusted and dependable solution through the Fog Computing to put the services and resources of the cloud closer to users, which facilitates the leveraging of available services and resources in the edge networks. Fog Computing is a scenario where a huge number of heterogeneous ubiquitous and decentralized devices communicate and potentially cooperate among them and with the network to perform storage and processing tasks without the intervention of third-parties.

Dew Computing is an on-premises computer software-hardware organization paradigm in the cloud computing environment where the on-premises computer provides functionality that is independent of cloud services and is also collaborative with cloud services. The goal of Dew Computing is to fully realize the potentials of on-premises computers and cloud services.

The definition and features of Cloud Computing, Fog Computing, and Dew Computing, the relationships among them, and their applications are still under heated discussion and are the focus of this special Issue.

This special issue intends to gather the preeminent minds of the field to discuss and envision the future development of cloud-related technologies, which would be of interest for researchers, entrepreneurs, and customers. This special issue seeks original, high quality and unpublished research contributions as well as implementation experiences from researchers in the related areas.

Guest Editors

- Dr. PAN Yi (yipan@gsu.edu), Georgia State University, USA
- Dr. LUO Guangchun (gcluo@uestc.edu.cn), University of Electronic Science and Technology of China, China

Important Dates

- First Submission Due: May 15, 2017
- Review and Final Decision Due: July 1, 2017
- Final Manuscript Due: July 25, 2017
- Publication Date: October 25, 2017

Preparation & Submission

For detailed information on the submission and preparation of manuscripts, please visit ZTE Communications Author Guidelines (http://wwwen.zte.com.cn/endata/magazine/ztecommunications/index_5104.html).

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 001
http://www.cnki.net/kcms/detail/34.1294.TN.20170215.0926.002.html, published online February 15, 2017

An Overview of Non-Stationary Property for Massive MIMO Channel Modeling

ZHANG Ping, CHEN Jianqiao, and TANG Tian
(Department of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract

Massive multiple-input multiple-output (MIMO) emerges as one of the most promising technologies for 5G mobile communication systems. Compared to the conventional MIMO channel models, channel researches and measurements show that significant non-stationary properties rise in massive MIMO channels. Therefore, an accurate channel model is indispensable for the sake of massive MIMO system design and performance evaluation. This article presents an overview of methods of modeling non-stationary properties on both the array and time axes, which are mainly divided into two major categories: birth-death (BD) process and cluster visibility region (VR) method. The main concepts and theories are described, together with useful implementation guidelines. In conclusion, a comparison between these two methods is made.

Keywords

birth-death process; cluster visibility region method; massive MIMO channel model; non-stationary properties

1 Introduction

Massive multiple-input multiple-output (MIMO) systems, which are equipped with tens or even hundreds of antennas, have been proposed to meet the increasing traffic demand of the future wireless communication systems. Compared to conventional MIMO systems, massive MIMO systems are able to substantially improve capacity, link reliability, energy efficiency and spectral efficiency [1]–[4]. For the sake of MIMO system design and performance evaluation, it is indispensable to develop accurate and efficient small-scale fading channel models. Generally, there are three generic approaches for conventional MIMO channel modeling, namely the stochastic model, deterministic-based model and geometry-based stochastic model (GBSM). The GBSM has been widely used as an efficient method to simulate the wireless propagation channels, because it generates the location of scatterers according to a certain probability distribution rather than specific channel environment. So far, many standardized GBSMs are proposed [5]–[7]. However, according to the channel measurements [8], [9], new characteristics of the massive systems make the conventional MIMO channel models not be effective to be applied in massive MIMO channels directly.

Non-stationary property is one of the new characteristics that makes massive MIMO channels different from conventional ones. Non-stationary property generally includes two aspects: angle of arrival (AOA) shifts at the receiver side and dynamic properties of clusters on both the array and time axes. AOA shifts can be described completely through spherical wave-front assumption according to geometrical relationships among transmitter, receiver and clusters [10]. This article does not focus on this aspect. Dynamic properties of clusters on both the array and time axes generally refer to the facts that each antenna at different physical location has different cluster sets and the location of cluster is changing over time. This property is mainly described through birth-death (BD) process and cluster visibility region (VR) method.

The BD process was first investigated to model the non-stationary properties of clusters on the time axis only [11]–[14]. After this property was observed on the large antenna array [8], [9], i.e., appearance and disappearance of clusters can occur on the array axis, the BD process was extended to the antenna array [15]–[17]. The authors in [15] proposed a 2D ellipse model that describes wide-band massive MIMO channels by many con-focal ellipses. Similarly, a 2D multi-ring channel model with the same center for massive MIMO was proposed in [16]. For both of them, the geometrical relationships related to the delays of multi-path components (MPCs) are assumed to mimic the distribution of clusters. In [17], a 3D twin-cluster model

This work was supported in part by the National Natural Science of Foundation for Creative Research Groups of China under Grant No. 61421061 and Huawei Innovation Research Program.

was proposed, where a cluster was divided into two representations of itself (one at the transmitter and the other at the receiver). Then the complicated scattering environment was abstracted by a virtual link between the two representations of this cluster. Although the descriptions of geometrical relationships are different in these modeling approaches, the methods of modeling non-stationary properties of clusters on both the array and time axes are all based on the BD process.

The cluster VR method models the non-stationary properties of clusters from the cluster perspective. VR is typically assigned to the cluster in such a way that when an terminal enters a VR, the cluster assigned to this VR is active (contributes to the impulse response). The cluster VR method was investigated on the time axis at first [18], and after the non-stationary properties of clusters on the large antenna array in massive MIMO was observed, it was extended [19]–[21]. In [19], a 3D two-cylinder regular-shaped GBSM for non-isotropic scattering massive MIMO channels was proposed. Non-stationary properties of clusters on the array axis were described by using a virtual sphere. In [20] and [21], the concept of VR in COST 2100 channel model was extended. In [20], the cluster VR was divided into two categories: the observed VR and the true VR. The cumulative distribution function (CDF) of the observed VR size could be written as a function of the probability density function (PDF) of the true VR size. On the other hand, the distribution of the true VR lengths could be obtained based on measurements. In [21], cluster VR on the array axis was characterized by the partially visible clusters and wholly visible clusters. Therefore, when using the cluster VR method to describe non-stationary properties of clusters, the definition of VR is most important.

This article aims to give a concise overview of methods of modeling non-stationary properties of clusters for massive MIMO, which mainly include two categories: BD process and cluster VR method. For the BD process, because of uniform modeling approach, a main flow chart that describes non-stationary properties of clusters on both the array and time axes is given. For the cluster VR method, some typical representatives of defining VR in different modeling approach are reviewed. A brief comparison of these two different methods is also given.

2 Birth-Death Process

The BD process is widely investigated to describe non-stationary properties of clusters on both the array and time axes in massive MIMO channel model. In this process, cluster evolution along array and time axes are calculated separately. The main flow of BD process is presented below.

Let N_{total} be the total number of clusters that are observable to at least one transmit antenna and one receive antenna. The value of N_{total} can be expressed as

$$N_{total}=\mathrm{card}\left(\bigcup_{l=1}^{M_T}\bigcup_{k=1}^{M_R}\left(C_l^T(t)\cap C_k^R(t)\right)\right), \tag{1}$$

where the operator $\mathrm{card}(\cdot)$ denotes the cardinality of a set, M_T and M_R denote the number of antenna elements at the transmitter and receiver, respectively, and $C_l^T(t)(C_k^R(t))$ denotes the cluster set in which clusters are observable to the l-th transmit antenna (the k-th receive antenna) at time instant t. Then, a cluster is observable to the l-th transmit antenna and the k-th receive antenna if and only if this cluster is in the set $\{C_l^T(t)\cap C_k^R(t)\}$. Cluster evolution on both the array and time axes determine cluster sets in $C_l^T(t)$ and $C_k^R(t)$.

Cluster evolution on the array axis is operated first. We assume the initial number of clusters N and the initial cluster sets of the 1-st transmit and receive antenna $C_1^T=\{c_x^T:x=1,2,...,N\}$ and $C_1^R=\{c_x^R:x=1,2,...,N\}$ at the initial time instant t are given, where c_x^T and c_x^R are two representations of Cluster$_x$, the subscript x represents the x-th cluster in cluster sets. Then, these clusters in cluster sets C_1^T and C_1^R evolve according to BD process on the array axis to recursively generate the cluster sets of the rest of antennas at the transmitter and receiver at the initial time instant t, which can be expressed as

$$\begin{aligned} &C_{l-1}^T(t)\xrightarrow{E}C_l^T(t) \quad (l=2,3,...,M_T)\\ &C_{k-1}^R(t)\xrightarrow{E}C_k^R(t) \quad (k=2,3,...,M_R)\end{aligned}, \tag{2}$$

where the operator $\xrightarrow{E}$ denotes cluster evolution on either the antenna array or time axes. The survival probabilities of the clusters inside the cluster sets on the array axis at the transmitter $P_{survival}^T$ and at the receiver $P_{survival}^R$ can be calculated as

$$\begin{aligned} P_{survival}^T&=e^{-\lambda_R\frac{\delta_T}{D_c^a}}\\ P_{survival}^R&=e^{-\lambda_R\frac{\delta_R}{D_c^a}}\end{aligned}, \tag{3}$$

where D_c^a is the scenario-dependent correlation factor on the array axis, λ_G (per meter) and λ_R (per meter) denote the cluster generation rate and the recombination rate respectively, and δ_T and δ_R are antenna spacing of transmit and antenna array respectively. According to the BD process, the average number of newly generated clusters N_{new}^T and N_{new}^R on the array axis can be calculated as

$$\begin{aligned} E[N_{new}^T]&=\frac{\lambda_G}{\lambda_R}\left(1-e^{-\frac{\delta_T}{D_c^a}}\right)\\ E[N_{new}^R]&=\frac{\lambda_G}{\lambda_R}\left(1-e^{-\frac{\delta_R}{D_c^a}}\right)\end{aligned}, \tag{4}$$

where $E[\cdot]$ denotes the expectation. Based on the BD process, clusters evolve gradually on the antenna array. To put it another way, some initial clusters may disappear for certain antennas, and some new clusters may appear. An example of cluster evolution on the receive antenna array is shown in **Fig. 1** [15]. It can be seen that these clusters evolve gradually from twenty clusters that are observable to the first antenna element originally. Four clusters of original twenty clusters disappear, name-

ly cluster_{10}, cluster_{14}, cluster_{15} and cluster_{20}. Meanwhile, five new clusters appear. As a result, different antenna elements may observe different cluster sets, like receive antennas C_4^R and C_{31}^R.

At the next time instant $t=t_{m+1}$, cluster evolution on the time axis is operated as

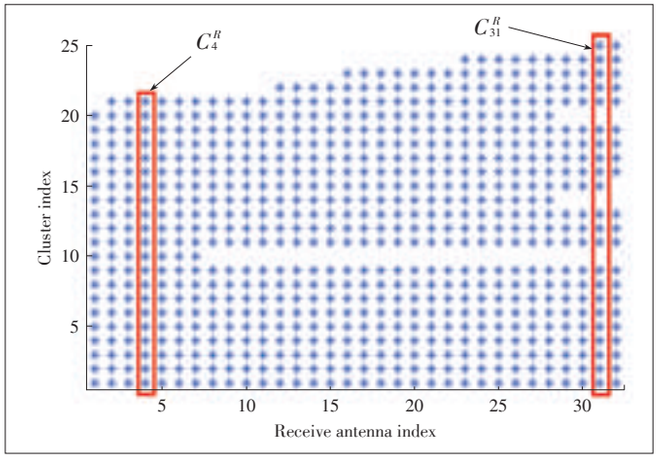
$$\begin{aligned} C_l^T(t_m) &\xrightarrow{E} C_l^T(t_{m+1}) \\ C_k^R(t_m) &\xrightarrow{E} C_k^R(t_{m+1}) \end{aligned}. \tag{5}$$

According to [22], to reproduce the evolution on the time axis, the channel fluctuation, which stands for a measure of variation of scattering environment in time domain, should be defined as q_{m+1} at $t=t_{m+1}$. q_{m+1} consists of two parts, namely the fluctuation caused by the movement of receiver and the fluctuation movement of clusters. Given the scenario-dependent correlation factor D_c^s, each cluster survives with probability $P_{survival}$ and mean number of newly generated clusters $E[N_{new}(t_{m+1})]$ can be presented as

$$P_{survival}(q_{m+1}) = e^{-\frac{\lambda_R q_{m+1}}{D_c^s}}, \tag{6}$$

$$E[N_{new}(t_{m+1})] = \frac{\lambda_G}{\lambda_R}\left(1 - e^{-\frac{\lambda_R q_{m+1}}{D_c^a}}\right). \tag{7}$$

After cluster evolution on the time axis, as (5)–(7) show, all clusters can be categorized as survived clusters or newly generated clusters. For the survived cluster, channel parameters should be updated from t_m to t_{m+1} according to the geometrical relationships. For the newly generated cluster, it is first attached to one transmit antenna element randomly, and then evolves along transmit antenna array to determine whether it can be observed based on the survival probabilities $P_{survival}^T$ on the array axis. The same judge process applies to the receive antenna array too. Finally, the cluster is added to antenna cluster sets if and only if it survives at both the transmitter and receiver side. The main flow of cluster evolution on both array and time axes based on the BD process is given in **Fig. 2**.



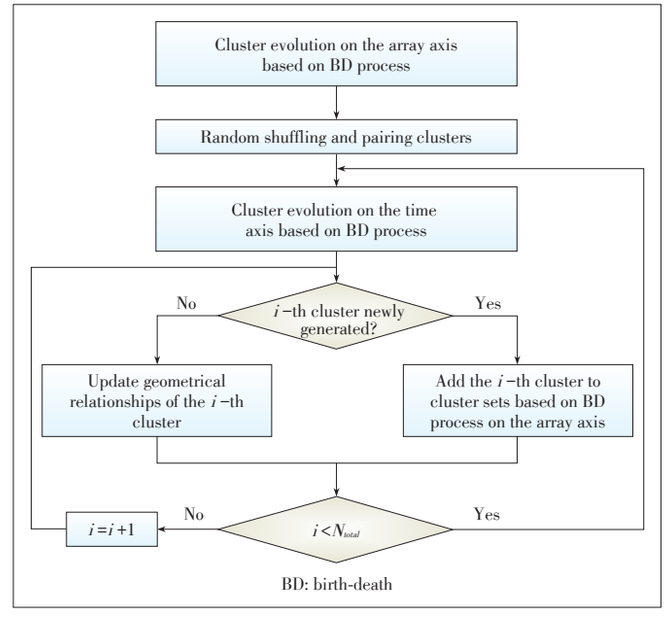
▲**Figure 1. An example of cluster evolution on the receive antenna array [15].**

3 Cluster Visibility Region Method

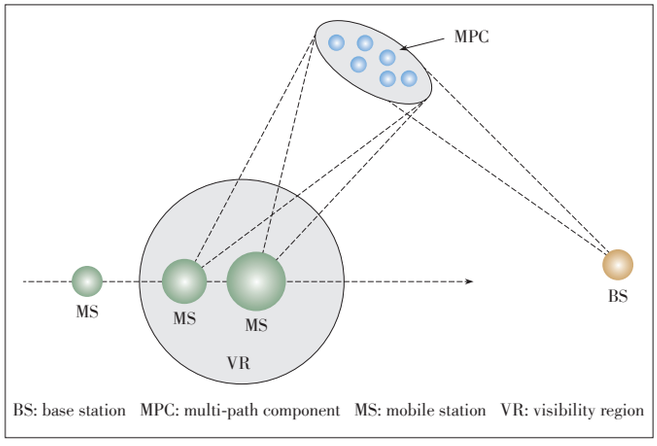
The cluster VR method describes non-stationary properties of clusters on both the array and time axes from the cluster perspective. In the exsiting channel models, different definitions of VR determine the methods of modeling cluster evolution. Therefore, the descriptions of VR in some representative channel models are reviewed separately.

3.1 COST 2100 Channel Model

In the COST2100 channel model, the cluster VR method is used to describe cluster evolution on the time axis. Specifically, clusters are divided into three categories: local clusters, single-bounce clusters and multiple-bounce clusters. VR confines the cluster activity within a limited geographical area. A cluster is depicted as an ellipsoid as viewed from mobile stations (MS) and from base stations (BS), and the size of the circle around the MS represents the visibility level of the cluster to the BS-MS channel, as illustrated in **Fig. 3**. After the MS moves into the VR, it receives signals scattered by the related cluster. As it moves towards the VR center, the cluster smoothly increases its visibility. The visibility is accounted mathematically by a VR gain, which grows form 0 to 1 upon entrance of the VR. By employing the VR size and the VR gain, cluster evolution at the receiver side can be described in the COST2100 channel model. Eventually, the channel impulse response (CIR) is obtained by the superposition of the MPCs from all active clusters determined by the position of the termi-



▲**Figure 2. The main flowchart of cluster evolution on both the array and time axes.**



▲Figure 3. VR concept.

nal. Additionally, parameterization and validation of the COST2100 channel model are given in [23] and [24] based on measurements.

3.2 3D Two-Cylinder Channel Model

The 3D two-cylinder channel model extends the concept of VR to the antenna array. All the scatterers are distributed on the surface of a cylinder as equivalent, and each antenna has its own visible area of scatterers by using a virtual sphere, where the radius is r and the center is the antenna element itself. In this case, scatterers on the surface of the cylinder can be seen by an antenna element only if the distance between them is less than r, as shown in **Fig. 4**. The radius r of the virtual sphere can be calculated according to the distance between two antenna elements and the correlation of their channel impulse responses. Furthermore, channel impulse responses are affected by the distribution of scatterers. If the distance is $\lambda/2$, correlation coefficient is 0.9, and scatterers are uniformly distributed on the surface of the cylinder, whose radius is R, the r can be expressed in terms of R as

$$r < R \times 2.7856 . \tag{8}$$

3.3 Channel Model Based on Extended Concepts of VR

The channel models proposed in [20] and [21] extend the concept of VR in COST2100 channel model to the antenna array. We take the extension in [20] as an example. In [20], VR size and visibility gain are discussed based on measurements. As shown in **Fig. 5**, the VR size of some clusters is entirely inside the array, while that of other clusters overlaps one or both ends of the antenna array. For the former case, the observed size of VR on the antenna array is the cluster's true VR size, while for the latter case, the true VR size may be much larger than what is observed on the antenna array.

As a result, VR size should be described in two stages. First, the relationship between the observed VR size and the true VR size is determined, as illustrated in **Fig. 6**. The CDF of the observed VR size $K_{\Delta}(y)$ can be written as a function of PDF of the true VR size $f_{\alpha}(\nu)$, and it can be expressed as

$$K_{\Delta}(y) = \begin{cases} K_{\Delta}'(y), & y \leqslant L \\ 1, & y > L \end{cases}, \tag{9}$$

$$K_{\Delta}'(y) = \int_{\Delta_0}^{y} f_{\alpha}(\nu)d\nu + 2y\int_{y}^{\infty} \frac{1}{L+\nu} f_{\alpha}(\nu)d\nu - 2\Delta_0 \int_{\Delta_0}^{\infty} \frac{1}{L+\nu} f_{\alpha}(\nu)d\nu, \tag{10}$$

where L is the length of the array, and Δ_0 is the smallest observation of the VR size on the array due to measurement data processing. Second, the particular distribution of the true VR

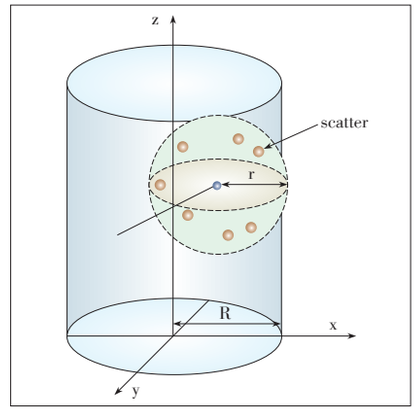


Figure 4. ▶ VR of the antenna array.

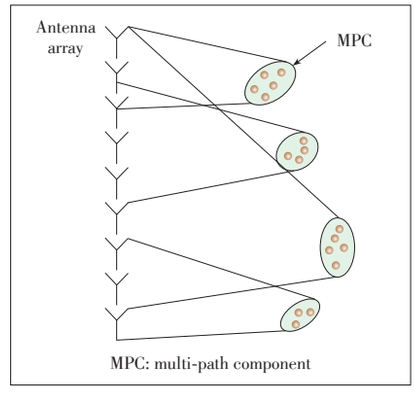
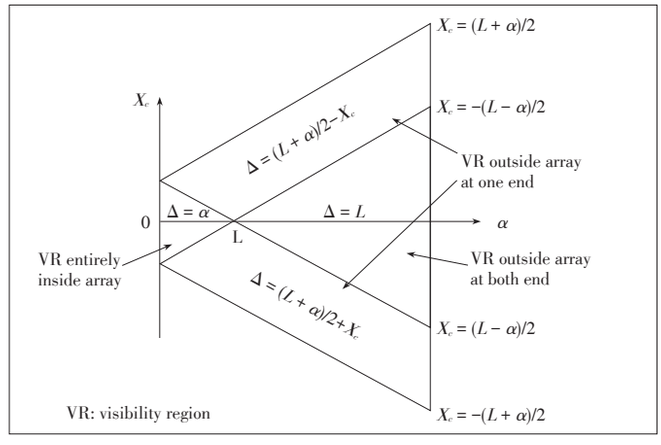


Figure 5. ▶ Extension of the concept of VR to the antenna array.



▲Figure 6. Relationship between a cluster's true VR size and observed VR size on the array Δ, depending on the array center position X_c.

size is calculated through the maximum likelihood estimate (MLE) approach based on the measurements. Finally, the VR size can be calculated.

4 Conclusions

In this article, we review methods of modeling non-stationary properties of clusters on both the array and time axes, namely BD process and cluster VR method. The main flow of BD process is given, and different definitions of VR in some representative channel models are reviewed. Based on mathematical derivation, the theory of BD process is more mature and complete, and its accuracy has been proved by lots of researches. However, it is too complex and lacks of intuitive geometric characteristics between clusters and the antenna array. On the other hand, the cluster VR method pays more attention to geometrical distribution of clusters, which makes it more natural to be integrated into the channel model based on stochastic geometry. However, further study is necessary to obtain the complete information of VR. From our point of view, a better scheme of modeling cluster evolution is to combine these two methods together. This scheme can reduce the complexity of modeling, and also reflect the geometrical characteristics of the channel.

References

[1] D. Gesbert, M. Shafi, D. Shiu, et al., "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281–302, Apr. 2003. doi: 10.1109/JSAC.2003.809458.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, et al., "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014. doi: 10.1109/MCOM.2014.6736761.

[3] F. Rusek, D. Persson, B. K. Lau, et al., "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2012. doi: 10.1109/MSP.2011.2178495.

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, et al., "An overview of massive MIMO benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014. doi: 10.1109/JSTSP.2014.2317671.

[5] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, 3GPP TR 25.996 version 12.0.0 Release 12, 2014.

[6] P. Kyösti, J. Meinilä, L. Hentilä, et al., "WINNER II channel models part II radio channel measurement and analysis results," IST-4-027756 WINNER II D1.1.2 V1.2, 2007.

[7] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, ITU M.2135-0, 2008.

[8] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. 6th European Conference on Antennas and Propagation*, Prague, Czech Republic, Mar. 2012, pp. 433–437.

[9] X. Gao, F. Tufvesson, O. Edfors, et al., "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. 46th Annual Asilomar Conference on Signals, Systems, and Computers*, California, USA, Nov. 2012, pp. 295–299.

[10] F. Bøhagen, P. Orten, and G. E. Øien, "On spherical vs. plane wave modeling of line-of-sight MIMO channels," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 841–849, Mar. 2009.

[11] T. Zwick, C. Fischer, D. Didascalou, et al., "A stochastic spatial channel model based on wave-propagation modeling," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 1, pp. 6–15, Jan. 2000.

[12] T. Zwick, C. Fischer, and W. Wiesbeck, "A stochastic multipath channel model including path directions for indoor environments," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1178–1192, Aug. 2002.

[13] C. C. Chong, C. M. Tan, D. I. Laurenson, et al., "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, Jul. 2005. doi: 10.1109/TWC.2005.850341.

[14] F. Babich and G. Lombardi, "A Markov model for the mobile propagation channel," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 63–73, Jan. 2000.

[15] S. B. Wu, C. X. Wang, H. Haas, et al., "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1434–1446, 2015.

[16] H. L. Wu, S. Jin, and X. Q. Gao, "Non-stationary multi-ring channel model for massive MIMO systems," in *International Conference on Wireless Communications & Signal Processing (WCSP)*, Nanjing, China, 2015, pp.1–6.

[17] S. B. Wu, C. X. Wang, H. Haas, et al., "A non-stationary 3D wideband twin-cluster model for 5G massive MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1207–1218, 2014.

[18] L. F. Liu, C. Oestges, J. Poutanen, et al., "The COST 2100 MIMO channel model," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, Dec. 2012. doi: 10.1109/MWC.2012.6393523.

[19] Y. Xie, B. Li, X. Y. Zuo, et al., "A 3D geometry-based stochastic model for 5G massive MIMO channels," in *11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE)*, Taiwan, China, 2015, pp. 216–222.

[20] X. Gao, F. Tufvesson, and O. Edfors, "Massive MIMO channels—measurements and models," in *2013 Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, Nov. 2013, pp. 280–284. doi: 10.1109/ACSSC.2013.6810277.

[21] X. R. Li, S. D. Zhou, E. Bjornson, et al., "Capacity analysis for spatially non-wide sense stationary uplink massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 7044–7056, 2015.

[22] T. Zwick, C. Fischer, D. Didascalou, and W. Wiesbeck, "A stochastic spatial channel model based on wave-propagation modeling," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 1, pp. 6–15, Jan. 2000.

[23] M. F. Zhu, G. Eriksson, and F. Tufvesson, "The COST 2100 channel model: parameterization and validation based on outdoor MIMO measurements at 300 MHz," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 888–897, 2013. doi: 10.1109/TWC.2013.010413.120620.

[24] J. Poutanen, K. Haneda, L. F. Liu, et al., "Parameterization of the COST 2100 MIMO channel model in indoor scenarios," in *Proc. 5th European Conference on Antennas and Propagation (EUCAP)*, Rome, Italy, 2011, pp. 3606–3610.

Manuscript received: 2016-11-11

Biographies

ZHANG Ping (pzhang@bupt.edu.cn) got his Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), China in 1990, and is currently a professor and doctorial tutor of BUPT, a member of the BUPT academic board, the director of State Key Lab of Networking and Switching Technology, a member of IMT-2020(5G) Group, and a consultancy expert of National Science Foundation Committee (NSFC). His personal research interest is mainly wireless network communications. Prof. ZHANG has published more than 400 papers in journals such as *IEEE Communications Magazine*, *IEEE Electronics Letters*, and *Chinese Journal of Electronics* and eight books at home and abroad. He holds about 161 patents and was awarded National Science and Technology Progress Prize 5 times.

CHEN Jianqiao (jqchen1988@163.com) received his M.S. degree from the Institute of Communications Engineering, Xidian University, China in 2014, and is currently pursuing the Ph.D. degree with the Department of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China. His research interests are in the area of channel modeling and wireless communications.

TANG Tian (tangtian@bupt.edu.cn) is a lecturer in Key Laboratory of Universal Wireless Communications (Ministry of Education), Beijing University of Posts and Telecommunications (BUPT), China. He received his Ph.D. degree from BUPT in 2004. He has long been engaged in the research and development of mobile communications, and participated in many national projects such as "863", "973", and NSFC. He has published more than ten academic papers. He was awarded the First Prize of China Institute of Communications in 2009 and Second Prize of National Technology Innovation in 2010. His research interests include wireless communication theory and technology of physical layer.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 002
http://kns.cnki.net/kcms/detail/34.1294.TN.20170227.1111.002.html, published online February 27, 2017

Measurement-Based Channel Characterization for 5G Wireless Communications on Campus Scenario

YANG Mi [1], HE Ruisi [1], AI Bo [1], XIONG Lei [1], DONG Honghui[1], LI Jianzhi[1], WANG Wei [2], FAN Wei [3], and QIN Hongfeng [4]

(1. Beijing Jiaotong University, Beijing 100044, China;
2. German Aerospace Center (DLR), Münchner Strasse 2082234 Wessling, Germany;
3. Aalborg University, Aalborg 9220, Denmark;
4. Xi'an R&D Center, ZTE Corporation, Xi'an 710114, China)

Abstract

The fifth generation (5G) communication has been a hotspot of research in recent years, and both research institutions and industrial enterprises put a lot of interests in 5G communications at some new frequency bands. In this paper, we investigate the radio channels of 5G systems below 6 GHz according to the 5G communication requirements and scenarios. Channel measurements were conducted on the campus of Beijing Jiaotong University, China at two key optional frequency bands below 6 GHz. By using the measured data, we analyzed key channel parameters at 460 MHz and 3.5 GHz, such as power delay profile, path loss exponent, shadow fading, and delay spread. The results are helpful for the 5G communication system design.

Keywords

channel measurement; 5G; channel characterization

1 Introduction

In the last decade, public networks have been evolving from voice-centric second-generation systems, e. g., Global System for Mobile Communications (GSM) with limited capabilities, to fourth-generation (4G) broadband systems that offer higher data rates, e.g., long-term evolution (LTE) [1]. In recent years, with the rapid development of data services, the fifth generation (5G) communication has attracted high attention both from research institutions and industrial enterprises. According to the IMT-2020 [2], in some key competencies, 5G needs to support 0.1-1 Gb/s rate, 106 devices/km^2 connection density, and below 1 ms end-to-end latency [3].

ITU has suggested a bandwidth for 5G communication systems up to 1490-1810 MHz. The bandwidth of the current plan, however, is only 687 MHZ, which is obviously insufficient. Facing the shortage of spectrum resources shortage, we can use a higher frequency band, or consider other frequency bands below 6 GHz to use the spectrum more efficiently. Since the low frequency band supports a larger propagation distance, it can effectively reduce the number of base stations and decrease the transmission power to save energy.

In the World Radio Communication Conference (WRC) in 2015, eight new frequency bands for International Mobile Telecommunication (IMT) was added in the proposal AI1.1, which are all below 6 GHz, including 470-698 MHz, 1427-1518 MHz, 3300-3400 MHz, 3400-3600 MHz, 3600-3700 MHz, 4800-4990 MHz, etc. At the same time, China also introduced candidate frequency bands to the international standard organizations, and mostly of them are below 6 GHz, e.g., 3.3-3.6 GHz, 4.4-4.5 GHz and 4.8-4.99 GHz.

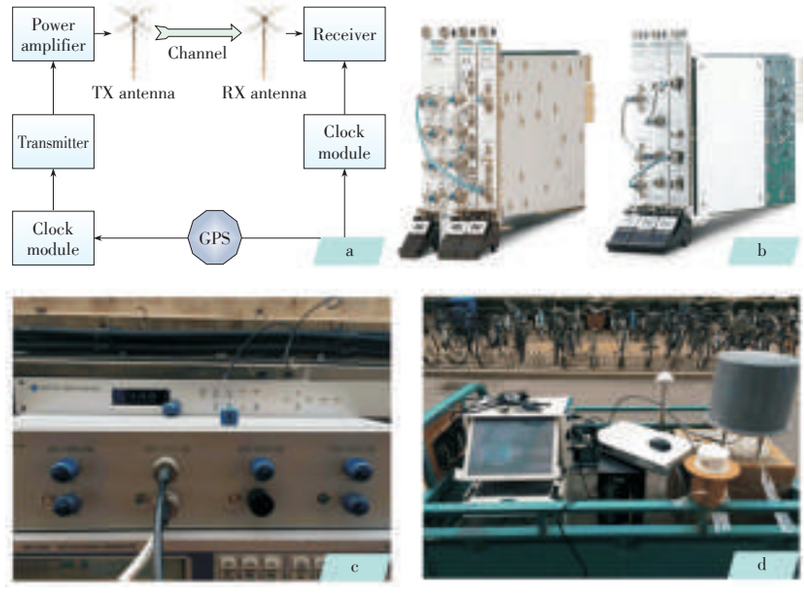
If we want to use below 6 GHz frequency bands in 5G communication systems, there are mainly two methods. One is to reuse the existing spectrums, and the other one is to use the new spectrums suggested in WRC 2015. The existing spectrums that can be reused include 800 MHz, 900 MHz, 1.8 GHz and 2.1 GHz. Among the new spectrums, the 3400-3600 MHz band has been considered as the 5G test frequency band, and it is also expected to be the first frequency band for 5G commu-

This work was supported by the National Natural Science Foundation of China under Grant 61501020, the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2016ZJ005, the China Postdoctoral Science Foundation under Grant 2016M591355, the Fundamental Research Funds for the Central Universities (No. 2016JBZ006), the Special Project of Cultivation and Development of Science and Technology Innovation Base in 2015, the National Natural Science Foundation of China under Grant U1334202, the Natural Science Base Research Plan in Shanxi Province of China under Grant 2015JM6320, and the Key Project from Beijing science and Technology Commission under Grant D151100000115004.

nications in China.

We have done related measurements and achieved some results on campus scenarios at 3.5 GHz [4]. In this paper, we carry out more measurements and analyses, and some rich and more meaningful results are obtained. We also present a channel measurement campaign for campus scenarios performed at two frequency bands (i.e., 460 MHz and 3.5 GHz). Furthermore, based on analysis of the measurement data, we present results on key channel parameters in terms of power delay profile, path loss exponent, shadow fading, and delay spread. The results can be used in the 5G communication system design.

The remainder of the paper is organized as follows. Section 2 describes the measurement system and measurement environment. Section 3 presents the measurement results of channel characterizations. Conclusions are drawn in Section 4.



GPS: Global Positioning System

▲**Figure 1. Our measurement system: (a) System architecture; (b) transmitter and receiver; (c) the clock module and power amplifier; and (d) measurement setup.**

2 Measurement Campaign

We describe our measurement campaign in the light of calibration, measurement system and measurement environment.

2.1 Calibration

According to [5], the received signal can be described in the frequency domain as

$$Y(d,f)=X(f)H_{TX}(f)H(d,f)H_{RX}(f), \quad (1)$$

where $X(f)$ is the transmitted signal, $Y(d, f)$ is the received signal, $H(d, f)$ is the transfer function of the radio channel, and $H_{TX}(f)$ and $H_{RX}(f)$ describe the front end effects (e.g., cables, power amplifiers) of the transmitter and receiver, respectively. In order to eliminate the influence of $H_{TX}(f)$ and $H_{RX}(f)$, reference measurements are necessary. The received reference signals in reference measurements can be expressed as

$$Y_{ref}(f)=X(f)H_{TX}(f)H_{ref}(f)H_{RX}(f). \quad (2)$$

Then the transfer function of the channel can be described as

$$H(d,f)=\frac{Y(d,f)}{Y_{ref}(f)}H_{ref}(f), \quad (3)$$

where $H_{ref}(f)$ is the attenuators used in reference measurements. Only the attenuator is used between Tx and Rx when the back-to-back measurement is assumed in reference measurement.

2.2 Measurement System

The measurement system is depicted in **Fig. 1**. Fig. 1a shows the measurement system architecture, including the transmitter, receiver, clock modules, power amplifier and antennas. Fig. 1b shows the transmitter and receiver, which are the core parts of the measurement system. They are based on National Instruments (NI) software radio equipments. The NI PXIe-5673E is a wide-bandwidth RF vector signal generator (VSG), which is used as the transmitter. On the other hand, the NI PXIe-5663E is a RF vector signal analyzer (VSA) with wide instantaneous bandwidth, which is used as the receiver. The transmitter and receiver support 85 MHz to 6.6 GHz frequency bands and more than 50 MHz instantaneous bandwidth, which meets our measurement requirements. An amplifier (Fig. 1c) is used to provide the 40 dBm maximum transmitted power. Two pairs of omnidirectional antennas (460 MHz and 3.5 GHz) are used in the measurements. Besides, two clock modules locked with the GPS provide synchronization between the transmitter and the receiver.

2.3 Measurement Environment

Main measurement parameters are shown in **Table 1**. The carrier frequencies are 460 MHz and 3.5 GHz, and the bandwidth is 30 MHz. **Fig. 2** shows the measurement environment and route. The measurements were conducted on the campus of Beijing Jiaotong University, China. The transmitter antenna is placed on the roof of the Siyuan Building with a height of about 60 m, and the receiver antenna is placed at a trolley with a height of 1.5 m. In Fig. 2a and Fig. 2b, the red line shows the line-of-sight (LOS) scenario and the blue line shows the non-LOS (NLOS) scenario. For the NLOS region, the LOS paths are mainly blocked by the buildings. Fig. 2b shows the measurement route seen from the transmitter location. The receiver's

▼Table 1. Main measurement parameters

Frequency	460 MHz and 3.5 GHz
Bandwidth	30 MHz
Maximum transmitted power	40 dBm
Scenario	Campus, LOS and NLOS
Height of transmitter antenna	60 m
Height of receiver antenna	1.5 m
Receiver speed	Low speed (1.2 m/s)

LOS: line-of-sight NLOS: non-LOS

moving speed is about 1.2 m/s, the length of the whole route is about 450 m, the nearest distance of the receiver and transmitter is 90 m, and their farthest distance (Fig. 2c) is 206 m.

3 Results

3.1 Power Delay Profile

Random and complicated radio-propagation channels can be characterized using the impulse-response approach [6], [7]. The power delay profile (PDP) describes the power profile at a certain delay interval [8], and shows how much power the receiver received with a certain delay interval. It has been widely used to describe the distribution of multi-path components (MPCs) in measured environments. The instantaneous PDP is denoted as

$$P(t,\tau)=|h(t,\tau)|^2, \tag{4}$$

where $h(\tau)$ is the measured channel impulse response at time t with delay τ. In order to get more accurate analysis results, elimination of the noise in the received signal is necessary. We capture part of the received signal to calculate the average power of the noise, and then set the noise threshold by adding 6 dB to the noise power. Only the signals larger than the noise threshold are considered to be valid, and the samples below the threshold are set to 0. **Fig. 3** shows the average PDPs (APDPs) that were averaged by using a sliding window with a length corresponding to the receiver traveled distance of 20 wavelengths.

Fig. 3a shows the APDPs at 460 MHz, while Fig. 3b shows the APDPs at 3.5 GHz. We can see that there are clear LOS components and a few scattering components in most locations. Because we have the same velocity of trolley and route for the 460 MHz and 3.5 GHz measurements, both of the two APDPs have similar shapes and change trends. **Fig. 4** shows the signal-to-noise ratio (SNR) calculated from the received signals through the whole route at 460 MHz (Fig. 4a) and 3.5 GHz (Fig. 4b). It is obviously that the SNR at 460 MHz is larger than that at 3.5 GHz with nearly 10 dB. We excluded the measured data whose SNR is too low in order to get more accurate results for the analysis of channel parameters.

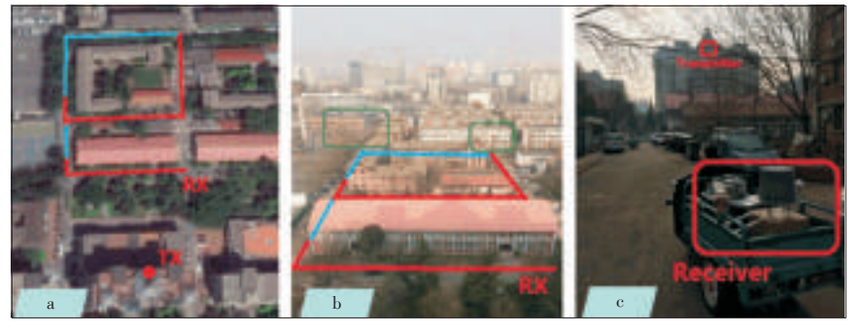
The two buildings in green circles in Fig. 2b are considered to be the reflectors which lead to the two multi-path components in Fig. 3a. The left building results in the multi-path component between 80 s to 170 s, and the right building leads to another multipath component (between 230 s to 300 s). At the same time, the multi-path components are more blurred at 460 MHz. The reason is the difference of the free-space transfer loss between two frequency bands. The lower frequency band (460 MHz) has larger receive power and more scattering components. In addition, there are some weak power areas in the middle of graphics (between 170 s to 230 s), they are mainly caused by the buildings and longer propagation distance in the measurement run.

3.2 Path Loss

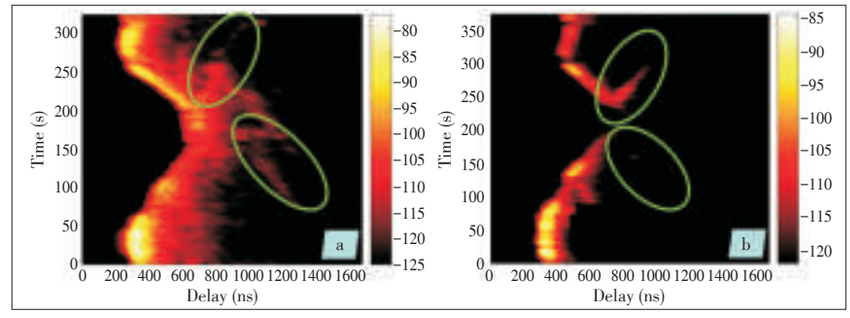
According to [9], path loss (PL) is a measure of the average RF attenuation to the transmitted signal when it arrives at the receiver. It is defined by

$$PL(dB)=10\log\frac{P_t}{P_y}, \tag{5}$$

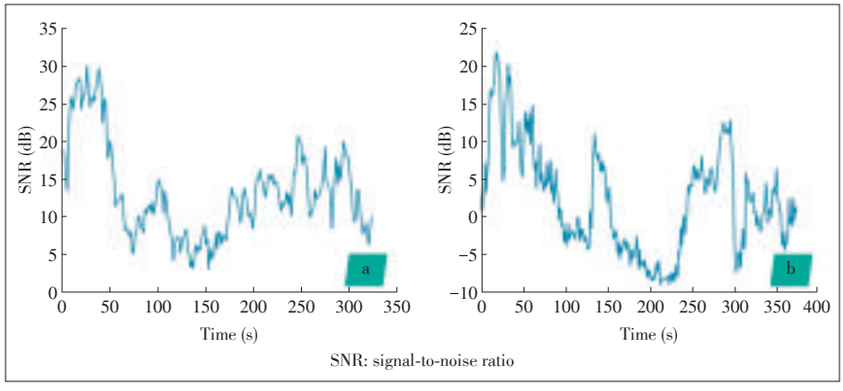
where P_t and P_y are the transmitted and received power, respec-



▲Figure 2. The measurement Environment: (a) Top view of the measurement route; (b) the measurement route seen from the transmitter location; and (c) the farthest distance between the receiver and transmitter.



▲Figure 3. Power delay profile: (a) APDP at 460MHz and (b) APDP at 3.5GHz. Green rings show the obvious scattering components.



▲Figure 4. SNR for the whole route at (a) 460 MHz and (b) 3.5 GHz. Two measurements at 460 MHz and 3.5 GHz have the same path, but the x-axis scaling are not identical due to the slight difference of the movement speeds.

tively. A general PL model uses γ to denote the relationship between the separation distance and the received power. So the path loss adopted in this paper follows [10] and [11]:

$$PL(dB) = PL(d_0) + 10\gamma \log(\frac{d}{d_0}) + X_\delta, \quad (6)$$

where γ is the path loss exponent and $PL(d_0)$ is the intercept value of the path loss model at the reference distance d_0 [12]. X_δ is a zero-mean Gaussian distributed random variable describing the random shadowing [13]. $\gamma = 2$ in free space. However, γ is generally higher for a realistic channel.

In this paper, we use the first path in PDP to determine the propagation distance between the transmitter and receiver. Here we should note the error of distance. The bandwidth is 30 MHz, resulting in a delay resolution of 33.33 ns corresponding to a distance 10 m. Because the true LOS path is located between two samples, there are less than 10-meter distance estimation error. Then, we transform the measured path loss from the time index to distance index.

Fig. 5 describes the scatter plot of path loss versus log-distance for 460 MHz, together with linear regression fit curve, and **Fig. 6** shows the corresponding results for 3.5 GHz.

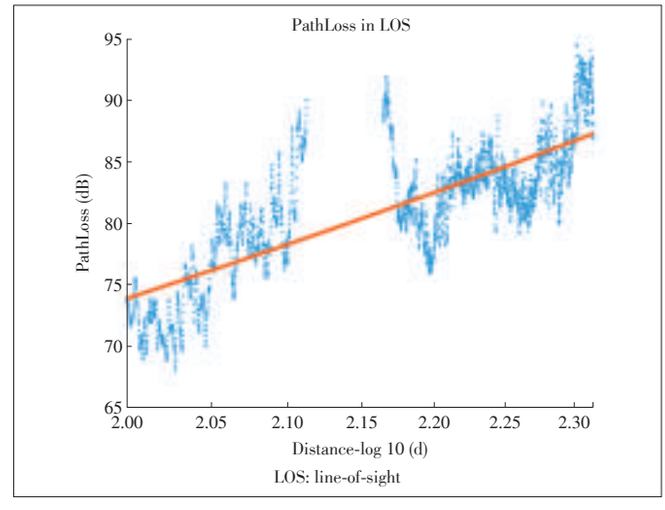
Because the amount of measurement data in the NLOS scenario is less and the maximum and minimum distance difference is too small to obtain accurate linear regression results, we only consider the LOS scenario. Based on the measurements, the γ and $PL(d_0)$ are shown in **Table 2**. It is found that γ =4.23 and $PL(d_0)$=−10.5 at 460 MHz, while γ =6.16 and $PL(d_0)$=−43.5 at 3.5 GHz. According to [14]–[16], γ should be between 2 to 5 in typical urban environments. A large value of γ at 3.5 GHz may be caused by the high frequency band and the difference between campus and urban.

3.3 Shadow Fading

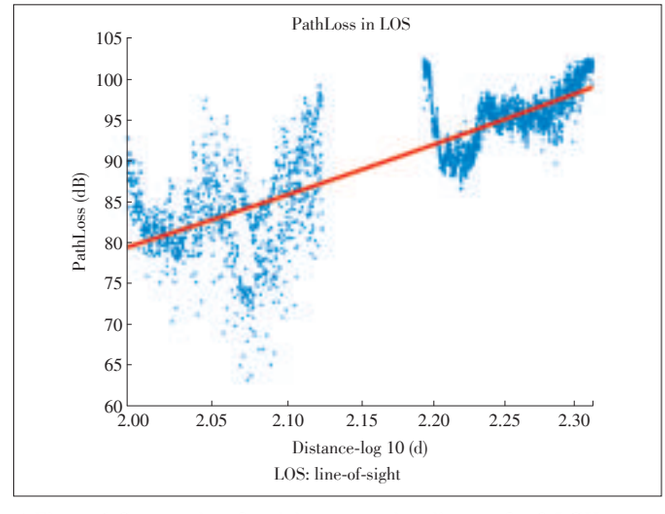
According to [17], after removing the distance-dependence from the received power, we obtain X_δ, which is the shadow fading component. Shadow fading in the dB scale can be modeled as a zero-mean Gaussian process with a standard deviation of σ [18]. **Fig. 7** shows the probability density function (PDF) of the measured shadow fading components, together with the Gaussian distribution fit. We can found that σ =3.304 dB at 460 MHz and σ = 4.208 dB at 3.5 GHz in the LOS scenario. It is noted that the model parameters above are limited by our measurement configurations.

3.4 Delay Spread

Root-mean-square (RMS) delay spread is the square root of the second central moment of a power-delay profile and is widely used to characterize the delay dispersion/frequency selectivity of the channel. It is the standard devi-



▲Figure 5. Scatter plot of path loss versus log-distance for 460 MHz, together with linear regression fit curve.

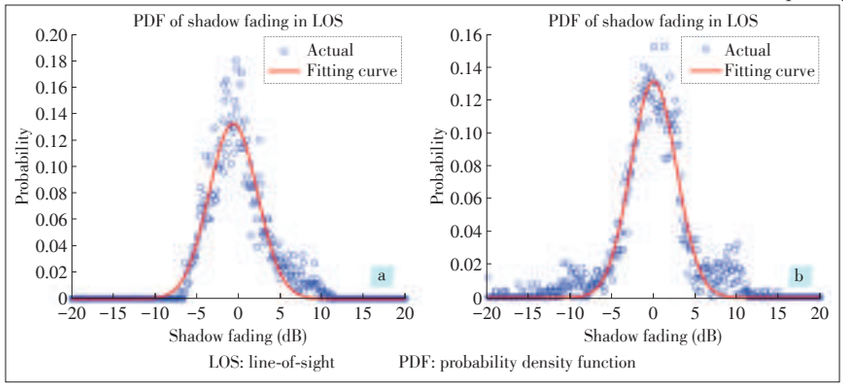


▲Figure 6. Scatter plot of path loss versus log-distance for 3.5 GHz, together with linear regression fit curve.

▼Table 2. γ and $PL(d_0)$

460 MHz γ in LOS	4.23
460 MHz $PL(d_0)$ in LOS	-10.5
3.5 GHz γ in LOS	6.16
3.5 GHz $PL(d_0)$ in LOS	-43.5

LOS: line-of-sight PL: path loss



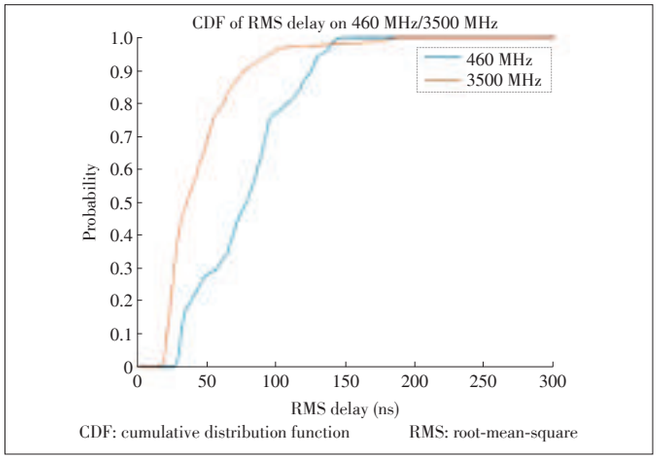
▲Figure 7. PDF plot of the measured shadow fading components, together with the Gaussian distribution fit: (a) 460MHz-LOS; (b) 3.5GHz-LOS.

ation about the mean excess delay [19] and defined as

$$\tau_{rms}(d)=\sqrt{\frac{\sum_p APDP(d,\tau_p)\tau_p^{\,2}}{\sum_p APDP(d,\tau_p)}-\left(\frac{\sum_p APDP(d,\tau_p)\tau_p}{\sum_p APDP(d,\tau_p)}\right)^2}, \quad (7)$$

where τ_p represents the delay and $APDP(d, \tau_p)$ describes the corresponding delay power of the pth path measured at the location d. The RMS delay spread is a good measure of the multipath spread. It is also used to give an estimate of the maximum data rate for transmission.

Fig. 8 shows the cumulative distribution function (CDF) of the estimated RMS delay spread for both LOS and NLOS scenarios. We present RMS delay spread for two scenarios on one CDF curve, so that we can compare the differences between two frequency bands for the entire path comprehensively. It is found that there is a mean value of 84.5 ns at 460 MHz band and 35.5 ns at 3.5 GHz band. The measurement at 3.5 GHz has a lower delay spread than at 460 MHz, the reason is the low frequency band has a lower propagation loss and better propagation characteristics. Therefore, the lower frequency band has a higher SNR for the same measurement route, at the same time, can capture rich multi-path components. In the NLOS scenario, the measured RMS delay spread at some locations is larger than 200 ns, which is far higher than the LOS scenario. On the other hand, in the LOS scenario without obvious multi-path components, the measured RMS delay spread has its minimum value (about 20–40 ns). Because of some obvious multipath components (highlighted in Fig. 3), there is a larger measured RMS delay spread compared with, which is consistent with many previous measurements.



▲Figure 8. CDF plot of the estimated RMS delay spread on 460 MHz and 3500 MHz.

4 Conclusions

In this paper, measurements-based channel characterizations are presented for campus scenarios at 460 MHz and 3.5 GHz carrier frequencies, with a bandwidth of 30 MHz. Using the measured data, we analyze key channel parameters, such as power delay profile, path loss exponent, shadow fading, and delay spread. A path loss exponent is found to be 4.23 for 460 MHz and 6.16 for 3.5 GHz in the LOS scenario. RMS delay spread has a mean value of 84.5 ns for 460 MHz and 35.5 ns for 3.5 GHz. The results in this paper are helpful for 5G channel modeling, system simulation, and communication system design.

References

[1] R. He, Z. Zhong, B. Ai, et al., "High-speed railway communications: from GSM-R to LTE-R," *IEEE Vehicular Technology Magazine*, vol. 11, no. 3, pp. 49–58, 2016. doi: 10.1109/MVT.2016.2564446.

[2] M. J. Marcus, "5G and 'IMT for 2020 and beyond' [spectrum policy and regulatory issues]", *IEEE Wireless Communications*, vol. 22, no. 4, pp. 2–3, 2015. doi: 10.1109/MWC.2015.7224717.

[3] ITU-R, "IMT vision—framework and overall objectives of the future development of IMT for 2020 and beyond," Rec. ITU-R M. 2083, Feb. 2014.

[4] R. He, M. Yang, L. Xiong, et al., "Channel measurements and modeling for 5G communication systems at 3.5 GHz band," in *URSI Asia-Pacific Radio Science Conference*, Seoul, South Korea, Aug. 2016, pp. 1855–1858. doi: 10.1109/URSIAP-RASC.2016.7601208.

[5] R. He, A. F. Molisch, F. Tufvesson, et al., "Vehicle-to-vehicle propagation models with large vehicle obstructions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2237–2248, 2014. doi: 10.1109/TITS.2014.2311514.

[6] T. K. Sarkar, Z. Ji, K. Kim, et al., "A survey of various propagation models for mobile communication," *IEEE Antennas and Propagation Magazine*, vol. 45, no.

3, pp. 51–82, 2003. doi: 10.1109/MAP.2003.1232163.
[7] T. S. Rappaport, S. Y. Seidel, and K. Takamizawa, "Statistical channel impulse response models for factory and open plan building radio communicate system design," *IEEE Transactions on Communications*, vol. 39, no. 5, pp. 794–807, 1991. doi: 10.1109/26.87142.
[8] R. He, W. Chen, B. Ai, et al., "On the clustering of radio channel impulse responses using sparsity-based methods," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 6, pp. 2465–2474, 2016. doi: 10.1109/TAP.2016.2546953.
[9] H. L. Bertoni, *Radio Propagation for Modern Wireless Systems*. London, UK: Pearson Education, 1999.
[10] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 42–49, 1995. doi: 10.1109/35.339880.
[11] X. Yin and X. Cheng. *Propagation Channel Characterization, Parameter Estimation, and Modeling for Wireless Communications*. Hoboken, USA: John Wiley & Sons, 2016.
[12] R. He, Z. Zhong, B. Ai, et al., "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. doi: 10.1109/LAWP.2011.2164389.
[13] R. He, Z. Zhong, B. Ai, and J. Ding, "Propagation measurements and analysis for high-speed railway cutting scenario," *Electronics Letters*, vol. 47, no. 21, pp. 1167–1168, 2011. doi: 10.1049/el.2011.2383.
[14] V. S. Abhayawardhana, I. J. Wassell, D. Crosby, M. P. Sellars, and M. G. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *IEEE 61st Vehicular Technology Conference, VTC 2005-Spring*, Stockholm, Sweden. doi: 10.1109/VETECS.2005.1543252.
[15] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," Rep. ITU-R M. 2135, 2008.
[16] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014. doi: 10.1109/JPROC.2014.2299397.
[17] A. F. Molisch, *Wireless Communications, 2nd ed.* Hoboken, USA: Wiley, 2010.
[18] R. He, Z. Zhong, B. Ai, et al. "Measurements and analysis of propagation channels in high-speed railway viaducts," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 794–805, 2013. doi: 10.1109/TWC.2012.120412.120268.
[19] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications Systems*, vol. 11, no. 4, pp. 360–393, 1963. doi: 10.1109/TCOM.1963.1088793.

Manuscript received: 2017-2-12

Biographies

YANG Mi (yangmi@bjtu.edu.cn) received the B.S. degree and the M.S. degree in electronic & communication engineering from Beijing Jiaotong University (BJTU), China. He is pursuing the Ph.D. degree with the State Key Laboratory of Rail Traffic Control and Safety, BJTU. His research interests include radio propagation models, vehicle-to-vehicle communications and software defined radio.

HE Ruisi (ruisi.he@bjtu.edu.cn) received his B.E. and Ph.D. degrees from Beijing Jiaotong University (BJTU), China in 2009 and 2015, respectively. He is an associate professor with the State Key Laboratory of Rail Traffic Control and Safety, BJTU. His research interests include radio propagation, long-term evolution-railways, and fifth-generation communications. Dr. He is an editor of the *IEEE Transactions on Wireless Communications*. He serves as the Early Career Representative of Commission C, International Union of Radio Science (URSI), and received the URSI Young Scientist Award in 2015.

AI Bo (boai@bjtu.edu.cn) received his M.S. and Ph.D. degrees from Xidian University, China in 2002 and 2004, respectively. He is a full professor and Ph.D. degree candidate advisor with the State Key Laboratory of Rail Traffic Control and Safety at Beijing Jiaotong University, China. He is the deputy director of the State Key Laboratory of Rail Traffic Control and Safety. He has authored or co-authored six books and published over 230 academic research papers. He holds 21 invention patents. He is an Institution of Engineering and Technology fellow and a vice chair of IEEE VTS Beijing Chapter. He is an associate editor of *IEEE Transactions on Consumer Electronics* and an editorial committee member of *Wireless Personal Communications*.

XIONG Lei (lxiong@bjtu.edu.cn) received his Ph.D. from Beijing Jiaotong University (BJTU), China in 2007. He is an associate professor at BJTU. He has authored or co-authored two books and more than 30 papers. He is an expert on railway communications in China. His research interests include rail mobile communications, channel simulation, and software defined radio.

DONG Honghui (hhdong@bjtu.edu.cn) received the B.E. degree from Xi'an Jiaotong University, China in 1999, the M.E. degree from the China Academy of Railway Sciences, China in 2002, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, China in 2007. He is currently an associate professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China. His current research interests include pattern recognition and intelligent systems, as well as transportation science and engineering.

LI Jianzhi (15111063@bjtu.edu.cn) received the B.S. degree in communication engineering, and the M.S. degree in electronic & communication engineering from Beijing Jiaotong University (BJTU), China in 2013 and 2015, respectively. Since then, he has been pursuing the Ph.D. degree at the State Key Laboratory of Rail Traffic Control and Safety, BJTU. His current research focuses on massive MIMO in real propagation environments, including massive MIMO channel measurements, channel characterization and modeling, and performance evaluation.

WANG Wei (Wei.Wang@dlr.de) received the bachelor's degree in communications engineering from Wuhan University, China in 2003; the master's degree from the University of Kiel, Germany in 2006; and the doctoral degree from the University of Erlangen-Nuremberg, Erlangen, Germany, with distinction (summa cum laude) in 2014. Since 2007, he has been a scientific staff member with the Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany. In March 2012, he was funded by EU COST-IC2100 for a short term scientific visiting at the University Aalborg, Denmark. His research interests includes channel measurement and modeling for localization/navigation applications, time variant channel parameters estimation and tracking, and radio based positioning and navigation. He has involved in several EU, ESA, GSA and DLR projects, e.g., GREAT, GRAMMAR, GalileoADAP, WHERE, WHERE2, and MarKom and so on. He has participated or led channel measurement campaigns for mobile, aeronautic, satellite and maritime propagation environments for these projects. He is a member of IEEE, VDE and COST 0802/2100. He received the best presentation paper award in ION GNSS 2012.

FAN Wei (wfa@es.aau.dk) received his B.E.achelor of Engineering degree from Harbin Institute of technology, China in 2009, master's double degrees with highest honors from Politecnico di Torino, Italy and Grenoble Institute of Technology, France in 2011, and Ph.D. degree from Aalborg University, Denmark in 2014. From February 2011 to August 2011, he was with Intel Mobile Communications, Denmark as a research intern. He conducted a three-month internship at Anite telecoms oy, Finland in 2014. He was a short-term visiting researcher at Keysight, Finland and Huawei, Sweden in 2016 and 2017. His main areas of research are over the air testing of MIMO terminals, radio channel modelling, virtual drive testing, and 5G phased array signal processing.

QIN Hongfeng (qin.hongfeng@zte.com.cn) received the master's and Ph.D. degrees from Northwestern Polytechnical University, China in 2000 and 2003, respectively. He is now working with ZTE Corporation as a senior engineer, senior expert and manager of baseband algorithm. He has authored or co-authored over 20 scientific research papers and books, and over 100 invention patents in his research area till now. His current research interests include wireless communications, radio propagation and channel modeling, and signal processing and application.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 003
http://www.cnki.net/kcms/detail/34.1294.TN.20170220.1110.002.html, published online February 20, 2017

A Survey of Massive MIMO Channel Measurements and Models

ZHANG Jianhua, WANG Chao, WU Zhongyuan, and ZHANG Weite
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract

Compared with conventional multiple-input multiple-output (MIMO), massive MIMO system with tens or even hundreds of antennas is able to give better performance in capacity and spectral efficiency, which is a promising technology for 5G. Considering this, massive MIMO has become a hot research topic all over the world. In this paper, the channel measurements and models of massive MIMO in recent years are summarized. Besides, the related 256 antenna elements with 200 MHz bandwidth at 3.5 GHz proposed by our team, the verification of rationality of the measurement method, and the spatial evolution of clusters in mobile scenario are provided.

Keywords

massive MIMO; channel measurement; channel model; virtual measurement; cluster

1 Introduction

Mobile traffic is predicted to grow more than 1000 times in the next 10 years, and the International Mobile Telecommunication (IMT) vision towards 2020 and beyond requires future 5G systems to deliver a 10 Gbps peak data rate. Considering this, more and more work is turning to the massive multiple-input multiple-output (MIMO) technology. Compared to the current state of the art, massive MIMO system has a large number of antennas, typically tens or hundreds, and provides better performance in efficiency, capacity, reliability and more [1]–[3]. It can improve channel capacity, can reduce latency on the air interface, and is robust against unintended man-made interference and intentional jamming [4], [5]. However, it brings increasing complexity of channel modeling.

Therefore, a series of massive MIMO measurement campaigns have been performed to evaluate the channel performance. For example, outdoor channel measurements at 2.6 GHz with a linear virtual array and a cylindrical array of 128-element antennas are reported in [6] that studied the sum-rate capacity, spectrum efficiency, precoding schemes, etc. In [7], an outdoor static measurement performed in a stadium is analyzed, with a linear 128 - element antenna virtual array at 1.4725 GHz and the angular power spectrum (APS) in the massive MIMO channel. The modeling of massive MIMO channel has also been studied. Moreover, spatial non-stationary properties should be considered in the model in response to the larger antenna array and near field effect. In this paper, massive MIMO channel measurements and modeling in recent years are reviewed. Section 2 discusses the recent measurement campaigns. The modeling work of massive MIMO is given in Section 3. Section 4 displays the work of our team, which analyzes the rationality of virtual measurement and the spatial evolution of clusters in the mobile scenario of massive MIMO. Finally, the conclusions are drawn in Section 5.

This research is supported in part by National Natural Science Foundation of China (61322110, 6141101115), in part by National Science and Technology Major Project of the Ministry of Science and Technology (2015ZX03002008), in part by National Key Technology Research and Development Program (2012BAF14B01), by "863" Program (2015AA01A703), and by Doctoral Fund of Ministry of Education (201300051100013).

2 Massive MIMO Channel Measurements

Channel measurements are indispensable in research of wireless communications. Here a series of massive MIMO measurement campaigns in recent two years are listed. **Table 1** gives thesetups and investigated channel characteristics of these measurements.

2.1 Capacity

With the increase of the antenna number, massive MIMO systems can improve spectral efficiency significantly. In [8], 128Tx - 16Rx massive MIMO is found to provide up to 434%

Table 1. Summary of massive MIMO channel measurements proposed in recent two years

Antenna array setup	Scenario	Carrier frequency (GHz)	Channel characteristics	Reference
Tx UPA/Rx ODA 16×16	UMi	3.5 and 2.35	Capacity; eigenvalues	[8]
Tx UPA/Rx ODA 32×56	UMa, O2I	6	Angle spread; delay spread; channel capacity	[40]
Tx UPA/Rx ODA 32×56	O2I	6	Delay spread; angular spread; capacity in different height	[41]
Tx UPA/Rx ODA 32×56	UMa	3.5	The rationality of virtual massive MIMO measurement	[34]
Tx UPA/Rx ODA 32×56	UMa mobility	3.5	Cluster number; cluster-AoA; cluster-AoD; radius of visibility region	[35]
Rx/cylindrical 24×2	O2I, UMi, UMa	2	Capability	[9]
Tx/cylindrical 32×4	indoor	19.85	SNR	[15]
Virtual linear 12×12	lecture hall	5.6	Condition number; delay spread variation	[11]
Tx/virtual circular 24	vehicle to infrastructure	2.6	SIR; power density	[42]
Rx/horn antenna 1	O2I	2.59	Correlation coefficient; SNR	[12]
Rx/2D virtual 12×12	UMa	2.53	Angle delay; angle spread.	[22]
UPA 4×4	indoor	2.4	SNR	[10]
Horizontal 64×1/ vertical 1×64	UMa, UMi	2.6	SNR	[43]
Planar 8×8	RMa	5.2	Power; SIR	[44]
Cylindrical	stadium	4.45	PDP of frequency correlation coefficient	[23]
Dipole	similar shopping hall	5.8	Condition number; scalar product	[13]
UCA 64×2	front square	3.33	PDP; PAS	[16]
Virtual 20×20	lecture hall	13–17	Channel gain; K-factor; delay spread; RMS delay spread	[45]
Tx/virtual linear 128×1	hall	2, 4, and 6	PL; PDP	[17]
Tx/ULA 128×8	stadium	1.4725	Channel gains; K-factors; (RMS) composite delay spreads	[7]

AoA: Angle of Arrival
AoD: Angle of Departure
O2I: Ourdoor to Indoor
ODA: Omnidirectional Array
PAS: Power Angular Spectrum
PDP: Power Delay Profile
PL: Path Loss
RMa: Rural Macro
RMS: Root Mean Square
SIR: Signal to Interference Ratio
SNR: Signal Noise Ratio
UCA: Uniform Cylinder Array
ULA: Uniform Linear Array
UMa: Urban Macro
UMi: Urban Micro
UPA: Uniform Planar Array

and 478% more capacity over traditional LTE single-user MIMO with 8Tx-8Rx configuration in macrocells and picocells, respectively. However, the benefit of diversity gains from user equipment (UE) with more antennas falls away as the dimensions of the base station (BS) array increase.

In [9], comparing with the 8 x 8 array, the 16 x 4 array improves the cell-average and cell-edge throughputs by 27% and 71% in Urban Macro (UMa) scenario and enhances the cell-average and cell-edge throughputs by 19% and 43% in Urban Micro (UMi) scenario. When comparing with 8 x 8, the 32 x 2 array enhances the cell-average and cell-edge throughputs by 60% and 112% in UMa and improves the cell-average and cell-edge throughputs by 80% and 118% in UMi scenario. In [10], an effective diversity gain measurement apparatus is proposed to assess diversity performance of multi-antenna systems.

2.2 Eigenvalue Properties and Antenna Array

The measurement in [8] demonstrates that, with the increasing height of user equipment, the BS elevation spreads decrease both in UMa and UMi while the azimuth spreads remain approximately the same. In [11], the condition number is shown to be suitable for measuring both the channel orthogonality between users and the channel harden effect according to the channel measurements in a large lecture hall.

The planar array geometry of a horizontal antenna element is compared with that of a vertical antenna element in [12]. The horizontal antenna arrangement appears to be best suited for massive MIMO operation and yields the lowest average correlation coefficient among the positions considered. According to [13], user proximity and user handgrip reduce the dispersion of the studied properties of the channel across frequencies.

Antenna selection aiming to reduce the number of RF transceiver chains is discussed in [14], in which switching structure and a convex-optimization scheme are presented.

2.3 Non-Stationary Properties

A sufficient interference reduction is obtained in [15] by zero forcing (ZF), whereas maximum ratio combining (MRC) cannot sufficiently reduce the interference when 24 elements are used, even in the 20 GHz band. Moreover, it shows that it is important to select the efficient antennas with high SNR when the antennas cannot be all used. In [6], significant variations in signal strength are characterized for several measured propagation scenarios in the 2.6 GHz frequency range and the change of power variations and correlation properties along with the array is illustrated.

In [16], a measurement campaign performs at the 3.33 GHz in outdoor scenarios, using an antenna array with 64 elements. The results show that the non-stationary properties of the channel over the large array size occur both in delay and spatial domains.

A measurement adopted frequency domain sounder in indoor scenarios in [17] indicates that the channel characteristics of massive MIMO are the non-stationary properties in spatial, delay and frequency domains and the independency between these channel parameters. The carrier frequency is within 2-6 GHz.

Several non-stationary properties of massive MIMO channels in a stadium are investigated in [18], in which the channel parameters appear stationary over the linear antenna array at the high frequency bandwidth (HFB) but not at low frequency bandwidth (LFB). That is because, more stronger Multipath Components (MPCs) appear at LFB due to the smaller path

loss and larger obstacle reflection coefficients.

2.4 Multiple Users

In [19], based on the measurements of multiple users (MU) massive MIMO system, spatial separation of closely-spaced users is demonstrated by the analysis and evaluation with the singular value spread.

Channel measurements are performed and corresponding singular value spreads and achieved sum-rate capacities are discussed in [20], and compared with conventional MIMO, massive MIMO is proven to provide better orthogonality between channels for different users and better channel stability.

2.5 Other Channel Characteristics

A flexible testbed is proposed in [21], where the base station operates with up to 100 coherent radio-frequency transceiver chains based on software radio technology.

In [22], an advance MIMO antenna array setup is used to conduct a measurement campaign in Uma propagation. The parameters of all multipath components are extracted with an iterative maximum likelihood high-resolution algorithm (RIMAX). An inter-user interference (IUI) cancellation scheme is also proposed in [22], which simplifies user scheduling method on massive antenna systems for wireless entrance. Besides, the system level simulations using the measured Channel State Information (CSI) confirm that enlarging the angular gap between users reduces spatial correlation and that the IUI cancellation proposal is effective under this condition.

The frequency correlation characteristics using the method of Fourier transform to the channel Power Delay Profile (PDP) are compared with those directly using cross-correlation to the transfer function in a stadium at 4.45 GHz in [23]. It can be determined that the uncorrelated scattering condition and constant means of transfer function are not held.

Angle properties of massive MIMO channels are studied with Multiple SIgnal Classification (MUSIC) and SAGE algorithms based on the channel measurement in outdoor scenarios in [7], in which the Directions of Arrival (DOAs) of the line-of-sight (LoS) signal change with varying geometrical positions in the environment.

The geometric method is used in [24] to characterize the attenuation behavior of the massive MIMO channel with an extended large attenuation matrix proposed. The results in [24] reveal that large attenuation is mostly determined by the transmission distance between the user and the nearest antenna and by the angle with respect to the antenna array as well.

3 Massive MIMO Channel Models

In recent years, many researchers all over the world have advanced the modeling work of massive MIMO channels. The channel models are usually classified into two categories: geometry-based stochastic models (GBSM) and correlation-based stochastic models (CBSM).

GBSMs have been studied extensively to evaluate the performance of wireless channels. These models have the advantage of comprising channel properties accurately. In [25], a novel non-stationary multi-ring channel model on both time and array axes is proposed for massive MIMO systems. With the multi-ring distribution of clusters, the propagation characteristics, such as power imbalance over the array, eigenvalue distribution and antenna correlation, match the conclusions drawn in measurements well. GBSM' direct involvement of scatters/clusters renders it as one of the most promising candidates for 3D MIMO channel modeling [26].

Considering the elevation, the spherical, cylindrical or other kinds of antenna arrays can be adopted. A3D two-cylinder regular-shaped GBSM for non-isotropic scattering massive MIMO channels is proposed in [27], which considers the non-stationary properties, 3D MIMO and spherical wave effect. A 3D wideband twin-cluster channel model is proposed in [28] for massive MIMO communication systems with carrier frequencies on the order of GHz. The near field effect and non-stationary properties are considered. An extension based on the cluster-based COST-2100 MIMO channel model is proposed in [6]. In this model, the channel cannot be seen as wide-sense stationary over the large array at the base station.

CBSMs have lower complexity. They can be categorized into the independent and identically distributed (i.i.d.) Rayleigh fading channel model and correlation channel model. The Kronecker-based stochastic model (KBSM) is one kind of correlation channel models. In [29], a novel KBSM for massive MIMO is proposed, which could capture antenna correlations well. In this model, the evolution of scatters is modeled by the birth-death process and the antenna correlation matrix is equal to the spatial correlation matrix and survival probability matrix. Unlike the i.i.d. Rayleigh channel model, KBSM in [29] considers antenna correlation for the channel model.

4 Analysis of Our Channel Measurement Campaigns

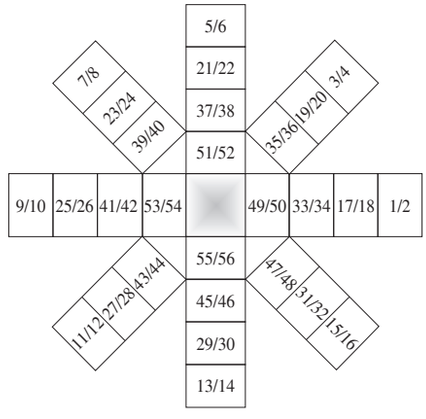
From 2010 to 2014, our team focused on 3D channel measurement and modeling [26] and contributed to 3D channel model standards, 3GPP TR 36.873 [30]. We has started to expand our channel sounder to support the massive MIMO channel measurement and modeling since 2015.

In our measurement work, a 32-element uniform planar array (UPA) was used in the Tx side, and a 16-element dual-polarized omnidirectional array (ODA) was used at the Rx side (**Fig. 1**). The parameters of the antennas are listed in **Table 2** and a straight route was set for the mobile scenario. The measurement campaigns were performed in Urban Macro (UMa), Urban Micro (UMi) and Ourdoor to Indoor (O2I). Here we give an analysis of the UMa measurement. (**Fig. 2**)

To form the massive MIMO array with different antenna

1/5	2/6	3/7	4/8
9/13	10/14	11/15	12/16
17/21	18/22	19/23	20/24
25/29	26/30	27/31	28/32

(a) Layout and schematic of antenna array at the Tx side



(b) Layout and schematic of antenna array at the Rx side

▲Figure 1. Antenna layouts used in our measurement work.

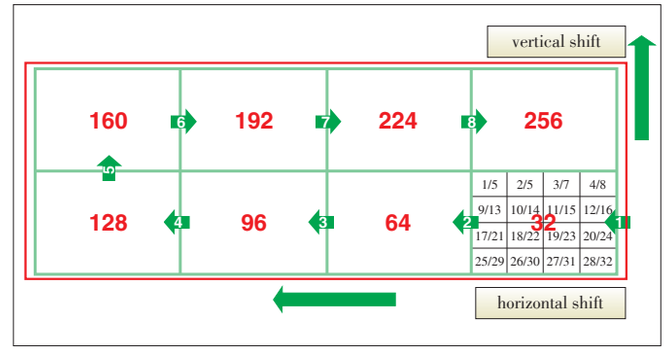
▼Table 2. The antenna parameters used in our measurement

Parameter		Value	
Antenna type		ODA (Rx)	UPA (Tx)
Number of antenna ports		56 (#1–#16 were chosen)	32
Overall radiation pattern		Omnidirectional	Hemispherical
Inter element spacing		41.0 mm	41.0 mm
Number of elements		28	16
Angle range	Azimuth	−180°–180°	−70°–70°
	Elevation	−70°–90°	−70°–70°
Polarized		±45°	±45°
Center frequency		3.5 GHz	
Bandwidth		200 MHz	
PN sequence		255 chips	

ODA: Omnidirectional Array UPA: Uniform Planar Array

numbers, our channel measurement campaigns were performed by using the virtual measurement method. **Fig. 3** shows the combining scheme, giving eight adjacent positions. For example, if we wanted to get the data of 64-element virtual antenna array, two groups of CIRs collected from two adjacent positions would be chosen and combined into one group of data. Then we used it as equivalent data collected from 64-element antenna array for further analysis. Similarly, reordering eight groups

▲Figure 2. The overview of the measurement area by Baidu Map (The red triangle is the Tx side; two yellow lines, R1 and R2, represent the measurement route in LoS and non-LoS (NLoS) conditions, respectively; and three red points are in LoS conditions while three blue points are in NLoS conditions).



▲Figure 3. The scheme of the antenna combining array in the virtual measurement.

of CIRs to one group could get the data of 256-element virtual antenna array and reordering 4 groups of CIRs to get the data of 128-element virtual antenna array. By the above process, we obtained the data of 256, 128, 64 and 32-element antenna arrays.

To estimate the channel parameters, the SAGE algorithm was used [31], which provides a joint estimation of parameter set $\theta_l=\{\tau_l,f_{d,l},\Phi_l,\Omega_l,\alpha_l\}$, $l=\{1,2,\cdots,L\}$. The τ_l, $f_{d,l}$, Φ_l, Ω_l and α_l denote the propagation delay, the doppler shift, the angle of departure, the angle of arrival and polarization of the l-th propagation sub-path, respectively. Specifically, $\Phi_l=[\theta_{T,l},\phi_{T,l}]$ and $\Omega_l=[\theta_{R,l},\phi_{R,l}]$, where $\theta_{T,l}$, $\phi_{T,l}$, $\theta_{R,l}$ and $\phi_{R,l}$ denote the elevation angle of departure (EoD), angle of departure (AoD), elevation angle of arrival (EoA) and angle of arrival (AoA), respectively. Every 4 snapshots are fed to SAGE to estimate one parameter set.

Finally, to observe the characteristics of clusters, the KpowerMeans clustering algorithm was used to get cluster-level pa-

rameters [32], [33]. The multiple path component distance (MCD) is the distance measure for different paths in Kpower-Means. It is a normalized value composing of delay and angular parts.

For delay distance, the definition is given as

$$MCD_{\tau,ij}=\eta\cdot\frac{\left|\tau_i-\tau_j\right|}{\Delta\tau}\cdot\frac{\tau_{sd}}{\Delta\tau}. \tag{1}$$

In (1), η is a scaling factor to adjust the weight of delay in the distance function. $\Delta\tau$ means the range of delay and $\Delta\tau=\max_{i,j}\left|\tau_i-\tau_j\right|$. τ_{sd} is the standard deviation of delay.

For angle distance, the definition is given as

$$MCD_{T/R,ij}=\frac{1}{2}\begin{vmatrix}\sin(\theta_i)\cos(\varphi_i) & \sin(\theta_j)\cos(\varphi_j)\\ \sin(\theta_i)\sin(\varphi_i) & -\sin(\theta_j)\sin(\varphi_j)\\ \cos(\theta_i) & \cos(\theta_j)\end{vmatrix}. \tag{2}$$

For Tx or Rx, the angle distance is obtained in the spherical coordinate system. θ means the elevation angle and φ means the azimuth angle.

The total distance is given by

$$MCD_{ij}=\sqrt{\left\|MCD_{T,ij}\right\|^2+\left\|MCD_{R,ij}\right\|^2+\left\|MCD_{\tau,ij}\right\|^2}. \tag{3}$$

We can see from above equations that MCD consists of angle of arrival, angle of departure, and delay.

4.1 Rationality of the Virtual Measurement

The rationality of the virtual measurement was proved in time and spatial domains [34], the power delay profile (PDP) calculated from combined CIRs and CIRs collected from the measurement campaigns could fit well. Estimated from the combination, the spatial angular characteristics, the elevation angle of departure (EoD), azimuth of departure (AoD), elevation angle of arrival (EoA), and azimuth of arrival (AoA) also matched well with those from the measurement campaigns.

We choose the spot P3 in LoS scenario and the spot P5 in NLoS scenario (Fig. 2) as examples. The other spots in Fig. 2 havesimilar characteristics.

Based on (4), the PDP reconstruction follows the form as (5):

$$\begin{aligned}h_{n,m}(t,\tau_n)=&\begin{bmatrix}F_{rx,n,V}(\Omega_l)\\ F_{rx,n,H}(\Omega_l)\end{bmatrix}^T\begin{bmatrix}\alpha_{l,VV} & \alpha_{l,VH}\\ \alpha_{l,HV} & \alpha_{l,HH}\end{bmatrix}\begin{bmatrix}F_{tx,n,V}(\Phi_l)\\ F_{tx,n,H}(\Phi_l)\end{bmatrix}\times\\ &\exp(jd_n2\pi\lambda_0^{-1}\sin(\Omega_l))\times\\ &\exp(jd_m2\pi\lambda_0^{-1}\sin(\Phi_l))\times\exp(j2\pi f_{d,l}t)\end{aligned}, \tag{4}$$

$$P(\tau_n)=\left\|\begin{bmatrix}F_{rx,V}(\Omega_l)\\ F_{rx,H}(\Omega_l)\end{bmatrix}^T\begin{bmatrix}\alpha_{l,VV} & \alpha_{l,VH}\\ \alpha_{l,HV} & \alpha_{l,HH}\end{bmatrix}\begin{bmatrix}F_{tx,V}(\Phi_l)\\ F_{tx,H}(\Phi_l)\end{bmatrix}\right\|_F^2, \tag{5}$$

where F_{rx} and F_{tx} represent the field patterns of antenna elements in Tx and Rx ends that are generated by the principle of MIMO over-the-air (OTA) test. $\begin{bmatrix}\alpha_{l,VV} & \alpha_{l,VH}\\ \alpha_{l,HV} & \alpha_{l,HH}\end{bmatrix}$ represents the polarized complex amplitude matrix (V stands for vertical polarization and H stands for horizontal polarization).

The SAGE algorithm is a parameter extraction and estimation procedure from the strong path to weak path. One path is characterized by five parameters by SAGE algorithm, and the angular information describes the features of the signal transmission in the space. The rationality of the virtual measurement is discussed by comparing the distribution of paths in the spatial domain with the practical measurement. The power azimuth spectrums (PAS) of practical measured data and virtual measured data in LoS and NLoS scenarios are given, respectively.

Fig. 4a shows the PAS results of angle of departure in the practical measurement in the LoS scenario, while **Fig. 4b** shows the PAS results in the virtual measurement in the LoS scenario. In Fig. 4, the X-coordinate represents the azimuth angle and Y-coordinate represents the elevation angle; the red zone of the label represents the strong path estimated by the SAGE algorithm in the angle of departure and the deep color area represents the weak path which we can mostly ignore because the value of power is lower than −25 dB. The color represents its power in these figures. It is obviously seen that the strong path region is the same in Figs. 4a and 4b and the power values of the strong path are almost equal. Figs. 4c and 4d represent the signal angles in the receiving end; the light color areas there are mostly similar, which means that the major paths are same in the practical and virtual measurements at the same measurement spot. Therefore, in the LoS scenario, the path distribution of the angle domain in virtual measurement is the same as that in practical measurement.

Similarly, **Fig. 5** depicts the PAS results in the NLoS scenario, which are more complex than those in the LoS scenario. According to Fig. 5, the region of the strong path increases in the NLoS scenario. In the transmitting end (Figs. 5a and 5b), the light color areas are more scattered. In the receiving end (Figs. 5c and 5d), it is observed that the number of paths get much more than that in the LoS scenario. With the path number increases, the corresponding departure and arrival angle values of each path in practical and virtual measurements are approximately equal. Therefore, a similar conclusion can be drawn in the NLoS scenario that the path distributions of virtual and practical measurements are almost the same.

4.2 Evolution of Clusters in the Mobile Scenario

The evolution of clusters presents the spatial variation of the massive MIMO channel during the movement [35]. When the mobile station (MS) moves, the power of different clusters varies, and some clusters appear or disappear. Based on the results of clustering, we can get the cluster number evolution

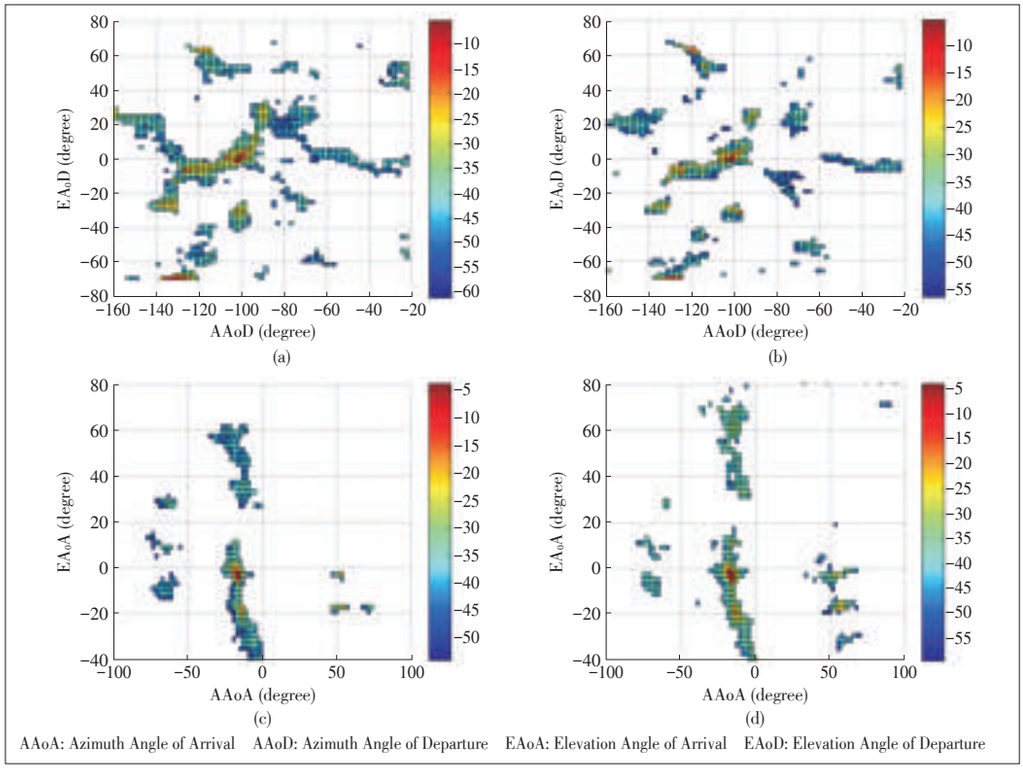


Figure 4.
The PAS results in the LoS scenario: (a) and (c) represent the results from the practical measurement and (b) and (d) represent the results from the virtual measurement.

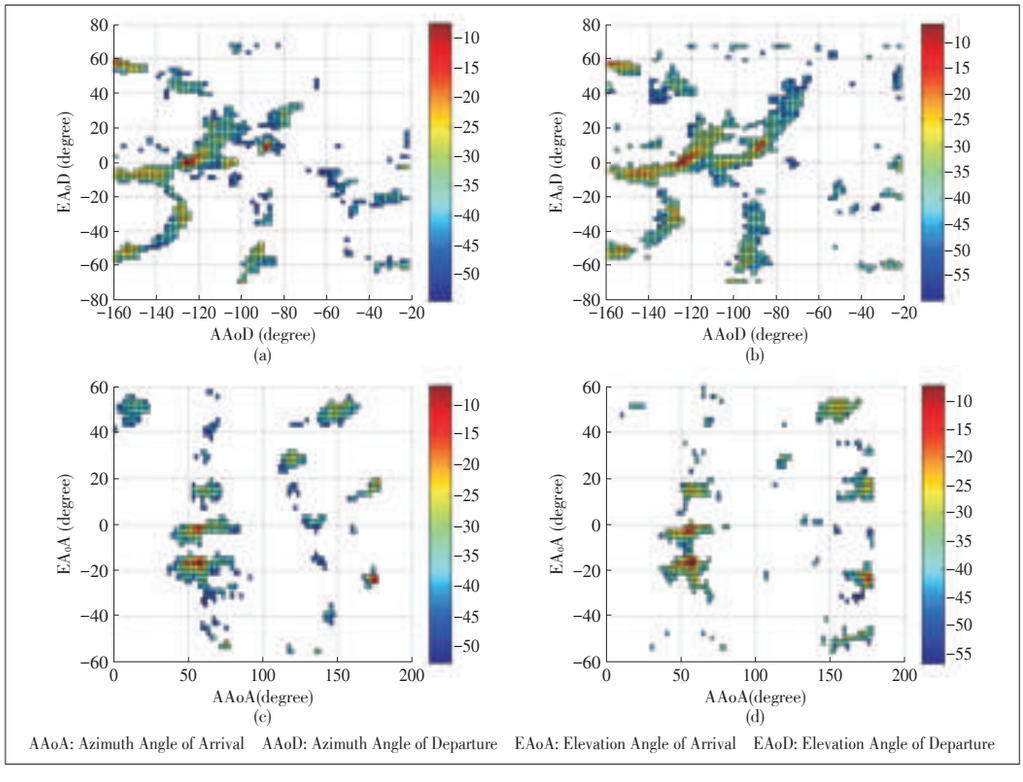
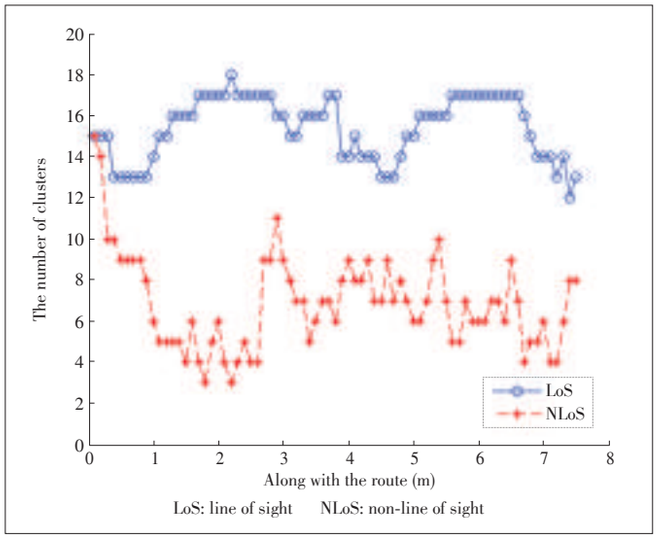


Figure 5.
The PAS results in the NLoS scenario: (a) and (c) represent the results from the practical measurement and (b) and (d) represent the results from the virtual measurement.

(**Fig. 6**) in R1 (LoS) and R2 (NLoS). The blue curve in Fig. 6 represents LoS conditions, with the average number of clusters is 13.7867; when the distance is 0.5 m, it gets the largest cluster number, 20. On the other hand, the average number is 5.8533 in NLoS conditions (the red curve in Fig. 6), and the largest number is 17 with the distance of 2.9 m and 3 m. There-



▲Figure 6. The evolution of clusters along with the measurement route.

fore, the cluster number in LoS conditions is more than that in NLoS conditions. This is because the cluster power is normalized and some MPCs' normalized power is too low to keep for clustering. In addition, the cluster number changes more sharply in NLoS conditions than in LoS conditions.

The evolution of clusters was simulated based on the birth-death process that can reflect the non-stationary properties of the clusters. The evolution of clusters $C(i)$, $\{c_1, c_2, \ldots, c_R\}$ is expressed as

$$C(i) \xrightarrow{E} C(i+1), (i = 1, 2, \cdots, R), \tag{6}$$

where R is the evolution number. It depends on the length of measurement route L and observation spacing δ_R along with the route, e.g., when the length of route is 8 m and the observation spacing is set as 0.1 m, the evolution number R is 80.

According to [36], the process of birth and death is assumed to be statistic independent, because the time variation of the channel can also be reflected when MS moves. The main variable in this model is the distance. The birth and death probabilities of each cluster are respectively expressed as

$$\lambda_i = \lambda e^{-\lambda \delta_R D}, \tag{7}$$

$$\mu_i = \mu e^{-\lambda \delta_R D}, \tag{8}$$

where λ is the cluster newly-generate rate, μ is the cluster disappear rate, δ_R is the observation spacing, and D is the distance correlation factor. Besides, the birth-death process should meet the following conditions

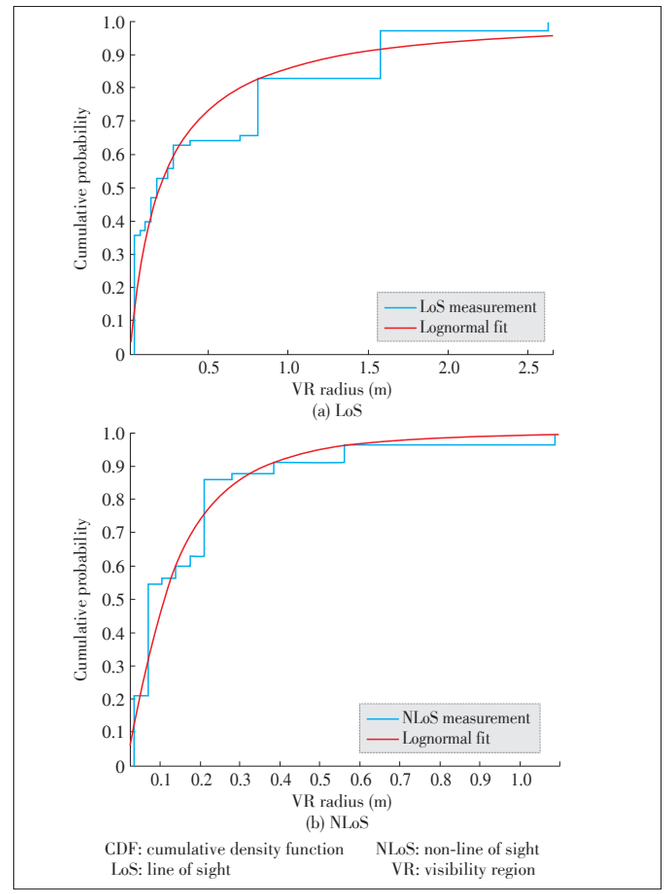
$$\begin{cases} p_{ii+1}(t) = \lambda_i(t) + o(t), (\lambda_i > 0, i = 0, 1, \cdots, N-1, \lambda_N = 0) \\ p_{ii-1}(t) = \mu_i(t) + o(t), (\mu_i > 0, i = 0, 1, \cdots, N, \mu_0 = 0) \\ p_{ii} = 1 - (\lambda_i + \mu_i)t + o(t) \\ p_{ij}(t) = o(t), |i-j| \geqslant 2 \end{cases}. \tag{9}$$

It is noted that the number of clusters is depend on the clustering algorithm. The ratio could be set to tell different clusters and the distance among different MPCs in a cluster. In this case, we wanted show more clusters in NLoS conditions so as to get more information of spatial evolution of clusters.

4.3 Visibility Regions

The concept of visibility region (VR) is proposed by the COST 259, COST 2100 MIMO channel model and more [37], [38]. It is an assumed circular region given in the measurement route. Each VR is related to only one cluster. When the MS moves inside one VR, the corresponding cluster would be active. Otherwise, it would be inactive. With these features, clusters can be observed clearly in mobile scenarios.

Based on the results of clustering, we inferred and calculated the lifetimes of clusters by converting the lifetime to the radius of VR [39], and the cumulative density functions (CDF) of radii of the VRs are presented. To simplify the method, we assume that the MS moves across the center of VRs, i.e., all of the circle centers are distributed along with the route. In LoS conditions, the radii of VRs range from 0.03 m to 2.63 m (**Fig. 7a**), while the radii in NLoS conditions range from 0.03 m to 1.0871 m (**Fig. 7b**), which are less than those in LoS condi-



▲Figure 7. The CDFs of radii of VRs in (a) LoS and (b) NLoS conditions.

tions because the appearance and disappearance of clusters occur more frequently. The lognormal distribution fitting is also built, with respect to the measurement curves. The mean and standard deviations of the fitting curves are 0.5180 m and 0.6505 m in LoS conditions and 0.1778 m and 0.2187 m in NLoS conditions, respectively (Fig. 7).

5 Conclusions

In this paper, we summarize the channel measurements and models of massive MIMO system in recent two years. With the increasing number of antennas, the capacity, efficiency and reliability of the system achieve better performance. The channel measurements of channel characteristics including capacity, eigenvalue properties, antenna arrays and MU are analyzed. Based on the channel model categories, the channel models proposed in recent two years are reviewed. The massive MIMO research work by our team is also presented. In our research work, the rationality of massive MIMO virtual measurement is verified and the evolution of clusters of massive MIMO in mobile scenarios is analyzed.

References

[1] N. Czink, X. Yin, H. Özcelik, et al., "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1465–1475, Apr. 2007. doi: 10.1109/TWC.2007.05595.

[2] L. Hentilä, M. Alatossava, N. Czink, and P. Kyösti, "Cluster-level parameters at 5.25 GHz indoor-to-outdoor and outdoor-to-indoor MIMO radio channels," in *Proc. 16th IST Mobile and Wireless Communication Summit*, Budapest, Hungary, 2007, pp. 1–5.

[3] J. Poutanen, K. Haneda, J. Salmi, et al., "Analysis of radio wave scattering processes for indoor MIMO channel models," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Tokyo, Japan, 2009, pp. 102–106.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010. doi: 10.1109/TWC.2010.092810.091092.

[5] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014. doi: 10.1109/MCOM.2014.6736761.

[6] X. Gao, F. Tufvesson, and O. Edfors, "Massive MIMO channels—measurements and models," in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, 2013, pp. 280–284.

[7] W. Li, L. Liu, C. Tao, et al., "Channel measurements and angle estimation for massive MIMO systems in a stadium," in *17th International Conference on Advanced Communication Technology (ICACT)*, Seoul, South Korea, 2015, pp. 105–108.

[8] Y. Yu, J. Zhang, M. Shafi, et al., "Measurements of 3D channel impulse response for outdoor-to-indoor scenario: capacity predictions for different antenna arrays," in *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Hong Kong, China, 2015, pp. 408–413.

[9] G. Liu, X. Hou, F. Wang et al., "Achieving 3D-MIMO with massive antennas from theory to practice with evaluation and field trial results," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–10, Apr. 2016. doi: 10.1109/JSYST.2015.2477503.

[10] W. J. Liao, B. Y. Dai, and B. R. Hsiao, "Effective diversity gain evaluation for large-scale MIMO antenna system," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1394–1397, 2016.

[11] J. Li and Y. Zhao, "Channel characterization and modeling for large-scale antenna systems," in *14th International Symposium on Communications and Information Technologies (ISCIT)*, Incheon, South Korea, 2014, pp. 559–563.

[12] M. Gauger, J. Hoydis, C. Hoek, et al., "Channel measurements with different antenna array geometries for massive MIMO systems," in *10th International ITG Conference on Systems, Communications and Coding*, Hamburg, Germany, 2015, pp. 1–6.

[13] A. O. Martinez, E. De Carvalho, J. O. Nielsen, and L. Jing, "Frequency dependence of measured massive MIMO channel properties," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, pp. 1–5.

[14] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Multi-switch for antenna selection in massive MIMO," in *IEEE Global Communications Conference (GLOBECOM)*, San Diego, USA, 2015, pp. 1–6.

[15] R. Kataoka, K. Nishimori, N. Tran, and T. Imai, "Basic performance of massive MIMO in indoor scenario at 20-GHz band," in *International Symposium on Antennas and Propagation (ISAP)*, Hobart, Tasmania, 2015, pp. 1–4.

[16] D. Fei, R. He, B. Ai, et al., "Massive MIMO channel measurements and analysis at 3.33 GHz," in *10th International Conference on Communications and Networking in China (ChinaCom)*, Shanghai, China, 2015, pp. 194–198.

[17] J. Li, B. Ai, R. He, et al., "Measurement-based characterizations of indoor massive MIMO channels at 2 GHz, 4 GHz, and 6 GHz frequency bands," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, pp. 1–5.

[18] L. Liu, C. Tao, D. W. Matolak, et al., "Stationarity investigation of a LOS massive MIMO channel in stadium scenarios," in *IEEE 82nd Vehicular Technology Conference (VTC Fall)*, Boston, USA, 2015, pp. 1–5.

[19] J. Flordelis, X. Gao, G. Dahman, et al., "Spatial separation of closely-spaced users in measured massive multi-user MIMO channels," in *IEEE International Conference on Communications (ICC)*, London, UK, 2015, pp. 1441–1446.

[20] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.

[21] J. Vieira, S. Malkowsky, K. Nieman, et al., "A flexible 100-antenna testbed for Massive MIMO," in *IEEE Globecom Workshops (GC Wkshps)*, Austin, USA, 2014, pp. 287–293.

[22] S. Sangodoyin, V. Kristem, C. U. Bas, et al., "Cluster-based analysis of 3D MIMO channel measurement in an urban environment," in *IEEE Military Communications Conference (MILCOM 2015)*, Tampa, USA, 2015, pp. 744–749.

[23] Y. Lu, C. Tao, L. Liu, et al., "Frequency correlation investigation of massive MIMO channels in a stadium at 4.45 GHz," in *17th International Conference on Advanced Communication Technology (ICACT)*, Seoul, South Korea, 2015, pp. 271–274.

[24] L. Liu, D. W. Matolak, C. Tao, et al., "Geometry based large scale attenuation over linear massive MIMO systems," in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, Davos, Switzerland, 2016, pp. 1–5.

[25] Q. U. A. Nadeem, A. Kammoun, M. Debbah, and M. S. Alouini, "On the mutual information of 3D massive MIMO systems: an asymptotic approach," in *2015 IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, China, 2015, pp. 2588–2592.

[26] X. Cheng, B. Yu, L. Yang, et al., "Communicating in the real world: 3D MIMO," *IEEE Wireless Communications*, vol. 21, no. 4, pp. 136–144, Aug. 2014. doi: 10.1109/MWC.2014.6882306.

[27] Y. Xie, B. Li, X. Zuo, M. Yang, and Z. Yan, "A 3D geometry-based stochastic model for 5G massive MIMO channels," in *11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE)*, Taipei, China, Aug. 2015, pp. 216–222.

[28] S. Wu, C. X. Wang, e. H. M. Aggoune, M. M. Alwakeel, and Y. He, "A non-stationary 3-D wideband twin-cluster model for 5G massive MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1207–1218, Jun. 2014. doi: 10.1109/JSAC.2014.2328131.

[29] S. Wu, C. X. Wang, E. H. M. Aggoune, and M. M. Alwakeel, "A novel Kronecker-based stochastic model for massive MIMO channels," in *IEEE/CIC International Conference on Communications in China (ICCC)*, Shenzhen, China, 2015, pp. 1–6.

[30] *Technical Specification Group Radio Access Network; Study on 3D channel model for LTE (Release 12. 2. 0)*, 3GPP TR 36.873, 2015.

[31] B. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Ingeman Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 434–450, 1999.

[32] D. Du, J. Zhang, C. Pan, and C. Zhang, "Cluster characteristics of wideband 3D MIMO channels in outdoor-to-indoor scenario at 3.5 GHz," in *IEEE 79th Vehicular Technology Conference*, Seoul, South Korea, 2014, pp. 1–6.

[33] C. Huang, J. Zhang, X. Nie, and Y. Zhang, "Cluster characteristics of wideband MIMO channel in indoor hotspot scenario at 2.35GHz," in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, Anchorage, USA, 2009, pp.

1–5.
[34] H. Yu, J. Zhang, Q. Zheng, et al., "The rationality analysis of massive MIMO virtual measurement at 3.5 GHz," in *IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, Chengdu, China, 2016, pp. 1–5.
[35] C. Wang, J. Zhang, L. Tian, M. Liu, and Y. Wu, "The Spatial evolution of clusters in massive MIMO mobile measurement at 3.5 GHz," presented at VTC2017-Spring, 2017 IEEE 85th Vehicular Technology Conference, Sydney, Australia, 2017.
[36] N. Czink, "The random-cluster model—a stochastic MIMO channel model for broadband wireless communication systems of the 3rd generation and beyond," Dissertation, Telecommunications Research Center Vienna (FTW), 2007.
[37] A. F. Molisch, H. Asplund, R. Heddergott, M. Steinbauer, and T. Zwick, "The COST259 directional channel model—part I: overview and methodology," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3421–3433, Dec. 2006.
[38] R. Verdone and A. Zanella, *Pervasive Mobile and Ambient Wireless Communications: COST Action 2100*. London, UK: Springer-Verlag, 2012.
[39] J. Poutanen, K. Haneda, J. Salmi, V. M. Kolmonen, and P. Vainikainen, "Modeling the evolution of number of clusters in indoor environments," in *Proc. 4th European Conference on Antennas and Propagation*, Barcelona, Spain, Apr. 2010, pp. 1–5.
[40] Y. Zhang, L. Tian, Y. Yu, et al., "3D MIMO channel characteristics and capacity evaluation for different dynamic ranges in outdoor-to-indoor scenario for 6 GHz," presented at IEEE 84th Vehicular Technology Conference VTC2016-Fall, Montreal, Canada, 2016.
[41] Q. Zheng, J. Zhang, H. Yu, Y. Zhang and L. Tian, "Propagation statistic characteristic of 3D MIMO channel in outdoor-to-indoor scenario with different antenna heights", presented at 19th International Symposium on Wireless Personal Multimedia Communications (WPMC 2016), Shenzhen, China, 2015, accepted.
[42] R. Zhang, Z. Zhong, J. Zhao, B. Li, and K. Wang, "Channel measurement and packet-level modeling for V2I spatial multiplexing uplinks using massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7831–7843, Oct. 2016.
[43] S. Zhang, A. Doufexi, and A. Nix, "Evaluating realistic performance gains of massive multi-user MIMO system in urban city deployments," in *23rd International Conference on Telecommunications (ICT)*, Thessaloniki, Greece, 2016, pp. 1–6.
[44] K. Maruta, A. Ohta, S. Kurosaki, T. Arai, and M. Iizuka, "Experimental investigation of space division multiplexing on massive antenna systems," in *IEEE International Conference on Communications (ICC)*, London, UK, 2015, pp. 2042–2047.
[45] J. Chen, X. Yin, and S. Wang, "Measurement-based massive MIMO channel modeling in 13–17 GHz for indoor hall scenarios," in *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–5. doi: 10.1109/ACCESS.2017.2652983.

Manuscript received: 2016-11-09

Biographies

ZHANG Jianhua (jhzhang@bupt.edu.cn) received her Ph.D. degree in circuit and system from Beijing University of Posts and Telecommunication (BUPT), China in 2003 and now is a professor of BUPT. She has published more than 100 articles in referred journals and conferences and 40 patents. She was awarded "2008 Best Paper" of *Journal of Communication and Network*. In 2007 and 2013, she received two national novelty awards for her contribution to the research and development of beyond 3G TDD demo system with 100 Mbps@20 MHz and 1 Gbps@100 MHz respectively. In 2009, she received the Second Prize for Science Novelty from Chinese Communication Standards Association for her contributions to ITU-R 4G (ITU-R M.2135) and 3GPP Relay channel model (3GPP 36.814). From 2012 to 2014, she did the 3D channel modeling work and contributed to 3GPP 36.873. She is also the member of the 3GPP 5G channel model for bands up to 100 GHz. Her current research interests include 5G, artificial intelligence, data mining especially in massive MIMO and millimeter wave channel modeling, channel emulator, and OTA test. She is an IEEE senior member and the drafting group (DG) chairwoman of ITU-R IMT-2020 channel model.

WANG Chao (chaowang@bupt.edu.cn) is working for a Ph.D. degree in the Key Lab of Universal Wireless Communications of Ministry of Education, Beijing University of Posts and Telecommunications, China. His current research field is massive MIMO channel measurement and modelling.

WU Zhongyuan (pyboon@foxmail.com) received the B.S. degree from China University of Petroleum, China in 2016. He is currently pursuing a M.S. degree from the Beijing University of Posts and Telecommunication, China. His research interests include massive MIMO channel measurement and modeling, and wireless channel characteristics in 5G system.

ZHANG Weite (zhangweite1994@foxmail.com) is currently a postgraduate student of Beijing University of Posts and Telecommunications (BUPT), China. He received the B.Sc. degree in communication engineering from Shanghai University, China in 2016. His research interests include propagation models, statistical properties of massive MIMO, and machine learning.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 004
http://www.cnki.net/kcms/detail/34.1294.TN.20170209.1001.002.html, published online February 9, 2017

Feasibility Study of 60 GHz UWB System for Gigabit M2M Communications

WANG Qi[1], GENG Suiyan[1], ZHAO Xiongwen[1,2], HONG Wei[2], and Katsuyuki Haneda[3]
(1. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China;
2. State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China;
3. Department of Radio Science Engineering, Aalto University, Espoo FI-00076, Finland)

Abstract

In this paper, the feasibility and performance of millimeter wave (mmWave) 60 GHz ultra-wide band (UWB) systems for gigabit machine-to-machine (M2M) communications are analyzed. Specifically, based on specifications, channel measurements and models for both line-of-sight (LOS) and non-LOS (NLOS) scenarios, 60 GHz propagation mechanisms are summarized, and 60 GHz UWB link budget and performance are analyzed. Tests are performed for determining ranges and antenna configurations. Results show that gigabit capacity can be achieved with omni-directional antennas configuration at the transceiver, especially in LOS conditions. When the LOS path is blocked by a moving person or by radiowave propagation in the NLOS situation, omni-directional and directional antennas configuration at the transceiver is required, especially for a larger range between machines in office rooms. Therefore, it is essential to keep a clear LOS path in M2M applications like gigabit data transfer. The goal of this work is to provide useful information for standardizations and design of 60 GHz UWB systems.

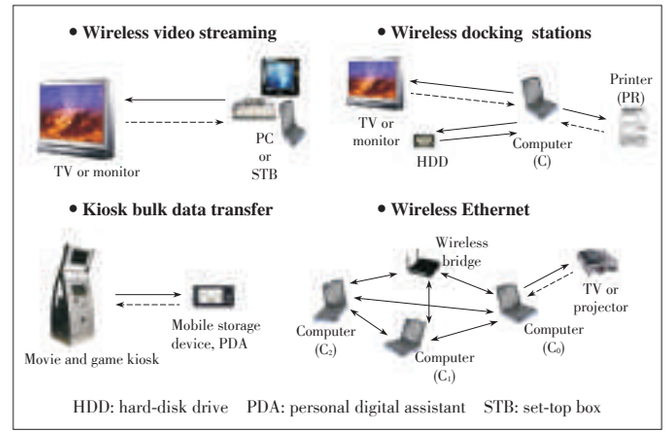
Keywords

mmWave 60 GHz; UWB; M2M; gigabit communications

1 Introduction

The millimeter wave (mmWave) 60 GHz band is seen as the major candidate for enabling wireless interface for gigabit applications, due to the intrinsic high transmission bandwidth available in the band [1], [2], [3]. The large bandwidth (as a thumb of rule the available bandwidth B is about 10% of the center frequency for transmission) makes 60 GHz radio particularly interesting for gigabit wireless communications [4]. There is 7 GHz unlicensed spectrum (from 57 to 64 GHz) around 60 GHz allocated worldwide today. The 60 GHz radio is often viewed as a shifted version of ultra-wide band (UWB), which is defined as any device emitting signals with fractional bandwidth greater than 0.2 or a bandwidth of at least 500 MHz at all times of transmission. Today, gigabit wireless applications are emerging, especially gigabit machine-to-machine (M2M) applications, e.g., wireless audio/video (A/V) cable replacement, wireless high data transfer, high quality multimedia services, etc. In high definition television (HDTV) applications, up to several Gb/s rate is required for supporting uncompressed exchange of information between TV, cameras, DVD and other appliances. **Fig. 1** shows examples of the 60 GHz radio for gigabit wireless applications in M2M networks.

The 60 GHz band regulation and standardization efforts are currently underway worldwide. At present, the international standards for 60 GHz band are ECMA-387, IEEE 802.15.3c and IEEE 802.11ad. In China, the study progress of 60 GHz



▲**Figure 1. Examples of the 60 GHz radio for gigabit M2M communications.**

This work is supported by the State Key Laboratory of Millimeter Waves, Southeast University, China under grant No. K201517. It is also supported by the Fundamental Research Funds for the Central Universities under Grant No. 2015 XS19.

band has been attracting more and more attention in recent years. In 2010, the 60 GHz wireless network project group (PG4) was established. The PG4 and the IEEE 802.11 working group formed a formal partnership. The IEEE 802.11-aj task group was founded according to the China millimeter wave band for next generation in September 2012. The IMT-2020 (5G) group was founded on February 19, 2013 for promoting the standard formulation in the 60 GHz band for 5G technology, and the first standard plan was completed in 2014 [5], [6].

However, the output power of 60 GHz devices is limited by regulations, and the free space path loss for a 60 GHz band carrier is much higher than that for a microwave carrier. Though it is possible to use high gain antennas to compensate for the high path loss at mmWave, the drawbacks of high gain antenna are that systems suffer from poor flexibility and limited mobility. In this work, we analyze 60 GHz UWB system link budget and performance, considering office room M2M applications (e.g. computer-to-computer data transfer) in the range of 1 m–5 m and employing the experimental path loss models of line-of-sight (LOS), non-LOS (NLOS) and LOS path blocked by moving persons studied in our previous works [7], [8]. Tests are also performed for determining ranges and antenna gains. The goal of this study is to provide useful information for the design of 60 GHz UWB systems in gigabit M2M communications and standardization. The 60 GHz band is planned and limited on transmit power, and the effective isotropic radiated power (EIRP) and antenna gain for various countries [9] including specifications in China [10] are summarized in **Table 1**.

2 MmWave 60 GHz Propagation Mechanisms

In design and optimization of wireless communications systems, channel models featuring the relevant characteristics of radiowave propagation are required. Ray tracing is a well-established tool for channel modeling. In the ray-tracing algorithm, reflection and diffraction are the main physical processes for LOS and NLOS environments. In our previous works [7], [8], mmWave 60 GHz propagation mechanism is studied based on direction-of-arrival (DOA) measurements. The DOA measurements require the detailed knowledge of the propagation channels. The measured power angle profiles (PAPs) and power delay profiles (PDPs) can then be connected with site-specific information of the measurement environments to find the origin of the arriving of signals. From [8] mmWave 60 GHz propagation mechanism can be concluded as:

- Direct path and the first-order reflected waves from smooth surfaces form the main contributions in LOS propagation environments.
- Diffraction is a significant propagation mechanism in NLOS cases. Moreover, the signal levels of diffraction and second-order reflection are comparable.
- Transmission loss through concrete or brick walls is very high.

The person blocking effect (PBE) is also measured in our previous work [7], as the movement of persons is quite usual in office rooms in reality. The PBE is a major concern for propagation research and system development and PBE at 60 GHz has been studied by many researchers [11], [12]. In [7], PBE is measured by employing DOA measurement techniques as described below.

▼Table 1. 60 GHz band plans and limits on transmit power, EIRP and antenna gain for various countries [9], [10]

Region	Frequency band	TX power (max)	EIRP	Antenna gain	Comments
USA	7 GHz (57 GHz–64 GHz)	500 mW	40 dBm (ave.) 43 dBm (max)	NS	For B>100 MHz, translate average PD from 9 μW/cm² to 18 μW/cm² at 3 m
Canada	7 GHz (57 GHz–64 GHz)	500 mW	40 dBm (ave.) 43 dBm (max)	NS	For B>100 MHz, translate average PD from 9 μW/cm² to 18 μW/cm² at 3 m
Japan	7 GHz (59 GHz–66 GHz) max 2.5 GHz	10 mW	NS	47 dBi (max)	
Australia	3.5 GHz (59.4 GHz–62.9 GHz)	10 mW	150 W (max)	NS	Limited to land and maritime
Korea	7 GHz (57 GHz–64 GHz)	10 mW	TBD	TBD	
Europe	9 GHz (57 GHz–66 GHz) min 50 MHz	20 mW	57 dBm (max)	37 dBi (max)	Recommendation by ETSI
China	5 GHz (59 GHz–64 GHz)	10 mW	44 dBm (ave.) 47 dBm (max)	NS	

EIRP: effective isotropic radiated power
ETSI: European Telecommunications Standards Institute
NS: no specification
PD: power density
TBD: to be determined

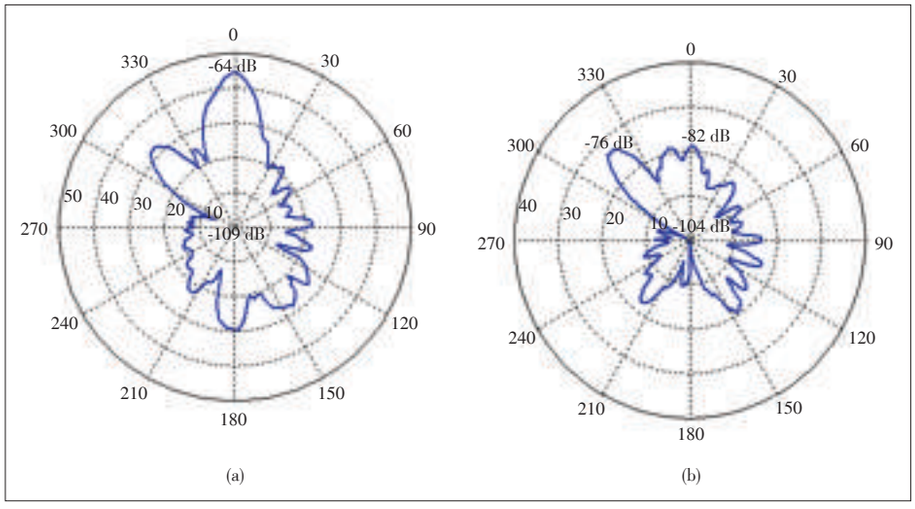
2.1 PBE Measurements

The PBE measurements were performed in a room at Aalto University, Finland, where the TX and RX positions are fixed with 5 m apart. When keeping a clear LOS path and a person blocked in the middle of the LOS path, measuring the PAPs of the clear LOS path and the blocked path, as shown in **Figs. 2a** and **2b**. It is seen that there is about 18 dB person attenuation in the blocked path (φ=0°). However, the PBE can be reduced to 12 dB by using selection diversity technique, i.e., selecting another stronger path (at φ=315°), which is considered as the first-order reflection from window glass in the room [8].

When the LOS path undergoes a deep fading (person blocking), the fading effects can be mitigated by selecting another independent strong signal. This selection diversity is a powerful communication receiver technique for link improvement. Therefore, the effective PBE=12 dB is considered in 60 GHz UWB system parameter analysis of this paper.

2.2 Radiowave Propagation Mechanisms in NLOS Cases

We know that in the LOS propagation environments, direct path and the first-order reflected waves from smooth surfaces form the main contributions of receiving signals [8]. This is also proved in [13] where a two-ray model (LOS path and first-order reflection from desktop) is proposed for 60 GHz M2M sys-



▲Figure 2. The measured power angle profiles of (a) clear LOS path and the (b) the LOS path blocked by a person in person block effect measurements.

tems. In NLOS cases, diffraction is a significant propagation mechanism, and the signal levels of diffraction and second-order reflection are comparable [8]. This indicates that radio links are relayed by direction and/or the second-order reflections in the NLOS propagation scenarios in 60 GHz band. As an example, **Fig. 3** shows the radiowave propagation in office room environment with NLOS scenario. In Fig. 3, the diffraction and second-order reflection rays are denoted by dot and real lines, respectively. It should be noted that in the NLOS case, signal power loss increases greatly with the increasing of distance between the TX and RX. Thus, propagation range is major concern in NLOS environments in system development.

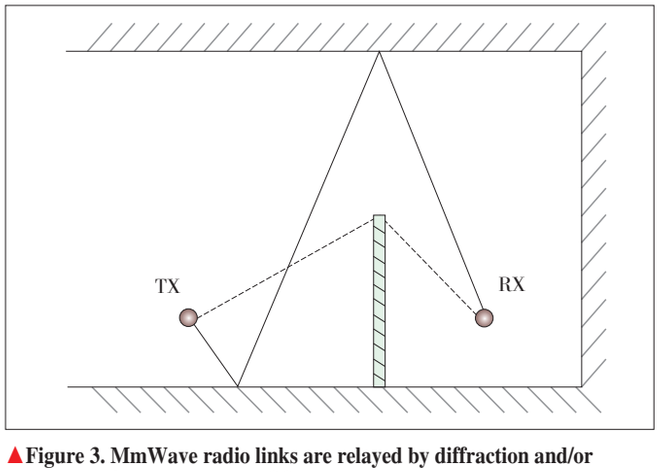
3 Analysis of 60 GHz UWB System Link Budget

In wireless communication systems, the upper bound of capacity is determined by Shannon theorem, which is function of bandwidth B and signal-to-noise ratio (SNR) with expression as:

$$C = B \cdot \log_2(1 + SNR). \quad (1)$$

A system capacity increases with B and SNR. However, increasing of bandwidth will lead to high noise power of system. For example, noise power is 18 dB higher with a UWB B=7 GHz channel than a narrowband B=100 MHz channel (when antenna noise temperature is T=290 K).

In this study, SNR=10 dB and B=1 GHz are considered for performing a basic feasibility study for achieving gigabit capacity of 60 GHz UWB systems.



▲Figure 3. MmWave radio links are relayed by diffraction and/or double-reflection in the NLOS case.

3.1 Parameters for 60 GHz UWB System Link Budget

In wireless communication systems, the performance and robustness is often determined by SNR from radio link budget:

$$SNR = P_t + G_t + G_r - PL - N_0 - IL, \quad (2)$$

where P_t is the transmitted power, G_t and G_r are the transmitter (TX) and receiver (RX) antenna gains, PL denotes path loss in propagation channel, N_0 is the total noise power at RX, and IL denotes the implementation loss of system. P_t is often limited by regulations of radio systems. In this work, it is chosen as $P_t = 10$ dBm, as it was specified by most of countries including China. The other system parameters are set as practical values, i.e., $IL = 6$ dB and noise figure NF = 6 dB in calculating total noise power: $N_0 = 10\log_{10}(kTB) + NF$, where k is Boltzmann's constant and T is the standard noise temperature T =290 K.

3.2 Path Loss Models in 60 GHz UWB Systems

Path loss characterizes channel large-scale fading, which is a key impact on the coverage and reliability of system. Path loss PL denotes the mean signal power loss and usually obeys the power distance law. Due to variations in the propagation environments, the signal power observed at any given pints will deviate from its mean, and this phenomenon is called shadowing. Because of shadowing, a fading margin FM is often considered in system design. Thus path loss PL is modeled as a combination of mean path loss and fading margin FM:

$$PL = \underbrace{PL_0(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right)}_{\text{mean path loss}} + FM, \quad (3)$$

where the free space path loss PL_0 is frequency-dependent, $PL_0 = 68$ dB at reference distance $d_0 = 1$ m, path loss exponent n is environment-dependent, and FM is mainly system-dependent. UWB system naturally leads to shadow fading improve-

ment relative to narrow band systems.

Based on our results that FM decreases with channel bandwidth B, and is less than 4 dB for a minimum bandwidth ($B=500$ MHz) in UWB channel with 90% link success probability [14], fading margin is considered as $FM=2$ dB for the 60 GHz UWB ($B=1$ GHz) system in this work.

The studies show that path loss exponent ranges in 2–3.5 in LOS and NLOS office room environments [8], [13]. In this work, path loss models of LOS ($n=2$), LOS path blocked by moving person and NLOS ($n=3.5$) are considered, and they are $PL_1(dB)=68+20\log_{10}(d)+FM$, $PL_2(dB)=68+20\log_{10}(d)+FM+PBE$, and $PL_3(dB)=68+35\log_{10}(d)+FM$ respectively. The path loss model of LOS+PBE is more feasible when comparing with the NLOS model. Since the blocking effect is modeled independently on mobile position, which reflects a real case that movement of persons is quite typical in office rooms. Whereas, the NLOS model accounts for high path loss due to large distances practically. **Table 2** summarizes the parameters used in the 60 GHz UWB system link budget analysis, in which the maximum range is chosen as 5 m considering office room M2M applications (e.g. computer-to-computer data transfer).

4 Analysis of 60 GHz UWB System Performance

As transmission power is restricted in regulations of 60 GHz radio systems, and further, path loss of the 60 GHz channel is high (e.g. free space path loss at 60 GHz is 22 dB higher than 5 GHz frequency band at $d_0=1$ m), the antenna gains become very important in guaranteeing radio links for achieving gigabit capacity of system. In the following, tests are being performed in order to determine ranges and combined antenna gains (sum of gains at TX and RX), when using the parameters and path loss models of LOS ($n=2$), LOS + PBE and NLOS ($n=3.5$) in Table 2.

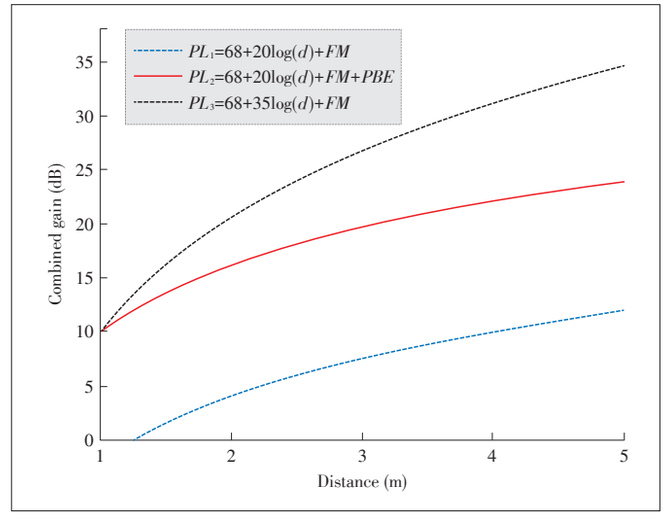
Fig. 4 shows the combined antenna gain vs. distance for the 60 GHz UWB system. It is seen that gigabit capacity can be achieved with omni-directional antennas configuration at the transceiver in LOS condition. However, at further distance of $d=5$, only with the LOS path loss PL_1 model antenna configuration of omni-directional is feasible for gigabit capacity. With another two path loss models of PL_2 (LOS + PBE) and PL_3 (NLOS $n=3.5$), antenna configuration of omni-directional is required for the 60 GHz UWB system.

Note that directional antenna is with high gain, for instance, the half power beam width (HPBW) is approximately 6.5 for an antenna with more than 30 dBi gain [9]. The drawbacks of high gain antenna are that systems suffer from poor flexibility and limited mobility. It should be noted that the path loss model of LOS + PBE in Fig. 4 is more feasible when comparing with the NLOS model, because the blocking effect is modeled independently on mobile position, which reflects the real case that movement of persons is quite typical in office rooms. Whereas, the NLOS model accounts for high path loss due to large distances practically. The results show that it is essential to keep a clear LOS path of 60 GHz UWB systems in gigabit M2M applications.

▼**Table 2. Radio link budget of 60 GHz UWB system**

60 GHz UWB system	
Data rate	> Gbps
Maximum range	5 m
Bandwidth	1 GHz
TX power	10 dBm
SNR	10 dB
Noise power	-78 dBm
Fading margin	2 dB
Implementation loss	6 dB
Effective person block effect	12 dB
Employed path loss models	LOS: $PL_1(dB)=68+20\log_{10}(d)+FM$ LOS + PBE: $PL_2(dB)=68+20\log_{10}(d)+FM+PBE$ NLOS: $PL_3(dB)=68+35\log_{10}(d)+FM$

LOS: line-of-sight PBE: person blocking effect UWB: ultra-wide band
NLOS: non-LOS SNR: signal-to-noise ratio



▲**Figure 4. Combined antenna gains in 60 GHz UWB system with link budget in Table 2.**

5 Conclusions

The feasibility and performance of mmWave 60 GHz UWB systems for gigabit M2M wireless communications are analyzed in this work. Specifically, based on specifications and experimental channel measurements and models for both LOS and NLOS scenarios, the 60 GHz propagation mechanisms are concluded, 60 GHz UWB radio link budget including person block effect and channel fading margin are provided, and system performance is analyzed further. Tests are also performed

for determining communication ranges and antenna configurations. Results show that when having a clear LOS path and employing omni-directional antennas at the transceiver, gigabit capacity can be achieved. When the LOS path is blocked by a moving person or radiowave propagation in NLOS situation, omni-directional and directional antennas at the transceiver are required for achieving gigabit capacity in the range of 5 m between machines in office rooms. The high gain antenna systems suffer from poor flexibility and limited mobility. Therefore, it is essential to keep a clear LOS path in gigabit M2M applications like data transfer in office rooms. The goal of this study is to provide useful information for the design of 60 GHz UWB systems in gigabit M2M communications.

References

[1] C. Gustafson, K. Haneda, S. Wyne, and F. Stufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp.1445–1455, Mar. 2014. doi: 10.1109/TAP.2013.2295836.

[2] X. Song, C. Jans, L. Landau, D. Cvetkovski, and G. Fettweis, "A 60 GHz los MIMO backhaul design combining spatial multiplexing and beamforming for a 100 Gbps throughput," in *Proc. IEEE Global Communications Conference*, San Diego, USA, Dec. 2015. doi: 10.1109/GLOCOM.2015.7417737.

[3] T. S. Rappaport, S. Sun, R. Mayzus, et al., "Millimeter wave mobile communications for 5G cellular: it will work!" *IEEE Access-Practical Innovations/Open Solutions*, vol. 1, pp. 335–349, May. 2013. doi: 10.1109/ACCESS.2013.2260813.

[4] T. Baykas, S. Chin-Sean; L. Zhou, et al., "IEEE 802.15.3c: the first IEEE wireless standard for data rates over 1 Gb/s," *IEEE Communications Magazine*, vol.49, no.7, pp.114–121, Jul. 2011. doi: 10.1109/MCOM.2011.5936164.

[5] X. Peng and L. Zhuo, "The 60 GHz Band Wireless Communications Standardizations (in Chinese)," *Information Technology & Standardization*, pp. 49–53, Dec. 2012.

[6] S. Geng, S. Liu, and X. Zhao, "60-GHz Channel Characteristic Interdependence Investigation for M2M Networks," in *International Conference on Communications and Networking*, Maoming, China, Aug. 2014, pp.402–406. doi: 10.1109/CHINACOM.2014.7054327.

[7] S. Geng, J. Kivinen, X. Zhao, and P. Vainikainen, "Measurements and analysis of wideband indoor radio channels at 60 GHz," in *3rd ESA Workshop on Millimeter Wave Technology and Applications*, Espoo, Finland, May. 2003, pp. 39–44.

[8] S. Geng, J. Kivinen, X. Zhao, and P. Vainikainen, "Millimeter-wave propagation channel characterization for short-range wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 3–13, Jan. 2009. doi: 10.1109/TVT.2008.924990.

[9] S. K. Yong and C. C. Chong, "An overview of multigigabit wireless through millimeter wave technology: potentials and technical challenges," *EURASIP Journal on Wireless Communications and Networking*, no. 1, 2007. doi: 10.1155/2007/78907.

[10] National Radio Administration. (2016, Oct.). *Chinese specifications of the 60 GHz band transmission power for short-range wireless applications (in Chinese)* [Online]. Available: www.miit.gov.cn

[11] M. Jacob, S. Priebe, A. Maltsev, et al., "A ray tracing based stochastic human blockage model for the IEEE 802.11ad 60 GHz channel model," in *Proc. 5th European Conference on Antennas and Propagation (EUCAP)*, Rome, Italy, Apr. 2011, pp.3084–3088.

[12] K. Dong, X. Liao, and S. Zhu, "Link blockage analysis for indoor 60 GHz radio systems," *Electronics Letters* , vol.48, no.23, pp.1506–1508, Nov. 2012. doi: 10.1049/el.2012.2994.

[13] Y. Shoji, H. Sawada, C. Chang-Soon, and H. Ogawa, "A modified SV-model suitable for line-of-sight desktop usage of millimeter-wave WPAN systems," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 10, pp. 2940–2948, Oct. 2009. doi: 10.1109/TAP.2009.2029286.

[14] S. Geng and P. Vainikainen, "Experimental Investigation of the Properties of Multiband UWB Propagation Channels," in *IEEE International Symposium on Wireless Personal Multimedia (PIMRC07)*, Athens, Greek, Sept. 2007. doi: 10.1109/PIMRC.2007.4394334.

Manuscript received: 2016-10-26

Biographies

WANG Qi (qiuqian12390@126.com) received the B.Sc. degree in electronic information technology from North China Electric Power University (NCEPU), China in 2012 and has been a successive postgraduate and doctoral student in electrical engineering and information technology with NCEPU since 2014. Her recent research interests include millimeter wave communications, massive MIMO channel modeling, and human blocking modeling.

GENG Suiyan (gsuiyan@ncepu.edu.cn) received the M.Sc. (Tech.) and Ph.D. degrees in 2003 and 2011 from the Helsinki University of Technology (TKK), Finland. From 1992 to 1998, she was a research engineer with the China Research Institute of Radiowave Propagation, China. From 2001 to 2011, she was a research engineer with the Radio Laboratory (Department of Radio Science and Engineering since the beginning of 2008), TKK. She is now an associate professor at North China Electric Power University, China. Her research topics include millimeter-wave and ultra-wideband radio wave propagation and stochastic channel modeling for future-generation radio systems and technologies.

ZHAO Xiongwen (huadian_zhaoxw@126.com) received his Ph.D. degree in 2002 with high honors from Helsinki University of Technology, Finland. He is now a full professor in wireless communications at North China Electric Power University, China and chairs several projects by the National Science Foundation of China, the State Key Laboratories and Industries on channel measurements, modeling and simulations. He is a reviewer of IEEE transactions, journals, letters, and conferences. He was a recipient of IEEE Vehicular Technology Society (VTS) Neal Shepherd Best Propagation Paper Award in 2014. He has served as the TPC members, session chairs, and a keynote speaker for numerous international and national Conferences. He is a senior member of IEEE.

HONG Wei (weihong@seu.edu.cn) received the B.S. degree from the University of Information Engineering, China in 1982, and the M.S. and Ph.D. degrees from Southeast University, China in 1985 and 1988, respectively, all in radio engineering. He is currently a professor and the dean of the School of Information Science and Engineering, Southeast University. He twice awarded the National Natrual Prizes (second and fourth class), thrice awarded the first-class Science and Technology Progress Prizes issued by the Ministry of Education of China and Jiangsu Province Government. He also received the foundations for China Distinguished Young Investigators and for "Innovation Group" issued by the National Science Foundation of China. Dr. HONG is Fellow of IEEE, Fellow of CIE, MTT-S AdCom Member (2014-2016), Vice-Presidents of Microwave Society and Antenna Society of CIE, and Chairperson of IEEE MTT-S/AP-S/EMC-S Joint Nanjing Chapter. He was an associate editor of *IEEE Transactions on MTT* during 2007–2010 and is the editor board members for *IJAP*, *China Communications*, *Chinese Science Bulletin*, etc.

Katsuyuki Haneda (katsuyuki.haneda@aalto.fi) received the D. Eng. degree from the Tokyo Institute of Technology, Japan in 2007. He is currently an assistant professor with the School of Electrical Engineering, Aalto University, Finland. His current research interests include high-frequency radios, such as millimeter wave and beyond, wireless for medical and post disaster scenarios, and in-band full-duplex radio technologies. Dr. Haneda was an active member of a number of European COST Actions, e.g., IC1004 "Cooperative Radio Communications for Green Smart Environments" and CA15104 "Inclusive Radio Communication Networks for 5G and beyond." He was a recipient of the Best Paper Award of the antennas and propagation track in the IEEE 77th Vehicular Technology Conference, Dresden, Germany, in 2013, and the Best Propagation Paper Award in the 7th European Conference on Antennas and Propagation, Gothenburg, Sweden, in 2013. He has been an associate editor of *IEEE Transactions on Antennas and Propagation* since 2012, and an editor of *IEEE Transactions on Wireless Communications* since 2013.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 005
http://www.cnki.net/kcms/detail/34.1294.TN.20170216.0933.002.html, published online February 16, 2017

Measurement-Based Spatial-Consistent Channel Modeling Involving Clusters of Scatterers

YIN Xuefeng[1], ZHANG Nan[2], Stephen Wang[3], and CHENG Xiang[4]
(1. Tongji University, Shanghai 201804, China;
2. ZTE Corporation, Shanghai 201203, China;
3. Ocado Technology, AL10 9NE, The United Kingdom;
4. Peking University, Beijing 100871, China)

Abstract

In this paper, the conventional method of establishing spatial channel models (SCMs) based on measurements is extended by including clusters- of- scatterers (CoSs) that exist along propagation paths. The channel models resulted utilizing this new method are applicable for generating channel realizations of reasonable spatial consistency, which is required for designing techniques and systems of the fifth generation wireless communications. The scatterers' locations are estimated from channel measurement data obtained using large- scale antenna arrays through the Space- Alternating Generalized Expectation- Maximization (SAGE) algorithm derived under a spherical wavefront assumption. The stochastic properties of CoSs extracted from real measurement data in an indoor environment are presented.

Keywords

channel modeling; large-scale antenna array; spatial consistency; space-alternating generalized expectation-maximization (SAGE) algorithm; spherical wavefront

1 Introduction

Channel modeling based on extensive measurement campaigns is important for characterizing wave propagations in real environments [1]. High-resolution channel estimation algorithms, such as the Space - Alternating Generalized Expectation - Maximization (SAGE) [2], [3] and Richter's Maximum Likelihood Estimation (RiMAX) algorithms [4], are widely adopted for processing channel measurement data. These algorithms are derived from generic parametric multipath models under the specular-path assumption [5], i.e. the dominant components in a channel are contributed by electromagnetic waves propagating along specular paths with planar wavefront when impinging the antenna arrays in the transmitter (Tx) and the receiver (Rx) [2], [6], [7]. An individual path can be described with the delay, direction (including azimuth and elevation) of departure (DoD), direction of arrival (DoA), complex-valued 2x2 dual-polarization matrix, and non-zero Doppler frequency in the case where the measurement environment is non-static [8], [9]. By grouping the paths in a channel as clusters [10], spatial channel models (SCMs) are established based on the statistics of path clusters. Typical SCMs are those specified in 3GPP TR25.996 [11], the WINNER II SCM-enhanced model [12], IMT-Advanced models [13], and COST 2100 models [14], [15].

Recently, channel modeling for designing the fifth generation (5G) wireless communication techniques and systems has been paid much attention [3], [16]–[19]. Spatial channels are exploited unprecedentedly in numerous 5G transmission techniques, such as the massive multiple-input multiple-output (MIMO) system using large-scale antenna arrays. The so-called "spatial consistency" property arises as a new characteristics which has not been effectively modeled in the conventional SCMs. Spatial consistency refers to the inherent relationship of channels observed by a user equipment (UE) in consecutive snapshots or drops when the UE moves in a clutter environment [17, Sect. 2.1]. In order to generate channels with reasonable spatial consistency, a 5G channel model needs to take into account the distribution of scatterers in the environment, a property yet considered in the conventional SCM modeling. To

This work is jointly supported by the key project "5G Ka frequency bands and higher and lower frequency band cooperative trail system research and development" of China Ministry of Industry and Information Technology under Grant number 2016ZX03001015 and by the HongKong, Macao and Taiwan Science & Technology Cooperation Program of China under Grant No. 2014DFT10290.

construct models of spatial consistency, the project "Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society (METIS)" has been proposed to include the geometric locations of the scatterers involved in the first and last hops of each path in SCMs [17]. Such a modeling strategy was adopted for establishing the initial map-based ray-tracing model, where the locations of scatterers are predicted by using a simple geometry-based method from the map of an environment. However, the map-based ray-tracing models are too site-specific to generate channel realizations in general cases. It is necessary to build measurement-based models that take into account the scatterers involved along paths estimated through channel sounding in real environments.

Recently, an algorithm of scatterer localization dedicated for MIMO channel sounding was introduced in [20]. This algorithm was derived under a spherical wavefront assumption and applicable for estimating the scatterers' locations along propagation paths. More specifically, this algorithm estimates two more distance parameters in addition to the conventional DoD, DoA, delay and Doppler frequency. These new parameters are the distance between the center of the Tx antenna array to the scatterer in the first hop, and the distance between the center of the Rx antenna array to the scatterer in the last hop of a path. The estimates of these distance parameters can be used together with the DoD and DoA to determine the location of the scatterers involved in the first and last hops of the path respectively. Experimental results illustrated in [20] clearly demonstrate the effectiveness of this algorithm applied to processing measurement data collected using large-scale antenna arrays.

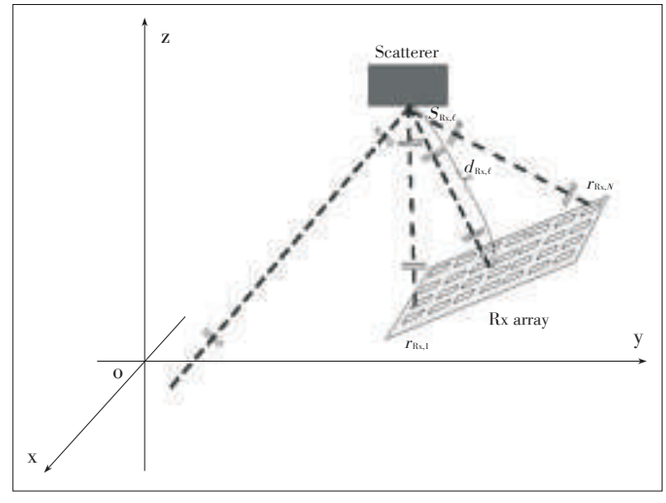
In this paper, we illustrate for the first time, an example of establishing channel models involving cluster of scatterers. We adapt the algorithm derived in [20] for estimating the scatterers' locations from single-input multiple-output (SIMO) channel measurements. A large-scale antenna array of 121 elements is used in the receiver side. The clusters of scatterers (CoSs) are extracted and their statistical properties are discussed. This knowledge can be incorporated into the frameworks of SCMs to associate each group of paths with two sets of scatterers located respectively in the first and last hops of the paths. The novelty of this work lies in the following aspects: 1) a new framework of SCMs for reproducing spatial-consistency by including the statistics of CoSs; and 2) an experimental example illustrating the procedure of establishing the CoS-included channel models. The enhanced SCMs resulted are applicable of generating non-stationary channel realizations for many 5G propagation scenarios, such as in the cases of massive MIMO, and of device to device communications with link ends surrounded by scatterers in their proximity.

The rest of this paper is organized as follows. Section 2 briefly reviews the scatterer localization algorithm derived based on the spherical wavefront assumption. In Section 3, the modeling procedure of including CoSs into SCMs is introduced. Section 4 elaborates the behavior of CoSs extracted from real measurements using a large-scale Rx antenna array in indoor environments. Finally, conclusive remarks are addressed in Section 5.

2 Spherical-Wavefront-Based Scatterer Localization Algorithm

The planar wave assumption has been widely used for deriving high-resolution channel parameter estimation algorithms. This assumption is only valid when the scatterer deviates from an antenna array larger than the Rayleigh distance $d_{\text{Rayleigh}}=2D^2/\lambda$, where D in meters is the aperture of the array and λ represents the wavelength of the carrier [21, Chapter 2.2.3]. In the scenarios where an antenna array has a sufficiently large aperture, the distance from the array to a scatterer may be less than d_{Rayleigh}. As a result, the planar wavefront assumption is inappropriate because the wavefront observed at the locations of individual antennas is non-uniform.

Fig. 1 illustrates a diagram of a wave incident to the Rx array when the scatterer that interacts with the wave deviates from the array center with a distance less than d_{Rayleigh}. It should be noticed that the planar array plotted in Fig. 1 is only used as an example. The antenna array has various configurations, such as spherical or cubic. The analysis in the following is conducted regardless of the exact shape of the array. It is obvious from Fig. 1 that when the scatterer is viewed as a point, the DoA of the ℓth path observed by the nth Rx antenna is determined by the antenna's location $r_{\text{Rx},n}$ and the location $s_{\text{Rx},l}$ of the scatterer involved in the last hop of the path. In another word, the DoAs of the ℓth path observed at all Rx antennas coincide with the radial directions of the antenna positions in a spherical coordinate system with the origin located at $s_{\text{Rx},l}$. Similar effect can be observed for the Tx antenna array in the case where the scatterer involved in the first hop of a path is so close to the Tx antenna array that the DoD observed at a Tx an-

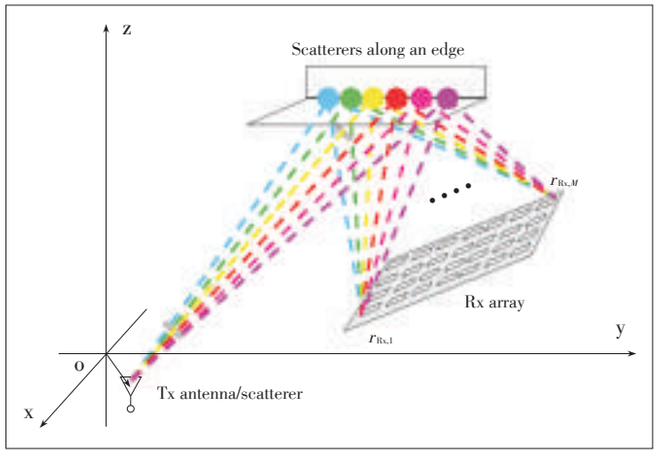


▲**Figure 1. Diagram of a wave impinging to the Rx array with a spherical wavefront when the distance $d_{\text{Rx},l} < d_{\text{Rayleigh}}$. A similar diagram can be made for the Tx array side.**

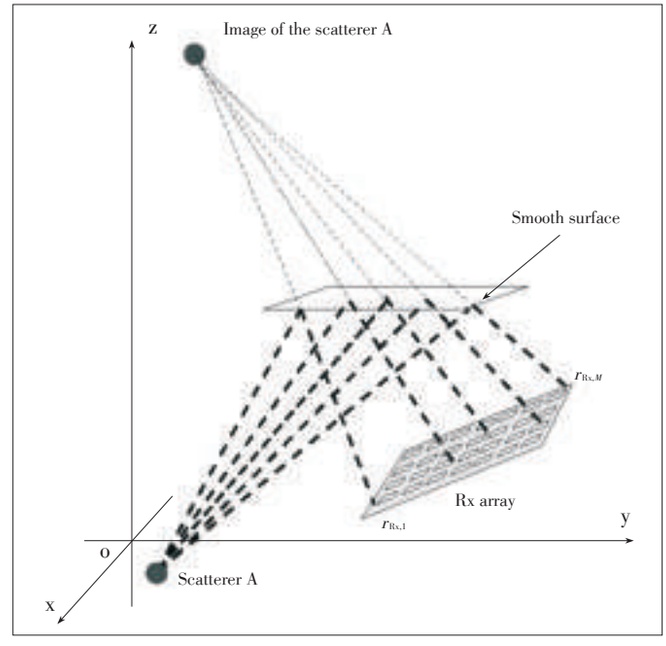
tenna is determined by the position $r_{\mathrm{Tx},m}$ of the Tx antenna and the location $s_{\mathrm{Tx},l}$ of the scatterer. The phenomenon of wave propagating with spherical wavefronts is often observed in the MIMO channel obtained by using large-scale antenna arrays. Since a planar wave can be viewed as a special case of waves with spherical wavefront when the distance between the scatterer emitting the wave and an antenna array is larger than d_{Rayleigh}, a new spherical-wavefront generic multipath model can be proposed and applied for parameter estimation of all paths in the channel [20]. The advantage of applying spherical-wavefront model is that the estimated parameters of a path can be used to localize the 3-dimensional (3D) locations of scatterers involved in the first and last hops of the paths.

The spherical wavefront assumption relies on the condition that the scatterer is a point source that has infinitesimal small volume. Thus, localization of a scatterer by using parameter estimates obtained under a spherical wavefront is accurate in the case where the physical extent of the scatterer is sufficiently small. However, this is not usually the case in reality. For example, in the scenario where scatterers are spaced closely along an edge as depicted in **Fig. 2**, due to the limited resolution of the measurement system in both direction and range [22], it is difficult to resolve these scatterers individually. Thus, the estimated 3D-location of a scatterer under the spherical wavefront assumption may not correspond to the exact location of any of these scatterers. Furthermore, the estimation result could be varying from snapshot to snapshot if the measurements involve random variations of the environment. In the case where the estimated locations of scatterers are close enough, they can be grouped as a CoS and used to describe the distribution of the scatterers along the paths which exhibit similar parameters.

It is worth mentioning that in the case where the true scatterers involved in the first or last hop of a path are located on a smooth surface where specular reflections occur, the localization of a scatterer under the spherical wavefront assumption can be erroneous significantly. **Fig. 3** illustrates an example of



▲**Figure 2. Multiple scatterers located along an edge in a small region.**



▲**Figure 3. A smooth surface introducing specular reflection for the impinging wave.**

such a scenario, where every antenna in the Rx array receives the wave reflected by different points on the smooth surface. The estimated 3D location of the scatterer under the spherical wavefront assumption corresponds to the image of the point scatterer with the smooth surface being the mirror. Although the scatterer localization does not provide correct results in such a case, from propagation point of view, the estimated scatterer still provides the contribution equivalent with the true scatterer to the propagation channel. Thus, the CoS formed by the estimated image-scatterers remains important for being considered in channel characterization.

3 SCMs Enhanced by Incorporating CoSs

The positions of scatterers in an environment can be used to predict channel characteristics via ray-tracing [23] or graph-modeling [24]. Similarly, the spatial consistency of channels observed in consecutive snapshots when a UE moves, or the non-stationarity of channel observed from different antennas in massive MIMO scenarios, can be reproduced based on the distribution of scatterers in the environment and their relationships with the propagation paths. In order to generate non-stationary channels with reasonable spatial consistency, the conventional SCMs established with multipath clusters need to incorporate the stochastic characteristics of the scatterers, or CoSs in the first and last hops of paths.

Parameterizing CoS-incorporated SCMs requires localization of the scatterers along multipath. This cannot be effectively performed through conventional MIMO channel measurements that adopt antenna arrays with small apertures, usually less

than 3 to 4λ [10, Ch. 9.2]. In many recently conducted measurement campaigns for massive MIMO channel characterization [25], [26], large-scale antenna arrays with apertures about dozens of wavelengths are applied. This facilitates the usage of spherical-wavefront estimation algorithms, like that proposed in [20] for extracting the locations of the scatterers, and then incorporates the characteristics of CoSs to SCMs. Some proposals are raised in the following for the exact procedure of including the stochastic behaviors of CoSs into SCMs.

The procedure of establishing an SCM incorporating the characteristics of CoSs is similar with that adopted for generating the standard SCMs [10], [12]. First, the measurement campaigns that make use of large-scale antenna arrays are conducted. In order to create sufficient randomness in the observations, the scatterers in the measurement environment need to be slightly moved in small ranges in measurement snapshots. Then, the measurement data is processed by using the channel parameter estimation algorithm derived under the spherical wavefront assumption, e.g. the SAGE algorithm [20]. The estimated paths are then clustered based on the parameter estimates that include the conventional geometrical parameters and the distances from the center of Tx array and of Rx array to the scatterers involved in the first and last hops of the path as well. Once the multipath clusters are obtained, the locations of the scatterers for the paths assigned in the same cluster are calculated, and the CoSs involved in the first and last hops of paths respectively are obtained. Based on the characteristics of a large number of estimated CoSs extracted from measurement data, the statistical parameters characterizing the stochastic behaviors of the first-hop and last-hop CoSs, e.g. the 3D extent of CoS and the distributions of scatterers in CoS, are calculated and included as parts of the CoS-incorporated SCMs.

When using the CoS-incorporated SCMs to create random channel realizations, the clusters of paths are first generated randomly. For a cluster, the values of distances between the center of the Tx antenna array and scatterers involved in the first hops of these paths are then generated based on the statistics of the first-hop CoSs. By assigning these distances to the paths in the cluster randomly, the positions of the scatterers in the first hops of paths can be calculated. Similar approaches can be applied to find the positions of the scatterers involved in the last hops of the paths. When the UE moves, or different Tx or Rx antennas in a large-scale array are selected, the corresponding channel is calculated based on the path parameters updated with the new positions of the UE, or the positions of the Tx and Rx antennas in the array respectively.

4 Experimental Results for Characteristics of CoSs

Recently, a measurement campaign for characterizing CoSs has been conducted in an office environment in Toshiba Research Europe ltd., Bristol, UK. The Toshiba Medav ultra-wideband (UWB) sliding-correlator channel sounder was applied in the measurements. A SIMO configuration was considered with the Tx equipped with a UWB biconical omnidirectional antenna and the Rx with a 121-element virtual planar antenna array. The sounding signal is generated by using a pseudo-noise m-sequence of 4095 chips, with bandwidth B = 500 MHz and center frequency of 9.5 GHz. The Rx antenna is an omni-directional biconical antenna, which is attached to an extendible fiberglass mast mounted with the movable part of a high-accuracy positioner. The Rx antenna moves along the vertices of 11 × 11 grids to form a virtual antenna array with positioning errors less than 1.2 μm. Each grid is square with edge on one side equal to half the wavelength of the wave with carrier frequency of 9.5 GHz. The vibration caused by relocating the antenna is minimized by calibrating the acceleration and de-acceleration of the motors. After the positioner reaches the desired location, the sounder waits for an additional period of approximately 1 s before measuring. This delay allows any residual vibration to dissipate before capturing CIRs. During the measurements, the environment was kept stationary, the Tx antenna was fixed with height of 1.2 m above the office floor, and the virtual Rx antenna array was located in a horizontal plane which is 1.8 m above the floor.

The measurements were conducted in a room which has the dropped ceiling made of polystyrene faux tile. The floor is a metallic structure covered with carpet tiles. The room has two glass walls, two slat walls, and some office furniture including metallic bench desks and shelves. **Fig. 4a** illustrates a photograph of the Rx antenna and the positioner in the office, and **Fig. 4b** shows the photograph of both the Tx and the Rx antennas.

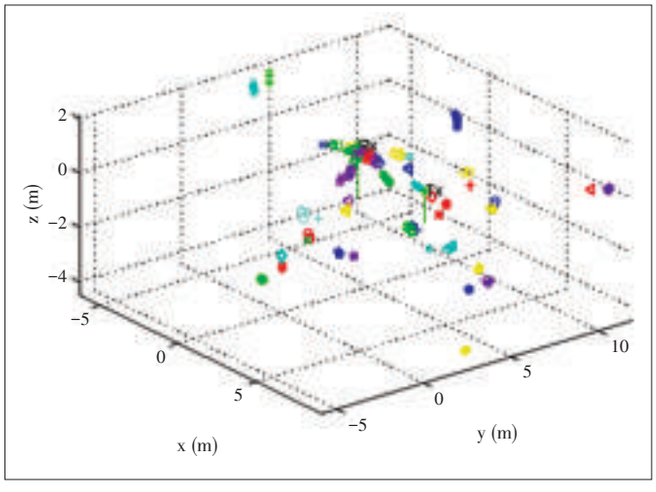
The SAGE algorithm derived in [20] was adapted to the SIMO case considered here. Thus, only the scatterers involved in the last hops of paths can be localized. Furthermore, as the Rx array is located horizontally, the estimation algorithm cannot distinguish whether the waves arrive above or below the plane. In such a case, the estimation range of the elevation of arrival (EoA) is limited to $[\ 90^o, 180^o\ ]$, i.e. below the array plane. However, it is likely that the paths with true EoA within $[\ 0^o, 90^o\ ]$ are estimated with an EoA symmetric to the true angle with respect to the array plane. The azimuth of arrival estimation range is set to $[\ -180^o, 180^o\ ]$. The number L of paths to be estimated is set to 40. It should be noticed that the number of paths can be estimated by applying the Akaike Information Criterion or Minimum Length Description principle [27]. In this preliminary study, we choose to set a large value for L as we usually did in many other works on high-resolution path parameter estimation [2], [28]. Furthermore, only vertical polarization is considered for estimating the complex attenuations of paths as both Tx and Rx antennas are vertically polarized in the measurements.

Since the SAGE algorithm is applied to estimate L = 40 paths per measurement snapshot, from totally 10 measurement

(a) View from the Tx (b) View from the Rx

▲Figure 4. The indoor environment where the measurements were conducted.

snapshots, we obtain the parameter estimates of 400 paths. During these measurement snapshots, the locations of the Tx antenna and Rx array are kept fixed, and the environment is unchanged. The randomness in each snapshot is generated by the thermal noise in the measurement equipment. Strictly speaking, the randomness should be generated by slightly changing the environment. A clustering algorithm similar to that introduced in [29] is applied to gather the paths as clusters. Totally 53 CoSs are found. **Fig. 5** depicts the locations of scatterers calculated based on the multipath parameter estimates. The spots plotted with same color and same legend marks in Fig. 5 represent the scatterers in the same CoS. These CoSs may not correspond exactly to the true scatterers involved in the last hops of paths due to specular reflections caused by smooth surfaces existing in the environment, as discussed in Section 2. Thus, it is possible that the distance between a CoS and the Rx array is larger than the physical size of the office where the measurements were conducted. From channel modeling point of view, all the CoSs extracted from the data need to be considered for constructing the stochastic channel model regardless whether the location of an estimated CoS coincides with real scatterers in the environment.

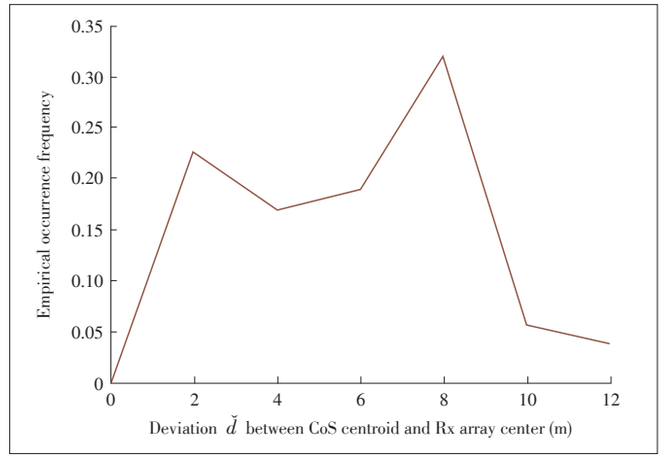


▲Figure 5. The results of CoSs obtained from estimation results.

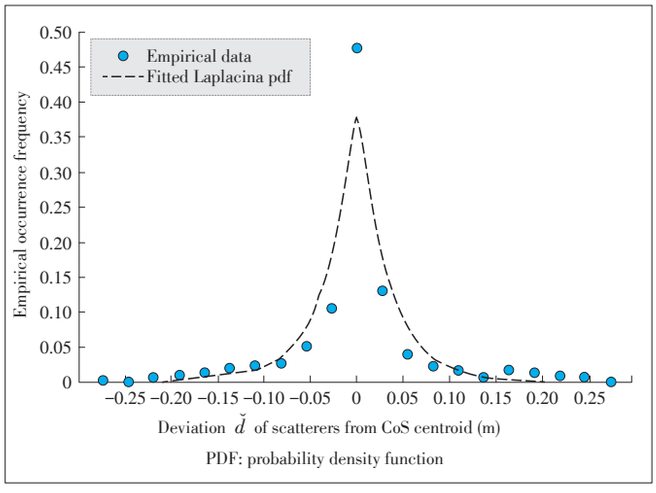
Based on the parameter estimates obtained from multiple snap-shots, the so-called inter-and intra-CoS statistical properties are extracted. The inter-CoS property is referred to as the distribution of the CoSs' centroid in a 3D coordinate system. In our case where the sounding signal has 500 MHz bandwidth and center frequency of 9.5 GHz, the estimated CoSs involved in the last hops of paths are distributed within the 3D volume of 15 m ×15 m ×6 m. It is obvious that these dimensions are larger than the extents of the office. We postulate that this is because some CoSs are actually the images of real scatterers in the office due to reflections on the smooth walls.

Another inter-CoS property investigated is the empirical distribution of the distance $\tilde{d}$ between the CoS centroid and the center of Rx antenna array. **Fig. 6** illustrates the empirical occurrence frequency (EOF) of the distance $\tilde{d}$. Fig. 6 shows that $\tilde{d}$ is distributed with two local maxima located at 2 m and 8 m respectively. This, according to our conjecture, is due to the reason that some scatterers in the environment exist on the ceiling and around the positioner in the vicinity of the Rx array and others are mainly distributed on the outskirts of the office. The former scatterers create the concentration of $\tilde{d}$ around 2 m, and the latter lead to the higher occurrence frequency at 8 m. It is worth mentioning that the EOF of $\tilde{d}$ observed in these measurements seem to justify the assumption adopted in some channel simulations based on sum-of-sinusoids (SoS), i.e. the scatterers are distributed along multiple rings centered at the ends of a communication link [30], [31].

An intra-CoS property considered here is the average empirical EOF of the deviation $\check{d}$ of the scatterers in a cluster from the centroid of the cluster. **Fig. 7** depicts the empirical EOFs



▲Figure 6. Empirical occurrence frequency of the distances of the centers of the CoSs to the center of the Rx antenna array.



▲**Figure 7. Empirical occurrence frequency of the deviation $\check{d}$ of scatterers from CoS centroid in two scenarios. Laplacian PDFs are fitted with the empirical data.**

of $\check{d}$ obtained in two scenarios. The probability density function (PDF) of a Laplacian distribution fitted to the empirical data is also demonstrated in Fig. 7. It can be observed that the EOF graph is symmetric with respect to $\check{d}=0$ m. The Laplacian PDF with position parameter $\mu=0$ m and scaling parameter $b=3.24$ exhibits certain deviations from the EoF. We postulate that this is because of the insufficient number of samples collected in the study. From Fig. 7 we conclude that the scatterers in a CoS are distributed symmetrically with respect to the CoS's centroid with standard deviation of $\sqrt{2b}\approx 4$ cm in the office environment considered.

5 Conclusions

In this paper, a new strategy of incorporating the locations of scatterers into SCM was proposed which is applicable provided the channel measurements are conducted with large-scale antenna arrays. In this scheme, a spherical wavefront parameter estimation algorithm is adopted to extract the distance from the center of transmitting antenna array to the scatterers involved in the first hop of the path, and the distance from the center of receiving antenna array to the scatterers involved in the last hop of the path. These two distance parameters together with direction of departure and direction of arrival can be used to localize scatterers at the first and last hops of the path respectively. The SCM incorporating the statistical characteristics of CoSs is applicable for generating channel realizations in multiple drops with demanded spatial consistency when an user equipment moves, or for non-stationary spatial channels observed through different antennas in massive MIMO scenarios. Channel measurement data were collected by using a Toshiba ultra-wideband sounder equipped an 11×11 virtual receiver antenna array and sounding signals of 500 MHz bandwidth at center frequency of 9.5 GHz. These data have been applied to illustrate the modeling procedure for the characteristics of CoSs. Based on the parameter estimates obtained from multiple snapshots, the inter-CoS and intra-CoS statistical properties were extracted, including the distribution of the centroid of CoSs, and the distribution of scatterers in individual CoSs for the indoor environment considered in the measurements. This work can supply significant guidelines for conducting measurements and modeling channels aiming at reproducing the spatial-consistency by including the statistical information of environments.

References

[1] E. Bonek, M. Herdin, W. Weichselberger, and H. Özcelik, "MIMO—study propagation first!" in *3rd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Darmstadt, Germany, Dec. 2003. pp. 150–153, doi: 10.1109/ISSPIT.2003.1341082.
[2] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. L. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal of Selected Areas in Communications*, vol. 17, no. 3, pp. 434–450, Mar. 1999. doi: 10.1109/49.753729.
[3] X. Yin, C. Ling, and M. Kim, "Experimental multipath-cluster characteristics of 28-GHz propagation channel," *IEEE Access*, vol. 3, pp. 3138–3150, 2016. doi: 10.1109/ACCESS.2016.2517400.
[4] A. Richter, M. Landmann, and R. S. Thomä, "Maximum likelihood channel parameter estimation from multidimensional channel sounding measurements," in *57th IEEE Vehicular Technology Conference (VTC)*, Jeju, South Korea, 2003, pp. 1056–1060. doi: 10.1109/VETECS.2003.1207788.
[5] J. Fuhl, J.-P. Rossi, and E. Bonek, "High-resolution 3-D direction-of-arrival determination for urban mobile radio," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 4, pp. 672 –682, 1997. doi: 10.1109/8.564093.
[6] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques—ESPRIT," in *19th Asilomar Conference on Circuits, Systems and Computers*, Pacific Grove, USA, Nov. 1985, pp. 83– 89. doi: 10.1109/ACSSC.1985.671426.
[7] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989. doi: 10.1109/29.32276.
[8] X. Yin, B. H. Fleury, P. Jourdan, and A. Stucki, "Polarization estimation of individual propagation paths using the SAGE algorithm," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Beijing, China, Sep. 2003. pp. 1795–1799, doi: 10.1109/PIMRC.2003.1260424.
[9] B. H. Fleury, X. Yin, K. G. Rohbrandt, P. Jourdan, and A. Stucki, "High-resolution bidirection estimation based on the sage algorithm: Experience gathered from field experiments," in *XXVIIth General Assembly of the Int. Union of Radio Scientists (URSI)*, Maastricht, Netherlands, vol. 2127, 2002.
[10] N. Czink, "The random-cluster model -a stochastic MIMO channel model for broadband wireless communication systems of the 3rd generation and beyond," Ph.D. dissertation, Technische Universitat Wien, Vienna, Austria, FTW Dissertation Series, 2007.
[11] *Universal Mobile Telecommunications System (UMTS);Spacial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, 3GPP TR 25.996 version 8.0.0 Release 8, 2008.
[12] *WINNER II Interim Channel Models*, IST-4-027756 WINNER D1.1.1 Std., 2007.
[13] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced (12/2009)*, ITU-R M.2135-1 Std, 2009.
[14] L. Liu, C. Oestges, J. Poutanen, et al., "The COST 2100 MIMO channel model," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 92–99, Dec. 2012. doi: 10.1109/MWC.2012.6393523.
[15] M. Zhu, G. Eriksson, and F. Tufvesson, "The COST 2100 channel model: parameterization and validation based on outdoor MIMO measurements at 300

MHz," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–10, 2013. doi: 10.1109/TWC.2013.010413.120620.
[16] J. Medbo, K. Börner, K. Haneda, et al., "Channel Modelling for the Fifth Generation Mobile Communications," in *Eighth European Conference on Antennas and Propagation (EuCAP)*, Hague, Holland, Apr. 2014, pp. 1–5. doi: 10.1109/EuCAP.2014.6901730.
[17] T. Jämsä, P. Kyösti, and K. Kusume, "Deliverable D1.2 Initial channel models based on measurements," Project Name: Scenarios, requirements and KPIs for 5G mobile and wireless system (METIS) , Document Number: ICT-317669-METIS/D1.2, Tech. Rep., 2014. [Online]. Available: www.metis2020.com
[18] Y. Ji, X. Yin, H. Wang, X. Lu, and C. Cao, "Antenna-de-embedded characterization for 13-17 ghz wave propagation in indoor environments," *IEEE Antennas and Wireless Propagation Letters*, vol. PP, no. 99, pp. 1–1, 2016. doi: 10.1109/LAWP.2016.2553455.
[19] X. Yin, L. Ouyang, and H. Wang, "Performance comparison of sage and music for channel estimation in direction-scan measurements," *IEEE Access*, vol. 4, pp. 1163–1174, 2016. doi: 10.1109/ACCESS.2016.2544341.
[20] J. Chen, S. Wang, and X. Yin, "A spherical-wavefront-based scatterer localization algorithm using large-scale antenna arrays," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1796– 1799, Sept. 2016. doi: 10.1109/LCOMM.2016.2585478.
[21] C. A. Balanis, *Antenna Theory Analysis and Design*, 3rd ed. Hoboken, USA: John Wiley & Sons, 2005.
[22] X. Yin and X. Cheng, *Propagation Channel Characterization, Parameter Estimation and Modeling for Wireless Communications*. Hoboken, USA: John Wiley & Sons, 2016.
[23] V. Degli-Esposti, D. Guiducci, A. de'Marsi, P. Azzi, and F. Fuschini, "An advanced field prediction model including diffuse scattering," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 7, pp. 1717– 1728, 2004. doi: 10.1109/TAP.2004.831299.
[24] T. Pedersen, G. Steinbock, and B. Fleury, "Modeling of reverberant radio channels using propagation graphs," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 12, pp. 5978–5988, 2012. doi: 10.1109/TAP.2012.2214192.
[25] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *IEEE 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, USA, 2012, pp. 295–299. doi: 10.1109/ACSSC.2012.6489010.
[26] X. Gao, F. Tufvesson, and O. Edfors, "Massive MIMO channels-Measurements and models," in *IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, 2013, pp. 280–284. doi: 10.1109/ACSSC.2013.6810277.
[27] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974. doi: 10.1109/TAC.1974.1100705.
[28] C.-C. Chong, C.-M. Tan, D. Laurenson, et al., "A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 139– 150, 2003. doi: 10.1109/JSAC.2002.807347.
[29] N. Czink, P. Cera, J. Salo, et al., "A framework for automatic clustering of parametric MIMO channel data including path powers," in *IEEE 64th Vehicular Technology Conference*, Montreal, Canada, Sept. 2006, pp. 1–5. doi: 10.1109/VTCF.2006.35.
[30] X. Cheng, C.-X. Wang, H. Wang, et al., "Cooperative MIMO channel modeling and multi-link spatial correlation properties," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 388– 396, Feb. 2012. doi: 10.1109/JSAC.2012.120218.
[31] Y. Wang and A. Zoubir, "Some new techniques of localization of spatially distributed sources," in *Forty-First IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, USA, 2007, pp. 1807– 1811. doi: 10.1109/ACSSC.2007.4487546.

Manuscript received: 2016-11-27

Biographies

YIN Xuefeng (yinxuefeng@tongji.edu.cn) received his Bachelor's degree in optoelectronics engineering from Huazhong University of Science and Technology, China, in 1995, and his M.Sc. degree in digital communications and Ph.D. in wireless communications from Aalborg University, Denmark, in 2002 and 2006, respectively. From 2006 to 2008, he worked as an assistant professor in Aalborg University. In 2008, he joined Tongji University as an associate professor in the College of Electronics and Information Engineering, China. Since 2017, he was promoted to a full professor and served as the vice dean for the college. His research interests include high-resolution parameter estimation for propagation channels, channel characterization and stochastic modeling for 5G wireless communications, radar signal processing and target recognition. He has published about 100 technical papers and co-authored the book "Propagation Channel Characterization, Parameter Estimation and Modeling for Wireless Communications" published by John Wiley and Sons Edition in 2016.

ZHANG Nan (zhang.nan152@zte.com.cn) received the Bachelor's degree in communication engineering and the Master's degree in integrated circuit engineering from Tongji University, China in July 2012 and March 2015, respectively. He is now a senior engineer at the Department of Algorithms, ZTE Corporation. His current research interests are in the field of 5G channel modeling and new air-interface.

Stephen Wang (stephen.wang@ocado.com) received an M.Sc. (distinction) in advanced photonics and communications from the University of Warwick, UK in 2005. He served as a wireless engineer in China Telecom and Nokia, respectively before obtained his Ph.D. in cognitive radio and radar systems from University of Bristol, UK in 2009. He then joined Shanghai Research Center for Wireless Communications as a senior researcher engineer and team leader in a collaborative research project with Nokia/NSN, coordinate related R&D activities in 3GPP LTE-HeNB standardization activities. From 2010–2016, he was with Toshiba Research Europe Limited, Telecommunications Research Laboratory, as a senior research engineer leading several research projects in mobile body area networks and 5G massive MIMO. From 2017, as an IT team leader in Ocado Technology, he has been leading wireless IoT technology and business for Ocado's IoT-enabled fulfillment centre and logistics. He is an inventor/co-inventor of 10+ patents. His research interests include industrial IoT, water resource/meter management and optimization, cognitive radio, energy-efficient resource optimization, channel propagation and modeling, and wearable healthcare applications. He is a regular reviewer of IEEE journals, guest editor and session chairs of flagship conferences, general co-chair of Wireless Internet: 8th International Conference (WICON 2014). He is a senior member of IEEE and a UK chartered engineer.

CHENG Xiang (xiangcheng@pku.edu.cn) received the Ph.D. degree from Heriot-Watt University and the University of Edinburgh, UK in 2009, where he received the Postgraduate Research Thesis Prize. He has been with Peking University, China since 2010, first as a lecturer, and then as an associate professor since 2012. His current research interests include mobile propagation channel modeling and simulation, next generation mobile cellular systems, intelligent transportation systems, and hardware prototype development. He has published more than 100 research papers in journals and conference proceedings. He received several best paper awards from international conferences, including the IEEE International Conference on ITS Telecommunications (ITST 2012), the IEEE International Conference on Communications in China (ICCC 2013), and the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC 2014). Dr. Cheng received the "2009 Chinese National Award for Outstanding Overseas PhD Student" for his academic excellence and outstanding performance. He has served as Symposium Leading-Chair, Co-Chair, and a member of the Technical Program Committee for several international conferences.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 006
http://kns.cnki.net/kcms/detail/34.1294.TN.20170227.1114.004.html, published online February 27, 2017

A Survey of System Software Techniques for Emerging NVMs

BAI Tongxin[1], DONG Zhenjiang[2], CAI Manyi[1], FAN Xiaopeng[1], XU Chengzhong[1], and LIU Lixia[2]
(1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China;
2. ZTE Corporation, Nanjing 210012, China)

Abstract

The challenges of power consumption and memory capacity of computers have driven rapid development on non-volatile memories (NVM). NVMs are generally faster than traditional secondary storage devices, write persistently and many offer byte addressing capability. Despite these appealing features, NVMs are difficult to manage and program, which makes it hard to use them as a drop-in replacement for dynamic random-access memory (DRAM). Instead, a majority of modern systems use NVMs through the IO and the file system abstractions. Hiding NVMs under these interfaces poses challenges on how to exploit the new hardware's performance potential in the existing system software framework. In this article, we survey the key technical issues arisen in this area and introduce several recently developed systems each of which offers novel solutions around these issues.

Keywords

non-volatile memory; persistent memory; file system; IO system

1 Introduction

Ever since the Internet and mobile computing dominated people's daily life, the continuing supply of big data, which relies on immense computing power to extract the hidden big values, has demanded higher speed of data storage. The big data trend challenges the design of computer systems, on both hardware and software, to sustain the development of new data intensive applications. For example, deep data analytics and in-memory computing demand shorter turn-around time between processing iterations, which translates to faster data transportation between storage and processors. As of its current stage, in-memory computing uses dynamic random-access memory (DRAM) as the main media for hot data storage given its advantage in speed and bandwidth. However, DRAM would soon hit energy wall when the total memory capacity keeps growing in a data center. According to a recent study, 100 petabytes main memory with DDR 3 DRAM would consume 52MW power, which is far beyond the energy budget for building a future exascale data center [1], [2].

The combined requirements on capacity, performance and power have motivated both industry and academia to pursue new technologies and build alternative memory devices to bridge the gap between fast DRAM and slow disks. In recent years, semiconductor manufacturers have invested heavily on non-volatile memory (NVM) devices. As the most mature NVM in its class, Flash is already widely used in commercial servers due to its high density and low static power consumption. However, Flash has its notable downside. Memory wear and block erasure make it ill-fit for random read and write for which DRAM has outstanding performance [3]. Alternative NVM technologies have advanced rapidly, each expected to improve upon some or all of the Flash weaknesses. Among the most influential new types of NVMs are Phase Change Memory(PCM), Spin Transfer Torque RAM(STT - RAM), Resistive RAM (RRAM), Racetrack Memory, and Domain Wall Memory (DWM). Most recently, Intel and Micron jointly announced 3D XPoint and claimed the new NVM delivers a performance of 1000 times shorter latency and longer endurance than of the conventional Flash. The density and performance boost is said to be enabled by a novel structure of memory cell with a stackable data access array.

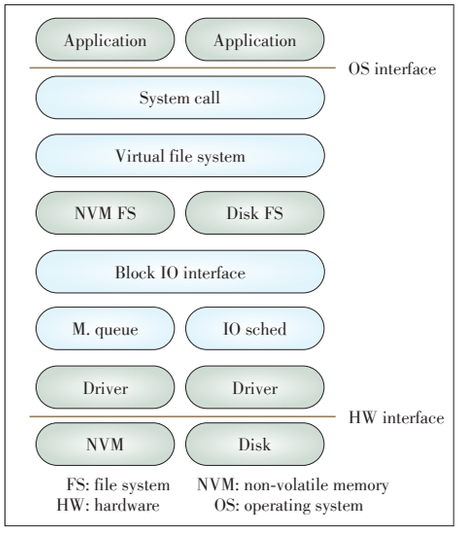
Other than higher read-write performance, most of the new NVMs support some extent of byte-addressable random access. The fine-grained access capability would revolutionize the way data is communicated between on-core and off-core memory. Potentially, CPU could directly address the data on the secondary storage without first sending instructions to a device controller. A secondary storage with byte-addressing capability is usually called Storage Class Memory (SCM). Once SCM is widely deployed, changes must be applied to the IO interface as well as the file system for they are traditionally optimized for slow devices. The fast storage would make many optimizations less effective, or even harmful, and favor a simpler design of the software stack for better performance.

The promising future of new hardware motivates software innovations which promote the lower level improvement to higher level usability. In this article we survey and introduce recent advances on how the NVM adaptation is addressed in sys-

tem software, particularly in IO and file system.

2 IO Subsystem for NVM

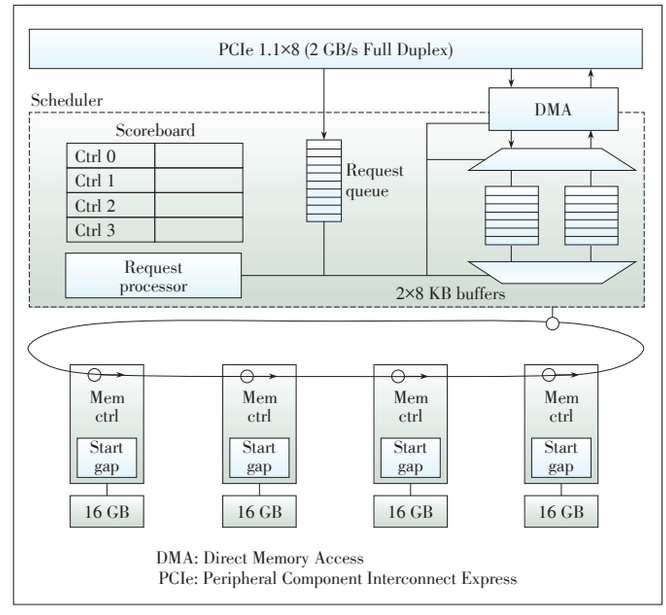
Fig. 1 summarizes a standard hierarchy of system components in which NVM devices shared the same IO interface with other block devices. The usual procedure of reading and writing data on a block device begins with a user program issuing a system call with arguments specifying the location and size into the target file in the file system. After the program is trapped to the kernel mode, the system call request is transitioned through multiple layers of kernel components, including the file system and the block IO interface, until it's finally translated into a sequence of low level instructions to the device driver. Since accessing block devices can pose a long latency, the IO requests are usually processed in an asynchronous way, leveraging Direct Memory Access (DMA) and interrupt handling to complete the data transfer while saving a large amount of CPU time. The software overhead during the request processing is caused by program state transition, file system management, IO scheduling, interrupt handling, buffer cache management and so on. Studies show that the software latency for processing blocked IO request is in the 10 microseconds on modern Intel processors running Linux operating system (OS). In reality, the exact software overhead varies for specific systems. Swanson et al. [4] measured that a single 512 byte IO request incurred about 19 microseconds of software overhead on Intel Nehalem 2.27GHz processor. In the test conducted by Yang et al. [5], a 512 byte IO operation cost 5 to 7 microseconds in software on 2.93GHz Intel Xeon processor. Compared to tens of milliseconds of latency due to disk access, the relative cost of software is very small. However, with NVM storage taking place, the relative cost of software increases significantly. The modern Flash solid-state drive (SSD) offers read-write latency of tens of microseconds. It's projected that newer generations of NVM can further reduce the state-of-the-art by 10 times [6]. With device latencygreatly reduced, the originally small software and interface overhead will dominate the cost of an IO operation. Therefore, in the future NVM systems, software and interface optimizations are the key to better storage performance. In the following we will introduce a number of recent developments on high performance IO interface and processing techniques for NVM storage.



◀**Figure 1. IO system software.**

2.1 Moneta

Caulfield et al. proposed Moneta [7] in 2010, an experimental NVM interface framework. Based on the simulations of Phase - Change Memory (PCM), the experiment results show that the Moneta interface helps random read and write performance with an increment of 18 times than that of the baseline. The software overhead is reduced by 60%. 4KB random read and write throughput can be maintained at the level of the 450k IOPs. Moneta is based on a design comprised of a token ring network and memory controller array, which improves the IO rate by exploiting the latency and parallelism advantage of the hardware. The detailed structure of Moneta IO is shown in **Fig. 2**. The IO scheduler is responsible for coordinating the data transmission and request scheduling. Data is transferred between memory and device through the PCIe interface. The requests are exchanged in the form of a command via a token ring network connecting the NVM memory controller and the request queue. When the system is running, the driver software issues IO requests which are transmitted via the Peripheral Component Interconnect Express (PCIe) interface to the Moneta scheduler and then are inserted into a first - in first - out (FIFO) queue.Requests larger than 8KB need to be decomposed and transmitted in sequence. To streamline data transfer, each Moneta memory controller is equipped with two 8KB



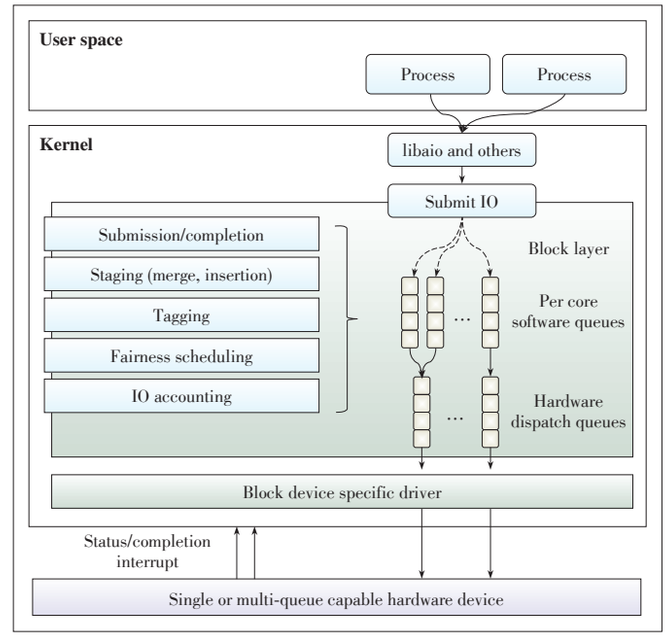
▲**Figure 2. Moneta IO architecture.**

buffers, used for caching the data read from and written to the storage after the requests are successfully processed. To make full use of the 2GB bandwidth of PCIe, it is not enough to expand the capacity of the device interface. As Caulfield et al. [7] have tried, a series of kernel optimization are applied to improve the efficiency of request processing in CPU. These software approaches include 1) avoiding the default IO scheduler, 2) using the Moneta built-in atomic read-write operations in place of locked kernel IO operations, 3) allowing multiple threads to process device interrupts in parallel, and 4) using spinlocks to avoid unnecessary context switch overhead caused by interrupts. Combining the above methods, Moneta reduces the IO access latency to around 1μs, achieving a significant improvement relative to the baseline latency of 10 μs. Correspondingly, the effective bandwidth increases to hundreds of mega bytes per second. Compared to traditional IO, Moneta excels with its comprehensive overhead reduction on interrupts, synchronization and scheduling.

For evaluation, the Moneta prototype was tested against a set of database applications. One interesting observation is that traditional database's optimization against disk storage can produce counter effect on Moneta's own optimization for NVM. In particular, PostgresSQL and MySQL receive less performance gainfrom the new framework than the simpler BerkeleyDB does.

2.2 Linux Multiqueue Block IO

The low latency and good parallelism of NVM storage would be underutilized features if the original Linux IO subsystem remains using a single request queue, which would easily become a performance bottleneck when the IO request rate approaches million per second. In order to solve the scalability problem in an NVM based system, Linux adopts a new block device interface blk-mq starting from kernel version 3.13. The internal structure of blk-mq is shown in **Fig. 3**. In the new IO framework, each IO request is processed in two phases separately. When an IO request arrives at the kernel through a system call, it is pushed onto a software staging queue which is dedicated to the CPU core on which the working thread is running. The request stays in the software queue while the kernel applies scheduling logics; then it's transmitted to a hardware dispatch queue waiting for the hardware to be ready to process it. To achieve high concurrency, a single storage device can be configured with multiple dispatch queues to better utilize the parallel processing capability of CPUs and in-band signaled interrupts of devices. In fact, the hardware queues can be allocated as many as several thousands, a number determined by how many virtual context a device can support. For example, a device supporting MSI-X can allocate 2048 queues for it can register 2048 interrupts. This new design of the IO subsystem promotes a fast and localized IOrequest processing scheme, especially by reducing unnecessary remote memory accesses in an NUMA environment.



▲**Figure 3. Internal structure of Linux blk-mq.**

The blk-mq block device interface successfully separates the software scheduling and the hardware message buffering functionalities which used to share a single request queue. The separation reduces synchronization and buffer congestion and hence leads to a much improved IO scalability.

2.3 Poll or Interrupt?

Despite that a redesign of IO subsystem internal structure fundamentally improves the IO scalability, there remain other system software overheads affecting IO performance. Most noticeably, the overhead comes from interrupt handling and context switch. In a typical OS setting, after an IO request is submitted to the kernel waiting to be processed, the calling thread returns or simply blocks for response. The completion of the request starts with a device interrupt notifying the CPU that the data is ready. After the device driver picks up the interrupt and finishes the handling procedure it will notify the IO subsystem which then will complete the remaining work and wake up the suspended thread. This asynchronous style of IO operation saves valuable CPU time since the program waiting for response can yield the CPU temporarily for other programs to use. In theory, however, the benefit of asynchronous IO only exists if the hardware latency is larger than the combined software overhead. When the device latency is reduced to microsecond level, the benefit will diminish.

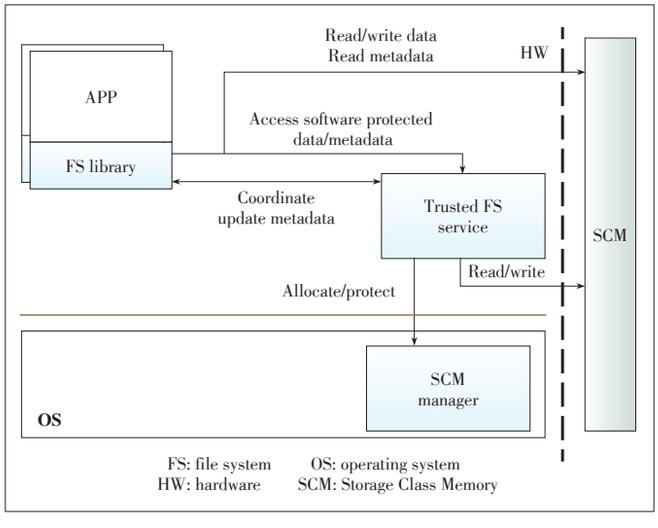
Polling provides a lighter weight means for checking device status than waiting for interrupt. To find out which is better with high performance NVM, Yang et al. compared the throughput and latency results using a simulated environment [8]. The study shows that synchronous completion requires shorter time and induces a better CPU utilization when that de-

vice latency is as low as several microseconds. Another interesting observation is, in the synchronous mode, a better hardware performance can reduce the software cost, whereas there is no such benefit for asynchronous IO. Furthermore, by studying throughput scalability, based on the case of 512 byte random reads, the study finds that the throughput of synchronous IO scales linearly with increased number CPUs. In contrast, asynchronous IO can only achieve 60%–70% throughput of the synchronous IO. Because asynchronous IO is suitable for processing long wait, when the system has a complex device setup, it can be suggested that the IO request be processed by a mixed mode of synchronous and asynchronous IO, achieving a load balance between CPU and the device.

2.4 NVM Express

NVM Express (NVMe) [9], [10] is a new software interface specification for accessing NVM devices attached to the PCIe bus. A working group on NVMe was formed in 2007. Technical work on the specification started in 2007 and the first release was finished in 2011. NVMe was designed from the ground up as an open device interface to exploit the low latency and parallelism of the future NVM devices, leading to an increased capacity of data path between CPU and storage. As a key design goal, the NVMe interface allows the processing power to fully utilize the internal parallelism available in the NVM devices and the bandwidth of the PCIe bus, hence effectively improving the IO performance. An example IO subsystem supporting NVMe is shown in **Fig. 4**.

NVMe can support up to 65,536 request queues, with request submission and completion stages allocating on different sets of queues (SQ and CQ).Separating the two stages reduces the likelihood of IO congestion, an issue often raised when using a single request queue in dealing with a large number of IO requests. The actual IO operations over the NVMe interface involve the software and the device exchanging commands and



▲Figure 4. Data organization in the BPFS file.

data on a dedicated memory mapped area in the host program's address space. Moreover, an NVMe device can support up to 2048 virtual contexts. With highly scalable multi-queue based IO scheduling, the NVMe interface supports a high throughput and concurrent data path between CPU and storage.

The ecology of NVMe-based systems is collaboratively built by the device manufacturers, chipsets providers as well as software venders. The first NVMe drives came from Samsung [11], LSI [12], Kingston [13] and Intel [14]. Early operating systems supported include Linux 3.3, Windows 8.1 and Windows Server 2012 R2. On chipsets, the major venders have already added NVMe support in their new product lines. To further promote software development using NVMe features, Intel announced the Storage Performance Development Kit (SPDK), a development toolkit providing the user level and poll based IO programming interface.

Performance studies on actual NVMe systems have recently been carried out by a group of researchers. Xu et al. measured and analyzed the performance differences of several database systems using NVMe SSD and Serial Advanced Technology Attachment (SATA) SSD on real machines [15]. The experiment results show that, compared to SATA SSD, the software overhead of NVMe SSD is reduced from 25% to 7% meanwhile 4 KB read throughput is increased from 70k IOPS to 750k IOPS. NVMe helps bring about 8 times performance gain in the tested database applications. This study is a solid proof that a redesigned device interface is necessary for exploiting the potential of new storage hardware. The performance demonstration also assures the system designers that NVM storage is ready to be a major investment for boosting overall service performance.

3 File Systems for NVM

A file system provides a data persistence service over a named space organized in directories. It strives to meet two practical goals: to maintain a consistent view of the data and to guarantee persisted data are reliably stored. In order to ensure the reliability and consistency, the file system needs careful organization of data layout as well as correct implementation of the operation semantics.

In a file system, stored objects are comprised of data and metadata, both of which may be accessed and modified when a single file operation is involved. The latency of accessing a conventional storage device can be long and unstable, which is a major concern of modern file system optimization. Techniques such as grouping metadata based on accessing pattern can improve data locality, leading to a better use of buffer cache. For reliability and consistency purposes, journaling and block level copy-on-write are typically employed to guard against system crash and power outage. These techniques themselves bring up additional operation overhead which may contribute to severe performance degradation. When a file sys-

tem is migrated to an NVM-based system, the change of storage structure can incur a host of new issues meanwhile eclipse the effect of existing optimizations. For example, fast NVMs make data prefetching and caching no longer a key mechanism for latency reduction. Moreover, the software overhead introduced by prefetching and caching can be significant in the setting of new systems.

New NVM storage exposes challenges as well as opportunities to modern file systems. New software structures and optimizations have been proposed and evaluated in several experimental systems. In this section we survey five state-of-the-art NVM file systems, outlining the new techniques around key issues developed in these systems.

3.1 File Operation Interface

Modern file systems usually offer two types of interfaces for reading and writing a file, the standard read() and write() system calls and then map the application programming interface (API). Realizing these two interfaces is affected by which particular hardware interface the NVM device is stalled on. If the device is installed on PCIe bus through NVMe interface, the legacy file system will function well as long the hardware interface is supported. Moreover, any IO system optimization against NVMe will benefit the file system as well. In this setting, data copying between the device and memory is necessary to realize both the read/write and the mmap APIs. In contrast, if the NVM is attached to the system through the memory bus, mmap would not involve extra copying, rendering faster external data operations.

The Byte-Addressable Persistent File System (BPFS) [16] is an NVM file system for the Windows operating system. Implemented using Windows Driver Model, BPFS provides users with the standard file operation interface. In the core, BPFS maintains an independent physical space for the file system, separated from the physical space of the user processes. Read and write operations involve data copying between the two physical spaces. Since the hardware is fast, buffer cache is no longer needed and the reduced software complexity helps improve the performance as a consequence. As an experimental file system, BPFS does not support mmap.

The Storage Class Memory File System (SCMFS) [17], [18] is an NVM file system developed for Linux, providing a compatible interface with common file systems in Linux. SCMFS leverages the processor's VMM features and simplifies the design of the file system to reduce the software overhead.

The Persistent Memory File System (PMFS) [19] is another example file system that exploits the processor's paging and memory ordering mechanisms to reduce the software overhead. PMFS provides both read/write API and mmap interface.

Aerie [20] is more of a file system framework than a single file system. It can be extended and customized based on particular application's requirements. As a framework, Aerie provides a flexible interface for the higher level software to work with. It can be used for implementinga POSIX compatible file system as well as building a user level library allowing applications to access files without going through the OS kernel.

The Non-Volatile Memory Accelerated (NOVA) file system [21] is log structured and designed for DRAM/NVM hybrid memory systems. NOVA is designed with full account of the device's byte addressability as well as the concurrency available in modern multicore systems. Implemented in the Linux kernel, NOVA supports both standard file operations and mmap APIs.

3.2 Internal Organizations and Management

Externally, files and directories are the main objects that a file system manages. Internally, the data in a file system are arranged in a collection of inodes, data files and logs. On the storage level, a data file can be laid out either in a sequence of extents, each of which being a consecutive run of blocks, or in a sequence of indirectly linked blocks [15]. Some data structures specific to file systems are crucial components for correctness and reliability purposes. For example, to guarantee crash consistency, a file system often uses journals to log uncommitted changes. Another widely used structure is the copy-on-write log tree, which provides a foundation for atomic updates of large data blocks.

As mentioned in previous sections, NVM has changed the cost ratio of software and hardware. As a result, the structures and algorithms in the conventional file systems, which were optimized against slow storages, do not work well for the fast NVM devices. To fully utilize the new hardware features, changes must be applied in software, potentially modifying the basic internal structures in the file systems. In the following, we highlight the key techniques and strategies for improved NVM usage introduced in the aforementioned file systems.

BPFS uses shadow paging to ensure reliable data update. Internally, the inodes, the catalog files and the data files are all stored in pages that are organized into a tree structure, as shown in Fig. 4. Unlike standard shadow paging that uses page level copy-on-write, the fine-grained access to NVM allows BPFS to manipulate data on a subpage level, a technical improvement that supports in-place modification of small data and partial copy-on-write, both of which can reduce the chance of page copying. For sake of performance, BPFS retains certain dynamic data structures in DRAM, including the storage management data structures and directory cache, to help speed up metadata querying speed.

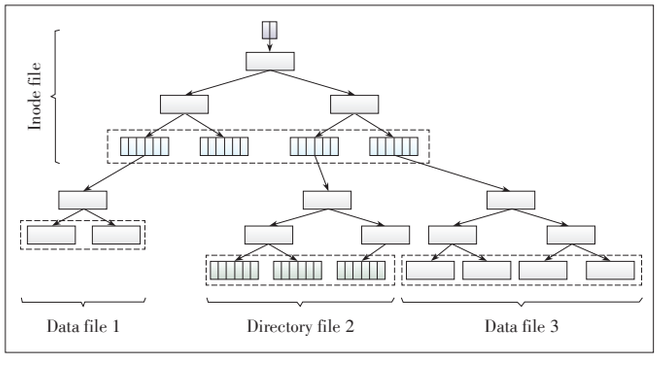
SCMFS leverages the existing virtual memory management (VMM) mechanisms provided by the OS and the hardware to simply the storage management of the new NVM file system.In SCMFS, the metadata and the address mapping table are stored in the physical address space, whereas the inodes, the catalog files and the data files are all mapped into virtual address space. In order to expedite the storage allocation and reclamation, SCMFS pre-allocates plenty of null files so that when

creating a new file it firstly looks for a suitable null file, and when deleting files it only marks them as null files. When the total size of null files is too large, garbage collection is triggered to recycle the storage of null files.

PMFS has the entire file system in the kernel address space. Under this arrangement, programs use the Direct Access (DAX) mechanism of Linux to access files, bypassing the buffer cache and incurring at most one copying between user space and kernel space for every piece of data. Moreover, zero-copy access is made possible to the memory mapped files whose user addresses are directly mapped to their in-kernel storage. Internally, PMFS organizes file storage based on B-tree structure, with 4KB, 2MB and 1GB as units of blocks. PMFS uses logs for basic consistency purpose.

Aerie decouples normal read and write operations from the management of the file system to reduce software overhead of directory lookups, metadata querying, synchronization operations and so on. Thus, different types of services can be assigned to different components which communicate and cooperate through Remote Procedure Call (RPC) as well as distributed lock service. As shown in **Fig. 5**, Aerie's distributed service architecture includes three major components: the storage manager in the OS kernel, the Trusted FS Service (TFS) and the FS Library (libFS). The storage manager is responsible for core functions and services that require privileged operations, such as allocating storage space for users, mapping the address space of the files, and modifying access permissions. TFS provides users with metadata modification, concurrency control and other critical services without special hardware support. TFS runs in an independent process and accepts RPC requests from user programs. LibFS provides ordinary file read-write operations and read-only operations of metadata. When privileged functions are required, LibFS will send RPC requests to the TFS. Aerie adopts an extent based multi-layerstructure for storage management. Seeking in a file is done by mapping the offset to a certain extent using a multi-level index. Moreover, user program scanning modifies an extent directly without going through the TFS.

NOVA augments the basic design of a log structured file system with features optimized against NVM. Based on the observation that logging is fast with NVM yet search is slow, NOVA builds an index in DRAM in addition to the logs in NVM to accelerate search operations. The traditional log structured file system suffers from the complexity of garbage collecting released logs into contiguous free regions. In NVM, random access is cheap so supplying a large contiguous region for logging is no longer necessary. In NOVA, logs are stored as linked lists so they don't need to be allocated in contiguous memory. Logs are chained up under individual inodes, which allows for high concurrency during access and recovery.



▲**Figure 5. The decoupled architecture of the Aerie file system.**

3.3 Consistency and Atomicity Maintenance

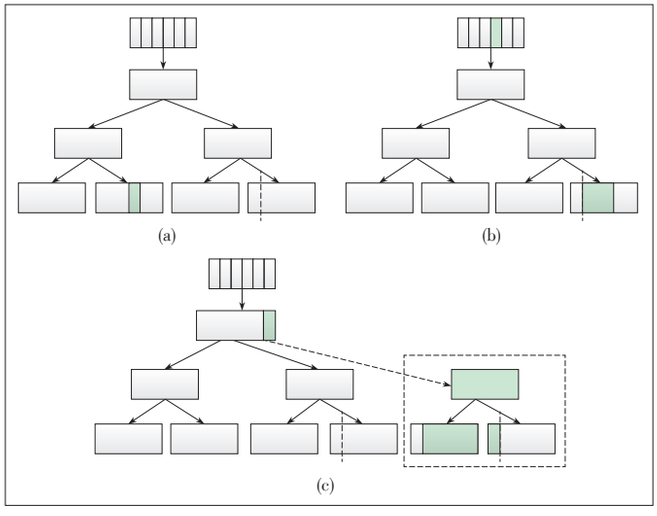
Journaling and shadow paging are techniques commonly used in file systems to achieve crash consistency. However, when implementing these techniques in an NVM system following the traditional way, performance issues may arise.

1) Issues with journaling. When journaling is enabled in a file system, write operations are amplified since every update requires writing into the storage twice, one to add an entry to the log, and the other to commit the change in the file. In terms of performance, the sequential characteristic of log appending is very favorable in the case of disk storage, but the new NVMs support fast random access, which significantly dampens the performance benefit of sequential logging.
2) Issues with shadow paging. With copy-on-write, an update to a logical page by the user program needs to be written to another free page, and only when changes are committed, the reference pointing to different methods for the old page will be replaced to ensure the atomicity of the modification. Since file systems usually organize internal storage in tree-like structures, a page getting modified implies its parent must be modified too. This may lead to a chain effect that an update on a single page triggers a series of page copying, causing severe write amplification.

Compared to traditional storage, NVM is fast and suitable for random access. The following outlines the solutions to the above issues, which take account of the hardware advantage.

BPFS proposes a technique called short-circuit shadow paging for updating persistent data. The new shadow paging scheme consists of three methods, in-place updates, in-place appends, and partial copy-on-write (**Fig. 6**). In-place updates can be applied for writes of 8 bytes or less, using hardware supported primitives. In-place appends refers to writing to the free area immediately beyond the file's end point. Since all the data beyond the file size is ignored, in-place writing to these locations is safe and once the writing is complete the file size is updated atomically. Partial copy-on-write allows atomic updates spanning multiple pages. The page copying only propagates to a point in which a single write suffices to commit the entire change.

In PMFS [19], based on the comparison study over the costs of different consistency techniques, the authors found that journaling at 64 byte granularity was most efficient for metadata



▲Figure 6. Three approaches to updating a file in BPFS: (a) in-place updates, (b) in-place appends, and (c) partial copy-on-write.

updates while it was less desirable than copy-on-write for large updates. Based on this observation, PMFS follows a hybrid strategy for consistency where in-place updates and fine-grained logging are used for metadata updates, and copy-on-write for file data updates. The metadata log update could be implemented by two different methods, namely undo logging and redo logging. In redo logging, the new data is firstly logged before committing to the file system. In undo logging, the old data is firstly logged before the new data is written in-place. In case of a failure, the system can be rolled back using the old data in the undo log. On the one hand, undo logging is costly for writes since a write barrier is required for every log entry in undo logging whereas only one write barrier is needed for a transaction in redo logging. On the other hand, redo logging is costly for reads and more difficult to implement since all the reads within a redo logging transaction have to search the redo log for the latest copy before reading from the file system. PMFS uses undo logging for metadata journaling.

Aerie needs to maintain the consistency of data updates in its particular distributed framework. If each metadata update requires one round RPC request from user program to the TFS, it's bound to hurt the service scalability. To solve this problem, Aerie applies an optimization that the client buffers the requests before periodically sending them to the TFS. In TFS, Aerie uses redo logging to realize atomic metadata updates.

NOVA modifies small metadata atomically in place. For single inode updates, NOVA relies on logs to record the changes. Modifications across multiple inodes resort to lightweight journaling to guarantee atomicity.

3.4 Hardware Primitives for Persistent Memory

Modern processors maintain consistency of memory operations by following certain a memory model. Without affecting the correctness of the program, memory requests could be delivered out of order after they are scheduled and buffered in the memory controller. Even the ordering enforcing instructions only guarantee the memory operations are properly ordered in the processor, disregarding in what order the data updates actually reach the memory. Unlike memory consistency, a file system is strict about when the data updates are safely stored. As previously mentioned, the file system consistency is realized by techniques such as journaling and copy-on-write, mostly implemented in software. For NVM, particularly byte-addressable NVM, pure software approaches to achieve consistency incur large overhead. To reduce the cost, new hardware primitives have been proposed and exploited.

BPFS proposes a write barrier instruction. Using the barriers, a program execution breaks down into a sequence of epochs. The order of persistent memory operations across epochs is strictly maintained. In addition to barriers, a file system needs atomic data updates to help crash recovery. Traditional file systems could verify atomicity by computing checksums. Leveraging hardware features, BPFS proposes a new atomic write primitive for small data updates. It is shown in a related research that such a light weight atomic operation requires merely 300 nanojoules reserved in the capacitor. With the new primitive, all data updates less than 8 bytes could be done in place.

PMFS uses atomic write instructions for modifying data of 8 bytes, 16 bytes and 64 bytes. The following scenarios explain when to apply the atomic instructions: 1) when reading a file, update the access time in the inode with the 8 byte atomic write instruction; 2) when appending to a file, use the 16 byte atomic write to update the size and access time in the inode; and 3) if Restricted Transactional Memory (RTM) [22] is available, use the RTM transactions for atomic updates within a cache line.

Aerie relies on the atomic instructions available in the x86 instruction set to realize three basic atomic primitives: 1) wlflush, which is implemented with the x86 clflush instruction, writes back the entire cache line; 2) bflush, which relies on the x86 mfence instruction, writes back the entire cache in processor to the storage, and 3) fence, which also uses the mfence instruction, enforces orderly writebacks. Based on these hardware primitives, Aerie manages a redo log for metadata updates.

However, it is worth to note that both clflush and mfence have limitations regarding memory writes to NVM. Clflush only flushes the cache line to the memory controller; it is left unknown whether the write eventually reaches the memory. Mfence only guarantees write orders are consistently recorded across CPUs; it has no constraints on the order of arrivals to the memory.

NOVA enforces write ordering upon memory operations by using a set of newly developed x86 instructions that have been proposed to tackle the above issues. These instructions include clflushopt (a more efficient version of clflush), clwb (cache line

write back without invalidation) and PCOMMIT (commit writes to NVM).

4 Conclusions

In recent years, the research on NVM and its software support has been a hot topic in computer systems area. The key advantage of NVM is its capacity to simultaneously achieve high density, low latency and low energy consumption. Hence it can potentially solve the energy scalability issues of large scale computer systems. The current NVM platform technologies, from the device interface to the software support, are not yet fully developed, leaving numerous challenges to be solved. From the software perspective, the most challenging issues arise in several areas, including IO optimization, memory management, file system as well as programming abstraction. To tackle these problems, researchers have explored novel ideas which involve restructuring the system software internals. In this article, we sampled a number of representative results in these areas and believe that new software techniques will emerge in response to the hardware's changing landscape in the future.

References

[1] B. Giridhar, M. Cieslak, D. Duggal, et al., "Exploring DRAM organizations for energy-efficient and resilient exascale memories," in *Proc. International Conference on High Performance Computing, Networking, Storage and Analysis*, Denver, USA, 2013, pp. 23:1–23:12. doi: 10.1145/2503210.2503215.
[2] Joel Hruska. (2014, Jul. 14). *Forget Moore's law: hot and slow DRAM is a major roadblock to exascale and beyond* [Online]. Available: http://www.extremetech.com/computing/185797-forget-moores-law-hot-and-slow-dram-is-a-major-roadblock-to-exascale-and-beyond
[3] Wikipedia. (2016, Dec. 19). *Memory* [Online]. Available: https://en.wikipedia.org/wiki/Flash
[4] S. Swanson, and A. M. Caulfield, "Refactor, reduce, recycle: restructuring the IO stack for the future of storage," *Computer*, vol. 46, no. 8, pp. 52–59, Aug. 2013.
[5] J. Yang, D. B. Minturn, and F. Hady, "When poll is better than interrupt," in *Usenix Conference on File and Storage Technologies*, San Jose, USA, Feb. 2012, pp. 3–3.
[6] Wikipedia. (2016, Oct. 18). *XPoint* [Online]. Available: https://en.wikipedia.org/wiki/3D
[7] A. M. Caulfield, A. De, J. Coburn, et al, "Moneta: a high-performance storage array architecture for next-generation, non-volatile memories," in *43rd Annual IEEE/ACM International Symposium on Microarchitecture*, Atlanta, USA, Dec. 2010, pp. 385–395.doi: 10.1109/MICRO.2010.33.
[8] A. Huffman. (2012). *NVM express revision 1.1* [Online]. Available: http://www.nvmexpress.org/wp-content/uploads/NVM-Express-1_1.pdf
[9] NVM Express. (2013, Apr.). *NVM express explained* [Online]. Available: http://nvmexpress.org/wp-content/uploads/2013/04/NVM_whitepaper.pdf
[10] Josh Linden. (2013, Jul. 18). *Samsung announces industry's first 2.5-inch NVMe SSD* [Online]. Available: http://www.storagereview.com/samsung_announces_industry_s_first_25inch_nvme_ssd
[11] Lyle Smith. (2013, Nov. 18). *LSI SF3700 SandForce flash controller line unveiled* [Online]. Available: http://www.storagereview.com/lsi_sf3700_sandforce_flash_controller_line_unveiled
[12] Dave Altavilla. (2014, Jan. 13). *Kingston hyper X predator PCI express SSD unveiled with LSI SandForce SF3700 PCIe flash controller* [Online]. Available: http://hothardware.com/news/kingston-hyperx-predator-pci-express-ssd-unveiled--with-lsi-sandforce-sf3700-flash-controller
[13] Intel. (2016) *Product comparison* [Online]. Available: http://www.intel.com/content/www/us/en/solid-state-drives/intel-ssd-dc-family-for-pcie.html
[14] J. Condit, E. B. Nightingale, C. Frost, et al, "Better IO through byte-addressable, persistent memory," in *ACM Symposium on Operating Systems Principles*, Big Sky, USA, Oct. 2009, pp. 133–146. doi: 10.1145/1629575.1629589.
[15] Q. Xu, H. Siyamwala, M. Ghosh, et al., "Performance analysis of NVMe SSDs and their implication on real world databases," in *8th ACM International Systems and Storage Conference*, Haifa, Israel, May 2015, article no. 6. doi: 10.1145/2757667.2757684.
[16] X. Wu and A. L. N. Reddy, "SCMFS: a file system for storage class memory," in *Conference on High Performance Computing, Networking, Storage and Analysis*, Seattle, USA, Nov. 2011, article no. 39. doi: 10.1145/2063384.2063436.
[17] X. Wu, S. Qiu, and A. L. Narasimha Reddy, "SCMFS: a file system for storage class memory and its extensions," *ACM Transactions on Storage (TOS)*, vol.9, no.7, Aug. 2013. doi:10.1145/2501620.2501621.
[18] D. S. Rao, S. Kumar, A. Keshavamurthy, et al., "System software for persistent memory," in *the Ninth European Conference on Computer Systems*, Amsterdam, The Netherlands, Apr. 2014, article no. 15. doi: 10.1145/2592798.2592814.
[19] H. Volos, S. Nalli, S. Panneerselvam, et al., "Aerie: flexible file-system interfaces to storage-class memory," in *the Ninth European Conference on Computer Systems*, Amsterdam, The Netherlands, Apr. 2014, article no. 14. doi: 10.1145/2592798.2592810.
[20] J. Xu and S. Swanson, "NOVA: A Log-structured File System for Hybrid Volatile/Non-volatile Main Memories," in *14th USENIX Conference on File and Storage Technologis (FAST 16)*, Santa Clara, USA, Feb. 2016, pp. 323–338.
[21] R. Arpaci-Dusseau and A. Arpaci-Dusseau. (2015, Mar.). *Operating systems: three easy pieces* [Online]. Available: http://pages.cs.wisc.edu/~remzIOSTEP
[22] Intel Corporation, "Intel64 software developer's manual," vol. 1, ch. 14, 2013.

Manuscript received: 2016-10-14

Biographies

BAI Tongxin (tx.bai@siat.ac.cn) obtained a PhD degree of Computer Science from the University of Rochester, USA. He is an associate professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interest concerns the interplay of programming language technology and big data systems.

DONG Zhenjiang (dong.zhengjiang@zte.com.cn) graduated from Harbin Institute of Technology, China. He is the vice president of Cloud Computing & IT Institute of ZTE Corporation, and the executive director of China Artificial Intelligent Association and of the Standing Committee of CCF on Service Computing. He is responsible for research and development of intelligent networks, customer services, IPTV and cloud computing, and has presided a number of national key R & D projects. His research achievements have been used in dozens of countries. His research interests are big data, AI and network security.

CAI Manyi (my.cai@siat.ac.cn) is currently a master student at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. Her current research interests include database systems and big data systems.

FAN Xiaopeng (xp.fan@siat.ac.cn) received his B.E. and M.E. degrees in computer science from Xidian University, China in 2001 and 2004 respectively. He received his Ph.D. degree in computer science from Hong Kong Polytechnic University, China in 2010. He is currently an associate professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include big data analytics, mobile cloud computing and software engineering. His recent research has focused on big data analytics in urban computing. He has published more than 30 papers in conferences and journals, and served as a TPC member for several conferences. He is a member of IEEE and ACM.

XU Chengzhong (cz.xu@siat.ac.cn) received his Ph.D. degree from the University of Hong Kong, China in 1993. He is currently a professor of the Department of Electrical and Computer Engineering of Wayne State University, USA. He also holds an adjunct appointment with the Shenzhen Institutes of Advanced Technology of Chinese Academy of Sciences as the Director of the Institute of Advanced Computing and Data Engineering. His research interest is in parallel and distributed systems and cloud computing. He has published more than 200 papers in journals and conferences with more than 7000 citations.

LIU Lixia (liu.lixia@zte.com.cn) received the M.S. degree from Ocean University of China in 2008. She is a senior engineer in the System Architecture Department of ZTE Corporation. Her research interests include text mining, bid data analysis and mining, machine learning, mathematical statistics, and cloud computing.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 007
http://kns.cnki.net/kcms/detail/34.1294.TN.20170301.1114.002.html, published online March 1, 2017

An Indoor Positioning Scheme for Visible Light Using Fingerprint Database with Multi-Parameters

CHEN Xiaohong, QIAN Chen, and WEI Wei
(School of Optoelectronic Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract

This paper proposes a novel indoor positioning scheme based on visible light communication (VLC). A new indoor VLC positioning scheme using fingerprint database with multi-parameters have been raised. We conduct simulation and experimental research on the illumination intensity distribution of several direction parameters. In the experiment, four LED matrixes are identified by LED-ID with room dimensions of 3.75×4.00×2.7 m^3. The results show that the mean of the location error is 0.22 m in the receiving plane, verifying the correctness and feasibility of the positioning scheme.

Keywords

visible light communication; direction parameter; fingerprint database with multi-parameters; indoor positioning

1 Introduction

With the popularity of smartphones and the increase of mobile data service, there is a growing demand for positioning and navigation, especially in complex indoor environments, such as airports, exhibition halls, supermarkets, and underground parking lots, where it is often necessary to ascertain the position information of facilities. The recent outdoor positioning is not suitable for indoor application of high precision positioning, due to the multipath fading and interference generated from other wireless devices in complex indoor environments. Therefore, seeking a kind of universal, low power, high precision indoor positioning technology is an imperative. As the supplement of indoor wireless location technology, visible light communication (VLC) has attracted widespread attention in recent years for indoor positioning service because of its features of unlicensed spectrum, immunity to electromagnetic interference, and free use in radio frequency restricted areas [1], [2]. The research of VLC provides a good basic support for the development of indoor positioning technology [3]–[5].

At present, the main algorithms adopted in indoor positioning system based on VLC are triangulation and scene analysis. Triangulation is a technique that uses the geometric properties of triangles to estimate the target localization [6], [7]. It has two branches: lateration and angulation. Lateration technique estimates the location of a target point according to its distances from multiple reference points. The distance is mainly derived by measuring time of arrival (TOA), time difference of arrival (TDOA) and received signal strength (RSS). Angulation measures angles relative to several reference points, namely angle of arrival (AOA) [8]. The system based on TOA [9], [10] is inconvenient in application because it requires strict clock synchronization between the reference points and the target point. Unlike TOA, TDOA-based systems measure the time difference of arrival. In [10], the difference in phase among three different frequency signals is used to calculate the value of TDOA and evaluate the performance of the system by the simulation. A RSS-based method measures the received signal strength and calculates the propagation loss that the mitted signal has experienced. An AOA-based method derives the angles of arrival relative to the reference points through the imaging receiver. This method does not need synchronizing, but the positioning accuracy will decrease sharply when the target point is away from the reference points because of the limitation of imaging resolution of the receiver.

In addition, scene analysis technique is performed in two phases: offline and online. During the offline, collection of the location information of sample points is called fingerprints, and a fingerprint database would be established. Then, the target's location will be found by matching real-time measurements to these fingerprints in the online phase. The method can combat the multipath effects in complex environments without parameters of estimated distance in an improved precision [11]. In view of the traditional fingerprint of visible light RSS, this paper puts forward a novel fingerprint (θ, RSS) by introducing the direction of the mobile terminal parameter θ combined with the traditional RSS. The approach is called the indoor positioning method for visible light using fingerprint database with multi-parameters.

2 Positioning Scheme

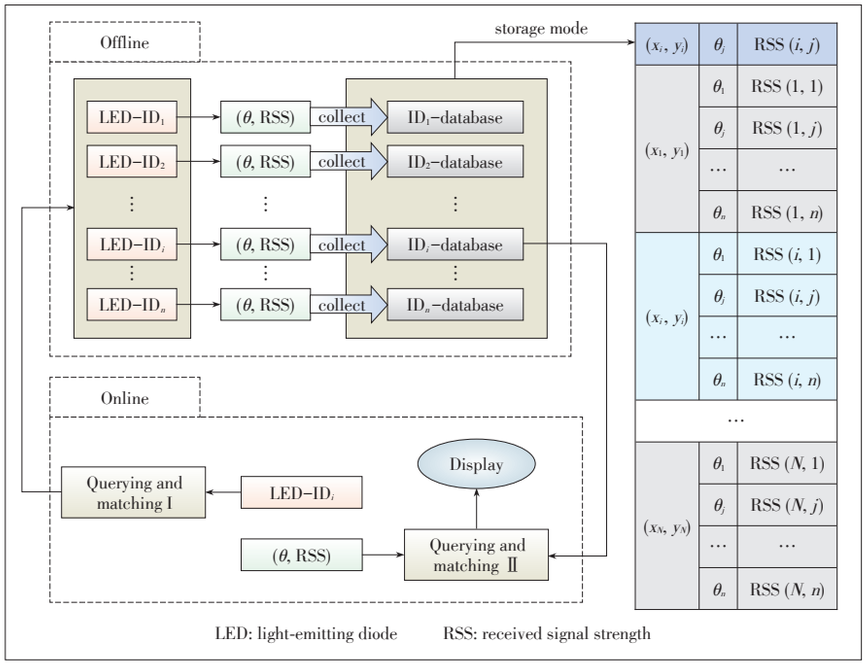
2.1 Principle and Establishment of the Fingerprint Database

Fig. 1 shows the proposed scheme that has offline and on-

This work was supported by ZTE Industry-Academia-Research Cooperation Funds under Grant No. 2014ZTE02-12.

An Indoor Positioning Scheme for Visible Light Using Fingerprint Database with Multi-Parameters

CHEN Xiaohong, QIAN Chen, and WEI Wei



▲**Figure 1. The principles of the proposed positioning scheme.**

line phases. We collect the direction parameter θ of the mobile terminal and the RSS value of each sample point under the coverage of each light-emitting diode (LED) light source, which generates a series of corresponding ID_i-databases in the offline phase. As shown in Fig 1, the LED-ID_i is the location information of the ith LED light source. Data in the fingerprint databases is stored in the form of (θ, RSS) and (x_i, y_i) are the coordinates of the sample points. If the direction parameter of each sample point is expressed as θ_j, then the received light intensity can be represented as RSS (i, j), and N and n are the number of sample points and the direction parameters, respectively. In the online positioning stage, the mobile terminal executes the operation of querying and matching according to the current LED-ID_i. After the ith database is downloaded to the mobile terminal, the location of the mobile terminal will be estimated by matching the received (θ, RSS) to these fingerprints.

2.2 Modeling

Compared to the ideal indoor illumination model, this paper provides a practical model for a small office building as shown in **Fig. 2**. The illumination lights are installed in a non-symmetrical layout. The detector is located in the plane with the level of 0.8 m and α is its elevation. A (3.06, 2.66, 2.7), B (3.07, 1.44, 2.7), C (0.68, 1.44, 2.7) and D (0.67, 2.66, 2.7) are the middle coordinates of each LED.

According to the proposed positioning scheme, we need to pre-store the illumination intensity under the direction θ of the sample point in the offline stage. It just needs to consider the I to IV quadrants on account of the upward receiving plane generally in 3D. We define the direction parameter as (α , β), where α and β denote elevation and azimuth.

The white LED is an incoherent source, thus the total illumination received from all the light sources is the sum of contribution from every LED, which can be expressed as:

$$E = \sum_{i=1}^{N} E_i , \tag{1}$$

where E_i is the illumination of the ith LED and N is the amount of LEDs.

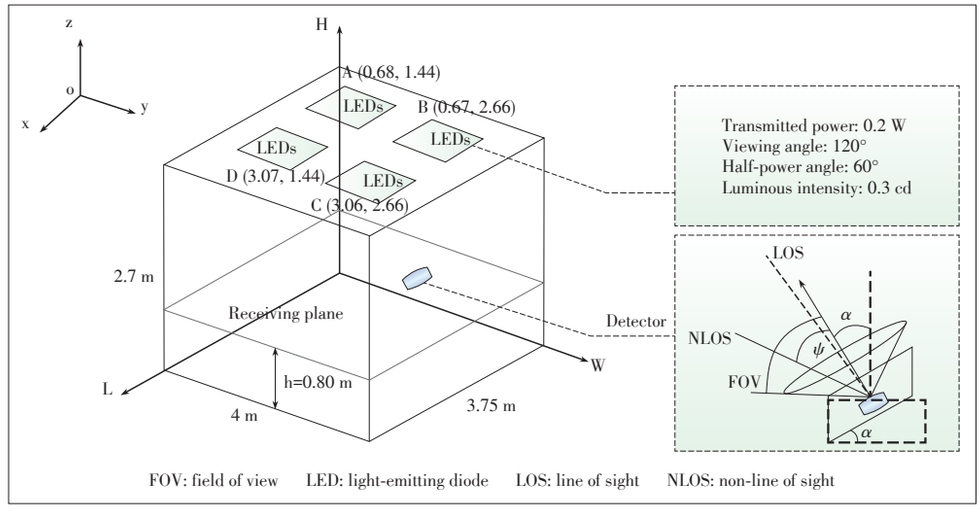
The radiation of a LED chip follows the Lambertian Radiation Pattern [12]. The luminous intensity formula is given by

$$I(\theta) = I(0)\cos^m(\theta) , \tag{2}$$

where $I(0)$ is the center luminous intensity of the LED, and the order m is given by $m = -\ln 2 / \ln \cos(\theta_{1/2})$ in which $\theta_{1/2}$ denotes the LED view angle at half power. Thus, the illumination intensity on the receiving plane with the distance D to light source can be expressed as:

$$E_{hor} = I(0)\cos^m(\theta) / D^2 \cdot \cos(\varphi) . \tag{3}$$

Every LED array is composed of 12×8 LED chips. The pa-



▲**Figure 2. The proposed indoor positioning model of visible light.**

rameters of each LED chip adopted in the simulation and experiment in this paper are shown in Fig. 2.

3 Simulation Analysis

In the proposed VLC positioning system where the downlink is considered only, the signal propagation path includes the line of sight (LOS) and non-line of sight (NLOS). As the direct visible light power to the detector accounts for 95% [13], only the LOS link was considered in the simulation. According to the above model, the paper studies the indoor illumination distribution with and without the direction by simulation.

3.1 Illumination Distribution

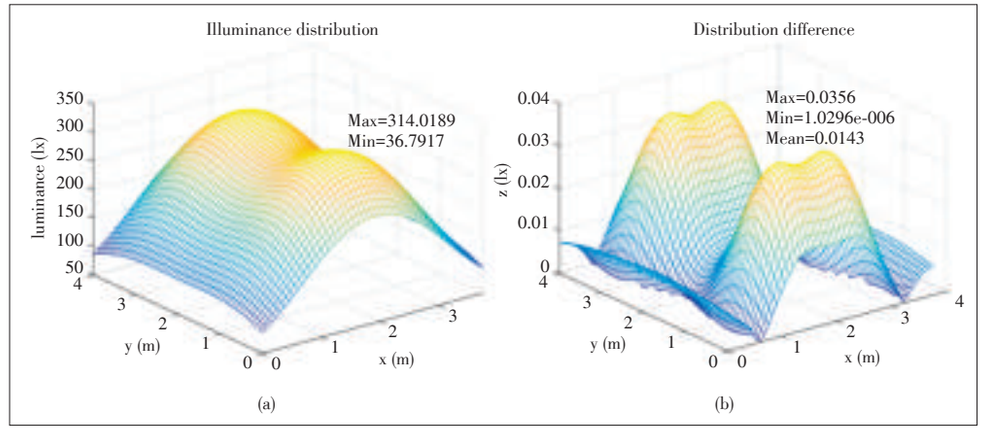
To explore the influence of the point light source and array source for this positioning method, the simulation for studying illumination distribution without direction parameters was carried out (**Fig. 3**).

Fig. 3a shows the illumination distribution of the point light source and the hump shape is due to the short interval of LEDs in the x axis direction. The maximum luminance is 314.0189 lx on the hump top derived from the superposition of light sources, while the minimum is appeared in the four corners. The illumination distribution decreases with the increasing distance to the center of LEDs. Under the same conditions, the source was regarded as one point and one array source respectively in the calculation process, and their difference distribution is shown in Fig. 3b. The figure shows that the maximum difference is only 0.0356 lx, while the mean is 0.0143 lx. The sensitivity of illuminometer adopted in the paper is 0.1 lx, therefore, the impact of the array source on the positioning results can be ignored.

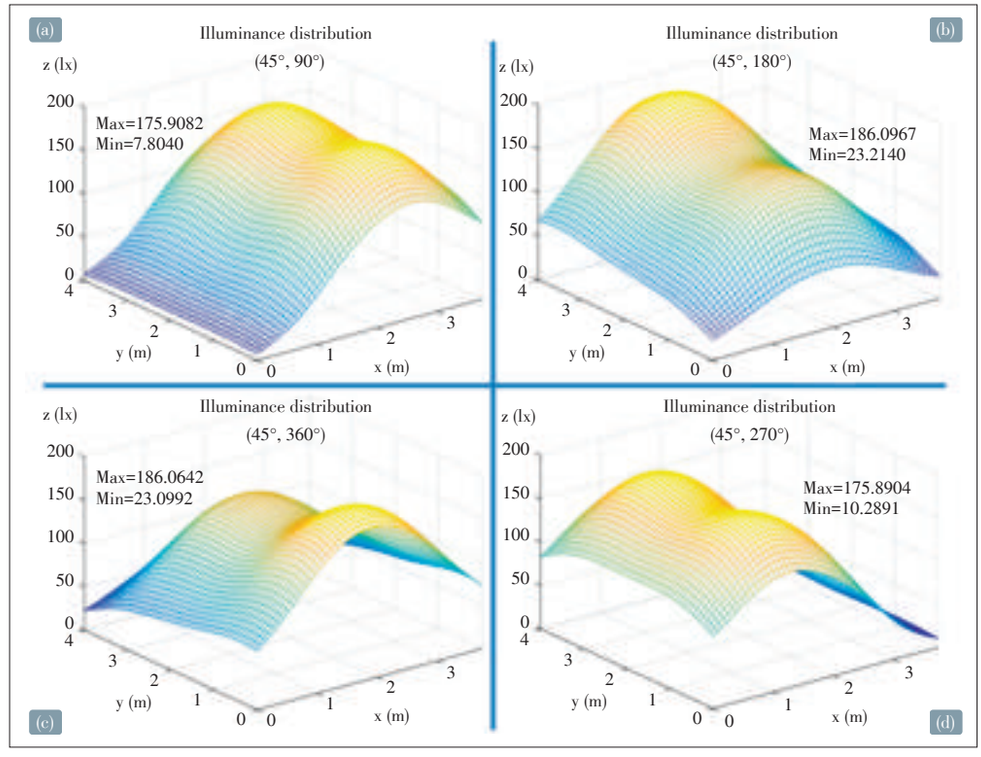
Fig. 4 shows the illumination distribution of several direction parameters of (45°, 90°), (45°, 180°), (45°, 270°) and (45°, 360°). It changes along with the different direction parameters at the same location in the room. Additionally, it reaches the maximum when facing the receiving surface of a mobile terminal, and reduces behind of the surface. Therefore, the proposed positioning fingerprint (θ, RSS) is significant.

3.2 Analysis of Positioning Results

To obtain the superior performance of the proposed positioning scheme visually, the isophotes are utilized to analyze the positioning results in the paper. First of all, the above simula-



▲**Figure 3. Simulation results: (a) the illumination distribution of the point light source and (b) the distribution difference of the point and array light sources.**



▲**Figure 4. The illumination distribution with four direction parameters: (a) (45°, 90°), (b) (45°, 180°), (c) (45°, 360°), and (d) (45°, 270°).**

tion results are indicated with isophotes, making the isophotes of different direction parameters intersect. Then, instead of the query matching process, the intersection is taken as an estimated position of the mobile terminal to verify the correctness and feasibility of the system.

Moreover, the normalization method is used to solve the issues of inconsistent range of illumination in this paper. The normalization is expressed as:

$$E_i^* = \frac{E_i - E_{\min}}{E_{\max} - E_{\min}}, \quad (4)$$

where E_i and E_i^* denote the illumination values before and after the normalization with the normalized range [0, 1]. E_{max} is the maximum and E_{min} is the minimum of the sample data. Furthermore, by using the illumination received with different direction parameters as sample data, we obtain E_{max} is 314.0189 lx and E_{min} is 7.8040 lx.

For example, when the real position was (2.85, 3.84), the location of the mobile terminal could be estimated by the illumination values of the direction parameters it received. The intersection point of the isophotes is the estimated position (**Fig. 5**). The first element of θ_i in Fig. 5 is the original illumination value, and the second is the normalized data. As it can be seen, the isophotes eventually intersect at the point of (2.85, 3.84), which verifies the feasibility and correctness of the proposed positioning method.

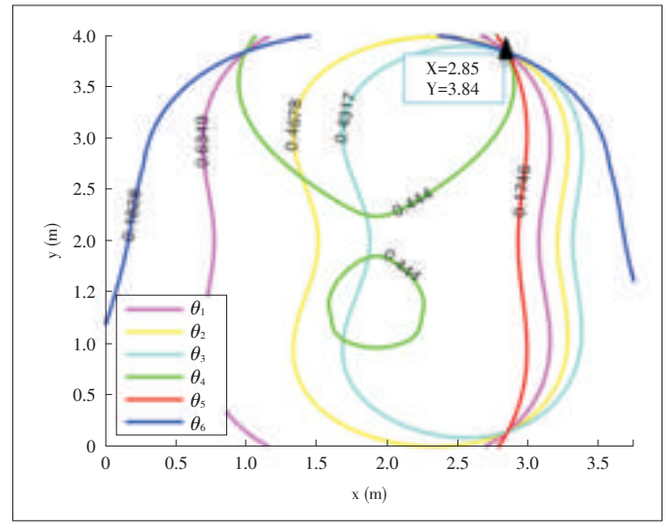
4 Experiment and Results

The verification test of the proposed model was carried out in the environment of a darkroom. The LED light sources were arrays that consists of 12×8 ZT5050WOS3 lamp beads. We used a CEM DT-1309 illuminometer with the sensitivity of 0.1 lx. The experimental model and the parameters of the light source were in accordance with the proposed model in Fig. 2. In the experiment of illumination distribution, 154 sample points were selected to be measured.

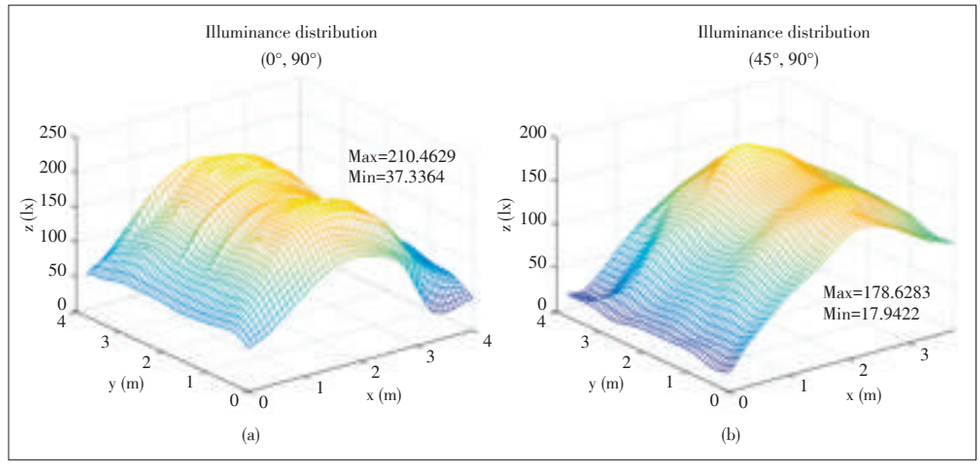
4.1 Analysis of Results

This paper investigates the spatial light intensity of five direction parameters of (0°, 90°), (45°, 90°), (45°, 180°), (45°, 270°) and (45°, 360°) with field measurements, and the obtained maximum and minimum are 210.4629 lx and 17.9422 lx, respectively. The measured results of (0°, 90°) and (45°, 90°) are shown in **Fig. 6** after interpolation processing. It can be seen that the variation trends are consistent with those in Fig. 4, but the overall illumination values are slightly lower than the simulation values. Taking the point of (2.85, 3.84) selected in the simulation as example, the received light intensity is normalized and then the results of positioning is analyzed.

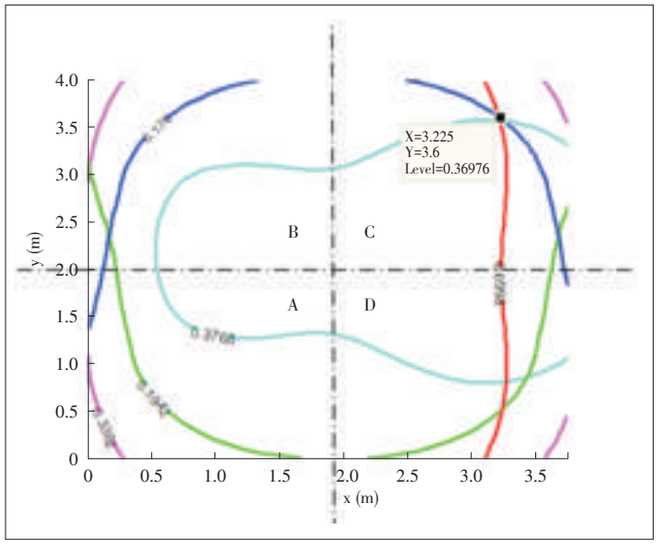
Fig. 7 shows the isophotes graph at the point of (2.85, 3.84), where there are several intersection points. The proposed positioning method based on multi-parameters fingerprint databases is a higher-precision positioning process based on the information of LED-ID. Consequently, the results in Fig. 7 are obtained on the premise of the received information of LED-ID for C#. Therefore, there is only one node for (3.225, 3.6) as the estimate point in the case. If there are multiple intersection points under the C# LED, we took the center location of the overlapping area as the target point, which should be set in advance in the principle. The positioning error of the target point is 0.45 m according to the Euclidean distance between the real and estimated positions of the mobile terminal.



▲**Figure 5. The isophotes graph at the point of (2.85, 3.84) based on the simulation.**

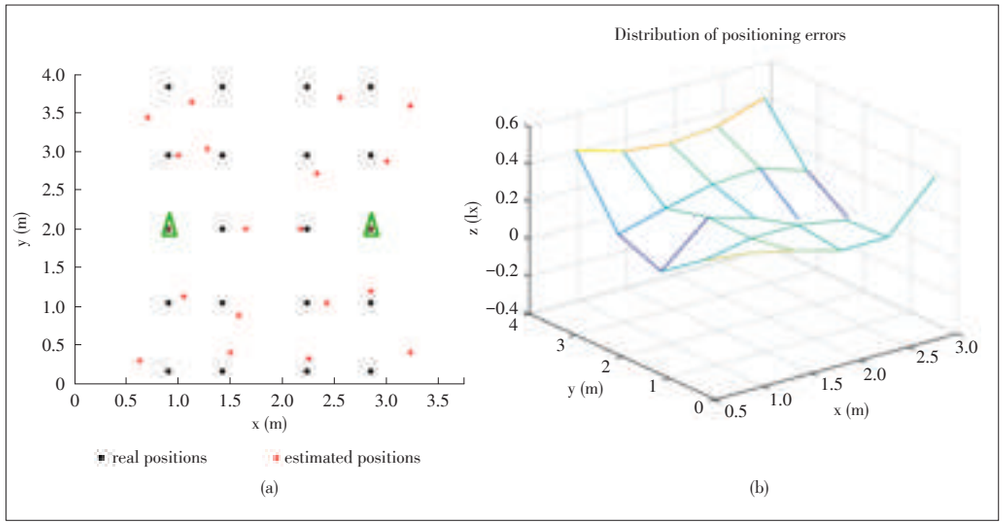


▲**Figure 6. The illumination distribution with direction parameters (a) (0°, 90°) and (b) (45°, 90°)**



▲Figure 7. The isophotes graph at the point of (2.85, 3.84) based on the experiment.

The accuracy is usually defined as the mean value of positioning errors. The positioning accuracy in our experiment is based on twenty points. **Figs. 8a** and **8b** show the real and estimated positions and their distribution of positioning errors, respectively. The positions of the two green triangles are the complicated superpositions of light sources, where the illumination intensity reaches the maximum, resulting in the poor positioning errors. Therefore, they are abandoned in order not to affect the overall positioning accuracy. In this way, when faced with complicated situations, the handling method must be defined in advance in the scheme to reduce the positioning error. As shown in Fig. 8b, in the edges and corners of the receiving plane, positioning errors also increase due to the decrease of the received light intensity. The maximum error is 0.4452 m and the average one is 0.2237 m.



▲Figure 8. The (a) real and (b) estimated positions of the mobile terminal at twenty points.

4.2 Analysis of Error

Factors affecting the location accuracy can be summarized into two main aspects. One is the sensitivity of detector and LED spacing. The illuminometer sensitivity adopted in this article is 0.1 lx. The mean deviation of illumination distribution between the point and array source is 0.01431 lx (Fig. 3b), an order of magnitude smaller than the sensitivity. Therefore, the difference was ignored in this paper under the experimental condition. The other influencing factor is the stability of light sources. The brightness of the LED is often instable with the change of supply voltage. On the other hand, the stability of LED light sources for the work played a vital role in the precision of the proposed positioning scheme. Although we tried to maintain the brightness stability by using constant current drive in the experiments, it still affected affect the positioning accuracy of the system, especially in the case of long working hours.

5 Conclusions

Compared to traditional indoor wireless location technology, the visible light positioning technology has the prominent features of ubiquitous coverage, energy saving, simple layout, low cost, etc. [14]. Therefore, it is expected to become one of the main means for indoor location. This paper presents an indoor positioning system model based on fingerprint databases. The simulation and experimental verification were conducted. The positioning errors of twenty points in the receiving plane with the accuracy of 0.22 m were also calculated, which visually verifies the correctness and feasibility of the proposed positioning scheme. Our work lays a theoretical and experimental basis for future research work.

References

[1] B. Liu, Y. Yang, L. Zhu, et al., "Light intensity stabilization in an indoor white LED wireless communication system," *Journal of Optoelectronics Laser*, vol. 21, no. 12, pp. 1794–1797, Dec. 2010.
[2] F. T. Wu and Q. L. Huang, "A precise model of LED lighting and its application in uniform illumination system," *Optoelectronics Letters*, vol. 7, no. 5, pp. 334–336, 2011. doi: 10.1007/s11801-011-1031-x.
[3] H. Chinthaka, N. Premachandra, T. Yendo, et al., "High-speed-camera image processing based LED traffic light detection for road-to-vehicle visible light communication," in *2010 IEEE Intelligent Vehicles Symposium*, San Diego, USA, Jun. 2010, pp. 793–798. doi: 10.1109/IVS.2010.5548097.

[4] T. Komine, J. H. Lee, S. Haruyama, and M. Nakagawa, "Adaptive equalization system for visible light wireless communication utilizing multiple white LED lighting equipment," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2892–2900, Jun. 2009. doi: 10.1109/TWC.2009.060258.
[5] K. Lee, H. Park, and J. R. Barry, "Indoor channel characteristics for visible light communications," *IEEE Communications Letters*, vol. 15, no.2, pp. 217–219, Jan. 2011. doi: 10.1109/LCOMM.2011.010411.101945.
[6] H. Liu, H. Darabi, P. Banerjee, et al., "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems Man & Cybernetics Part C Applications & Reviews*, vol. 37, no. 6, pp. 1067–1080, Oct. 2007. doi: 10.1109/TSMCC.2007.905750.
[7] L. Brás, N. B. Carvalho, P. Pinho, et al., "A review of antennas for indoor positioning systems," *International Journal of Antennas and Propagation*, vol. 2012, no. 2012, pp. 1497–1500, Dec. 2012. doi: 10.1155/2012/953269.
[8] W. Zhang, M. Kavehrad, "Comparison of VLC based indoor positioning techniques," *Proceedings of SPIE—The International Society for Optical Engineering*, vol. 8645, no. 12, pp. 1372–1375, Jan. 2013. doi: 10.1117/12.2001569.
[9] K. Panta and J. Armstrong, "Indoor localization using white LEDs," *Electronics Letters*, vol. 48, no. 4, pp. 228–230, Feb. 2012. doi: 10.1049/el.2011.3759.
[10] S. Y. Jung, S. Hann, and C. S. Park, "TDOA-based optical wireless indoor localization using LED ceiling lamps," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1592–1597, Jan. 2012. doi: 10.1109/TCE.2011.6131130.
[11] Q. Wan, X. S. Guo, and Z. X. Chen, "Theory, methods and applications of indoor localization," *Beijing, China: Publishing House of Electronic Industry*, 2012.
[12] S. Hann, J. H. Kim, S. Y. Jung, et al., "White LED ceiling lights positioning systems for optical wireless indoor applications," in *European Conference on Optical Communication*, Torino, Italy, Sept. 2010, pp. 1–3. doi: 10.1109/ECOC.2010.5621490.
[13] T. Komine and M. Nakagawa, "Fundamental analysis for visible light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 100–107, Jun. 2004. doi: 10.1109/TCE.2004.1277847.
[14] R. Shen and J. Zhang, "Visible light communication based indoor positioning methods," *Journal of Information Engineering University*, vol. 15, no. 1, pp. 41–45, Feb. 2014. doi: 10.3969/j.issn.1671-0673.2014.01.008.

Manuscript received: 2015-12-10

Biographies

CHEN Xiaohong (cxh@njupt.edu.cn) received her M.S. degree from Nanjing University of Posts and Telecommunications (NUPT), China in 2015. She is currently an assistant experimentalist with the School of Optoelectronic Engineering, NUPT. Her main research fields include visible light communications and image processing. She has published two papers in *Infrared Physics & Technology*.

QIAN Chen (qianc@njupt.edu.cn) received his B.S. degree in original radio engineering from Nanjing University of Posts and Telecommunications (NUPT), China. He is an associate professor and the associate dean of the School of Optoelectronic Engineering, NUPT. Dr. QIAN is mainly engaged in optical fiber, optical fiber communication, and visible light communication.

WEI Wei (weiwei@njupt.edu.cn) received her B.S., M.S., and Ph.D. degrees from Xi'an Jiaotong University, China in 1982, 1987, and 1998, respectively. She is a professor and the dean of the School of Optoelectronic Engineering Nanjing University of Posts and Telecommunications, China. Dr. WEI is mainly engaged in optical fiber and optoelectronic materials and devices in recent years. She obtained many research funds from NSFC (6 grants), National "863", National "973", and other programs at the provincial and ministerial levels. She has published over 100 peer-reviewed journal papers in *Advanced Materials*, *Applied physics Letter*, *Optical Express*, *Nanotechnology*, *Optical Letter*, *Journal of Chemistry Materials*, *Macromolecules*, and *Applied Physics B*, which were cited for over 800 times. She has held 16 patents.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 008
http://www.cnki.net/kcms/detail/34.1294.TN.20170206.1519.002.html, published online February 6, 2017

A Method for Constructing Open-Domain Chinese Entity Hypernym Hierarchical Structure

CAI Hongbo[1], CHEN Hong[2], and LIU Shen[1]
(1. Harbin Institute of Technology, Harbin 150001, China;
2. ZTE Corporation, Nanjing 210012, China)

Abstract

Entity relation is an essential component of some famous knowledge bases, such as Freebase, Yago and Knowledge Graph, while the hyponymy plays an important role in entity relations that show the relationship between the more general terms (hypernyms) and the more specific instances of the terms (hyponyms). In this paper, we present a comprehensive scheme of open-domain Chinese entity hypernym hierarchical construction. Some of the most important unsupervised and heuristic approaches for building hierarchical structure are covered in sufficient detail along with reasonable analyses. We experimentally evaluate the proposed methods and compare them with other baselines. The result shows high precision of our method and the proposed scheme will be further improved with larger scale corpora.

Keywords

Entity hypernym; hierarchization; Apriori Algorithm; suffix hypernym; open-domain

1 Introduction

Up till now, there exists a vast amount of free text on the Web, including newswire, blogs, product reviews, emails, governmental documents, and so on. How could a computer help the human to understand all of the data? A popular idea is turning unstructured text into structured one that could represent information concisely. Therefore, in recent years, some famous knowledge bases were constructed, directly showing the structural relations between entities.

Among hyponymy relations, we can classify and cluster entities by constructing the whole knowledge structure that includes the same level relations and affiliations as well. The relation between "植物 (plant)" and "动物 (animal)" is the same level, but the one between "植物 (plant)" and "生物 (living thing)" is affiliation, called hyponymy academically. If we can reorganize and supplement such relations between entities, we will get lots of information from the entity relation knowledge graphs, such as the position of a certain entity in the entity classification system. We can also know that "动物 (animal)" and "植物 (plant)" both belong to "生物 (living thing)". It is of great significance in search recommendation of search engines. For example, in a search engine, if a user put in "姚明 (YAO Ming)" and he/she will get the recommended information of "林书豪 (Jeremy Lin)" and "易建联 (YI Jianlian)" according to the entity knowledge graph, which is one of the functions of Knowledge Graph of Google. Baidu, a search engine company in China, is doing the similar project, which has begun loading online. Entity is the basic unit in natural language processing. Entity relation extraction is a traditional problem of natural language processing, which also makes some benefits for many other natural language processing tasks and information retrieval. Constructing an accurate and comprehensive entity relation graph is a great academic significance and has practical value for artificial intelligence.

There are some other issues relevant to entity relation graphs such as knowledge maps [1] that connect knowledge with locations. However, our entity relation graphs are based on hyponymy.

Entity relations are also relevant to entity relation graphs, and many scholars have proposed related solutions. Qian [2] exploited constituent dependencies to produce the dynamic syntactic parse tree and combined the entity semantic information to improve the relation extraction performance. Fader [3] proposed Reverb system based on [4]. The Reverb system firstly recognizes the word that describes the relation, and then makes the noun phrases in context of the word be the relation arguments to constitute relation triple. Although entity relation extraction and entity relation graphs have a lot in common, the entity relation graph mainly constructs the structure of all the entities so that many other entities, instead of the certain entity itself, are involved when building the relation between every two entities. Che [5] proposed entity relation extraction based on similarity computation.

The relevant research also contains the discovery of new knowledge and academic hotspot research, which has significance for the discovery of new research points. Chen [6] started to use the method of constructing knowledge graphs to investigate the development direction of academic research as well as some promising research areas recent years.

All the previous research work concentrated on entity rela-

This work was supported by ZTE Industry-Academia-Research Cooperation Funds.

tion construction of restricted domain. There is still less research for entity relation construction of open-domain. We propose a method for constructing entity hypernyms hierarchical structure for open-domain entity type diversification, hierarchizing the hypernyms of open-domain entities.

We firstly mine the hierarchical relation between entity hypernyms by using the association between frequent itemsets. We then use the suffix information of entity hypernyms to hierarchize and complete the hierarchical structure. We propose three hierarchical methods to hierarchize entity hyponymy in different ways.

2 Hyponymy Hierarchization

All the entities we used are from Sogou Cell Dictionary[1] and Baike[2]. We obtain a large number of entities and hypernyms according to the method of hypernym discovery based on the Internet [7]. There is no hierarchy between the hypernyms. For example, "花 (flower)" and "植物 (plant)" are the hypernyms of "百金花 (centaury)", and actually "植物 (plant)" is also a hypernym of "花 (flower)". "猫科动物 (Felidae)", "哺乳动物 (mammal)" and "动物 (animal)" are the hypernyms of "美洲豹 (catamount)", however, "动物 (animal)" is a hypernym of "哺乳动物 (mammal)" while "哺乳动物 (mammal)" is also a hypernym of "猫科动物 (Felidae)". Before hierarchizing, all hypernyms are at the same level.

There are a large number of hyponymy relations between hypernyms. We need to obtain the hyponymy between hypernyms by data mining based on the entities and hypernyms.

2.1 Hyponymy Hierarchization Based on Apriori Algorithm

2.1.1 Problem Analysis

There are a large number of hyponymy relations between hypernyms [8]. By observing the data, we find that if B is a hypernym of A, most entities belong to A also belong to B. But only a small number of entities belong to B and also belong to A. For example, "植物 (plant)" is a hypernym of "单子叶植物 (monocotyledon)". Then most entities that have the hypernym "单子叶植物 (monocotyledon)" also have the hypernym "植物 (plant)" while part of entities that belong to "植物 (plant)" belong to "单子叶植物 (monocotyledon)". We could discover the hyponymy between A and B by calculating the association between A and B.

2.1.2 Frequent Itemset Association Rules Mining

The association between hypernyms is pretty similar to the frequent itemset in data mining, which we can obtain by Apriori Algorithm. Apriori Algorithm is a traditional algorithm in data mining. It aims at identifying the frequent individual items and can be used to judge whether there is hyponymy between two hypernyms. In Apriori Algorithm, there are two important parameters: confidence and support. They play important roles in our experiment.

After simple analyzing, we can know that the probability for the low support frequent itemsets contain hypernyms is relatively low as well as the low confidence frequent itemsets. Instead, the probability could be high if the confidence and support are also high.

The input for Apriori Algorithm is the entities and their hypernyms. Each entity can have several hypernyms. The output is the confidence and support for each hypernym relation between hypernyms.

Table 1 shows the sample input: each line is a hypernym relation, and mainly 2 parts.

Table 2 shows the sample output: each line is a hypernym relation, and mainly 4 parts.

The confidence is the threshold for estimating the accuracy of the hypernym relation, and its value ranges from 0 to 1. The support is the threshold for the statistical support of the hypernym relation and its value is an integer.

The confidence and support between each two hypernyms need to be calculated as follows.

$$confidence(A,B)=\frac{count(A,B)}{count(A)}, \quad (1)$$

$$support(A,B)=count(A,B), \quad (2)$$

where A and B are both hypernyms. Other variables and functions are shown in **Table 3**.

If hypernyms A and B always co-occur, this may indicate that A is one of the hypernyms of B and meanwhile B is also

▼**Table 1. Sample input**

Entity	Hypernym
百金花 (centaury)	植物 (plant)
百金花 (centaury)	花 (flower)
百金花 (centaury)	中药 (traditional Chinese medicine)
日本角鲨 (Squalus japonicus)	生物 (living thing)
日本角鲨 (Squalus japonicus)	动物 (animal)
黄色白茧蜂 (Phanerotoma flava Ashmead)	昆虫 (insect)

▼**Table 2. Sample output**

Hypernym A	The hypernym of A	Confidence	Support
川菜 (Sichuan Cuisine)	饮食 (diet)	1	38
石竹亚纲 (Caryophyllidae)	植物 (plant)	0.968994	64
种子植物门 (Spermatophyta)	植物 (plant)	0.977477	1501
石竹目 (Caryophyllales)	种子植物门 (Spermatophyta)	1	63
鸟类 (bird)	动物 (animal)	1	58
冬青属 (ilex L.)	双子叶植物纲 (Dicotyledoneae)	0.907029	21

[1] http://pinyin.sogou.com/dict/

[2] Baidu Baike (http://baike.baidu.com/) and Hudong Baike (http://www.baike.com/)

▼Table 3. Definitions of variables and functions

Variables and functions	Definition
Confidence (A, B)	The probability that B is one hypernym of A for pre-estimation accuracy.
Support (A, B)	The statistical support that B is one hypernym of A.
Count (A, B)	The number of times A and B co-occur in the same entity hypernym set.
Count (A)	The number of times A occurs in all the entity hypernym set.

one of the hypernyms of A. We consider A and B have strong association that they should be synonyms. In this situation, there are no hypernym relations between A and B.

2.1.3 Algorithm Improvement and Optimization

We use two thresholds, confidence and support, to determine whether a hypernym relation can stand only when the confidence and support reach the specific thresholds. After some experiments, we find that if the support of a hypernym relation is just a little lower than the threshold while its confidence is much higher the threshold, it should also probably be correct. For this case, we improve the algorithm and use new evaluation methods.

1) Linear optimization

For a hypernym relation that needs to be judged, we set support as x, confidence as y and the hypernym exponent as H. We use four more parameters, C_1, C_2, S_1 and S_2 to determine whether B is one of the hypernyms of A. The meaning of each parameter is shown as follows:

C_1: forward direction confidence, the number of times for A and B co-occurring divided by the numbers of times for A occurring. It is a pre-estimation whether B is one of the hypernyms of A.

C_2: backward direction confidence, the number of times for A and B co-occurring divided by the numbers of times for B occurring. It is a pre-estimation whether A is one of the hypernyms of B.

S_1: minimum support. We consider the hypernym relation disconfirmed if x is lower than $S1$.

S_2: basic support. We consider the hypernym relation confirmed if x is higher than S_2 and y is higher than C_1.

H: the hypernym exponent for B to A, which can be estimated as follows:

$$H = y - \max\left(1 - \frac{(x - S_1)(1 - C_1)}{S_2 - S_1}, C_1\right), \quad (3)$$

of which the image description is shown in **Fig. 1**.

For each hypernym relation, we firstly calculate its support x and confidence y, then check whether $x>S_2$ (above the blue line) or $S_1 \leqq x \leqq S_2$ (above the red line). If a hypernym relation satisfies the condition, H will be positive, otherwise negative.

Actually, the blue line shows a traditional evaluation method while the red line shows the newly designed in this paper.

In summary, we calculate H according to its formula and determine a hypernym relation should be reserved or not by the positive or negative result.

2) Logarithmic optimization

For each hypernym relation, we get the same x, y and H as for the linear optimization. We use three parameters, C_1, C_2, and S, to determine whether B is one of the hypernyms of A. The meaning of each parameter is shown as follows:

C_1 and C_2: the same as those for linear optimization.

S: the support threshold.

H: the hypernym exponent for B to A, which can be estimated as:

$$H = y \log(x) - C_1 \log(S). \quad (4)$$

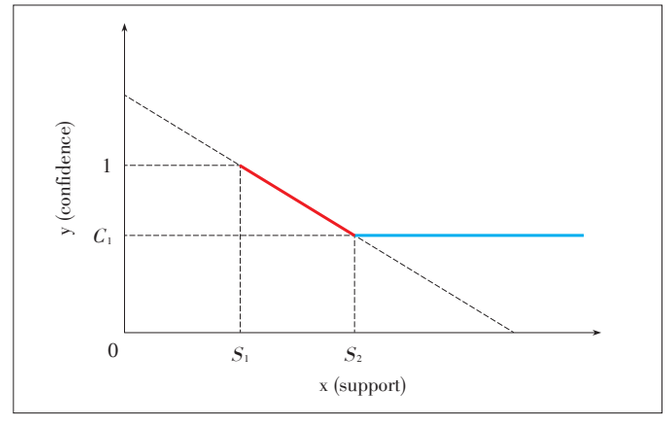
Similar to linear optimization, we calculate H according to its formula and determine a hypernym relation should be reserved or not by the positive or negative result.

2.1.4 Tongyici Cilin

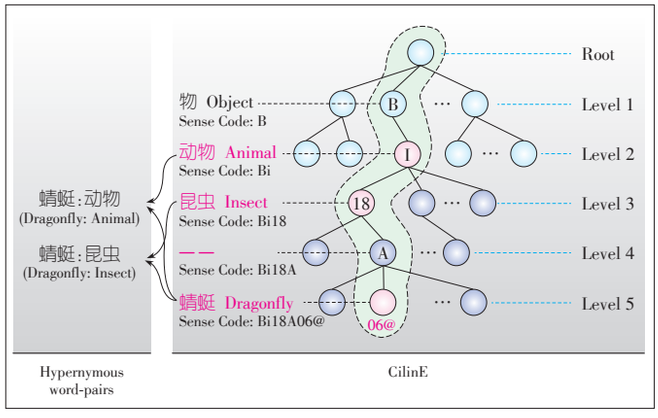
HIT-SCIR (Harbin Institute of Technology - Research Center for Social Computing and Information Retrieval) Tongyici Cilin (Extended) [9] (Cilin) is a Chinese Semantic Dictionary built by the Research Center for Social Computing and Information Retrieval in Harbin Institute of Technology. It includes 77,343 words, constructed into a 5-level hyponymy structure.

Cilin is built artificially and contains lots of commonsense hypernyms, which is complementary to the hypernym relations dug out automatically, is suitable to solve the problem that there are some hiatuses in the topmost hypernym chain [10].

There are 5 levels (not including the root) for the data in Cilin (**Fig. 2**), of which the first level has 12 categories such as "人(human)", "物 (object)", "时间与时空 (time and space)", "抽象事物 (abstract thing)", "特征 (property)", "动作 (action)", "心理活动 (mental activity)", "活动 (activity)", "现象与状态 (phenomena and state)", "关联 (relevance)", "助词 (auxiliary word)" and "敬语 (honorific)". Because we mainly process the words about entities, we keep the previous four categories from function words.



▲Figure 1. The determination for hypernym relation.



▲**Figure 2. Cilin hierarchical structure.**

We need to extract all the hypernym relations and find lots of polysemy that means one word belongs to different categories. After observing the data, we find that lots of polysemy is not accurate, such as "林肯 (Lincoln)" in "人 (human)" while also in "汽车 (automobile)", we just discover the "林肯 (Lincoln)" in "人 (human)" in data. So we pick up the polysemy for special processing in next filtration. Furthermore, there are some mistakes in some names of categories in Cilin, and we try to fix these mistakes in our experiments.

We keep the words relevant to the entities that have good quality for relation filtration, after extracting the hypernym relations in Cilin. For example, the "辈分 (generation)" in "人 (human)" and the "性能 (performance)" in "抽象事物 (abstract thing)" will be filtered.

When we are extracting the hypernyms of "哈尔滨工业大学 (Harbin Institute of Technology)", we may get "大学 (university)" and "高校 (college and university)", which are synonyms. We use the synonyms in Cilin and combine them.

2.2 Hyponymy Hierarchization Based on Suffix Hypernym

2.2.1 Problem Analysis

In hyponymy hierarchization experiment, we firstly use Apriori Algorithm to discover the hypernym relations between hypernyms by digging out the association of frequent itemset. However, Apriori Algorithm cannot discover the hypernym relations for those words occur just a few times. For this situation, we hierarchize the hypernyms based on suffix hypernym in case to obtain new hypernym relations.

We find that some hypernyms are the suffixes of their hyponyms. For example, "医院 (hospital)" is the suffix of words such as "哈工大校医院 (HIT University Hospital)", then "医院 (hospital)" is a hypernym of "哈工大校医院 (HIT University Hospital). Therefore we utilize the suffix information to discover and complete the entity hierarchical construction.

2.2.2 Suffix Hypernym

We define suffix hypernym that if word A is the suffix of word B, A is most likely to be a hypernym of B and A is the suffix hypernym. By observing the data, "运动员 (athlete)" is usually the suffix of other words such as "篮球运动员 (basketball athlete)" and "足球运动员 (football athlete)" and it is one of their hypernyms as well. Thus, "运动员 (athlete)" is a suffix hypernym.

The method, unlike Apriori Algorithm, is designed by the characteristic of Chinese that the suffix is usually the head word. The suffix hypernym occurs a lot so that we design the following steps to discover the hierarchical relations between hypernyms:

Step 1: count the frequency of each word that be the suffix of others among all the hypernyms;

Step 2: choose the words that have a high statistical frequency more than the threshold as suffix hypernym;

Step 3: do the suffix matching among hypernyms in order to obtain new hypernym relations.

2.3 Hyponymy Hierarchization Based on Classification

2.3.1 Problem Analysis

Most entities can be classified into "人 (human)", "物 (object)", "时间 (time)", "空间 (space)" and "抽象事物 (abstract thing)", which are the 5 hypernyms all from the top of Cilin. The roots of some hypernyms are not in the 5 top hypernyms, so we propose the hyponymy hierarchizing algorithm based on classification in order to put the hypernyms with no roots in the top 5 hypernyms into the 5 basic hypernyms. According to the data analysis, considerable entities do not reach "人 (human)", "物 (object)", "时间 (time)", "空间 (space)" and "抽象事物 (abstract thing)" at all. There is still a lot to do to enrich the hypernyms in the whole entity relation graph, especially the hypernym relations near the root.

With a hypernym non-polysemy assumption, we consider a hypernym belongs to "人 (human)", "物 (object)", "时间 (time)", "空间 (space)" and "抽象事物 (abstract thing)" but has no polysemy. The assumption is important and we need to analyze its correctness.

An entity may have several hypernyms, for example, "苹果 (apple)" belongs to "水果 (fruit)", "电影 (movie)" as well as "手机 (mobile phone)". So the entity "苹果 (apple)" may be polysemic. However, a hypernym, no matter it is "水果 (fruit)", "电影 (movie)" or "手机 (mobile phone)", will not be polysemic, because the hypernym itself stands for a category. Thus, the hypernym non-polysemy assumption is valid.

A hypernym in hierarchy has several fathers and children and they should also belong to one of "人 (human)", "物(object)", "时间 (time)", "空间 (space)" and "抽象事物 (abstract thing)". We actually put the hypernyms into small sets, in which the words are all belong to the same top hypernym among the 5 ones. The model is called cheat-in-exam model. First, we assume that there is an exam for the students in a class. All the students finish the exam by themselves without

copying others and get the class average score. For the second time, we set the students in groups and each group includes 4 students. Everyone in the same group is allowed to copy each other and we get another class average score. Usually, the second class average score is higher than the first one. This is the cheat-in-exam model.

2.3.2 Good Hypernym

When the confidence of the hypernym of an entity is more than 0.985, we call this hypernym a good hypernym.

We need the good hypernyms for hyperynm hierarchizing based on classification because most of the input data of hypernyms are noise.

After the hypernym non-polysemy assumption, we use the cheat-in-exam model to optimize our algorithm and also consider the calculated classification of the hypernym in the same set.

For similarity calculation, we use Backward Maximum Matching Algorithm according to the characteristics of Chinese. We use the words with their roots in the top 5 hypernyms as priori knowledge to guide other hypernyms to hierarchize.

Our algorithm use the idea of K nearest neighbors, that is, if the root of a hypernym is not in the top 5 hypernyms, we find the closed hypernyms with their roots in the top 5 hypernyms to determine its top hypernym.

We find the longest suffix of the hypernym with unknown top hypernym among the hypernyms with certain top hypernym in the top 5 and also find out the longest suffixes of its parents and children to determine the classification of the set of their own. This is the a cheat-in-exam model.

3 Experiments

3.1 Experiment Data

The data of entities and hypernyms we used are shown in **Table 4**.

3.2 Experiment of Hyponymy Hierarchization Based on Apriori Algorithm

Our statistics shows that there are 136,039 hypernyms in all 700 thousand words. We have 30,453 good hypernyms, 22.4% of all the hypernyms. Most good hypernyms occur with more than 0.985 confidence. Thus, the hypernyms with confidence lower than 0.985 probably are noise instead of hypernyms.

We adjust the confidence and support and find that the accuracy of result will be improved with the increase of confidence but the number of hypernym relations decreased. It is the same for the adjustment of support.

▼Table 4. The data of entities and hypernyms

The number of entities	The number of hypernyms	The average number of hypernyms for each entity
745,620	9,010,192	12.1

Using Apriori Algorithm to discover the association of frequent itemset, we obtained 8327 hypernym relations between hypernyms. **Table 5** shows the parameters setup for this experiment.

There are some indirect edges in the 8327 hypernym relations. For example, "被子植物 (angiosperm)" belongs to "植物(plant)" and "生物 (living thing)" while "植物 (plant)" belongs to "生物 (living thing)", so we can get the hypernym relation that "被子植物 (angiosperm)" belongs to "生物 (living thing)" because "植物 (plant)" belongs to "生物 (living thing)". We filter such redundant relations such as relation A→C while A→B and B→C exist.

There are 5422 relations reserved after filtering indirect edges of the original 8327 relations. We randomly picked up 200 relations of 5422 relations for manual evaluation and the precision is 97.0%.

The experiment results (**Table 6**) are basically consistent with expectations, while the result of linear optimization is the best with high precision and many hypernym relations.

3.3 Experiment of Hyponymy Hierarchization Based on suffix hypernym

We obtained 8747 hypernym relations by hyponymy hierarchization based on suffix hypernym. We did two more steps for these relations: filtering the indirect edges and duplicating relations.

There are 7503 reserved after pre-processing of the original 8747 hypernym relations. We randomly picked up 300 relations of 7503 relations for manual evaluation and the precision is 96.7% with 290 correct relations.

As shown in **Table 7**, we obtained a large number of hypernym relations between hypernyms by discovering the suffix hypernyms and using them to hierarchize the hypernyms. However, these relations are usually limited in some domains and have limited forms. We chose the two-character-suffix for high performance instead of one-character-hypernyms such as "人

▼Table 5. Parameters setup

Parameter	Definition	Value
C_1	Forward direction confidence	0.9
C_2	Backward direction confidence	0.8
S_1	Minimum support	5.0
S_2	Basic support	10.0

▼Table 6. Experiment results of hyponymy hierarchization based on Apriori Algorithm

	Precision	The relation number (no indirect edge)	Percentage of number increase
No optimization	97.0% (±1%)	3396	-
Linear optimization	97.0% (±1%)	5422	60.0%
Logarithmic optimization	84.5% (±1%)	4073	19.9%

(human)" and "物 (thing)", although "人 (human)" and "物 (thing)" are also hypernyms of many entities.

3.4 Experiment of Hyponymy Hierarchization Based on Classification

Most Chinese words have two characters or more. A Chinese word and its last character probably have different meanings such as "亚洲地理 (Asian geography)" and "理 (idea)". According to the data, the precision reaches about 80% even if just one character is matched during the process of backward maximum matching of most hypernyms. A higher precision will be reached when backward maximum matching is used for two or more characters. As for matching more than two characters, the result shows low performance that only 6.26% can be classified for good hypernyms and 3.30% for all hypernyms, even when the cheat-in-exam model is used.

We use the words with two or more characters backward maximum matched as closed words and 44.33% of good hypernyms are able to be classified.

After the experiments above, we classified 5119 hypernyms out of 11,547 unclassified good hypernyms. We randomly picked up 200 relations for manual evaluation and 187 of them are correct with the precision of 93.5%.

In summary, we obtained 5119 hypernym relations with the precision of 93.5% from good hypernyms, and 19,970 hypernym relations from all hypernyms.

The experiment results (**Table 8**) show that the classification reorganizes the entire entity relation graph, especially the part closed to the root. Although the hyponymy hierarchization based on classification discovers lots of hypernym relations, which are all closed to the root, the information provided is relatively limited.

Hyponymy hierarchization based on Apriori Algorithm hierarchizes hypernyms by using the associations between hypernyms; hyponymy hierarchization based on suffix hypernym hierarchizes hypernyms by using the suffix information; and hyponymy hierarchization based on classification hierarchizes hypernyms also by using the suffix information to complete the hypernym hierarchical structure. These three methods hierarchize the entity hypernyms in three different ways and the experiment results show that they all have high precision and obtain good results.

▼Table 7. Experiment results of hyponymy hierarchization based on suffix hypernym

Precision	The relation number	The relation number after filtering indirect edges and duplicating relations
96.7%	8747	7503

▼Table 8. Experiment results of hyponymy hierarchization based on classification

Precision	The relation number	The number of good hypernyms
93.5%	19,970	5119

4 Conclusions

We obtained a large number of entities and their hypernyms by extracting the data from the Internet, and then constructed the hierarchical structure of hypernyms based on Apriori Algorithm, suffix hypernym and classification and completed the hypernym hierarchical structure. We achieved good experiment results with high precision.

Acknowledgement

We thank the anonymous reviewers for their helpful comments.

References

[1] Q. Tang, F. Gao, and J. Wang, "Knowledge map concepts analysis and research," *Information Studies: Theory and Application*, vol. 34, no.1, pp. 121–125, 2011.
[2] L. Qian, G. Zhou, F. Kong, Q. Zhu, and P. Qian, "Exploiting constituent dependencies for tree kernel-based semantic relation extraction," in *Proc. 22nd International Conference on Computational Linguistics*, Manchester, UK, Aug. 2008, pp. 697–704.
[3] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proc. Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, Jul. 2011, pp. 1535–1545.
[4] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, amd O. Etzioni, "Open information extraction from the web," in *20th International Joint Conference on Artificial Intelligence*, Hyderabad, India, Jan. 2007, pp. 2670–2676.
[5] W. Che, T. Liu, and S. Li, "Automatic entity relation extraction," in *The First National Conference on Information Retrieval and Content Security (NCIRCS' 2004)*, Shanghai, China, 2004, 1–6.
[6] C. Chen, Y. Chen, J. Hou, and Y. Liang, "CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of the China Society for Scientific and Technical Information* vol. 28, no. 3, pp. 401–421, 2009.
[7] R. Fu, B. Qin, and T. Liu, "Exploiting multiple sources for open-domain hypernym discovery," in *Proc. 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, USA, 2013, pp. 1224–1234.
[8] R. Fu, "Open-domain named entity recognition and hierarchical category acquisition," Ph.D. dissertation, Harbin Institute of Technology, Harbin, China, 2014.
[9] W. Che, Z. Li, and T. Liu, "LTP: a Chinese language technology platform," in *Proc. 23rd International Conference on Computational Linguistics, Demonstrations Volume*, Beijing, China, Aug. 2010, pp. 13–16.
[10] C. Friedman, G. Hripcsak, W. DuMouchel, et al., "Natural language processing in an operational clinical information system," *Natural Language Engineering*, vol. 1, no. 1, pp. 83–108, 1995.

Manuscript received: 2015-11-05

Biographies

CAI Hongbo (hbcai@ir.hit.edu.cn) received his master's degree from School of Computer Science and Technology, Harbin Institute of Technology, China. He is an engineer at Alibaba Corporation. His main research interest is entity relation graphs.

CHEN Hong (chen.hong3@zte.com.cn) received her B.S. degree from the Department of Information Engineering, Nangjing University of Posts and Telecommunications, China in 2007. She is a senior research engineer at ZTE Corporation. Her research interests incloud social network analysis and intelligent question answering. She holds five patents .

LIU Shen (sliu@ir.hit.edu.cn) received his master's degree from School of Computer Science and Technology, Harbin Institute of Technology, China. He is an engineer at Miaozhen System. His main research interest is entity relation extraction.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 009
http://kns.cnki.net/kcms/detail/34.1294.TN.20170306.1830.002.html, published online March 6, 2017

Moving Target Detection and Tracking for Smartphone Automatic Focusing

HU Rongchun, WANG Xiaoyang, ZHENG Yunchang, and PENG Zhenming
(School of Opto-Electronic Information, University of Electronic Science and Technology of China, Chengdu 610051, China)

Abstract

In this paper, a non-contact auto-focusing method is proposed for the essential function of auto-focusing in mobile devices. Firstly, we introduce an effective target detection method combining the 3-frame difference algorithm and Gauss mixture model, which is robust for complex and changing background. Secondly, a stable tracking method is proposed using the local binary patter feature and camshift tracker. Auto-focusing is achieved by using the coordinate obtained during the detection and tracking procedure. Experiments show that the proposed method can deal with complex and changing background. When there exist multiple moving objects, the proposed method also has good detection and tracking performance. The proposed method implements high efficiency, which means it can be easily used in real mobile device systems.

Keywords

moving target detection; frame-difference method; background modeling method; camshift tracking; meanshift tracking; auto-focusing

1 Introduction

Smartphones have become an important part in model life. Most of the daily scenes contain changing background and diverse moving objects, which causes blur and low imaging quality. To get images and videos with high quality, people have paid much attention to the auto-focusing technique of camera in mobile devices.

This work was supported by ZTE Industry-Academia-Research Cooperation Funds.

In this paper, we propose a non-contact cell phone camera auto-focus method, which contains object detection in captured video, intelligent tracking and camera focus. For the moving object detection module, we adopt the 3-frame difference method and Gauss mixture model (GMM) [1]. For the intelligent tracking module, we adopt the optimized camshift algorithm [2], and the detected target of the front-end module is used as an input box to realize the intelligent tracking of multiple targets. The focus module conducts auto focus based on the coordinates provided by the object detection and tracking module, which is a non-contract focus method. The whole detection and tracking procedure can provide accurate object locating in real-time, and has robustness against diverse scene.

2 Target Detection

We present an object detection method combining the 3-frame difference method and GMM. The 3-frame difference method is suitable for detecting moving objects from statistic background, while the GMM method is suitable for dynamic background.

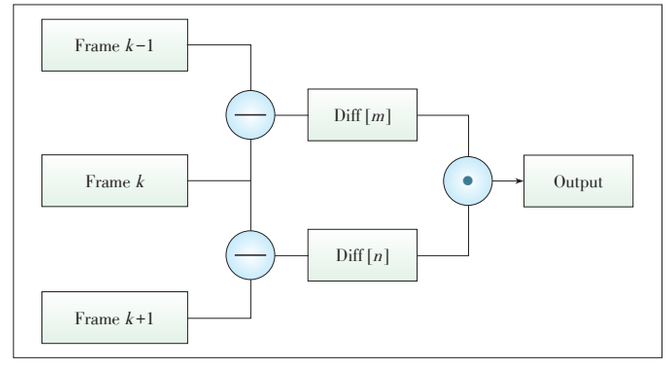
2.1 Three-Frame Difference Method

As we all know, frame difference is an effective way to detect moving objects at a low time costing level. The frame difference method, which detects the moving objects in video sequences by calculating the differences between two or more frames, can fully represent the feature of moving objects. Our proposed 3-frame difference method is shown in **Fig. 1**.

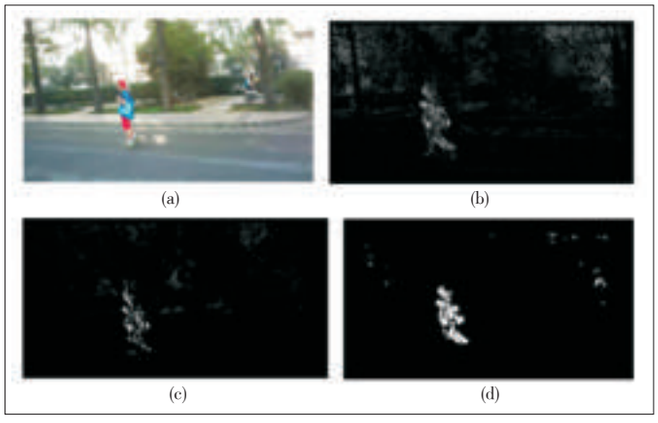
In Fig. 1, Diff[m] denotes the frame difference of fame k-1 and frame k, while Diff[n] denotes the frame difference of fame k+1 and frame k. After the frame process, the output is the dot product of Diff[m] and Diff[n]. **Fig. 2** shows the foreground detection results of the refined 3-frame difference method.

2.2 Gauss Mixture Model

The frame difference method can get an estimation of moving objects in videos. However, it is also sensitive to light change and noise corruption. The Gauss mixture model is an adaptive background modeling method, which has good perfor-



▲Figure 1. The refined 3-frame difference method.



▲Figure 2. Object detection results in the foreground: (a) Input frame; (b) by 3-frame difference; (c) by multiplying 2-frame difference; and (d) the binary image.

mance in complex changing scenes.

GMM represents every pixel by a mixture of Gauss models with different parameters. The number of Gauss models is set empirically. Each Gauss model has a weight value, which varies during the modeling procedure. By adjusting the parameters of Gauss models and the weight value, GMM can deal with slightly changing background and noise interruption. For every pixel in an image, the probability density function of GMM is defined as follows:

$$f\left(X_t = x\right) = \sum_{i=1}^{K} \omega_{i,t} * \eta\left(x, \mu_{i,t}, \Sigma_{i,t}\right), \quad (1)$$

$$\eta\left(x, \mu_{i,t}, \Sigma_{i,t}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|\Sigma_{i,t}\right|^{1/2}} e^{-\frac{1}{2}\left(X_t - \mu_{i,t}\right)^T \Sigma_{i,t}^{-1}\left(X_t - \mu_{i,t}\right)}, \; i = 1, 2, ..., K, \quad (2)$$

where X_t is the color of pixel. K is the number of Gauss distribution chosen to model the current image, $\eta\left(x, \mu_{i,t}, \Sigma_{i,t}\right)$ is the ith Gauss distribution at the time t, with an average value $\mu_{i,t}$ and a covariance value $\Sigma_{i,t}$. $\omega_{i,t}$ is the weight value of the ith Gauss distribution at time t, with $\sum_{i=1}^{K} \omega_{i,t} = 1$.

As the parameters of the mixture model of each pixel change, we can determine which of the Gaussians of the mixture are most likely produced by background processes. Here we choose the Gaussian distributions that have the most supporting evidence and the least variance. The distributions numbered 1 to B are chosen as the background model and B is expressed as

$$B = \arg\min_b \left(\sum_{k=1}^{b} \omega_k > T \right), \quad (3)$$

where T is a measure of the minimum portion of the data, which should be accounted for by the background. By identifying the background, we get an estimation of moving objects. **Fig. 3** shows the background modeling result of GMM. In Fig. 3b, the object information can easily be seen. It is not a pure background image due to the fact that there are no enough image frames to make a good mixture Gauss model. In Fig. 3c, the background is clear and with no object information. It is because that the mixture Gauss model is well build after enough image frames.

2.3 Target Detection in Foreground

After the separation of background and foreground by the above two methods, we need eliminate the false alarms in foreground images.

The post-processing of foreground images includes noising smoothing [3], threshold segmentation [4] and morphological processing including erosion and dilation [5]. After this procedure, the processed image becomes a binary image, in which 0 represents the background pixel and 1 represents the object pixel. Here we get the binary images obtained by both the frame difference method and GMM. The final binary image is obtained by combining them together:

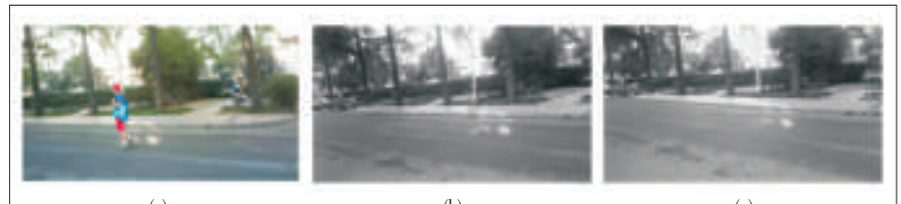
$$Result = Output\ of\ Diff \ \ Output\ of\ GMM, \quad (4)$$

where *Output of Diff* is the binary image from the frame difference method, and *Output of GMM* is the binary image of the GMM detection. Then we process the region in which the pixels are equal to 1 by edge detection. It gives the coordinate of the pixel block's boundary and section area. The coordinate can be used to realize auto-focusing.

We introduce the temporal filtering theory before the final detection results are output. We also conduct a coincidence comparison to the candidate foreground area which is detected in several continuous frames. The target detected in several continuous frames is determined as the true target while the others are treated as false alarms.

3 Smart Tracking

The smart tracking module includes the feature extraction of moving targets and target tracking. The key point feature is selected to perform the tracking procedure. As to the tracker, we choose the camshift tracking algorithm because it makes good



▲Figure 3. The background estimation results of GMM: (a) The original image; (b) background estimation 1; and (c) background estimation 2.

use of the color information of objects. However, the camshift does not take the texture and spatial structure of objects into consideration, which may cause inaccurate tracking result. Therefore, the Local Binary Pattern (LBP) feature descriptor [6] is introduced to the traditional camshift tracker, which builds a new effective joint histogram model of target appearance.

3.1 Camshift Algorithm

The basic idea of the camshaft, a continuous adaptive mean-shift algorithm [7], is to process all the frames with a mean-shift operator, and the last frame result (the central location and the window size of the search window) is regarded as the initial value of the mean-shift's searching window in next frame. Then the iteration continues. Assume that the video consists of n frames, the processing steps are shown in **Algorithm 1**.

Algorithm 1 Camshift Tracking

Input:
Labeled target area in the first frame;
Output:
Coordinate of targets in each frame.
1. Initialize: select the target area in the video;
2. Calculate 2D color probability distribution in the selected area;
3. **for** k=1:n **do**
4. Tracking the selected object using the meanshift tracker;
5. Calculate the object coordinate in the current frame;
6. Mark the object;
7. **end for**

3.2 LBP Operator

LBP is an operator to describe the partial texture features of an image. LBP at the coordinates (x,y) can be calculated as

$$LBP(x,y)=\sum_{n=0}^{p-1}2^n \operatorname{sgn}(i_n-i_c), \tag{5}$$

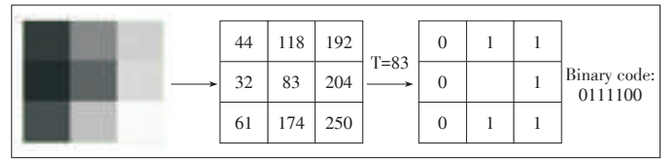
where i_c denotes the gray scale value of a pixel at the coordinates (x,y), i_n denotes the adjacent pixel's gray scale value, and p denotes the number of adjacent pixels. We usually adopt the 3×3 window with p=8. The sign function (sgn) is written as

$$\operatorname{sgn}(x)=\begin{cases}1 & (x\geqslant 0)\\ 0 & (x<0)\end{cases}. \tag{6}$$

Fig. 4 shows the processing steps of LBP, the LBP texture map is then obtained.

3.3 Target Tracking Algorithm Based on Combined Color and Texture Features

Algorithm 2 describes target tracking based on the combined color and texture feature.



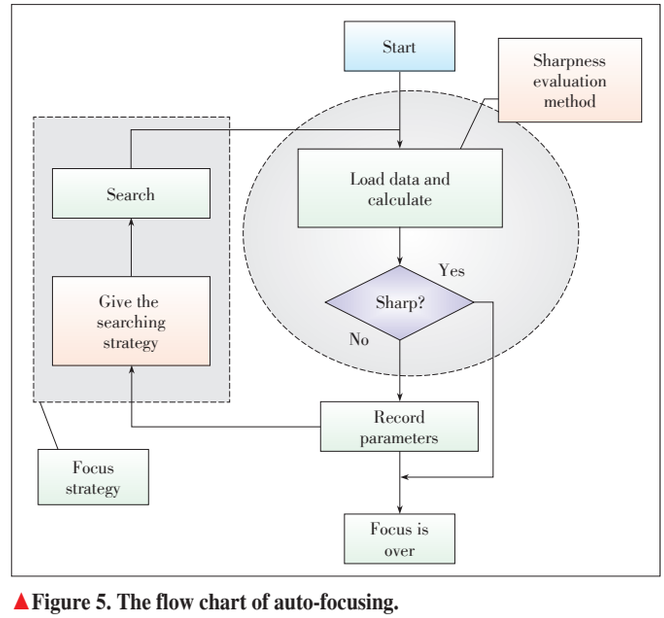
▲**Figure 4. The processing steps of LBP feature.**

Algorithm 2 Target tracking algorithm based on combined color and texture feature

Input:
The selected moving target;
Output:
Coordinate of targets in each frame.
1. Initialize: The size and position of each window;
2. **for** k=1:n **do**
3. Extract the texture feature;
4. Use the texture histogram with the image to get back projection;
5. Get the texture probability distribution map;
6. Carry out AND operation of the texture probability distribution map and the tone probability distribution;
7. Camshift tracking;
8. Update the size of search area;
9. **end for**

4 Auto-Focusing

When the target detection and tracking are completed, we need to conduct auto-focusing (**Fig. 5**) according to the given coordinates from the detection. An original focal length is first



▲**Figure 5. The flow chart of auto-focusing.**

given to calculate the sharpness value according to the sharpness evaluation method. The mountain climbing searching (MCS) algorithm is then conducted until the auto-focus system finds the sharpest target.

4.1 Focusing Sharpness Evaluation

The high-frequency components show the detail of an image. However, the first-order difference operator is not really sensitive to the high-frequency components, but works well on detecting the low-frequency components such as the target's outline. Our proposed method adopts the refined 8-neighbour Laplacian operator [8]. The Laplacian operator is an edge detecting operator defined by the x, y second order partial derivatives of the $f(x, y)$ image, which can be described as an approximate Laplacian operator template:

$$L_4(x,y)=\nabla^2\approx\begin{bmatrix}0 & 1 & 0\\ 1 & -8 & 1\\ 0 & 1 & 0\end{bmatrix}, \tag{7}$$

where ∇^2 is the second derivative differential operator for the Laplacian operator.

The focusing sharpness evaluation function is then given by

$$J_{laplacian8}=\sum_M\sum_N L_4(x,y)\,. \tag{8}$$

4.2 Focusing Searching Strategy

Our method uses MCS [9] to focus the search. First, an initial focus value M is given. The sharpness value $J(M)$ is calculated by the focusing sharpness evaluation method, and the focal length value is changed for calculating a new value N and its sharpness $J(N)$. $J(M)$ and $J(N)$ are then compared. At the same time, the search direction is determined and the iterative search is continued based on the side on which the sharpness value is greater. In **Fig. 6**, we search from point M to point N. In the searching process, the searching results such as focal length and the sharpness J are saved for the next iteration. A search climbs in one direction until the sharpness value reaches the maximum and begins to decline, and the searching process for the first time follows the solid line $M-N-P-P1$ in Fig. 6. Then the second search starts from $P1$ and makes a reverse lookup until the sharpness value begins to decline over the peak once again. The dashed line $P1-P-P2$ reflects the second searching process. Every time when the searching is done, we reduce the searching step in the next iteration. The iteration is repeated until the maximum focus value is found, and then the focusing is over.

5 Experiments

Our experiments were conducted at a DELL T5600 power station (CPU: Intel XeonE5 - 2603, 1.8GHz. Memory: 8GB. GPU: NVIDIA Tesla C2075), with which the process reached 25 fps.

We selected several specific videos that contain complex backgrounds and multiple targets. In this way, we examined the performance of the proposed algorithm under different circumstances.

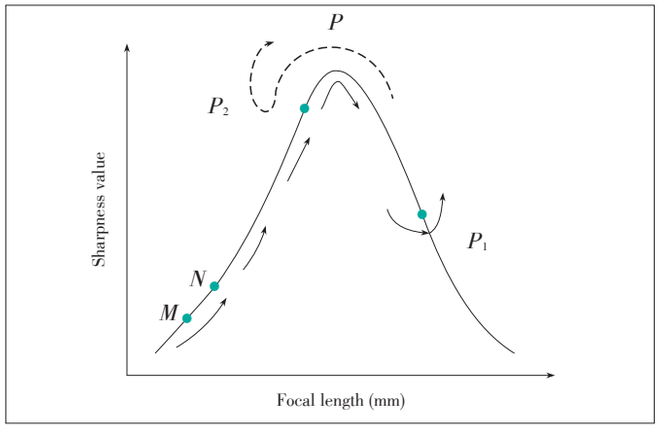
5.1 Target Detection

We captured one or more targets in the video sequences for target detection. The pictures in **Fig. 7** show that the football players were moving fast and the background contains trees, buildings and the sky. In this case, the moving targets could be easily detected by the proposed method (according to their sizes and positions).

To evaluate the performance of target detection, we introduced the detection accuracy rate (DAR) as a metric, which is defined as

$$DAR=\frac{Target\ detected\ frame\ amounts}{Frame\ amount}\,. \tag{9}$$

The DAR under single-target and multi-target modes for test videos are shown in **Tables 1** and **2**. The multi-target mode has a higher DAR than the single target mode, although the latter reaches a high DAR. This experiment demonstrates the effectiveness of the proposed detection method. The combi-



▲Figure 6. The mountain climbing search method.

▲Figure 7. Multi-target detection.

▼Table 1. Performance of target detection (single target)

File name	corridor.avi	office1p.avi	office2p.avi	outdoor.avi
Frame number	386	440	430	854
Target-detected frames	332	348	378	769
DAR	86%	79%	88%	90%

DAR: detection accuracy rate

▼Table 2. Performance of target detection (multi-target)

File name	basketball.avi	football.avi	football2.avi	spring.avi
Frame number	1098	1165	2035	1655
Target-detected frames	1010	1048	1933	1522
DAR	92%	90%	95%	92%

DAR: detection accuracy rate

nation of the frame difference method and GMM achieved good performance in highly complex background with multiple moving targets.

5.2 Smart Tracking

The tracking strategy is that one target was selected to track manually or automatically (according to the target's size or position) among the detected targets (**Fig. 8**). Then the tracking is kept until a new detection starts.

Similar with the target detection evaluation metric, we introduced the tracking accuracy rate (TAR), which is defined as

$$TAR = \frac{Target\ tracked\ frame\ amounts}{Frame\ amount}. \quad (10)$$

Table 3 shows the *TAR* of four different videos. As we can see, the refined camshift method can track well in changing background.

5.3 Combining Detection with Tracking

In our final combining test, the two detection and tracking modes were used to evaluate the performance of the proposed algorithm. The two evaluation modes are as follows.

1) MT Mode: It is for the multi-target scene. All the targets in the video sequences are detected and one of them is selected as the tracking object based on its size or position. The accuracy rate of target detection relies on the multi-target detection algorithms, while the tracking accuracy relies on the single target tracking algorithm.
2) S Mode: It is for the single target scene. The target is detected and tracked automatically.

Table 4 shows the performance of the two evaluation modes.

In practice, the algorithm efficiency has great influence on the performance of an auto-focusing system. The proposed method is more than 25 *fps* and has reached real-time requirements. **Table 5** shows the *fps* of each tested video under the MT and S modes. We can see that the proposed detection and tracking algorithm performs better in the S mode due to

▲Figure 8. The smart tracking is very smooth and robust.

▼Table 3. Performance of target tracking (single target)

File name	corridor.avi	office1p.avi	office2p.avi	outdoor.avi
Frame number	386	440	430	854
Target-tracked frames	379	426	427	828
TAR	98%	97%	99%	97%

TAR: tracking accuracy rate

▼Table 4. Performance of target detection and tracking

File name	Scene type	MT mode accuracy rate	S mode accuracy rate
corridor.avi	Single target	-	>95%
office1p.avi	Single target	-	>95%
office2p.avi	Single target	-	>95%
outdoor.avi	Single target	-	>95%
runner.avi	Single target	-	>95%
basketball.avi	Multi-target	90%	-
football.avi	Multi-target	90%	-
football2.avi	Multi-target	90%	-
penquan.avi	Multi-target	90%	-

▼Table 5. Efficiency of target detection and tracking

File name	Scene type	*fps* of MT mode	*fps* of S mode
corridor.avi	Single target		>178
office1p.avi	Single target	-	>175
office2p.avi	Single target	-	>141
outdoor.avi	Single target	-	>200
runner.avi	Single target	-	>178
basketball.avi	Multi-target	>55	-
football.avi	Multi-target	>70	-
football2.avi	Multi-target	>65	-
penquan.avi	Multi-target	>55	-

the single target property. In the MT mode, the efficiency of the proposed method is also higher than 25 *fps*.

6 Conclusions

In summary, the proposed method works well on detecting and tracking moving objects in mobile phone video sequences. The proposed target detection method is based on the 3-frame difference method and GMM. The tracking method is a combination of the LBP feature and camshift tracker. Auto-focusing is realized by using the coordinates of the detection and tracking modules. The proposed method can perform well in many multi-target scenes. The experiments of detection and tracking show that the proposed method can be used to achieve the function of non-contact auto-focusing in mobile devices.

References

[1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for Real-time tracking," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, USA, Jun. 1999, pp. 246–252. doi: 10.1109/CVPR.1999.784637.
[2] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "Time-domain filter bank analysis: a new design theory," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1412–1429, 1992.
[3] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proc. Eleventh ACM International Conference on Multimedia*, Berkeley, USA, Nov. 2003. doi: 10.1145/957013.957017.
[4] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, Aug. 2002. doi: 10.1109/TIP.2002.1014998.
[5] J. M. Menon and L. J. Stockmeyer, "Garbage collection in log-structured information storage systems using age threshold selection of segments," U.S. Patent No. 5,933,840, Aug. 3, 1999.
[6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, USA: Prentice Hall, 2002.
[7] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," *Analysis and Modeling of Faces and Gestures*. Berlin, Germany: Springer Berlin Heidelberg, 2007, pp. 235–249. doi: 10.1007/978-3-540-75690-3_18.
[8] G.-Y. Gong, W. Z. He, and X.-H. Gao, "Optimized mountain climb-searching of auto-focusing in infrared imaging system," *Laser & Infrared*, vol. 11, no. 026, 2007.
[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp.603–619, 2002. doi: 10.1109/34.1000236.

Manuscript received: 2015-11-16

Biographies

HU Rongchun (hrc@swust.edu.cn) received his master's degree in information and communication engineering from University of Science and Technology of China (UESTC) in 2007. He is a lecture and pursuing the Ph.D. degree in signal and information processing at UESTC. His research interests include machine learning and image processing.

WANG Xiaoyang (xywang_2012@163.com) received her B.E. degree in electronic science and technology from University of Electronic Science and Technology of China. She is currently a Ph.D. candidate in signal and information processing there. Her research interests include image processing, computer vision, and compressive sensing theory and applications.

ZHENG Yunchang (zhengyunchang@foxmail.com) received the B.E. and M.S. degrees in electronic science and technology from University of Electronic Science and Technology of China. His research interests include machine learning and computer vision.

PENG Zhenming (zmpeng@uestc.edu.cn) received his Ph.D. degree in geo-detection and information technology from Chengdu University of Technology, China in 2001. From 2001 to 2003, he was a postdoctoral researcher with the Institute of Optics and Electronics (IOE), Chinese Academy of Sciences. He is a professor with University of Electronic Science and Technology of China. His research interests include image processing, radar signal processing, and target recognition and tracking. Prof. PENG is a member of the IEEE and the Aerospace Society of China.

DOI: 10.3969/j. issn. 1673-5188. 2017. 01. 010
http://www.cnki.net/kcms/detail/34.1294.TN.20170213.1000.002.html, published online February 13, 2017

Nonbinary LDPC BICM for Next-Generation High-Speed Optical Transmission Systems

LIN Changyu[1], Ivan B. Djordjevic[1], WANG Weiming[2], and CAI Yi[3]

(1. Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721, USA;
2. ZTE Inc., 6 Huashiyuan Rd, Donghu H.D.Z., Wuhan 430223, China;
3. ZTE (TX) Inc., 55 Madison Ave., Morristown, NJ 07960, USA)

Abstract

We propose a nonbinary byte-interleaved coded-modulation scheme with inner and outer turbo-like iterative decoder. The net coding gain is 0.6 dB higher than the state-of-the-art binary single parity check (SPC) low-density parity-check (LDPC) based turbo-product counterpart, with adjustable iterations and lower error-floor. We provide the details of Bahl-Cocke-Jelinek-Raviv (BCJR) based inner code decoder and optimum signal constellation design (OSCD) method in this paper. The single-mode fiber (SMF) channel simulation is also discussed.

Keywords

fiber optics and optical communications; modulation; forward error correction (FEC); coded modulation

1 Introduction

The capacity demands due to Internet traffic growth is increasing exponentially. After the 100 Gb/s Ethernet (100GbE) standard has been adopted, the higher data rate age is coming for the next-generation Ethernet standards, possibly for 400 GbE and 1TbE. The most significant issues of the current optical networks include limited bandwidth resources and nonlinearity effects in transmission. For the given physical links and network equipment, an effective solution to the optical signal noise ratio (OSNR) requirement is based on forward error correction (FEC), as the response to the demands of high speed reliable transmission.

The low-density parity-check (LDPC) codes with large coding gains have been considered as the promising solution to the FEC. However, using the LDPC codes only might encounter error floor issues, especially for very high rate LDPC codes. Hard decision (HD) algebraic codes, such as Bose-Chaudhuri-Hocquenghem (BCH) and Reed-Solomon (RS) code, are typically used for clearing error floors. With the stronger computation capacity, the soft decision (SD) decoding algorithm enables more possibilities. Traditionally, the inner code is LDPC code, the outer code can be a shorter simple code such as BCH and RS. With SD decoding enabled, it has been showed that the reversed order decoding can provide a larger coding gain [1]. Considering the linearity of both inner and outer codes, the coding gain can be further increased with iterative decoding as a turbo product code (TPC). Although the single parity check (SPC) code, a special case of BCH code, cannot guarantee to correct any error in HD decoding, it can be a very good candidate in SD TPC decoding, which increases the accuracy of the input probabilities to LDPC decoder. Recently, this idea has been experimentally verified in [2] and a large coding gain is obtained.

To further improve the coding performance, we consider the nonbinary (NB) coding scheme, which has already been proved to have a larger gain especially for higher modulation formats [3]. The nonbinary LDPC code is getting more popular with the advanced hardware and it is proved to be hardware friendly in [5], in which the SD nonbinary LDPC and HD-RS code is implemented in FPGA. As a consequence, instead of binary SPC and LDPC, we consider a larger Galois Field (GF) with four elements, which is a good compromise between performance and complexity [4], and this method can be generalized to any size GF.

Without putting additional redundancy to the data, extra gain can be obtained by using turbo equalizer (TE) decoding scheme [6]. The number of iterations can be adjusted based on the quality demands. To verify the robustness of the proposed nonbinary scheme, a high nonlinear polarization-division-multiplexed (PDM) single-mode fiber (SMF) transmission has been simulated. We also consider the suboptimal and fast algorithm for the decoder, which significantly decreases the complexity with little reduction in performance.

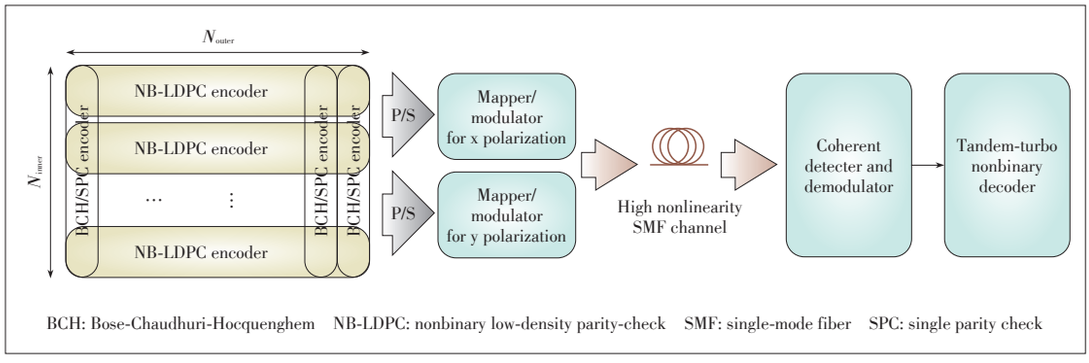
2 Proposed Tandem-Turbo-Product Nonbinary Coding Scheme

The encoder of the proposed nonbinary (NB) tandem-TPC code is shown in **Fig. 1**. Based on two linear codes, NB LDPC and NB BCH/SPC, we encode each row as a codeword of NB LDPC code and each column is a codeword of NB BCH/SPC code. The overall coded block size is N_{inner}-by-N_{outer} with elements from GF(q). The encoded nonbinary block is then passed to the PDM-16QAM modulator with Gray mapping.

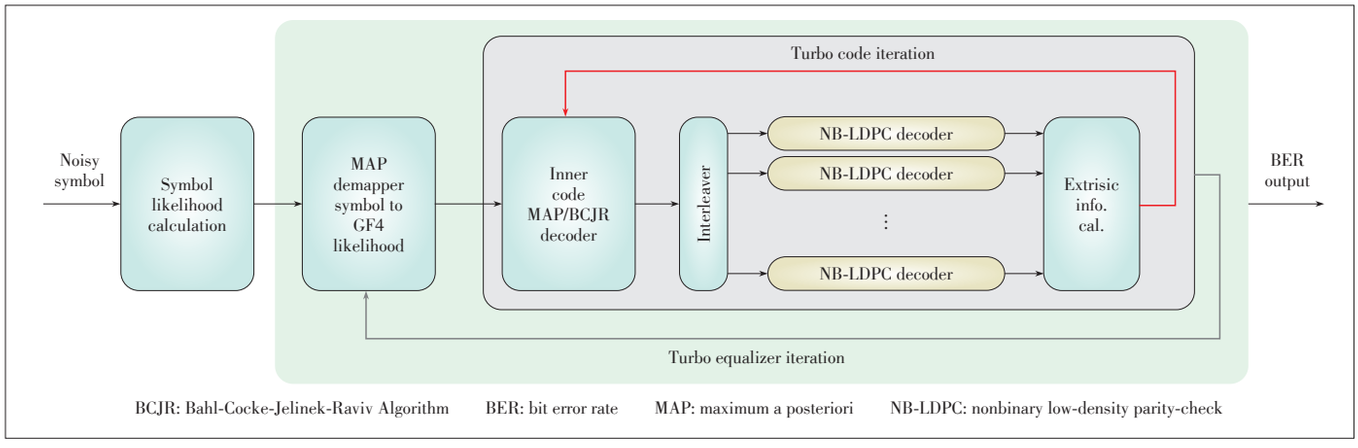
The structure of the decoder is provided in **Fig. 2**. The PDM-16QAM symbol likelihood is calculated for the following tandem-turbo-product decoder. At the first iteration, there is no

Nonbinary LDPC BICM for Next-Generation High-Speed Optical Transmission Systems

LIN Changyu, Ivan B. Djordjevic, WANG Weiming, and CAI Yi



◀Figure 1. System diagram of the tandem-turbo-product coded modulation scheme over high nonlinearity SMF channel.



▲Figure 2. Tandem-turbo-product nonbinary BICM decoder with adjustable iterations.

extrinsic information for the TE, so the demapper simply computes the GF(q) log-likelihood ratio (LLR) for the TPC decoder. The TPC is consisted of two NB linear block codes, BCH and LDPC, and the rows are decoded as LDPC code with row-layered sum-product algorithm and the columns are decoded by maximum a posteriori (MAP) decoder. The MAP decoder can be done with look-up table (LUT) or Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. The LLRs are updated with both the decoders in each iteration as typical in TPC code. The TE comes to help if the output of the TPC is unsatisfactory at the first round. The extrinsic information of the TPC is remapped to symbol level and attenuated before being combined with the channel symbol likelihood in the demapper. The updated LLRs after the demapper are passed to the TPC decoder and the first TE decoding iteration is started. There are three adjustable iterations, LDPC iteration, TPC iteration, and TE iteration and the trade-off between complexity and performance can be further optimized for further performance improvement.

The symbol log-likelihood ratio (SLLR) of the transmitted symbol can be updated by turbo equalizer as

$$\lambda(s_i)=\log\left(\frac{P(s_i|r)}{P(s_0|r)}\right)=\log\left(\frac{P(r|s_i)P(s_i)}{P(r|s_0)P(s_0)}\right)=\log\left(\frac{P(r|s_i)}{P(r|s_0)}\right)+\lambda_{ext}(s_i), \quad (1)$$

where the s_i is the transmitted symbol ($i=0,1,\ldots,15$ *for* $16QAM$), r is the received symbol and s_0 is the referent symbol point. The second equation is due to Bayes' theorem and $\lambda_{ext}(s_i)$ is the prior reliability of symbol s_i, which can be computed as

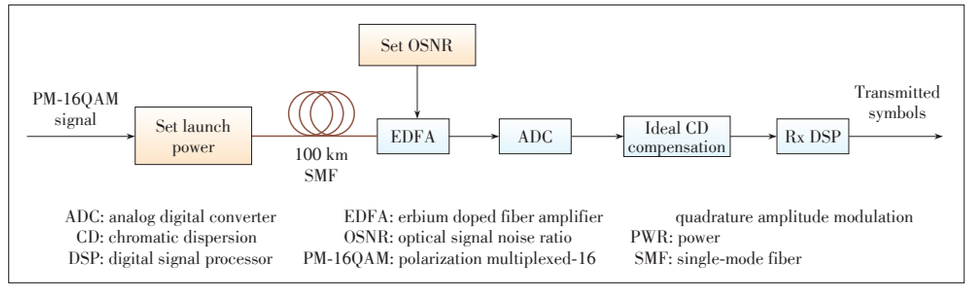
$$\lambda_{ext}(s_i)=\log\frac{P(s_i)}{P(s_0)}=\sum_{j=1,c_j\in s_i}^{m}\log\left(\alpha_t-L_p(c_j)\right), \quad (2)$$

in which $L_p(c_j)$ is the prior GF(q) LLR and c_j is the GF(q) representation of the mapped symbol. α_t is the attenuation coefficient for the t th TE iteration. The prior/extrinsic LLR is the difference between output and input of the TPC decoder as $L_p(c_j)=L(c_j^{out})-L(c_j^{in})$ and the GF(q) LLR for $k\in\{0,1,\ldots,q-1\}$ can be obtained by

$$L(c_j=k)=\log\frac{\sum_{c_j=k,c_j\in s_i}\lambda(s_i)}{\sum_{c_j=0,c_j\in s_0}\lambda(s_0)}. \quad (3)$$

The summation in $L(c_j)$ in (3) can be calculated by max* [7] operation or be simplified by replacing the max* operation by max-operation only, by ignoring the correction terms, which can significantly reduce the complexity. The TPC decoder accepts the GF(4) LLRs and begin the TPC iteration, and in each

TPC iteration, the inner code decoder and LDPC decoder exchange decoded LLRs to improve the performance. The (vertical) inner code is NB SPC/BCH code with BCJR decoder, the trellis decoding algorithm is based on MAP decoding rule, which is of low complexity for SPC. The (horizontal) outer code is NB LDPC code, and it is decoded by row-layered sum-product algorithm.



▲**Figure 3. PM-16QAM SMF transmission link setup with high nonlinearity effect.**

3 SMF Modeling and Channel Simulation

The SMF is modeled by solving the nonlinear Schrödinger equation (NLSE) with split-step (Fourier) method. The total length of SMF is 100 km with step size 100 m, the attenuation is 0.2 dB/km, the dispersion coefficient is 17 ps/nm/km and the nonlinear parameter is $1.2\,(\mathrm{W\cdot km})^{-1}$. Simulation of the channel is done for both Additive White Gaussian Noise (AWGN) channel and SMF-based transmission system (**Fig. 3**). For SMF simulation, the transmitter side uses eight times upsampling with ideal pre-filter. The power of the PDM-16QAM signal is set before launching into the SMF. A Gaussian noise is loaded after the transmission of SMF with a given OSNR. On the receiver side, the ideal chromatic dispersion (CD) compensation is done before the typical digital signal processor (DSP) blocks and the transmitted symbols are passed to the tandem-turbo-product decoder.

4 Optimal Signal Constellation Design and Performance

New constellation points are obtained as the center of mass of such obtained clusters. This procedure is repeated until convergence or until the predetermined number of iterations has been reached. It can be shown that this algorithm is optimum in minimum mean square error (MMSE) sense.

The MMSE-optimum signal constellation design (OSCD) algorithm can be formulated as follow:

1) Initialization: Choose an arbitrary auxiliary input distribution. Choose an arbitrary signal constellation as initial constellation and set the size of this constellation to M.
2) Apply the Arimoto-Blahut algorithm to determine optimum source distribution.
3) Generate long training sequences $\{x_j; j=0, \cdots, n-1\}$ from optimum source distribution, where n denotes the length of the training sequence used for signal constellation design. Let A_0 be the initial M-level signal constellation set of subsets of constellation points.
4) Group the samples from this sequence into M clusters. The membership to the cluster is decided by Euclidean distance squared of sample point and signal constellation points from previous iteration. Each sample point is assigned to the cluster with smallest distance squared. Given the mth subset (cluster) with N candidate constellation points, denoted as $\hat{A}_m=\{y_i\,;\,i=1,\ldots,N\}$, find the MMSE of partition $P(\hat{A}_m)=\{S_i\,;\,i=1,\ldots,N\}$, as follows

$$D_m = D\left(\left\{\hat{A}_m, P\left(\hat{A}_m\right)\right\}\right) = n^{-1}\sum_{j=0}^{n-1}\min_{y\in\hat{A}_m} d(x_j,y), \tag{4}$$

where d is Euclidean distance squared between the jth training symbol and symbol y being already in the subset (cluster). With $D(\cdot)$, we denote the distance function.
5) If the relative error $|D_{m-1}-D_m|/D_m \leqslant e$, where e is the desired accuracy, the final constellation is described by $\{\hat{A}_m\}$. Otherwise continue.
6) Determine the new constellation points as the center of the mass for each cluster. With the mean square-error criterion, $x(S_i)$ is the Euclidean center of gravity or centroid given by

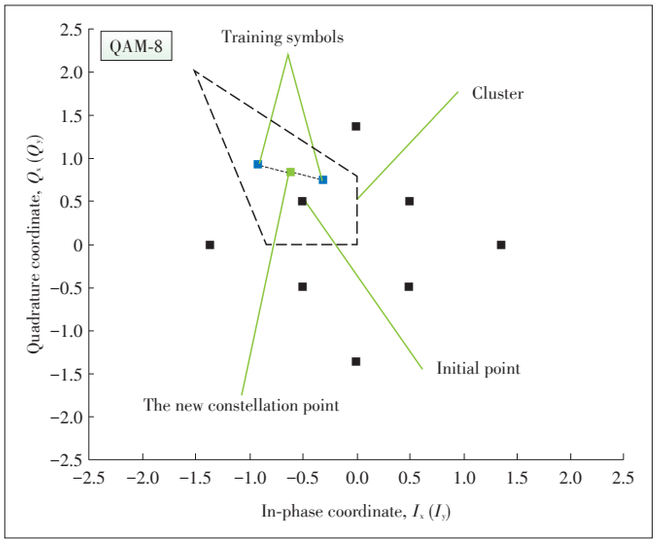
$$x(S_i) = \frac{1}{||S_i||}\sum_{j:x_j\in S_i} x_j\,, \tag{5}$$

where $||S_i||$ denotes the number of training symbols in the region S_i as shown in **Fig. 4**. If there is no training sequence in the region, set $x(S_i)=y_i$, the old constellation point. Define $\hat{A}_{m+1}=x(P(\hat{A}_m))$, replace m by $m+1$, and go to step 3.

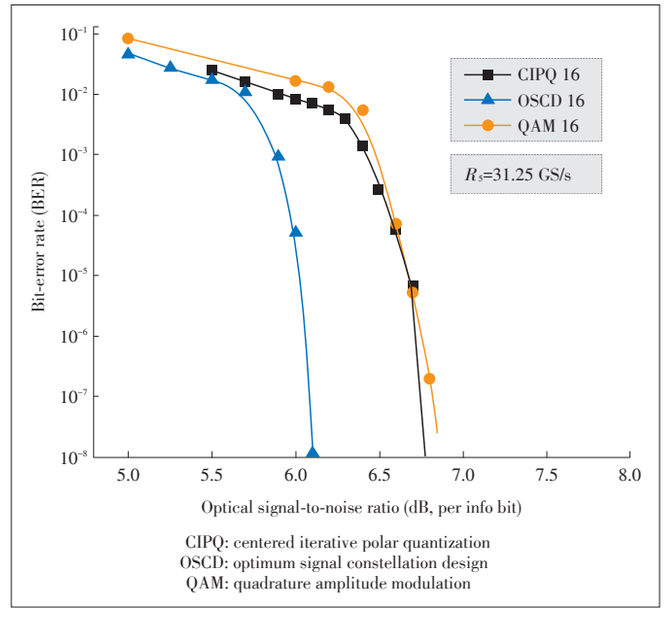
Repeat the steps 4–6 until convergence.

Fig. 5 shows the signal constellations obtained for following signal constellation sizes: 16, 32, 64 and 128. The results are obtained for ASE noise dominated scenario.

It should be noticed that these signal constellations remind to the format of IPQ-signal constellations, except for the center point. Alternatively, the IPQ-approach can be used by placing the first single point in the origin and then the IPQ-procedure is applied. Howeverthat the IPQ-procedure uses some approximations to come up with closed form solutions, which are valid assumptions for reasonable large signal constellation sizes. Therefore, it is a suboptimum solution for medium signal constellation sizes. We will later show that signal constellations obtained by MMSE-OSCD algorithm significantly outperform IPQ-inspired signal constellations containing the point located in



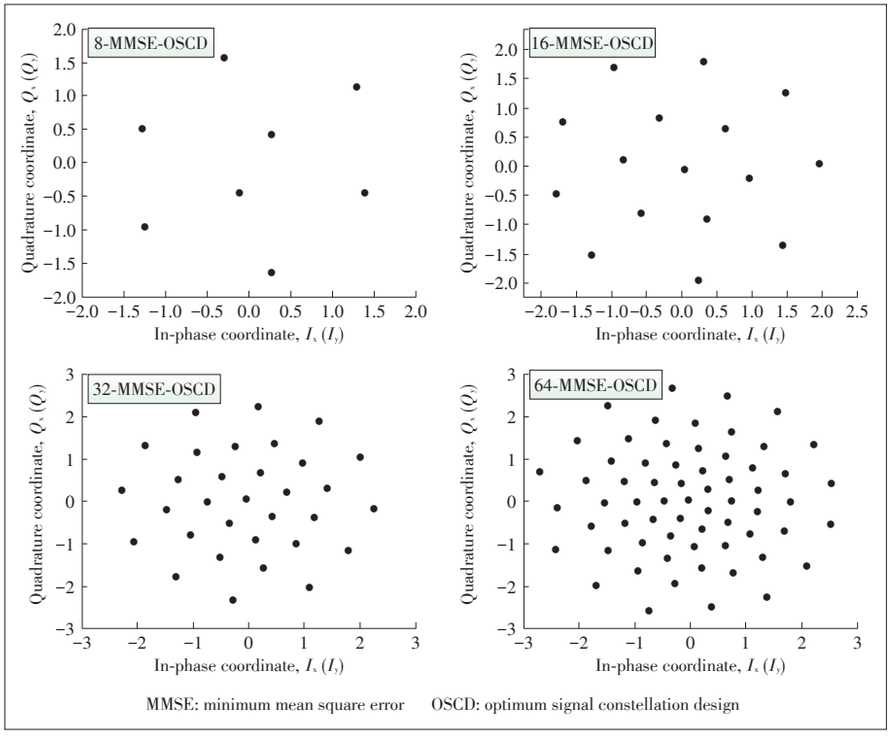
▲Figure 4. The illustration of the OSCD algorithm.



▲Figure 6. BER performance of proposed MMSE-OSCD algorithm based constellations against QAM and CIPQ.

the origin.

The results of Monte Carlo simulations of proposed MMSE-OSCD-algorithm based constellations are summarized in **Fig. 6**. When measured at BER of 10^{-8}, the 16-ary MMSE-OSCD algorithm based signal constellation outperforms 16-QAM by almost 1 dB. The channel symbol rate was set to 31.25 GS/s, and QC LDPC (16935, 13550) of girth-8 and column-wieght-3 was used in simulations. The improvement of MMSE-OSCD over centered iterative polar quantization (CIPQ) and quadrature amplitude modulation (QAM) decreases as the signal constellation size grows.



▲Figure 5. The MMSE-OSCD constellation.

5 MAP/BCJR Decoder for the Inner Code

Optimum or sub-optimum MAP algorithms are usually good candidates to be used in "soft-in/soft-out" decoders. However, the MAP algorithm, especially for nonbinary codes, is a computationally complicated decoding method. Here we present a near optimum MAP decoding rule for nonbinary codes based on the dual space of the code. Because the complexity of this proposed algorithm is related to the inverse of the code rate, it can be attractive for the codes with high coding rates. The shorter code word length is also a plus for implementation consideration.

Here we focus on the inner code only and start with some notation introduction. Vectors and matrices are denoted in boldface letters and their elements in lower case, e.g., v_i is the i th element of the vector $v=(v_0,v_1,\cdots,v_{N-1})$. Also an (N, K) code is referred to a linear block code with the length of N and dimension of K with the parity-check matrix that is denoted by H. The codes are defined over a finite field of order q, F_q, therefore the elements of codewords belong to set {0,1, ... , q-1}. Code-

words are transmitted over a discrete time memory-less and noisy channel, and the soft decision received vector for a codeword is denoted by $r=(r_0,r_1,\ldots,r_{N-1})$.

The APP algorithm for the block code can be developed as

$$\hat{v}_i=\arg\max_{v_i\in\{0,1,\cdots,q-1\}}\Pr(v_i|r,v\cdot H^T=0)\ . \tag{6}$$

Based on typical BCJR algorithm, the equation above is equal to

$$\hat{v}_i=\arg\max_{v_i\in\{0,1,\cdots,q-1\}}\sum_{(s',s)\in\varepsilon_{v_i}^i}\alpha_i(s')\gamma_i(s',s)\beta_{i+1}(s)\ , \tag{7}$$

where is a state at the i th-level of the code trellis and is a state at the $(i+1)^{\text{th}}$-level of the code trellis. Moreover, $\varepsilon_{v_i}^i$ is the set of all edges of the trellis of the code, which connects the states to the states and corresponds to the code symbol v_i. If $r_{i,j}=(r_i,r_{i+1},\ldots,r_{j-1})$, for $0\leqslant i\leqslant j\leqslant N$, we then have

$$\alpha_i(s)=\Pr(s_i=s,r_{0,i})\ , \tag{8}$$

$$\beta_i(s)=\Pr(r_{i,N}|s_i=s)\ , \tag{9}$$

$$\gamma_i(s',s)=\Pr(s_{i+1}=s,r_i|s_i=s')\ . \tag{10}$$

Let $\Omega_{i-1}^{(c)}(s)$ be as a set of all states at the $(i-1)th$-level of the code trellis, which are adjacent to state s. According to the BCJR algorithm for state at ith-level of the trellis, and can be calculated recursively.

For $0\leqslant i\leqslant N$,

$$\alpha_i(s)=\sum_{s'\in\Omega_{i-1}^{(c)}(s)}\alpha_{i-1}(s')\gamma_i(s',s)\ . \tag{11}$$

By knowing the fact that $\alpha_o(s_0)=1$, all α_i for $0\leqslant i\leqslant N$ can be calculated. This is called the forward recursion. The backward recursion is similarly obtained.

For $0\leqslant i\leqslant N$,

$$\beta_i(s)=\sum_{s'\in\Omega_{i+1}^{(d)}(s)}\gamma_i(s,s')\beta_{i+1}(s')\ , \tag{12}$$

where $\Omega_{i+1}^{(d)}(s)$ is a set of all states at $(i+1)$ th-level of the code trellis and these states are adjacent to state s. Using the fact that $\beta_N(s_f)=1$, β_i all $0\leqslant i\leqslant N$ for can be found. The branch transition probability for a block code with statistically independent information bits can be written as

$$\begin{aligned}\gamma_i(s,s')&=\Pr(s_{i+1}=s,r_i|s_i=s')=\\&\Pr(s_{i+1}=s|s_i=s')\Pr(r_i|s_{i+1}=s,s_i=s')=\\&\Pr(v_i)\Pr(r_i|v_i).\end{aligned} \tag{13}$$

Different methods have been suggested to carry out the MAP decoding algorithm, but in all of these methods, there are three common major steps:

1) Perform the forward recursion process and store all the values calculated for α_i for $0\leqslant i\leqslant N$.
2) Perform the backward recursion process and store all the values calculated for β_i for $0\leqslant i\leqslant N$.
3) For each received codeword symbol, r_i, find the MAP probability using (2). To calculate (2), the transition probabilities are needed.

Due to the independence of forward and backward recursions from each other, step 1 and step 2 can be done simultaneously.

6 Results and Discussion

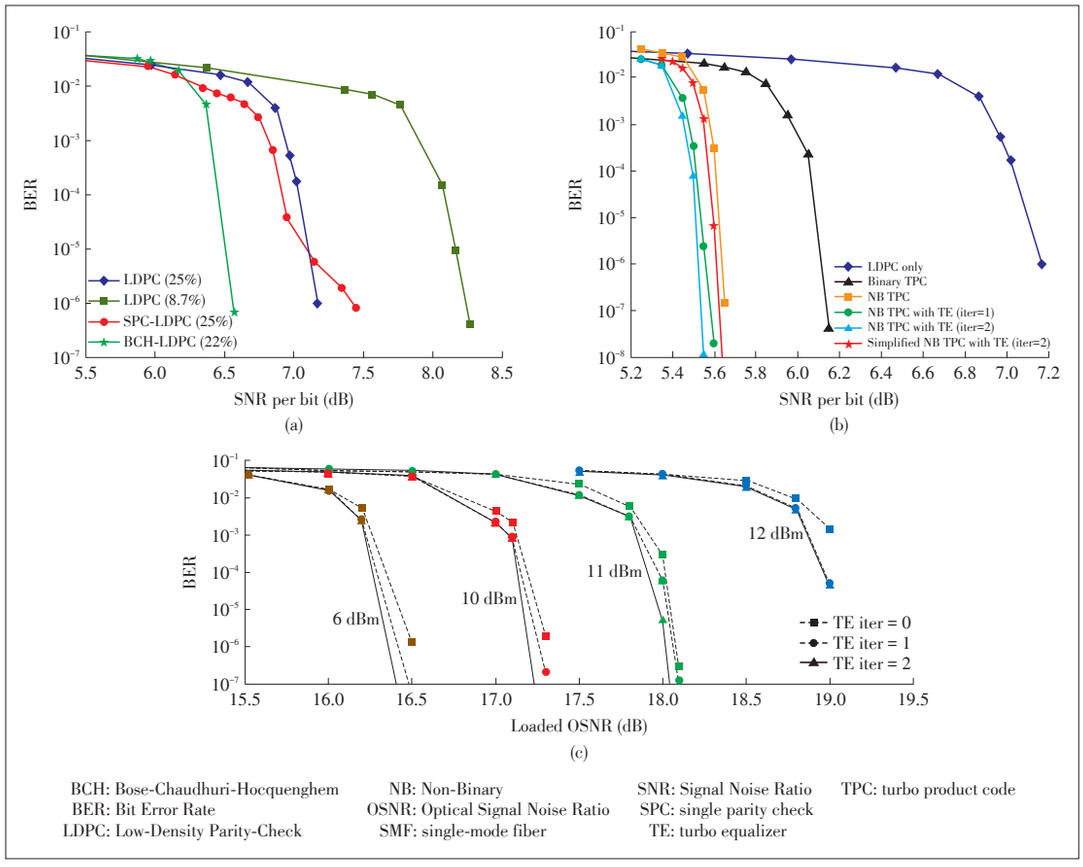
For binary coding, **Fig. 7a** shows that in the BER performance TPC (SPC (7,8) and LDPC (17104,18611)) of 25% overhead outperforms LDPC (17104,18611). Though this TPC performs even better than the LDPC (13550,16935) with the same overhead in the low SNR region, it begins to show some error floor at $\text{BER}=10^{-5}$. By replacing the SPC with BCH (57,64), about 0.3 dB gain is obtained compared to the SPC-LDPC case without any error floor and with 22% overhead.

In **Fig. 7b**, the nonbinary TPC (NB-SPC (7, 8) and NB-LDPC (6744, 8430)) has about 0.4 dB gain in OSNR compared to the binary counterpart (SPC (7,8) and LDPC (13550, 16935)) with half the length. With two TE iterations, we are able to extend the gain to 0.6 dB without increasing the overhead. It should be noticed that both binary and nonbinary performance are evaluated at the 4-th TPC iteration, and the maximum number of iterations for LDPC of are 30 and 18 for binary and nonbinary LDPC codes, respectively. When reducing the $\max^*$-operation to max-operation of the nonbinary TPC code with TE, only less than 0.2 dB loss in SNR is observed.

With four different launch powers, ranging from 6 dBm to 12 dBm, the proposed tandem-TPC scheme is tested with studied SMF link. After the transmission, to achieve the same BER performance, the higher linear ONSR is required for larger nonlinear transmission cases. **Fig. 7c** shows the BER performance of nonbinary twin-turbo code (NB-SPC (7, 8) and NB-LDPC (6744, 8430)) with TE, and it demonstrates excellent robustness even in highly nonlinear scenarios.

7 Conclusions

We proposed the nonbinary tandem-TPC-TE based coded modulation scheme. When TE and TPC are employed in tandem, a smaller number of iterations are needed for NB LDPC with half the length compared to the binary counterpart. The proposed NB TPC provides 0.6 dB improvement in Net Coding Gain (NCG) for the same BER performance with adjustable iteration/performance. The complexity is even lower when low-complexity decoding algorithm is used with small loss in gain.



◀Figure 7.
BER performance in: (a) BCH vs SPC based TPC; (b) nonbinary TPC vs binary TPC; and (c) SMF nonlinear transmission with loaded linear noise.

The robustness has been verified in highly nonlinear PDM-16QAM SMF transmission. The proposed NB TPC represents a promising NB coded-modulation scheme for the next generation optical transmission systems.

References

[1] I. B. Djordjevic, L. Xu, and T. Wang, "Reverse concatenated coded modulation for high-speed optical communication," *IEEE Photonics Journal*, vol. 2, no. 6, pp. 1034–1039, Dec. 2010. doi: 10.1109/JPHOT.2010.2091678.

[2] H. G. Batshon and H. Zhang, "Multidimensional SPC-based bit-interleaved coded-modulation for spectrally-efficient optical transmission systems," in *Proc. SPIE 9008, Optical Metro Networks and Short-Haul Systems VI, 90080F*, San Francisco, USA, 2013. doi: 10.1117/12.2037021.

[3] M. Arabaci, I. B. Djordjevic, L. Xu, and T. Wang, "Nonbinary LDPC-coded modulation for high-speed optical fiber communication without bandwidth expansion," *IEEE Photonics Journal*, vol. 4, no. 3, pp. 728–734, Jun. 2012. doi: 10.1109/JPHOT.2012.2195777.

[4] D. Zou and I. B. Djordjevic, "FPGA implementation of concatenated non-binary QC-LDPC codes for high-speed optical transport," *Optics Express*, vol. 23, no. 11, pp. 14501–14509, May 2015. doi: 10.1364/OE.23.014501.

[5] Y. Zhang and I. B. Djordjevic, "Multilevel nonbinary LDPC-coded modulation for high-speed optical transmissions," in *Asia Communications and Photonics Conference*, Shanghai, China, Nov. 2014. doi: 10.1364/ACPC.2014.ATh1E.6.

[6] M. Cvijetic and I. B. Djordjevic, *Advanced Optical Communication Systems and Networks*. Boston, USA: Artech House, Jan. 2013.

[7] C. Lin, I. B. Djordjevic, and D. Zou, "Achievable information rates calculation for optical OFDM few-mode fiber long-haul transmission systems," *Optics Express*, vol. 23, no. 13, pp. 16846–16856, 2015. doi: 10.1364/OE.23.016846.

Manuscript received: 2016-12-31

Biographies

LIN Changyu (lcy@email.arizona.edu) received his Ph.D. from University of Arizona, USA. He is a senior optical engineer with Infinera Corporation. His research interests include OFDM system, MIMO signal processing, fiber nonlinearity, coded-modulation and quantum information.

Ivan B. Djordjevic (ivan@email.arizona.edu) is a tenured associate professor in the Department of Electrical and Computer Engineering, College of Engineering, University of Arizona, USA, with a joint appointment in the College of Optical Sciences. Prior to joining University of Arizona, he was with University of the West of England and University of Bristol, UK; Tyco Telecommunications, Eatontown, USA. His current research interests include optical communications/networks, error control coding, coded modulation, turbo equalization, OFDM applications, quantum information processing and quantum biology.

WANG Weiming (wang.weiming6@zte.com.cn) received his master's degree from Huazhong University of Technology. He is a senior algorithm/FPGA engineer with ZTE Corporation, China. His research interests include digital signal processing, error control coding and coded-modulation.

CAI Yi (Anthony.tsai@ztetx.com) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, USA, in 2001. Dr. Cai is a director of DSP with ZTE Optics Labs, USA. Prior to joining ZTE Corporation, he worked with Tyco Electronics Subsea Communications as an outstanding researcher. His research focuses on the application of coherent detection, forward error correction, and advanced modulation formats in optical fiber communications.

ZTE Communications Guidelines for Authors

• Remit of Journal

ZTE Communications publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

• Manuscript Preparation

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3000 to 8000, and no more than 8 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

• Abstract and Keywords

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

• References

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to ZTE Communications Editorial Style. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

• Copyright and Declaration

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors, b) the manuscript has not been published elsewhere in its submitted form, c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

• Content and Structure

ZTE Communications seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

• Peer Review and Editing

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

• Biographical Information

All authors are requested to provide a brief biography (approx. 100 words) that includes email address, educational background, career experience, research interests, awards, and publications.

• Acknowledgements and Funding

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

• Address for Submission

magazine@zte.com.cn
12F Kaixuan Building, 329 Jinzhai Rd, Hefei 230061, P. R. China

ZTE COMMUNICATIONS

ZTE Communications has been indexed in the following databases:

- Cambridge Scientific Abstracts (CSA)
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Inspec
- Ulrich's Periodicals Directory
- Wanfang Data—Digital Periodicals

ZTE COMMUNICATIONS
Vol. 15 No. 1 (Issue 55)
Quarterly
First English Issue Published in 2003

Supervised by:
Anhui Science and Technology Department

Sponsored by:
Anhui Science and Technology Information Research Institute and ZTE Corporation

---

Editorial Correspondence:
Add: 12F Kaixuan Building, 329 Jinzhai Road,
Hefei 230061, P. R. China
Tel: +86-551-65533356
Fax: +86-551-65850139
Email: magazine@zte.com.cn

Published and Circulated (Home and Abroad) by:
Editorial Office of *ZTE Communications*

Printed by:
Hefei Tiancai Color Printing Company

Publication Date:
February 25, 2017

Publication Licenses:
ISSN 1673-5188
CN 34-1294/ TN

Advertising License:
皖合工商广字0058号

Annual Subscription:
RMB 80