An Automatic Clustering Algorithm For Multipath Components Based On Kernel-Power-Density

Ruisi He[1], Qingyong Li[2], Bo Ai[1], Yang Li-Ao Geng[2], Andreas F. Molisch[3], Vinod Kristem[3], Zhangdui Zhong[1], and Jian Yu[2]
[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
[3]Department of Electrical Engineering, University of Southern California, Los Angeles, USA
Email: ruisi.he@ieee.org

***Abstract*—In the real-world environments, multipath components (MPCs) of wireless channels are generally distributed as groups, i.e., clusters. Modeling the clustered MPCs is important and necessary for channel modeling and an automatic clustering algorithm is thus required. This paper proposes a novel Kernel-power-density (KPD) based algorithm for MPC clustering. It uses the Kernel density to incorporate the modeled behavior of MPCs and takes into account the power of the MPCs. The proposed algorithm only considers the K nearest MPCs in the density estimation to better identify the local density variations of MPCs. Simulations validate the KPD algorithm and almost no performance degradation is found even with a large number of clusters and large cluster angular spread. The KPD algorithm enables applications with no prior knowledge about the clusters such as number and initial locations. It can be used for the cluster based channel modeling for 4G/5G communications.**

I. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) systems have been widely used to increase data rates and improving quality of wireless transmission [1]. Realistic channel models are needed in MIMO system design such as transceiver designs and performance analysis [2]–[7]. Recently, clustered based channel modeling has been an important trend in the development of MIMO channels, as it has maintains accuracy while reducing complexity. A large body of channel measurements [8] has also validated that the multipath components (MPCs) are generally distributed in groups, i.e., clustered, in the real-world environments. Therefore, the cluster based channel models have been widely adopted in standardized channel models, such as COST 259 [9] and 3GPP Spatial Channel Model [10]. Modeling the clustered MPCs is thus important and necessary for channel modeling.

Even though the concept of clustered MPCs is widely accepted in channel modeling, finding good clustering algorithms is very much an open and research-active topic. Several clustering algorithms have been proposed over the past few years. In [11], MPCs are clustered within delay domain by using region competition algorithm. In [12], [13], a sparsity-based method is proposed to cluster MPCs, which exploits the feature that the power of the MPCs is exponentially decreasing with increasing delay. In [14], the K-Power-Means (KPM) algorithm is proposed. It considers the impact of MPC power in computing the cluster centers. In [15], the Fuzzy-c-means algorithm is used to cluster MPC and is found to outperform the KPM when using random initialization. In [16], a hierarchical agglomerative clustering algorithm is used to search for clusters jointly in the delay-angle-space domain.

Despite the impressive progress made in automated clustering, the existing works have several limitations: i) the attributes of MPCs are not well incorporated into the clustering algorithm, and ii) the number of clusters and many user-specified parameters are usually required. In general, an algorithm with fewer user-specified parameters and easier adjustment is needed for MPC clustering. In this paper, a novel clustering framework is proposed by using a density based method. A Kernel density is introduced to incorporate the modeled behavior of MPCs, and the proposed algorithm requires no prior knowledge of the number of clusters. This paper only presents the main idea of clustering algorithm and more analysis and detailed validations are presented in the extended version of [17].

The rest of the paper is organized as follows. Section II describes the wireless propagation channels. Section III shows the framework of the proposed clustering algorithm. Section IV validates the algorithm using simulations. Finally, Section V concludes the paper.

II. CHANNEL DESCRIPTION

We consider the double-directional channel model [18] in this paper, which contains the information of power α, delay τ, direction of departure (DOD) Ω_T, and direction of arrival (DOA) Ω_R of the MPCs. For each snapshot, the double-directional channel impulse response h can be expressed as (1), where M is the number of cluster and N_m is the number of MPCs in the m-th cluster. $\alpha_{m,n}$ and $\phi_{m,n}$ are the amplitude gain and phase of the n-th MPC in the m-th cluster, respectively. τ_m, $\Omega_{T,m}$, and $\Omega_{R,m}$ are the arrival time, DOD, and DOA of the m-th cluster, respectively. $\tau_{m,n}$, $\Omega_{T,m,n}$, and $\Omega_{R,m,n}$ are the excess delay, excess DOD, and excess DOA of the n-th MPC in the m-th cluster, respectively, where excess delay is usually taken with respect to the first component in the cluster, while excess angles are taken with respect to the mean. $\delta(\cdot)$ is the Dirac delta function and t is time.

$$h(t,\tau,\Omega_T,\Omega_R)=\sum_{m=1}^{M}\{\sum_{n=1}^{N_m}\alpha_{m,n}e^{j\phi_{m,n}}\delta(\tau-\tau_m-\tau_{m,n})\times\delta(\Omega_T-\Omega_{T,m}-\Omega_{T,m,n})\times\delta(\Omega_R-\Omega_{R,m}-\Omega_{R,m,n})\} \quad (1)$$

$$\rho_x=\sum_{y\in K_x}\exp(\alpha_y)\cdot\exp\left(-\frac{|\tau_x-\tau_y|}{\left(\sigma_{\tau_y,y\in K_x}\right)^2}\right)\cdot\exp\left(-\frac{|\Omega_{T,x}-\Omega_{T,y}|}{\sigma_{\Omega_{T,y},y\in K_x}}\right)\cdot\exp\left(-\frac{|\Omega_{R,x}-\Omega_{R,y}|}{\sigma_{\Omega_{R,y},y\in K_x}}\right) \quad (2)$$

All the multipath parameters in (1) can be estimated by using high-resolution algorithms. Therefore, the clustering analysis can be applied to the extracted MPCs. In this paper, the set of all the MPCs for one snapshot is noted as Φ and each MPC is represented as x.

III. Clustering Algorithm

A Kernel-power-density (KPD) based algorithm for MPC clustering is proposed in this section. The KPD is a density-based clustering algorithm, it uses the Kernel density and only considers the neighboring points when computing the density. The steps of KPD are presented as follows:

(1) *Calculating Density*: For each MPC sample, say x, calculate the density ρ using the K nearest MPCs as in (2), where y is an arbitrary MPC that $y\neq x$. K_x is the set of the K nearest MPCs for the MPC x. $\sigma_{(\cdot),y\in K_x}$ is the standard deviation of the K nearest MPCs in the domain of $(\cdot)$. In (2), we use the Gaussian Kernel density for the delay domain as the physical channels does not favor a certain distribution of delay; we use the Laplacian Kernel density for the angular domain as it has been widely observed that the angle of MPC follows the Laplacian distribution [19]. The heuristical term of $\exp(\alpha)$ in (2) shows that MPCs with strong power increase the density, which is intuitive.

(2) *Calculating Relative Density*: For each MPC sample, calculate the relative density ρ^* using the K nearest MPCs' density, as follows:

$$\rho_x^*=\frac{\rho_x}{\max\limits_{y\in K_x\cup\{x\}}\{\rho_y\}}\ . \quad (3)$$

By using the relative density, we normalize the density over different regions, which ensures that different clusters have similar level of density. It can be seen from (3) that $\rho^*\in(0,1]$.

(3) *Searching Key MPCs*: For each MPC x, if $\rho_x^*=1$, label it as the key MPC $\hat{x}$. We thus obtain the set of key MPCs as follows:

$$\hat{\Phi}:=\{x|x\in\Phi,\rho_x^*=1\}\ . \quad (4)$$

The key MPCs can be considered as the initial cluster centroids.

(4) *Clustering*: For each MPC x, define its high-density-neighboring [20] MPC $\tilde{x}$ as:

$$\tilde{x}:=\mathop{\arg\min}_{y\in\Phi,\rho_y^*>\rho_x^*}\{d(x,y)\} \quad (5)$$

where d represents the Euclidean distance. Similar to the idea of density-reachable in [21], we connect each MPC to its high-density-neighboring MPC and the connectedness path is defined as

$$p_x:=\{x\to\tilde{x}\}\ . \quad (6)$$

We thus obtain a connectedness map, ζ_1, as follows:

$$\zeta_1:=\{p_x|x\in\Phi\}\ . \quad (7)$$

Note that two MPCs can be connected to each other over multiple paths. Those MPCs which are connected and reachable to the same key MPC in ζ_1 are grouped as one cluster.

(5) *Cluster Merging*: For each MPC, connect it to its K nearest MPCs and the connectedness path is defined as

$$q_x:=\{x\to y,y\in K_x\}\ . \quad (8)$$

We thus obtain another connectedness map, ζ_2, as follows:

$$\zeta_2:=\{q_x|x\in\Phi\}\ . \quad (9)$$

If i) two key MPCs are reachable in ζ_2 and ii) any MPC in any path connecting the two key MPCs has $\rho^*\geq\chi$, where χ is a density threshold, we merge the two key MPCs' clusters as one new cluster.

In the above KPD algorithm, K determines how many MPCs are used to calculate density. A small K reduces the size of local region and we use $K=\sqrt{T/2}$ as suggested in [22]; χ determines whether two clusters can be merged. A large χ leads to a large number of clusters. For simplicity, we suggest to set $\chi=0.8$, which is found to have a reasonable performance as reported in Section IV. Detailed analysis of parameter selection can be found in the extended version of this work in [17], where further algorithm validations with simulations and measurements are presented.

IV. Validation

To validate the KPD algorithm, we use a simulated channel based on the 3GPP Spatial Channel Model Extended (SCME) MIMO channel model [23] where the ground truth is available [1]. The F measure [24] is used to evaluate the clustering performance, which is a robust external quality measure that can be used to balance the precision and recall. The value of the F measure ranges from 0 to 1, and a larger value indicates higher clustering quality.

Fig. 1 and Fig. 2 show the details of KPD implementations. In Fig. 1, 5 clusters are generated and cluster 3 is close to cluster 4. As shown in Fig. 1(b), the estimated density ρ has a large dynamic range and it is difficult to identify cluster 1

[1] We disregard the elevation domain in simulation for simplicity.

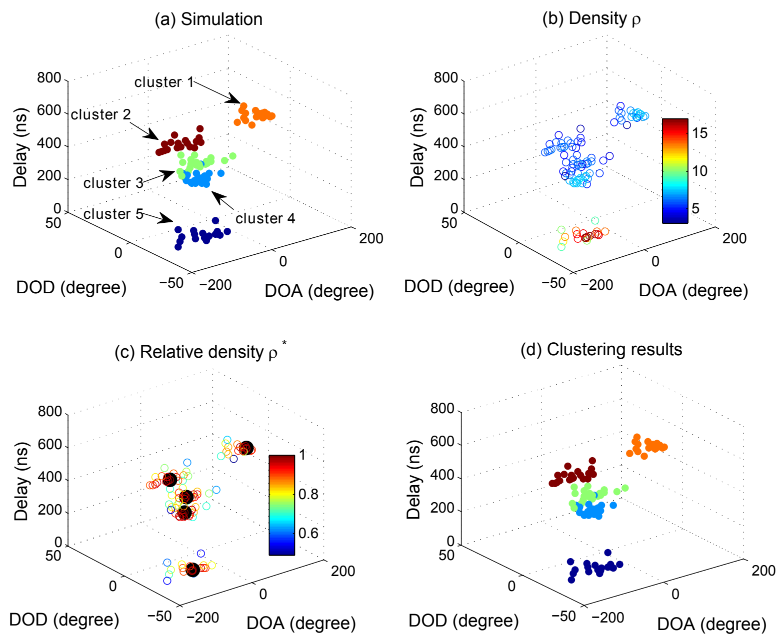

Fig. 1. Illustration of KPD clustering using the simulation with 5 clusters. (a) Simulated 5 clusters of MPCs. (b) Plots of the estimated density ρ, where the color bar indicates the level of ρ. (c) Plots of the estimated density ρ^*, where the color bar indicates the level of ρ^*. The solid black points are the key MPCs with $\rho^* = 1$. (d) Clustering results with the KPD algorithm, where the clusters are plotted with different colors.

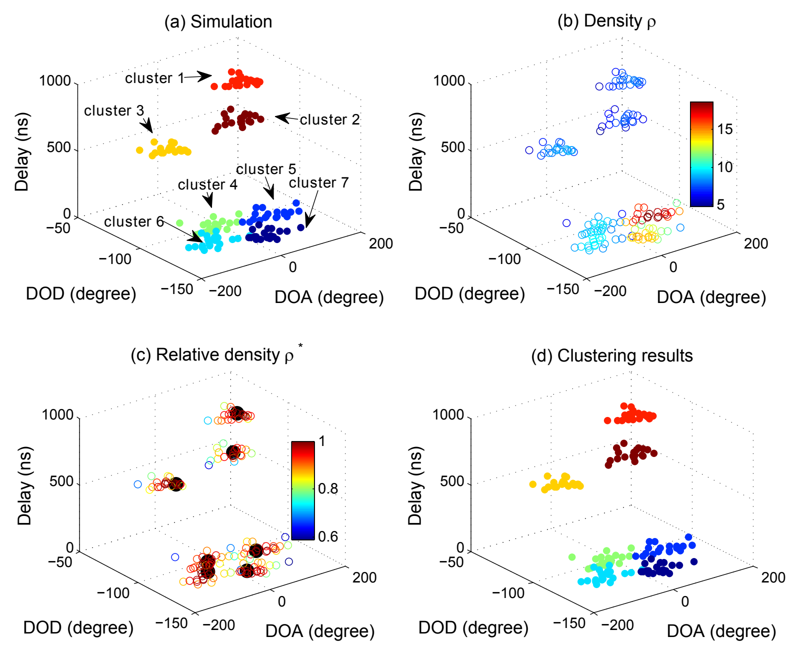

Fig. 2. Illustration of KPD clustering using the simulation with 7 clusters. (a) Simulated 7 clusters of MPCs. (b) Plots of the estimated density ρ, where the color bar indicates the level of ρ. (c) Plots of the estimated density ρ^*, where the color bar indicates the level of ρ^*. The solid black points are the key MPCs with $\rho^* = 1$. (d) Clustering results with the KPD algorithm, where the clusters are plotted with different colors.

and cluster 3 by setting a density threshold. However, after calculating the relative density (i.e., normalizing the local density), it is easier to identify each cluster by using the key MPCs, as shown in Fig. 1(c). The final clustering result in Fig. 1(d) has 100% correct identification. In Fig. 2, 7 clusters are generated and clusters 4-7 are close to each other. As shown in Fig. 2(b) and Fig. 2(c), the local density variations can be better observed by using the relative density. With KPD algorithm, all the 7 clusters are successfully identified in Fig. 2(d). Note that the step of clustering merging does not occur in the simulated channels of Fig. 1 and Fig. 2, however, this step is necessary to guarantee that an acceptable performance

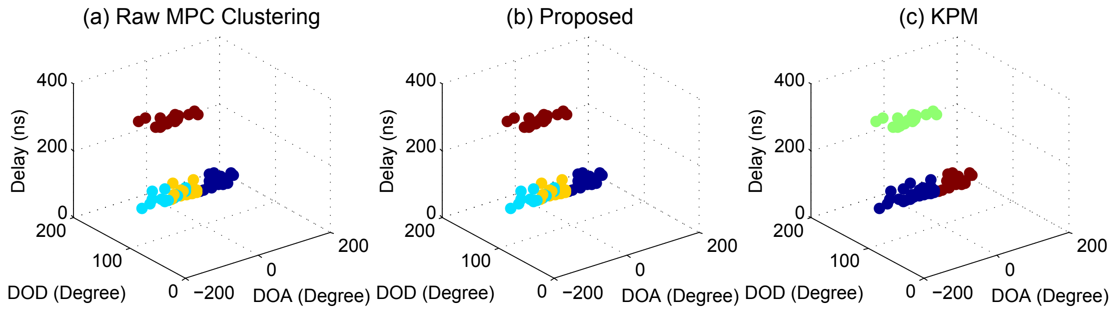

Fig. 3. Clustering algorithm validation with simulated channels with 4 clusters. (a) Simulated MPCs with SCME channel model, where the raw clusters are plotted with different colors. (b) Clustering results with the proposed KPD algorithm. (c) Clustering results with the KPM algorithm.

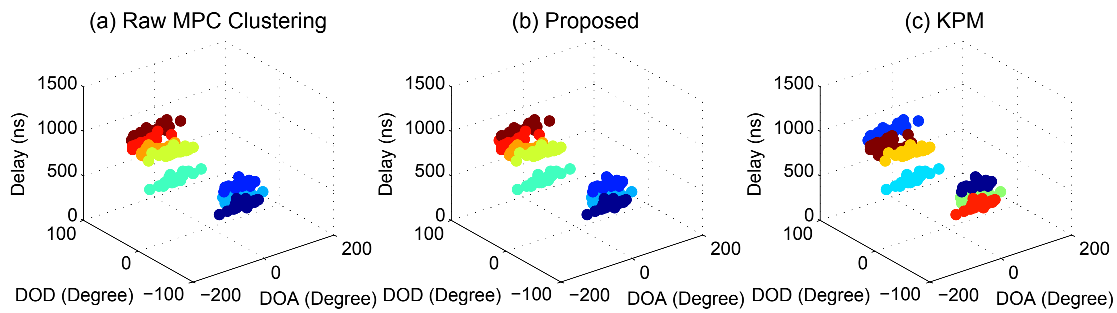

Fig. 4. Clustering algorithm validation with simulated channels with 8 clusters. (a) Simulated MPCs with SCME channel model, where the raw clusters are plotted with different colors. (b) Clustering results with the proposed KPD algorithm. (c) Clustering results with the KPM algorithm.

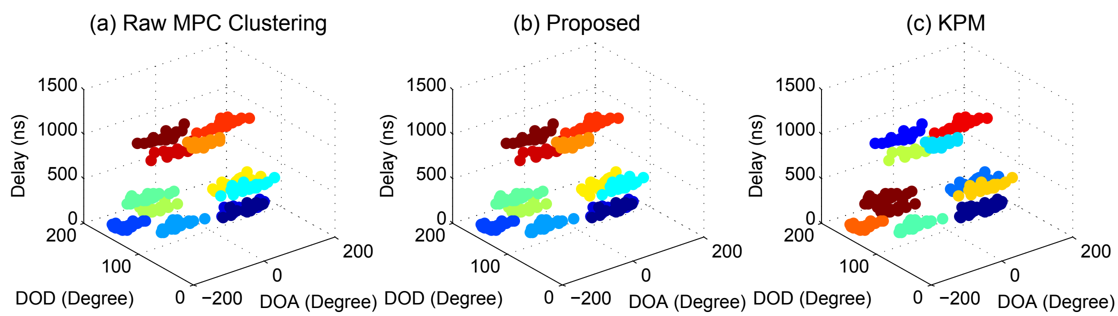

Fig. 5. Clustering algorithm validation with simulated channels with 12 clusters. (a) Simulated MPCs with SCME channel model, where the raw clusters are plotted with different colors. (b) Clustering results with the proposed KPD algorithm. (c) Clustering results with the KPM algorithm.

is achieved.

Figs. 3-5 show the raw clusters in the simulated channel and the clustering results by using the KPD and KPM algorithms for comparisons. We generated 4, 8, and 12 clusters with different power and delay/angular positions in Fig. 3, Fig. 4, and Fig. 5, respectively [2]. It can be seen that the KPM usually leads to a wrong cluster number decision: 3, 7, and 10 clusters are identified by KPM for Fig. 3, Fig. 4, and Fig. 5, respectively. The values of F measure are also summarized in Table I, where we can see that the KPD leads to a larger value of F measure and thus has a better performance. Note that the above simulation is affected by the randomly generated channels.

Furthermore, we test the performance of the algorithm under different cluster numbers and cluster angular spreads, and we still use the SCME channel model to generate MPCs. Fig. 6 shows the impact of cluster number, where 300 random channels are simulated for each cluster number case. It can be seen that the proposed KPD algorithm shows a better performance, and the value of the F measure decreases only slightly for larger cluster numbers. The KPM algorithm shows good performance only for a small number of clusters.

[2]Each figure is based on one random simulation of SCME channel model.

TABLE I

F MEASURE EVALUATION

Simulation	KPD	KPM
Fig. 1	1.00	0.85
Fig. 2	0.99	0.92
Fig. 3	0.96	0.91

Then we test the impact of cluster angular spread on the clustering accuracy. Different spreads are introduced by adding white Gaussian noise with variances of $\{1^\circ, 2^\circ, \ldots, 15^\circ\}$ to the MPCs DOA and DOD [25]. 300 random channels are simulated for each cluster angular spread case. In Fig. 7

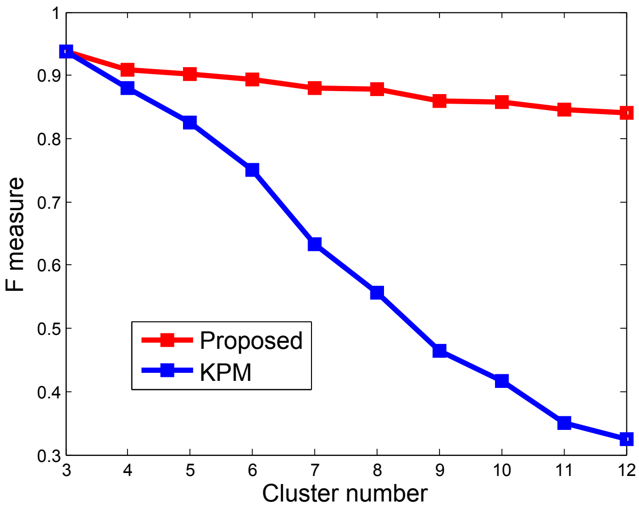

Fig. 6. Impact of cluster number on the F measure.

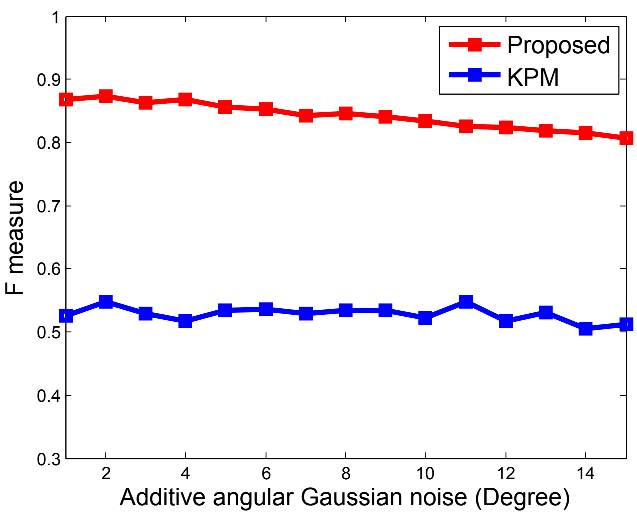

Fig. 7. Impact of cluster angular spread on the F measure, where 8 clusters are simulated.

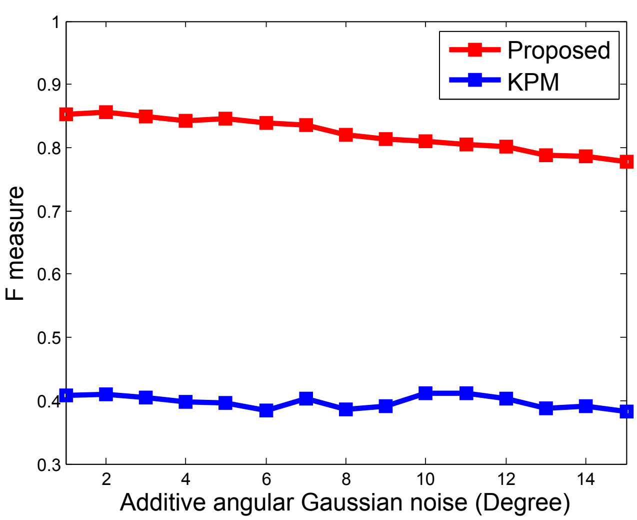

Fig. 8. Impact of cluster angular spread on the F measure, where 10 clusters are simulated.

and Fig. 8, the numbers of clusters are fixed to 8 and 10, respectively. It is noteworthy that the F measure generally decreases with the increasing cluster angular spread, and the KPD algorithm shows a better performance for arbitrary cluster sizes.

V. Conclusion

In this paper, a KPD based algorithm is proposed for MPC automatical clustering. It uses the Kernel density to incorporate the modeled behavior of MPCs and only considers the K nearest MPCs in the relative density estimation, which is able to better identify the local density variations of MPCs. A heuristic approach to merge clusters is introduced to improve the clustering performance. The synthetic MIMO channel data validates the proposed algorithm and it is found that the KPD algorithm provides a trustworthy clustering result with a small number of user input.

VI. Acknowledgments

This work is supported in part 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, and the Key Project from Beijing science and Technology Commission under Grant D151100000115004.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] M. Zhao, Y. Cai, Q. Shi, B. Champagne, and M. Zhao, "Robust transceiver design for MISO interference channel with energy harvesting," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4618–4633, 2016.

[3] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications-a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[4] Y. Cai, R. C. de Lamare, L.-L. Yang, and M. Zhao, "Robust MMSE precoding based on switched relaying and side information for multiuser MIMO relay systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5677–5687, 2015.

[5] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[6] Y. Cai, R. C. de Lamare, B. Champagne, B. Qin, and M. Zhao, "Adaptive reduced-rank receive processing based on minimum symbol-error-rate criterion for large-scale multiple-antenna systems," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4185–4201, 2015.

[7] L. Zhang, Y. Cai, R. C. de Lamare, and M. Zhao, "Robust multibranch tomlinson–harashima precoding design in amplify-and-forward MIMO relay systems," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3476–3490, 2014.

[8] A. F. Molisch and F. Tufvesson, "Propagation channel models for next-generation wireless communications systems," *IEICE Transactions on Communications*, vol. 97, no. 10, pp. 2022–2034, 2014.

[9] 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, 2006.

[10] 3GPP TSG RAN, "Spatial channel model for multiple input multiple output (MIMO) simulations," 3GPP, Tech. Rep., pp. 1-41, 2004.

[11] C. Gentile, "Using the Kurtosis measure to identify clusters in wireless channel impulse responses," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 6, pp. 3392–3395, 2013.

[12] R. He, W. Chen, B. Ai, A. F. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin, "On the clustering of radio channel impulse responses using sparsity-based methods," *IEEE Transactions Antennas and Propagation*, vol. 64, no. 6, pp. 2465–2474, 2016.

[13] R. He, W. Chen, B. Ai, A. Molisch, W. Wang, Z. Zhong, J. Yu, and S. Sangodoyin, "A sparsity-based clustering framework for radio channel impulse responses," in *Proc. IEEE VTC'16*, 2016, pp. 1–5.

[14] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo, "A framework for automatic clustering of parametric MIMO channel data including path powers," in *Proc. IEEE VTC'06*, 2006, pp. 1–5.

[15] C. Schneider, M. Bauer, M. Narandžić, W. Kotterman, and R. S. Thomä, "Clustering of MIMO channel parameters-performance comparison," in *Proc. IEEE VTC'09*, 2009, pp. 1–5.

[16] 3GPP TSG RAN WG1 R1-163115, "A hierarchical agglomerative clustering algorithm for channel modelling," 3GPP, Tech. Rep., pp. 1-6, 2016.

[17] R. He, Q. Li, B. Ai, Y. Geng, A. F. Molisch, V. Kristem, Z. Zhong, and J. Yu, "A Kernel-power-density based algorithm for channel multipath components clustering," *IEEE Transactions Wireless Communications*, under review.

[18] M. Steinbauer, A. F. Molisch, and E. Bonek, "The double-directional radio channel," *IEEE Antennas and Propagation Magazine*, vol. 43, no. 4, pp. 51–63, 2001.

[19] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 347–360, 2000.

[20] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Proc. ACM KDD'96*, 1996, pp. 226–231.

[22] Y. L. A. Geng, Q. Li, R. Zheng, F. Zhuang, and R. He, "RECOME: a new density-based clustering algorithm using relative KNN kernel density," in *Proc. SIAM Data Mining*, under review.

[23] D. S. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *Proc. IEEE VTC'05*, 2005, pp. 3132–3136.

[24] B. Larsen and C. Aone, "Fast and effective text mining using linear-time document clustering," in *Proc. ACM SIGKDD ICKDDM'99*, 1999, pp. 16–22.

[25] N. Czink, P. Cera, J. Salo, E. Bonek, J.-P. Nuutinen, and J. Ylitalo, "Automatic clustering of MIMO channel parameters using the multi-path component distance measure," *WPMC'05*, 2005.