

Experimental Validation of Clustering-based Multipath Component Classification with 28 GHz Indoor Measurements

Hyunsoo Lee, Adam P. Dubs, Nakyoung Lee, and Sunwoo Kim
Department of Electronic Engineering, Hanyang University, Seoul, South Korea
{geniusoo, duanandi16, nageong2379, remero}@hanyang.ac.kr

Abstract—In high-frequency wireless systems, reliable classification of multipath components (MPCs) remains challenging due to environment-dependent propagation. This paper introduces a real-time two-stage clustering method that couples a constraint-based preprocessing step with probabilistic clustering on a reduced feature set. The first stage suppresses noise and prunes features derived from delay, angle, and power-domain observations, yielding compact representations. The second stage conducts lightweight probabilistic assignment that accommodates residual uncertainty without strong prior assumptions. Experimental validation on measurement data indicates that the method further improves stability and suggests that real-time MPC classification can be achieved without sacrificing fidelity.

Index Terms—Millimeter-wave (mmWave), multipath component, real-time clustering, feature reduction

I. INTRODUCTION

In high-frequency wireless systems, exploiting propagation characteristic such as the strong directionality of millimeter-wave (mmWave) and terahertz bands, is a key enabler of future networks [1]–[3]. However, received signal typically comprise a line-of-sight (LoS) and multiple reflections, collectively referred to as multipath components (MPCs). Furthermore, MPCs vary with the receiver (RX) location and the surrounding environment. To address these challenges, both clustering and non-clustering technique have been used to classify MPCs.

Non-clustering approaches, such as the tapped-delay-line channel model, are well established. They represent channels as multiple taps with distinct delays, optionally augmented with angular information, and have been adopted in 3GPP channel models [2], [4], [5]. However, these approaches require extensive measurement campaigns to validate model fit and often need re-parameterization to reduce error in applications such as robotics communications and integrated sensing and communications [6]–[8].

By contrast, clustering technique such as *K-means* and *DBSCAN* operate on feature representations derived from the received signal [9]–[11]. Common strategies include distance-based partitioning that minimizes the Euclidean distance to reference centroids subject to hyperparameter constraints, and probabilistic modeling that fits clusters via iterative inference. However, computational cost increases with feature dimensionality, and model-based clustering may fail when initialization or prior assumptions are inaccurate.



Fig. 1. The experiment layout for 28 GHz indoor measurements.

Prior studies typically select technique by prioritizing computational efficiency or accuracy for tasks such as channel modeling, indoor measurements, or machine learning pipelines. In particular, measurement-driven experiments often validate performance via offline data collection and post-processing, without considering real-time constraints. This paper proposes a MPC-classification method that combines low computational cost with high accuracy and provides experimental validation. The method adopts a two-stage pipeline for real-time operation: a constraint-based preprocessing step that suppresses noise and reduces the feature set, followed by a probabilistic approach operating on the reduced features.

Notations: The operators $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $|\cdot|$ denote the conjugate, transpose, conjugate transpose, and the absolute value.

II. SYSTEM ARCHITECTURE

A. Testbed Overview

The testbed is composed of two chassis from National Instruments (NI) and two radio frequency (RF) devices from TMYTEK [12]–[15]. The chassis for a transmitter (TX) and a RX are comprised of NI units such as a FPGA module, a timing module to synchronize between the TX and RX, and an embedded controller for aforementioned units. Note that the FPGA module of the TX is connected to a digital-to-analog converter (DAC), and the DAC is connected to a local

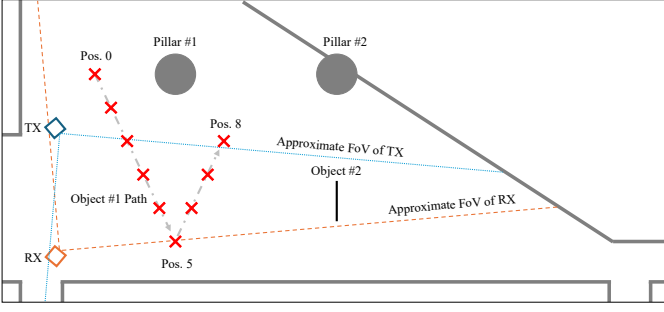


Fig. 2. The schematic of measurement setup.

oscillator (LO) module, respectively. Here, a LO module of the RX is connected to an analog-to-digital converter (ADC), and the ADC is connected to the FPGA module of the RX, respectively. Each RF device for the TX and RX is comprised of a RF converter and an uniform planar array for analog beamforming equipped with 16 antenna elements. Likewise, each RF converter for the TX and RX performs up-converting to 28 GHz band and down-converting to 10.56 GHz band, respectively. The detailed system specification is described in Table I.

B. Received Signal Model

This paper considers a multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) system, where the TX and RX are equipped with N_T and N_R elements of uniform planar array (UPA), respectively. Note that $N_T = N_T^v \times N_T^h$ and $N_R = N_R^v \times N_R^h$ denote the number of antenna elements on the horizontal and the vertical side. The antenna element spacing of the UPA is the half-wavelength $\lambda_c/2$, where λ_c is derived from $c = f_c \lambda_c$, and c is the speed of light.

The received signal for the p -th TX beam at the q -th RX beam can be expressed as

$$y^{p,q}(\mu, \nu) = \mathbf{w}_{R,q}^H \mathbf{H}(\mu, \nu) \mathbf{f}_{T,p}^* x^{p,q}(\mu, \nu) + n^{p,q}(\mu, \nu) \quad (1)$$

where $x^{p,q}(\mu, \nu)$ and $y^{p,q}(\mu, \nu)$ denote the transmitted data and the received data at the μ -th symbol of the ν -th subcarrier, respectively. $n^{p,q}(\mu, \nu)$ denotes the additive Gaussian noise. Moreover, $\mathbf{f}_{T,p} \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{w}_{R,q} \in \mathbb{C}^{N_R \times 1}$ denote the p -th TX precoder and the q -th RX combiner.

TABLE I. TESTBED HARDWARE SPECIFICATION

Parameters	Value
Center frequency	28 GHz
Bandwidth	3.072 GHz
Subcarrier spacing (Δf)	480 kHz (# of subcarrier: 6400)
Antenna size	TX: 16 (4×4); RX: 16 (4×4)
TX beam AoD range	-45° to 45° with 6° interval
RX beam AoA range	-45° to 45° with 2° interval
CP duration (T_{CP})	0.417 μ s
Size of integration	9

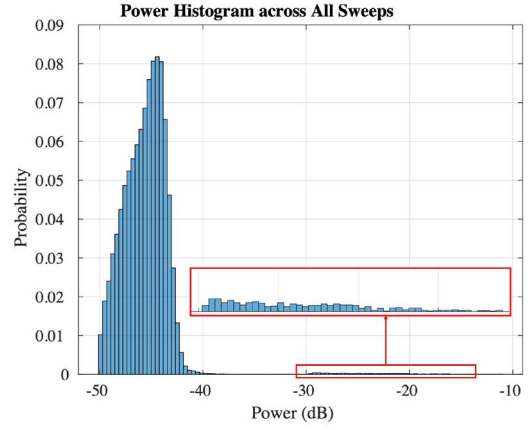


Fig. 3. The power histogram of 28 GHz MIMO OFDM testbed in the indoor measurement. The power measurement results less than -30 dB represent the mixture of sidelobe components, noise components, and the multi-bounce paths.

Note that the channel matrix $\mathbf{H}(\mu, \nu) \in \mathbb{C}^{N_R \times N_T}$ under N_p propagation paths can be expressed as

$$\mathbf{H}(\mu, \nu) = \sum_{n=1}^{N_p} \alpha_n(\mu, \nu) e^{j2\pi(\mu f_\nu^D T_{\text{Tot}} - \tau_n \nu \Delta f)} \times \mathbf{a}_R(\Phi_{R,n}) \mathbf{a}_T(\Phi_{T,n})^T, \quad (2)$$

where $\alpha_n(\mu, \nu)$ denotes the complex attenuation. f_ν^D and $T_{\text{Tot}} = T_{\text{sym}} + T_{\text{CP}}$ are the Doppler frequency and the total symbol duration, where T_{sym} and T_{CP} denote the OFDM base symbol duration and the cyclic prefix duration, respectively. Here, $\mathbf{a}_T \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{a}_R \in \mathbb{C}^{N_R \times 1}$ denote the TX steering vector and the RX steering vector, with the angle-of-departure (AoD) $\Phi_{T,n} = [\phi_{T,n}, \theta_{T,n}]^T$ at the TX and the angle-of-arrival (AoA) $\Phi_{R,n} = [\phi_{R,n}, \theta_{R,n}]^T$ at the RX, where ϕ and θ are the azimuth and the elevation, respectively. Moreover, τ_n and Δf denote the propagation delay of the n -th path and the subcarrier spacing.

III. CLUSTERING TECHNIQUE

A. Feature Construction for Clustering

Sequential beam sweeps are employed: for each transmit beam $p \in 1, \dots, P$, all receive beams $q \in 1, \dots, Q$ are scanned in order (TX $p = 1 \rightarrow$ RX $q = 1, \dots, Q$; then TX $p = 2 \rightarrow$ RX $q = 1, \dots, Q$; etc.).

Let t index a sweep instance after coherent time integration over a fixed window. For each (t, p) , samples $y^{t,p,q}(\mu, \nu)$ are collected over q, μ , and ν . Let $\hat{P}^{t,p,q}$ denote the aggregated power for TX/RX beam pair (p, q) at sweep t , obtained from $y^{t,p,q}(\mu, \nu)$ across (μ, ν) , $\hat{P}^{t,p,q} = \sum_{(\mu, \nu)} |y^{t,p,q}(\mu, \nu)|^2$. Clustering operates on log-power RX-sweep vectors used as features, $\mathbf{z}$, as a feature for clustering, and it can be represented as

$$\mathbf{z}_{t,p} = [\log \hat{P}^{t,p,1}, \dots, \log \hat{P}^{t,p,Q}]^T \in \mathbb{R}^Q. \quad (3)$$

Equivalently, for each t the raw I/Q data can be organized as a tensor $\mathcal{Y}^{(t)} \in \mathbb{C}^{P \times Q \times N_\mu \times N_\nu}$ with axes p, q, μ , and ν .

To enable real-time operation, feature size is reduced via power-thresholding based on the empirical distribution of measured powers, suppressing sidelobes and low-power multiple-bounce paths. In particular, MPCs are difficult to distinguish from noise components arising from heavily attenuated multi-reflection paths and sidelobes. The thresholding step removes such components and improves computational efficiency.

B. Gaussian Mixture Model

Measurement samples collected from experimentation are assumed to follow Gaussian mixture model (GMM). Let $s \in \mathbb{R}^d$ denote the feature vector derived from the log-power RX sweep after preprocessing, $s \equiv \tilde{z}_{t,p}$ with $d = Q$. A K -component GMM assumes that s is drawn from a finite mixture of Gaussian densities:

$$p(s | \Theta) = \sum_{k=1}^K \pi_k \mathcal{N}(s | \mu_k, \Sigma_k), \quad (4)$$

where $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$, $\pi_k \geq 0$, and $\sum_{k=1}^K \pi_k = 1$. Here, $\mu_k \in \mathbb{R}^d$ and $\Sigma_k \in \mathbb{R}^{d \times d}$ denote the mean and covariance of the k -th component. Given the dataset $\mathcal{S} = \{s_n\}_{n=1}^N$, maximum-likelihood (ML) estimation seeks

$$\begin{aligned} \hat{\Theta} &\in \arg \max_{\Theta} \mathcal{L}(\Theta), \\ \mathcal{L}(\Theta) &= \sum_{n=1}^N \log \left(\sum_{k=1}^K \pi_k \mathcal{N}(s_n | \mu_k, \Sigma_k) \right). \end{aligned} \quad (5)$$

Direct maximization is intractable due to the log-sum structure. The expectation-maximization (EM) algorithm addresses this by introducing a latent component indicator $c_n \in \{1, \dots, K\}$ and iterating:

a) *E-step*:

$$\gamma_{nk} \triangleq P(c_n=k | s_n, \Theta) = \frac{\pi_k \mathcal{N}(s_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(s_n | \mu_j, \Sigma_j)}.$$

b) *M-step*:

$$\begin{aligned} N_k &= \sum_{n=1}^N \gamma_{nk}, \quad \mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} s_n, \\ S_k &= \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} (s_n - \mu_k^{\text{new}})(s_n - \mu_k^{\text{new}})^\top, \\ \lambda_k &= \epsilon \cdot \frac{\text{tr}(S_k)}{d}, \\ \Sigma_k^{\text{new}} &= S_k + \lambda_k I, \quad \pi_k^{\text{new}} = \frac{N_k}{N}, \end{aligned}$$

where ϵ is a regularization numerical number to keep the stability for data clustering, $\epsilon \in [10^{-3}, 10^{-2}]$. For real-time operation, Σ_k may be constrained to be diagonal (or shared across k), which reduces per-iteration complexity to $\mathcal{O}(N K d)$.

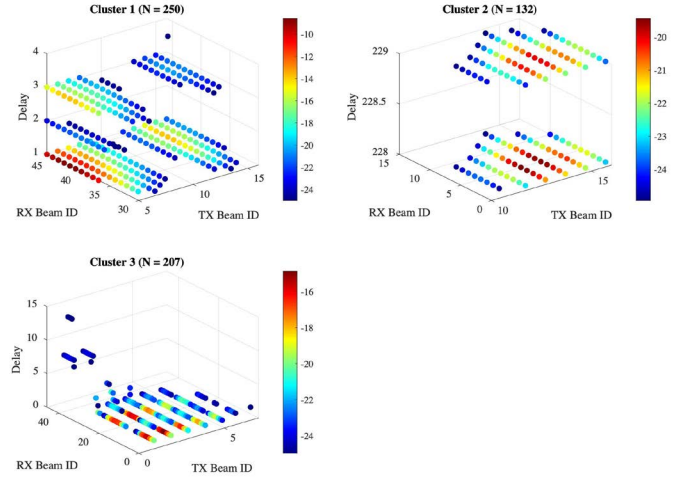


Fig. 4. The result of the proposed clustering technique from single round. Each blue point (•) represents the residual sidelobe components.

IV. EXPERIMENTATION AND RESULTS

A. Experimentation Layout

Experiments are conducted on the MIMO OFDM testbed in the indoor environment. The TX and RX are placed 6.3 m apart and oriented approximately 40° off the LoS axis. As shown in Fig. 2, two objects are placed in the scene: *object #2* placed at position $(-3.3, 13.7)$ [m] and *object #1* placed at distinct 9 positions across rounds. Each round comprises 15 full beam sweeps, at 50 Hz (50 Sweeps per second). Both the TX and RX sweep from -45° to 45° with sequential sweeps: for each TX beam $p \in 1:P$, the RX sweeps $q \in 1:Q$ in order, as mentioned in Section III. The number of TX and RX beams are $P = 16$ and $Q = 46$, with angular resolutions of 6° and 2° , respectively. The total of TX-RX beam pairs is 736.

B. Results and Analysis

Fig. 3 shows the power histogram derived from indoor measurements. Bins above -30 dB reflect a mixture of the LoS path, reflections from *Object 1*, *Object 2*, and residual sidelobe components. Together, these components account for 99.8% of samples above -30 dB.

Fig. 4 presents clustering results from a representative round; a single-sweep result is also depicted. As illustrated in Fig. 1, numerous spurious reflection points can be appeared near the TX and RX due to factors such as the fabric floor, lattice-pattern ceiling, furniture, and pillars. These factors induce strong sidelobe components, shown as blue and cyan points in Fig. 4.

Fig. 5 aggregates results across all rounds; each point represents the mean over sweeps within a round. Here, the LoS path corresponds to the beam pair $p = 4$ and $q = 44$. The proposed clustering technique identifies the LoS path and MPCs while suppressing reflections from the floor, ceiling, and pillars.

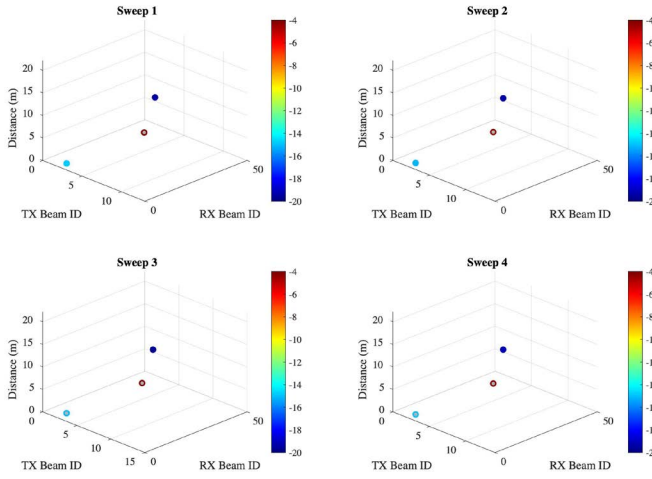


Fig. 5. The data clusters of each sweep. The brown point (*), cyan point (•), and blue point (•) represent the LoS path, and reflected paths from *Object 1* and *Object 2*, respectively. Each point has a mean value of samples in the cluster.

V. CONCLUSION

This paper presents a two-stage data clustering technique with 28 GHz indoor measurements via experimental validation. The proposed data clustering technique first processes the indoor measurement to construct a feature set for data clustering, and reduces the size of the feature set by power thresholding. The experimentation is conducted in an indoor environments where the LoS path component is sparse and high diffuse reflection. The experiment result shows that the proposed two-stage data clustering technique reduces the computational complexity using preprocessed feature set, and classifies the strong MPCs. Further extension is developing the clustering technique to estimate channel parameter for sensing capabilities such as target tracking and radio simultaneous localization and mapping.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00409492).

REFERENCES

- [1] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, 2015.
- [2] O. Kanhere, H. Poddar, and T. S. Rappaport, "Calibration of NYURay for Ray Tracing Using 28, 73, and 142 GHz Channel Measurements Conducted in Indoor, Outdoor, and Factory scenarios," *IEEE Trans. Antennas Propag.*, 2024.
- [3] A. Ghosh, R. Takahashi, and M. Kim, "Comparison of Clustering Techniques Using an Indoor Measurement at 300 GHz," *IEEE Trans. THz Sci. Technol.*, vol. 13, no. 6, pp. 678–687, 2023.
- [4] A. F. Molisch, *Wireless Communications*. John Wiley & Sons, 2012.
- [5] R. He *et al.*, "Clustering Enabled Wireless Channel Modeling Using Big Data Algorithms," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 177–183, 2018.
- [6] M. Kashef, R. Candell, and Y. Liu, "Clustering and Representation of Time-Varying Industrial Wireless Channel Measurements," in *Proc. IEEE 45th Annu. Conf. Ind. Electron. Soc. (IECON)*, 2019, pp. 2823–2829.
- [7] Z. Zhang *et al.*, "A Cluster-Based Statistical Channel Model for Integrated Sensing and Communication Channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11 597–11 611, 2024.
- [8] Y. Qin *et al.*, "Time-Varying Channel Measurement and Analysis at 105 GHz in an Indoor Factory," in *Proc. 18th Eur. Conf. Antennas Propag. (EuCAP)*, 2024, pp. 1–5.
- [9] D. B. Fogel, "An Introduction to Simulated Evolutionary Optimization," *IEEE Trans. Neural Netw.*, vol. 5, no. 1, pp. 3–14, 1994.
- [10] 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. KDD*, vol. 96, no. 34, 1996, pp. 226–231.
- [11] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 169–194, 1998.
- [12] NI, "mmWave Transceiver System," National Instruments, Tech. Rep., Jul. 2018. [Online]. Available: <https://www.ni.com/en-us/shop/wireless-design-test/what-is-mmwave-transceiver-system.html>
- [13] TMYTEK, "UDBox 5G," TMYTEK, Tech. Rep., Nov. 2022, V1.0.8. [Online]. Available: <https://www.tmytek.com/products/frequency-converters/udbox5g>
- [14] TMYTEK, "BBBox 5G One," TMYTEK, Tech. Rep., Oct. 2022, V1.0.8. [Online]. Available: <https://www.tmytek.com/products/beamformers/bbox>
- [15] TMYTEK, "BBBox 5G Lite," TMYTEK, Tech. Rep., Oct. 2022, V1.4.7. [Online]. Available: <https://www.tmytek.com/products/beamformers/bbox>