International Journal of Microwave and Wireless Technologies, page 1 of 8.
doi:0000000000

RESEARCH PAPER

Fast Identification and Clustering of Multi-Path Components for Multi-band Industrial Wireless Channels

MENGFAN WU[1,2], MATE BOBAN[1], FALKO DRESSLER[2]

Multi-path components are both the challenge and the resources to exploit in high-frequency wireless communication, especially in environment with complex reflections. On one hand, late-arriving multi-path components cause inter-symbol interferences in digital communication. On the other hand, techniques such as Multiple-Input, Multiple-Output and Rake receivers have been widely applied to utilize the information carried in the math-path components. To this end, identifying and clustering multi-path components is the foundation in tackling the challenges and boosting the utilization with reliable and correct information. Past research focuses either on extracting the path information, or on clustering the extracted components. In this paper, we propose a complete work flow that performs identification as well as clustering of multi-path components. We extend our previous work in clustering algorithm to indoor propagation measurements of three different frequency bands, as well as multiple transmitter-receiver locations. We verify that the fast attenuation of THz-band signals results in clear separations of peaks in measurements, which in turn facilitates the identification and clustering solutions. The ease of application highlights the wide-applying potential of high-frequency communication.

Keywords: Authors should not add keywords, as these will be chosen during the submission process see: https://www.cambridge.org/core/journals/international-journal-of-microwave-and-wireless-technologies/information/author-instructions/submitting-your-materials for the full list

I. INTRODUCTION

Wireless communication is a trending topic in transforming and modernizing industrial applications, including automation, manufacturing, monitoring, and maintenance. With its ease of deployment, together with mobility and flexibility, wireless communication is preferred over conventional wired system to be deployed in complex and dynamic environments like manufacturing floors with machinery. Nevertheless, industrial settings also pose challenges to wireless communication due to the presence of industrial objects, often with metallic surfaces. The resulting complex propagation phenomena, such as reflection, scattering, and diffraction, can severely affect the quality and reliability of signals. As a result, analyzing and understanding the propagation environment in industrial settings is essential for advancing wireless communication in industrial scenarios.

While the conventional frequency bands (sub-6 GHz) are effective for numerous traditional applications, challenges exist for deploying them in industrial applications. Limited capacity and interference, considering the multitude of devices to be equipped with communication capability, can constrain the performance of wireless communication in terms of throughput and latency. To overcome these challenges, there is a pressing need to explore and exploit new frequency bands, particularly in the terahertz (THz) and millimeter-wave (mmWave) ranges. With wider spectrum, THz and mmWave communication could alleviate congestion in convential bands and could also provide new propagation characteristics to be harnessed. Moreover, the short communication range of mmWave and THz bands is less serious a problem because the typical communication in industrial settings occurs over short distances.

Regardless of frequency, the complex propagation effects in industrial settings yield multi-path effect, where signals arriving at the receiver due to different points of reflection, etc., can lead to both constructive and destructive interference. Therefore, modeling wireless channels that incorporate multi-path components becomes essential for optimizing wireless communication systems in industrial applications. To this end, wireless channel models are often built upon modelling electromagnetic propagation with a finite number of multi-path components (MPCs), with each of the components characterized by its unique amplitude, delay, and angles of arrival and departure. In

[1]Munich Research Center, Huawei Technologies
[2]School of Electrical Engineering and Computer Science, TU Berlin

Corresponding author: Mengfan Wu
Email: mengfan.wu@huawei.com

practice, MPCs with similar characteristics are grouped into clusters, so as to facilitate processing received signals in batches and reduce channel estimation complexity. Moreover, distinguishing clusters of MPCs is beneficial for system like rake receivers and MIMO to select useful and dominant MPCs with less interferences. Clusters with distinct spatial directions can be exploited with beam forming to enhance channel capacity.

In terms of the identification of MPCs, classic approaches include using parameter estimators on measurements, simulating the propagation ray tracing, and manual inspection on the signals. MPCs are often represented as the peak points in the received signal with angular information. For measurements with noise, parameter estimators for reconstructing MPC information could yield unreliable results even with high complexity in implementation and computation. The availability of validation with ray tracing platform is also not guaranteed, which is dependent on the amount of information known about the propagating environment. In the case of an unknown wireless environment, or when the number of MPCs reaches certain multitude, identifying peaks and performing manual clustering of MPCs becomes time-consuming and erroneous.

In light of the challenges mentioned, there is a need for efficient processing of measurements to extract the information of MPC clusters. In this paper, we approach the task from from a data processing and manipulation perspective, without fitting parameter models. Leveraging well-known clustering algorithm, we also investigate the effectiveness of different distance metrics in distinguishing MPCs. Our contributions are summarized as follows:

(i) we propose a complete work flow for processing raw measurements, applying clustering algorithm, and extracting cluster information.
(ii) we compare two different metrics for distinguishing MPCs and evaluate their performance with different clustering algorithm.
(iii) we analyze the clustering results across different frequency bands, demonstrating the strong performance of our solution, particularly with high-frequency signals.

II. RELATED WORK

In [1], Czink et al. introduced KPowerMeans, a modified version of the widely used K-means algorithm [2], which uses the power of multipath components as the scaling factor of the original multipath component distance (MCD). The newly-derived distance metric then substitutes Euclidean distance in the K-means algorithm. This algorithm was also utilized in [3] and [4] to identify multipath clusters at mmWave and THz frequencies. Li et al. [5] collected channel measurements at THz frequencies and employed the DBSCAN algorithm [6] to identify the multipath clusters.

Chen et al. [7] utilized THz channel measurements and ray tracing simulations to cluster and match multipath components. Initially, the DBSCAN algorithm was employed to cluster the multipath components observed in the actual measurements. Subsequently, the identified clusters were matched with those observed in a ray tracing simulator using the multipath component distance (MCD) metric.

He et al. [8] introduced a clustering algorithm that identifies independent clusters by utilizing a kernel density measure. Schneider et al. [9] presented a novel clustering approach based on Fuzzy-c-means. Gentile [10] proposed a new clustering algorithm that utilizes the region competition algorithm [11], an optimization technique originally developed for image segmentation, and the kurtosis measure. In [12], clustering is treated as a sparsity-based optimization problem that takes advantage of the physical property that the power of multipath components decreases exponentially with respect to the delay.

As preprocessing step for raw measurements, for identifying the components in the signals with super-positioned MPCs, various parameter estimators can be applied. Classic solutions include the (MD-MUSIC) algorithm [13], the space-alternating generalized expectation-maximization (SAGE) [14] algorithm, the variational Bayesian version of SAGE [15], and the RiMAX algorithm [16]. Parameter estimators generalize the received signals into parametric representations, which are prone to errors with the presence of noise. Furthermore, the computational demand of such algorithms is usually high, making the scalability an issue when dealing with a large number of MPCs to be extracted.

Considering the challenges of implementing channel parameter estimator, we treat the raw measurements as matrices of data and apply simple preprocessing, by filtering out points with base receiving power, appending angular information, etc., without iterative optimization involved in preprocessing. MPCs and their corresponding clusters are automatically formed by applying clustering algorithm on the data consisting of power information from various angles.

An earlier version of this paper, in which the measurement is collected at only one transmitter(TX)-receiver(RX) location pair and only on THz frequency, was presented at EuCAP 2024 and was published in its proceedings [17]. In this paper, we apply the clustering workflow on measurements from various TX-RX locations and on three different frequency bands. Moreover, we propose a new implementation of the distance metrics in the clustering algorithm, plus a novel post-processing function for extracting MPC cluster information. In the end, we evaluate the clustering results not only by time range separation, but also by peak separation when more than one MPCs arrive at the same time.

III. MEASUREMENT SETUP

To study the feasibility of high-frequency wireless communication in industrial settings, a measurement campaign

Table 1. Caption

Parameter	Sub-6 GHz	mmWave	THz
Frequency (GHz)	6.75	74.25	305.27
Calibrated Bandwidth (GHz)		5	
Azi. Scanning Angles ($\alpha_{TA,RA}$)		$\{-180°, -165°, \ldots, 165°\}$	
Ele. Scanning Angle ($\alpha_{TE,RE}$)		$0°$	

Table 2. Variables, parameters and acronyms used in algorithms and experiments

T	Measured time steps
P	Number of polarization pairs of transmitter and receiver
I, J M, N	Number of the scanning angles of transmitter azimuth/elevation and receiver azimuth/elevation
$\mathbf{x}$	Identified cluster of MPCs
H	Measurement data
α_*	$* \in \{TA, TE, RA, RE\}$, the corresponding angles of signals in ray-tracing experiments

[18] was conducted in a small factory scenario. The measurement environment is a realistic replica of manufacturing sites, and therefore consists of strong reflective surfaces of metal and glass, creating a complex propagating environment for wireless communication with non-negligible MPCs.

To mimic the potential wireless communication equipped on various devices in the room, and also to study the propagating environment on line-of-sight (LOS) and non-line-of-sight (NLOS) conditions, 3 transmitter locations and 12 receiver locations are selected and 20 measurements on 20 pairs of transmitter-receiver settings are conducted. The directions of transmitting and receiving are fixed on the horizontal plane, and rotate with a 15 degree separation, aiming to capture the full range of propagation scenarios, including line-of-sight and non-line-of-sight cases. The measurements cover the terahertz frequency (300 GHz), mmWave frequency (74.25 GHz), as well as Sub-6 GHz frequency (6.75 GHz), providing a comprehensive understanding of the channel characteristics at different bands. For each frequency band, the measurement is conducted with two pairs of antenna polarization directions (horizontal-horizontal and vertical-vertical). Detailed of measurement configuration are documented in Table 1.

For each transmitter and receiver location, we obtained measurement data of the following format:

$$H \subset \mathbb{R}_{\geq 0}^{K \times P \times I \times J \times M \times N}$$

Details of the variables related to dimension of the data are listed in Table 2. In our case, the elevation scanning angle is $0°$ only, therefore $J = 1$ and $N = 1$.

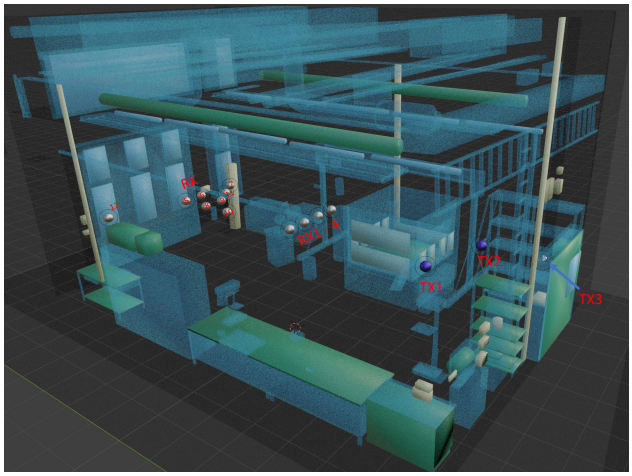


Fig. 1.: Transmitter and receiver locations, marked by blue and silver spheres respectively. Measurement is conducted in a room simulating manufacturing environment. Materials are plotted transparent for visibility of the transmitter/receiver locations.

IV. CLUSTERING ALGORITHM

To perform direct identification and clustering of the measurement signal, we need to convert the signal matrix into data points that fit the corresponding clustering algorithm. With the clustered output, another procedure of post-processing the results is needed to extract characteristic information of the MPC clusters. We divide this section into three parts that describe the three procedures.

A) Pre-processing of Measurement Signals

We describe the pre-processing of signal matrix in Algorithm 1, separated in terms of two different metrics of evaluating the distance between signals.

For classic Euclidean distance, the power of the received signal at the specific time step, as well as the power difference between the current time step and the previous/following steps, are included in the feature. The angular difference between signals are represented by the sinusoidal values (cosine and sine) of the transmitter and receiver angles, which aims to differentiate symmetrical angles and keep the physical continuity of angles at numerical jumps (e.g. $359°$ to $0°$). The complete process of getting the sinusoidal features is described in Algorithm 2. The number of sinusoidal features are reduced by using only cosine/sine values when all of the angular values are in the monotonic range of cosine/sine functions.

For power distance [1], the angle information (departure/arrival) of a signal is fed to compute the multipath component distance, as later shown in Equation (4).

The conditions to exclude points with $p_* = 0, \forall * \in \{f, b, t\}$ in Algorithm 1 are to filter out points in the measurement where there is no signal received, thus to reduce the size of data later fed into clustering algorithm and speed up the process.

Algorithm 1 Preprocessing on Measurement Matrix

Input: H, `pid`, $\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}$ Measurement data of the required format, selected polarization index, lists of transmitter's azimuth and elevation angles and receiver's azimuth and elevation angles

Output: l

Preprocessing before using Euclidean distance

$l \leftarrow \varnothing$
for $\mathtt{i}, \mathtt{j}, \mathtt{m}, \mathtt{n} \in [I] \times [J] \times [M] \times [N]$ **do**
  $h \leftarrow H[:][\mathtt{pid}][\mathtt{i}][\mathtt{j}][\mathtt{m}][\mathtt{n}]$
  for $t \in K$ **do**
    $p_t \leftarrow\ h[t]$
    $p_f \leftarrow\ p_t - h[t-1], p_b \leftarrow\ p_t - h[t+1]$
    $\boldsymbol{\theta} \leftarrow \mathtt{GetAngles}(\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}, i, j, m, n)$
    if not ($p_f = 0$ and $p_b = 0$ and $p_t = 0$) **then**
      $l \leftarrow l \cup \{(t, \boldsymbol{\theta}, p_f, p_b, p_t)\}$
    end if
  end for
end for

Preprocessing before using power distance

$l \leftarrow \varnothing$
for $\mathtt{i}, \mathtt{j}, \mathtt{m}, \mathtt{n} \in [I] \times [J] \times [M] \times [N]$ **do**
  $h \leftarrow H[:][\mathtt{pid}][\mathtt{i}][\mathtt{j}][\mathtt{m}][\mathtt{n}]$
  for $t \in K$ **do**
    $p_t \leftarrow\ h[t]$
    $p_f \leftarrow\ p_t - h[t-1], p_b \leftarrow\ p_t - h[t+1]$
    if not ($p_f = 0$ and $p_b = 0$ and $p_t = 0$) **then**
      $l \leftarrow l \cup \{(p_t, t, \mathbf{TA}[\mathtt{i}], \mathbf{TE}[\mathtt{j}], \mathbf{RA}[\mathtt{m}], \mathbf{RE}[\mathtt{n}])\}$
    end if
  end for
end for

B) Clustering Algorithm

We compare two clustering algorithms: K-means [2] and DBSCAN [6]. Both algorithms have been elaborated in the past; therefore, we refer to the corresponding papers for the details of the algorithm, whereas below we elaborate on our implementation aspects for both algorithms. Terminology of related terms are detailed in Table 3.

In the original versions of the two algorithms, n-dimensional points are input and the Euclidean distance based on all dimensions is used as the metric to determine the proximity of two points:

$$d_{euc}(m,n) = \left(\sum_{i=1}^{n} (m_i - n_i)^2 \right)^{1/2} \tag{1}$$

In [1], power distance is introduced to evaluate the proximity of a MPC component to its cluster center:

$$d_{pow}(m,n) = P_m \cdot \mathrm{MCD}_{mn}, \tag{2}$$

Algorithm 2 `GetAngles`

Input: $\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}, i, j, m, n$ lists of transmitter's azimuth and elevation angles and receiver's azimuth and elevation angles, index of the corresponding sampled angle

Initialization

for $\mathbf{L} \in \{\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}\}$ **do**
  Shift all values in $\mathbf{L}$ to be in $[0, 2\pi]$
  if $x \in [0, \pi], \forall x \in \mathbf{L}$ or $x \in [\pi, 2\pi], \forall x \in \mathbf{L}$ **then**
    `cos_only` $\leftarrow$ `True`
  end if
  Shift all values in $\mathbf{L}$ to be in $[-\pi/2, 3\pi/2]$
  if $x \in [-\pi/2, \pi/2], \forall x \in \mathbf{L}$ or $x \in [\pi/2, 3\pi/2], \forall x \in \mathbf{L}$ **then**
    `sin_only` $\leftarrow$ `True`
  end if
end for

Per-point processing

for $\mathbf{L} \in \{\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}\}$ **do**
  $\boldsymbol{\theta} \leftarrow \varnothing$
  $k \leftarrow$ the corresponding index for the list
  if `cos_only` **then**
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\cos \mathbf{L}[k]\}$
  else if `sin_only` **then**
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\sin \mathbf{L}[k]\}$
  else
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\cos \mathbf{L}[k], \sin \mathbf{L}[k]\}$
  end if
end for
return $\boldsymbol{\theta}$

Table 3. Notations in equations

$d(m,n)$	Distance between point m and n
d_{euc}	Euclidean distance
d_{pow}	Originally defined power distance
d'_{pow}	Newly defined power distance
P_m	Power of point m
MCD_{mn}	Multipath component distance between m and n
MCD_{AoA}	Angle of arrival component of MCD
MCD_{AoD}	Angle of departure component of MCD
MCD_{τ}	Time difference component of MCD

n is the cluster centroid in the K-means algorithm and m is the input points to be clustered.

$$\mathrm{MCD}_{mn} = \sqrt{\|\mathrm{MCD}_{AoA,mn}\|^2 + \|\mathrm{MCD}_{AoD,mn}\|^2 + \|\mathrm{MCD}_{\tau,mn}\|^2} \tag{3}$$

For components in computing MCD_{mn}:

$$\|\mathrm{MCD}_{AoA/AoD,mn}\| = \frac{1}{2}\left|\begin{pmatrix}\sin(\theta_m)\cos(\phi_m)\\ \sin(\theta_m)\sin(\phi_m)\\ \cos(\theta_m)\end{pmatrix} - \begin{pmatrix}\sin(\theta_n)\cos(\phi_n)\\ \sin(\theta_n)\sin(\phi_n)\\ \cos(\theta_n)\end{pmatrix}\right| \tag{4}$$

$$\mathrm{MCD}_{\tau,mn} = \zeta \cdot \frac{|\tau_m - \tau_n|}{\Delta\tau_{\max}} \cdot \frac{\tau_{std}}{\Delta\tau_{\max}} \tag{5}$$

DBSCAN [6] does not define a centroid for each cluster, but finds neighbors for each input point. Therefore, when calculating the distance, the power of both signal points should be considered, rather than just one as in Equation (2) which results in $d(m,n) \neq d(n,m)$. To this end, we derive:

$$d'_{pow}(m,n) = (|P_m - P_n| + P_{base}) \cdot \mathrm{MCD}_{mn} \tag{6}$$

for computing the power distance when applying DBSCAN algorithm. The computation of absolute difference ensures symmetry when swapping the input sequence of the two points. The addition of P_{base} serves as the base scalar when two points have similar power level but differ in angles.

For algorithms using both distance metrics, it is necessary to adapt clustering parameters and further scale the inputs before feeding data points. The process flow is shown in Algorithm 3 and the scaling factor for each of the feature is shown in Table 4. Line 5-8 in Algorithm 3 shift the power values of all points when using decibel values to be ≥ 0, thus guaranteeing that computed power distance by Equation (2) to be ≥ 0.

Algorithm 3 `Clustering`

1: get l from measurement preprocessing
2: **if** Euclidean distance **then**
3: $l \leftarrow$ `Scaling`(l)
4: **else if** Power distance **then**
5: **if** Use decibel values for power **then**
6: $P_{dB} \leftarrow \{10 \cdot \log v[0], \forall v \in l\}$
7: $p_{\min} \leftarrow \min P_{dB}$
8: $v[0] \leftarrow \log v[0] - p_{\min}, \forall v \in l$
9: **end if**
10: choose time scaling factor ζ
11: **end if**
12: choose centroid model (k-means) or density model (DBSCAN)
13: perform clustering, get $c_x \forall x \in l$

C) Post-processing

We propose Algorithm 4 to process the clustered points or the original matrix of the raw signal. For simplicity, the dimension of the matrix processed is $T \times I \times M$, which suits our measurement data with only one measuring angle of transmitting and receiving elevations. The algorithm can be further extended to be compatible with data of higher dimensions where the elevation angles are also included.

Algorithm 4 tracks the peak (with power greater than its neighboring angles) of the arriving signals in each time frame and consistently log the location of peaks, with considerations that the peak might shift in angular domains. It is worth noting that Algorithm 4 can also be used as a pre-processing step to extract the peaks in the raw signals. The advantages of using Algorithm 4 on clustered results lie in the shorter processing time, where it is easier to find local maxima in a frame (in Line 6) when processing sparse matrix yielded by clustered data.

The information of peaks identified by Algorithm 4 in each cluster, e.g. angular values and duration, is then valuable for further microwave applications.

Algorithm 4 Track Local Maxima Over Time

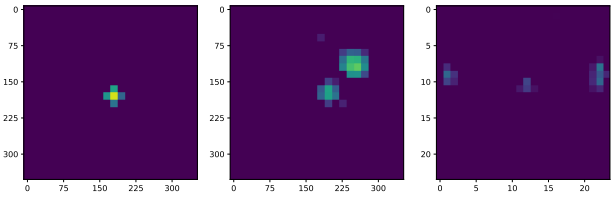
1: **Input:** data points $(p_t, t, \mathtt{TxAz}_t, \mathtt{RxAz}_t)$ in the same cluster
2: Get `Frames` $(T \times I \times M)$ from signals of the same cluster
3: $\mathbf{P}_{\text{found}} \leftarrow \varnothing$, $\mathbf{P}_{\text{live}} \leftarrow \varnothing$
4: **for** `Frm` $\in$ `Frames` **do**
5: get time stamp `t` and local `maxima` in `Frm`
6: $\mathbf{P}_{\text{new}} \leftarrow \varnothing$
7: **for** each detected peak `(i, j)` in `maxima` **do**
8: `matched` $\leftarrow$ `False`
9: **for** $p \in \mathbf{P}_{\text{live}}$ **do**
10: **if** `(i, j)` is within 1-pixel shift of p **then**
11: Update p with new `(i, j)`
12: Add updated p to $\mathbf{P}_{\text{new}}$
13: `matched` $\leftarrow$ `True`
14: **break**
15: **end if**
16: **end for**
17: **if** not `matched` **then**
18: Create p_{new} with $p.\mathtt{t}_{\text{start}} = \mathtt{t}$ and information `(i, j)`
19: $\mathbf{P}_{\text{new}} \leftarrow \mathbf{P}_{\text{new}} \cup p_{\text{new}}$, $\mathbf{P}_{\text{live}} \leftarrow \mathbf{P}_{\text{live}} \cup p_{\text{new}}$
20: **end if**
21: **end for**
22: **for** $p \in \mathbf{P}_{\text{live}}$ **do**
23: **if** p is not in $\mathbf{P}_{\text{new}}$ **then**
24: $p.\mathtt{t}_{\text{end}} \leftarrow \mathtt{t}$, $\mathbf{P}_{\text{found}} \leftarrow \mathbf{P}_{\text{found}} \cup p$
25: **end if**
26: **end for**
27: $\mathbf{P}_{\text{new}} \leftarrow \mathbf{P}_{\text{live}}$
28: **end for**
29: Get the ending time step of signal cluster: $\mathtt{t}_{\text{end}}$
30: **for** $p \in \mathbf{P}_{\text{live}}$ **do**
31: $p.\mathtt{t}_{\text{end}} \leftarrow \mathtt{t}_{\text{end}}$, $\mathbf{P}_{\text{found}} \leftarrow \mathbf{P}_{\text{found}} \cup p$
32: **end for**
33: **Return** $\mathbf{P}_{\text{found}}$

Table 4. Scaling Factors in Clustering with Euclidian Distance

Name	Value
p_t, p_f, p_b in dB	0.2
t	$7.788 \cdot 10^8$
$\boldsymbol{\theta}$	8.0

Table 5. Searching Range of Hyper-Parameter in Clustering

DBSCAN	e for Eucl. dist.	{2.0, 2.5, 3.0}
	e for power dist.	{0.001, 0.003, 0.006}
	P_{base}	{ 0.005, 0.01, 0.02}
	ζ	{0.1, 0.2, 0.4, 0.8}
K(Power)-Means	k	{3, 4, . . . , 10}



(a) Single MPC (b) Two MPCs in same cluster (c) Two MPCs in separate clusters

Fig. 2.: Heatmaps of multi-path components arriving in three different ways

V. EXPERIMENTS

A) Parameter Settings

The scaling factors for each feature used to compute Euclidean distance and power distance are listed in Table 4. For algorithm of K(Power)-Means, the number of clusters k is the hyper-parameter. For DBSCAN, the neighbor-threshold distance e is the parameter to determine if a cluster should expand to include another point. The searching range of k and e are listed in Table 5.

B) Ground Truth

For any time step in the measurements with signals received from any angle, we inspect the heatmap and evaluate if the signal belongs to a certain cluster.

For the transmitter-receiver location pairs where there is always only one MPC arriving at a time, as shown in Figure 2a, or more than one MPCs close to each other as shown in Figure 2b, we manually identify the time range of the corresponding MPC cluster and then cluster the MPCs based on their time-wise and angular proximity. If there are more than one MPCs arriving at a time, which are also distant in arriving angle and thus do not belong to the same cluster (shown in Figure 2c), we inspect the clustered results and check if the MPCs are separated correctly.

C) Accuracy of Time Range Separation

We only evaluate accuracy of time range separation for measurements where there are no MPCs belonging to separate clusters arriving at the same time. With algorithm Algorithm 5, each time step is assigned with a cluster number that is the majority of clusters for points in the corresponding frame.

Algorithm 5 Identifying MPCs' Time Range

for $t \in K$ **do**
  $\mathbf{x}_t \leftarrow \{x'\}, \forall x'$ with $x_{\texttt{time}} = t$
  if $|\mathbf{x}_t \geq 1|$ **then**
    $c^t \leftarrow \texttt{mode}(\{c_x, x \in \mathbf{x}_t\})$
  else
    $c^t \leftarrow -1$
  end if
end for
return $\mathbf{c} = \langle c^t \rangle, \forall t \in K$

Ground truth labelling is performed for selected pairs of Tx-Rx locations and on all of the three frequency bands. Cases of TX1-RX1 and TX1-RX8 are in LOS condition. Cases of TX2-RX3 and TX2-RX6 are in NLOS conditions. The accuracy is then computed with Adjusted Rand Index (ARI) [19] and shown in Table 6. Higher values (close to 1) in ARI indicates better match between the experiment result and the ground truth. We see that using DBSCAN with power distance outperforms the other three methods in 7 out of 12 cases with the highest average accuracy. Specifically, DBSCAN with power distance performs the best in 3 out of 4 cases on THz frequency, and in 3 out of 4 cases in mmWave frequency as well.

We also observe that there is a general trend that the clustering accuracy is higher, for any method, when the frequencies of the signals are higher. The underlying reason is that faster attenuation of high-frequency signals results in clear separation of arriving MPCs, thus formulating a simpler task for the clustering algorithm.

D) Accuracy of Simultaneous Peak Separation

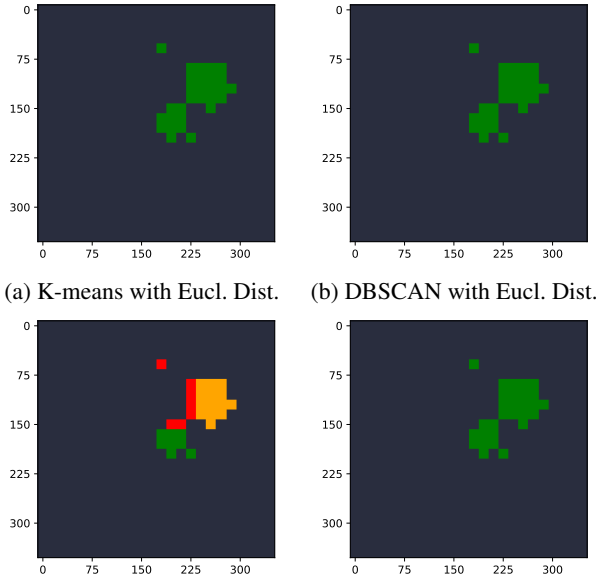
We plot the clustering results of the four methods for the MPC components shown in Figure 2b and Figure 2c.

In Figure 3, where the two arriving MPCs are also close on the angular domain and thus belong to the same cluster, all clustering methods achieve the correct clustering results, except for K-means with power distance. Even though it is also practically viable to separate the MPCs, K-means with power distance also fail to correctly cluster the signals detected at the edge angles of MPCs (as seen in pixels marked in red in Figure 3).

In the case where two arriving MPCs are far away from each other in angular domain (as in Figure 4) and fall into different clusters, none of the four methods gives perfect solutions. We notice that DBSCAN algorithm (in Figure 4b

Table 6. Evaluation Accuracy for Selected Tx-Rx Locations

			k-means		DBSCAN	
TX	RX	Freq.	eucl.	p-dist.	eucl.	p-dist.
1	1	THz	0.787	0.835	0.951	0.951
		mmW.	0.928	0.727	0.797	0.967
		Sub-6G	0.694	0.395	0.755	0.650
1	8	THz	0.712	0.978	0.862	0.780
		mmW.	0.657	0.769	0.832	0.995
		Sub-6G	0.514	0.453	0.494	0.801
2	3	THz	0.990	0.985	1.000	1.000
		mmW.	0.733	0.656	0.686	0.921
		Sub-6G	0.740	0.352	0.650	0.638
2	6	THz	0.806	0.750	0.998	0.999
		mmW.	1.000	0.651	0.899	0.899
		Sub-6G	0.775	0.425	0.353	0.651
Average			0.778	0.665	0.773	0.854

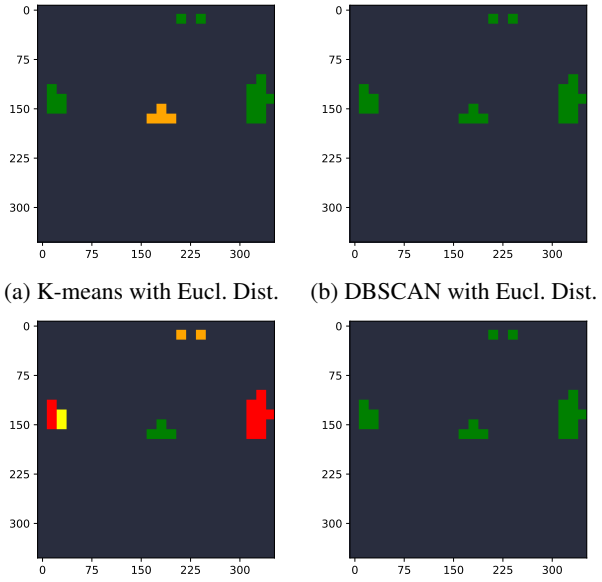


(a) K-means with Eucl. Dist. (b) DBSCAN with Eucl. Dist.

(c) K-means with Power Dist. (d) DBSCAN with Power Dist.

Fig. 3.: Clustering results for two MPCs in the same cluster arriving at the same time

and Figure 4d) determines all arriving signals at the same time step to be in the same cluster, irrespective of the distance metrics, while k-means algorithm separates the signals that are almost 180° apart in transmission azimuth. However, k-means with Euclidean distance (in Figure 4a) does not separate the two noise signals at 15° receiving azimuth, while k-means with power distance (in Figure 4c) fails to determine the signals of transmitter azimuth 30° (shown in yellow color) to be in the same cluster as other neighboring signals marked in red.



(a) K-means with Eucl. Dist. (b) DBSCAN with Eucl. Dist.

(c) K-means with Power Dist. (d) DBSCAN with Power Dist.

Fig. 4.: Clustering results for two MPCs in the same cluster arriving at the same time

VI. CONCLUSIONS

In this paper, we applied two classic clustering algorithms to the multi-band industrial wireless measurements. Different from previous solutions where multi-path component is firstly identified from measurements, we propose a new process flow of pre-processing measurements, performing clustering, and post-processing the clustered results to facilitate extracting MPC cluster information. Inspired by the prior literature, we experimented with the metric of power distance together with traditional Euclidean distance in the two clustering algorithms. By evaluating the clustering results from time range identification as well as peak separation, we conclude that DBSCAN algorithms are performing well in identifying MPC clusters' delay when there is only one MPC arriving at a time. For multiple MPCs arriving at the same time, k-means performs better when separating the MPCs. We observe that our algorithm is especially suitable for clustering high-frequency signals, e.g. in THz range, thanks to the clear separation of arriving MPCs due to stronger attenuation. Future work includes automated parameter tuning of the clustering algorithms, as well as detailed comparison of the clustering results in LOS and NLOS scenarios.

ACKNOWLEDGMENT

This work has been performed in part in the framework of the HORIZON JU-SNS-2022 project TIMES, cofunded by the European Union.

COMPETING INTERESTS

The author(s) declare none.

REFERENCES

[1] 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 *64th IEEE Vehicular Technology Conference (VTC 2006)*, Montréal, Canada: IEEE, Sep. 2006, pp. 1–5.

[2] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[3] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-Wave Multipath Clustering and Channel Modeling," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1445–1455, Mar. 2014.

[4] C.-L. Cheng, S. Sangodoyin, and A. Zajic, "THz Cluster-Based Modeling and Propagation Characterization in a Data Center Environment," *IEEE Access*, vol. 8, pp. 56 544–56 558, Mar. 2020.

[5] Y. Li, Y. Wang, Y. Chen, Z. Yu, and C. Han, "Channel Measurement and Analysis in an Indoor Corridor Scenario at 300 GHz," in *IEEE International Conference on Communications (ICC 2022)*, Seoul, South Korea: IEEE, May 2022.

[6] E. Martin, K. Hans-Peter, and X. Xiaowei, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD 1996)*, Portland, OR, Aug. 1996, pp. 226–231.

[7] Y. Chen, Y. Li, C. Han, Z. Yu, and G. Wang, "Channel Measurement and Ray-Tracing-Statistical Hybrid Modeling for Low-Terahertz Indoor Communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8163–8176, Dec. 2021.

[8] R. He, Q. Li, B. Ai, et al., "A Kernel-Power-Density-Based Algorithm for Channel Multipath Components Clustering," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7138–7151, Nov. 2017.

[9] C. Schneider, M. Bauer, M. Narandzic, W. A. T. Kotterman, and R. S. Thomae, "Clustering of MIMO Channel Parameters - Performance Comparison," in *69th IEEE Vehicular Technology Conference (VTC 2009-Spring)*, Barcelona, Spain: IEEE, Apr. 2009.

[10] 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, Jun. 2013.

[11] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.

[12] 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, Jun. 2016.

[13] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758–1774, 1992.

[14] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[15] D. Shutin and B. H. Fleury, "Sparse Variational Bayesian SAGE Algorithm With Application to the Estimation of Multipath Wireless Channels," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3609–3623, Aug. 2011.

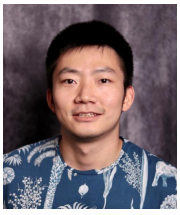
[16] R. S. Thomae, M. Landmann, G. Sommerkorn, and A. Richter, "Multidimensional high-resolution channel sounding in mobile radio," in *21st IEEE Instrumentation and Measurement Technology Conference*, vol. 1, Como, Italy: IEEE, May 2004, pp. 257–262.

[17] M. Wu, T. Zugno, M. Boban, and F. Dressler, "Direct Clustering and Multi-Path Component Identification on THz Channel Measurements in a Factory Environment," in *18th European Conference on Antennas and Propagation (EuCAP 2024)*, Glasgow, United Kingdom: IEEE, Mar. 2024, pp. 1–5.

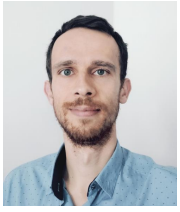
[18] D. Dupleich, A. Ebert, Y. Völker-Schöneberg, et al., "Characterization of Propagation in an Industrial Scenario from Sub-6 GHz to 300 GHz," in *2023 IEEE Global Communications Conference (GLOBECOM 2023)*, Kuala Lumpur, Malaysia: IEEE, Dec. 2023, pp. 1475–1480.

[19] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, Dec. 1985.

Bibliographies (no more than 150 words each, accompanied by a 75 x 88 pixel photo)



Mengfan Wu received a bachelor degree in telecommunications from Tongji University, Shanghai, in 2018 and received his M.Sc. degree in Electrical Engineering and Information Technology from ETH Zurich in 2021. He is a doctoral student at Huawei Munich Research Center and TU Berlin. His main research interests are federated learning algorithms and machine learning solutions for high-frequency telecommunications.



Mate Boban received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University. He is a technical expert and team leader with Huawei Munich Research Center, Germany. Currently, he is a co-chair of COST INTERACT WG1 (Radio Channels) and a vice-chair of ETSI Industry Specification Group (ISG) on THz. He has co-chaired several IEEE conferences and workshops, was an Associate Editor of IEEE TMC, and has been actively involved in EU-funded 5G and 6G projects. His current research interests are in channel modeling and machine learning for radio access networks.

Falko Dressler (Fellow, IEEE) is full professor and Chair for Data Communications and Networking at the School of Electrical Engineering and Computer Science, TU Berlin. He received his M.Sc. and Ph.D. degrees from the Dept. of Computer Science, University of Erlangen in 1998 and 2003, respectively. Dr. Dressler has been associate editor-in-chief for IEEE Trans. on Mobile Computing and Elsevier Computer Communications as well as an editor for journals such as IEEE/ACM Trans. on Networking, IEEE Trans. on Network Science and Engineering, Elsevier Ad Hoc Networks, and Elsevier Nano Communication Networks. He has been chairing conferences such as IEEE INFOCOM, ACM MobiSys, ACM MobiHoc, IEEE VNC, IEEE GLOBECOM. He authored the textbooks Self Organization in Sensor and Actor Networks published by Wiley & Sons and Vehicular Networking published by Cambridge University Press. He has been an IEEE Distinguished Lecturer as well as an ACM Distinguished Speaker. Dr. Dressler is an IEEE Fellow as well as an ACM Distinguished Member. He is a member of the German National Academy of Science and Engineering (acatech). He has been serving on the IEEE COMSOC Conference Council and the ACM SIGMOBILE Executive Committee. His research objectives include adaptive wireless networking (sub-6GHz, mmWave, visible light, molecular communication) and wireless-based sensing with applications in ad hoc and sensor networks, the Internet of Things, and Cyber-Physical Systems.