Direct Clustering and Multi-Path Component Identification on THz Channel Measurements in a Factory Environment

Mengfan Wu*†, Tommaso Zugno*, Mate Boban*, Falko Dressler†
*Munich Research Center, Huawei Technologies
†School of Electrical Engineering and Computer Science, TU Berlin
Email: {mengfan.wu, tommaso.zugno, mate.boban}@huawei.com, dressler@ccs-labs.org

***Abstract*—Multi-path components pose both a challenge and an opportunity in high-frequency wireless communication, especially in environments with complex propagation conditions. In this paper, we derive a clustering algorithm to be applied directly to the measurements of indoor THz propagation. We show that such method does not require preprocessing to identify the peaks of multi-path components, but rather extract the time range of clustered multipath components in measurements. Ray-tracing experiments are performed together with classic clustering methods to validate our solution on the corresponding measurements. Our solution facilitates the identification of both clusters and multi-path components directly on measurements without the need to reconstruct scenario in ray-tracing to identify the sources of specular components.**

***Index Terms*—Clustering techniques, THz, channel modelling**

I. Introduction and Related Work

Many channel models assume that wireless propagation happens through a finite number of multipath component (MPC)s, each representing a plane wave travelling along a different path. Each MPC is characterized by its complex amplitude, delay, direction of arrival and departure. For the purpose of channel modeling, MPCs that exhibit similar characteristics are grouped into a cluster. In this context, different machine learning techniques have been exploited to identify clusters.

In [1], authors presented KPowerMeans, a variant of the popular K-means algorithm which uses power-distance [2] to compute clusters centroids. The same algorithm was used in [3] and [4] for the identification of multipath clusters at mmWave and THz frequencies. Li et al. [5] collected channel measurements at THz frequencies and applied the DBSCAN algorithm [6] to identify the multipath clusters. Chen et al. [7] exploited THz channel measurements and ray tracing simulations for clustering and matching of MPCs. First, the MPCs observed in real measurements are clustered using the DBSCAN algorithm. Then, the identified clusters are matched with those observed in a ray tracing simulator based on the MPC distance (MCD) metric. In [8], authors proposed a clustering algorithm that identifies independent clusters based on kernel power density. In [9], a novel clustering approach based on Fuzzy-c-means is presented. In [10], authors proposed a new clustering algorithm based on the Kurtosis Measure and the region competition algorithm [11], an optimization technique originally developed for image segmentation. In [12], clustering is treated as a sparsity-based optimization problem which exploits the physical property of MPCs, whose power decreases exponentially with respect to delay.

In this paper, we present a novel data generating process to convert THz measurement data to data points to be processed by widely-used clustering algorithms (K-means, KPowerMeans, DBSCAN). The time ranges of similar MPCs are obtained through postprocessing of clustering results. We validate our solution by comparing the outcome with clustering results of the MPCs identified by ray-tracing platform, where the propagation environment is reconstructed. Our solution facilitates the utilization of MPCs in propagation environments where accurate ray-tracing is not available.

This work has been performed in part in the framework of the HORIZON-JU-SNS-2022 project TIMES, cofunded by the European Union. The views expressed are those of the authors and do not necessarily represent the project.

II. Clustering Method

A. Background

The measurement is performed in an industrial scenario where the transmitting antenna pointing towards a receiving antenna placed inside of a machine. For each transmitter and receiver location, we obtained measurement data of the following format:

$$H \subset \mathcal{C}^{K\times P\times I\times J\times M\times N}$$

Denotation of the dimensions are explained in Table I.

The scanning angles of transmission and receiving antenna, together with frequency band and bandwidth is shown in Table Table II. Note that the direction of the receiving antenna is turned on the horizontal plane to receive reflected signals from all directions of the machine.

For both measurement data and simulation in ray-tracing, we limit the tracked time range to 33.7 nanoseconds so that we focus on rays with propagation distance shorter than 10 meters.

B. Algorithm of Direct Clustering

In Algorithm 1, we show the preprocessing flow and clustering steps of identifying time segments of each MPC based on measurement data. For identifying with Euclidean distance, the

TABLE I
VARIABLES, PARAMETERS AND ACRONYMS USED IN ALGORITHMS AND EXPERIMENTS

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

TABLE II
MEASUREMENT PARAMETERS

Parameter	Values
Frequency	300.7 GHz - 306.9 GHz
Bandwidth	6.2 GHz
Tx Azimuth $\mathbf{TA}$	$\{-15°, 0°, 15°\}$
Tx Elevation $\mathbf{TE}$	$\{-30°, -15°, 0°\}$
Rx Azimuth $\mathbf{RA}$	$\{-180°, -165°, \ldots, 165°\}$
Rx Elevation $\mathbf{RE}$	$\{-45° \ -30°, \ldots, 45°\}$

angular information of obtained signals are obtained from Algorithm 2, which tries to limit the number of angular features while guaranteeing the continuity of angles. Specifically, when sampled angles in $\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}$ are distinguishable by either the sine or cosine value, the corresponding trigonometric values are added to the angular features. Otherwise, both of the trigonometric values are added. This is to avoid confusion of angles that are symmetric along x(/y) axis thus having same sine(/cosine) values but are distinguishable by cosine(/sine) values. Furthermore, the physical proximity of 359° to 0° are also guaranteed.

Apart from Euclidean distance, power-distance, as firstly introduced in [2], is also widely-applied distance when measuring the difference of MPCs. When applied in clustering algorithm, the characteristics of scaling the angular distance based on the power of the point implies that the clustering center should be close to the point with the greatest power, Note that the power in the power-distance can be defined as proportional to the square of the amplitude of the signal, or the shifted decibel values where all of the power is guaranteed to be non-negative.

For the clustering methodology, we compare the representative method of centroid models (K-means [13]) and density models (DBSCAN [6]).

C. Clustering with Ray-Tracing

The same environment is reconstructed by performing ray-tracing in Sionna [14]. Tracked rays limited within the range of the scanning angles in Table II are shown in Figure 1. Given the highly accurate 3D information of the environment available, we can obtain accurate angle information from ray-tracing. Therefore, for Euclidean methods, the points to be clustered are configured as $(p, t, \boldsymbol{\theta})$. For power-distance, the points to be clustered are configured as $(p, t, \alpha_{TA}, \alpha_{TE}, \alpha_{RA}, \alpha_{RE})$.

Algorithm 1 Cluster On Measurements

Input: $\mathcal{C}^{K\times P\times I\times J\times M\times N}$, $\texttt{pid}$, $\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}$ Measurement data of the required format, selected polarization index, lists of transmitter's azimuth and elevation angles and receiver's azimuth and elevation angles

Preprocessing

for Euclidean distance
$l \leftarrow \emptyset$
for $\texttt{i}, \texttt{j}, \texttt{m}, \texttt{n} \in [I] \times [J] \times [M] \times [N]$ **do**
  $\boldsymbol{\theta} \leftarrow \texttt{GetAngles}(\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}, i, j, m, n)$
  for $t \in K$ **do**
    $p_t \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    $p_f \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
      - $H[t-1][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    $p_b \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
      - $H[t+1][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    if $\neg(p_f == 0$ 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

for power distance
$l \leftarrow \emptyset$
for $\texttt{i}, \texttt{j}, \texttt{m}, \texttt{n} \in [I] \times [J] \times [M] \times [N]$ **do**
  for $t \in K$ **do**
    $p_t \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    $p_f \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
      - $H[t-1][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    $p_b \leftarrow H[t][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
      - $H[t+1][\texttt{pid}][\texttt{i}][\texttt{j}][\texttt{m}][\texttt{n}]$
    if $\neg(p_f == 0$ and $p_b == 0$ and $p_t == 0)$ **then**
      $l \leftarrow l \cup \{(p_t, t, \mathbf{TA}[\texttt{i}], \mathbf{TE}[\texttt{j}], \mathbf{RA}[\texttt{m}], \mathbf{RE}[\texttt{n}])\}$
    end if
  end for
end for

Clustering

get l from measurement preprocessing
if Euclidean distance **then**
  $l \leftarrow \texttt{Scaling}(l)$
else if Power distance **then**
  if Use decibel values for power **then**
    $P_{dB} \leftarrow \{\log v[0], \forall v \in l\}$
    $p_{\min} \leftarrow \min P_{dB}$
    $v[0] \leftarrow \log v[0] - p_{\min}, \forall v \in l$
  end if
  choose time scaling factor
end if
choose centroid model or density model
perform clustering, get $c_x \forall x \in l$

Identifying MPCs' Time Range

for $t \in K$ **do**
  $\mathbf{x}_t \leftarrow \{x'\}, \forall x'$ with $x'[1] == t$
  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 $\mathbf{c} = \langle c^t \rangle, \forall t \in K$

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

$\boldsymbol{\theta} \leftarrow \emptyset$

for $\mathbf{L} \in \{\mathbf{TA}, \mathbf{TE}, \mathbf{RA}, \mathbf{RE}\}$ **do**

  $k \leftarrow$ the corresponding index for the list

  if $\cos(x) \geq 0, \forall x \in \mathbf{L}$ or $\cos(x) \leq 0, \forall x \in \mathbf{L}$ **then**

    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\sin \mathbf{L}[k]\}$

  else if $\sin(x) \geq 0, \forall x \in \mathbf{L}$ or $\sin(x) \leq 0, \forall x \in \mathbf{L}$ **then**

    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\cos \mathbf{L}[k]\}$

  else

    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} \cup \{\cos \mathbf{L}[k], \sin \mathbf{L}[k]\}$

  end if

end for

return $\boldsymbol{\theta}$

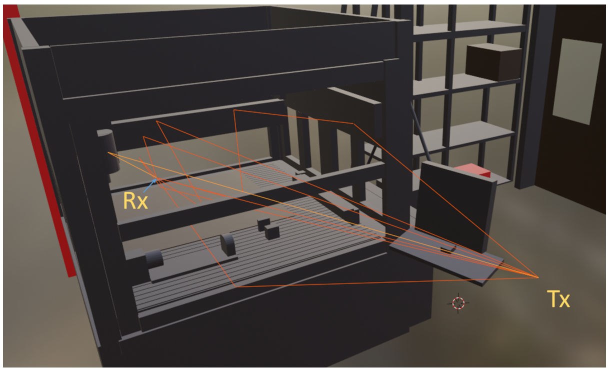


Fig. 1. Ray-tracing result

III. Experiments

Scaling factors of features for clustering and hyperparameter settings of different clustering methods are shown in Table III. We perform parameter searching for k in K(Power)Means and e in DBSCAN and pick the most representative results.

A. Comparison of Results

1) Ray Tracing vs Measurement: We firstly validate the rays computed by ray-tracing simulation in Figure 2. According to Figure 2(a), the LOS and S2 peaks are accurately identified in ray tracing (r0 and r3), The S4 and S5 peaks in measurements are also identified with good timing, but with offset on power (r5 and r6 in ray-tracing). r1 is possibly not present because it is a reflection at the edge of the machine according to ray-tracing. A light mismatch between measurement and simulation may cause this ray to be missing in measurement. Similarly, the mismatch between r2 and S1 is possibly due to the inaccurate modelling the of the cylindrical milling head. In general, we confirm that the ray-tracing experiment performed is a good match of the measurement.

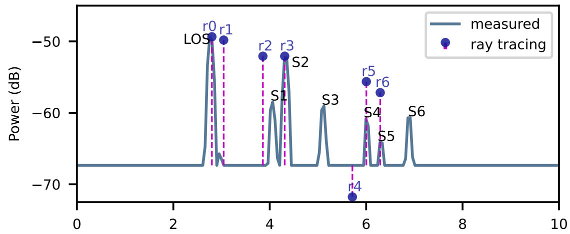
2) Clustering on Ray-Tracing Points: We firstly perform clustering on all of the points of ray-tracing result (without angular selection) and then observe r0 to r6 from them. We obtain clusters as {(r0, r1), (r2, r3), (r4, r5, r6)}.

If clustering algorithm is solely performed on points r0 to r6, we obtain identical result. As we increase the number of clusters in K-means or decrease the distance threshold in DBSCAN, r4 is firstly separated from r5, r6 to form a single-point cluster. r2 and r3 are separated if we change the hyper-parameters further.

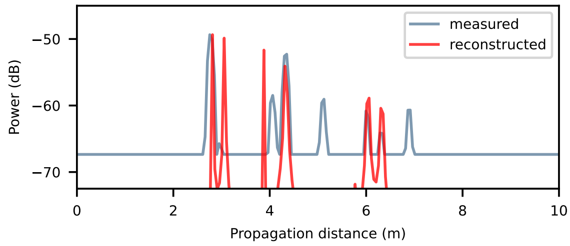
3) Direct Clustering: We show the clustered time steps of measurements in Figure 3. After observing the reflective surface and manually identifying the time range of each MPC, we compare the clustering results of different method in Table IV. For each MPC / MPC-cluster, an evaluation time range $\mathcal{R}_{ev}$ is defined around the center of the peaks. Out of the time steps in this range, we identify the true range of MPC (clusters): $\mathcal{R}_{true}$. The results given by Algorithm 1: $\mathcal{R}_p$ are then used for evaluation. The accuracy is then computed as:

$$acc. = \frac{|\mathcal{R}_{true} \cap \mathcal{R}_p| + |(\mathcal{R}_{ev} - \mathcal{R}_{true}) \cap (\mathcal{R}_{ev} - \mathcal{R}_p)|}{|\mathcal{R}_{ev}|}$$

We notice that all of the methods successfully separates line of sight component, single-refection components (S1-S3), and



(a) Ray-tracing MPCs in scatters marked from r0 to r6. Peaks in measurements are marked by LOS and S1-S6



(b) Ray-tracing MPCs are reconstructed into power-delay profile

Fig. 2. Comparing Measurement with ray-tracing results.

TABLE III
SCALING FACTORS AND HYPER-PARAMETER SETTINGS IN CLUSTERING

		Name	Value
Scaling		p_t, p_f, p_b in dB	0.2
		t	$2.48 \cdot 10^9$
		$\boldsymbol{\theta}$	8.0
		ζ for scaling time component in power dist.	5
Hyper-parameter	DBSCAN	e for Euclidean dist.	{2.0, 2.5, 3.0}
		e for power dist.	{0.08, 0.01, 0.013}
	K(Power)-Means	k	{3, 4, 5, 6, 7}

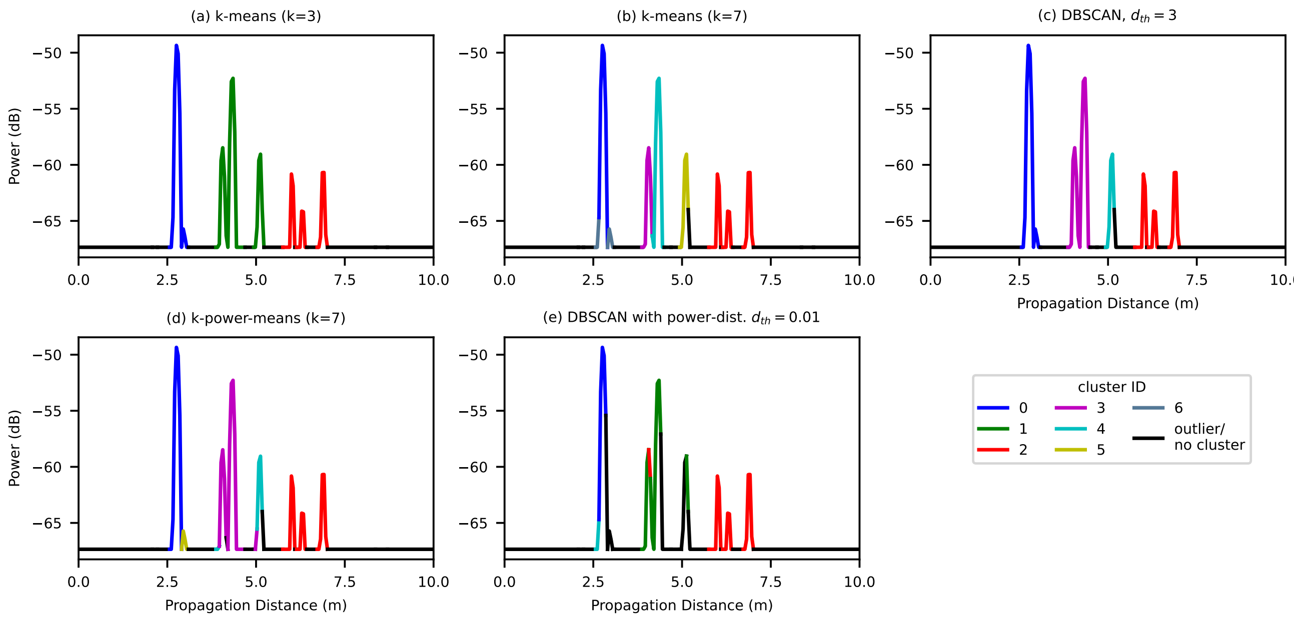


Fig. 3. Clustered power-delay profile of different clustering methods directly applied on Measurement

TABLE IV
CLUSTERING ACCURACY OF TIME STEPS OF DIFFERENT CLUSTERING METHODS DIRECTLY APPLIED ON MEASUREMENT

		Distance type	Euclidean distance			power distance, $\tau = 5$	
		Methods	K-means (k=3)	K-means (k=7)	DBSCAN, $d_{th} = 3$	KPowerMeans (k=7)	DBSCAN, $d_{th} = 0.01$
MPCs	Line of sight		0.95	0.76	0.95	0.86	0.71
	Single bounce reflection	S1	0.86	0.92	0.95	0.70	0.81
		S2		1.00			
		S3		1.00	0.91	0.83	
	High-order reflection	S4	0.84	0.84	0.84	0.84	0.91
		S5					
		S6					

multi-reflection (S4-S6) components, which corresponds to the clustering results of ray tracing points in Section III-A2. The difference between the results of the methods lies in whether the MPCs of single-reflection are clustered together, where the separation of these MPCs are dependent on the hyper-parameters of clustering algorithm, e.g. the setting of total cluster number k in K-means and KPowerMeans. However, the multi-refelction points (S4-S6) are not separated even if the hyperparameters are set to allow more precise clustering, e.g. setting $k = 7$ which is the total number of peaks in measurements. As the power of the MPCs in multi-reflection are much lower than the others, the point distance between S4-S6 are likely to be smaller than their distance to other MPCs, especially when power plays an important role e.g. in power distance. The inability to discern low-power MPCs is tolerable because they play a less important role than other MPCs in communication applications due to the high signal-noise ratio. Note that such clustering technique is especially useful for THz band. Due to the sparsity of channels on THz frequency, the peaks in power-delay profiles are well separated and therefore suitable to be applied with our solution. In terms of the usage of Euclidean distance or power distance, we compare Figure 3 (a)-(c) with Figure 3 (d)-(e), and find that the methods using Euclidean distance perform better by segmenting peaks clearly while the power-distance-based methods tend to yield small fractions for signals of single bound reflection.

IV. CONCLUSION

In this paper, we introduced a novel workflow for clustering MPCs directly from channel sounding measurements. The clustering results are validated by comparing the measurement data with ray-tracing results where the propagation environment is reconstructed. Through our algorithm, the signal segments with similar propagation characteristics are identified directly without processing measurements to identify the peaks of each MPC. The results can be directly utilized by communication applications, and are especially useful when it is hard to reconstruct the propagating environment in ray-tracing platforms.

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] M. Steinbauer, H. Ozcelik, H. Hofstetter, C. Mecklenbrauker, and E. Bonek, "How to quantify multipath separation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 85, no. 3, pp. 552–557, 2002.

[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 *Proceedings of the Second International Conference on Knowledge Discovery in Databases and Data Mining (KDD)*, 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] J. MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, 1967, pp. 281–297.

[14] J. Hoydis, S. Cammerer, F. A. Aoudia, et al., "Sionna: An Open-Source Library for Next-Generation Physical Layer Research," arXiv, Information Theory (Cs.IT); Artificial Intelligence (Cs.AI); Machine Learning (Cs.LG) 10.48550/2203.11854, Mar. 2023.