Unsupervised Machine Learning-Based User Clustering in THz-NOMA Systems

Yushen Lin, *Student Member, IEEE*, Kaidi Wang, *Member, IEEE*, and Zhiguo Ding, *Fellow, IEEE*

***Abstract*—In this letter, different unsupervised machine learning (ML)-based user clustering algorithms, including K-Means, agglomerative hierarchical clustering (AHC), and density-based spatial clustering of applications with noise (DBSCAN) are applied in non-orthogonal multiple access (NOMA) assisted terahertz (THz) networks. The key contribution of this letter is to design ML-based approaches to ensure that the secondary users can be clustered without knowing the number of clusters and degrading the performance of the primary users. The studies carried out in this letter show that the proposed schemes based on AHC and DBSCAN can achieve superior performance on system throughput and connectivity compared to the traditional clustering strategy, i.e., K-means, where the number of clusters is determined in an adaptive and automatic manner.**

***Index Terms*—Machine learning (ML), non-orthogonal multiple access (NOMA), user clustering.**

I. Introduction

With the fifth-generation (5G) communication successfully standardized and deployed globally, the research for sixth-generation (6G) wireless communication recently attracts significant attention as the current 5G communication networks can no longer meet all the requirements of tremendous traffic growth envisioned in the future wireless networks, i.e., ultra-high throughput, ultra-massive connectivity, and ultra-low latency, etc [1]. Non-orthogonal multiple access (NOMA) combined with beamforming (BF) and terahertz (THz) technologies is seen as a promising solution to meet the growing demand. The THz spectrum offers a vast amount of bandwidth for communication, and NOMA can boost spectral efficiency by enabling multiple users to share the same bandwidth resources [2], [3].

This letter differs from existing works on THz-NOMA [4], [5], [6], [7] in that it focuses on the application of machine learning(ML)-based clustering techniques for legacy THz-NOMA systems, where beams have been pre-configured to serve primary users. The main contributions of this letter include: i) investigating how existing beams can be utilized to serve secondary users by applying unsupervised ML-based user clustering algorithms, specifically agglomerative hierarchical clustering (AHC) and density-based spatial clustering of applications with noise (DBSCAN); ii) proposing methods that offer a more robust performance, particularly when the number of clusters needs to be determined dynamically, as compared to the benchmark K-Means algorithm. Simulation results are presented to demonstrate the feasibility and effectiveness of these ML approaches in THz-NOMA networks and to evaluate the trade-off between system performance and complexity.

Manuscript received 10 March 2023; accepted 25 March 2023. Date of publication 29 March 2023; date of current version 11 July 2023. This work was supported by the U.K. EPSRC under Grant EP/W034522/1, and in part by H2020 H2020-MSCA-RISE-2020 under Grant 101006411. The associate editor coordinating the review of this article and approving it for publication was D. Mishra. *(Corresponding author: Yushen Lin.)*

Yushen Lin and Kaidi Wang are with School of Electrical and Electronic Engineering, The University of Manchester, M13 9PL Manchester, U.K. (e-mail: yushen.lin@student.manchester.ac.uk; kaidi.wang@ieee.org).

Zhiguo Ding is with Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, UAE, and also with the Department of Electrical and Electronic Engineering, University of Manchester, M13 9PL Manchester, U.K. (e-mail: zhiguo.ding@manchester.ac.uk).

Digital Object Identifier 10.1109/LWC.2023.3262788

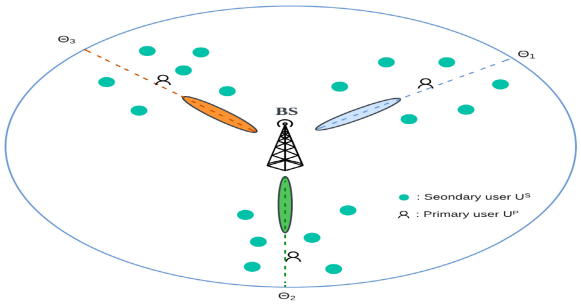
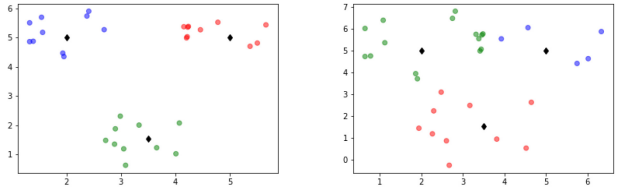


Fig. 1. A schematic diagram of the considered system model.



(a) Focused user deployment. (b) Overlapping user deployment.

Fig. 2. Clustering results from different algorithms under focused and overlapping user deployment with $N = 30$.

Users = N Secondary Users = N - M (U(s))
Clusters = Primary Users = M (U(p))
Antennas (Base Station) = L (M <= L)

II. System Model and Problem Formulation

Consider a system with one base station (BS) and N single-antenna users, where the BS is located at the centre of a disc and the users are randomly distributed. The collection of all users and secondary users are denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$ and $\mathcal{J} = \{1, 2, \ldots, J\}$, respectively. It is assumed that N users are divided into M clusters and served by the BS with L antennas where $M \leq L$. As shown in Fig. 1, in each cluster, one user is selected to play the role of the primary user, and the rest of the users become secondary users [2]. Thus, in the considered system, there are M primary users and $J = N - M$ secondary users, where $J \geq M$, denoted by U^P and U^S, respectively.

A. Spatial and Beamforming Model

The illustrations of the user deployments are followed by the Poisson cluster process (PCP), where the point is identically

and independently distributed (i.i.d.) around the parent point as shown in Fig. 2. It resembles physically-clustered user scenarios, i.e., meeting rooms or cafés. Hybrid BF is used to generate beams for the primary users, which means that the signal vector sent by the BS could be expressed as follows:

$$\tilde{\mathbf{s}}_v^P = [\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_M]\mathbf{D}_{bm}\mathbf{s}_v^P, \quad (1)$$

$\tilde{\mathbf{f}}_m$ denotes the analog BF vector for U_m^P, and $\mathbf{D}_{bm}$ denotes the $M \times M$ digital BF matrix, and $\mathbf{s}_{v^P} = [s_v 1^P, \ldots, s_v M^P]^T$.

The analog BF vector can be selected from a pre-configured codebook [8], [9]:

$$\tilde{\mathbf{f}}_m \in \left\{ M^{-\frac{1}{2}} \mathbf{a}_d\left(\frac{(2\pi \times 0)}{\Psi}\right), \ldots, M^{-\frac{1}{2}} \mathbf{a}_d\left(\frac{2\pi(\Psi-1)}{\Psi}\right) \right\}, \quad (2)$$

where Ψ is the size of codebook, and $\mathbf{a}_d(\theta)$ is a $L \times 1$ vector to denote the angle of departure (AoD). Specifically, $\mathbf{a}_d(\theta)$ is given by:

$$\left[1, e^{-j2\pi \frac{A_s}{\lambda}\sin(\theta)}, \ldots, e^{-j2\pi(L-1)\frac{A_s}{\lambda}\sin(\theta)}\right]^T, \quad (3)$$

where A_s and λ denote antenna spacing and wavelength, respectively. Without loss of generality, the value of $\frac{A_s}{\lambda}$ is assumed to be $\frac{1}{2}$ in the system. U_m^P's BF vector $\mathbf{f}_m$ can be chosen by using the codebook and finding a vector whose θ is closest to the corresponding primary users' AoD. The composite BF vector of primary users $\mathbf{f}_m$ could be written as:

$$\mathbf{f}_m = [\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_M]\mathbf{d}_m, \quad (4)$$

where $\mathbf{f}_m$ can be considered as bandwidth resources to be allocated to U^S, and $\mathbf{d}_m$ denotes the m-th column of $\mathbf{D}_{bm}$.

B. Channel and Signal Model

The channel vector between the BS and U_m^P is given by [10],

$$\mathbf{h}_\mathbf{m}^\mathbf{P} = \mathbf{a}_d(\theta_m)\frac{g_m^P\sqrt{PL_m^P}}{(1+r_{m,P})}, \quad (5)$$

where $g_m^P \sim \mathcal{CN}(0, \sigma_m^2)$, PL_m^P and $r_{m,P}$ represent the channel, path loss, and distance, respectively. Specifically, the path loss is defined as follows [3]:

$$PL_m^P = \left(\frac{c}{4\pi f_c}\right)^{-2} e^{\zeta r_{m,P}(r_{m,P}^{\Gamma_{PL}}+1)}, \quad (6)$$

where c denotes the speed of light, f_c denotes carrier frequency, ζ represents the molecular absorption coefficient, and Γ_{PL} is the path loss exponent. The signal received at U_m^P could be expressed as follows [2]:

$$y_m^P = \frac{|\varrho_m^P|^2}{\sqrt{PL_m^P}}\mathbf{a}_d^H(\theta_m^P)\sum_{i=1}^{M}\mathbf{f}_i\left(\sqrt{p_i^P}s_i^P + x_s\right) + n_m^P, \quad (7)$$

where ϱ_m^P denotes the fading coefficient, θ_m^P denotes the AoD of the corresponding primary user and n_m^P denotes the additive while Gaussian noise with noise power σ^2. Moreover, the transmitted signal x_s denotes as $\sum_{j=1}^{J} b_{ji}\sqrt{p_{ji}^S}s_{ji}^S$, where p_{ji}^S is the transmit power of secondary user and s_{ji}^S is the message to secondary user U_j^S via beam i.

According to the decoding strategy in [2], all the primary users decode their own signals and treat the signals for secondary users as noise. The data rate of primary user U_m^P can be expressed as follows:

$$R_m^P = \log\left(1 + \frac{q^p p_m^P}{q^p \sum_{j=1}^{J} b_{jm}p_{jm}^S + \Upsilon^m + \sigma^2}\right), \quad (8)$$

where q^p denotes $\frac{|\varrho_m^P|^2}{PL_M^P}|\mathbf{a}_d^H(\theta_m^P)\mathbf{f}_p|^2$, the inter-beam interference Υ^m is:

$$\Upsilon^m = \frac{|\varrho_m^P|^2}{PL_m^P}\sum_{i=1,i\neq m}^{M}|\mathbf{a}_d^H(\theta_m^P)\mathbf{f}_i|^2\left(p_i^P + \sum_{j=1}^{J} b_{ji}p_{ji}^S\right), \quad (9)$$

where b_{jm} and b_{ji} are the beam allocation indicator as suggested in [11].

At the secondary user, the signal for U_m^P via the same beam can be decoded and removed, and then U_j^S can decode its own signal at the following data rate:

$$R_{j,m}^S = \log\left(1 + \frac{\frac{|\varrho_j^s|^2}{PL_j^S}p_{jm}^S|\mathbf{a}_d^H(\theta_j^S)\mathbf{f}_m|^2}{\Upsilon^{jm} + \sigma^2}\right), \quad (10)$$

where the path loss PL_j^S, and inter-beam interference of U_j^S Υ_{jm} are defined similarly with those for U_m^P.

C. Problem Formulation

In this subsection, the formulated optimization problem is presented. Afterward, how the different clustering algorithms. In the considered THz-NOMA system, the sum rate of the secondary users could be expressed as follows:

$$R_{sum}^S = \sum_{j=1}^{J}\sum_{m=1}^{M} b_{jm}R_{jm}^S. \quad (11)$$

A sum-rate maximization problem can be formulated as:

$$\max_{M,\mathcal{C}} \; R_{sum}^S \quad \text{(P1a)}$$

$$\text{s.t.} \; \mathcal{C}_m \cap \mathcal{C}_{m'} = \emptyset, \mathcal{C}_m, \mathcal{C}_{m'} \in \mathcal{C}, m \neq m', \quad \text{(P1b)}$$

$$\sum_{j=1}^{J}\sum_{m=1}^{M} b_{jm}p_{jm}^S \leq P_T, \quad \text{(P1c)}$$

$$R_m^P \geq \bar{R}_m^P, \forall m, \quad \text{(P1d)}$$

where the set of cluster index sets denotes by $\mathcal{C}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$, and $\mathcal{C}_m$ is denoted by the set of users in cluster m. P_T represents the total transmission power of the BS, and $\bar{R}_m^P$ denotes the QoS primary users' target data rate. Constraint (P1b) ensures that each user can only be assigned to one cluster. Constraint in (P1c) represents the total transmission power budget. Constraint (P1d) ensures that no secondary user is scheduled on beam f_k but primary user U_m^P still suffers interference from the secondary users on the other beam. It is worth mentioning that the power utilization of primary users is not the main concern based on the (P1a). The design of BF with power allocation is not the main scope and is assumed

to be given in this letter, which can be treated as promising research directions.

Humans can easily identify the three clusters in the user deployment shown in Fig. 2(a). However, it is a non-trivial task for the BS to determine the appropriate number of clusters. Since the user clustering problem is an integer programming, which is an NP-complete and non-convex optimization and highly challenging to solve [12]. Moreover, this is further compounded by the degradation of the sum capacity due to inter-beam interference. These challenges motivate the application of ML-based user clustering algorithms to assist the BS to make proper decisions in an automatic and adaptive manner.

III. Description of Proposed ML-Based User Clustering Approaches

In this section, how the different clustering algorithms, i.e., K-means, DBSCAN, and AHC can be applied in the proposed system is demonstrated.

A. Benchmark Scheme: K-Means Clustering

K-means is a partitioning-based clustering algorithm designed to find a pre-defined number of clusters, which are represented by their centroids. The rationale behind the benchmark K-means is described as follows: denote Z by the pre-configured number of clusters, and a_j, $j \in J$, by the AoD of secondary users. Choose K centroids, and each AoD of users is then assigned to the closest centroid by comparing the distance between users' AoD and the center of each cluster $dist(a_j, \mu_m)$, and each set of users assigned to the same centroid forming a cluster. The criterion of recomputing the center of each cluster can be expressed as follows:

$$\mu_m = \frac{1}{|\mathcal{C}_m|} \sum_{j \in \mathcal{C}_m} a_j. \tag{12}$$

Then repeat the assignment and update the centroid until the threshold is met or equivalently.

Proposition 1: The objective function in K-means, i.e., $dist(a_j, \mu_m)$, can be expressed as follows:

$$\min_{m=1,\ldots,M} -2\mathcal{R}[a_j^H \mu_m] + \mu_m^H \mu_m. \tag{13}$$

Proof: This proposition can be proved by expanding the original objective function:

$$\|a_j - \mu_m\|^2 = -2\mathcal{R}[a_j^H \mu_m] + a_j^H a_j + \mu_m^H \mu_m, \tag{14}$$

where the term $a_j^H a_j$ can be neglected as it has no impact on the optimal solution of (P1a) as the AoD for the users' channels is fixed. ■

The proposed methods use the AoD of the users' channels as feature, rather than the users' path loss. This is because the users with similar path loss may have significantly different azimuth angles. Therefore they should not be grouped into the same cluster in the considered massive MIMO-NOMA system.

B. DBSCAN-Based User Clustering

Unsupervised learning techniques that identify unique groups or clusters in the data are referred to as density-based clustering. The motivation for applying DBSCAN in the proposed system is driven by its ability to determine Z based on the density of the data points and the adjustable parameters *eps* and *minPts*, providing an advantage over K-means in a dynamic manner.

The following definitions and concepts are provided to better understand DBSCAN:

Otherwise considered Noise

Definition 1 (minPts): The minimum of observations that be grouped together for a region to be considered dense.

Definition 2 (Eps): A measurement of distance will be used to find observations in the neighborhood of any given point.

Definition 3 (Neighborhood): The distance between two points a and b, determined by the Euclidean distance.

Definition 4 (Noise): Let $\mathcal{C}_1, \ldots, \mathcal{C}_n$ be the clusters w.r.t. $minPts_j$, $j = 1, \ldots, n$. Noise is determined as a set of objects in the observations that do not exist in any cluster $\mathcal{C}_j$, i.e., noise $= \{a \in D | \forall j : a \notin \mathcal{C}_j\}$, where D is all the users' AoD.

The procedure of DBSCAN can be abstracted as follow. It starts by randomly selecting a point in the dataset. If there are at least *minPts* points within a radius of *Eps* to the point, then each point to be involved in the same cluster, otherwise, assign it to *noise*. The clusters are then expanded by recursively repeating the *neighborhood* calculation for each neighboring point. The DBSCAN algorithm has been modified to determine the value of *Eps* by calculating the average distance between each point and its k-nearest neighbors (KNN). The average k-distances are then plotted in ascending order on a k-distance graph. The optimal value of *Eps* is the point of maximum curvature where the plot has the greatest slope [13].

C. AHC-Based User Clustering Algorithm

The aforementioned clustering algorithms have common limitations, such as the number of clusters or the hyper-parameters must be pre-configured by a human user. In this case, AHC has the advantage because these parameters are not required input and the results are reproducible. The process of AHC can be summarized as follows:

- It begins by dividing the dataset into singleton nodes and combining the two currently closest nodes into one node until only one node remains, which contains the entire dataset. For example, the procedure will treat each user's AoD as one cluster, and therefore, the number of clusters will be reduced from N in the proposed system.
- A cluster is formed by merging the two closest data points, resulting in $N - 1$ clusters.
- The merging step is repeated the last step until the threshold is achieved.

The input of AHC is a condensed matrix with a dissimilarity index and the output is traditionally a stepwise dendrogram containing clustering results.

Definition 5 (Dissimilarity Index): On a set $\mathcal{F}$ is a map $: \mathcal{F} \times \mathcal{F} \to [0, \infty)$ which is reflexive and symmetric, where $\mathcal{F}$ has N elements.

Definition 6 (Stepwise Dendrogram): Denote one finite set $\mathcal{F}_0$ with cardinality $N = |\mathcal{F}_0|$, a stepwise dendrogram is a list of $N-1$ triples (u_i, v_i, δ_i) where $i = 0, \ldots, N-2$ such that $\delta i \in [0, \infty)$ and $u_i, v_i \in \mathcal{F}_i$, where F_{i+1} is recursively defined as $(\mathcal{F}_i \setminus \{u_i, v_i\}) \cup n_i$ and $n_i \notin \mathcal{F} \setminus \{u_i, v_i\}$ is a label for a new node, where n_i is the new node formed by joining the nodes u_i and v_i at the distance δ_i.

The strategy used to determine if the clusters are to be merged is called Lance-Williams algorithms (LWA). The advantage of using the LWA is that it does not need to keep the original data points. Denote $\mathcal{C}_p$ and $\mathcal{C}_q$ by the next two clusters to be merged together, $D(p, q)$ by the distance between $\mathcal{C}_p$ and $\mathcal{C}_q$, and $D(pq, r)$ by the distance between the cluster $\mathcal{C}_p \cup \mathcal{C}_q$ and cluster $\mathcal{C}_r$. For more general argument, the distance $D(pq, r)$ can be recursively computed as [14]:

$$D(pq, r) = \chi_p D(p, r) + \chi_q D(q, r) + \beta D(p, q) + \gamma |D(p, r) - D(q, r)|, \quad (15)$$

where χ_p, χ_q, β, and γ denote the coefficients that depend on the numbers of data in the cluster C_p, C_q, and C_r. Different coefficients correspond to different distance update strategies. In this letter, Ward's method [14] is applied, while these coefficients are defined by $\chi_p = \frac{N_p+N_r}{N_p+N_q+N_r}$, $\chi_q = \frac{N_q+N_r}{N_p+N_q+N_r}$, $\beta = \frac{-N_r}{N_p+N_q+N_r}$ and $\gamma = 0$, where N_p, N_q, and N_r are the numbers of data in $\mathcal{C}_p$, $\mathcal{C}_q$, and $\mathcal{C}_r$, respectively. The cluster dissimilarity in Ward's method between clusters is given by:

$$D(\mathcal{C}_p, \mathcal{C}_q) = \frac{N_p N_q}{N_p + N_q} ||\mu_p - \mu_q||^2, \quad (16)$$

where μ_p and μ_q denote the centroid of $\mathcal{C}_p$ and $\mathcal{C}_q$ cluster by averaging the normalized users' AoD:

$$\mu_p = \frac{1}{N_p} \sum_{l=1}^{N_p} a_{p,l} \quad (17)$$

Proposition 2: The AHC with Ward's method fulfills the reducibility property, and merge cost $D(\mathcal{C}_p \cup \mathcal{C}_q, \mathcal{C}_r \cup \mathcal{C}_s)$ is independent of the situation where $\mathcal{C}_r, \mathcal{C}_s$ are merged after $\mathcal{C}_p, \mathcal{C}_q$ are merged or another way round.

Proof: The achieved reducibility property of the proposed method proved can be by checking the equation (16). For a more general argument of representing the cluster dissimilarity in Ward's method [14]:

$$D(C_k, C_l) = \frac{N_k N_l}{N_k + N_l} \left(\frac{2}{N_k N_l} \sum_{a \in C_k} \sum_{b \in C_l} ||a - b||^2 - \frac{1}{N_k^2} \sum_{a \in C_k} \sum_{a' \in C_k} ||a - a'||^2 - \frac{1}{N_l^2} \sum_{b \in C_l} \sum_{b' \in C_l} ||b - b'||^2 \right). \quad (18)$$

This equation holds regardless of whether the data is Euclidean or not, since it can be proved inductively from the recursive distance update equation of Ward's method. This shows that dissimilarities in Ward's method are likewise unrelated to the sequence of merging steps. ■

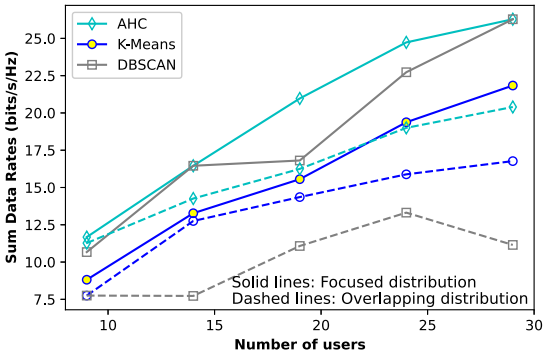


Fig. 3. Sum data rate against the number of secondary users of different clustering algorithms with $L = 4$.

The primary improvement in the AHC algorithm is in the process of determining the number of clusters, which can be summarized as follows. The "L method" is employed as the evaluation metric plotted against the number of clusters, to calculate the metric for a range of trial cluster counts. To locate the knee point, which represents the optimal number of clusters, two lines are fitted via linear regression. The fit is further refined through an iterative process, which proceeds as follows. Denote the left and right sequences of observations as L_f and R_f partitioned at $x = f$, such that L_f includes observations with $x = 2 \cdots f$, and R_f is with $x = f + 1 \cdots e$, where $f = 3 \cdots e - 2$. Then, the expression for obtaining the knee point of the evaluation graph can be obtained as follows:

$$\hat{f} = \underset{f}{\operatorname{argmin}}\, RMSE(f), \quad (19)$$

where the total root mean squared error *RMSE*(f), which separates L_f and R_f at point $x = f$, is given by:

$$RMSE(f) = \frac{f-1}{e-1} RMSE(L_f) + \frac{e-f}{e-1} RMSE(R_f). \quad (20)$$

It is worth mentioning that the different linkage methods yield the same clustering results in Fig. 2(a), but different in Fig. 2(b). The modification made to the linkage equations to a general form, as represented by (15) allowing further investigation of the impact of the various linkage methods on the performance of overlapping distribution, which may be include in future works.

IV. Simulation Results

Computational simulations are presented in this section to evaluate the performance of the ML-based user clustering methods. For all conducted simulations [2]: $f_c = 300$ GHz, $\sigma^2 = -90$ dBm, $\zeta = 5e^{-3}$, $\varrho_m^P = 2$, $\Psi = 30$, bandwidth = 24 GHz. The radius of focused and overlapping deployment are 1 m and 2.5 m, respectively.

The user deployment shown in Fig. 2 illustrates that the three clustering results of focused user deployment show the same cluster results, with $M = 3$, $Eps = 0.2$, and $minPts = 5$, respectively. Only the overlapping case of the result in AHC is illustrated in Fig. 2(b) due to space limitations.

In Fig. 3, the impact of the number of users on the performance of the three considered ML cluster algorithms

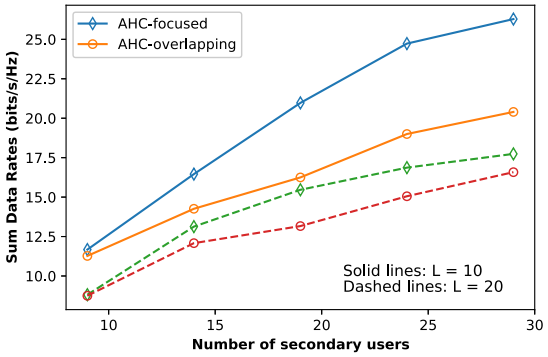
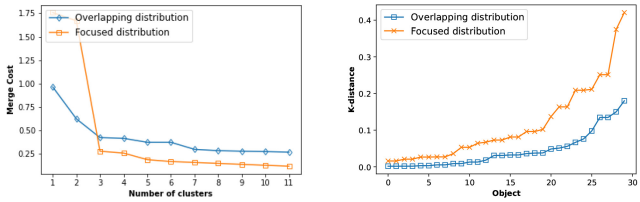


Fig. 4. Sum data rate versus the number of secondary users of AHC with the different number of antennas with $L = 3$.



(a) Merge cost in AHC. (b) KNN in DBSCAN.

Fig. 5. Analysis of merge cost in AHC and Nearest Neighbour in DBSCAN.

is studied, where $L = 4$. The number of clusters and the value of *Eps* are set to random for K-Means and DBSCAN. As can be seen from the figure that AHC outperforms the other two algorithms, while K-Means and DBSCAN are no longer promising when the number of clusters and hyper-parameters are randomly chosen. The dashed lines in Fig. 3 show that the overall system throughput in the case with overlapping deployment is lower than that with the focused user deployment as the beams have to cover the larger areas and users are more separated compared to the deployment in Fig. 2(b). In the original DBSCAN algorithm, *Eps* and *minPts* are required to be pre-configured, and both have different impacts on the clustering results. *Eps* controls the local neighborhood of the points and *minPts* controls the tolerant of the algorithm to noise. When *Eps* is too small, the majority of the points will not be clustered. Otherwise, it causes close clusters to be merged into a single cluster, which means that all points will form as a singleton cluster.

The impact of the antenna number on the performance in the proposed system is depicted in Fig. 4. It can be seen from the figure that the sum data rate drops when the number of antennas increases and such reduction can be explained as follows. Increasing the number of antennas makes it more difficult for secondary users to match a beam when spatial beams become more directional.

In Fig. 5, the analysis of merge cost in AHC and KNN in DBSCAN is presented to show the strategies of selecting the number of clusters in AHC and hyper-parameters in DBSCAN as proposed in Section III. Recall that Fig. 2 shows the distributions of users, which could result in different curvatures in the graphs of merge cost Fig. 5(a) and 5(b). The figure shows that the value of *minPts* is at the maximum curvature around 0.2 with the focused user deployment. However, determining the maximum curvature can be more challenging for the overlapping user deployment compared to merge cost in AHC. The average run time complexity of AHC and DBSCAN are both $\mathcal{O}(N^2)$ as stated in [13], [14]. Some spatial access methods, e.g., R*-trees, could descend complexity to $\mathcal{O}(n \log n)$ in DBSCAN. However, such methods is not recommended due to the same time-consuming nature of the KNN algorithm used to determine *minPts* and *Eps* when N is large.

V. Conclusion

In this letter, in order to ensure that the secondary users can be well clustered without knowing the number of clusters and degrading the performance of the primary users, three unsupervised ML-based user clustering algorithms, namely K-Means, AHC, and DBSCAN, were applied in NOMA-assisted THz networks. The simulation results demonstrated that the proposed algorithm based on AHC can outperform the other two, i.e., K-means and DBSCAN, where the number of clusters and hyper-parameters are selected in an adaptive and autonomous manner. For future works, the optimization of beamforming design and power allocation can be included.

References

[1] X. You et al., "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, pp. 1–74, Feb. 2021.
[2] Z. Ding and H. V. Poor, "Joint beam management and power allocation in THz-NOMA networks," 2022, *arXiv:2205.12938*.
[3] Z. Ding, "Potentials and limits of using preconfigured spatial beams as bandwidth resources: Beam selection versus beam aggregation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2575–2579, Dec. 2022.
[4] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.
[5] Z. Ding and H. V. Poor, "Design of THz-NOMA in the presence of beam misalignment," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1678–1682, Jul. 2022.
[6] J. Tan and L. Dai, "THz precoding for 6G: Challenges, solutions, and opportunities," *IEEE Wireless Commun.*, early access, May 9, 2022, doi: 10.1109/MWC.015.2100674.
[7] S. B. Melhem and H. Tabassum, "User pairing and outage analysis in multi-carrier NOMA-THz networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5546–5551, May 2022.
[8] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[9] Y. Zou, W. Rave, and G. Fettweis, "Analog beamsteering for flexible hybrid beamforming design in mmWave communications," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Sep. 2016, pp. 94–99.
[10] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, "Unsupervised machine learning-based user clustering in millimeter-wave-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7425–7440, Nov. 2018.
[11] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4281–4298, Jun. 2019.
[12] J. Al-Karaki, A. Kamal, and R. Ul-Mustafa, "On the optimal clustering in mobile ad hoc networks," in *Proc. 1st IEEE Consum. Commun. Netw. Conf.*, 2004, pp. 71–76.
[13] 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. 2nd Int. Conf. Knowl. Disc.*, 1996, pp. 226–231.
[14] G. J. Szekely and M. L. Rizzo, "Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method," *J. Classification*, vol. 22, pp. 151–183, Sep. 2005.