Clustering-based Adaptive Beam Footprint Design for 5G Urban Macro-Cell

Puneeth Jubba Honnaiah, Eva Lagunas, Nicola Maturo, Symeon Chatzinotas
Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg
email: \{puneeth.jubba-honnaiah, eva.lagunas, nicola.maturo, symeon.chatzinotas\}@uni.lu

Abstract—In a dense 5G urban-eMBB environment, the user density and traffic loads follow a spatio-temporal variability. To meet high traffic demands, the 5G base stations exploit spatial multiplexing by means of Active Antenna Systems (AAS) and beamforming. However, pedestrian and vehicular users are highly mobile, rendering non-dynamic beamforming designs totally inefficient in terms of meeting the users' demand requests. In particular, the latter results in either overload or underutilized beams in a cell. Hence, a practical approach to meet such spatio-temporal heterogeneous demand is to consider dynamic and adaptive beam footprint design that takes into account both the actual users' position as well as the traffic loads. In this paper, we first study and evaluate the state-of-the-art fixed-cell beamforming (based on ITU-R M.2412-0) in a test environment and highlight its drawbacks. Next, we propose a adaptive macro-cell beam footprint design where the beams are dynamically shaped based on the spatial users distribution and their demand requests. Numerical simulations demonstrate the high system performance achieved by the proposed methodology.

Index Terms—Beam steering, Beam footprint design, Beamforming, Clustering, IMT-2020, Dense urban-eMBB layout.

I. INTRODUCTION

The advancements in beam steerable antennas and pattern reconfigurable antennas have made it possible to steer the signal power in the direction of interest [1], [2]. We can obtain such directional beams by altering the magnitude and phase of individual antenna signals in an array of multiple antennas [3]. Such beamforming techniques impact the beam radiation pattern and footprint coverage. Consequently, they can help in mitigating the co-channel interference, improving the signal-to-interference-plus-noise ratio (SINR) at the receiver side and increasing the throughput at the network nodes [4].

International Mobile Telecommunications-2020 (IMT-2020) systems are the successor of IMT-Advanced mobile systems and include the advanced abilities of IMT to be more efficient, fast, flexible, and reliable. The authors of [5] propose test environments and evaluation configurations for Indoor Hotspot-eMBB, Dense Urban-eMBB, and Rural-eMBB of IMT-2020. Accordingly, 3GPP defines a set of Layer 1 (PHY) and Layer 2 (MAC) beam management procedures to ensure the alignment between a beam and the users by beam sweeping, beam measurement, beam determination and beam reporting. Such beam-management procedures of 5G NR depend on measurement-plus-reporting in a closed-loop manner. Any delay in measurement or reporting may make the beam-correspondence between gNB and UE very difficult and the system may have to frequently perform beam-recovery or beam-switching leading to loss of performance.

Several authors in literature have discussed multiple beamforming and direction-of-arrival (DOA) estimation algorithms. The authors of [6] build on compact antenna structures with fast beamforming capabilities and, by combining them with a simple training procedure, enhance the average throughput of a multitenor system. The authors of [7] consider (DoA) estimation problem using electronically steerable parasitic array radiator (ESPAR) antenna based on compressive sensing. The authors of [8] propose a conformal circular beam-steering array with omnidirectional coverage and unidirectional beam.

Focusing on beam footprint design, the authors of [5] have proposed a Fixed Beam Footprint (FBF) methodology where the footprints are designed equidistant from each other. However, the users in a Dense Urban-eMBB environment are highly mobile and have different Quality-of-Service requirements. Hence their distribution and QoS demands change with time. This adds a challenge in FBF design, such that some beams will be overloaded and some will be underused.

Hence, in this paper, we consider a novel alternative approach for Adaptive Beam Footprint (ABF) design that is focused on network design where each beam covers an area with multiple users and hence independent of feedback failures. Furthermore, by the introduction of adaptability, we simplify the Radio resource management (RRM) procedures such as power allocation and user scheduling.

First, we consider the user locations and their traffic demand in previously mentioned Dense Urban-eMBB test environment and group the users into clusters based on
their positions and demand. Subsequently, we determine the optimal steering direction of the beams for every cluster that could best reduce the co-channel interference and meet the traffic demand of all the users in the cell. Furthermore, numerical results show that by the proposed ABF design, we can distribute the demand across all the beams more evenly.

II. SYSTEM MODEL

We consider a Dense Urban-eMBB test environment [5] like the one illustrated in Figure 1, which shows an example of three fixed beams that are spatially placed equidistant from each other. The macro sites of the macro-layer are regular hexagonal grid-like structure. A macro-layer base station is placed such that it serves three macro sites. The height of BS antenna is 25 meters. The distance between two macro-layer base station in 200 meters. The users in a macro-layer of the Dense Urban-eMBB environment include mobile users (both pedestrian and vehicular users) whose number, position and QoS requirements change with time. The device deployment include 80% indoor users and 20% outdoor (in-car) users who are randomly and uniformly distributed over the area under macro-layer. The mobility model includes fixed and identical speed of all users of the same mobility class in a random and uniformly distributed direction. The users speed of interest is assumed to be 3 km/hour for indoor users and 30 km/hour for outdoor (in-car) users [5].

III. PROPOSED ADAPTIVE BEAM FOOTPRINT DESIGN

The benchmark FBF design outlined in [5] defines three sites around a macro base station where each site represents a macro cell. Accordingly, the antenna bearing orientation proposed in the benchmark has three fixed beams whose centre directions of the main antenna lobe have 120 degrees separation from each other. This ensures omnidirectional (360 degrees) coverage using three directional beams and serves the total coverage region. However, users across the macro cells are not uniformly distributed and the traffic demand for a pedestrian user defers from a vehicular user. Also, as the users are mobile, their geographic positions vary with time.

Hence, in this section, we discuss steps involved in achieving proposed Adaptive Beam Footprint design. The dynamic and adaptive nature is introduced in both, finding the number of beams and also in designing the beam locations and footprint.

1) Firstly, we find the best value for the number of beams $K$ using the CH criteria.
2) Then, with an adapted value of $K$, we obtain cluster sets that evenly distributes the traffic throughput demand across all the beams. Then, we define the cluster centres as beam centres.
3) Lastly, we compute the Elevation and Azimuth angles at the base station for the beam centres defined in previous step.

A. Evaluate the optimal number of clusters

The authors of [5] suggest in having three fixed beams and does not comment on optimizing over the number of beams. Hence, choosing the number of beams defined as $K$ is not standardized and left for operators implementation preference. Accordingly, choosing fewer number of beams creates beam overload and does not fully exploit the spatial diversity. On the other hand, choosing too many beams will result in higher inter-beam interference due to the overlap of main beam with adjacent beams’ side-lobes. Also, having fixed beams for a dynamic demand scenario is not the optimal approach. Hence there is a need for effective algorithm to define Optimal number of beams.

The best value for the number of beams is indistinct, and the possible values depend on the shape and scale of user distribution. Therefore, in this work, we use cluster analysis [9], [10] to define the optimal number of beams as $K$. There exist a plethora of literature [10] on clustering methods with application in computer vision and pattern recognition. The use of such clustering methods based on partition for Macro-Cell beam footprint design in Dense urban-eMBB layout 5G systems, which we propose in this paper, is a novel approach.

Generally, increasing the value $K$ will reduce the amount of error in the user clustering. Hence, zero error could be achieved when each user is considered as a cluster. However, having a beam for every user is not an Optimal solution. Such solutions fail to exploit spatial diversity benefits and cause inter-beam interference. Hence the optimal decision of $K$ is a balance between the highest compression of the users using a single cluster, and the highest efficiency by assigning each user to its own cluster. Authors of [11]–[13] have focused on finding the best number of clusters for a given data set. However, in our case, we consider the Calinski-Harabasz (CH) criterion [14] and define an optimization problem as,

$$\text{maximize } \frac{\sum_{k=1}^{K} n_k \| M_k - \overline{M} \|}{(K - 1) \sum_{k=1}^{K} \sum_{m \in G_k} \| x_m - M_k \|}; \quad (1a)$$

subject to $K \leq M$, \quad (1b)

which maximizes the CH criteria for an optimal value of $K$. $M$ is the total number of users, $n_k$ is the number of users in a cluster $k$, $\overline{M}_k$ is the centroid of the cluster $k$, $\overline{M}$ is the
mean geographical location of all the users, $\|\mathcal{M}_k - \mathcal{M}\|$ is the Euclidean distance between the two vectors and $x_m$ is the user position of the user $m$. The constraint (1b) ensures that the number of beams do not exceed the total number of users. We choose Calinski-Harabasz (CH) criterion because it is one of the accurate ways to compute $K$ for clustering methods that use squared Euclidean distance. The larger the CH ratio, the better is the data partitioning and hence we solve (1a) to find any value of $K$ that maximizes Calinski-Harabasz index value and chose it as the optimal number of beams. As the number of users in a Dense urban-eMBB layout is small, we solve this problem using simple exhaustive search.

B. Clustering of distributed users

In the benchmark design, irrespective of change in the number of users in a beam/beam-demand, the position of the beams are fixed. Hence, the offered throughput remains relatively similar across all the beams. Therefore, such rigid plans will fail to distribute traffic demand across all the beams evenly and result in either under-use the offered throughput (beam capacity is unused) or overload the beam (beam demand is unmet). Also, a beam with very high demand suffers power limitations from system hardware. Furthermore, it is important to support the regulatory standards by not exceeding EMF radiations in beams with higher demand. Hence, a beam footprint has to be designed such that the occurrence of unbalanced aggregated-beam demand is avoided. To do so, we first need to consider the users in a euclidean sense and cluster them into sets of adjacent users. This clustering should distribute the total system demand evenly among all the sets. Then we find the Euclidean centroids of these clusters and direct the main antenna lobes of the $K$ beams at these centroids.

To do so, we consider metric spaces and endow universe $\mathcal{N}$ with a metric space $(\mathcal{X}, r)$ such that $\mathcal{N} \subseteq \mathcal{X}$, where $\mathcal{X}$ is a set of all points in a 2D Euclidean space and $r$ is a distance metric on $\mathcal{X}$. Then, we define, $D_k = \sum_{m \in \mathcal{G}_k} d_m$, as the demand of beam $k$ and, $D_{sys} = \sum_{m=1}^{M} d_m$, as system demand, where $d_m$ is the requested traffic demand of user $m$, and $\mathcal{G}^k$ represents the clustered users belonging to beam $k$. Accordingly, we define the clustering problem as,

\[
\begin{align*}
\text{minimize} & \quad \sum_{k=1}^{K} \sum_{m \in \mathcal{G}^k} r(x_m, c_{G^k}) \left( \frac{D_k}{D_{sys}} \right) ; \\
\text{subject to} & \quad \bigcup_{k=1}^{K} \mathcal{G}^k = \mathcal{N}, \\
& \quad \mathcal{G}^i \cap \mathcal{G}^j = \emptyset, \forall i \neq j, \\
& \quad \mathcal{G}^i \neq \emptyset, \forall i, \\
& \quad c_{G^k} = \frac{1}{\sum_{m \in \mathcal{G}^k} d_m} \sum_{m \in \mathcal{G}^k} d_m x_m, \\
\end{align*}
\]

whose objective function is to minimise the weighted distance of each cluster member from the cluster center, where $r(x_m, c_{G^k}) = ((x_m - c_{G^k})(x_m - c_{G^k})')$ is the Euclidean distance between any user $m$ and the cluster center $c_{G^k}$. Weights are added to distribute the demand evenly among all the beams and are defined as the ratio of beam demand to the system demand.

The first constraint in (2b) ensures that all the users are under the coverage region. The second constraint in (2c) assures that any user will be served by only one beam. The third constraint mentioned in (2d) ensures that the beams have at least one user and to avoid planning beams with zero demand. In the constraint (2e), $c_{G^k}$ is two element vector in the 2D Euclidean space representing the weighted cluster centroid of the cluster $k$.

Distributing the total traffic demand across all the antenna beams evenly is not a straightforward solution because of the mobile nature of the users. Hence, we group the users into clusters and serve them under a beam. In other words, we partition the $M$ users into $K$ number of clusters $\{\mathcal{G}^1, \mathcal{G}^2...\mathcal{G}^K\}$ with $\{c_{G^1}, c_{G^2}...c_{G^K}\}$ as cluster centres with each cluster to be converted to a beam in a later stage. We use the same number $K$ for the number of clusters and beams, since each beam will serve one cluster.

Among the various available clustering algorithms such as k-means, k-medoids, Partitioning Around Medoids (PAM) and Clustering LARge Applications (CLARA) [15], we employ weighted k-means clustering using iterative Lloyd’s algorithm [16] approach. The weighted version of the k-means considers both user geographical positions and demand requirements and hence provides better demand distribution among all the beams. Also, Lloyd’s algorithm is known as a centroid tessellation, which is beneficial for the beam design, since the beam center is likely to point in the direction of the dominant group of users. Hence, the demand can be better satisfied. The steps of Lloyd’s Iteration Partition Clustering is shown in the Algorithm 1.

In Step-1 of the Algorithm 1, we chose the cluster seeds or the initial $K$ cluster centres using the k-means++ algorithm [17] for faster computation which converges better than random seeding. In Step-2, we compute distance matrix $(\mathcal{R}_{K \times M})$ using the Distance Metric (using a Weighted Euclidean Distances (WED) expressed as,

\[
\text{WED}(x_m, c_{G^k}) = ((x_m - c_{G^k})(x_m - c_{G^k})') \left( \frac{\sum_{m \in G^k} (d_m)}{\sum_{m=1}^{M} (d_m)} \right),
\]

where $d_m$ is the broadband traffic demand of any user $n.)$ between all the users and the cluster centres. Then using the distance matrix $\mathcal{R}_{K \times M}$, we group the users into clusters to their nearest cluster centre. The WED will ensure that the clustering is based not only on their
geographical location but also on their broadband traffic demand. In Step 4, we compute a weighted version of the mean user position as a cluster centre $c_{G^k}$ using,

$$c_{G^k} = \frac{1}{\sum_{m \in G^k} d_m} \sum_{m \in G^k} d_m x_m.$$  

(4)

The Step-2 to Step-4 are repeated in Lloyd’s iteration fashion for better clustering. The algorithm stops either when the cluster assignments do not change or when the maximum number of iterations are reached. Upon the termination of the algorithm, all the broadband users will be grouped into $K$ clusters $\{G^1, G^2, ..., G^K\}$.

**Algorithm 1** Loyd’s Iteration Partition Clustering Algorithm

**procedure** CLUSTERING($K, X, d, D, C_s, DM, M$) 

1. Total number of beams, 
2. Broadband user set, 
3. User demand in Mbps, 
4. Initial seeds for cluster centres, 
5. Distance Metric, 
6. distance matrix, 
7. Maximum number of iterations 
8. [Step 1] Choose initial cluster centres/ seeds $\{c_{G^1}, c_{G^2}, ..., c_{G^K}\}$ as per $k$-means $+$ Algorithm. 
9. while (Cluster assignments do not change) OR (I is not reached) do 
10. [Step 2] Compute distance $R_{K \times M}$ between each of $\{c_{G^1}, c_{G^2}, ..., c_{G^K}\}$ and all of $\{x_1, x_2, ..., x_M\}$ using $DM$ shown in (3). 
11. [Step 3] Assign $\{x_1, x_2, ..., x_M\}$ users to $K$ clusters $\{G^1, G^2, ..., G^K\}$ based on the minimum distance between the users and cluster centre using $R_{K \times M}$. 
12. [Step 4] Compute new cluster centres $\{c_{G^1}, c_{G^2}, ..., c_{G^K}\}$ by using (4). 
13. end while 
14. end procedure

C. Elevation and Azimuth angle computation

In this step, we express the previously obtained cluster centres using Azimuth and Elevation angle. Azimuth angle can be defined using the Figure 2 as a sector angle. In the benchmark, for three beam scenario, this angle is fixed and 120 degrees apart. The Elevation angle can be defined using the Figure 2 as a tilt angle and is again fixed in the benchmark.

However, in the proposed case, the adaptive Azimuth angle can be obtained from the cluster center position using $\theta' = \arcsin(\frac{OL}{OC})$. The value $\theta'$ takes the value Azimuth angle $\theta$ differently in different quadrants. This can be understood using the Figure 3 where $OL$ is the distance between the latitude (in degrees) of the macro base station and the latitude (in degrees) of the cluster center. The distance $OC$ is the distance between the center of macro base station and center of the cluster.

The adaptive Elevation angle can be obtained from Figure 4 using $\phi' = \arctan(OC/h)$, and $\phi = 180 - \phi'$, where $h$ is the height of the macro base station (25 meters suggested in [5]). Thus the obtained $\theta_k$ and $\psi_k$ could be used to position a beam $k$.

IV. Simulation and Result Analysis

A. Configuration Parameters

We consider a dense Urban-eMBB environment with the base station at Interdisciplinary Centre for Security, Reliability and Trust (SnT) (49.6276-degree latitude and 6.1594-degree longitude). We define the coverage region of 100 meters around the base station. We consider 15 random users in the coverage region and demand of the users varies from 15 mbps to 35 mbps.

![Azimuth angle and Elevation angle in Dense urban-eMBB layout](image)

![Azimuth angle computation](image)

![Elevation angle computation](image)
B. Demand Distribution

The beam demand $D_k$ (in bps) of a beam $k$ is the summation of traffic demand of all the users belonging to the beam $k$. The total system demand (in bps) is the summation of all the beam demand that has to be met by macro base station.

Using the proposed ABF plan, the optimum number of beams $K$, changes in every iteration. However, to have a fair comparison of demand distribution between the benchmark FBF plan and the proposed ABF plan, we use $K = 8$. Figure 5 is the benchmark Fixed Beam Footprint design where the 8 beams are placed equidistant from each other. For any uneven distribution of users (denoted by red dots), it can be noticed that certain beams are overloaded while some are unused. The beams 1 and 5 does not have any users (zero demand to be met) and hence the beam capacity is unused while the beam 4 and 8, have relatively more number of users and hence overloaded. Figure 6 is the proposed Adaptive Beam Footprint design where the 8 beams are directed towards the cluster centroids. It can be seen that we have no beams with zero demand and also the users are more evenly distributed among all the beams.

The main objective of this work is to distribute $D_{sys}$ evenly among $K$ beams. To verify this, we have to compute beam demands $D_k \forall k$ for both benchmark Fixed design and proposed Adaptive design. We run the simulation for 1000 iterations for evaluation.

We can see in Figure 7 that the Fixed Footprint Design distributes the demand most unevenly. We also simulated the experiment for an adaptive design using only k-means algorithm which considers only the geographical locations of the user for Adaptive Beam Design. We can see that Adaptive design using only k-means performs better than the Fixed footprint design. However, the proposed Adaptive Beam Footprint design, which considers both user location and demand has clearly reduced beam demand for beams with high demands and distributed it to the beams with relatively lower demand. In other words, the proposed design distributes the total system demand more evenly and ensures that beams are always assigned with an adequate beam demand that have to be met. Furthermore, it ensures that no beam is planned with zero demand. This is majorly because in rigid fixed plan, the geographical user locations and their traffic demand is ignored and hence some beams are pointed to geographic locations where no users are present.

The Jain’s Fairness Index ($\mathcal{J}$) [18] is a well-known fairness metric, which in this context measures how evenly the demand is distributed across all the beams and is defined as, $\mathcal{J} = (\sum_{k=1}^{K} D_k)^2/(K \sum_{k=1}^{K} D_k^2)$. The value ranges between $\frac{1}{K}$ and 1, where $\frac{1}{K}$ signifies that the system is least fair and 1 signifies that the system is most fair. The Jain’s fairness index is computed for both benchmark fixed plan and proposed adaptive design for 1000 iterations and as shown in Figure 8. It is evident that the proposed adaptive design performs better with higher value in comparison to the benchmark fixed design for all iterations.

C. Mean Absolute Error

Mean Absolute Error (MAE) is a measure of errors between ideal demand distribution and observed demand distribution and is expressed as, $MAE = \frac{1}{K} \sum_{k=1}^{K} |D_k - \frac{D_{sys}}{K}|$. The MAE values for both benchmark FBF design and proposed ABF design is computed for 1000 iterations and a cumulative distribution function is plotted as shown in Figure 9. The CDF plots shows...
that the error between the ideal demand distribution and observed demand distribution is less in the proposed ABF design than the benchmark FBF design.

V. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel Adaptive Beam Footprint (ABF) design where, instead of steering the beams to fixed locations, the macro base stations steer the beams adaptively to the positions proposed by ABF design. The proposed ABF design is highly successful in distributing the traffic demand evenly among all the beams and hence avoiding beam overloading or underuse of a beam. The algorithm has been simulated in a Report ITU-R proposed Dense 5G urban-eMBB test environment with mobile pedestrian and vehicular users whose user density and traffic loads vary spatio-temporally. Numerical simulations conducted shows the benefit of using the ABF design instead of FBF design. A possible extension of this work would be to design beamforming vectors to generate directional beams for the proposed adaptive footprint and furthermore, verify the user satisfaction by computing the offered capacity to the users.

REFERENCES

[1] H. Kawakami and T. Ohira, “Electrically steerable passive array radiator (espar) antennas,” IEEE Antennas and Propagation Magazine, vol. 47, no. 2, pp. 48–50, 2005.

[2] I. Uchendu and J. R. Kelly, “Survey of beam steering techniques available for millimeter wave applications,” Progress In Electromagnetics Research B, vol. 68, pp. 35–54, 2016. [Online]. Available: https://doi.org/10.2528/PIERB16030703

[3] Y. Ojio, H. Kawakami, K. Gyoda, and T. Ohira, “Improvement of elevation directivity for ESPAR antennas with finite ground plane,” in IEEE Antennas and Propagation Society International Symposium. 2001 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting (Cat. No.01CH37229). IEEE. [Online]. Available: https://doi.org/10.1109/APS.2001.95938

[4] C. Deng, Y. Li, Z. Zhang, and Z. Feng, “A hemispherical 3-d null steering antenna for circular polarization,” IEEE Antennas and Wireless Propagation Letters, vol. 14, pp. 803–806, 2015. [Online]. Available: https://doi.org/10.1109/lawp.2014.2382107

[5] ITU, “Guidelines for evaluation of radio interface technologies for IMT-2020,” International Telecommunication Union (ITU), Technical Report Version Report ITU-R M.2412-0, 10 2017. [Online]. Available: https://www.itu.int/dms_pub/itu-t/rep/ituf-1542017-PDF-E.pdf

[6] V. Barousis, A. G. Kanatas, N. D. Skentos, and A. Kalis, “Pattern diversity for single rf user terminals in multiuser environments,” IEEE Communications Letters, vol. 14, no. 2, pp. 151–153, 2010.

[7] H. Yazdani, A. Vosoughi, and N. Rahnavard, “Compressive sensing based direction-of-arrival estimation using weighted greedy block coordinate descent algorithm for espar antennas,” in MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM), 2017, pp. 169–173.

[8] A. Wang, D. Tripplet, and C. Stevens, “Broadband/multiband conformal circular beam-steering array,” in IEEE MTT-S International Microwave Symposium Digest, 2005., 2005, pp. 1883–1886.

[9] T. Iguchi, D. G. Mixon, J. Peterson, and S. Villar, “On the tightness of an SDP relaxation of k-means,” CoRR, vol. abs/1505.04778, 2015. [Online]. Available: http://arxiv.org/abs/1505.04778

[10] Q. Du, V. Faber, and M. Gunzburger, “Centroidal voronoi tessellations: Applications and algorithms,” SIAM Review, vol. 41, no. 4, pp. 637–676, Jan. 1999. [Online]. Available: https://doi.org/10.1137/s0036144593539552

[11] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 50, no. 2, pp. 73–102, 1988.

[12] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 2, pp. 411–423, 2001. [Online]. Available: https://doi.org/10.1111/1467-9868.00293

[13] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[14] T. Calinski and J. Harabasz, “A dendrite method for fair assessment and discrimination for resource allocation in shared computer systems,” in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA ’07. USA: Society for Industrial and Applied Mathematics, 2007, p. 1027–1035.

[15] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” in Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA ’07. USA: Society for Industrial and Applied Mathematics, 2007, p. 1027–1035.

[16] S. Lloyd, “Least squares quantization in PCM,” IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, Mar. 1982. [Online]. Available: https://doi.org/10.1109/tit.1982.1056489

[17] D. Arthur and S. Vassilvitskii, “K-means+: The advantages of careful seeding,” in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA ’07. USA: Society for Industrial and Applied Mathematics, 2007, p. 1027–1035.