SOURCE LOCALIZATION FOR MOLECULAR COMMUNICATION VIA DIFFUSION

by

Oğuzhan Yetimoğlu

B.S., Computer Engineering, Boğaziçi University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering
Boğaziçi University
2022

ACKNOWLEDGEMENTS

Words cannot express my gratitude to my professors for their invaluable patience and feedback. I must firstly thank my supervisors. Prof. Tuna Tuğcu and Assist. Prof. Hüseyin Birkan Yılmaz really do deserve the greatest of thanks, since they have provided me with incredible support, encouragement and guidance in both the writing of this thesis and the work which preceded it.

I wish to show my appreciation to all my professors in the computer engineering department, and my professors and colleagues in Nanonetworking Research Group. They supported me with their help, knowledge and experience throughout my graduate education.

I owe so much thanks to my mother Sevcan Yetimoğlu, my father Halil Yetimoğlu, and my sister Zeynep Simay Yetimoğlu who were continuously supporting me throughout my life. Their belief in me has kept my spirits and motivation high during this process.

I would like to thank Abdurrahman Dilmaç for all the support and encouragement he give me, and the patience and unwavering faith he has in me. He believed in me even when I don't believe in myself. For that I am eternally grateful.

Away from work, special thanks go to my best friends Onur Deniz, Hasan Ali Düzağaç, and İbrahim Sina Aşık for both of whom friendships goes as far as editing.

This research was partially supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under grant number (1190049) andby the Turkish Directorate of Strategy and Budget under the TAM Project number 2007K12-873.

ABSTRACT

SOURCE LOCALIZATION FOR MOLECULAR COMMUNICATION VIA DIFFUSION

Molecular communication is a type of communication that provides communication in mediums such as underwater, where traditional communication paradigms are insufficient, and in which information is carried by molecules in fluid environments. However, the fact that the information particles in molecular communication propagates by diffusion prevents us from solving problems such as finding the location of the transmitter with traditional methods. Especially, when the number of transmitters is more than one, the solution becomes more complicated. The solution described in this thesis is to find the locations of the transmitters by employing the coordinates of the molecules hitting the receiver in a fluid environment where there are multiple transmitters and a spherical receiver that absorbs the hitting molecules completely. For this localization solution, first the coordinates of the hitting molecules are clustered with models such as Gaussian, Bayesian mixture models, and K-means. By calculating the average values of the clustered data, the direction of the corresponding transmitter is determined. At the same time, the distance is determined based on the number of data belonging to separate clusters and the probability of the particles hitting the receiver. The results show that the most promising cluster algorithm is K-Means. By calculating the direction and the distance of the locations via clustered data, we can estimate the transmitter locations.

ÖZET

DİFÜZYON İLE MOLEKÜLER İLETİŞİM İÇİN LOKALİZASYON

Moleküler iletişim, geleneksel iletişim paradigmalarının yetersiz olduğu sualtı gibi ortamlarda iletişimi sağlayabilecek, akışkan ortamlarda bilginin moleküllerle taşındığı bir iletişim paradigması türüdür. Öte yandan moleküler iletişim moleküllerin difüzyonu ile gerçekleşmesi, vericinin yerini bulma gibi problemlerin çözümünü geleneksel yöntemlerle çözebilmemizi engeller. Özellikle verici sayısının birden fazla olduğu durumlarda çözüm daha karışık hale gelmektedir. Bu tezde anlatılan çözüm ise çoklu vericilerin ve küresel, üzerine çarpan molekülleri tamamen emen bir alıcının olduğu akışkan bir ortamda alıcıya çarpan moleküllerin koordinatlarını kullanarak vericilerin yerlerini bulacak bir çözüm önerilmiştir. Bu yer bulma çözümü için, önce çarpan moleküllerin koordinatları K-ortalama, Gauss ve Bayes karışım modelleri gibi kümeleme modelleri ile kümelenir. Kümelenen verilerin ortalama değerleri hesaplanarak her kümeye ait ilgili vericinin yönü tespit edilir. Aynı zamanda herbir kümeye ait verinin sayısı, ve moleküllerin alıcıya çarpma olasılığından yola çıkarak mesafe tespit edilir. Sonuçlar, en umut verici küme algoritmasının K-Means olduğunu göstermektedir. Lokasyonların yönünü ve mesafesini kümelenmiş verilerle hesaplayarak verici konumlarını tahmin edebiliriz.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

D	Diffusion Coefficient
$\mathbf{d}_k$	Direction vector from center of Rx to k-th Tx
$\hat{\mathbf{d}}_k$	Estimated direction vector from center of Rx to k-th Tx
$\mathrm{erfc}(\cdot)$	Complementary Error Function
$F_{hit}(\cdot,\cdot,\cdot)$	Function of hitting molecules
$\hat{F}_{hit}(\cdot)$	Estimation of function of hitting molecules
K	Number of Tx
k	Tx variable
N	Number of molecules per Tx
$\mathcal{N}(\cdot,\cdot)$	Gaussian distribution
n	Molecule variable
r_r	Radius of Rx
r_0	Distance between center of Rx and Tx
r_{0k}	Center of Rx
t	Time
Δt	Time interval
t_{end}	Simulation end time
$\mathcal{L}$	Locations of particles
$\mathcal{T}$	Sampling time of hitting function
$W(\cdot)$	Wiener process
$\mathbb{W}$	Probability matrix
w_{kj}	Probability of particle j that belongs to k-th cluster
$\lvert\cdot\rvert$	Cardinality operator
$\Gamma(\cdot)$	Clustering function
$\delta_k(\cdot)$	Function of location estimation
ϵ_d	Average percent of Euclidean errors
ϵ_ϕ	Average angle error

μ	Mean
$\Delta\xi_n$	Change of coordinates in 3D
σ	Standard deviation
Ψ_k	Location of $k-$th transmitter
$\hat{\Psi}_k$	Estimated location of $k-$th transmitter

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three Dimensional
DiM	Dirichlet Mixture Model
DoA	Direction of Arrival
ISI	Inter-symbol interference
GMM	Gaussian Mixture Model
MC	Molecular Communications
MCvD	Molecular Communications via Diffusion
ML	Machine Learning
MLE	Maximum Likelihood Estimation
RTT	Round Trip Time
Rx	Receiver Node
Tx	Transmitter Node

1. INTRODUCTION

Molecular communications (MC) is promising, small-scale technology that enables molecules as an information carriers in fluidic medium. Molecular communication appears as a credible alternative in scenarios when conventional communication systems are unsuitable owing to the fluidic propagation environment and difficulties in the design of in-body antennas. Additionally, significant advancements have been made in other areas of MC use, including industry, the environmental issues, genetics engineering, and communication technology. Also, targeted medication delivery [1–3], nanorobots, pathogen transmission [4, 5] and cancer research are just a few examples of the micro-scale research that has significantly risen in the healthcare sector in recent years [6]. As a result, MC has gained a great deal of attention in literature related to science and engineering [7].

The term "molecular communication via diffusion" (MCvD) refers to an MC system in which chemicals move through diffusion 1.1. The information carriers in this systems are molecules, which propagate by random walk of individual particles carried by the thermal energy in the system. Cell-to-cell communication—a process that is already employed in the nature—is an example of MCvD [8]. Paracrine signaling is one method of inter-cellular communication (i.e., the signaling between nearby cells) in the human body.

1.1. Source Localization via Diffusion

The need and potential use cases for MC grow as applications drive communication networks to smaller. As a result, MC networks have recently received a lot of attention. The fact that molecular receivers in MCvD systems receive a significant amount of diffusion noise together with a heavy tail signal (inter-symbol interference [ISI]) is one of the key issues. When constructing an MCvD system with right adjusted parameters, information about transmitter location is essential to reducing noise in

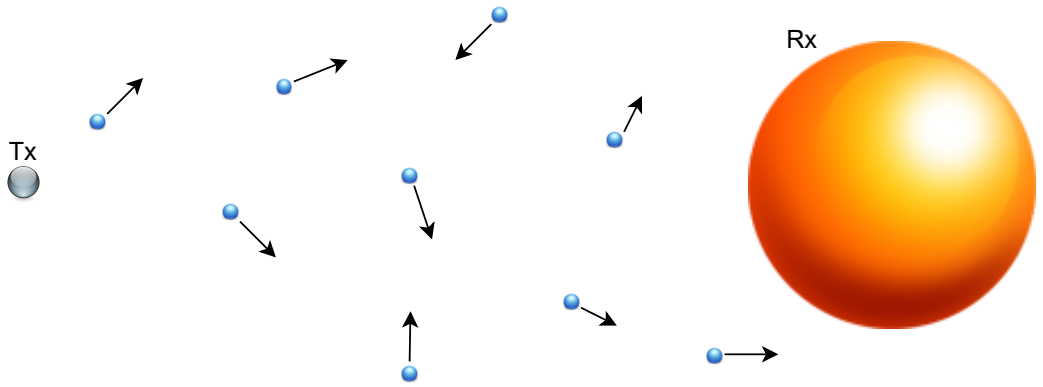


Figure 1.1. An illustration of molecular communications.

the received signal. Transmitter localization's another promising use case is in medical area (healthcare), where unhealthy cells can act as transmitters. Unhealthy cells can be localized to allow the required post-treatment procedures by following the chemicals emitted by such entities.

1.2. Related Work

Wang et al. in [9] present an algorithmic distance estimation scheme with fully absorbing receiver. Receiver estimates the distance between the transmitter and itself via counting the number of particles within a predefined time frame. In [10], the proposal is transmitter positioning method for MCvD in the system that includes a single transmitter and multiple spherical absorption receivers. It utilizes Levenberg-Marquardt method for estimating distances between the transmitter and every receiver.

In [11], the researchers use Poiseuille flow and ring-shaped observation receivers to analytically calculate the location information of a point transmitter in a vessel-like medium. The goal of this work is locating anomalous and diseased cells by the particles they release. Based on the assumption that the MC system is perfectly synchronized, two scenarios are taken into consideration. It has been demonstrated that in the situation of perfect synchronization, high-quality localization estimation is feasible using just one ring-shaped observation receiver. In the same study, a different

approach using two ring-shaped observation receivers is also suggested for the scenario where the receiver is not synchronized with the transmitter.

In [12], using signal-dependent noise and ISI, the position of only one point transmitter with an transparent 3D spherical receiver is determined analytically. Both scenarios, where the positions of the observing receiver are either known or unknown, result in the proper localization of the point transmitter.

In [13, 14], experiment-based distance estimation is investigated in macro-scale environments. In [13], various methods including data analysis based and machine learning techniques are utilized. In [14], a localization algorithm based on sensors is presented. In this paper, ethanol molecules are evaporated from a petri dish at 23°C and employed as information carriers. Cluster groups are formed from 24 sensors, and some sensors are allowed to participate in more than one cluster group. For each cluster, localization is estimated using data collected from sensors and an estimation of wind velocity.

In [15], two estimation techniques for the distance between the transmitter and the receiver are investigated for the scenario with only one transmitter. The first technique estimates distance based on the knowledge that the molecule concentration achieves a maximum (i.e., via peak time). The second technique makes use of signal energy, also known as the sum of molecule concentration. The authors conclude that utilizing signal energy for distance prediction is preferable to using peak time for estimation. Nevertheless, this approach has increased complexity higher than the former. In [16], channel characteristics diffusion coefficient, including propagation distance, and medium co-flow velocity are also evaluated. The same work also analytically derives closed-form formulations of the lower limits of Cramer-Rao.

In [11,15,17–21], estimation is performed by employing peak of the signal. Luo et al. in [20] investigate a novel effective distance estimation method, in which receiver can localize transmitter via information of peak values of two different molecules (different

diffusivity) emitted by a transmitter. Similar distance estimation technique (different molecules with different diffusivity) is assigned in [21].

In [18], researchers propose a method to estimate the distance between bionanomachine pairs that both transmit and receive. To estimate the distance, one of the bionanomachine pairs emits A molecules with coefficient D_A into the medium. When the other bionanomachine receives the A molecules, it reacts by emitting B molecules with coefficient D_B into the medium. By using detection times of peak concentration, diffusion coefficients, and round trip (RTT) time protocol, which is the proposed method by the researchers, the distance between bionanomachines is estimated. Another paper that uses RTT to predict distance is [21].

In [22], the propagation distance is estimated using a high-accuracy protocol. Multiple symbols are transmitted by the source in the form of molecules. Also, by using the Newton-Raphson formula depending on the molecule concentration, the receiver calculates the distance.

The authors of [23] use maximum likelihood estimation (MLE) to determine the parameters (diffusion coefficient, medium co-flow velocity, and distance of propagation) of the inverse Gaussian distributed channel. [12] demonstrates that localization via concentration of particles in the system with high accuracy in cases where single transmitter may be attained using an iterative MLE approach. In another paper that uses MLE to localize the transmitter [19], the estimation of distance is investigated in the context of a single transmitter scenario. The researchers obtain the Cramer-Rao lower bound for the error for distance estimation's variance. Also, the lower bound takes molecular degradation and a steady channel flow into account. In [24], distance is estimated based on the difference of dynamic environment. It means that not only molecules propagate, but also the transmitter and the receiver in the environment with a constant diffusion coefficient. To accomplish the aim, a novel two-step scheme based on MLE is proposed. Among localization techniques, MLE has higher accuracy. On the other hand, it needs more computational power.

Table 1.1. Survey table.

Category	**Method**	**Highlights**
Distance est.	Algorithmic distance estimation scheme [9]	Sync. and unsync. conditions
	Various ML methods [13]	Experiment in macroscale env.
	Concentration-peak time or received concentration energy [15]	
	MLE method [16]	Inverse Gaussian distributed channel
	Signal peak [17]	Only one-way transmission, no clock sync. between nanomachines
	Signal peak [18]	Single spike of molecules
	MLE method [19]	Cramer-Rao lower bound
	Signal peak [20, 21]	Two different types of molecules, sync. does not required
	MLE, Newton–Raphson method [22]	ISI considered
	MLE method [23]	Inverse Gaussian distributed channel
	Two-step scheme based on MLE [24]	Initial distance estimation, Tx and Rx moves
Location est.	Levenberg-Marquardt method and multi-point positioning method [10]	Single Tx and multiple spherical absorption Rx
	Signal peak [11]	Poiseuille flow and ring-shaped observation Rx
	Iterative maximum likelihood estimation [12]	Locations of the Rx are both known and unknown
	Sensor network-based localization algorithm [14]	Experiment in macroscale env., Gaussian plume model

In addition, two different techniques are provided in [17] to estimate the distance in a 1-D diffusion-based channel without clock synchronization. In the first technique, the transmitter and receiver's distance are determined using the concentration's peak value. In the second technique, the distance by modifying the transmission mechanism is estimated.

1.3. Contribution of This Thesis

Source localization problem has been researched in many ways for traditional communication paradigms and wide range of transmitter types such as radioactive, acoustic, radio-based, or optical. Also, it has been investigated deeply for multiple transmitter cases. However, in MC literature, researches are concentrated mostly on estimation of the distance instead of localization. Additionally, these are focused on single transmitter cases. In this theses, we proposed a solution for multiple transmitter localization. The main contributions of this thesis can be defined as follows:

- Proposing an approach to estimate each distance between an absorbing receiver and a corresponding transmitter by using the location coordinates of absorbed particles by the absorbing receiver.
- Proposing an approach to estimate each DoA for each transmitter by calculating the mean value of location coordinates of absorbed particles by the absorbing receiver.
- Evaluating the performances of different clustering algorithms with various error metrics based on the angle differences and the distance differences by investigating
    - the effects of the number of transmitters,
    - the effects of minimum angle between transmitter pairs,
    - the effects of distance variation,
    - the effects of quantization of the locations on the surface.

1.4. Thesis Outline

This thesis is outlined as follows. In Chapter 2, we introduce our method for transmitter localization in diffusive channels. First, we describe the system model, propagation model and channel model of the system of our study. Second, we propose our method for localization problem. In Chapter 3, we discuss the results of our method with alternative clustering techniques. In Chapter 4, we present a summary for the results and describe open questions with possible future works.

2. TRANSMITTER LOCALIZATION IN DIFFUSIVE CHANNELS

2.1. System Model

In this section, we focus on the micro-scale environment that includes a receiver and multiple transmitters. In such environments, the signal received by the absorbing transmitter is affected by signal emission process, signal absorption process, and the type of propagation (e.g., flow-based propagation, diffusion-based propagation).

2.1.1. System Topology

In this thesis, diffusion-based propagation is considered as a type of propagation. The medium is assumed to be 3D, unbounded, diffusion-based with no flow. Also, the topology consists of a single absorbing receiver (i.e., memorizes each emitted molecule) with spherical shape and multiple point transmitters, where the locations are random around the spherical receiver Figure 2.1. The transmitters emit same number of molecules synchronously.

Figure 2.1 demonstrates an example configuration of the system topology with and absorbing receiver, 2 point transmitters, and the molecules emitted from these transmitters. The point transmitters emit molecules from their locations. The distance between a location of transmitter k and the center of the absorbing sphere is r_{0k}. The radius of the absorbing receiver is r_r.

In the topology, we assume that there is no molecule-molecule interaction due to the narrow number of molecules in the medium. Also, we assume there is no source noise around receiver (closer than $13\,\mu m$) except diffusion noise. Because, in this

topology, the received signal drops significantly. after 13 μm. Additionally, we consider that the the point transmitters and the spherical receiver are perfectly synchronized (some of the works that use synchronized communication components in literature [25, 26]).

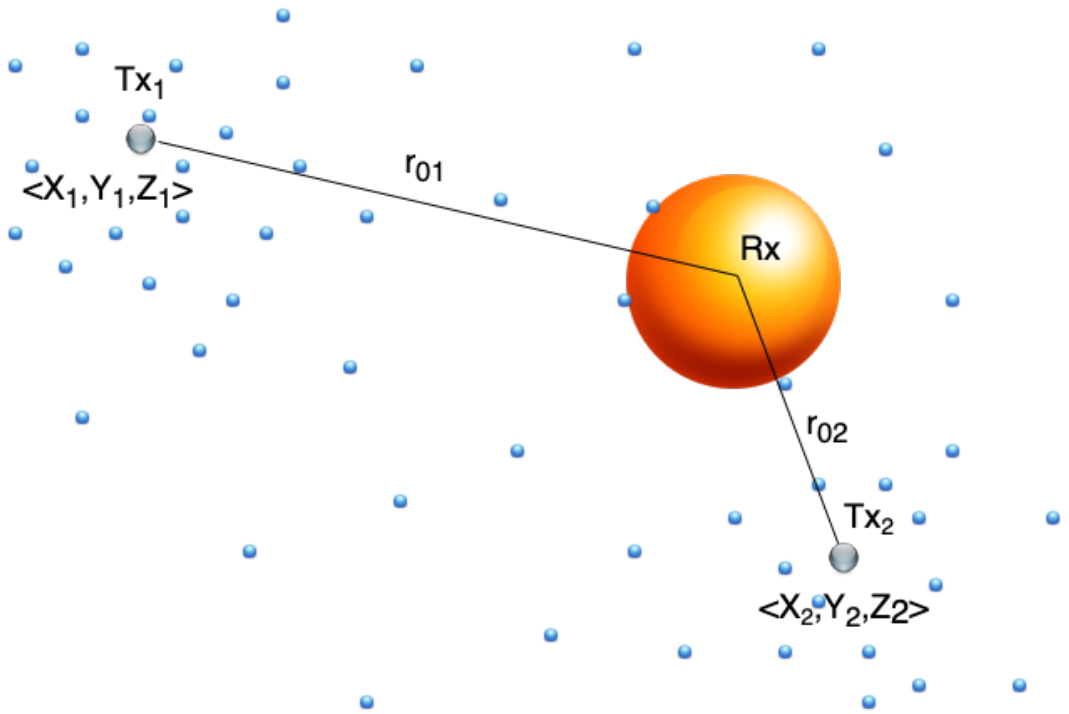


Figure 2.1. Two transmitter points, the propagating particles (emitted from transmitter points, some of them absorbed), and absorbing spherical receiver.

2.2. Propagation Model

Emitted molecules propagate randomly through the medium. This propagation can be modeled by Wiener process $W(t)$ [27] and it is defined by

$$\begin{aligned} &W(0) = 0, \\ &W(t) \text{ is almost surely continuous}, \\ &W(t) \text{ has independent increments}, \\ &W(t_t) - W(t_s) \sim \mathcal{N}(0,\, t_s - t_t), \text{ for } 0 \leq t_t \leq t_s, \end{aligned} \tag{2.1}$$

where $\mathcal{N}(\mu,\, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2. For each molecule, (2.1) explains that the probability of location change in one direction obeys Gaussian distribution. The location change at each dimension in 3D space, can be calculated by Δx, Δy, and Δz in time interval Δt, are given by

$$\Delta \xi_n = < \Delta x_n, \Delta y_n, \Delta z_n > \tag{2.2}$$

with

$$\begin{aligned} \Delta x_n &\sim \mathcal{N}(0, 2D\Delta t), \\ \Delta y_n &\sim \mathcal{N}(0, 2D\Delta t), \\ \Delta z_n &\sim \mathcal{N}(0, 2D\Delta t), \end{aligned} \tag{2.3}$$

where D denotes the diffusion coefficient of molecules in the environment. Moreover, the location change of each molecules is calculated by Gaussian distribution for each time interval Δt (the probability of the step size is determined by time interval).

2.3. Channel Model

In the literature, hitting distributions of the particles absorbed by fully absorbing receiver is well researched. In [28], cumulative distribution of hitting particles absorbed

by the receiver is analytically solved with respect to time (t), the distance between point transmitter and center of the absorbing receiver (r_0), and the radius of absorbing receiver (r_r) is given as,

$$F_{hit}(t, r_0, r_r) = \frac{r_r}{r_0}\text{erfc}\left(\frac{r_0 - r_r}{\sqrt{4Dt}}\right), \tag{2.4}$$

where erfc($\cdot$) and D denote the complementary error function and diffusion coefficient, respectively.

2.4. Multiple Transmitter Localization in MCvD

This thesis consists of a localization proposal with multiple point transmitter that emit same number of molecules at time t and an absorbing receiver in unbounded, 3D, diffusive, no flow fluidic environment. Additionally, the fluidic environment has a constant temperature and the viscosity.

We assume that the number of point-shaped transmitters in each scenario, the molecule numbers transmitted by point transmitters, and the time of transmission are known. It is also assumed that the diffusion coefficient of the fluidic environment (D) and the radius of the absorbing receiver (r_r) are constant and known.

It is challenging to locate the points of each transmitter by regarding to the locations of particles (denoted by $\mathcal{L}$) absorbed by the fully absorbing receiver. In order to estimate the location of the particles, we minimize the difference between each real ($\Psi_k = < X_k, Y_k, Z_k >$)and the predicted ($\hat{\Psi}_k = < \hat{X}_k, \hat{Y}_k, \hat{Z}_k >$) location of corresponding transmitter (k-th) pairs. This minimization is shown as,

$$\min_{\delta(\mathcal{L})} \sum_{k=1}^{K} \|\Psi_k - \hat{\Psi}_k\|, \tag{2.5}$$

where $\delta_k(\mathcal{L}) = \hat{\Psi}_k$ and $\delta(\cdot)$, and K are the function of location estimation (locations

of absorbed particles) and the number of point transmitter.

Each point transmitter emits same kind of particles from their coordinates. It causes absorbing receiver to absorb particles concurrently. So, first, we need to cluster (both soft and hard clustering) each absorbed particles ($\mathcal{L}$) into clusters, which correspond to one of point transmitters. $\Gamma(\cdot)$ (clustering function) finds the probability matrix ($\mathbb{W}$), which expresses that each data point belongs to corresponding transmitter. Note that, the sum of each row equals to 1 and also $\mathbb{W}$ is zero-one matrix for hard-decision clustering, while it is $(0,1)-$matrix for soft-decision clustering.

2.4.1. Mixture Models and K-Means for Clustering

In [29–34], the researchers study numerous methodologies for analyzing spherical data based on mixture models. In order to solve the localization of multiple transmitters problem in MCvD, we modified mixture model-based techniques. Popular clustering algorithms that employ soft-decision methods include the Gaussian mixture model (GMM) and the Dirichlet model (DiM). The probability that a data point belongs to corresponding cluster is determined using soft decision algorithms. In other words, these decision algorithms estimate the likelihood of each designation rather than rigidly identifying a data point. An iterative expectation-maximization techniques used in this work are both GMM and DiM. Based on the data gathered from the previous iteration, cluster probabilities associated to a data point are estimated in the expectation step.

In the maximization step, these probabilities are used to estimate parameters for the cluster distribution. GMM attempts to fit a set of (particularly, K) 3D Gaussian distributions to spherical data to cluster it. Gaussian distribution is used by the Dirichlet model as well, by utilizing a Dirichlet process. Other than these EM model-based algorithms, we used the K-Means algorithms to solve localization problem for multiple transmitters in MCvD. Moreover, K-Means algorithm uses a hard-decision logic to assign data points into clusters so that each data point is rigorously assigned to a single cluster (in our case, transmitter).

Pseudo-codes of the algorithms adapted in this thesis are presented in Figure 2.2 and Figure 2.3. In Figure 2.2, the logic behind EM-based models (soft decision) is shown with expectation and maximization steps. K-means algorithm is given in Figure 2.3.

```
Data: Hitting Molecules L
Result: Γ(L) = W
Initialization: means, covariance matrices, mixing coefficients of the clusters
while not converge do
    if Expectation Step then
        for each data point do
            calculate probability of each cluster using parameters of the corresponding
              cluster
    if Maximization Step then
        Update distribution parameters, i.e., the means, the covariance matrices and
          the mixing coefficients of the clusters from calculated probabilities
```

Figure 2.2. Expectation-maximization algorithm.

```
Data: Hitting Molecules L
Result: Γ(L) = W
Initialization: centroids of the clusters
while not converge do
    for each data point do
        assign the data point to the closest cluster
    Update centroids
```

Figure 2.3. K-means algorithm.

2.4.2. Direction and Distance Estimation

To estimate a location of a point, we need the direction and the distance estimation for this point, which DoA and r_{0k} for $k-th$ transmitter in our case. Hence, first, we need to cluster (via GMM, DiM and K-Means) hitting molecules ($\mathcal{L}$) properly to reach this valuable information. By doing so, we calculate the centers of each cluster, which is an estimation of DoA for corresponding transmitter. We mean the coordinates of each cluster individually for K-Means, due to the fact that it is a hard-decision algorithm. We use the F_{hit} function from (2.4) and the number of molecules received

to estimate r_{0k}.

To estimate the DoA for each transmitter, we mean the coordinates of each absorbed data points that belongs to corresponding cluster. To mean the data points for soft-decision clustering, we calculate average weighted sum of data points of a cluster with probabilities coming from expectation step. The estimated direction from the spherical receiver's center to the k-th transmitter ($\hat{\mathbf{d}}_k$) is calculated by taking mean of the location (with respect to each dimension) of the data points with the relevant w_{kj} as

$$\hat{\mathbf{d}}_k = \frac{\sum_{j=1}^{|\mathcal{L}|} w_{kj} \{< x_j, y_j, z_j > - < X_{Rx}, Y_{Rx}, Z_{Rx} >\}}{\sum_{j=1}^{|\mathcal{L}|} w_{kj}}, \tag{2.6}$$

where w_{kj} indicates the probability of particle j that belongs the k-th cluster (transmitter), $|\cdot|$ is the cardinality operator, and $< X_{Rx}, Y_{Rx}, Z_{Rx} >$ is the absorbing receiver center, respectively. Besides, 2.6 is also usable for K-Means. Due to K-Means algorithm is a hard-decision algorithm for clustering, the weights (w_{kj}) take a value either 0 or 1. When K-Means predict that $n-th$ particle belongs to $k-th$ transmitter, w_{kj} will take 1, otherwise 0.

To find the distance of $k-$th transmitter (r_{0k}), the number of the data points that belong to cluster r_{0k} with time steps and equation 2.4 can bu utilized. First, the coordinates of the data points are discretizated cumulatively with respect to their absorption time by absorbing receiver. Next, w_{kj}, which is provided by clustering algorithms, yields the estimated number of particles until time t_s by

$$N_k^{t_s} = \sum_{j \in \mathcal{L}_{t_s}} w_{kj}, \tag{2.7}$$

where $\mathcal{L}_{t_s}$ denotes the particles absorbed until time t_s. The calculated values of $N_k^{t_s}$ are used for estimating F_{hit} in (2.4) such that

$$\hat{F}_{hit}^k(t_s) = \frac{N_k^{t_s}}{N}. \tag{2.8}$$

At the right side of (2.4), both parameters D and r_r are already known. Next step is

to estimate expected distance of transmitter point k as

$$r_{0k} = \underset{r_0}{\operatorname{argmin}} \sum_{t_s \in \mathcal{T}} \left\{ F_{hit}(t_s, r_0, r_r) - \hat{F}^k_{hit}(t_s)) \right\}^2 , \tag{2.9}$$

where $\mathcal{T}$ denotes the hitting function's sampling time.

3. RESULTS

3.1. Performance Evaluation

Using a particle-based simulator, the coordinates of hitting molecules ($\mathcal{L}$) are generated. The particles in the simulator propagate via diffusion by computing their 3D displacements in accordance with (2.2) at each simulation time step. Later, applying the algorithms GMM, DiM, and K-Means in the machine learning library scikit-learn [35], point transmitter clusters are identified from data made up of hitting molecules.

Depending on the quantity of transmitters, we examine a variety of different scenarios. From each transmitter point, 10^4 molecules are released into the environment. Other parameters are selected as $10^{-5}s$ and $79.4\frac{\mu m^2}{s}$ for time step Δt and diffusion coefficient D, respectively. Moreover, estimations are obtained by averaging 500 repetitions. The radius of the receiver (r_r) is $5\mu m$, and the transmitters are initialized in random near the receiver with a distance (from the center of the absorbing receiver to a point transmitter) between $8\mu m$ and $12\mu m$ ($r_0 \in [8, 12]$). The distance range of signals is what led to these values being determined for the receiver's radius and distances. An upper bound for the distance between a transmitter and a receiver in an MC network is computed so that the bit error rate stays under 10^{-3}. Expressly, the the range of distance is chosen to maintain adequate levels of communication quality. Furthermore, the direction information becomes considerably less reliable the further apart the transmitter and receiver are. Hence, when the signal absorbed by receiver is poor, it is difficult to precisely identify the transmitter's location. Additionally, additional factors besides distance values are frequently employed for simulations in MC studies. The reader is encouraged to refer to [11, 36] for such examples.

For the sake of comparison, the percentage Euclidean error is calculated by di-

Table 3.1. Simulation parameters.

Parameter	**Symbol**	**Value**
Radius of Rx	r_r	$5\mu m$
Diffusion coefficient	D	$79.4\frac{\mu m^2}{s}$
No. of released molecules per Tx	N	10^4
Distance btw centers of Rx and Tx	r_0	$8 \sim 12\mu m$
Number of Tx	K	2, 3, 4, 5, 6
Simulation End Time	t_{end}	$0.5s$
Time step	Δt	$10^{-5}s$

Table 3.2. Average percentage Euclidean errors of the algorithms.

Number of Tx	**GMM**	**DiM**	**K-Means**
2	23.58%	21.83%	8.51%
3	25.11%	22.93%	14.74%
4	29.29%	28.98%	19.70%
5	33.64%	33.85%	24.04%
6	36.98%	37.20%	28.67%

viding the Euclidean error of our prediction for each transmitter by the actual distance between the corresponding transmitter and the center of the receiver. Then, for each simulation, the average percent of Euclidean errors (ϵ_d) is calculated as

$$\epsilon_d = \frac{1}{K}\sum_{k=1}^{K}\frac{\|\Psi_k - \hat{\Psi}_k\|}{r_{0k}}. \tag{3.1}$$

Similar to that, the average angle error (ϵ_ϕ) may be calculated as follows

$$\epsilon_\phi = \frac{1}{K}\sum_{k=1}^{K}\cos^{-1}\left(\frac{\mathbf{d}_k \cdot \hat{\mathbf{d}}_k}{\|\mathbf{d}_k\|\,\|\hat{\mathbf{d}}_k\|}\right) \tag{3.2}$$

where $\mathbf{d}_k$ is the direction vector of the grand truth of transmitter location, and $\hat{\mathbf{d}}_k$ is the direction vector of the estimated transmitter location. ϵ_ϕ denotes the angular

Table 3.3. Average angle errors of the algorithms.

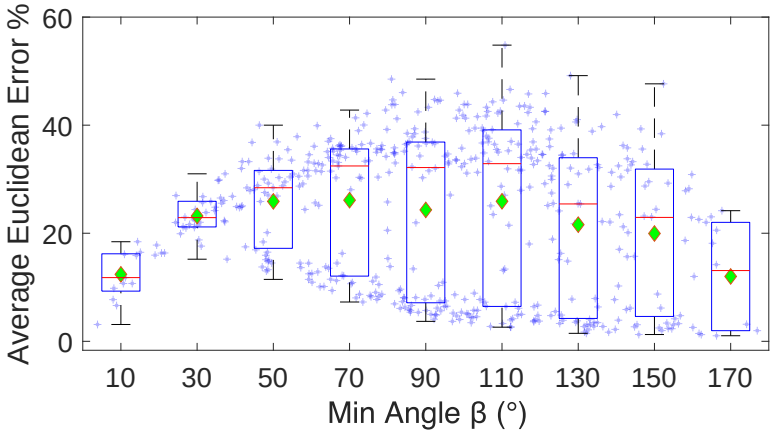
Number of Tx	**GMM**	**DiM**	**K-Means**
2	9.80°	9.14°	4.70°
3	10.75°	9.95°	8.24°
4	14.96°	15.18°	11.06°
5	18.10°	18.65°	13.56°
6	20.51°	20.96°	16.14°

difference regarding to the center of the spherical-shaped receiver between the actual (grand truth) and the locations predicted by the algorithms.

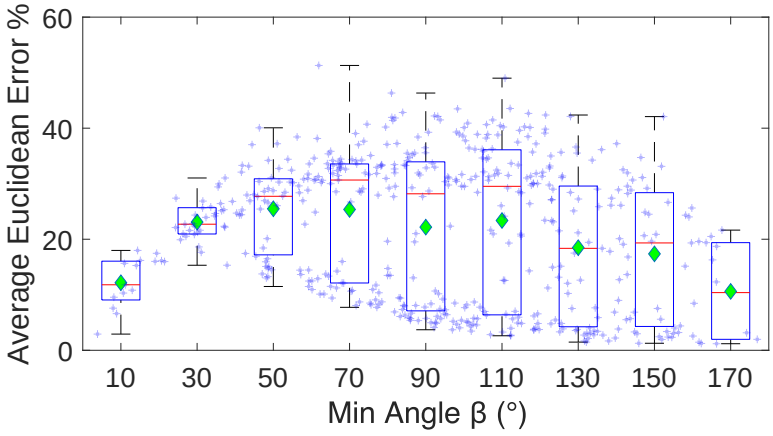
3.1.1. Effects of the Number of Transmitters

If a comparatively smaller intersection among clusters are randomly initialized, each technique outlined above can differentiate the clusters on absorbed particles (spherical data) achieved from the locations of the points of particles absorbed more easily. Additionally, particular to the soft-decision algorithms, it is likely that they have a tendency to categorize a collection of absorption points belonging to multiple transmitters as a single cluster, especially if the grouping resembles a cluster with a pleasing distribution or structure. As a result, the clustering becomes complicated, and consequently, there is a positive correlation between the number of point transmitters (and the amount of intersection among actual clusters) and the estimation errors of the algorithms. Mean percentage Euclidean errors and mean angle differences between the actual location of a transmitter point and its predicted location for various number of transmitters and various algorithms are showed in Table 3.2 and Table 3.3, respectively.

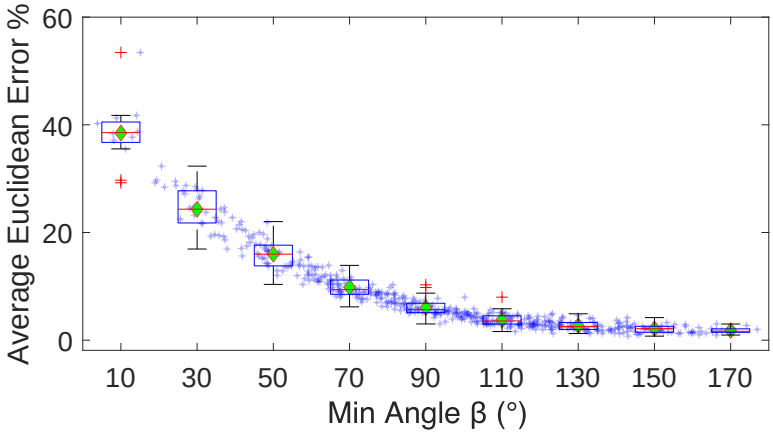
K-Means provides the best accurate transmitter position estimations across all scenarios of all algorithms considered. The distribution of the hitting molecules' coordinates may be to responsible for this. Using the presumption that the input data set contains a mixture of Gaussian distributions, GMM and DiM attempt to fit data



(a) Gaussian Mixture Model



(b) Dirichlet Mixture Model



(c) K-Means Model

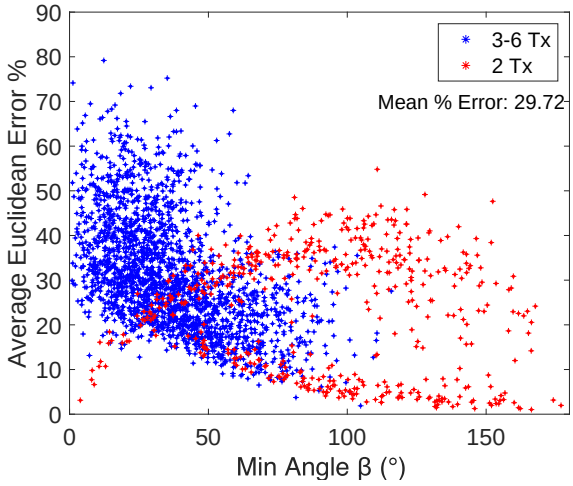
Figure 3.1. Average percentage of Euclidean errors for 2-transmitter scenarios ($K = 2$) regarding to the minimum angle β.

points to clusters. However, in practice, the spherical receiver's distribution of the coordinates of molecules that have been hit differs from a Gaussian distribution. The authors in [36] calculate the appropriate distribution analytically with various assumptions. The aim of the K-Means algorithm is to create clusters by taking into account the distance between 3D data points, as stated in Figure 2.3.

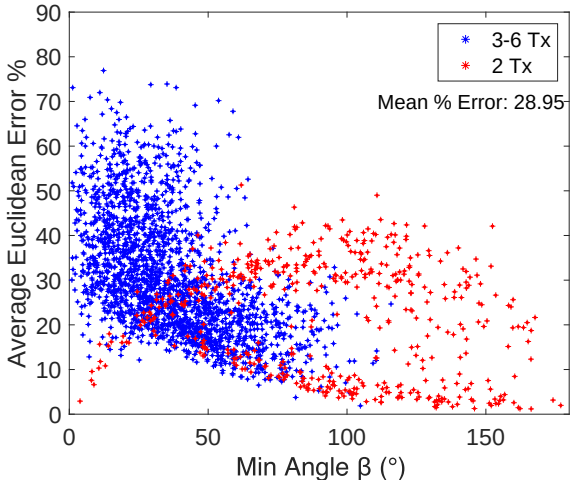
3.1.2. Effects of Minimum Angle Between Transmitter Pairs

Out of the K transmitters, we examined the impact of the minimum angle (β) between every binary combination of transmitters. Differentiation of the clusters considers the molecules released from two transmitters with a narrow angle (small difference) with regard to the center of the receiver being too close angularly. Therefore, it is expected that estimation errors and angle differences have a negative correlation. However, when the 2-transmitter scenario is examined, this is found only for K-Means, as shown in Figure 3.1. in Figure 3.1, each simulation has a point for the average percentage of Euclidean error on the figure. The box plots on the figures demonstrate statistical features of Euclidean error in 20-degree intervals between 0° and 180°. Red lines and green diamond points denote the median and mean values of each interval, respectively. The top and bottom of each box indicate data in between 25-th and 75-th percentiles for each interval. Additionally, red dots with plus shapes and bars highlight extreme points and outliers. Through all cases, the standard deviations of average Euclidean errors are 13.50%, 12.45%, and 8.12% for GMM, DiM, and K-Means, respectively.. The angle difference between the point transmitters may be thought of as being almost insignificant to soft decision algorithms. Even though GMM and DiM may be valuable up to around 30° minimum angle difference when the absorbing receiver assumes there to be two transmitters, K-Means are often preferred.

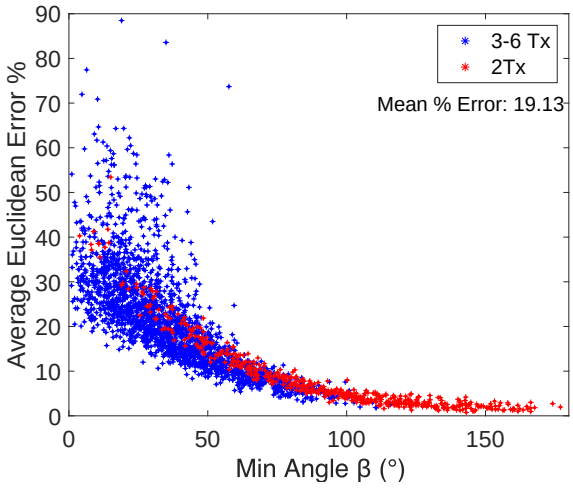
Each binary combination of transmitter creates a distinct angle difference in a scenario when there are more than two transmitters. We group these examples together based on the smallest angle difference between the transmitters to explore how angle differences influence estimation error. In Figure 3.2, the impact of β on the



(a) Gaussian Mixture Model



(b) Dirichlet Mixture Model

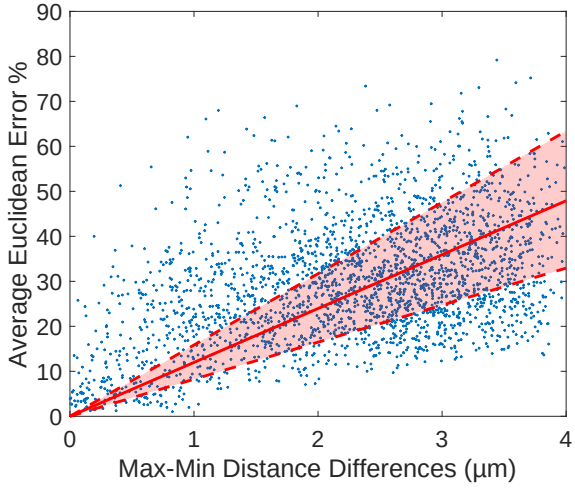


(c) K-Means Model

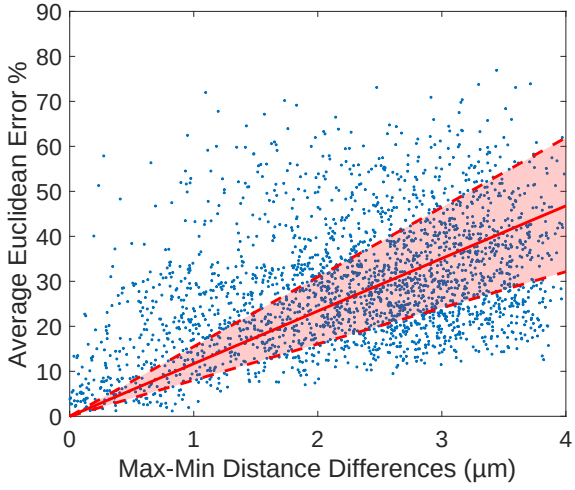
Figure 3.2. Mean percentage of Euclidean errors for each scenario regarding to minimum angle β. As indicated on the figures' legends, blue dots show the mean percentage Euclidean error for simulation scenarios that has transmitters, which are more than two, and red dots represent the mean percentage Euclidean error for simulation scenarios that has exact 2-transmitters.

estimation performance is depicted. As the transmitters are further apart, it is expected that the Euclidean error percentage would decrease. Only the K-Means algorithm, which employs a hard decision logic for the clustering, shows this behavior, though. When all instances are taken into account, soft decision algorithms do not provide

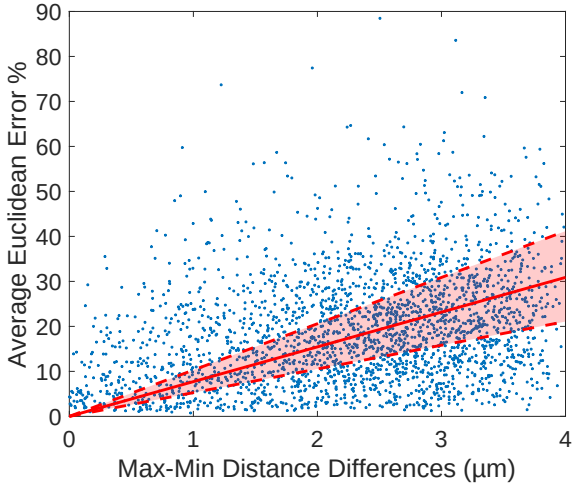
the expected negative correlation, but when two transmitter cases are excluded, the expected patterns are shown.



(a) Gaussian Mixture Model



(b) Dirichlet Mixture Model



(c) K-Means Model

Figure 3.3. Mean percentage Euclidean errors of all scenarios regarding to the difference of maximum-minimum distance between transmitters. Blue dots on the figure indicate the mean percentage Euclidean for a simulation.

Compared to GMM and DiM, K-Means produces results that are more favorable. As a result, K-Means algorithm performs confidence interval analysis for ϵ_ϕ, where confidence interval is an estimation of the difference between sample mean and the real mean with a specific confidence level.

In Table 3.4, there is a specified sample size for the number of binary combination of transmitters and each minimum angle interval, and the mean confidence interval reveals the range of actual error with confidence level 95%. By applying this information, we can more precisely evaluate the algorithm's performance in terms of the number of transmitters and the value of minimum angle.

3.1.3. Effects of Distance Variation

The number of particles absorbed by receiver and emitted from a transmitter that is closer to the receiver has to be greater than those emitted from a transmitter that is further away, according to the system topology and channel model. As a result, a cluster relating to the nearby transmitter is assumed to have more data points. Thus, we give more importance to how the distance variation, which is the difference between the r_0 values of the closest and furthest transmitters, affects the estimation error.

The number of points (or hitting molecules) in the clusters associated to the closest and distant transmitters differs as the distance variation increases. Furthermore, if the number of molecules received from the furthest transmitter is insufficient to identify its cluster, the clustering algorithms are likely to classify such molecules as belonging to another cluster. Higher percentages of Euclidean errors might result from this.

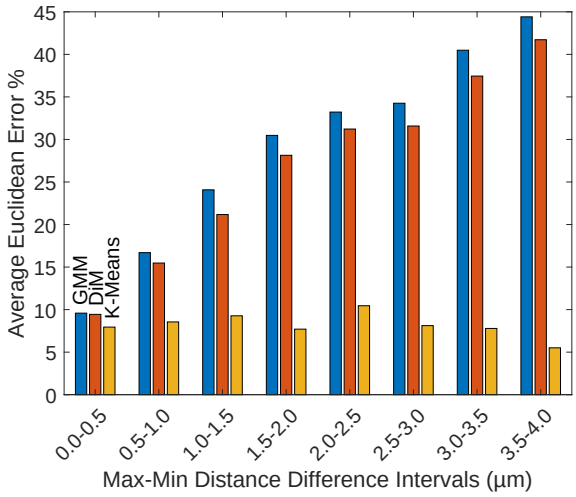
As was previously noted, the differences between the maximum and minimum r_0 values are employed for comparison since they may point to a potential estimate disruption. The linear regression curves and related confidence intervals are demonstrated in Figure 3.3 with the percentage Euclidean errors fitted on them. When all cases are taken into consideration and the wide confidence intervals are used, there is low correlation for all algorithms.

The 2- and 4-transmitter scenarios, as well as all scenarios, are examined under maximum-minimum distance difference (difference between the further transmitter and

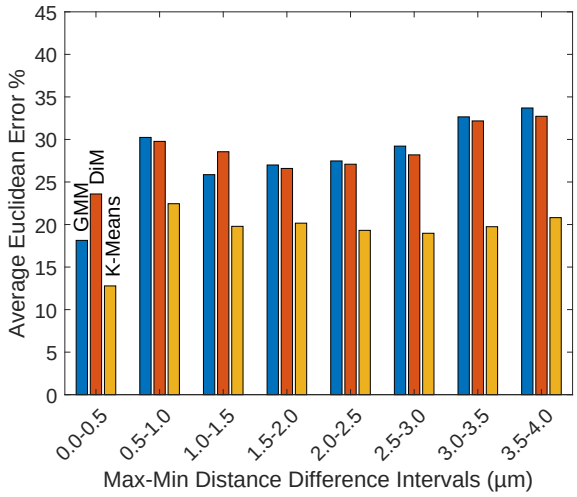
Table 3.4. Mean angle estimation errors of K-Means with the associated mean confidence intervals (confidence level 95%). Please be informed that the lack of greater minimum angle intervals for a given number of transmitters because taking into account the number of transmitters reduces the probability of getting greater minimum angles when initiating transmitters via simulation at random.

Mean Conf. Intervals for ϵ_ϕ of Each Minimum Angle Intervals		
Num of Tx	**Min Angle**	**Conf. Interval for ϵ_ϕ**
2	$0° - 30°$	$18.557° \pm 1.712°$
	$30° - 60°$	$9.899° \pm 0.524°$
	$60° - 90°$	$4.824° \pm 0.222°$
	$90° - 120°$	$2.244° \pm 0.116°$
	$120° - 150°$	$1.289° \pm 0.104°$
	$150° - 180°$	$0.909° \pm 0.141°$
3	$0° - 30°$	$15.356° \pm 1.533°$
	$30° - 60°$	$9.413° \pm 0.621°$
	$60° - 90°$	$4.264° \pm 0.196°$
	$90° - 120°$	$2.219° \pm 0.326°$
4	$0° - 30°$	16.166 ± 0.856
	$30° - 60°$	$9.144° \pm 0.505°$
	$60° - 90°$	$4.551° \pm 0.272°$
5	$0° - 30°$	$16.543° \pm 0.742°$
	$30° - 60°$	$10.574° \pm 0.638°$
	$60° - 90°$	$4.992° \pm 0.716°$
6	$0° - 30°$	$18.099° \pm 0.683°$
	$30° - 60°$	$12.513° \pm 0.844°$
	$60° - 90°$	$5.977° \pm 3.144°$

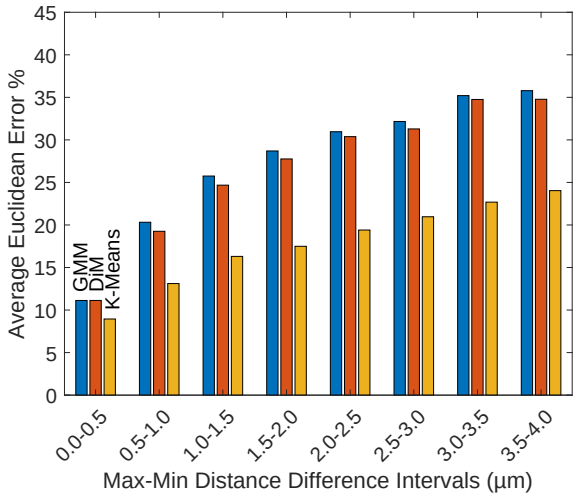
the closest one) intervals for all algorithms for further examination (Figure 3.4). Despite the fact that the mean percentage Euclidean errors in the 2- and 4-transmitter cases do not show any correlation for the K-Means algorithm, a correlation may be established



(a) 2 Transmitter Cases



(b) 4 Transmitter Cases



(c) All Cases

Figure 3.4. Mean percentage Euclidean error in various distance variation intervals for each algorithm.

when all cases are taken into consideration. The first result is that K-means performs much better across all intervals compared to soft decision algorithms. There is the negative correlation between the number of transmitters and the performance difference between the soft decision algorithms and K-Means.

Moreover, interval statistics i.e., median, mean, confidence interval, and outliers is investigated using the box plot shown in Figure 3.5 without distinguishing the in-

stances based on the number of transmitters. A statistical tendency suggests that when distance variation increases, the median and upper end of the confidence interval may as well.

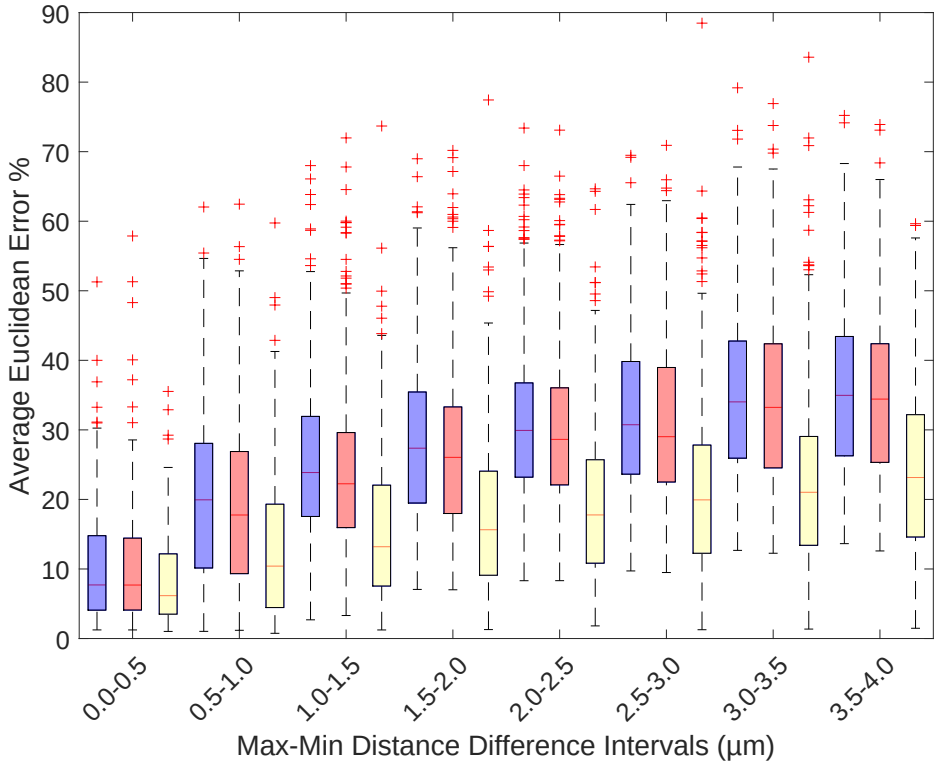


Figure 3.5. Percentage Euclidean error in various distance variation intervals for all scenarios. For the corresponding intervals, blue boxes, red boxes, and yellow boxes denote GMM, DiM, and K-Means, respectively.

3.1.4. Effects of Quantization of the Locations on the Surface

It has been assumed up until now that an absorbing receiver has the ability to store the precise locations of absorbing molecules that are absorbed on its spherical surface. The spherical surface is divided into equal-area regions to satisfy receiver memory and processing capacity concerns (i.e., quantized) [37, 38]. This limits the absorbing receiver's interactions to a certain set of sites and the quantity of absorbed molecules that correspond to them. In this configuration, we evaluate the average angle

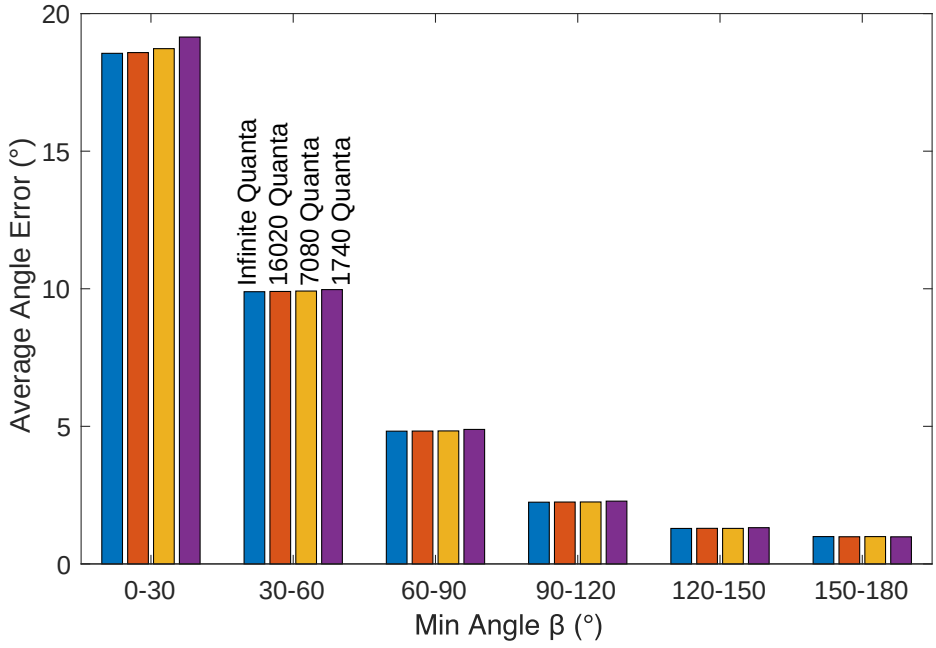


Figure 3.6. Mean angle errors in minimum angle variations for 2-transmitter scenarios. Blue boxes represent infinite quanta (no quantization), red boxes represent 16020 quanta (89 pieces azimuth, 180 pieces elevation), yellow boxes represent 7080 quanta (59 pieces azimuth, 120 pieces elevation), purple boxes represent 1740 quanta (29 pieces azimuth, 60 pieces elevation).

error to discover the quantization effect.

To investigate the impact of quantization, we studied four scenarios with various numbers of quantization bins. The constant angle intervals for each quantization bin for comparable situations are as follows: indefinitely small (i.e., the data without quantization), 2°, 3°, and 4°. To evaluate the performance variation brought on by the locations' quantization over minimum angle values, the K-Means algorithm for the 2-transmitter case is selected. Although the average angle error for each minimum angle interval is similar, when the number of quantization bins is decreased, as shown in Figure 3.6, the mean angle error either slightly increases or stays the same. The angle error caused by the quantization process is therefore negligible.

4. CONCLUSION AND FUTURE WORK

This thesis proposes a solution to the multiple transmitters localization problem in diffusion-based 3D mediums with a single absorbing receiver. Utilizing algorithms for clustering, directional data averaging, and cumulative distributions of hitting probability, we estimate the coordinates of each transmitter. The results demonstrate that by using the information provided by molecules received by a single receiver, we can precisely determine the locations of multiple transmitters. K-Means is the clustering algorithm that offers the most accurate estimation of actual values out of the three that were taken into consideration. Additionally, the error value increases for all algorithms as the number of transmitters increase and the angle between transmitters narrows.

As a future work, by changing the expectation-maximization method to better take into consideration the ground truth of hitting locations' distribution, we want to create a method that estimates the positions of multiple transmitters. Another goal is to precisely predict the transmitter number, such that it is not expected that the receiver already knows this information.

REFERENCES

1. Rudsari, H. K., M. Zoofaghari, M. Veletic, J. Bergsland and I. Balasingham, "The End-to-end Molecular Communication Model of Extracellular Vesicle-based Drug Delivery", *arXiv preprint arXiv:2207.01875*, 2022.

2. Islam, T., E. Shitiri and H.-S. Cho, "In-body Sequential Multidrug Delivery Scheme Using Molecular Communication", *IEEE Access*, Vol. 10, pp. 39975–39985, 2022.

3. Zhao, Q. and L. Lin, "Adaptive Release Rate in Drug Delivery Based on Mobile Molecular Communication", *IEEE Wireless Communications and Networking Conference*, Nanjing, 2021.

4. Schurwanz, M., P. A. Hoeher, S. Bhattacharjee, M. Damrath, L. Stratmann and F. Dressler, "Duality Between Coronavirus Transmission and Air-Based Macroscopic Molecular Communication", *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, Vol. 7, No. 3, pp. 200–208, 2021.

5. Gulec, F. and B. Atakan, "A Molecular Communication Perspective on Airborne Pathogen Transmission and Reception via Droplets Generated by Coughing and Sneezing", *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, Vol. 7, No. 3, pp. 175–184, 2021.

6. Borges, L. F., M. T. Barros and M. Nogueira, "Toward Reliable Intra-body Molecular Communication: An Error Control Perspective", *IEEE Communications Magazine*, Vol. 59, No. 5, pp. 114–120, 2021.

7. Nakano, T., A. W. Eckford and T. Haraguchi, *Molecular Communication*, Cambridge University Press, 2013.

8. Bi, D., A. Almpanis, A. Noel, Y. Deng and R. Schober, "A Survey of Molecular Communication in Cell Biology: Establishing a New Hierarchy for Interdisciplinary

Applications", *IEEE Communications Surveys and Tutorials*, Vol. 23, No. 3, pp. 1494–1545, 2021.

9. Wang, X., M. D. Higgins and M. S. Leeson, "An Algorithmic Distance Estimation Scheme for Diffusion Based Molecular Communication Systems", *IEEE International Conference on Communications*, pp. 1134–1139, London, 2015.

10. Miao, Y., W. Zhang and X. Bao, "Cooperative Source Positioning for SIMO Molecular Communication via Diffusion", *IEEE International Conference on Communication Technology*, pp. 495–499, Xi'an, 2019.

11. Turan, M., B. C. Akdeniz, M. Ş. Kuran, H. B. Yilmaz, I. Demirkol, A. E. Pusane and T. Tugcu, "Transmitter Localization in Vessel-Like Diffusive Channels Using Ring-shaped Molecular Receivers", *IEEE Communications Letters*, Vol. 22, No. 12, pp. 2511–2514, 2018.

12. Kumar, S., "Nanomachine localization in a diffusive molecular communication aystem", *IEEE Systems Journal*, Vol. 14, No. 2, pp. 3011–3014, 2020.

13. Gulec, F. and B. Atakan, "Distance Estimation Methods for a Practical Macroscale Molecular Communication System", *Nano Communication Networks*, Vol. 24, p. 100300, 2020.

14. Gulec, F. and B. Atakan, "Localization of a Passive Molecular Transmitter with a Sensor Network", *Springer International Conference on Bio-inspired Information and Communication Technologies*, pp. 317–335, Cham, 2020.

15. Wang, X., M. D. Higgins and M. S. Leeson, "Distance Estimation Schemes for Diffusion Based Molecular Communication Systems", *IEEE Communications Letters*, Vol. 19, No. 3, pp. 399–402, 2015.

16. Chen, Y., Y. Li, L. Lin and H. Yan, "Parameter Estimation of Diffusive Molecular Communication with Drift", *IEEE Access*, Vol. 8, pp. 142704–142713, 2020.

17. Huang, J.-T., H.-Y. Lai, Y.-C. Lee, C.-H. Lee and P.-C. Yeh, "Distance Estimation in Concentration-based Molecular Communications", *IEEE Global Communications Conference*, pp. 2587–2591, Atlanta, GA, 2013.

18. Moore, M. J., T. Nakano, A. Enomoto and T. Suda, "Measuring Distance From Single Spike Feedback Signals in Molecular Communication", *IEEE Transactions on Signal Processing*, Vol. 60, No. 7, pp. 3576–3587, 2012.

19. Noel, A., K. C. Cheung and R. Schober, "Bounds on Distance Estimation via Diffusive Molecular Communication", *IEEE Global Communications Conference*, pp. 2813–2819, Austin, TX, 2014.

20. Luo, Z., L. Lin, Q. Fu and H. Yan, "An Effective Distance Measurement Method for Molecular Communication Systems", *IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*, pp. 1–4, Hong Kong, 2018.

21. Sun, Y., M. Ito and K. Sezaki, "An Efficient Distance Measurement Approach in Diffusion-based Molecular Communication Based on Arrival Time Difference", *Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication*, pp. 1–6, 2017.

22. Lin, L., Z. Luo, L. Huang, C. Luo, Q. Wu and H. Yan, "High-accuracy Distance Estimation for Molecular Communication Systems via Diffusion", *Elsevier Nano Communication Networks*, Vol. 19, pp. 47–53, 2019.

23. Lin, L., C. Yang, S. Ma and M. Ma, "Parameter Estimation of Inverse Gaussian Channel for Diffusion-based Molecular Communication", *IEEE Wireless Communications and Networking Conference*, pp. 1–6, Doha, 2016.

24. Huang, S., L. Lin, W. Guo, H. Yan, J. Xu and F. Liu, "Initial Distance Estimation and Signal Detection for Diffusive Mobile Molecular Communication", *IEEE Transactions on NanoBioscience*, Vol. 19, No. 3, pp. 422–433, 2020.

25. ShahMohammadian, H., G. G. Messier and S. Magierowski, "Blind Synchronization in Diffusion-Based Molecular Communication Channels", *IEEE Communications Letters*, Vol. 17, No. 11, pp. 2156–2159, 2013.

26. Lin, L., J. Zhang, M. Ma and H. Yan, "Time Synchronization for Molecular Communication with Drift", *IEEE Communications Letters*, Vol. 21, No. 3, pp. 476–479, 2016.

27. Srinivas, K. V., A. W. Eckford and R. S. Adve, "Molecular Communication in Fluid Media: The Additive Inverse Gaussian Noise Channel", *IEEE Transactions on Information Theory*, Vol. 58, No. 7, pp. 4678—4692, 2012.

28. Yilmaz, H. B., A. C. Heren, T. Tugcu and C. Chae, "Three-Dimensional Channel Characteristics for Molecular Communications With an Absorbing Receiver", *IEEE Communications Letters*, Vol. 18, No. 6, pp. 929–932, 2014.

29. Straub, J., J. Chang, O. Freifeld and J. Fisher III, "A Dirichlet process mixture model for spherical data", *Artificial Intelligence and Statistics*, Vol. 38, pp. 930–938, 2015.

30. Taghia, J., Z. Ma and A. Leijon, "Bayesian Estimation of the Von-Mises Fisher Mixture Model With Variational Inference", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 36, No. 9, pp. 1701–1715, 2014.

31. Gopal, S. and Y. Yang, "Von Mises-Fisher Clustering Models", *International Conference on Machine Learning*, pp. 154–162, Bejing, 2014.

32. McGraw, T., B. C. Vemuri, B. Yezierski and T. Mareci, "Von Mises-Fisher Mixture Model of the Diffusion ODF", *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 65–68, Arlington, VA, 2006.

33. Franke, J., C. Redenbach and N. Zhang, "On a Mixture Model for Directional Data on the Sphere", *Scandinavian Journal of Statistics*, Vol. 43, No. 1, pp. 139–155,

2016.

34. Banerjee, A., I. Dhillon, J. Ghosh and S. Sra, "Generative Model-Based Clustering of Directional Data", *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 19–28, New York, NY, 2003.

35. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine Learning in Python", *The Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.

36. Akdeniz, B. C., N. A. Turgut, H. B. Yilmaz, C. Chae, T. Tugcu and A. E. Pusane, "Molecular Signal Modeling of a Partially Counting Absorbing Spherical Receiver", *IEEE Transactions on Communications*, Vol. 66, No. 12, pp. 6237–6246, 2018.

37. Hancock, M., "Equal Area Sphere Partitioning", `http://notmatthancock.github.io/2017/12/26/regular-area-sphere-partitioning.html`, 2017, [accessed on May 16, 2021].

38. Beckers, B. and P. Beckers, "A General Rule for Disk and Hemisphere Partition into Equal-area Cells", *Elsevier Computational Geometry*, Vol. 45, No. 7, pp. 275–283, 2012.

APPENDIX A: COPYRIGHT PERMISSION GRANTS