MODELING AND ANALYSIS OF MOLECULAR SIGNALS IN MULTISCALE MOLECULAR COMMUNICATION

A Dissertation Submitted to
the Graduate School of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Electronics and Communication Engineering

by
Fatih GÜLEÇ

July 2021
İZMİR

ACKNOWLEDGMENTS

I would like to express my gratitudes and sincere thanks to my supervisor, Prof. Dr. Barış Atakan, who led me into the area of molecular communication. I feel very lucky that I have the chance to work with him. I am indebted to Prof. Atakan for his continued support, guidance, understanding and fruitful discussions. I would like to express my sincere thanks to Prof. Dr. Cüneyt Güzeliş and Prof. Dr. Bilge Karaçalı for accepting to drive this thesis further with their valuable comments. My cordial thanks also extend to Prof. Dr. Mustafa Aziz Altınkaya and Prof. Dr. Tuna Tuğcu for taking part in my thesis committee and for their valuable suggestions and comments.

I would also like to acknowledge the support of Tübitak between 2019 and 2021 to conduct my studies via the grant number 119E041.

I would like to thank Dr. Eylem Kurulgan Demirci, Dr. Abdurrahman Gümüş, Çağın Ekici and my undergraduate student Ahmet Alaca for their support in laboratory equipment.

I would like to thank my colleagues Bilal Orkan Olcay and Aslı Taşçı for their friendship and motivational coffee breaks. I would like to express my sincere thanks to my ex-room mate and colleague Serhat Çağdaş for his sincere friendship and fruitful discussions.

Last but not least, I am thankful to my mother, my father and my grandmother for their endless support. I also thank my beloved dog Güneş and my girl friend Burcu Ovacık for their company and love during the writing process of this thesis. Finally, I would like to thank my ex-wife and friend Sinem Ercan. Without her support, I could not have started the journey that allowed me to become an academic.

ABSTRACT

MODELING AND ANALYSIS OF MOLECULAR SIGNALS IN MULTISCALE MOLECULAR COMMUNICATION

This thesis focuses on modeling, analysis, and novel experimental techniques in molecular communication (MC). The objective of this thesis is to develop novel engineering solutions and modeling approaches to enable MC applications. The first part of the thesis is about microscale MC studies. In this part, a model of how a receiver nanomachine measures and reconstructs a molecular signal is proposed with a probabilistic approach.

In the second part, macroscale MC studies with active transmitters are given. An experimental setup which includes a sprayer emitting alcohol molecules as a transmitter and an alcohol sensor as the receiver is employed. Using the data collected by this setup, five statistical methods, a feature extraction algorithm and the fluid dynamics-based distance estimation algorithm are proposed for distance estimation. Furthermore, a novel droplet-based signal reconstruction approach to channel modeling is proposed. Moreover, MC is utilized to propose an end-to-end system model which considers pathogen-laden cough/sneeze droplets as the input and the infection state of the human as the output. In addition, the concept of mobile human ad hoc network which exploits the similarity of airborne transmission-driven human groups with mobile ad hoc networks and uses MC as the enabling paradigm is introduced.

Finally, macroscale MC studies with passive transmitters are detailed in the third part. A novel experimental platform which consists of an evaporating alcohol source and a sensor network is proposed. A sensor network based clustered localization algorithm is proposed to estimate the location of the passive transmitter.

ÖZET

ÇOK ÖLÇEKLİ MOLEKÜLER HABERLEŞMEDE MOLEKÜLER SİNYALLERİN MODELLENMESİ VE ANALİZİ

Bu tez, moleküler haberleşmede (MH) modelleme, analiz ve yeni deneysel tekniklere odaklanmaktadır. Bu tezin amacı, MH uygulamalarını mümkün kılmak için yeni mühendislik çözümleri ve modelleme yaklaşımları geliştirmektir. Tezin ilk bölümü mikro ölçekli MH çalışmaları hakkındadır. Bu kısımda, olasılıksal bir yaklaşımla bir alıcı nanomakinenin bir moleküler sinyali nasıl ölçtüğüne ve yeniden oluşturduğuna dair bir model önerilmektedir.

İkinci bölümde, aktif vericilerle makro ölçekli MH çalışmaları verilmektedir. Verici olarak alkol molekülleri yayan bir püskürtücü ve alıcı olarak bir alkol sensörü içeren deneysel bir düzenek kullanılmıştır. Bu düzenekle toplanan veriler kullanılarak, mesafe kestirimi için beş istatistiksel yöntem, bir öznitelik çıkarma algoritması ve akışkan dinamiğine dayalı mesafe kestirim algoritması önerilmektedir. Ayrıca, kanal modellemeye yönelik yeni bir damlacık bazlı sinyalin yeniden oluşturulması yaklaşımı önerilmektedir. Bu çalışmaların yanı sıra MH, patojen yüklü öksürük/hapşırık damlacıklarını girdi olarak ve insanın enfeksiyon durumunu çıktı olarak kabul eden uçtan uca bir sistem modeli önermek için kullanılır. Ek olarak, havadan bulaşma güdümlü insan gruplarının mobil tasarsız ağlarla benzerliğinden yararlanan ve kolaylaştırıcı paradigma olarak MH'yi kullanan mobil insan tasarsız ağı kavramı tanıtılmaktadır.

Son olarak, pasif vericilerle makro ölçekli MH çalışmaları üçüncü kısımda detaylandırılmaktadır. Buharlaşan bir alkol kaynağı ve bir sensör ağından oluşan yeni bir deneysel platform önerilmektedir. Pasif vericinin konumunu kestirmek için bir sensör ağı tabanlı kümelenmiş lokalizasyon algoritması önerilmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

1.1. Molecular Communication

One of our most important characteristics as humans is to constantly try to make sense of what we live and our thoughts in the context of a story (Harari, 2014). We send our thoughts to each other as a message by encoding it into voices, texts and pictures. For example, human communities living as hunter-gatherers about 15,000 years ago were sending us a message with the pictures they drew in Spain's Altamira cave (Gombrich, 1995). The evolution of speech is actually one of the examples of our need to send a message to each other. Over time, new ways to quickly send our messages to greater distances have been developed. Now, thanks to technological tools such as smart phones, we can send our messages to the other end of the world at the speed of light via electromagnetic waves. So, do only humans send messages?

1.1.1. Molecular Communication in Nature

At microscale (nm to cm), the message exchange of unicellular organisms such as bacteria is based on molecular communication (MC), which is one of the oldest communication methods in nature. Namely, messages are transmitted via special molecules and interpreted by the sensing biological entities. For example, bacteria transmit signaling molecules for other bacteria to indicate their existence to the other members of the bacterial community. According to the concentration level of the sensed molecules by these receiver members, they decide whether there are sufficient bacteria in the bacterial community or not (quorum sensing) (Miller and Bassler, 2001). In addition, organized tissue formation of cells in the human body is realized by the information that is transmitted to organs in different parts of the body with signaling molecules called hormones. Hence,

multicellular organisms such as humans also use MC for the information transfer among biological cells (Atakan, 2014).

At macroscale (cm to m), many insect species communicate with each other using chemicals called pheromones. For example, when ants find a food source, they carry the food up to their nest in a string like a line. This "ant highway" is created by pheromones that let each other know that they find food. The information generated in this way is sensed by the other members of the ant colony and transmitted to the nest, allowing the food found to be transferred to their nests quickly with the other members of the colony receiving this information (Bossert and Wilson, 1963). Furthermore, plants are known to use MC with each other through volatile organic compounds similar to pheromones. For example, when the tomato plant detects the danger of herbivorous insects that can eat its leaves, it can notify the surrounding tomato plants of the presence of this insect with the molecules it secretes (Coppola et al., 2017).

1.1.2. Molecular Communication from the Perspective of Communication Engineering

After introducing the use of MC in nature, we can move on to where and how MC can be used in communication engineering. MC is a biologically inspired communication paradigm where chemical signals are used instead of electrical signals to carry information. MC is an emerging area combining biology, chemistry, biophysics and communication engineering. Biologists have been researching in this area for a long time, but the research on MC by communication engineers is still in its infancy. MC can be performed in aqueous or gaseous environments. Moreover, MC can be examined in microscale and in macroscale (Farsad et al., 2016).

1.1.2.1. Microscale Molecular Communication

With the development of bio/nanotechnology, it has become possible to produce nanomachines (NMs) which are defined as the artificial devices composed of the nanometer-scale components. In the existing literature, the term NM mostly refers to bionanoma-

chines, bionanorobots and genetically engineered cells (Atakan, 2014). While a single NM is a very low-end machine with extremely limited capabilities, the interconnection of the NMs forms nanonetworks and makes sophisticated bio/nanotechnology applications possible. It is important for the NMs to communicate among each other to perform more complex tasks (Akyildiz et al., 2008, 2011). MC is one of the most prominent communication paradigms for the interconnection of the NMs (Hiyama et al., 2005). A diffusion-based MC system which is composed of a transmitter NM (TN) and receiver NM (RN) is illustrated in Figure 1.1. In this MC system, information symbols which are encoded in messenger molecules are released by the TN to be received by the RN. In microscale, if there is not any fluid flow in the MC channel, the movement of the released molecules is governed by free diffusion, i.e., Brownian motion (Berg, 1993).

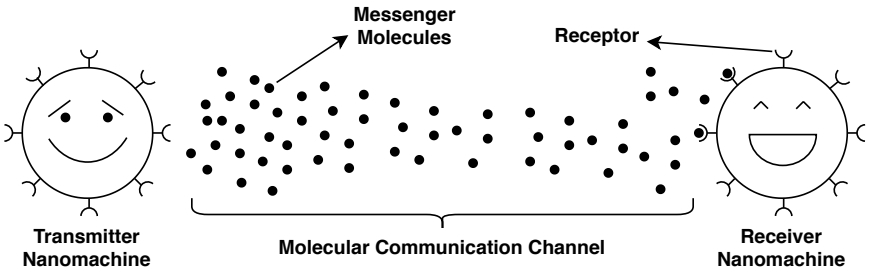


Figure 1.1. A diffusion-based MC system.

Nanonetworks can have biomedical, industrial, environmental and military applications (Akyildiz et al., 2008). Although there is not a commercialized application yet, the most important applications of nanonetworks are biomedical applications such as targeted drug delivery in the human body, cancer treatment and health monitoring (Atakan et al., 2012). As shown in Figure 1.2, after a swarm of NMs forming a nanonetwork is released into the body through a pill or injector, NMs can only target infected cells to deliver drug without harming healthy cells in a coordinated manner by communicating with each other. In this way, the harmful effects of methods such as "chemotherapy", which is widely used in cancer treatment and harms healthy cells, are eliminated.

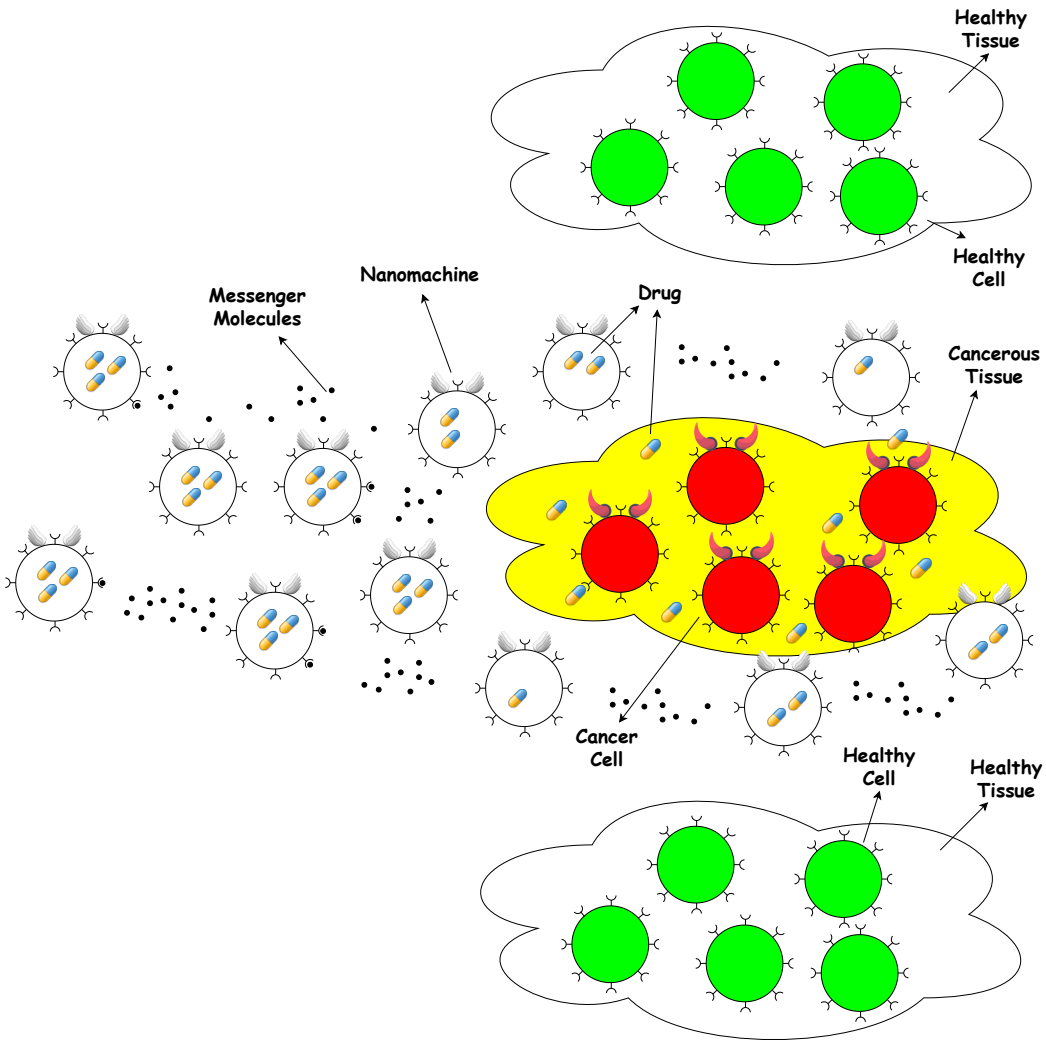


Figure 1.2. Targeted drug delivery application of a nanonetwork enabled by MC.

1.1.2.2. Macroscale Molecular Communication

Although MC is mostly studied in microscale, it paves the way for novel practical macroscale applications which are introduced as follows. The pioneering experimental study in macroscale MC is about establishing a communication link using an electrical sprayer with a fan behind it, an alcohol sensor and alcohol molecules as the transmitter (TX), receiver (RX) and messenger molecules, respectively (Farsad et al., 2013). For this experimental setup, several channel models are proposed in (Farsad et al., 2014; Kim et al., 2015) and its data rate is improved via multiple input multiple output (MIMO) technique in (Koo et al., 2016). Furthermore, it is shown that MC can be used in macroscale

environments such as pipe networks where there is significant attenuation for electromagnetic wave-based communication (Qiu et al., 2014; Guo et al., 2015). In (Farsad et al., 2017), an experimental platform consisting of pumps, pipes and a pH meter is proposed. This platform encodes the information symbols according to the pH level, as also used in a macroscale fluidic platform explained in (Khaloopour et.al., 2019). Magnetic nanoparticles which are sensed by a susceptometer (as a RX) are employed to encode information symbols in (Unterweger et al., 2018). Another experimental platform is accomplished by using an odor generator as the TX and a mass spectrometer as the RX (Giannoukos et al., 2017; McGuiness et al., 2018, 2019). (Ozmen et al., 2018) proposes a platform with a chemical vapor TX and photoionization detectors. Moreover, laser induced florescence technique is implemented for reception in an aqueous experimental platform for macroscale MC (Abbaszadeh et.al., 2019). Similarly in Guo et al. (2020), it is shown that vertical underwater MC via buoyancy is possible by the help of particle image velocimetry for reception. Furthermore, MC is proposed to be used in mobile robot platforms (Wang et al., 2015; Zhai et al., 2018). In (Zhai et al., 2018), an algorithm is proposed for mobile RX robots to move towards a static sprayer, similar to a bacteria swarm.

In this thesis, we classify the TXs (or sources) as passive and active in macroscale MC. Active TXs emit molecules with an initial force which causes an initial velocity, whereas there is not any initial force to emit molecules in passive TXs.

1.2. Contributions

The contributions of this thesis in the field of micro- and macroscale molecular communication can be listed as follows.

- A realistic model about the sensing of molecular signals by the receiver nanomachines is developed.
- The first distance estimation methods including three novel data analysis-based methods is developed and implemented experimentally for macroscale MC systems.
- The first feature extraction algorithm that processes molecular signals in macroscale MC is introduced for statistical estimation methods.

- A novel fluid dynamics-based distance estimation algorithm is developed for macroscale MC systems.
- An end-to-end system model is introduced for macroscale MC systems and validated by experimental data.
- The airborne transmission of pathogens between an infected and a susceptible human is modeled as a MC system for the first time in the literature.
- Mobile human ad hoc network architecture is conceptualized as a framework merging biology, fluid dynamics, medicine and epidemiology for the spread of infectious diseases through airborne pathogen transmission.
- The first macroscale experimental MC platform for passive sources is developed and implemented by using a sensor network.
- Sensor network-based clustered localization algorithm, which is the first localization algorithm for passive sources, is introduced.

1.3. Organization of the Thesis

In this section, the organization of the thesis is given via the summaries of each chapter as follows.

Part 1: Studies in Microscale Molecular Communication

Chapter 2 - Signal Reconstruction in Diffusion-based Molecular Communication: In this chapter, a model of how a reciver nanomachine (RN) measures and reconstructs the molecular signal is proposed. The signal around the RN is assumed to be a Gaussian random process instead of the less realistic deterministic approach. After the reconstructed signal is derived as a Doubly Stochastic Poisson Process (DSPP), the distortion between the signal around the RN and the reconstructed signal is derived as a new performance parameter in MC systems. The derived distortion which is a function of system parameters such as RN radius, sampling period and the diffusion coefficient of the channel, is shown to be valid by employing random walk simulations. Then, it is shown that the original signal can be satisfactorily reconstructed with a sufficiently low-level of

distortion. Finally, optimum RN design parameters, namely RN radius, sampling period and sampling frequency, are derived by minimizing the signal distortion. The simulation results reveal that there is a trade-off among the RN design parameters which can be jointly set for a desired signal distortion.

Part 2: Studies in Macroscale Molecular Communication with Active Sources

Chapter 3 - Distance Estimation Methods for a Practical Macroscale Molecular Communication System: Accurate estimation of the distance between the transmitter (TX) and the receiver (RX) in MC systems can provide faster and more reliable communication. In addition, distance information can be used in determining the location of the molecular source in practical applications. Existing theoretical models in the literature are not suitable for distance estimation in a practical scenario. Furthermore, deriving an analytical model is a nontrivial problem, since the liquid in the TX is sprayed as droplets rather than molecules, these droplets move according to Newtonian mechanics, the size of the droplets change during their propagation and droplet-air interaction causes unsteady flows. Therefore, five different practical methods comprising three novel data analysis based methods and two supervised machine learning (ML) methods, Multivariate Linear Regression (MLR) and Neural Network Regression (NNR), are proposed for distance estimation at the RX side. In order to apply the ML methods, a macroscale practical MC system, which consists of an electric sprayer without a fan, alcohol molecules, an alcohol sensor and a microcontroller, is established, and the received signals are recorded. A feature extraction algorithm is proposed to utilize the measured signals as the inputs in ML methods. The numerical results show that the ML methods outperform the data analysis based methods in the root mean square error sense with the cost of complexity. The nearly equal performance of MLR and NNR shows that the input features such as peak time, peak concentration and the energy of the received signal have a highly linear relation with the distance. Moreover, the peak time based estimation, which is one of the proposed data analysis based methods, yields better results with respect to the other proposed four methods, as the distance increases. Given the experimental data and fluid dynamics theory, a possible trajectory of the molecules between the TX and RX is given. Our findings show that distance estimation performance is jointly affected by unsteady flows and the non-linearity of the sensor. According to our findings based on fluid dynamics, it is evaluated that fluid dynamics should be taken into account for more accurate

parameter estimation in practical macroscale MC systems.

Chapter 4 - Fluid Dynamics-based Distance Estimation Algorithm for Macroscale Molecular Communication: In this chapter, a novel approach based on fluid dynamics is proposed for the derivation of the distance estimation in practical MC systems. According to this approach, transmitted molecules are considered as moving and evaporating droplets in the MC channel. With this approach, the Fluid Dynamics-based Distance Estimation (FDDE) algorithm which predicts the propagation distance of the transmitted droplets by updating the diameter of evaporating droplets at each time step is proposed. FDDE algorithm is validated by experimental data. The results reveal that the distance can be estimated by the fluid dynamics approach which introduces novel parameters such as the volume fraction of droplets in a mixture of air and liquid droplets and the beamwidth of the TX. Furthermore, the effect of the evaporation is shown with the numerical results.

Chapter 5 - A Droplet-based Signal Reconstruction Approach to Channel Modeling in Molecular Communication: In this chapter, a novel droplet-based signal reconstruction (SR) approach to channel modeling, which considers liquid droplets as information carriers instead of molecules in the MC channel, is proposed for practical sprayer-based macroscale MC systems. Our proposed approach takes a two-phase flow which is generated by the interaction of droplets in liquid phase with air molecules in gas phase into account. Two-phase flow is combined with the SR of the RX to propose a channel model. The SR part of the model quantifies how the accuracy of the sensed molecular signal in its reception volume depends on the sensitivity response of the RX and the adhesion/detachment process of droplets. The proposed channel model is validated by employing experimental data.

Chapter 6 - A Molecular Communication Perspective on Airborne Pathogen Transmission and Reception via Droplets Generated by Coughing and Sneezing: Infectious diseases spread via pathogens such as viruses and bacteria. Airborne pathogen transmission via droplets is an important mode for infectious diseases. In this chapter, the spreading mechanism of infectious diseases by airborne pathogen transmission between two humans is modeled with a molecular communication perspective. An end-to-end system model which considers the pathogen-laden cough/sneeze droplets as the input and the infection state of the human as the output is proposed. This model uses the

gravity, initial velocity and buoyancy for the propagation of droplets and a receiver model which considers the central part of the human face as the reception interface is proposed. Furthermore, the probability of infection for an uninfected human is derived by modeling the number of propagating droplets as a random process. The numerical results reveal that exposure time affects the probability of infection. In addition, the social distance for a horizontal cough should be at least 1.7 m and the safe coughing angle of a coughing human to infect less people should be less than -25°.

Chapter 7 - Mobile Human Ad Hoc Networks: A Communication Engineering Viewpoint on Interhuman Airborne Pathogen Transmission: Pathogens such as viruses and bacteria play a vital role in human life, since they cause infectious diseases which can lead to epidemics. Recent coronavirus disease 2019 epidemic has shown that taking effective prevention measures such as wearing masks are important to reduce the human deaths and side effects of the epidemic. It is therefore requisite to accurately model the spread of infectious diseases whose one of the most crucial routes of transmission is airborne transmission. The transmission models in the literature are proposed independently from each other, at different scales and by the researchers from various disciplines. In small scale, airborne transmission models using fluid dynamics do not consider the facts about the survival of pathogens and their interactions with the human cells. In a larger scale, the epidemiological models omit the knowledge from research areas such as biology and fluid dynamics and make rough estimations by fitting statistical data. Therefore, there is a need to merge all these research attempts. To this end, in this chapter, we propose a communication engineering approach that melts different disciplines such as epidemiology, biology, medicine, and fluid dynamics in the same pot to model airborne pathogen transmission among humans. In this approach, we introduce the concept of mobile human ad hoc networks (MoHANETs). This concept exploits the similarity of airborne transmission-driven human groups with mobile ad hoc networks and uses molecular communication as the enabling paradigm. The aim of this chapter is to present a unified framework using communication engineering, and to highlight future research directions for modeling the spread of infectious diseases among humans through airborne pathogen transmission. In this chapter, we first review the airborne pathogen transmission mechanisms. Then, the MoHANET is given with a layered structure. In these layers, the infectious human emitting pathogen-laden droplets through air and the exposed human to

these droplets are considered as the transmitter and receiver, respectively. Our preliminary results show that the proposed MoHANET architecture can be employed to predict the dynamics of infectious diseases by considering the propagation of pathogen-laden droplets, their reception and mobility of humans. Moreover, the experimental methods for the proposed approach are reviewed and discussed.

Part 3: Studies in Macroscale Molecular Communication with Passive Sources

Chapter 8 - Localization of a Passive Molecular Transmitter with a Sensor Network: Macroscale MC, which has a potential for practical applications, is a promising area for communication engineering. In a practical scenario such as monitoring air pollutants released from an unknown source, it is essential to estimate the location of the TX. This chapter presents a novel Sensor Network-based Clustered Localization Algorithm (SNCLA) for passive transmission by using a novel experimental platform which mainly comprises a clustered sensor network (SN) with 24 sensor nodes and evaporating ethanol molecules as the passive TX. In SNCLA, Gaussian plume model is employed to derive the location estimator. The parameters such as transmitted mass, wind velocity, detection time and actual concentration are calculated or estimated from the measured signals via the SN to be employed as the input for the location estimator. The numerical results show that the performance of SNCLA is better for stronger winds in the medium. Our findings show that evaporated molecules do not propagate homogeneously through the SN due to the presence of the wind. In addition, the estimation error of SNCLA decreases for higher detection threshold values.

Chapter 9 - Conclusions and Future Research Directions: In this chapter the concluding remarks of the thesis are summarized and future research directions are given.

CHAPTER 2

SIGNAL RECONSTRUCTION IN DIFFUSION-BASED MOLECULAR COMMUNICATION

2.1. Introduction

In diffusion-based molecular communication (MC), the transmitter nanomachine (TN) sends information symbols to the receiver nanomachine (RN) by emitting different levels of molecule concentrations. Then, the RN senses the surrounding molecule concentration levels to reconstruct the signal emitted by the TN. In order to understand the performance of the MC systems more clearly, the accuracy of the signal reconstruction needs to be investigated.

In fact, the accuracy of concentration sensing is studied in the biophysics domain where the cell is considered as a molecule concentration measuring device. In the biophysics literature, there are two approaches about how the cell infers information about its environment by sensing the molecule concentration. The first approach is perfect monitoring where the cell is modeled as a permeable sphere and counts the molecules inside its volume. The second approach is perfect absorbing, where the cell is assumed to count the molecules hitting its surface. The first work about the molecule concentration measurement of a cell is given by Berg and Purcell (Berg and Purcell, 1977). In this pioneering paper, how a cell can measure the constant molecule concentration as a perfect monitoring device and as a perfect absorbing device with receptors are proposed. The uncertainty, which is defined as the mean square fluctuation of the measured molecule concentration, is derived for a constant molecule concentration outside the cell. In (Endres and Wingreen, 2008), the cell is modeled as a perfect absorber which counts the molecules with the receptors on its surface. Furthermore, the cell is modeled as a gradient sensing device for perfect monitoring and perfect absorbing models. It is stated that the perfect absorbing model is better than the perfect monitoring model both in concentra-

tion measurement and gradient sensing, since the previously counted molecules are not counted again and removed from the medium (Endres and Wingreen, 2008). Another method for the concentration measurement is the maximum likelihood estimation (MLE) which is derived by using the probability of time series for the receptor occupancy by the molecules (Endres and Wingreen, 2009). The uncertainty of the estimate is found to be better by a factor of two according to the Berg-Purcell limit given in (Berg and Purcell, 1977). This corresponds to the fact that cells can sense the molecule concentration two times more accurate with MLE. The comparison of these models and sensing limits are given in (Aquino et al., 2016).

In the existing literature of diffusion-based MC, there are two assumptions about the reception of the RNs. The first one is that the molecule concentration around the RN is assumed to be constant (Pierobon and Akyildiz, 2010, 2011b). In (Pierobon and Akyildiz, 2010), the reception process is given as a transformation process. However, there is no derivation of what this process is related to. In (Pierobon and Akyildiz, 2011b), the molecule concentration outside the RN, which is modeled as a receiver with receptors, is given as a deterministic function and an additive reception noise is defined before the reception which is employed to model the random effects of the molecule-receptor binding process. Similarly, an additive counting noise is defined (Pierobon and Akyildiz, 2011a) and employed in the literature (Lin et al., 2018; Mosayebi et al., 2018; Lin et al., 2019) to model the error between the constant signal outside the RN and the reconstructed signal, but the distortion between them is not derived in these studies. Furthermore, the additive noise approach is used in several studies about non-coherent signal detection (Li et al., 2017, 2016b,a; Liu et al., 2019). However, these papers focus on mitigating the effect of inter-symbol interference due to the long-tail nature of the channel impulse response, not the sensing model of the RN. The second assumption about the reception of the RNs, is that the molecule concentration is assumed to be sensed perfectly by the RNs (Kilinc and Akan, 2013; Srinivas et al., 2012; Noel et al., 2014b; Mustam et al., 2017). In (Atakan and Akan, 2010), the reception of the RN is modeled by defining a virtual reception volume which we utilize for our study. However, (Atakan and Akan, 2010) only focuses on the derivation of a realistic channel model and the molecular communication rate on that channel and issues of signal reconstruction are not addressed.

Due to the stochastic nature of the molecule movements, the mentioned two as-

sumptions in the literature cannot be realistic. None of the studies in the diffusion-based MC literature assumes the molecule concentration as a random process and derives an error, which can occur during the molecule sensing process of the RN. In this study, a novel approach is proposed without any need for these assumptions. The existing molecule concentration around the RN is modeled as a Gaussian random process resulting the reconstructed signal as a Doubly Stochastic Poisson Process (DSPP). In the neuro-spike communication channel (Aghababaiyan et al., 2018a; Maham and Kizilirmak, 2018; Aghababaiyan et al., 2018b), the DSPP is also used to model the input spike-trains rather than modeling of concentration around the RN. Furthermore, in those works, how accurate a molecular concentration around the RN is reconstructed is not investigated. With the random process approach, the signal reconstruction performance is investigated by deriving a signal distortion function, which is the Mean Square Error (MSE) between the existing signal around the RN and the reconstructed signal. By using the random walk simulation of the molecules, the derived distortion function, which consist of the system parameters such as the RN radius, diffusion coefficient and sampling period, is validated. In addition, the distributions of the original and the reconstructed signals are generated. The results show that the RN can reconstruct the surrounding signal with a small distortion, if the system parameters are appropriately selected. Besides the derivation of the distortion during the signal reconstruction, our work contributes to the literature by revealing the relation between the signal reconstruction and the RN design. We obtain optimum RN design parameters using the cases where the signal distortion function is minimum with respect to the corresponding parameter. The trade-off among the RN design parameters is shown with the numerical results. Through the extensive analytical and numerical analyses, the optimal design parameters of the RN such as the RN radius and sampling frequency are investigated by minimizing the signal distortion.

The remainder of the chapter is organized as follows. In Section 2.2, the motivation to find the accuracy of the molecular signal reconstruction is given. In Section 2.3, the system model for the signal reconstruction is introduced. The distortion of the reconstructed signal in the MSE sense is derived in Section 2.4. The validation of the system model and the numerical results are presented in Section 2.5. In Section 2.6, the optimum design parameters of the RN are derived and analyzed. This study was first introduced in (Atakan and Gulec, 2019).

2.2. Motivation

In this section, our motivation for the system model is explained by using a one dimensional scenario. Due to its simplicity, one dimensional movement of the molecules is used to show the concept that the random movements of the molecules cause the molecule concentration (or the molecular signal) to be a random process. This concept can be applied to realistic three dimensional scenarios. Let us consider a TN emitting molecules instantaneously as a single spike to send an information symbol to the RN. The molecule concentration around RN can be given by

$$C = \frac{Q}{\sqrt{4\pi Dt}} e^{\frac{-r^2}{4Dt}}, \tag{2.1}$$

where Q is the number of the molecules, C is the molecule concentration, D is the diffusion coefficient, r is the distance from the TN and t is the propagation time (Bossert and Wilson, 1963).

Actually, (2.1) characterizes the average behavior of molecule concentration at a certain distance from the TN. However, the instantaneous molecule concentration changes randomly due to the random walk phenomenon which is utilized to model the diffusion of the molecules. According to this phenomenon, molecules make random steps to the right or to the left direction with equal probability on the x axis. Their successive steps are statistically independent from their previous steps. The random movements cause the molecule concentration to change randomly at each instant of the diffusion. Therefore, the molecule concentration generates a random process rather than a constant value given in (2.1). This is illustrated in Fig. 2.1 where the solid line represents the theoretical model in (2.1) and the oscillating line represents the random process generated via a random walk simulation. Due to the randomness of the molecules, the molecule concentration needs to be modeled by a random process instead of a deterministic function. Therefore, our main motivation is to investigate how accurate the RN reconstructs the surrounding signal, which is modeled as a random process, as given in the next section.

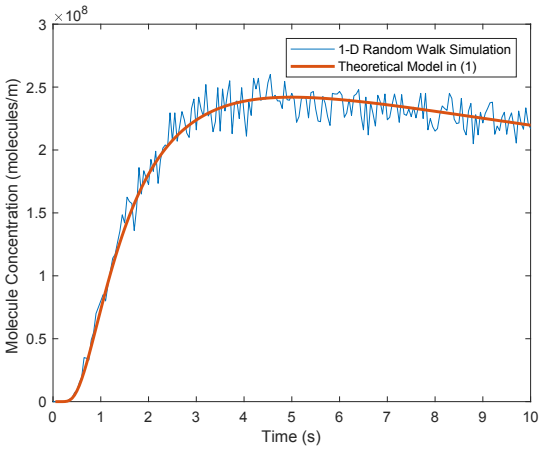


Figure 2.1. The comparison of the theoretical model and the random walk simulation for the molecule concentration with the parameters $Q = 10^4$, $D = 10^{-11}$ m^2/s , $r = 10$ μm, (step time) $\tau = 10^{-3}$ s, (step length) $\delta = 0.0447$ μm.

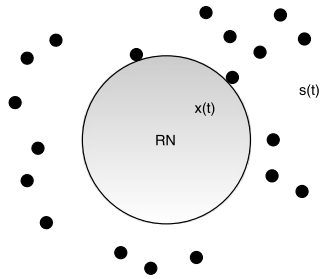


Figure 2.2. Signal reconstruction model for the RN.

2.3. System Model

A model to perceive and reconstruct the signal around the RN is proposed in this section. The RN is assumed to be a perfect absorber meaning that a molecule is received, when it hits its surface. In this way, the RN senses the molecule concentration during certain observation periods, i.e., samples the signal. It is also assumed that no chemical reaction occurs during the movement of the molecules. Let $s(t)$ be the random process that shows the number of the molecules outside the RN and $x(t)$ be the counted number of the molecules by the RN. This model is illustrated in Fig. 2.2. Due to the large number of the molecules, $s(t)$ is assumed to be a Gaussian random process with mean $m(t)$, variance $v(t)$ and autocovariance $R(t_1, t_2)$. The RN counts the molecules along a sampling period T and it is assumed that the number of the molecules does not change in this period. An

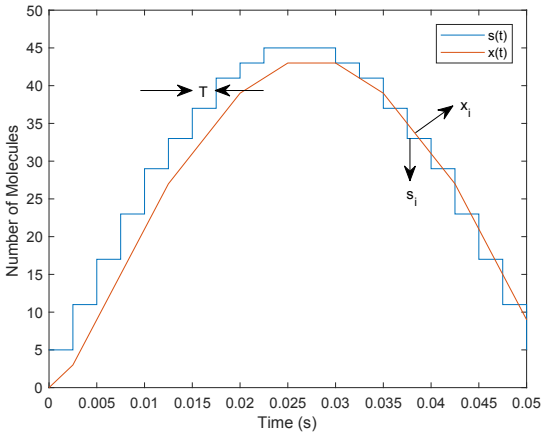


Figure 2.3. Existing signal outside the RN and the sampled signal.

analogy with the delta modulation can be established to understand the relation between $s(t)$ and $x(t)$. As illustrated in Fig. 2.3, s_i and x_i, which are measured along T, are the i^{th} samples of $s(t)$ and $x(t)$, respectively. Since $s(t)$ is assumed to be a Gaussian random process, s_i is assumed as a Gaussian random variable with $N(\mu_s, \sigma_s^2)$.

In order to define a molecular signal as the changing concentration levels around the RN, it is essential to specify a volume in which the RN is located similar to the virtual reception volume approach employed in (Atakan and Akan, 2010) as illustrated in Fig. 2.4. In this figure, V_N denotes the volume of the RN and V_R stands for the reception volume, in which the molecular signal exists. a and b are the radii of the spherical volumes

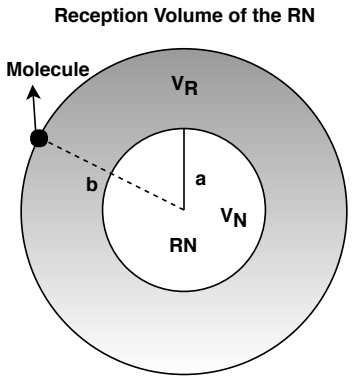


Figure 2.4. Reception volume of the RN and a single molecule.

V_N and V_R, respectively. When a molecule is received by the RN, it is consumed and not released to V_R again. The probability that the molecule emitted at the distance y from the center of the RN can reach its surface within time t can be found through the first-hitting time probability of the random walk as follows (Ziff et al., 2009; Bai et al., 2016; Yilmaz et al., 2014)

$$F(y,t) = \frac{a}{y}\text{erfc}\left(\frac{y-a}{\sqrt{4Dt}}\right), \tag{2.2}$$

where D is the diffusion coefficient, a is the radius of the RN and erfc(.) is the complementary error function. This capture can be used to find the hitting rate of the molecules to the surface of the RN. Since the capture of a molecule is an event with two possible outcomes as "capture" or "escape", x_i can be assumed as a random variable with a binomial distribution. Due to the large number of molecules around the RN, the binomial distribution can be approximated to Poisson distribution. Thus, x_i is assumed to have a Poisson distribution in our study. On the other hand, the spatial distribution of the molecules around the RN is given as follows. It is assumed that the molecules in the reception volume are uniformly distributed over the interval (a, b) where $b > a$. By using the mean distance of the molecules to the surface of the RN, i.e., $\frac{a+b}{2}$, the average number of molecules which can hit the surface of the RN within a sampling interval can be given as,

$$\lambda = F\left(\frac{a+b}{2}, T\right) s_i, \tag{2.3}$$

where $F(\frac{a+b}{2}, T)$ is abbreviated as $F(T)$ in the rest of the chapter. Here, λ is the rate of the Poisson random variable x_i depending on another random variable s_i. All the symbols and variables used in Section 2.3 and 2.4 are summarized in Table 2.1.

A Poisson process is generally employed as a counting process and it can be defined as a random process which has independent increments being Poisson distributed. The rate of the Poisson process can be generalized as the time-varying intensity function, i.e., $\lambda(t)$. When the intensity function is constant, the Poisson process is homogeneous, whereas the Poisson process is nonhomogeneous when the intensity function is time-dependent. The intensity function may be unknown in some situations and cannot be treated as constant. In such circumstances, it is reasonable to consider the intensity function as a random process. When the intensity function of the nonhomogeneous Poisson process is another random process, the Poisson process is called a Cox process or a Doubly Stochastic Poisson Process (DSPP) (Cox, 1955).

Table 2.1. Symbols and Variables used in Section 2.3 and 2.4.

Symbol/Variable	Definition
V_R	Reception volume
V_N	Volume of the RN
a	Radius of the RN
b	Radius of the reception volume
T	Sampling period
$s(t)$	Gaussian random process showing the number of the molecules in V_R
$m(t)$	Mean of $s(t)$
$v(t)$	Variance of $s(t)$
$R(t_1, t_2)$	Autocovariance of $s(t)$
$x(t)$	Doubly stochastic Poisson process showing the number of the molecules counted by the RN
$\lambda(t)$	Intensity of $x(t)$
s_i	Random variable showing the i^{th} sample of $s(t)$
μ_s	Mean of s_i
σ_s^2	Variance of s_i
λ	Rate of x_i
$F(y,t)$	The first-hitting time probability of a molecule emitted at a distance y within time t
$\mathcal{E}$	Signal distortion

In our system model, the intensity function of $x(t)$ can be found by replacing the Gaussian random variable s_i with the Gaussian random process $s(t)$ in (3). This intensity function is given by

$$\lambda(t) = \frac{F(T)}{T} s(t), \tag{2.4}$$

where the intensity function is normalized by a factor of $\frac{1}{T}$ due to the definition of the rate, which is the number of molecules hitting the surface of the RN within unit time. Since the intensity of $x(t)$ is another random process given in (2.4), $x(t)$ is a DSPP. Furthermore, $s(t)$ can be assumed as a stationary random process, since it is constant along a sampling period T.

Using the knowledge that $x(t)$ is a DSPP which is defined for $t \geq t_0$ and $s(t)$ is a stationary Gaussian random process, the mean of $x(t)$ can be given as (Valderrama et al., 1995)

$$\begin{aligned} E\{x(t)\} &= \int_{t_0}^{t} E\{\lambda(u)\} du = \int_{t_0}^{t} E\left\{ \frac{F(T)}{T} s(u) \right\} du \\ &= \frac{F(T)}{T} \int_{t_0}^{t} m(u) du, \end{aligned} \tag{2.5}$$

where $E\{.\}$ represents the expectation operator. To find the variance of $x(t)$, the autocovariance of $\lambda(t)$ is needed which is derived as

$$\begin{aligned} Cov\{\lambda(t_1), \lambda(t_2)\} &= E\left\{[\lambda(t_1) - E\{\lambda(t_1)\}]\,[\lambda(t_2) - E\{\lambda(t_2)\}]\right\} \\ &= E\left\{\frac{F(T)}{T}\,[s(t_1) - m(t_1)]\,\frac{F(T)}{T}\,[s(t_2) - m(t_2)]\right\} \\ &= \frac{(F(T))^2}{T^2} R(t_1, t_2), \end{aligned} \tag{2.6}$$

where $Cov\{.\}$ shows the covariance operator. Then, by using (2.6), the variance of $x(t)$ is given as (Valderrama et al., 1995)

$$\begin{aligned} Var\{x(t)\} &= 2\int_{t_0}^{t}\int_{t_0}^{t_2} Cov\{\lambda(t_1), \lambda(t_2)\} dt_1 dt_2 + \int_{t_0}^{t} E\{\lambda(u)\} du \\ &= \frac{2\,(F(T))^2}{T^2}\int_{t_0}^{t}\int_{t_0}^{t_2} R(t_1, t_2) dt_1 dt_2 + \frac{F(T)}{T}\int_{t_0}^{t} m(u) du, \end{aligned} \tag{2.7}$$

where $Var\{.\}$ is the variance operator. The second moment of $x(t)$ can be calculated by using (2.5) and (2.7) in the formula $E\{x^2(t)\} = Var\{x(t)\} + (E\{x(t)\})^2$ as given by

$$\begin{aligned} E\{x^2(t)\} = \frac{2\,(F(T))^2}{T^2}\int_{t_0}^{t}\int_{t_0}^{t_2} R(t_1, t_2) dt_1 dt_2 + \frac{F(T)}{T}\int_{t_0}^{t} m(u) du \\ + \left(\frac{F(T)}{T}\int_{t_0}^{t} m(u) du\right)^2. \end{aligned} \tag{2.8}$$

The statistical properties of the DSPP $x(t)$ is employed to derive the signal reconstruction distortion in the next section.

2.4. Derivation of the Signal Distortion

In this section, the signal distortion between the signal outside the RN and the reconstructed signal is derived as a Mean Square Error (MSE). The signal distortion, i.e., the MSE ($\mathcal{E}$), is given by

$$\mathcal{E} = \frac{E\{(s_i - x_i)^2\}}{V^2} = \frac{E\{s_i^2\} - 2E\{s_i x_i\} + E\{x_i^2\}}{V^2}, \tag{2.9}$$

where s_i is a Gaussian random variable with $N(\mu_s, \sigma_s^2)$ showing the number of the molecules in the volume between the outer boundary of the reception volume and the RN, $(V_R - V_N)$, x_i is a doubly stochastic Poisson variable showing the counted number of the molecules in V_N, V is the volume. Here, $\mathcal{E}$ is derived by using concentrations of the molecules inside

and outside the RN. Therefore, (2.9) includes the volume V where those molecules are located. Also in (2.9), V_R is calculated as $\frac{4}{3}\pi b^3$ and similarly V_N is calculated as $\frac{4}{3}\pi a^3$. Since the counted number of the molecules by the RN depends on the number of the molecules outside the RN, random variables x_i and s_i are dependent. The second term of the numerator in (2.9) consists of the product of these two dependent random variables. The expected value of this product can be found by

$$E\{s_i x_i\} = \rho_{s_i x_i}\sigma_{x_i}\sigma_{s_i} + E\{x_i\}E\{s_i\}, \tag{2.10}$$

where $\rho_{s_i x_i}$ is the correlation coefficient between x_i and s_i. After substituting (2.10) in (2.9), the volumes can be clarified to find the molecule concentrations by writing explicitly the volumes for each term as given by

$$\mathcal{E} = \frac{E\{s_i^2\}}{(V_R - V_N)^2} - 2\rho_{s_i x_i}\frac{\sigma_{x_i}}{V_N}\frac{\sigma_{s_i}}{(V_R - V_N)} - 2\frac{E\{x_i\}}{V_N}\frac{E\{s_i\}}{(V_R - V_N)} + \frac{E\{x_i^2\}}{V_N^2}. \tag{2.11}$$

In (2.11), the numerator of the first term can be found as $E\{s_i^2\} = \sigma_s^2 + \mu_s^2$. To find the second and the third term of the numerator in (2.9), the first and second moments of $x(t)$ is employed. $E\{x_i^2\}$ can be found as given in equation (2.12) by setting $t_0 = iT$ and $t = (i+1)T$ in (2.8) where $i \geq 0$ is an integer. Subsequently, the signal distortion becomes as given by (2.13).

$$E\{x_i^2\} = E\{x^2(t)\}\Big|_{t_0=iT}^{t=(i+1)T} = \frac{2\left(F(T)\right)^2}{T^2}\int_{iT}^{(i+1)T}\int_{iT}^{t_2} R(t_1,t_2)dt_1dt_2 + \frac{F\left(T\right)}{T}\int_{iT}^{(i+1)T} m(u)du + \left(\frac{F\left(T\right)}{T}\int_{iT}^{(i+1)T} m(u)du\right)^2. \tag{2.12}$$

$$\mathcal{E} = \frac{\sigma_s^2 + \mu_s^2}{(V_R - V_N)^2} - 2\Bigg[\frac{\mu_s}{(V_R - V_N)}\frac{F\left(T\right)}{T}\int_{iT}^{(i+1)T}\frac{m(u)}{V_N}du + \frac{\rho_{s_i x_i}\sigma_s}{(V_R - V_N)V_N}\sqrt{\int_{iT}^{(i+1)T}\int_{iT}^{t_2}\frac{2\left(F(T)\right)^2 R(t_1,t_2)}{T^2}dt_1dt_2 + \int_{iT}^{(i+1)T}\frac{F\left(T\right)m(u)}{T}du}\Bigg] + \frac{1}{V_N^2}\Bigg[\frac{2\left(F(T)\right)^2}{T^2}\int_{iT}^{(i+1)T}\int_{iT}^{t_2} R(t_1,t_2)dt_1dt_2 + \frac{F\left(T\right)}{T}\int_{iT}^{(i+1)T} m(u)du + \left(\frac{F\left(T\right)}{T}\int_{iT}^{(i+1)T} m(u)du\right)^2\Bigg]. \tag{2.13}$$

Since $s(t)$ is assumed to be a stationary random process, the mean, variance and autocovariance can be written as $m(t) = \mu_s$ and $v(t) = R(t_1,t_2) = \sigma_s^2$ for one sample

measured along T. Using this assumption, $\mathcal{E}$ can be simplified in a time interval from 0 to T as given in (2.14) which is used to obtain the numerical results in the next section.

$$\mathcal{E} = \frac{\sigma_s^2 + \mu_s^2}{(V_R - V_N)^2} - \frac{2\rho_{s_i x_i}\sigma_s\sqrt{F\left(T\right)\left[F\left(T\right)\sigma_s^2 + \mu_s\right]} - 2F(T)\mu_s^2}{V_N(V_R - V_N)} + \frac{F\left(T\right)\left[F\left(T\right)\sigma_s^2 + \mu_s + F\left(T\right)\mu_s^2\right]}{V_N^2}. \quad (2.14)$$

2.5. Validation of the Signal Distortion Function & Numerical Results

In this section, the derived $\mathcal{E}$ given in (2.14) is validated through random walk simulations. In addition, the numerical results are given and analyzed by evaluating the signal distortion function. Furthermore, the distributions of the signal outside the RN and the reconstructed signal are observed. The simulation parameters for the numerical results are given in Table 2.2 and the random walk simulation parameters are given in Table 2.3. In the simulation experiments, the signal outside the RN is calculated by dividing the number of molecules s_i to the volume between the boundary of V_R and the boundary of V_N as shown in Fig. 2.4. Similarly, the concentration received by the RN is found by dividing the received number of molecules x_i to V_N.

Table 2.2. Simulation Parameters.

Parameters	**Values**
μ_s	10^2
σ_s^2	10^2
Correlation coefficient (ρ_{sx})	0.75
Diffusion coefficient (D)	10^{-12} m^2/s
Radius of the V_N (a)	1 μm
Radius of the V_R (b)	$2 - 3$ μm
Sampling period (T)	0 - 0.25 s

The aim of the numerical result part is to observe the accuracy of the signal reconstruction of the RN for different MC system parameters, such as radius of the reception

Table 2.3. Random Walk Simulation Parameters.

Parameters	**Values**
μ_s	10^2
Diffusion coefficient (D)	10^{-12} m^2/s
Radius of the V_N (a)	1 μm
Radius of the V_R (b)	$2-3$ μm
Time (t)	$0-0.25$ s
Step time (τ)	10^{-3} s
Step length (δ)	0.0447 μm

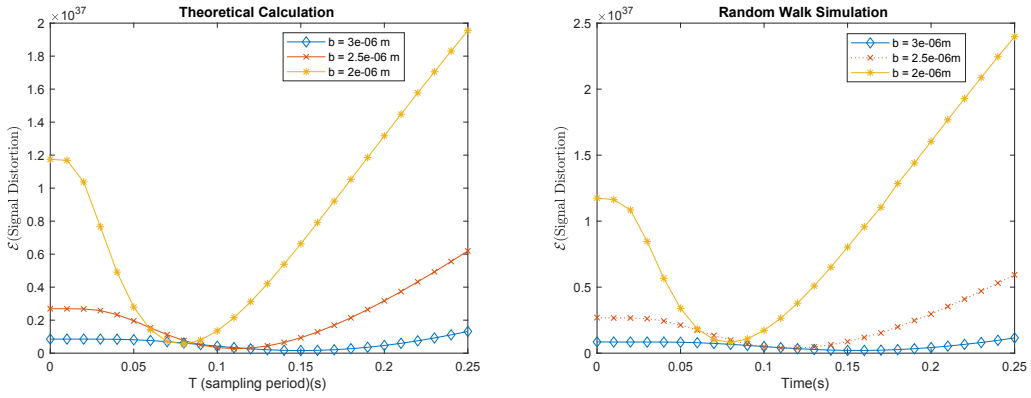
volume, sampling period and diffusion coefficient. In the light of these results, the signal reconstruction performance can be improved by adjusting the MC system parameters appropriately.

2.5.1. Validation of the Theoretical System Model

A random walk simulation is performed to show that the signal distortion function is valid. In the random walk simulations, a molecule is assumed to make a random movement in every τ seconds with a step length of δ meters on x, y and z axes separately. Every step of a molecule, whose step length can be calculated by $\delta = \sqrt{2D\tau}$, is independent from its other steps (Berg, 1993).

During the random walk simulation, the following assumptions are made. The initial distance of the molecules to the center of the RN is assumed as $(a+b)/2$ which is the midpoint of the boundaries of V_R and V_N, as given in the theoretical model. A molecule in V_R is received, when it hits the RN. After the reception of a molecule, it is not released to V_R again. Every received molecule is only counted once, since the RN is assumed to be a perfect absorber. Until the end of the random walk simulation, no additional molecule can enter the V_R. Using these assumptions, the numerical and the random walk simulation results for the signal distortion ($\mathcal{E}$) as a function of the sampling period are given in Fig. 2.5 (a) and (b), respectively. In both of the figures, the signal distortion functions have approximately the same convex structure and nearly the same minimum points. This shows that the signal distortion function given in (2.14) is valid. Small differences are observed in signal distortion values between Fig. 2.5 (a) and Fig. 2.5 (b) which can be clarified as follows. In the random walk simulation, the molecules

start walking from the midpoint between the boundaries of V_R and V_N and during the sampling period they can move out of the V_R. However, in the theoretical model, it is assumed that during the sampling period the number of the molecules is constant and the number of the molecules inside the V_R can decrease if and only if the molecules are received by the RN.



(a) (b)

Figure 2.5. Signal distortion vs. sampling period for different V_R radii (a) theoretical calculation (b) random walk simulation.

The numerical results given in this section validates the signal distortion function derived in Section 2.4 for the system model given in Section 2.3. Using this signal distortion function, the effect of the system parameters is examined in the next subsection.

2.5.2. Numerical Results

In this part, we first observe the signal distortion function for three different V_R radii as given in Fig. 2.5 (a). The different molecular signals (concentration values) are obtained by changing the radius of V_R, i.e., b, while keeping the number of molecules in V_R constant. As the V_R becomes smaller, the signal distortion values fall more steeply to their minimum values, since the hitting rate of the molecules increases in a smaller volume. This also shows that the RN can respond faster for smaller values of V_R. As the RN samples for a longer duration, it can capture more molecules. However, after reaching the minimum of the signal distortion function which gives the optimum sampling period on

the x-axis, the RN captures more molecules than it needs to calculate the valid molecule concentration around it. Therefore, the signal distortion increases after the minimum of the signal distortion function. This result reveals that the RN can increase its signal reconstruction performance by increasing the sampling period, only until the minimum signal distortion point.

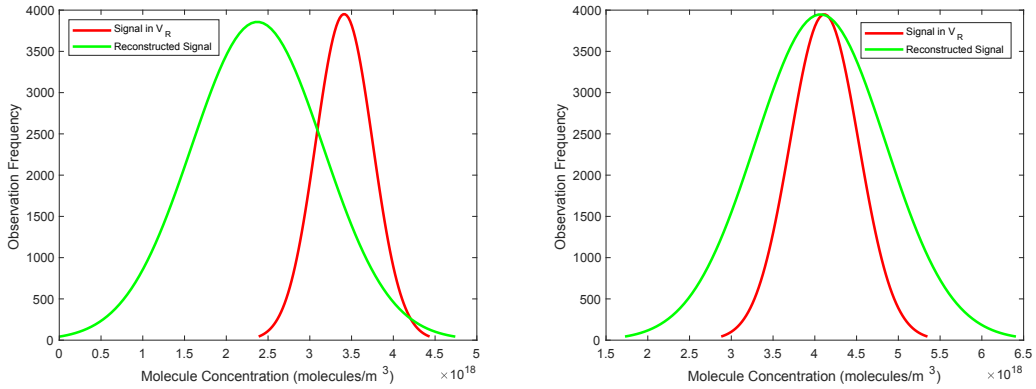


Figure 2.6. Distributions of the signal in V_R and the reconstructed signal with the parameters $D = 10^{-12}$ m^2/s, $\mu_s = 10^2$, $\sigma_s^2 = 10^2$, $b = 2$ μm (common parameters), (a) $T = 0.06$ s, $a = 1$ μm, (b) $T = 0.12$ s, $a = 1.3$ μm .

In Fig. 2.6 (a) and (b), the fitted distributions of the original signal samples, $s/(V_R - V_N)$, and the reconstructed signal samples, x/V_N, are shown for different parameters. The results are obtained as follows. First, a Gaussian distribution for s is generated. Then, by using the distribution of s, the distribution of x is calculated by generating a Poisson distribution with the rate given in (2.3) for 10^6 samples. In Fig. 2.6 (a), the signal in V_R has a Gaussian distribution with $\mathcal{N}(3.41 \times 10^{18}, 1.17 \times 10^{35})$, while the reconstructed signal has also a Gaussian distribution with $\mathcal{N}(2.37 \times 10^{18}, 6.22 \times 10^{35})$, since the distribution of the reconstructed signal converges to Gaussian distribution due to the large number of samples. The difference between the mean values of the distributions corresponds to the distortion of the signal reconstruction. Furthermore, the figure reveals that the signal reconstruction can cause errors, when an information transfer from the TN to the RN takes place. However, when the system parameters are appropriately set, the signal can be reconstructed with a small distortion as observed in Fig. 2.6 (b).

In Fig. 2.7, the effect of the diffusion coefficient on the signal distortion is illustrated. For a larger diffusion coefficient, $F(T)$ increases more rapidly. This rapid increase

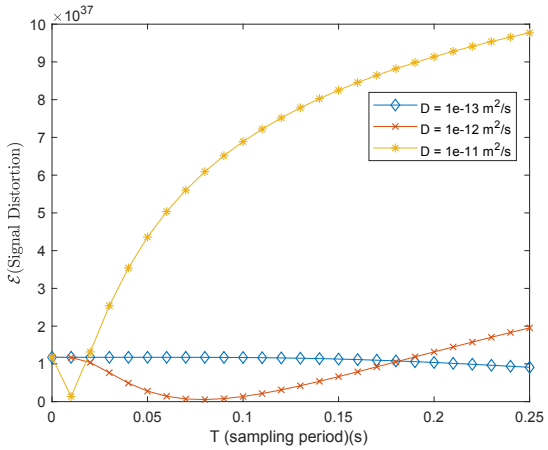


Figure 2.7. Signal distortion vs. sampling period for varying diffusion coefficients with $a = 1$ μm and $b = 2$ μm.

causes the signal distortion with the larger D to reach its minimum value more quickly. If the RN continues to count the molecules after the minimum of the signal distortion function, the concentration difference grows, since the captured molecules increase and the remaining molecules in V_R decrease. After the minimum point of the signal distortion function, the signal distortion goes up as fast as the magnitude of the diffusion coefficient , due to the faster movement of the molecules in a less dense environment. To obtain minimum signal distortion, the optimization of the RN design parameters is essential, as discussed in the next section.

2.6. Receiver Nanomachine Design

In this section, optimum parameters such as RN radius (a), sampling period (T) and sampling frequency (f) are examined for the RN design. First, we focus on the optimum sampling period, i.e., T_{opt}. The sampling period needs to be estimated in order to calculate the receiving time and the information rate of the RN.

By means of the signal distortion function given in (2.14), the optimum parameters can be derived. Since the signal distortion function is convex as observed with the numerical results in Section 2.5, T_{opt} can be found by solving the equation for T as given

by

$$\frac{\partial \mathcal{E}(T)}{\partial T} = 0. \tag{2.15}$$

When the sampling period is set as T_{opt}, the minimum signal distortion is obtained. The derivative of $\mathcal{E}$ with respect to T is given in (2.16) where $z = \frac{b-a}{4\sqrt{DT}}$. However, the derived expression of T_{opt} is a long equation and cannot be written in this chapter. Instead, the numerical comparison of the signal distortion for a constant T and T_{opt} is given in Fig. 2.8. To derive T_{opt}, an approximation such that $\text{erfc}(x) = e^{-c_1 x - c_2 x^2}$, where $c_1 = 1.09500814703333$ and $c_2 = 0.75651138383854$, is used (Tsay et al., 2013). Furthermore, it is assumed that $e^{-c_1 z-(c_2+1)z^2} \approx e^{-c_1 z-(c_2+2)z^2}$, $e^{-2c_1 z-2c_2 z^2} \approx e^{-2c_1 z-(2c_2+2)z^2}$, $e^{-2c_1 z-(2c_2+1)z^2} \approx e^{-2c_1 z-(2c_2+2)z^2}$ to solve the equation for T_{opt}. The numerical results in Fig. 2.8 show that the derived T_{opt} gives better results and validate that the sampling period is optimum, when it is compared with a constant sampling period.

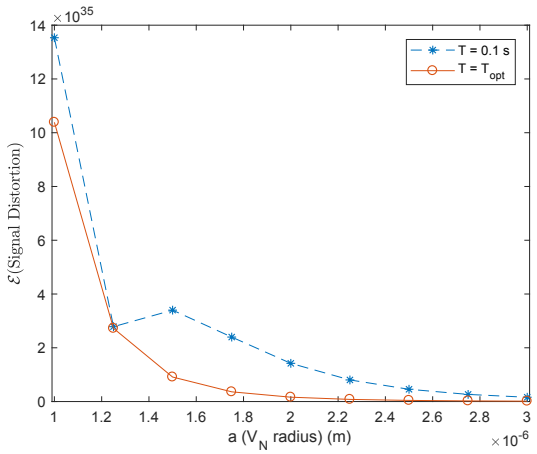


Figure 2.8. Constant and optimum sampling period comparison for $D = 10^{-12}$ m^2/s, $b = 2a$ and $\mu_s = \sigma_s^2 = 10^2$.

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial T} = &-\frac{9(b-a)D\mu_s^2 e^{-z^2}}{16a^2(a+b)(b^3-a^3)\pi^{5/2}(DT)^{3/2}} \\ &+\frac{9(b^2-a^2)D\mu_s e^{-z^2}+36a(b-a)D\mu_s^2 e^{-z^2}\text{erfc}(z)+36a(b-a)D\sigma_s^2 e^{-z^2}\text{erfc}(z)}{32(a+b)^2a^5\pi^{5/2}(DT)^{3/2}} \\ &-\frac{18\rho_{sx}\sigma_s^3 a(b-a)De^{-z^2}\text{erfc}(z)+9(b^2-a^2)D\mu_s e^{-z^2}+18a(b-a)\sigma_s^2 D\text{erfc}(z)}{32a^2\pi^{5/2}(a+b)(b^3-a^3)(DT)^{3/2}\sqrt{2a\mu_s(a+b)\text{erfc}(z)+4a^2\sigma_s^2\left(\text{erfc}(z)\right)^2}}.\end{aligned} \tag{2.16}$$

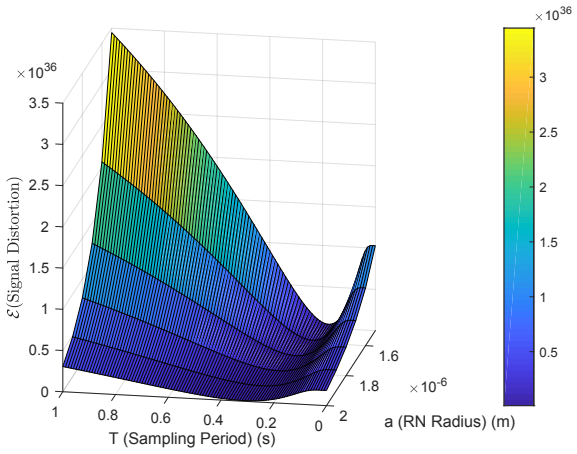


Figure 2.9. Signal distortion vs. a and T surface plot for $\mu_s = 10^2$, $\sigma_s^2 = 10^2$, $D = 10^{-12}$ m^2/s and $b = 2a$.

Another critical parameter of the RN is its radius. It is essential to choose the optimum radius for the minimum MSE signal reconstruction. Similar to T_{opt}, the optimum RN radius, i.e., a_{opt}, can be found by solving

$$\frac{\partial \mathcal{E}(a)}{\partial a} = 0. \tag{2.17}$$

If the second derivative of $\mathcal{E}(a)$ is positive, its minimum can be calculated. The first derivative of the $\mathcal{E}(a)$ with respect to a and the solution for a_{opt} are too long equations to write here. Instead, the numerical results regarding the relationship among a, T, f and the signal distortion are examined.

Assuming that $b = 2a$, the combined effect of the system parameters on the signal distortion is shown in Fig. 2.9. As the size of the RN grows, the RN needs more time to collect the sufficient number of molecules required for the desired concentration value around it and thus, to reduce the signal distortion. This stems from the fact that, while the surface of the RN increases proportionally with a^2, its volume increases proportionally with a^3. To balance this situation, the RN extends its sampling period. Regarding the T_{opt} values, as the RN radius increases, the RN needs more time to reconstruct the signal, but has a lower signal distortion due to the larger capture probability of the molecules. This shows that there is a trade-off among a, T and the signal distortion. When a larger RN size is chosen to have a lower signal distortion, a longer optimum sampling period is required. On the contrary, if the RN is desired to respond faster for reception, its cost

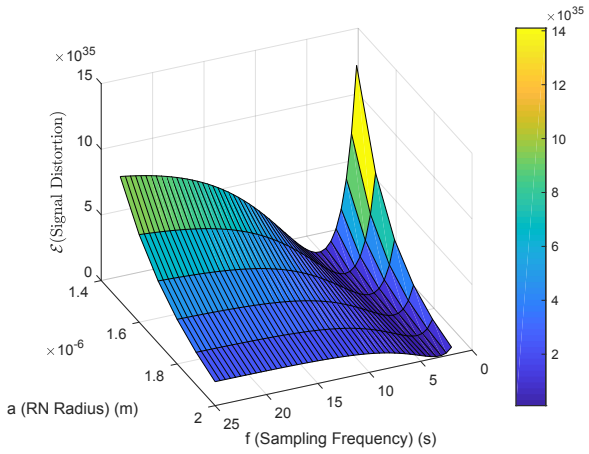


Figure 2.10. Signal distortion vs. a and f surface plot for $\mu_s = 10^2$, $\sigma_s^2 = 10^2$, $D = 10^{-12}$ m^2/s and $b = 2a$.

is a worse signal reception. Therefore, RN sampling period and its corresponding signal distortion are essential to determine the optimum RN radius, i.e., a_{opt}.

The sampling frequency, which is calculated by $f = \frac{1}{T}$, can be used to design a RN efficiently. The combined effect of the sampling frequency, RN radius and signal distortion can be seen in Fig. 2.10. An analysis similar to the sampling period can be made for the sampling frequency. As the RN radius increases, the sampling frequency and signal distortion decrease at the optimum frequency points, i.e., f_{opt}. The relation among the signal distortion, sampling frequency and the RN radius is required to be considered for the RN design. When a smaller RN is chosen, the signal distortion and f_{opt} becomes higher. Hence, the cost for a smaller RN with the minimum signal distortion is a lower quality communication and a more complex structure for a faster signal processing.

It is not always possible to design the RN according to the optimum parameters. In such cases, a signal distortion constraint can be defined. Due to the convexity of the signal distortion function, a range within the minimum and maximum values of the sampling period (or sampling frequency) can be determined for the given signal distortion constraint. Subsequently, the RN radius can be chosen according to this range of the sampling period (or sampling frequency).

CHAPTER 3

DISTANCE ESTIMATION METHODS FOR A PRACTICAL MACROSCALE MOLECULAR COMMUNICATION SYSTEM

3.1. Introduction

Beside the usage of molecular communication (MC) in microscale as given in Chapter 2, it can also be employed in macroscale (cm to m). In nature, animals like bees, flies or ants use MC to send messages over several meters with the help of the pheromones. Similar to the existing methods in nature, pheromones can be exploited for long range communication among NMs as proposed in (Giné and Akyildiz, 2009). On the other hand, the macroscale MC concept was employed in a practical MC system, which consists of an electric sprayer controlled by a microcontroller as the transmitter (TX), a metal-oxide alcohol sensor connected to a microcontroller as the receiver (RX) and alcohol molecules as the messenger molecules, to send information symbols over a few meters (Farsad et al., 2013). The channel and noise models are proposed for this macroscale MC system in (Farsad et al., 2014) and (Kim et al., 2015). It was observed that there is a nonlinearity in the channel which differs from conventional communication systems. A similar MC system is proposed to be employed in confined structural environments consisting of several pipe types. It was shown that MC provides reliable communication, whereas electromagnetic wave based communication cannot (Qiu et al., 2014). The data rate of the practical MC system can be increased by multiple input multiple output (MIMO) MC system configuration as proposed and experimentally shown in (Lee et al., 2015) and (Koo et al., 2016). In these two papers, an air compressor is used instead of a fan to emit molecules. In (Zhai et al., 2018), a method to mitigate inter-symbol interference is proposed for an experimental MC system similar to the platform given in (Farsad et al., 2013), except the fan behind the TX. (McGuiness et al., 2018) proposes to employ

an odor generator as the TX and a mass spectrometer with a quadrupole mass analyzer as the RX for macroscale MC.

The macroscale experimental studies focus mostly on channel modeling which does not agree with theoretical studies in the literature. However, except the propagation delay, these channel models do not clearly state the relationship among the physical parameters such as the distance between the TX and RX or the number of released molecules by the TX. By knowing the distance before the communication starts, a higher communication rate can be achieved via arranging the parameters such as the number of released molecules (Atakan and Akan, 2007; Eckford, 2007; Nakano et al., 2013; Pierobon and Akyildiz, 2013). Furthermore, the deployment of the receivers can be arranged autonomously in a nanonetwork with the distance information known by the RX. The distance estimation can also be used to find the location of a molecular source in various environmental monitoring applications. Therefore, it is important to estimate the distance between the TX and RX, which is the main purpose of this chapter.

The first proposed distance estimation protocols are based on two-way transmission. In (Moore et al., 2010), four distance estimation protocols are proposed which are based on measuring the round trip time (RTT) and signal fading in amplitude or frequency for a 1-D diffusion channel. In these protocols, the TX transmits a signal and the RX transmits a feedback signal with a different type of molecule, when the transmitted signal is received. Then, the TX estimates the distance, when it receives the feedback signal. These protocols are expanded in (Moore et al., 2012). The RTT protocol is improved for a 2-D diffusion channel with a more realistic microscale MC model in (Moore and Nakano, 2012). Since using feedback signals is time consuming and increases interference in the communication channel, it is reasonable to estimate the distance at the RX side with a single transmission. Accordingly, two distance estimation schemes with a one-way transmission are proposed in (Huang et al., 2013). Here, the RX measures the received peak concentration or the time interval between the first and second peaks to estimate the distance. These schemes require less time with respect to two-way transmission protocols, but they are derived for a 1-D diffusion channel, and the estimation accuracy is not improved significantly. In (Noel et al., 2014a), a lower bound for distance estimation accuracy is derived for the diffusion channel without any flow and initial drift. In another study, two distance estimation schemes, where emission time is considered as a parame-

ter, are proposed for a 3-D scenario in a diffusion channel (Wang et al., 2015b). The first scheme uses the peak time of the received signal, and the second scheme, which gives more accurate results, uses the received energy to estimate the distance. The same authors propose an algorithmic distance estimation scheme and two parameter optimization methods in (Wang et al., 2015a) for a 3-D diffusion channel. When compared with (Wang et al., 2015b), this scheme has a worse distance estimation performance, requires a more complex receiver and needs more time for estimation. In (Lin et al., 2019), a distance estimation protocol is proposed to estimate the distance in a 3-D diffusion channel in the presence of additive noise and inter-symbol interference by using maximum likelihood estimation. This proposed method has a high accuracy with a cost of high complexity.

All of the distance estimation methods in the literature consider an ideal microscale channel model where the transmitted molecules do not have an initial velocity, the molecules move according to Brownian motion and the TX transmits and the RX receives the signals perfectly. However, this channel model is not realistic for macroscale MC. Furthermore, there is no distance estimation method for macroscale situations in the literature. In this chapter, we propose five distance estimation methods for a practical macroscale scenario, where molecules propagate indoors with an initial velocity and without any constant flow in a 3-D medium, i.e., air. Three of the proposed methods are novel data analysis based methods, and two of them include supervised machine learning (ML) methods. To collect data for the purpose of using them in these methods, an experimental setup similar to the tabletop platform in (Farsad et al., 2013) is employed. In this setup, an electric sprayer without a fan is the TX that emits alcohol molecules and a metal-oxide alcohol sensor is the RX. The TX and RX are controlled with a microcontroller via a computer to record the received signals. A novel algorithm is proposed to extract features from the measured signals by the RX. For the first time, the distance estimation is made with supervised ML methods for a practical macroscale MC system. Multivariate linear regression (MLR) and neural network regression (NNR) are used as ML methods which use the extracted features as inputs. The experimental data are used as the training and test data of these methods. Afterwards, the collected data are analyzed and three distance estimation methods are proposed which are less complex but less accurate than ML methods. The first data analysis based method is power based distance estimation, which employs the relation between the distance and the received power to transmitted power ratio. In

the second data analysis based method, which is the peak time based distance estimation, the relation between the peak time of the received signal and the distance is exploited. At long distances (170 cm and longer), this method has the best results among the data analysis based methods. Combined distance estimation is the third method, which uses a combination of the power based and peak time based estimation. Furthermore, all of these applied methods are compared by discussing the numerical results. It is shown that ML methods perform better than the data analysis based methods.

The existing theoretical models which are based on the diffusion of the molecules are insufficient to explain the experimental results. Actually, the TX emits liquid droplets instead of molecules. During the propagation of the droplets, their size change dynamically, they are affected by the Newton's laws of motion and interact with the air molecules. As detailed in Section 2.2, it is difficult to consider all of these effects for deriving an analytical expression to estimate the distance. This gives us the motivation to employ ML and data analysis based methods for distance estimation. Moreover, an analysis is made on the experimental data. Experimental findings show that the movement of the molecules in a practical scenario is also affected by some flows accompanying diffusing molecules. In particular, the analyses reveal that unsteady flows can affect the propagation of the molecules in addition to the diffusion, even if there is no fan behind the TX. Correspondingly, a possible trajectory of the transmitted molecules in the communication channel is given. In addition, our analysis based on experimental findings indicate that the nonlinear characteristic of the sensor can cause measurement errors at lower distances. The main result of these analyses is that a fluid dynamics perspective is needed for channel modeling and parameter estimation in MC for macroscale practical scenarios. Also, open research directions are indicated by analyzing the numerical results.

The rest of the chapter is organized as follows. In Section 3.2, the experimental setup, which is used for data collection, is explained. Section 3.3 gives the motivation for using the ML and data analysis based methods. Data collection procedure is explained in Section 3.4. The proposed feature extraction algorithm is given in Section 3.5. The supervised ML methods, which are applied to estimate the distance in this chapter, are briefly introduced in Section 3.6. In Section 3.7, the proposed distance estimation methods based on data analysis are given. In Section 3.8, the numerical results for the implemented methods are presented and compared. Moreover, received signals and the

motion of the molecules are analyzed with a fluid dynamics perspective. The study given in this chapter was first published in (Gulec and Atakan, 2020a).

3.2. Experimental Setup

In this section, the experimental setup of a practical macroscale MC system is explained in order to collect data for distance estimation. This system consists of a transmitter (TX), a receiver (RX) and a propagation channel similar to the tabletop molecular communication system in (Farsad et al., 2013). However, unlike (Farsad et al., 2013), there is no fan in our system. TX transmits molecular signals by spraying ethyl alcohol molecules having an initial velocity with an Instapark electric sprayer. This sprayer has its own battery in it and can be controlled with a microcontroller via a custom switch circuit. The RX receives the molecular signals with an MQ-3 alcohol sensor which gives the best performance among the low-cost metal oxide alcohol sensors (Farsad et al., 2013). The TX and RX are both controlled with an Arduino Uno microcontroller board which is connected to a computer. The block diagram of the system is given in Figure 3.1. As shown in this diagram, only one microcontroller board and one computer is used for simplicity and perfect synchronization between the TX and RX.

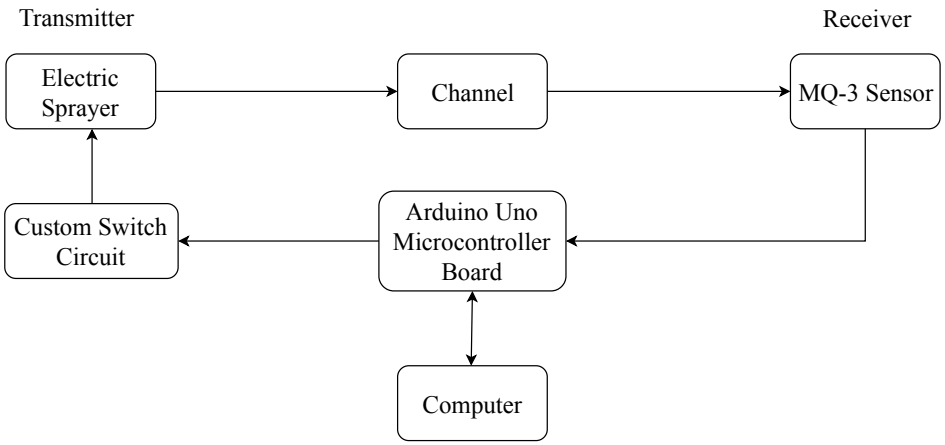


Figure 3.1. Block diagram of the experimental setup.

The TX is controlled with a computer via an Arduino microcontroller board so that the spraying duration, i.e., emission time (T_e), can be arranged. The TX and RX were

100 cm high from the ground and the experiments were carried out in a laboratory. The molecule concentration in the environment is sensed by the MQ-3 sensor and analog voltage values corresponding to the measured molecule concentration are sent to the Arduino microcontroller board. This analog signal is digitized with a 10-bit analog to digital converter by the microcontroller board. Finally, the received signal is sent to the computer from Arduino microcontroller board to be recorded. The components of the TX and RX can be seen in Figure 3.2. The motivation of our proposed distance estimation methods in this chapter for this experimental setup is given in the next section.

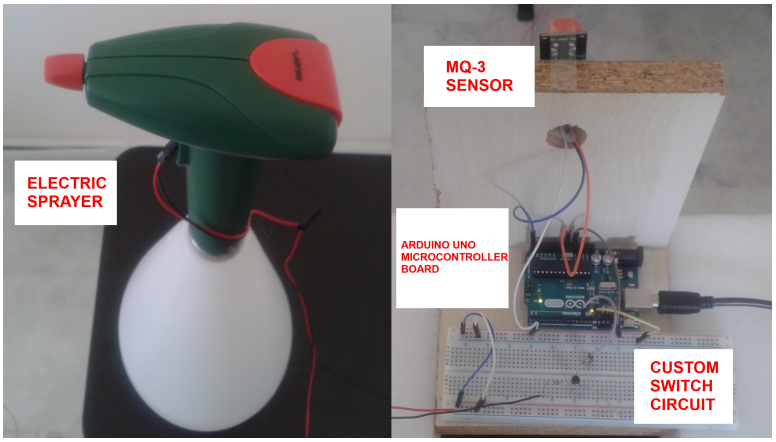


Figure 3.2. Components of the transmitter (left) and the receiver (right).

3.3. Motivation

In this section, the motivation of using ML and data analysis based methods for distance estimation in a practical macroscale scenario is given. The distance estimation methods in the literature are based on the theoretical models considering that the transmitted molecules only diffuse in the medium with or without a constant flow. However, these models are not realistic for our scenario due to the challenges given as follows:

- The liquid ethanol is sprayed as droplets rather than molecules (Rothe and Block, 1977; Ghosh and Hunt, 1994).
- The diameter of a spherical droplet ($100-800$ μm (Al Heidary et al., 2014)) is much

larger than the diameter of an ethanol molecule (0.44 nm (Carmo and Gubulin, 1997)). Therefore, it is not appropriate to characterize the motion of transmitted droplets as the Brownian motion (Al Heidary et al., 2014; De Cock et al., 2017).

- After the transmission of the droplets, they break up into smaller droplets (Ghosh and Hunt, 1994). Furthermore, the ethanol droplets can evaporate in the MC channel. Hence, the diameters of the droplets change as they move through the air (Al Heidary et al., 2014).

- The evaporation of the droplets generates diffusing molecules. These molecules are subject to boundaries for an indoor practical scenario. Moreover, it is very difficult to determine whether the indoor boundaries with various geometries are absorber or reflector.

- The transmitted droplets start their movement with an initial velocity. The interaction of these droplets with the air molecules cause unsteady flows, e.g., turbulent flows, which cause the velocity of droplets fluctuate randomly (Begg et al., 2009; Maxey, 1987) (see Section 3.8.3.2).

In our case, the droplets are sprayed with an initial velocity and potential energy due to its height (100 cm) from the ground to reach the RX at least 100 cm away from the TX on the horizontal axis. The initial energy is converted into kinetic energy and the air resistance (or aerodynamic drag) decreases the velocity of the droplets. Furthermore, the gravity affects the vertical positions of the droplets. Here, a mechanistic model from the fluid dynamics literature is applied to show the motion of the spherical liquid droplets for different sizes in 2-D without considering dynamically changing droplet size, the evaporation of droplets and unsteady flows. This model is based on the Newton's second law of motion and calculates the change in the velocity as given by (De Cock et al., 2017)

$$\frac{\partial v_x}{dt} = -\frac{3C_D\rho_a v_x^2}{4\rho_d a} \tag{3.1}$$

$$\frac{\partial v_z}{dt} = \frac{3C_D\rho_a v_z^2}{4\rho_d a} - g, \tag{3.2}$$

where v_x and v_z are the velocity of the droplet in x and z axes, respectively, a is the droplet diameter, ρ_a and ρ_d are the densities of the air and droplet, respectively, g is the gravity acceleration and C_D is the drag coefficient which is related with the air resistance. This

coefficient is given as $C_D = 24/Re + 1.5$ (Barati et al., 2014) where Re is the Reynolds number. Re is defined as $Re = \frac{v\,a}{\nu_a}$ where v is the velocity of the droplet and ν_a is the kinematic viscosity of air. The droplets are assumed to start their movement as parallel to the ground with the same initial velocity. For the numerical results, the initial velocity on the horizontal axis is given as $v_x(0) = 10$ m/s which is used as an average velocity value for the droplets in (De Cock et al., 2017). The positions of the droplets in x and z axes using this model are given in Figure 3.3. The numerical values for the experimental parameters are given in the caption of Figure 3.3.

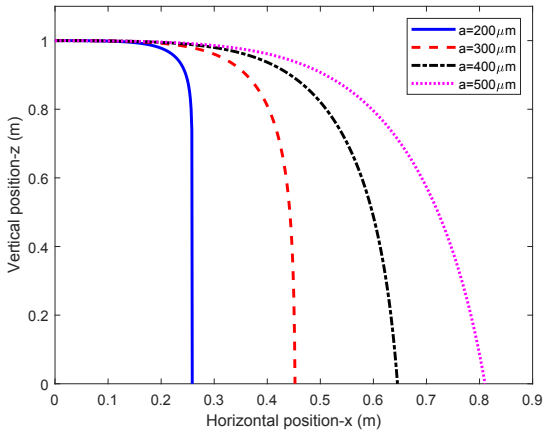


Figure 3.3. Trajectory of different-sized droplets in x and z-axis by using Newtonian mechanics with the parameters $v_x(0) = 10$ m/s, $v_z(0) = 0$ m/s, $\rho_a = 1.2$ kg/m^3, $\rho_d = 789$ kg/m^3, $\nu_a = 1.516 \times 10^{-5}$ m^2/s, $g = 9.81$ m/s^2.

In this figure, larger droplets travel a larger distance, since they decelerate less than smaller droplets. This figure also illustrates that droplets with different sizes are affected by the gravity. However, this model is not sufficient for all of the droplets which arrive at the RX for a distance of at least 100 cm, since unsteady flows, changing size and evaporation of droplets are not considered.

For distance estimation, it is highly difficult to derive an analytical expression, which can take all of the aforementioned effects into account and this derivation is beyond the scope of this chapter. Therefore, it is reasonable to use the alternative practical methods such as data analysis, linear regression and neural networks for estimating the distance accurately. With these methods, these effects on the collected signals are implicitly modeled and employed for distance estimation. Next, the experimental procedure is

explained to collect data and implement distance estimation methods.

3.4. Data Collection

This section provides detailed information on the collection of data by using the experimental setup given in Section 3.2 to be used for the proposed distance estimation methods. When the TX transmits a signal, the molecules begin to propagate in the air with an initial velocity supplied by the TX. When the TX stops transmitting, the RX starts to record the molecular signal for 100 seconds for distances up to 180 cm and 300 seconds for longer distances. During the experiments, only one single puff is transmitted from the TX to estimate the distance for one measurement. The transmissions are repeated for three different emission times as $T_e = \{0.25, 0.5, 0.75\}$ s and eleven different distances between 100 cm and 200 cm with 10 cm steps. The parameter ranges and replication numbers for the experiment are given in Table 3.1. A break was taken for at least 5 minutes between consecutive measurements. During this break, the laboratory is ventilated by opening the door and the window to reduce the molecule concentration level inside the laboratory. By the help of this “break and ventilate" procedure, the effect of the remaining molecules from previous transmissions is minimized. Without applying this procedure, the noise level increases and thus, it gets more difficult to detect the signal by the RX side. In order to estimate the distance between the TX and RX, it is essential to extract the features from the signals to employ them as inputs to ML methods, as discussed in the next section.

Table 3.1. The experimental parameters and their ranges.

Parameter	**Value**
T_e	{0.25, 0.5, 0.75} s
d	{100, 110, 120,..., 190, 200} cm
Replication for each T_e in every distance	10
Replication for each distance	30
Number of total measurements	330

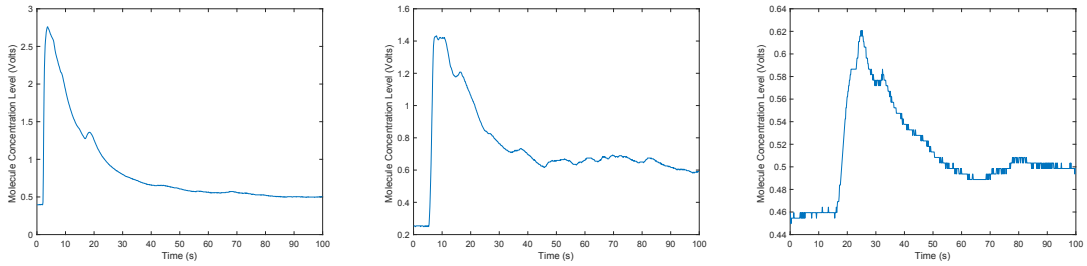
3.5. Feature Extraction

In this section, a novel feature extraction algorithm is proposed. This algorithm processes the received molecular signals to produce features, which can be defined as the input variables characterizing the output variable to be estimated for ML algorithms (Christopher, 2016). The variables which are used in Sections 3.5 and 3.7 are summarized in Table 3.2.

Table 3.2. Summary of the variables used in Sections 3.5 and 3.7.

Variable	**Definition**
T_e	Emission time
d	Actual Distance
$\hat{d}$	Estimated distance
A_o	Initial offset level
A_{thr}	Threshold amplitude
K	Detection threshold
$C[n]$	Measured molecule concentration
$y[n]$	Output of the moving average filter
W_1	The number of samples before the n^{th} sample of the moving average filter
W_2	The number of samples after the n^{th} sample of the moving average filter
t_{peak}	Peak time
C_{peak}	Peak molecule concentration
t_{low}	Time showing the 10% reference point on the rising edge of the measured signal
C_{low}	Molecule concentration showing the 10% reference point on the rising edge of the measured signal
t_{high}	Time showing the 90% reference point on the rising edge of the measured signal
C_{high}	Molecule concentration showing the 90% reference point on the rising edge of the measured signal
R	Rise time on the rising edge of the measured signal
ΔC	Molecule concentration level difference between C_{high} and C_{low}
G	Gradient on the rising edge of the measured signal
N_R	Number of the received signal samples up to t_{peak}
E_R	Received energy up to t_{peak}
P_R	Received power up to t_{peak}
P_T	Transmitted power
$\overline{P_R}$	Average received power
$\overline{P_T}$	Average transmitted power
a_1, b_1	Curve fitting parameters of power based estimation
$\overline{t_{peak}}$	Average peak time
a_2, b_2	Curve fitting parameters of peak time based estimation

The choice of the features is important, since it affects the performance of the ML algorithms. Eight features which are t_{low}, C_{low}, R, ΔC, G, t_{peak}, C_{peak} and E_R are extracted from the collected data. Three of the measured signals at 100, 160 and 200 cm are shown in Figure 3.4 in order to justify the choice of the selected features. In this figure, the delay time, which is defined as the beginning time of the detection of the transmitted molecules, increases, as the distance increases. t_{low} gives a more reliable information about the delay time due to the fluctuations of the signal. C_{low} also decreases, as the distance increases. Since the signal attenuates as the distance increases, the peak points (t_{peak}, C_{peak}) and the received energy (E_R) of the signals contain information about the distance. Furthermore, as the distance increases, the slope of the signals' rising edge decreases. This shows that R, ΔC and G are informative features. To the best of our knowledge, the proposed feature extraction algorithm which is summarized in Figure 3.5, is the first one that processes the real measured molecular signals.



(a) (b) (c)

Figure 3.4. Received signal at (a) $d = 100$ cm , (b) $d = 160$ cm (c) $d = 200$ cm.

As the first step of the feature extraction algorithm, A_o, which is defined as the initial offset level characterized by the average of the measured sensor voltage, is determined. A_o is calculated by averaging the first p samples of the signal, before the transmitted molecules start to be sensed by the RX. In order to avoid considering the transmitted molecules by the TX, p is selected as 5 empirically for our experimental scenario. The molecule concentration level is measured as the voltage values received from the MQ-3 sensor. A_o corresponds to the average molecule concentration level which is measured before the transmitted signal arrives. After A_o is determined, the detection threshold is given by the equation

$$K = A_o + A_{thr}, \tag{3.3}$$

where A_{thr} is set to 0.1 volts. The value of A_{thr} is chosen sufficiently high according to our observations in order to detect the transmitted signal. The derivation of a selection criterion for A_{thr} requires to handle the related detection theory together with the probability distributions of the observed molecular signals. This is beyond the scope of this chapter.

Subsequently, smoothing the received signal is needed due to the fluctuations in the signal originated from the random movements of the molecules. The signal is smoothed by filtering with a moving average filter whose input-output relationship is given by (Oppenheim, 1999)

$$y[n] = \frac{1}{1 + W_1 + W_2} \sum_{k=-W_1}^{W_2} C[n-k]. \tag{3.4}$$

Here, $(W_1 + W_2 + 1)$ gives the window size which is chosen as 7 by setting $W_1 = W_2 = 3$. This window size is chosen empirically in order to eliminate the ripples on the signal without changing its general shape. Increasing the window size makes the signal smoother. However, if the window size is greater than 7, the values of the extracted features begin to diverge from their original values.

Subsequent to smoothing operation, the detection block is accomplished. In this block, a decision is made whether the signal is above the predetermined K or not. If there is no detection, then the RX waits for the new signal. Otherwise, the first peak of the signal is detected. According to our observations, there can be more than one peak in the received signal, after the redundant peaks are eliminated by the smoothing filter. The number of these peaks increase as the distance increases, since the variance of the time distribution that the molecules take to reach the sensor is getting larger. The first peak time is considered as the reference time to calculate the features, since it is assumed that the majority of the molecules arrive at the RX until the first peak time. In order to find the first peak point of the signal, the first derivative of the signal's samples is taken. The first negative sample index at the point where the sign of the first derivative changes from positive to negative values gives the peak of the signal. After that, the nearest local minimum point before the first peak of the signal is detected. The same method to find the peak point of the signal can be used to find the first minimum point of the signal. When the signal is inverted, the first peak point gives the first minimum point of the signal. The features extracted from the first peak of the signal are shown in Figure 3.6. Here, the

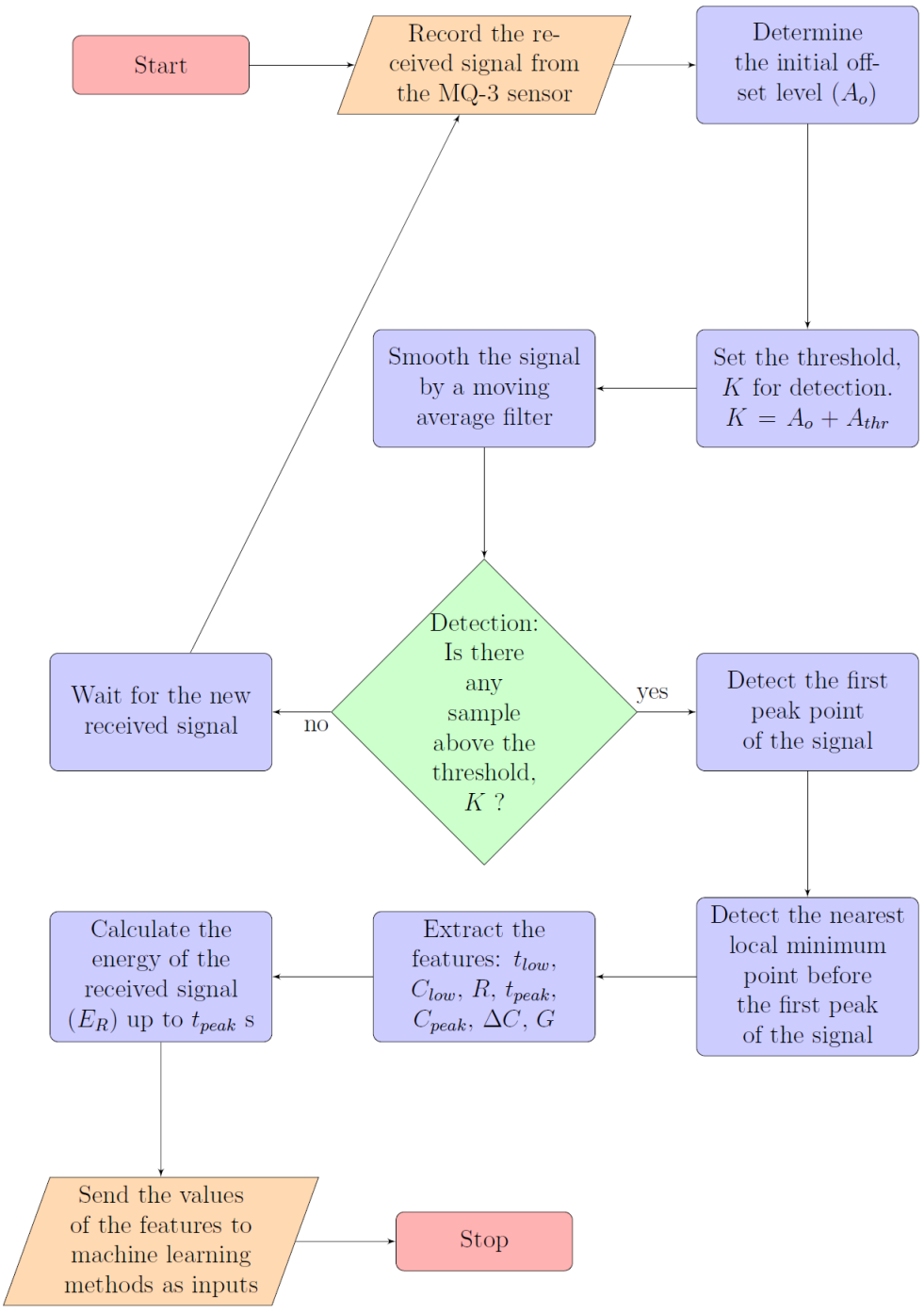


Figure 3.5. Feature Extraction Algorithm.

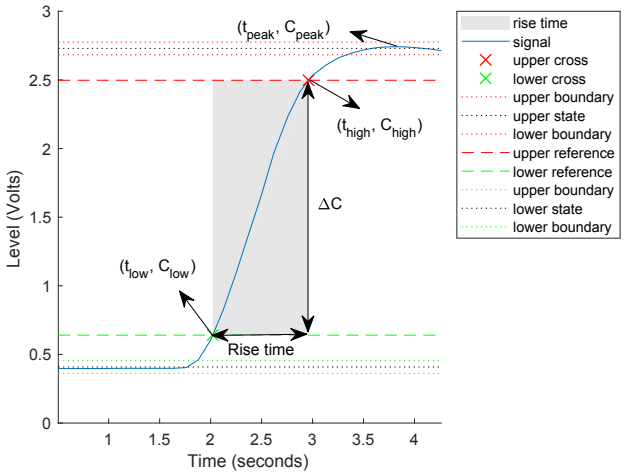


Figure 3.6. Extracted features from the received signal.

peak time and molecule concentration values are recorded as t_{peak} and C_{peak}, respectively. Afterwards, the part of the signal which is between the peak and minimum points is treated as the rising edge of a pulse. The positive-going bilevel waveform consisting of high and low levels is obtained to extract the information which can change according to distance. As illustrated in Figure 3.6, 10% reference point of the signal gives two features as t_{low} and C_{low} and similarly 90% reference point of the signal is chosen to represent t_{high} and C_{high}. These reference points are found by calculating the points at 10% and 90% of the time indices for the time interval between the peak and minimum points. The rise time (R) of this waveform is calculated by measuring the time between t_{low} and t_{high}. Another extracted feature for this signal is ΔC which is defined as the difference in the amplitude level during the rise time measurement points.

The seventh feature is the gradient, which is derived by $G = \Delta C/R$. G gives the change during the rise time, which can be discriminating according to different distances. Finally, the received energy up to t_{peak} is calculated by the formula

$$E_R = \sum_{n=1}^{N_R} |C[n] - A_o|^2. \tag{3.5}$$

In this formula, the received molecule concentration level is decreased by the existing noise level to eliminate its effect. Furthermore, the energy is found up to t_{peak}, since the time-dependent function of the concentration values has a long tail and it gets much

longer, as the distance increases. In the next section, the ML algorithms that is used in this chapter is introduced.

3.6. Supervised Machine Learning Methods for Distance Estimation

ML has a key role in science and engineering, since it facilitates the estimation of the desired outputs without changing the system model for changing inputs. It provides the ability to improve the estimation models adaptively which is called “learning". When the inputs and outputs are known during the training of the ML system, it is called “supervised" learning. In supervised ML, classification and regression are two ways of estimation where the former tries to estimate the discrete results for a discrete output and the latter deals with the estimation of the results of a continuous output. In this chapter, ML techniques, such as MLR and NNR are used for the first time to estimate the distance between a TX and an RX in a MC system. Next, the mentioned ML methods are introduced briefly.

3.6.1. Multivariate Linear Regression

MLR is a simple method to estimate the output according to the model equation given below

$$\hat{d} = \theta_0 + \sum_{j=1}^{m} \theta_j x_j, \tag{3.6}$$

where θ_0 is the bias, m is the number of the features, x_j's and θ_j's represent the features and the unknown coefficients, respectively (Friedman et al., 2008). (3.6) can also be expressed in matrix form as

$$\hat{\mathbf{d}} = \mathbf{x}\boldsymbol{\theta}, \tag{3.7}$$

where $\mathbf{x}$ is an $N \times (m+1)$ matrix of features, $\boldsymbol{\theta}$ is a column vector of $(m+1)$ elements representing the coefficients, $\hat{\mathbf{d}}$ is a column vector of N elements showing the estimated distance and N is the number of samples. Each row of $\mathbf{x}$ shows one sample having an extra 1 as the first element of each row. N represents the number of training samples for the training period and the number of test samples while the distance is estimated using

(3.7). The model coefficients are chosen during the training period to minimize the cost function given by the formula below

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(d_i - \hat{d}_i)^2, \tag{3.8}$$

By using the least squares method in the training period, the coefficients are determined by

$$\boldsymbol{\theta} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{d}, \tag{3.9}$$

where $\mathbf{d}$ is a column vector with N elements showing the actual distances for training. After finding the coefficients by the training according to (3.9), the distance can be estimated by (3.7) where N shows the number of test samples.

3.6.2. Neural Network Regression

NNR is a ML method inspired by the neurons in the brain. It has a layered structure which includes an input layer, one or more hidden layers and an output layer. These layers consist of nodes like neurons connecting the input layer to the output layer with weights (Friedman et al., 2008). As illustrated in Figure 3.7, the weights (Θ's) are arranged as the elements of a function that maps the input values (x_i) to an output value, which is the estimated distance ($\hat{d}$) in our case, with minimum error.

The relation between the input layer and the hidden layer (a_j) can be given by

$$a_j^{(k)} = g\left(\sum_{i=0}^{n}\theta_{ji}^{(k)}x_i\right), \tag{3.10}$$

where $\theta_{ji}^{(k)}$ shows the weight that connects the i^{th} node to the j^{th} node between the layers (k) and $(k+1)$, $a_j^{(k)}$ is the j^{th} hidden node in the k^{th} layer, x_i is the input values defined from 0 to n and $g(.)$ is the activation function. The relation between the hidden layer and the output layer can be defined as

$$\hat{d} = g\left(\sum_{j=0}^{p}\theta_{ji}^{(k)}x_i\right), \tag{3.11}$$

where the number of the hidden nodes in the hidden layer is defined from 0 to p. Here, the input and hidden nodes with the subscript "0" are defined as the bias units. In this

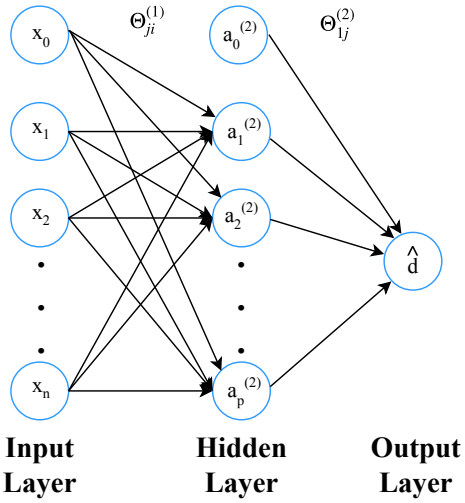


Figure 3.7. General structure of a neural network for regression with one output node and a single hidden layer.

chapter, the hyperbolic tangent sigmoid function is used as the activation function which is given by

$$g(z) = \frac{2}{1 + e^{-2z}} - 1. \tag{3.12}$$

The training of the neural network can be made with backpropagation algorithms. In these algorithms, the weights are initialized generally with random values and the output of the network is calculated in the forward direction. Then, using the output and the values of the hidden nodes, the errors are calculated in the reverse direction and the weights are arranged iteratively to obtain minimum error between the estimated and actual output values. In this chapter, Levenberg-Marquardt (LM) backpropagation algorithm, which is one of the most popular and fastest algorithms, is employed to train the neural network (Hagan and Menhaj, 1994). Before giving the update mechanism of the LM algorithm, some definitions are made as follows. A pattern is defined as the number of the input-output pairs of the neural network. Each different path from the input layer to the output layer gives a different pattern which is indexed from 1 to r. The training error for the i^{th} pattern is given by

$$e_i = r_i - o_i, \tag{3.13}$$

where r_i is the desired output and the o_i is the trained output of the neural network for the

i^{th} pattern. Moreover, the error vector (e) is given as

$$\mathbf{e} = \begin{bmatrix} e_1 & e_2 & \dots & e_r \end{bmatrix}^T, \tag{3.14}$$

where e_i is the training error for the i^{th} pattern. Similarly, the sum of the squares function is given by

$$E(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{r} e_i^2, \tag{3.15}$$

where $\boldsymbol{\theta}$ is the weight vector of N elements representing all of the weights in the neural network. The LM algorithm updates these weights to converge the minimum training error by minimizing $E(\boldsymbol{\theta})$. For this operation, Jacobian matrix is used which can be given as

$$\mathbf{J} = \begin{bmatrix} \frac{\partial e_1}{\partial \theta_1} & \frac{\partial e_1}{\partial \theta_2} & \cdots & \frac{\partial e_1}{\partial \theta_N} \\ \frac{\partial e_2}{\partial \theta_1} & \frac{\partial e_2}{\partial \theta_2} & \cdots & \frac{\partial e_2}{\partial \theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_r}{\partial \theta_1} & \frac{\partial e_r}{\partial \theta_2} & \cdots & \frac{\partial e_r}{\partial \theta_N} \end{bmatrix}, \tag{3.16}$$

where θ_i represents the weights defined from 1 to N. The weight update rule of the LM algorithm in matrix form is given by

$$\boldsymbol{\theta}_{l+1} = \boldsymbol{\theta}_l - \left(\mathbf{J}_l^T \mathbf{J}_l + \mu \mathbf{I}\right)^{-1} \mathbf{J}_l \mathbf{e}_l, \tag{3.17}$$

where l is the index of iterations, $\mathbf{I}$ is the identity matrix, μ is a positive coefficient related with the learning rate.

In LM algorithm, $E(\boldsymbol{\theta})$ is computed and checked to see, if it decreases. If $E(\boldsymbol{\theta})$ decreases, then μ is reduced by a step size γ which is a user defined parameter. Otherwise, μ is increased by γ. The algorithm continues to iterate until it reaches to a predetermined threshold value for $E(\boldsymbol{\theta})$ (Hagan and Menhaj, 1994). After the weights are determined by the training with LM algorithm, the distance can be estimated by using (3.11). Next, novel distance estimation methods based on data analysis are introduced.

3.7. Distance Estimation Methods Based on Data Analysis

The features extracted by the proposed algorithm in Section 3.5 are employed to derive novel distance estimation methods in this section. For the first method, the relation among the average received power, average transmitted power and the distance

are exploited. Next, the average peak time and distance relation is used for the second estimation method. Finally, these two methods are combined to derive the third distance estimation method.

3.7.1. Power Based Distance Estimation

By using the received energy up to t_{peak}, which is calculated in Section 3.5, the received power (P_R) is calculated with the formula

$$P_R = \frac{E_R}{N_R} = \frac{1}{N_R} \sum_{n=1}^{N_R} |C[n] - A_o|^2. \tag{3.18}$$

P_R is averaged over 10 measurements for each T_e, i.e., 0.25 s, 0.5 s and 0.75 s, at each distance. Hence, we obtain three average power ($\overline{P_R}$) values according to three different T_e's for each distance.

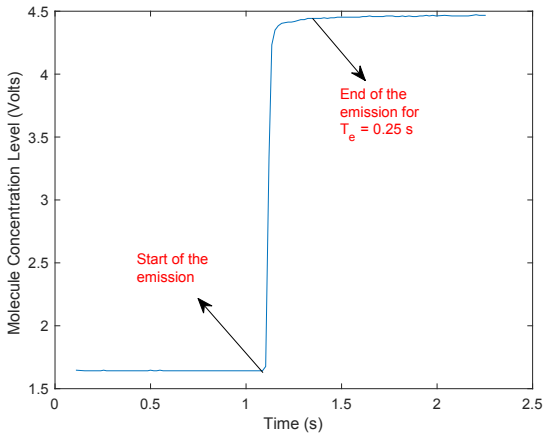


Figure 3.8. Transmitted signal for $T_e = 0.25$s.

On the other hand, the transmitted power (P_T) is measured and averaged over 4 measurements for three different values of T_e to obtain the average transmitted power ($\overline{P_T}$). The measurements are made by placing the sensor to a very close proximity of the TX, which is assumed to generate a rectangular molecular pulse within T_e. The measured transmitted signal for 0.25 s emission time, is shown in Figure 3.8. $\overline{P_T}$ values are 7.3598, 9.3666 and 11.0108 W for 0.25, 0.5 and 0.75 s emission times, respectively. Afterwards,

$\overline{P_R}$ and $\overline{P_T}$ values are employed to derive a relation with the distance between the TX and RX. Figure 3.9 shows that there is a decreasing exponential relation among these parameters for three different emission times.

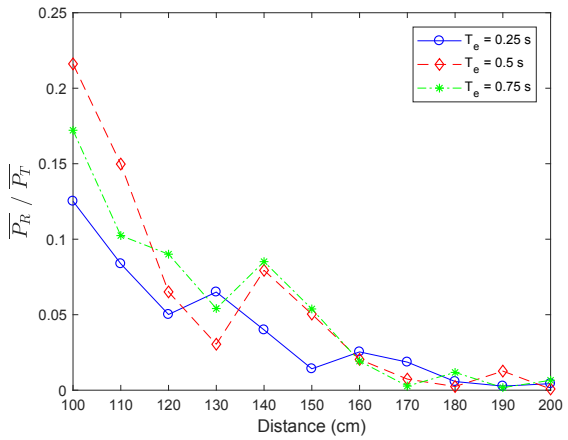


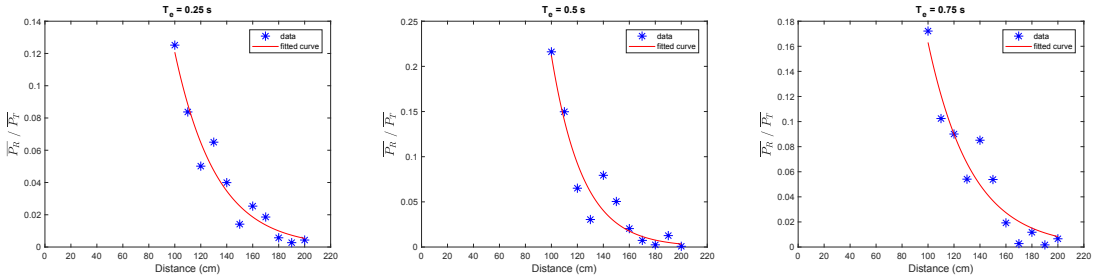
Figure 3.9. $\overline{P_R}/\overline{P_T}$ vs. distance for different emission times.

The relation between $\overline{P_R}/\overline{P_T}$ and the distance is exploited to derive a function by using curve fitting techniques with the data illustrated in Figure 3.9. The function defining the relation between $\overline{P_R}/\overline{P_T}$ and the distance, is best fitted for the curve fitting model with the equation

$$\frac{\overline{P_R}}{\overline{P_T}} = a_1 e^{b_1 d}, \tag{3.19}$$

where a_1 and b_1 are the curve fitting parameters. The curve fitting is made by using non-linear least squares method which minimizes the sum of squared errors. LM algorithm, which is explained in detail in Section 3.6, is used to estimate the a_1 and b_1 coefficients in an iterative way. Furthermore, other algorithms such as the Newton or Gauss-Newton (GN) can be employed for curve fitting (Hansen et al., 2013). While the Newton algorithm has a more stable convergence than the GN algorithm, it has a cost of calculating second order derivatives. The GN algorithm is less complex, since it calculates first order derivatives instead of second order derivatives and thus, it has a faster (but less stable) convergence. The LM algorithm performs better than the GN algorithm in terms of stability for convergence with the cost of higher complexity. By using the LM algorithm, the measured data and their corresponding fitted curves are given in Figure 3.10 and the

calculated a_1 and b_1 parameters with the Root Mean Square Error (RMSE) values for each T_e are given in Table 3.3.



(a) (b) (c)

Figure 3.10. Measured $\overline{P_R}/\overline{P_T}$ vs. the distance and their fitted curves for (a) $T_e = 0.25$ s, (b) $T_e = 0.5$ s and (c) $T_e = 0.75$ s.

As observed from the figure, the measured $\overline{P_R}/\overline{P_T}$ decreases exponentially by the distance. This relation can be employed to estimate the distance by pulling out d in (3.19) which is given by

$$\hat{d} = \frac{1}{b_1} ln\left(\frac{P_R}{P_T a_1}\right), \tag{3.20}$$

where average power values are replaced with the measured power values and $ln(.)$ shows the natural logarithm.

Table 3.3. Curve fitting parameters for power based distance estimation.

$\mathbf{T_e}$	$\mathbf{a_1}$	$\mathbf{b_1}$	**RMSE**
0.25 s	2.724	-0.03116	0.0096
0.5 s	13.4001	-0.04146	0.0211
0.75 s	3.1888	-0.02973	0.0171

3.7.2. Peak Time Based Distance Estimation

In this method, similar to the power based distance estimation, the extracted t_{peak} feature is averaged over the measured values for each distance. The averaging is performed separately for three different emission times as 0.25, 0.5 and 0.75 s. As observed

in Figure 3.11, there is an increasing exponential relation between the average peak time and the distance.

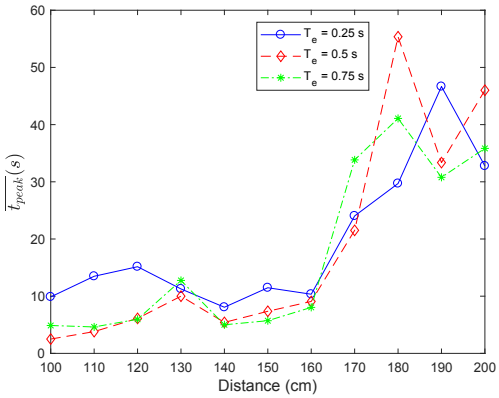


Figure 3.11. Average peak time vs. distance for different emission times.

A function can be obtained for the data in Figure 3.11 by curve fitting with the formula given by

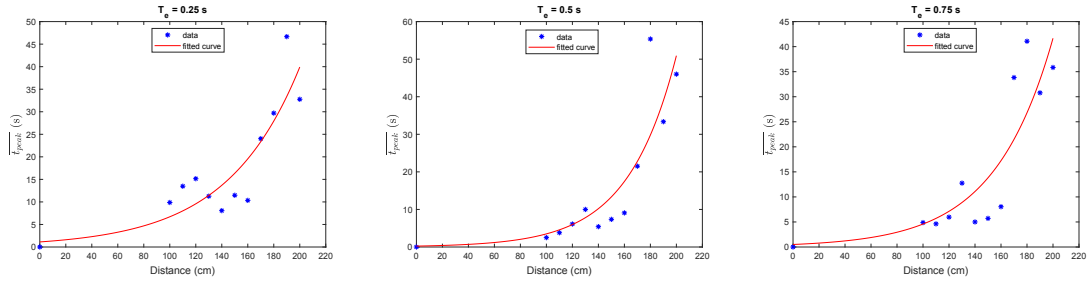
$$\overline{t_{peak}} = a_2 e^{b_2 d}, \tag{3.21}$$

where $\overline{t_{peak}}$ is the average peak time, a_2 and b_2 are curve fitting parameters. An insertion is made to the data for $\overline{t_{peak}} = 0$ at $d = 0$ to obtain a more accurate curve. The same procedure is used for curve fitting as applied for power based distance estimation. According to (3.21), the fitted curves are given in Figure 3.12 for different emission times. By using these curve fitting models, their parameters and RMSE values are calculated as given in Table 3.4. Via the substitution of the measured peak time (t_{peak}) with $\overline{t_{peak}}$, the relation between the average peak time and the distance are employed to derive a distance estimation model as given by

$$\hat{d} = \frac{1}{b_2} ln\left(\frac{t_{peak}}{a_2}\right). \tag{3.22}$$

Table 3.4. Curve fitting parameters for peak time based distance estimation.

$\mathbf{T_e}$	$\mathbf{a_2}$	$\mathbf{b_2}$	**RMSE**
0.25 s	1.1259	0.0178	6.6574
0.5 s	0.2401	0.0268	9.1950
0.75 s	0.5045	0.0221	7.7602



(a) (b) (c)

Figure 3.12. Measured average peak time vs. the distance and their fitted curves for (a) $T_e = 0.25$ s, (b) $T_e = 0.5$ s and (c) $T_e = 0.75$ s.

3.7.3. Combined Distance Estimation

The power based and peak time based distance estimation methods given in Section 3.7.1 and 3.7.2, respectively, can be combined to obtain a novel estimation method. When the ratio P_R/P_T to t_{peak} is written by using the curve fitting models in (3.19) and (3.21), the following equation is obtained.

$$\frac{\frac{P_R}{P_T}}{t_{peak}} = \frac{a_1 e^{b_1 d}}{a_2 e^{b_2 d}}. \tag{3.23}$$

By the manipulation of (3.23), d can be derived as

$$\hat{d} = \frac{1}{b_1 - b_2} ln \left(\frac{P_R a_2}{P_T t_{peak} a_1} \right), \tag{3.24}$$

which combines the power based and peak time based distance estimation methods. Next, the numerical results are given for all of the methods given up to now.

3.8. Results and Analysis

In this section, numerical results are presented for distance estimation to analyze the performance of ML methods given in Section 3.6 and the proposed data analysis based methods given in Section 3.7. First, the numerical results of the ML and data analysis based methods are discussed. Then, these methods are compared and an analysis about the experimental results are given.

3.8.1. Numerical Results

The aforementioned nine features, which are t_{low}, C_{low}, R, ΔC, G, t_{peak}, C_{peak}, T_e and E_R, are given as the input for training the ML methods. Here, all of the features except T_e, are extracted as explained in Section 3.5. Since T_e is assumed to be known by the RX, it is also given as a feature. The Monte Carlo simulations are performed 10^5 times for the ML algorithms to average the estimated distance values and their errors. The data are divided randomly as 70% for training and 30% for testing in MLR. In NNR, the data are randomly divided as 70% for training, 15% for validation and 15% for testing. Furthermore, the hidden layer of the neural network has one node in the simulation model.

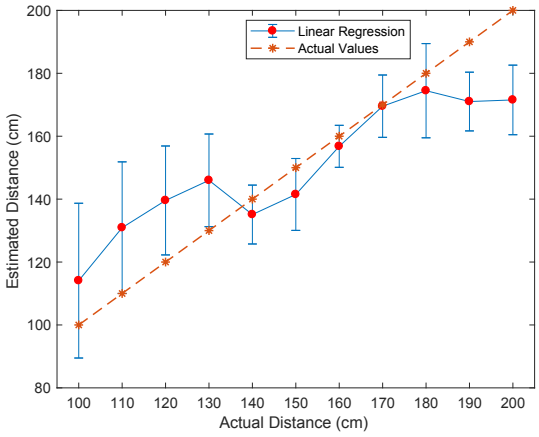


Figure 3.13. Mean estimated distance and their standard deviations with linear regression for each distance.

The performance comparison of the ML methods can be made for each distance. The mean estimated distances and their standard deviations with respect to actual distance are given in Figures 3.13, and 3.14 for MLR and NNR, respectively. The points at each distance show the mean estimated values and the vertical bars show their standard deviations. Moreover, a perfect estimation line is given to see how far the estimated values approach to the actual values. The reason for giving these figures is to see especially the standard deviations, i.e., to visualize how consistent the estimated values are. The NNR has slightly better results than MLR for the mean estimated values and standard deviations. However, since the difference between MLR and NNR results is small, it can

be deduced that the relationship between the features used as input and the distance is highly linear. The cost of the NNR method is its complexity and longer time requirement. Hence, estimating the distance with MLR can be considered as an efficient ML method due to its simplicity and performance for our practical scenario and extracted features.

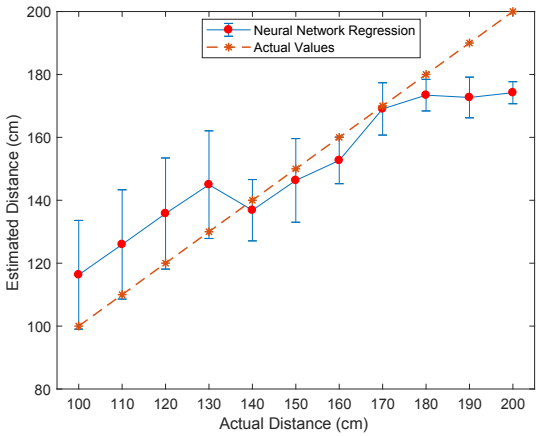


Figure 3.14. Mean estimated distance and their standard deviations with neural network regression for each distance.

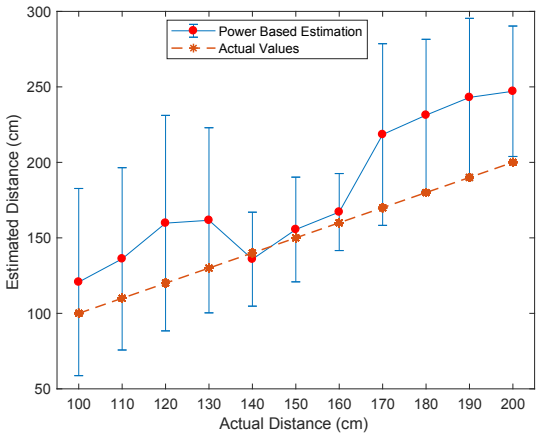


Figure 3.15. Mean estimated distances and their standard deviations for power based distance estimation.

The performance evaluation of the proposed data analysis based distance estimation methods given in Section 3.7, are made by using the features of the measured signals.

The mean estimated distance values and their standard deviations are shown in Figures 3.15, 3.16 and 3.17 for power based, peak time based and combined estimation methods, respectively.

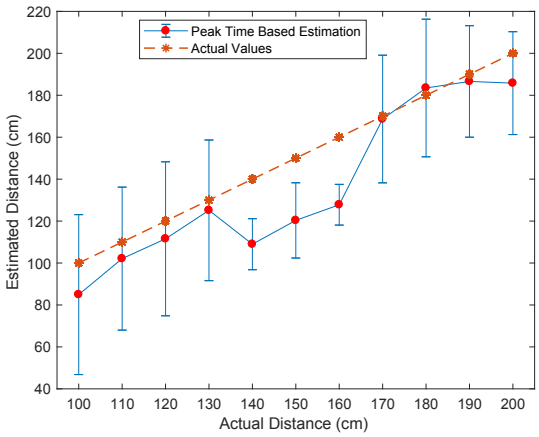


Figure 3.16. Mean estimated distances and their standard deviations for peak time based distance estimation.

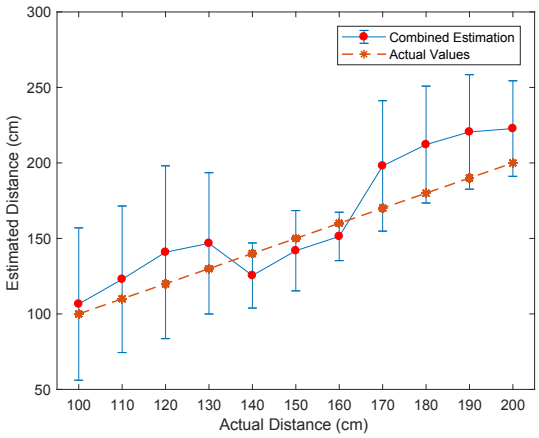


Figure 3.17. Mean estimated distances and their standard deviations for combined distance estimation.

The values for every actual distance on these figures show the mean (dots) and standard deviation (vertical bars) values of the estimated distance over 30 measurements. Due to the smaller number of features employed in data analysis based methods, the standard deviations are relatively higher, when it is compared with the ML methods. There

can be some unused parameters in the data analysis based methods to estimate the distance more accurately with respect to the ML methods. Next, all of the results is given in a comparable format to evaluate the performance of the distance estimation methods given in this chapter.

3.8.2. Comparison

In Table 3.5, the Root Mean Square Error (ρ) values are given for each method, which is defined by

$$\rho = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{d}_i - d_i)^2}, \tag{3.25}$$

where M is the number of the samples of the test data. In data analysis based methods, M is taken as 330 by using whole data samples. In ML methods, M represents the number of the samples of the test data as defined in the beginning of this section and ρ is averaged over 10^5 Monte Carlo simulation trials. The results for ML methods are much better than the data analysis based methods. However, the cost of the superiority of the ML methods emerge as time complexity which means that the ML methods require more time and processing power for distance estimation calculations due to the longer equations used in these methods. NNR has the highest complexity among the proposed methods. Furthermore, the training phase of the ML methods can restrict real-time applications due to this complexity. Therefore, there is a trade-off between the complexity and performance for the ML and data analysis based methods, which needs to be optimized for the implementation of an application. On the other hand, the fact that the data analysis based methods have less time complexity, makes the implementation of these methods easier than the ML methods.

Table 3.5. Distance estimation performance for the ML and data analysis based methods.

Method	**RMSE**
Linear Regression	21.4470
Neural Network Regression	20.1157
Power Based Estimation	62.3077
Peak Time Based Estimation	33.2434
Combined Estimation	44.3157

While the distance values increase, the same absolute error deviation has different meanings according to the actual distance and the error values need to be normalized to determine which method is better. Hence, mean absolute percentage error (MAPE) is preferred as the performance metric rather than the absolute error deviation. For each actual distance, the MAPE (ϵ) values shown in Figure 3.18 are calculated according to the formula as given by

$$\epsilon = \frac{100}{M} \sum_{i=1}^{M} \frac{|\hat{d}_i - d|}{d}. \tag{3.26}$$

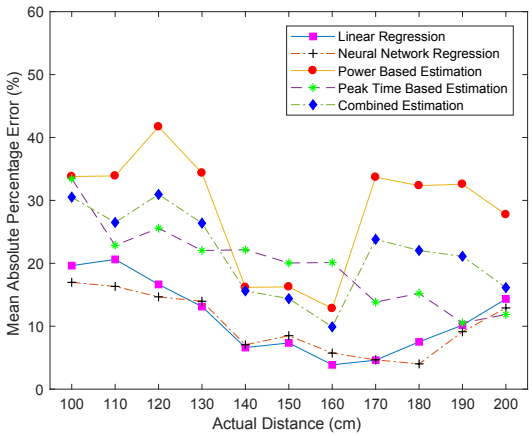


Figure 3.18. Mean absolute percentage errors vs. actual distance for the ML and data analysis based methods.

In our practical scenario, our observations show that some of the droplets fall to the ground, before they arrive to the RX. The droplets are sprayed from the TX as sufficiently large droplets to be affected by the gravity (De Cock et al., 2017). Moreover, there are other effects such as boundary conditions in the diffusion process and the initial velocity of the molecules. Since ML has more parameters to model these effects analytically, they perform better than the data analysis based methods except the peak time based estimation method at long distances. In general, the peak time based estimation provides the best performance among the data analysis based methods. However, the peak time based estimation method has the worst performance between 140 and 160 cm due to the non-linearity of the sensor as detailed in Section 3.8.3.

Unlike the other methods, peak time based estimation has a nearly linear estimation performance and its MAPE decreases, as the distance increases. At 190 and 200 cm,

it has very close values with ML methods and even performs better than the ML methods at 200 cm. This result shows us that peak time based estimation can be applied to longer distances. Furthermore, it reveals that the peak time provides more accurate distance estimation, as the RX gets farther from the TX. Hence, the distance can be estimated with more sensitive sensors for farther distances.

In order to determine the optimal distance estimation method for a practical scenario, the processing capacity of the RX is essential. If the RX does not have the capacity to implement ML algorithms, then the peak time based estimation method can be used with a cost of slightly higher error than ML methods. Otherwise, MLR can be used as a more efficient method between ML methods due to its lower complexity and high accuracy. NNR has the best performance among all the proposed methods in terms of accuracy, although it imposes the highest complexity. Next, our observations and numerical results are analyzed to explain the phenomena behind the propagation of the molecules and its reception by the RX.

3.8.3. Analysis of the Results

The lowest MAPE values are between 140 and 180 cm for ML methods and between 140 and 160 cm for data analysis based methods, except the peak time based estimation. The reason why the estimated values are better within the interval 140 and 160 cm, lies in the measured signals. This can be more clearly understood with the average velocity profile of the transmitted molecules in the medium as shown in Figure 3.19. In order to calculate the average velocity of the molecules, it is assumed that the majority of the molecules arrive at the RX, when the received molecule concentration is at its peak point. Although some of the molecules reach the RX before or after this peak point, it is reasonable to choose t_{peak} as the average arrival time of the molecules. Hence, the average velocity of the molecules is defined as the ratio of the distance to the time duration at which the peak concentration is read from the sensor. In Figure 3.19, average velocities are obtained according to this definition for each T_e and also the mean velocity is given as the average of the velocity values given for three different values of T_e.

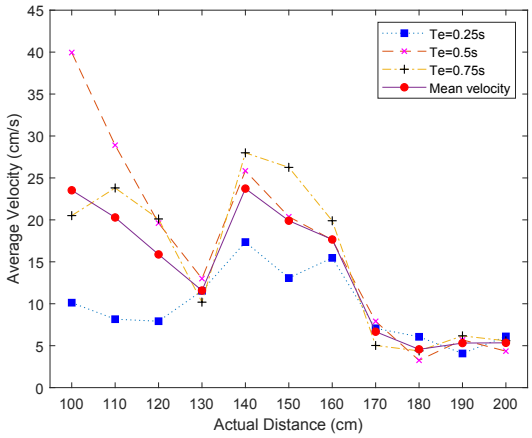


Figure 3.19. Average velocity profile of the transmitted molecules with respect to the actual distance.

3.8.3.1. The Effect of the Sensor's Non-linearity

The larger velocities of the molecules for the measurements between 140 and 160 cm can be related to the non-linearity of the sensor characteristics as given in Figure 3.20. In this figure, R_s is the sensor resistance which changes according to the molecule concentration and R_o is the sensor resistance measured at the concentration level 0.4 mg/L. The sensitivity of the sensor is shown for different gases in logarithmic scale in Figure 3.20 (a) (Hanwei-Electronics, 2018). The values in this figure for alcohol is plotted in linear scale in Figure 3.20 (b) to see the non-linearity of the sensor more clearly for our case. The MQ-3 sensor makes more sensitive measurements in lower concentrations with respect to higher concentrations. Due to this fact, the sensor can have errors for the measurements at lower distances. Another possible explanation to understand the phenomena behind the motion of the molecules in our scenario, can be made by employing the dynamics of the fluid which is considered as follows.

3.8.3.2. The Effect of Factors Related to Fluid Dynamics

In ideal conditions where the molecules only make an isotropic random walk without being exposed to any force, the diffusion process can be used to model the movement of the molecules. However, the molecules come out of the TX as liquid droplets with

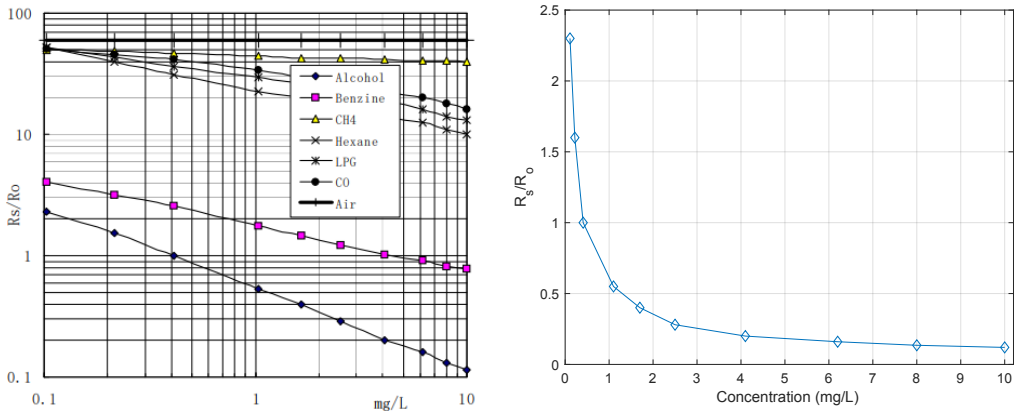


Figure 3.20. Sensitivity characteristics of the MQ-3 sensor (R_s/R_o vs. Concentration) (Hanwei-Electronics, 2018) (a) for different gases in logarithmic scale (b) for alcohol in linear scale (plotted using original values from the datasheet).

an initial velocity in our case. Thus, several processes concerning the dynamics of the fluid can occur and the diffusion process is not sufficient to understand the motion of the molecules. Therefore, the analysis is needed to be made in terms of fluid dynamics.

When a liquid is sprayed into the air, the droplets initially come out of the sprayer with a certain angle which is the beamwidth (spray angle) of the sprayer as shown in Figure 3.21 with α. The interaction between the transmitted droplets and the air molecules leads to a contraction in the effective beamwidth, which is defined as the angle of the propagating droplets as given in Figure 3.21 with β. This interaction is detailed as follows.

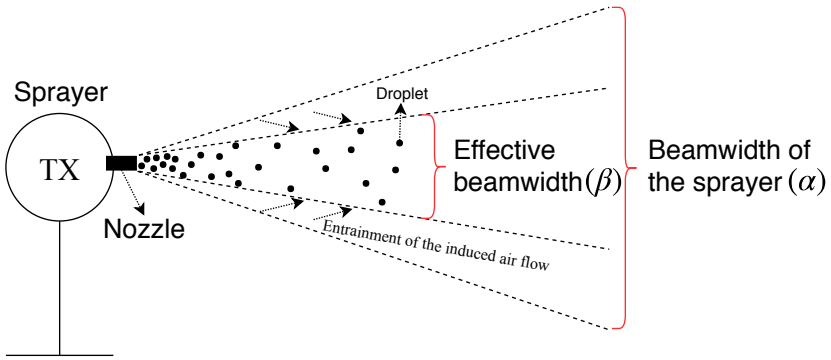


Figure 3.21. Beamwidth of the sprayer (TX) and the effective beamwidth.

As the droplets move in the air due to the force applied by the sprayer, they gain a momentum, which can be defined as the product of the liquid's mass and velocity. According to the law of momentum conservation, when two objects are collided, their total momentum does not change before and after the collision (Kleppner and Kolenkow, 2013). The collisions of the droplets with the air molecules cause aerodynamic drag which can be defined as the air force resisting the movement of the object (Anderson Jr, 2016). The ethyl alcohol droplets slow down due to the aerodynamic drag. Besides, the momentum lost by the droplets is gained by the air molecules due to the momentum conservation. The momentum exchange among the droplets and air molecules generates an air flow towards the center of the beamwidth along the horizontal axis (Rothe and Block, 1977). This flow drags the droplets into the inner regions from the outer regions. Namely, the angle at which the droplets propagate is reduced relative to the initial α angle, and this leads to a contraction. This contraction of the droplets' propagation angle facilitates turbulent flows which is detailed as follows (Ghosh and Hunt, 1994).

The flows can be classified into two types as *laminar* flow, which has parallel and regular streamlines, and *turbulent* flow, which has random and irregular streamlines (Anderson Jr, 2016). Even if the effect of the turbulent flow can be observed, its mathematical modeling is still a big problem due to its complicated nature. The flow type of the fluid depends on the Reynolds number (Re). In our case, Re is given as (Ghosh and Hunt, 1994),

$$Re = \frac{2|V_l - V_a|a}{\nu_a}, \tag{3.27}$$

where a is the radius of the spherical droplet, ν_a is the kinematic viscosity of the air, V_l and V_a are the velocities of the liquid and airflow. Re influences the type of viscous flows. As Re increases, it is more likely to have a turbulent flow in the medium. There is a critical value of Re to determine the flow type (Avila et al., 2011). Above this critical value, which depends on the parameters of the transmitted molecules and the medium, the flow becomes turbulent. In order to determine where there is a turbulent flow according to (6.26), the radii of the liquid droplets, the velocities of the liquid droplets and air molecules are needed to be measured, which is beyond the scope of this chapter. High-speed photography techniques or Phase Doppler Anemometry (PDA) can be used to measure the liquid droplet size and velocities. In (Begg et al., 2009), the movement of the liquid gasoline fuel droplets after they come out of a sprayer (injector) is analyzed.

After the spraying operation, vortex ring-like structures, which include turbulent flows, are observed experimentally for the liquid droplets. As they move away from the sprayer, the average liquid droplet size gets smaller due to the break-up of droplets into smaller sized droplets and evaporation (De Cock et al., 2017; Al Heidary et al., 2014). Diffusion does not have an important effect, if the droplet is large enough. Hence, diffusion is more effective, after the velocity of the droplets decreases to a more steady value. The findings and results in (Begg et al., 2009) can be used to analyze the movement of the droplets for our scenario, since the densities of the gasoline fuel and ethyl alcohol have close values (Mařík et al., 2014).

In the vicinity of the nozzle of the TX, the flow of the droplets can be defined as unsteady flow, since the velocity of the droplets are not constant with respect to time due to the interaction with the air molecules (Munson et al., 2009a). Moreover, if a constant flow is stopped suddenly, the flow shows unsteady characteristic (Munson et al., 2009a). Similarly in our case, since the TX suddenly stops its transmission at the end of T_e, it can be deduced that the flow in our case becomes an unsteady flow. As soon as molecules are emitted, they are affected by the diffusion which can be neglected in the region where molecules are entrained by the higher velocities of unsteady flows such as turbulent flows. As the droplets move away from the TX, the aerodynamic drag decreases their velocities until a point where they start to move with only diffusion. As the result of these movements and findings in (Begg et al., 2009), the transmitted droplets can be expected to have a trajectory as illustrated in Figure 3.22. Here, the motion of the droplets is divided into two zones as the unsteady flow zone and diffusion zone. In the first zone, the droplets where they travel a distance d_f, are affected by the unsteady flows such as turbulent flows. The droplets are exposed to random fluctuating aerodynamic drag forces due to the turbulent flows (Maxey, 1987). When the effect of the flows ends, the droplets start to move with diffusion where they travel a distance d_B until the RX. Although gravity affects the motion of the droplets in both zones, its effect is greater in the unsteady flow zone, since the size of the droplets is larger. By using the trajectory in Figure 3.22, the velocity profile in Figure 3.19 can be explained as follows. For the measurements between 100 and 130 cm, the non-linearity of the sensor can cause more measurement errors due to the lower sensitivity in higher concentrations. In addition to this, the average distance traveled in the unsteady flow zone can be larger for the measurements between

140 and 160 cm than the average distance traveled in the unsteady flow zone for the measurements at other distances due to effects of the unsteady flows. Thus, the droplets move with larger velocities which cause shorter arrival time to the RX and a longer d_f than the measurements at other distances. Consequently, these factors such as non-linear characteristic of the sensor and unsteady flows jointly affect the performance for distance estimation.

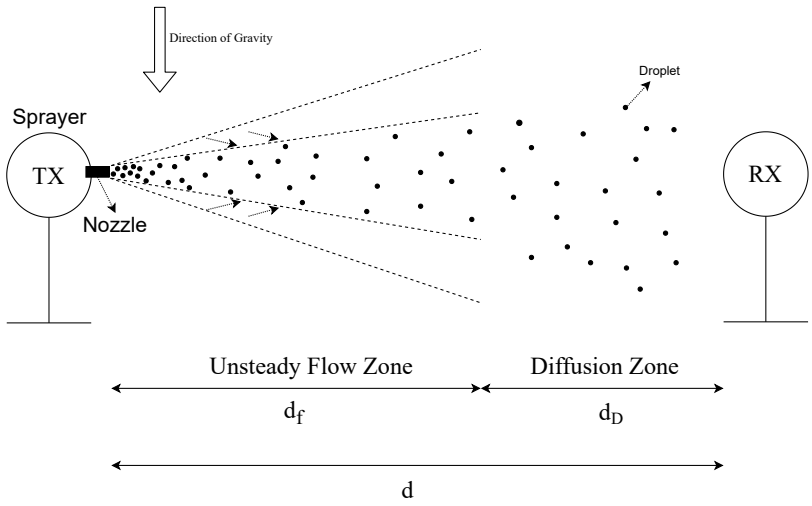


Figure 3.22. The trajectory of the transmitted molecules between the TX and the RX.

As given by the analysis about the unexpected average velocities of the molecules, their motion for macroscale practical scenarios can be modeled by exploiting fluid dynamics. The channel modeling and parameter estimation with the fluid dynamics point of view are open research issues for practical MC scenarios. Next, the role of emission time, i.e., T_e, is examined for distance estimation by using the trajectory given in Figure 3.22.

3.8.3.3. The Effect of the Emission Time

Figure 3.19 for short distances up to 130 cm, the average velocity of the molecules for $T_e = 0.5$ s are greater than the average velocity of the molecules for $T_e = 0.25$ s. However, for $T_e = 0.75$ s, the average velocity is less than the molecules transmitted with $T_e = 0.5$ s (except the nearly equal value at 120 cm). This result shows that there

is an optimum T_e for the fastest propagation of the molecules according to the distance. Furthermore, it shows that sending more molecules than this optimum T_e can decrease the average velocity of the molecules. When T_e is larger than the optimum T_e, a higher molecule concentration occurs in the unsteady flow zone. This increases the collision rate of the molecules and thus, reduces their velocity. For distances between 140 and 160 cm, the average velocity is directly proportional with T_e due to the above-mentioned longer average distance traveled in the unsteady flow zone. For long distances greater than 170 cm, T_e does not significantly affect the average velocity of the molecules. Hence, it can be deduced from the results that an appropriate T_e should be chosen for a faster propagation according to the distance. More experimental research is needed for the derivation of the optimum T_e, which is left as a future work.

CHAPTER 4

FLUID DYNAMICS-BASED DISTANCE ESTIMATION ALGORITHM FOR MACROSCALE MOLECULAR COMMUNICATION

4.1. Introduction

In molecular communication (MC), the distance between the TX and RX is a significant channel parameter, since higher data rates can be achieved via an accurate distance estimation by arranging the MC system parameters properly (Atakan and Akan, 2007; Nakano et al., 2013). Furthermore, the location of a molecular TX in a molecular network can be found by an accurate distance estimation in a practical application as experimentally shown in Chapter 8 (Gulec and Atakan, 2020b). In addition, the location of an infected human which spreads a disease by emitting airborne pathogen-laden droplets through sneezing or coughing in a crowded place can be predicted by employing biological sensors as the RX and the infected human as the TX (Khalid et al., 2019, 2020). The distance estimation methods in the literature can be discussed as two-way and one-way methods. In two-way methods, the TX sends a molecular pulse signal to the RX and the RX sends a feedback signal to the TX, when it is received. Subsequently, the distance is calculated by the TX based on this feedback signal (Moore et al., 2010, 2012; Moore and Nakano, 2012). In one-way distance estimation methods which need less time with respect to two-way methods, the distance is estimated by the RX. In (Huang et al., 2013), the received peak concentration or the peak time between the consecutive transmissions are measured by the RX for distance estimation in a 1-D channel. In (Wang et al., 2015b), two distance estimation schemes are proposed for a 3-D diffusion channel where the received peak time and the received energy are used for the first and second scheme, respectively. In addition, an algorithmic distance estimation method is proposed in (Wang et al., 2015a) by the same authors. The study in (Lin et al., 2019) proposes a method which uses maxi-

mum likelihood estimation for a 3-D diffusion channel. (Noel et al., 2015) proposes joint estimation methods for channel parameters in a diffusive channel with a steady flow and degradable molecules. Furthermore, the Cramer-Rao lower bound is derived for distance estimation in (Noel et al., 2014a). In all of these estimation methods, an ideal microscale channel where molecules propagate via diffusion is considered. Our study given in Chapter 3 proposes five methods including three data analysis based and two machine learning methods for distance estimation in a practical macroscale MC system for the first time. However, the physical meanings of the estimation parameters on which these methods depend are not known.

In fact, the liquid is sprayed as droplets rather than molecules in practical sprayer-based macroscale scenarios (Ghosh and Hunt, 1994; Al Heidary et al., 2014) similar to the saliva emitted as droplets through sneezing/coughing in airborne transmission of pathogen-laden droplets. This similarity gives us the motivation to employ the droplet-based MC systems to reveal the infection mechanism of contagious diseases which spread through airborne transmission of droplets. The airborne pathogen transmission plays a vital role in infections of the pathogens such as severe acute respiratory syndrome (SARS) virus (Peiris et al., 2003), influenza virus (Killingley and Nguyen-Van-Tam, 2013) and SARS coronavirus 2 (SARS-CoV-2) which causes coronavirus disease 19 (COVID-19). Due to the importance of the global COVID-19 pandemic, it is essential to highlight that the distance estimation methods proposed for droplet-based macroscale MC systems can also be employed to find the location of a pathogen source in public places via biological sensors.

In the literature, practical droplet-based MC systems having impulsive input signals are assumed to have a channel impulse response which is based on molecular diffusion with drift (Farsad et al., 2014; Kim et al., 2015). The channel parameters are derived by fitting experimental data and the physical meanings of these fitted parameters are unknown and not measurable. However, the fact that the sprayer-based systems emit droplets instead of molecules implies that droplets are subject to Newton's laws of motion and molecular diffusion is negligible. Furthermore, droplets interact with air particles during their propagation due to their initial velocity. Therefore, considering only the diffusion of molecules is not sufficient and a fluid dynamics perspective is needed for a more accurate distance estimation in macroscale scenarios as initially discussed in

Chapter 3. In this chapter, the Fluid Dynamics-based Distance Estimation (FDDE) algorithm is proposed for a practical macroscale MC system. In the FDDE algorithm, droplets are considered as information carriers in the channel and a two-phase flow model is used. Here, liquid phase of droplets and gas phase of air particles represent these two phases. In this model, liquid droplets and gaseous air particles move together as a mixture. Moreover, it is considered that droplets evaporate as the time elapses. The TX is modeled as a directed emitter with a predefined beamwidth. Droplets are assumed to move in a cone shaped volume determined by this beamwidth. Then, the laws of mass and momentum conservation are utilized to estimate the average propagation distance of the droplets from the TX by analytical derivations. The FDDE algorithm employs these derivations by also considering reducing droplet diameter due to evaporation as the droplets propagate in the MC channel. Subsequently, the proposed FDDE algorithm is validated by experimental data. It is shown that the distance between the TX and RX in the MC channel depend on novel parameters such as the beamwidth of the TX, volume fraction of the droplets in the mixture volume and densities of the air particles and liquid droplets. Moreover, the results show that the effect of the evaporation for shorter distances is small. With the validation of the FDDE algorithm, it is revealed that modeling the motion of the droplets is realizable with physically measurable parameters instead of fitted parameters which are based on experimental data.

The rest of the chapter is organized as follows. The proposed FDDE algorithm is presented in Section 4.2. The experimental work explaining the measurement of the required parameters is given in Section 4.3.1. The numerical results including validation of the proposed FDDE algorithm and its comparison with the methods given in Chapter 3 are presented in Section 4.3.2. The study given in this chapter was first presented in (Gulec and Atakan, 2021b).

4.2. Fluid Dynamics-Based Distance Estimation Algorithm

In this section, the FDDE algorithm is proposed to estimate the distance between the TX and RX with a fluid dynamics approach. The experimental setup detailed in Chapter 3 is employed for the validation of the proposed FDDE algorithm. When the droplets are transmitted into the air, air particles are entrained by the droplets and this

induces an air flow. During this interaction generating this air flow in the vicinity of the nozzle of the TX, the velocity difference among the droplets and air particles is large and this difference fluctuates over time. Here, it is useful to introduce Reynolds number (Re) which determines the flow type of the fluid, i.e., laminar or turbulent. The changing relative velocity of the droplets with respect to air particles cause Re to grow which is given by

$$Re = \frac{d|v_d - v_a|}{\nu_a} \tag{4.1}$$

where d is the diameter of the droplet, ν_a is the kinematic viscosity of the air, v_d and v_a are the velocities of the droplets and air particles, respectively. As Re increases, deriving analytical solutions for the motion of particles becomes difficult, since the turbulent diffusivity increases (Ghosh and Hunt, 1994). However, as the distance between the TX and RX increases, the droplets can be assumed to have nearly the same velocity with the entrained air particles. This situation makes Re zero and turbulent flows are not considered (Sazhin et al., 2001). Hence, a tractable analytical solution for the average velocity and traveling distance of the droplets is feasible. For a macroscale scenario without a constant flow, the distance between the TX and RX is sufficiently large that the effect of the initial interaction among the droplets and air particles can be negligible for the total traveling distance of the droplets. The motions of the droplets and air particles with the same velocities can be modeled by using the two-phase flow model. In this model, the two phases represent the liquid and gas phases of the fluid particles, i.e., liquid droplets and gaseous air particles. The evaporated droplets, i.e., gaseous ethanol molecules in our case, are not considered for the gas phase, since their effect is small with respect to the surrounding air in the medium. As illustrated in Figure 4.1, droplets and air particles move together as a mixture between the TX and RX. This mixture is assumed to propagate in a cone shaped volume which has a beamwidth of 2θ (see Figure 4.1). In fluid dynamics literature, two-phase flow models are applied to estimate the average distance of the fuel droplets sprayed by a fuel injector in a car engine (Sazhin et al., 2001) or the coverage of the sprayed pesticide droplets in agriculture (Al Heidary et al., 2014). In this study, the two-phase flow model given in (Sazhin et al., 2001) is adopted and modified to estimate the propagation distance of the droplets transmitted from the TX. This modification is accomplished by considering that the liquid droplets change their sizes via evaporation. Furthermore, the distance estimation procedure is given in an algorithmic way by updating the diameter

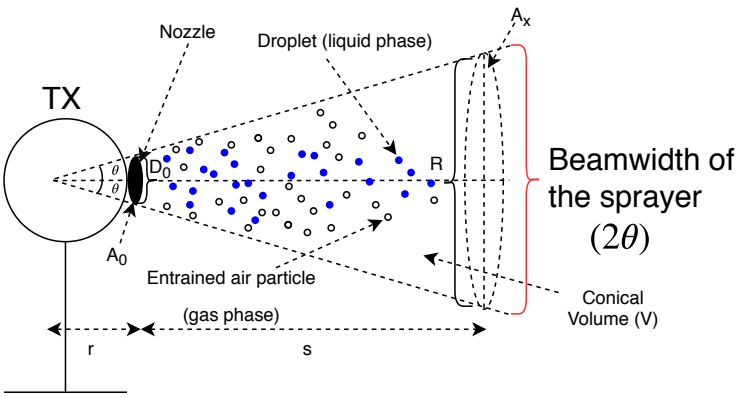


Figure 4.1. Two-Phase Flow Model.

at each time step. Next, the evaporation of the droplets, which is detailed as follows, is considered for the FDDE algorithm.

4.2.1. Evaporation of Droplets

The transmitted droplets can evaporate as they move away from the sprayer (Al Heidary et al., 2014). The evaporation can be described in terms of the reduction in the droplet diameter. The diameter change with respect to time is modeled as (Mokeba et al., 1997)

$$\frac{\partial d}{\partial t} = -2\frac{M_v D_v \rho_a \Delta P}{M_a d \rho_d P_a}(2 + 0.6Sc^{1/3}Re^{1/3}), \tag{4.2}$$

where M_v is the molecular weight of the evaporating vapor, M_a is the molecular weight of the air, D_v is the diffusion coefficient of the droplet's vapor in the air, d is the droplet diameter, P_a is the partial air pressure, ΔP is the pressure difference between the droplet surface and the diffusing vapor in the air and Sc is the Schmidt number which is given by $Sc = \nu_a/D_v$. The parameters and subscripts used for the proposed algorithm are summarized in Table 4.1. For an algorithm which evaluates the motion of droplets by using (4.2), the diameter change (Δd) in every time step can be given as below

$$d_i = d_{i-1} + \Delta d \tag{4.3}$$

$$\Delta d = -2\frac{M_v D_v \rho_a \Delta P}{M_a d_i \rho_d P_a}(2 + 0.6S_c^{1/3}Re^{1/3})\Delta t, \tag{4.4}$$

Table 4.1. Definitions of the parameters and subscripts

Parameter	**Definition**	**Parameter**	**Definition**
Re	Reynolds number	M	Molecular weight
θ	Half-beamwidth of the TX	d	Average diameter of the droplet
D	Diffusion coefficient	ρ	Density
Δt	Time step	P	Partial pressure
Sc	Schmidt number	m	Mass
ΔP	Vapor pressure difference	V	Conical mixture volume
α_d	Volume fraction of droplets in mixture	T_e	Emission time of the TX
D_0	Diameter of the nozzle orifice	A	Circular cross-sectional area
v	Average velocity	$\dot{m}$	Mass flow rate
s	Average distance	R	Diameter of A_x
p	Momentum	t_s	Total simulation time
$\dot{p}$	Time rate of change of the linear momentum	r	Distance between the nozzle and starting point of the flow
Q	Volumetric flow rate		
Subscript		**Subscript**	
d	droplet (liquid ethanol)	i	value at the i^{th} time step
v	vapor ethanol	x	mixture
a	air	0	initial value

where d_i is the diameter at the time t_i which increases with steps of Δt for each iteration and represents the elapsed time at the i^{th} time step. In order to estimate the distance values in a practical MC system, it is essential to relate these values with the measured sensor voltage values. For this purpose, t_i is assumed as the peak time of the signal measured by the sensor, since the majority of droplets are assumed to reach the RX at this peak time.

Here, a new parameter (α_d) which shows the volume fraction of the droplets in the mixture of droplets and air particles (see Figure 4.1) is introduced. The evaporation of droplets causes α_d to be time-dependent. As the diameters of droplets change with time, α_d is needed to be updated. To derive a model for this, let N be the number of droplets, d_i the average diameter of a spherical droplet, V the mixture volume and α_{d_i} the volume fraction of droplets in the mixture volume during the i^{th} time step. Here, V is considered as a constant cone-shaped volume with the beamwidth of the TX for the distance between the TX and RX as shown in Figure 4.1. Then, the volume fractions for consecutive time

steps can be given as

$$\alpha_{d_{i-1}} = \frac{N\frac{4}{3}\pi\left(\frac{d_{i-1}}{2}\right)^3}{V} \tag{4.5}$$

$$\alpha_{d_i} = \frac{N\frac{4}{3}\pi\left(\frac{d_i}{2}\right)^3}{V}. \tag{4.6}$$

Hence, the relation between consecutive volume fractions of droplets is given by

$$\alpha_{d_i} = \alpha_{d_{i-1}}\left(\frac{d_i}{d_{i-1}}\right)^3. \tag{4.7}$$

For this derivation, it is essential to consider that the number of droplets in the mixture volume increases during the emission of droplets. Therefore, it is assumed that the volume fraction of droplets at $t = T_e$ is a pre-known constant, i.e., α_{d_0}, and α_{d_i} increases linearly between $t = 0$ and $t = T_e$. After $t = T_e$, the volume fraction is given by (4.7). Hence, the volume fraction of droplets can be defined as a time-dependent function:

$$\alpha_d(t = t_i) = \alpha_{d_i} = \begin{cases} \frac{\alpha_{d_0}}{T_e}t_i & ,0 \leq t_i \leq T_e \quad (4.8) \\ \alpha_{d_{i-1}}\left(\frac{d_i}{d_{i-1}}\right)^3 & ,t_i > T_e. \quad (4.9) \end{cases}$$

4.2.2. Propagation of Evaporating Droplets in Two-Phase Flow

In order to clearly explain the propagation of evaporating droplets in two-phase flow, some definitions related with fluid dynamics are first given as follows. As illustrated in Figure 4.1, the liquid phase and the gas phase form two phases of the two-phase flow model in which liquid droplets and gaseous air particles move together as a mixture. This model can be generated by employing the laws of mass and momentum conservation. In fluid dynamics, the conservation of mass is applied by using the mass flow rate ($\dot{m}$) which is defined as the amount of mass flowing through a surface per unit time (kg/s) (Munson et al., 2009b). From the conservation of mass, which states that the net mass flow rate ($\dot{m}$) is zero in a closed system (a system with no external forces acted on and no external matter exchange), the equation below can be written (Munson et al., 2009b)

$$\dot{m}_0 = \dot{m}_i, \tag{4.10}$$

where $\dot{m}_0$ and $\dot{m}_i$ is the mass flow rate of droplets for the initial state and i^{th} time step, respectively. The mass flow rate is expressed by (Munson et al., 2009b)

$$\dot{m}_i = \rho_i Q_i = \rho_i A_i v_i, \tag{4.11}$$

where Q_i is the volumetric flow rate (m^3/s) giving the fluid volume per unit time, ρ_i is the density of the fluid (kg/m^3) and v_i is the fluid velocity (m/s) which is perpendicular to the cross-sectional area of the fluid A_i (m^2) for the i^{th} time step. In 4.11, if ρ_i is constant, $\dot{m}_i$ depends on the volumetric flow rate. Therefore, α_d can be incorporated into 4.11 in order to quantify the volume fraction of droplets in the mixture. Hence, the mass flow rate of air particles ($\dot{m}_{a_i}$) is given by taking out the volume of droplets from the mixture volume as given by

$$\dot{m}_{a_i} = \rho_a A_{x_i} v_{x_i} - \alpha_{d_i} \rho_a A_{x_i} v_{x_i} \tag{4.12}$$

$$= (1 - \alpha_{d_i}) \rho_a A_{x_i} v_{x_i}, \tag{4.13}$$

where ρ_{x_i} is the density of the mixture, v_{x_i} is the velocity of the mixture and A_{x_i} is the circular cross-sectional area of the mixture volume at the i^{th} time step as illustrated in Figure 4.1. Then, the mass flow rate of droplets at the i^{th} time step can be expressed by subtracting the mass flow rate of air particles from the mixture ($\dot{m}_{x_i}$) as given by

$$\dot{m}_i = \dot{m}_{x_i} - \dot{m}_{a_i} \tag{4.14}$$

$$= \rho_{x_i} A_{x_i} v_{x_i} - (1 - \alpha_{d_i}) \rho_a A_{x_i} v_{x_i} \tag{4.15}$$

Thus, (4.10) can be written by using (4.15) as given by

$$\rho_d A_0 v_0 = \rho_{x_i} A_{x_i} v_{x_i} - (1 - \alpha_{d_i}) \rho_a A_{x_i} v_{x_i}, \tag{4.16}$$

where v_0 is the initial velocity of droplets, A_0 is the circular cross-sectional area of the nozzle and the term in the left-hand side shows the initial mass flow rate of droplets.

As observed in Figure 4.1, the geometrical relation between A_{x_i} and A_0 can be represented as given by

$$\tan(\theta) = \frac{D_0}{2r} = \frac{R_i}{2(s_i + r)}, \tag{4.17}$$

where D_0 is the diameter of the nozzle orifice, r is the distance between the nozzle and the assumed theoretical starting point of the flow, θ is the half-beamwidth of the TX, s_i is the distance traveled by the mixture or droplets and R_i is the diameter of A_{x_i} at the i^{th}

time step. By using (4.17), A_{x_i} can be represented in terms of A_0, s_i, D_0 and θ as given by

$$A_{x_i} = A_0 + \pi D_0 \, s_i \, \tan(\theta) + \pi s_i^2 \tan^2(\theta). \tag{4.18}$$

Next, the conservation of the momentum is exploited to derive the estimated distance. The momentum (p) can be defined for the i^{th} time step as given by (Munson et al., 2009b)

$$p_i = m_i v_i. \tag{4.19}$$

The law of momentum conservation states that the total momentum does not change for a closed system. In our model, it is assumed that the fluid motion is linear (not rotational) and only on the horizontal axis. As in the case of conservation of mass, the time rate of change of the linear momentum, which is defined as the change of the momentum in time, is employed to apply the momentum conservation in fluid dynamics (Munson et al., 2009b). To that end, the linear momentum equation, which implies that the system's time rate of change of the linear momentum ($\dot{p}$) is constant for a closed system, is used as given by (Munson et al., 2009b)

$$\dot{p}_0 = \dot{p}_i, \tag{4.20}$$

where the left-hand side and right-hand side statements are the time rate of changes of the linear momentum at the initial state of droplets and the mixture state, respectively. By substituting (4.19) into (4.20), the linear momentum equation is expressed in terms of mass flow rate as given by (Munson et al., 2009b)

$$\dot{m}_0 v_0 = \dot{m}_i v_i \tag{4.21}$$

$$\rho_d A_0 v_0^2 = \rho_{x_i} A_{x_i} v_{x_i}^2, \tag{4.22}$$

where the mass flow rate derived in (4.11) is substituted into (4.21). The equations (4.16), (4.18) and (4.22) are manipulated to derive the traveling distance of the mixture. Firstly, (4.16) is multiplied by $1/(A_{x_i} v_{x_i})$ and ρ_{x_i} is derived as shown below.

$$\rho_{x_i} = \frac{\rho_d A_0 v_0}{A_{x_i} v_{x_i}} + (1 - \alpha_{d_i})\rho_a. \tag{4.23}$$

Secondly, (4.22) can be written as

$$\rho_{x_i} = \frac{\rho_d A_0 v_0^2}{A_{x_i} v_{x_i}^2}. \tag{4.24}$$

Then, (4.24) is substituted into (4.23) and all the terms are taken to the left-hand side as given by

$$\frac{\rho_d A_0 v_0^2}{A_{x_i} v_{x_i}^2} - \frac{\rho_d A_0 v_0}{A_{x_i} v_{x_i}} - (1-\alpha_{d_i})\rho_a = 0. \tag{4.25}$$

Afterwards, we let $\tilde{v}_i = v_0/v_{x_i}$ and write (4.25) in terms of $\tilde{v}_i$ as

$$\tilde{v}_i^2 - \tilde{v}_i - \frac{(1-\alpha_{d_i})\rho_a A_{x_i}}{\rho_d A_0} = 0. \tag{4.26}$$

For physically meaningful parameter values, one of the roots of (4.26) is negative and the positive root is given by

$$\tilde{v}_i = \frac{v_0}{v_{x_i}} = \frac{1}{2}\left(1 + \sqrt{\frac{4(1-\alpha_{d_i})\rho_a A_{x_i}}{\rho_d A_0}}\right). \tag{4.27}$$

By using (4.27), v_{x_i} is found as

$$v_{x_i} = \frac{2v_0}{1 + \sqrt{\frac{4(1-\alpha_{d_i})\rho_a A_{x_i}}{\rho_d A_0}}}. \tag{4.28}$$

By substituting $A_0 = \frac{\pi D_0^2}{4}$ and (4.18) into (4.28), v_{x_i} is derived as

$$v_{x_i} = \frac{2v_0}{1 + \sqrt{k_1 + k_2 s_i + k_3 s_i^2}}. \tag{4.29}$$

where

$$k_1 = 1 + 4(1-\alpha_{d_i})\frac{\rho_a}{\rho_d}, \quad k_2 = \frac{16(1-\alpha_{d_i})\rho_a \tan(\theta)}{D_0 \rho_d},$$
$$k_3 = \frac{16(1-\alpha_{d_i})\rho_a \tan^2(\theta)}{D_0^2 \rho_d}. \tag{4.30}$$

The average velocity of droplets can be expressed as the derivative of the distance with respect to time. Hence, v_{x_i} is given in terms of the difference of the consecutive distance and time values as given by

$$v_{x_i} = \frac{\Delta s_i}{\Delta t} = \frac{s_i - s_{i-1}}{\Delta t}, \tag{4.31}$$

where Δs_i represents the average distance traveled by the droplets for the time duration $\Delta t = t_i - t_{i-1}$. By incorporating (4.30) into (4.29), the following equation is derived:

$$\sqrt{k_1 + k_2 s_i + k_3 s_i^2} = \frac{2v_0 \Delta t}{\Delta s_i} - 1. \tag{4.32}$$

Then, both sides of (4.32) are squared as given by

$$k_1 + k_2 s_i + k_3 s_i^2 = \frac{4v_0^2(\Delta t)^2 - 4v_0 \Delta t \Delta s_i + (\Delta s_i)^2}{(s_i - s_{i-1})^2}. \tag{4.33}$$

Next, when $\Delta s_i = s_i - s_{i-1}$ is substituted into (4.33), a quartic equation can be obtained as shown by

$$k_3 s_i^4 + (k_2 - 2k_3 s_{i-1}) s_i^3 + (k_1 - 2k_2 s_{i-1} + k_3 s_{i-1}^2 - 1) s_i^2 + (k_2 s_{i-1}^2 - 2k_1 s_{i-1} + 4v_0 \Delta t + 2s_{i-1}) s_i + k_1 s_{i-1}^2 - 4v_0^2 (\Delta t)^2 - 4v_0 \Delta t s_{i-1} - s_{i-1}^2 = 0. \quad (4.34)$$

s_i is found by solving the quartic equation in (4.34) in terms of its previous value s_{i-1} and Δt. For physically meaningful parameter values, one of the roots is negative and the two of the roots are complex numbers in (4.34). Only one real positive root is left as the solution of s_i which is too long to write in this chapter. The results for this solution are shown numerically in the next section.

4.2.3. Operation of the Algorithm

Finally, an estimation algorithm given in Algorithm 1 is proposed by using the derived parameters to evaluate the average traveling distance of droplets. In this algorithm, the initial parameters are given at the beginning such as T_e, α_{d_0}, D_0, v_0, Δt, θ and total simulation time (t_s). D_0 is assigned as the initial droplet diameter. Then, the diameter of droplets are calculated for each time step. During the emission of droplets, droplet diameters are assumed to be constant. Otherwise, the reduction in the droplet diameter is calculated by using Δd to obtain α_{d_i} as given in (4.8) or (4.9). Subsequently, the average distance is calculated by using the real positive root of (4.34). These steps are repeated until the end of the simulation. The FDDE algorithm can be utilized to model the movement of droplets emitted from the TX as given with the numerical results in the next section.

4.3. Numerical Results

In this section, numerical results using the FDDE algorithm given in Section 4.2 are presented. This algorithm is validated by experimental results. Firstly, measurements of some practical parameters to be used as input in this algorithm are given.

Algorithm 1 FDDE Algorithm

input: T_e, α_{d_0}, D_0 v_0, Δt, θ, t_s
$t_0 = 0$; $i = 1$; $d_0 = D_0$
Calculate Δd by (4.4)
while $t_{i-1} \leq t_s$ **do**
    if $t_{i-1} \leq T_e$ **then**
        $d_i = d_{i-1}$
        Calculate α_{d_i} by (4.8)
    else
        $d_i = d_{i-1} + \Delta d$
        Calculate α_{d_i} by (4.9)
    end if
    Calculate s_i by using the real positive root of (4.34)
    $i = i + 1$
    $t_i = t_{i-1} + \Delta t$
end while

4.3.1. Measurements

The parameters of the sprayer such as θ, D_0 and v_0 are required to be measured to compare the practical experiment results with the proposed algorithm. D_0 is measured by a precise digital caliper, which is a measurement tool with a resolution of 0.01 mm for the length of an object. θ is measured by using image analysis with the software ImageJ. The image used for this analysis is given in Figure 4.2.

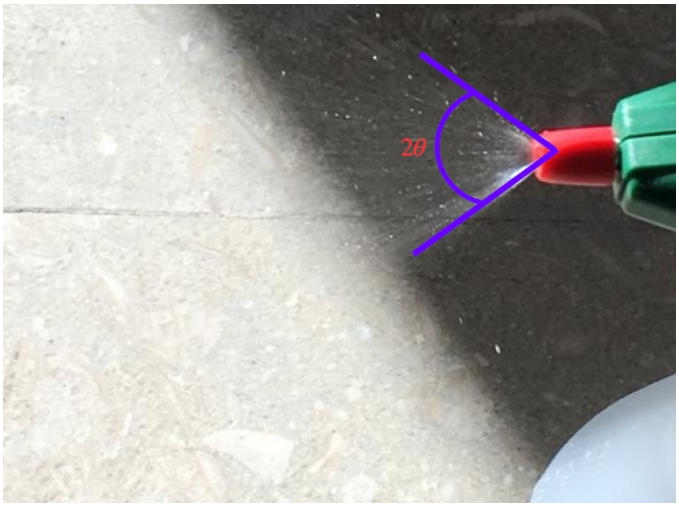


Figure 4.2. Measurement for the beamwidth of the TX.

Then, the volumetric flow rate (Q), which gives the fluid volume flowing through the sprayer per unit time, is measured in order to calculate the initial average velocity of droplets (v_0). For the measurement of Q, a precision balance, which is an equipment measuring the mass with a precision of 0.001 g, is used. The mass of the sprayer filled with the liquid ethanol is measured before and after a short spraying. Meanwhile, the elapsed time for spraying is recorded. In order to find the volume of the sprayed liquid ethanol, the mass is divided by the density of the liquid ethanol (ρ_d) whose value is obtained from (PubChem Database, 2019). Hence, Q can be calculated by dividing the mass difference to the elapsed time for consecutive measurements (Munson et al., 2009b). Q can be presented by the formula

$$Q = \frac{\Delta V}{\Delta t_v}, \tag{4.35}$$

where ΔV and Δt_v show the volume and time difference between the initial and final measurement values. In order to eliminate the random effects, ten measurements are made to calculate Q. Afterwards, the initial average velocity of the droplets can be found by (Munson et al., 2009b)

$$v_0 = \frac{Q}{A_0}, \tag{4.36}$$

where A_0 is the cross-sectional area of the nozzle calculated by $A_0 = \pi(D_0/2)^2$. The results of ten measurements for Q and v_0 are shown in Fig. 4.3. The average of these ten v_0 values is employed for the comparison of the theoretical and experimental results. The measured and calculated values of θ, D_0 and v_0 are given in Table 4.2.

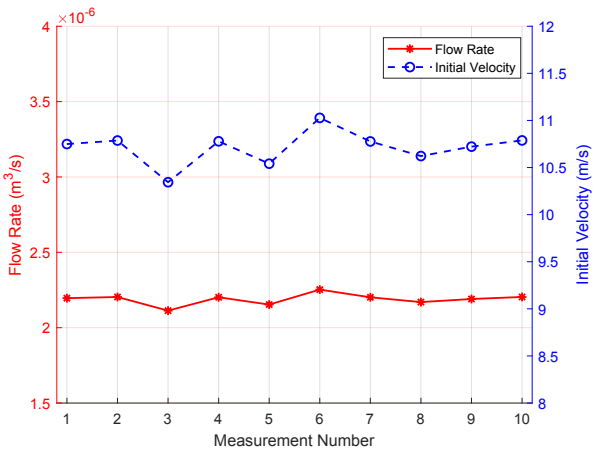


Figure 4.3. Measurement results for flow rate and initial velocity.

For the validation of the FDDE algorithm, the experimental data are collected by using the experimental setup given in Chapter 3. The measurements are made for nine different distances ranging from 1 to 1.8 m in steps of 0.1 m. At each distance, five different emissions are made with $T_e = 0.25$ s from the TX (a single puff) at $20°$C. A sufficient amount of duration (at least five minutes) is left between two consecutive measurements in order to eliminate the effects of the previous transmissions. The room that the measurements are made is ventilated along this duration. In the next part, the experimental parameters given in Table 4.2 are used for the numerical evaluation of the proposed methods.

4.3.2. Results and Analysis

In Table 4.2, Re is given as 0, since the droplet and entrained air velocities are assumed to be equal for the two-phase flow. Furthermore, α_{d_0} is chosen as a small number (0.001) in accordance with the observations about the mixture volume of droplets and air particles. Furthermore, the chemical properties of ethanol and air in Table 4.2 are obtained from (PubChem Database, 2019; Lugg, 1968) and (Mokeba et al., 1997), respectively. ΔP is used as the same value in (Mokeba et al., 1997), since the densities of liquid in (Mokeba et al., 1997) and our study have close values.

Table 4.2. Experimental parameters

Parameter	**Value**
Half-beamwidth of the TX (θ)	$38°$
Diameter of the nozzle orifice (D_0)	510 μm
Average initial velocity of the droplets (v_0)	10.8 m/s
Time step (Δt)	0.01 s
Reynolds number (Re)	0
Emission time of the TX (T_e)	0.25 s
Volume fraction of droplets at $t = T_e$ (α_{d_0})	0.001
Density of liquid ethanol (ρ_d)	789 kg/m^3 (PubChem Database, 2019)
Density of air (ρ_a)	1.2 kg/m^3 (Mokeba et al., 1997)
Molecular weight of ethanol (M_v)	46.069×10^{-3} kg/mol (Mokeba et al., 1997)
Molecular weight of air (M_a)	28.9647×10^{-3} kg/mol (PubChem Database, 2019)
Diffusion coefficient of vapor ethanol (D_v)	11.81×10^{-6} m^2/s (Lugg, 1968)
Partial pressure of air (P_a)	10^5 Pa (Mokeba et al., 1997)
Vapor pressure difference (ΔP)	790 Pa (Mokeba et al., 1997)
Kinematic viscosity of air (ν_a)	1.516×10^{-5} m^2/s (Mokeba et al., 1997)

When the peak time of the sensor voltage, which is assumed as t_i, is measured by the RX, the distance can be estimated by the FDDE algorithm as shown in Figure 4.4. Here, Δt is chosen sufficiently small so that peak time values measured by the sensor correspond to t_i values exactly. Each value in Figure 4.4 denotes the mean value of five estimations for the corresponding distance. In order to evaluate the accuracy of the estimations for each distance, Mean Absolute Percentage Error (ϵ) is chosen as the performance metric as given by

$$\epsilon = \frac{100}{N} \sum_{k=1}^{N} \frac{|\hat{s}_k - s|}{s} \tag{4.37}$$

where N is the number of estimations, $\hat{s}$ and s are the estimated and actual distance values.

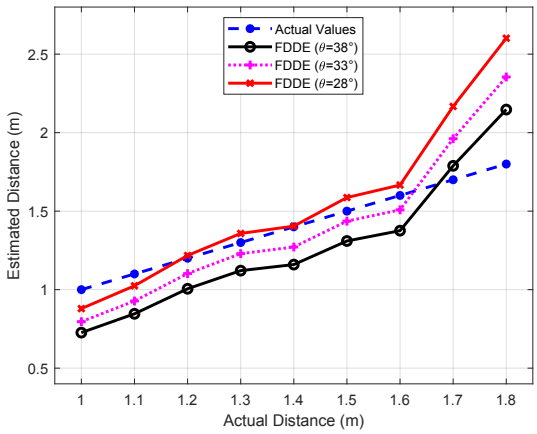


Figure 4.4. Experimental and estimated distance values with the FDDE algorithm.

Figure 4.5 shows the performance of the FDDE algorithm in terms of ϵ. Figs. 4.4 and 4.5 show the effect of θ by using its measured value ($38°$) and two different values ($33°$, $28°$). This effect is related with the physical phenomena about the spraying pattern and interaction of droplets with air particles. The sprayer affects the spraying pattern depending on the parameters such as θ, the spatial dispersion and size of the droplets (Al Heidary et al., 2014). Our observations during the experiments show that the majority of droplets propagate in a narrower beamwidth with respect to the measured value of θ. Furthermore, the interactions among droplets and air particles induce an air flow in the vicinity of the nozzle towards the center of the beamwidth along the horizontal axis. This

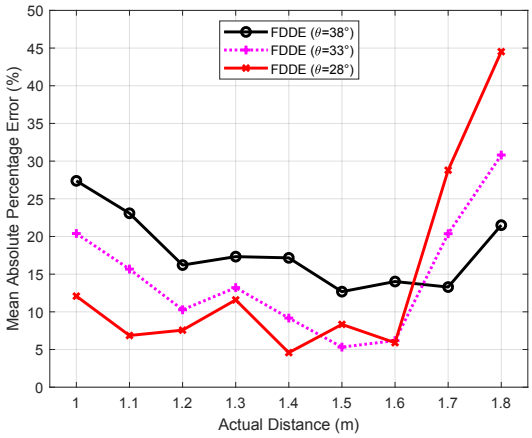


Figure 4.5. Mean Absolute Percentage Error between experimental and estimated distance values.

flow entrains droplets from the outer region into the inner region within the beamwidth (Ghosh and Hunt, 1994). This fact also supports the narrowing beamwidth after spraying droplets. Hence, as θ decreases, the FDDE algorithm gives more accurate results for shorter distances. The narrowing beamwidth can be estimated via high-speed photography techniques by visual analysis of the spray pattern or Phase Doppler Anemometry (Begg et al., 2009). However, the estimation and measurement of this narrowing beamwidth is beyond the scope of this chapter.

In Figs. 4.4 and Figure 4.5, the performance of the FDDE algorithm for each θ value changes dramatically for larger distances. This change can be explained with the effect of diffusion for the droplets at longer distances as also discussed in (Gulec and Atakan, 2020a). The effect of diffusion is negligible for large droplets at smaller distances. However, as the distance increases, the average velocity of droplets is affected by diffusion due to decreasing droplet sizes. Therefore, the FDDE algorithm gives better results for shorter distances.

In Fig. 4.6, the performance of FDDE algorithm is compared with the statistical distance estimation methods in Chapter 3. These statistical methods are linear regression (LR) and neural network regression (NNR) as the machine learning, and power based (PBE), peak time based (PTBE) and combined estimation (CE) as the data analysis based methods. For shorter distances (up to 1.5 m), the FDDE algorithm has a better perfor-

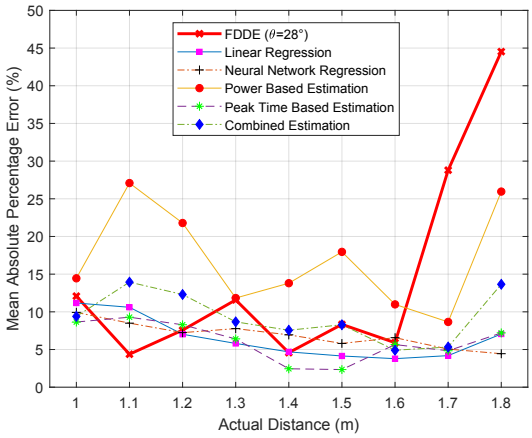


Figure 4.6. The performance comparison of FDDE algorithm with the distance estimation methods given in Chapter 3.

mance than PBE and CE and nearly the same performance with machine learning methods. However, for longer distances (after 1.6 m), the error for FDDE algorithm increases sharply due to the decrement in average velocity of droplets and effect of diffusion.

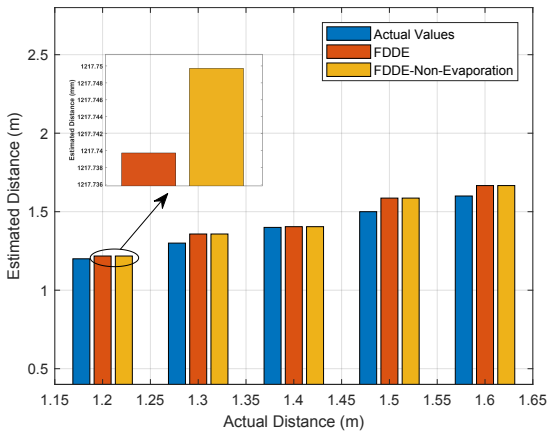


Figure 4.7. The effect of evaporation in FDDE algorithm.

The cost of better performance of statistical methods (especially for longer distances) is the long process of data collection and feature extraction to generate the estimation parameters, and learning process by using these data. However, FDDE algorithm can give good results with simpler calculations and without collecting data. Therefore, the complexity of the FDDE algorithm is less than statistical methods in terms of time and

computation. Moreover, FDDE algorithm can be implemented by using physically meaningful and measurable parameters instead of deriving parameters according to collected data for distance estimation.

Figure 4.7 shows the effect of the droplet evaporation for distance estimation with the FDDE algorithm. The results in this figure for the non-evaporation case is obtained by ignoring the evaporation of droplets after T_e. For our experimental scenario, the effect of evaporation is small as it can be seen in magnified view of the results for 1.2 m in Figure 4.7. The effect of evaporation for other distances in this figure are similar to the results obtained for 1.2 m. Although this effect is very small for short distances, it can be more influential for longer distances and higher air temperatures due to the increment of the evaporation.

CHAPTER 5

A DROPLET-BASED SIGNAL RECONSTRUCTION APPROACH TO CHANNEL MODELING IN MOLECULAR COMMUNICATION

5.1. Introduction

In a macroscale MC application, employing an accurate channel model enables a more efficient information transfer between the TX and RX. In addition, the location of a molecular TX can be estimated via accurate channel models (Qiu et al., 2015; Gulec and Atakan, 2020b,a). As an exemplary application, an infected human emitting pathogen-laden droplets into the air through sneezing or coughing can be considered as a molecular TX in public places. By deploying biological sensors as molecular RXs, this scenario can be evaluated as a macroscale MC system (Khalid et al., 2019). Actually, the communication between a sneezing/coughing human and a biological sensor is very similar to the system given in (Farsad et al., 2013) and used in our study, since sneezing/coughing can be considered as spraying droplets into the air with an initial velocity. Therefore, an accurate channel model in such a practical setup can be employed to model the spreading mechanism of infectious diseases through airborne transmission and detecting diseases in public places. In (Farsad et al., 2014) and (Kim et al., 2015), channel models for the platform in (Farsad et al., 2013) are derived by modifying the solution of the diffusion equation and model coefficients are estimated by fitting experimental data. Since the models in (Farsad et al., 2014) and (Kim et al., 2015) are based on the molecule diffusion assumption, correction factors are needed to modify the diffusion equation. In these studies, it is not clearly known what fitted parameters of the modified diffusion equation are physically related to. Since the TX sprays liquid droplets rather than molecules and droplets are larger with respect to molecules, only diffusion is not sufficient for channel modeling in practical scenarios. Thus, initial velocities and interactions of droplets

with air molecules and the surface of the sensor (RX) need to be considered. Hence, a fluid dynamics perspective considering droplets as information carriers is required for a more accurate channel model as discussed in Chapter 3. In (Abbaszadeh et al., 2018) and (Li et al., 2020), turbulent flows and vortex rings without considering droplets are used with this perspective. However, the accuracy of the received signals by the sensor is not considered. Therefore, the droplet-sensor interaction and the accuracy of the sensed concentration by the sensor need to be taken into account.

In contrast to the models in (Farsad et al., 2014) and (Kim et al., 2015) based on the diffusion of molecules, this chapter proposes a more realistic system model which reveals the physical meanings of the channel parameters. In this approach, droplets sprayed by the TX are considered as information carriers in the channel rather than diffusing molecules. The interaction of these droplets with air generates a two-phase flow where the first phase is liquid and the second phase is gas. In two-phase flow, droplets and air molecules move together in liquid and gas phases, respectively. The TX is modeled as a directed emitter with a predefined beamwidth. Droplets are assumed to move in a cone shaped volume determined by this beamwidth. The model also quantifies how accurate the transmitted signal is reconstructed by the RX by examining movements of droplets, interactions among droplets and surface of the sensor. Therefore, this part of the model is called as signal reconstruction (SR). The SR involves two consecutive processes which are the adhesion/detachment process of droplets and the sensitivity response of the sensor. The resulting end-to-end system response is derived and validated by experimental data. It is revealed that the end-to-end system response of a practical MC system depends on the distance, parameters of the sensor measurement circuit, sensitivity characteristic of the sensor, beamwidth of the TX, spray coefficient and reaction rate constants in the adhesion/detachment process. The study presented in this chapter was first introduced in (Gulec and Atakan, 2021a).

5.2. A Practical Channel Model Based on Fluid Dynamics and Signal Reconstruction

In this section, we introduce the end-to-end system model which was validated by using the experimental setup given in Figure 5.1. The details of this experimental setup

was given in Chapter 3.

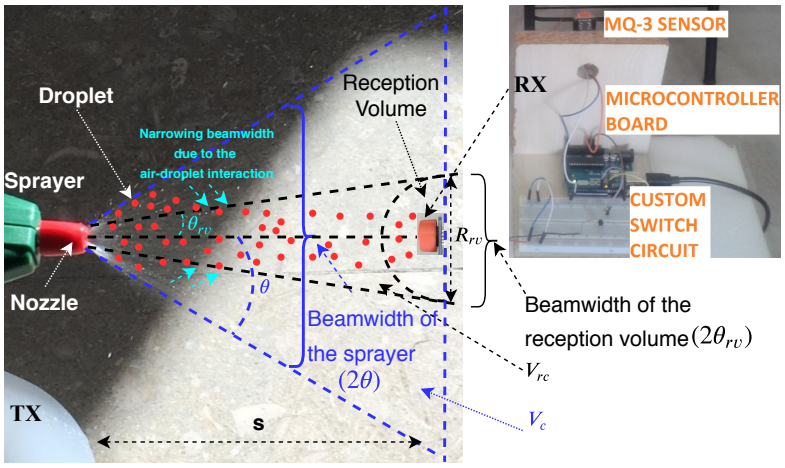


Figure 5.1. The experimental setup and parameters.

For the system model shown in Figure 5.2, the transmitted mass of droplets is taken as the input and the measured sensor voltage is the output of the system. The end-to-end system impulse response ($E_{out}(t)$) is defined as the system's output to a short spray emission along T_e (without a fan) which can be considered as an impulsive input signal. For the derivation of $E_{out}(t)$, the propagation time of droplets (t_0) is needed to be estimated. However, we only focus on the signal after t_0. Solely, the effect of two-phase flow on the signal in the reception volume (RV) of the RX is investigated. In addition, the initial offset voltage of the sensor is eliminated by subtracting the sensor voltage at t_0 from the whole signal.

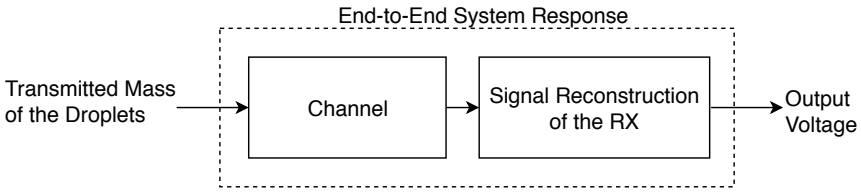


Figure 5.2. Block diagram of the end-to-end system model.

5.2.1. The Effect of Two-Phase Flow on the Initial Concentration in the Reception Volume

As shown in Figure 5.1, droplets are assumed to move in a cone-shaped volume (Ghosh and Hunt, 1994) and the beamwidth (2θ) is defined as the initial spraying angle of the sprayer's nozzle. The interaction among droplets and air molecules creates a two-phase flow where droplets and air molecules move together as a mixture (Ishii and Hibiki, 2010). Here, the first phase is the liquid phase of droplets and the second phase is the gas phase of air molecules. Due to this interaction, the majority of droplets move in a narrower beamwidth, as they propagate in the channel (Ghosh and Hunt, 1994). This narrower beamwidth, denoted by $2\theta_{rv}$, is assumed to encompass the RV as illustrated in Figure 5.1. Hence, the beam of droplets forms two concentric cones. The volume and base diameter of the inner cone are denoted by V_{rc} and R_{rv}, respectively. The volume of the outer cone is V_c and the distance between the TX and RX is s as depicted in Figure 5.1. Our approach based on fluid dynamics allows us to accurately model the initial concentration around the sensor by taking into account the droplet-air interaction and the narrowing beamwidth.

If the spatial distribution of droplets is assumed to be homogeneous, the mass of droplets in the inner cone can be found via multiplying the total transmitted mass, i.e., m_{TX}, by the ratio of the cone volumes, i.e., V_{rc}/V_c. However, due to the interactions among droplets and air molecules, the propagation of droplets is far from homogeneity (Ghosh and Hunt, 1994). Therefore, in order to quantify the inhomogeneous propagation of droplets, we define the spray coefficient (γ). By employing the ratio of the volumes, i.e., V_{rc}/V_c, and γ, a scaling factor (η) can be introduced to obtain the mass of droplets in V_{rc} as given by

$$\eta = \frac{V_{rc}}{V_c}\gamma = \frac{\frac{\pi}{3}s\left(s\tan\theta_{rv}\right)^2}{\frac{\pi}{3}s(s\tan\theta)^2}\gamma = \left(\frac{\tan\theta_{rv}}{\tan\theta}\right)^2\gamma, \tag{5.1}$$

where $1 \leq \gamma \leq V_c/V_{rc}$. Here, $\gamma = 1$ means that the spatial distribution of droplets are homogeneous within the beamwidth of the outer cone (2θ), and γ=V_c/V_{rc} means that all droplets propagate in the beamwidth of the inner cone ($2\theta_{rv}$). Then, the initial droplet concentration in the beamwidth of the RV (C_0) is derived by employing (5.1) as follows:

$$C_0 = \frac{m_{TX}\eta}{V_c} = \frac{m_{TX}\left(\frac{\tan\theta_{rv}}{\tan\theta}\right)^2\gamma}{\frac{\pi}{3}s(\frac{R_{rv}}{2})^2} = \frac{3m_{TX}\gamma}{\pi s^3\left(\tan\theta\right)^2}. \tag{5.2}$$

Here, C_0 is assumed as the initial value of the time-dependent concentration in the RV ($C(t)$) which is derived in the next part of the chapter. In (5.2), m_{TX} can be presented by using the volumetric flow rate (Q) of the TX which gives the fluid volume flowing through the sprayer per unit time (Munson et al., 2009b). Here, $Q = V_{TX}/T_e$ where V_{TX} shows the emitted liquid volume. Since the transmitted mass can be written as $m_{TX} = V_{TX}\rho_d$ where ρ_d is the density of the liquid forming droplets before spraying, (5.2) is given by

$$C_0 = \frac{3QT_e\rho_d\gamma}{\pi s^3 \left(\tan\theta\right)^2} = \frac{3V_{TX}\rho_d\gamma}{\pi \left(s\tan\theta\right)^2 s}. \tag{5.3}$$

In fluid dynamics, the effects of some parameters such as the length or volume can be reduced by using dimensionless parameters. These dimensionless parameters are also utilized for a more concise expression (Munson et al., 2009b). To this end, the characteristic length (L_c) is defined as

$$L_c = \frac{V_{TX}}{A_c} = \frac{V_{TX}}{\pi \left(s\tan\theta\right)^2} \tag{5.4}$$

where A_c is the cross-sectional area of the outer cone. By using L_c, a dimensionless parameter (s^*) as the normalized distance is defined as $s^* = s/L_c$. Hence, (5.3) is simplified as given by

$$C_0 = \frac{3\rho_d\gamma}{s^*}. \tag{5.5}$$

5.2.2. Signal Reconstruction of the Receiver

For our scenario, the reconstruction of the molecular signal around the RX is subject to an error due to the random adhesions/detachments of droplets to/from the sensor and the sensitivity of the sensor. Hence, the SR is modeled as the combination of the adhesion/detachment process and the sensitivity response of the sensor as shown in Figure 5.3.

Let X and Y represent the droplets in the RV and the adhered droplet-sensor complex, respectively. Moreover, Z is defined as the detached state of the droplet which is assumed not to be sensed by the sensor. In order to provide a better understanding, an analogy between the adhesion/detachment process of droplets and the reception process of

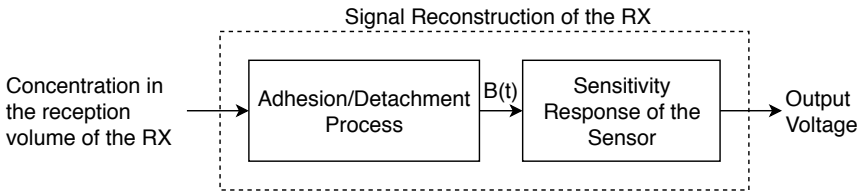


Figure 5.3. Block Diagram of the Signal Reconstruction.

a biological cell which captures/releases molecules with its receptors can be established. Here, droplets are assumed to arrive the RV instantaneously after the emission in order to relate the chemical kinetics of droplets in the RV and the sensor measurement as a function of time. The adhesion and detachment can be modeled as first order reactions which are given by

$$\mathrm{X} \xrightarrow{\mathrm{k_1}} \mathrm{Y} \xrightarrow{\mathrm{k_2}} \mathrm{Z}, \tag{5.6}$$

where k_1 and k_2 are the rate constants of the corresponding reaction. Let $C(t)$ and $B(t)$ denote the concentrations of X and Y in kg/m^3, respectively. Based on the reaction system in (5.6) and the rate law (Atkins and De Paula, 2010), the concentrations $C(t)$ and $B(t)$ can be characterized as

$$\frac{dC(t)}{dt} = -k_1 C(t) \tag{5.7}$$

$$\frac{dB(t)}{dt} = k_1 C(t) - k_2 B(t), \tag{5.8}$$

where the initial conditions are defined as $C(0) = C_0$ and $B(0) = 0$. The solution of (5.7)-(5.8) for $B(t)$ can be given as

$$B(t) = \frac{k_1 C_0}{k_2 - k_1}[e^{-k_1 t} - e^{-k_2 t}]. \tag{5.9}$$

Subsequent to adhesion/detachment process, $B(t)$ is converted to an electrical signal by the metal-oxide MQ-3 sensor. Sensors of this type measure the concentration around them by changing their resistance so that each concentration value corresponds to a sensor resistance (R_s) value as shown in Figure 5.4. In order to normalize the measured R_s, R_o is defined as the sensor resistance measured at 0.0004 kg/m^3 (Hanwei-Electronics, 2018). For each concentration value, R_s/R_o determines how sensitive the sensor can measure. This sensitivity characteristic, whose values are taken from its datasheet (Hanwei-Electronics, 2018), can be employed to obtain a sensitivity function $(f(t))$ which maps

concentration values to R_s/R_o values via curve fitting technique. The datasheet values are fitted by using nonlinear least squares method which minimizes the sum of square errors. $f(t)$ can be fitted as given by

$$f\left(B\left(t\right)\right) = 0.0116\left(B(t)\right)^{-0.5855} - 0.0743 = \frac{R_s}{R_o}, \tag{5.10}$$

where $B(t)$ is the input to the sensitivity response of the sensor as shown in Figure 5.3 and derived in (5.9). The scalar curve fitting parameters are estimated by employing Levenberg-Marquardt algorithm which has an estimation performance with the Root Mean Square Error (RMSE) value of 0.0371 (Hagan and Menhaj, 1994).

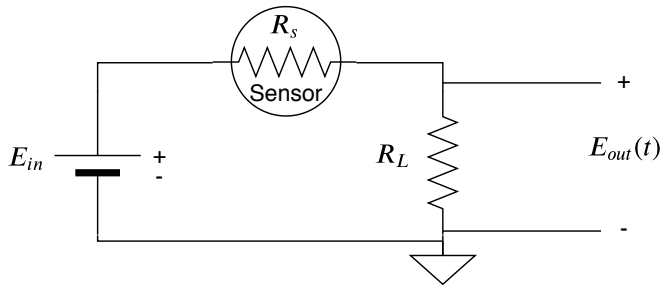


Figure 5.4. Sensor Measurement Circuit.

In our experimental setup, the sensor measurement is made with a circuit as given in Figure 5.4 where R_L is the load resistance, E_{in} is the input voltage and $E_{out}(t)$ is the output voltage which also gives the end-to-end system impulse response for an impulsive input signal. Using this circuit, R_s can be derived via Kirchhoff's voltage law (Nilsson and Riedel, 2010) as given by

$$R_s = \left(\frac{E_{in}}{E_{out}(t)} - 1\right) R_L, \tag{5.11}$$

where E_{in} is given as 5 V in (Hanwei-Electronics, 2018). By combining (5.10) and (5.11), the relation between $f(B(t))$ and the parameters of the sensor measurement circuit can be written as

$$f(B(t)) = \left(\frac{E_{in}}{E_{out}(t)} - 1\right) \frac{R_L}{R_o}. \tag{5.12}$$

By using (8.16), $E_{out}(t)$ is given as

$$E_{out}(t) = \frac{E_{in} R_L}{R_o \left(f(B(t)) + \frac{R_L}{R_o}\right)}. \tag{5.13}$$

Finally, the end-to-end impulse response can be expressed by substituting (5.9) into (5.13) as given by

$$E_{out}(t) = \frac{E_{in}R_L}{R_o\left[f\left(\frac{k_1C_0}{k_2-k_1}(e^{-k_1t}-e^{-k_2t})\right)+\frac{R_L}{R_o}\right]}, \tag{5.14}$$

where it is important to note that C_0 is a function of the spray coefficient γ as given in (5.3). Therefore, the impulse response in (5.14) involves three novel channel parameters, i.e., k_1, k_2 and γ. These parameters are affected by the change of other parameters in the channel such as the distance between the TX and RX. Especially, γ depends on the type of the sprayer's nozzle which affects the spraying pattern (Al Heidary et al., 2014) and interactions among droplets and air molecules. For numerical results, k_1, k_2, and γ are manually configured by making the Mean Square Error (ϵ) between the samples of $E_{out}(t)$ and experimental signal ($F(t)$) as small as possible according to the formula which is given as

$$\epsilon = \frac{1}{N}\sum_{n=1}^{N}\left(E_{out}[n]-F[n]\right)^2 \tag{5.15}$$

where N shows the number of samples, $E_{out}[n]$ and $F[n]$ are discrete-time representations of $E_{out}(t)$ and $F(t)$, respectively. Next, it is shown by numerical results that the proposed model given in (5.14) can be used for practical scenarios.

5.3. Numerical Results

In this section, the proposed channel model is validated by experimental data. First, the measurement of some practical parameters are explained.

5.3.1. Measurements

In order to measure Q, the sprayer is placed on a precision balance. For each measurement, the liquid is sprayed from the sprayer for a short interval (Δt_v). The mass of the sprayer is measured before and after spraying. Hence, the mass difference is found by measuring the mass values before and after spraying. By dividing the mass difference to ρ_d, the volume difference (ΔV) is calculated for each measurement. Thus, Q can be

presented by (Munson et al., 2009b)

$$Q = \frac{\Delta V}{\Delta t_v}. \tag{5.16}$$

Here, the average of ten measurements is considered for Q. In addition, R_o is calculated by using $E_{out}(t)$ value when the concentration value is 0.0004 kg/m^3. The detection scope of the sensor which is between 5×10^{-5} and 10^{-2} kg/m^3, is scaled for $E_{out}(t)$ values between 0 and 5 V (Hanwei-Electronics, 2018). Hence, $E_{out}(t) = 0.2$ V for the concentration value of 0.0004 kg/m^3. Then, R_o is calculated by (5.11). Furthermore, θ is measured with ImageJ software using the image given in Figure 5.1. Next, the values of the experimental parameters given in Table 5.1 are used for the validation of the proposed model.

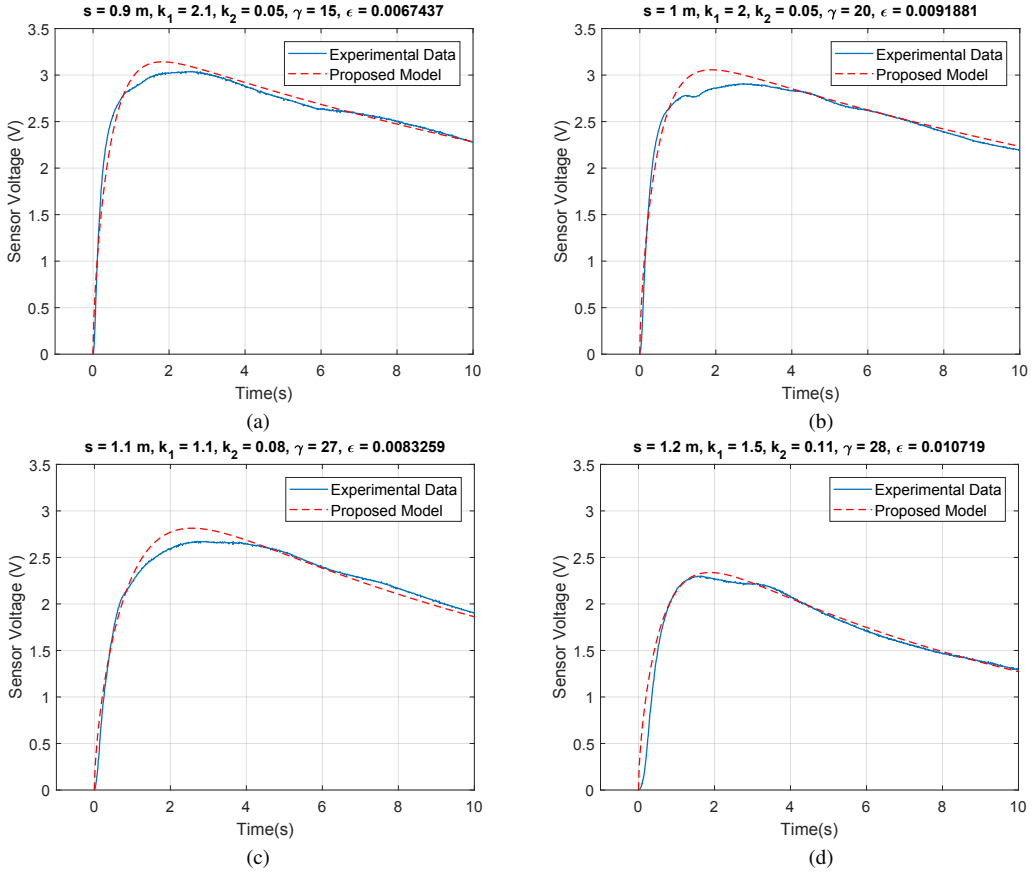


Figure 5.5. The comparisons of the proposed model with experimental data for the parameters given in the title of each signal.

Table 5.1. Experimental parameters

Parameter	**Value**
Distance between the TX and RX (s)	$\{0.9, 1, 1.1, 1.2\}$ m
Volumetric flow rate (Q)	2.204×10^{-6} m^3/s
Density of liquid ethanol (ρ_d)	789 kg/m^3
Emission time of the TX (T_e)	0.5 s
Load resistance (R_L)	1 kΩ
Sensor resistance at 0.0004 kg/m^3 (R_o)	24 kΩ
Half-beamwidth of the sprayer (θ)	38°

5.3.2. Results for Channel Modeling

In Figure 5.5, the comparisons of experimental data and the proposed model are shown. For performance evaluation, ϵ is calculated for a duration of 10 s. Figure 5.5 validates that the proposed channel model can be employed to estimate the received signal.

As proposed in (Atakan and Gulec, 2019) for diffusion-based MC in microscale, the SR of the RX may result in error due to the random movements of the molecules in the reception volume. Accordingly for our macroscale scenario, k_1 and k_2 represent the average characteristic of the random movements of the droplets in the RV. In addition, γ describes the effect of the two-phase flow on the initial concentration in the RV.

As shown in Figure 5.5, γ increases as s increases, since the shape of droplets' spatial dispersion gets narrower with s. The increment of γ with s validates the narrowing beamwidth at longer distances due to the droplet-air interaction (Ghosh and Hunt, 1994). Moreover, the slope of the rising edge of the signals in Figure 5.5 between 0 and peak time is proportional to k_1. This slope decreases as s increases, since k_1 is proportional to the energy of droplets entering the RV. At longer distances, the droplet-sensor interaction in the RV decreases due to the decreasing droplet energy. In addition, as k_2 increases, the adhered droplets on the sensor detach faster and the sensor voltage drops faster from its peak value to its initial level.

In Figure 5.6, twenty different manually fitted signals (with a ϵ of less than 0.021) are used to observe the relation of k_1 and k_2 with the normalized distance (s^*). Droplets are more unlikely to stay adhered to the sensor due to the decreasing energy of droplets for longer distances, as shown with the change of average k_1 values in Figure 5.6. In contrast to k_1, k_2 is almost constant, since the energy of droplets is not so effective on

their detachment from the sensor in the RV. Although physically measurable parameters (k_1, k_2, γ) are proposed for the channel model, there is a need for further research effort to investigate the relation of these parameters with the environmental conditions.

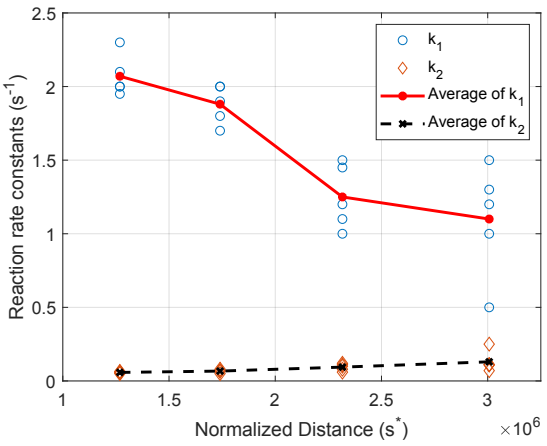


Figure 5.6. The relation of k_1 and k_2 with the normalized distance.

CHAPTER 6

A MOLECULAR COMMUNICATION PERSPECTIVE ON AIRBORNE PATHOGEN TRANSMISSION AND RECEPTION VIA DROPLETS GENERATED BY COUGHING AND SNEEZING

6.1. Introduction

Molecular communication (MC) can be helpful for practical macroscale applications which include several experimental platforms using molecules or droplets for the information transfer (Unterweger et al., 2018; McGuiness et al., 2018; Abbaszadeh et.al., 2019; Khaloopour et.al., 2019; Hamidović et al., 2019). Furthermore, MC is utilized to solve practical problems such as finding the distance to a molecular source (Gulec and Atakan, 2020a) as covered in Chapter 3 and 4.

Hence, MC can be employed as a tool to model the biological phenomena which consider droplets such as the transmission and reception of pathogens (viruses, bacteria, etc.) which cause contagious diseases via droplets. This concept is first proposed in (Khalid et al., 2019) which consider the infectious human as a blind transmitter (TX) emitting pathogen-laden droplets and sensors as the receiver (RX) for outdoor environments. This study is improved in (Khalid et al., 2020) by taking silicon nanowire field effect transistor-based biosensors into account to model pathogen detection for airborne pathogen transmission.

Airborne transmission and self-inoculation (direct contact) are two modes of infectious disease transmission via pathogens (Bourouiba et al., 2014). Droplets can be classified as large droplets and droplet nuclei (aerosols) which have sizes of larger and smaller than 10 μm, respectively (Ai and Melikov, 2018). All expiratory activities such as coughing, sneezing, breathing and speaking can generate large droplets and aerosols. While large droplets can be effective in short-range, aerosols can spread pathogens to

longer distances due to their interactions with air. Airborne transmission via droplets is a significant infection mechanism for pathogens such as influenza virus (Killingley and Nguyen-Van-Tam, 2013), severe acute respiratory syndrome (SARS) virus (Peiris et al., 2003) and new SARS coronavirus-2 (SARS-CoV-2) which causes coronavirus disease 2019 (COVID-19) (Prather et al., 2020). By the time this chapter is written, the global pandemic of COVID-19 still continues and there is no cure for COVID-19. It is essential to emphasize that this study is related with the COVID-19 outbreak, since one of the main mechanisms of this disease is airborne transmission.

In fluid dynamics literature for airborne pathogen transmission, the movement of droplets are modeled by considering the gravity, interaction of droplets with air, evaporation, airflows and droplet size (Mittal et al., 2020). For coughing, (Wei and Li, 2017) proposes a two-stage jet flow model, which is based on experimental data. In (Bourouiba et al., 2014), the propagation of human sneeze and cough is modeled as a cloud consisting of droplets and air and justified by experimental data for a sneezing/coughing human. Analytical approaches are proposed in (Wei and Li, 2015; Liu et al., 2017; Xie et al., 2007) to estimate the trajectories of large droplets and aerosols by considering evaporation and turbulent flows for coughing. In (Wang et al., 2020), a Lagrangian particle model and a discrete random walk model are proposed for the movement of expiratory droplets and turbulent airflows, respectively.

The data collected for airborne pathogen transmission are based on physical experimental setups. These setups consist of humans, thermal manikins or respiratory machines for the emission of droplets, and air samplers or imaging devices to measure the droplet concentration after the emission (Ai and Melikov, 2018). Despite the reliability of the collected data from physical setups, computational fluid dynamics (CFD) simulators are preferred more due to their low cost and high resolution in time and space. In CFD simulations, 3-D Navier-Stokes equations are employed with boundary conditions and solved by numerical methods. In (Pendar and Páscoa, 2020), the breakup process of droplets, droplet-air interaction and resulting turbulent flows are considered for the movement of droplets which are emitted by coughing and sneezing with and without a mask. The study in (Dbouk and Drikakis, 2020) takes into account the effect of evaporation and wind for the dispersion of droplets. (Busco et al., 2020) proposes a CFD method to model the dispersion of droplets for sneezing and validates it with experimental data.

The aforementioned studies for airborne pathogen transmission focus on the dispersion of droplets for coughing/sneezing. Very few studies such as (Vuorinen et.al, 2020) model the reception of droplets via considering the inhalation rate of a human. In this chapter, we investigate airborne pathogen transmission and reception mechanisms between humans with a MC perspective for indoor environments. The infectious human which emit a cloud consisting of pathogen-laden droplets and air by coughing/sneezing is considered as the TX and the uninfected human is defined as the RX unlike the studies in (Khalid et al., 2019, 2020) where the RX is a biosensor. Furthermore, the effects of gravity, buoyancy, and air-droplet interactions are taken into account for the indoor propagation of the cloud, which are not considered in (Khalid et al., 2020). The propagation of the cloud in the MC channel is modeled by modifying the deterministic model in (Bourouiba et al., 2014) in a probabilistic way. The cloud travels under the influence of initial velocity, buoyancy and gravity. The number of droplets in this cloud is modeled as a random process. A receiver model which takes the central part of human face as the interface with pathogens into account is proposed. The propagation and reception models are employed for the proposed end-to-end system model in order to give the infection state of the RX as the system output. The MC perspective gives the opportunity to handle the airborne pathogen transmission modeling as a system comprised of TX, channel and RX. Moreover, this perspective enables to approach the estimation of a human's infection state as a binary detection problem according to a threshold defined as the sufficient number of pathogen-laden droplets for infection. The contributions of this chapter can be summarized as follows:

- In order to model the spread of the infectious diseases between humans, MC perspective leading us to use and adapt the well-known communication engineering techniques is proposed.

- An end-to-end system model which combines a channel model involving the propagation of pathogen-laden droplets as a cloud and a RX model including the interaction of these droplets with the uninfected human is proposed.

- A probabilistic approach which enables the derivation of the probability of infection for an uninfected human is employed.

Furthermore, the proposed model is evaluated by numerical results. Our key findings for

a coughing TX is given as follows:

- Increased exposure time to pathogens increases the probability of infection.
- For a horizontal cough, the social distance should be at least 1.7 m.
- It is safer to cough with an initial angle less than -25° to infect less people.

The rest of the chapter is organized as follows. In Section 6.2, the proposed end-to-end system model is presented. Section 6.3 provides the derivation of the probability of infection and the numerical results are given in Section 6.4. The study given in this chapter was first introduced in (Gulec and Atakan, 2021c).

6.2. End-to-End System Model

This section provides a detailed explanation of the proposed end-to-end system model for droplet-based MC between two humans via sneezing/coughing in four steps. As given in Figure 6.1, this model incorporates the airborne pathogen-laden droplet transmission with the reception of these droplets by the human that is considered as the RX. Figure 6.1 also shows that the end-to-end system impulse response is defined as the infection state which is the output of the end-to-end system, since a sneeze/cough can be considered as an impulsive input signal. Transmitted droplets via sneezing/coughing are modeled as a cloud which is a mixture of air and droplets. The propagation of the cloud is defined as a two-phase flow where the first and second phase represent the liquid phase of droplets and the gas phase of the air, respectively (Munson et al., 2009b). As the first step of the end-to-end system model, the trajectory of the cloud is derived. In this study, the model given in (Bourouiba et al., 2014) for the propagation of the cloud is adopted and modified. In the second step, we derive an end-to-end system model with a probabilistic approach instead of the deterministic approach in (Bourouiba et al., 2014). The third step details the RX model which includes signal reconstruction, integration, quantization and detection parts. In the last step, the algorithm for the implementation of the system model is given.

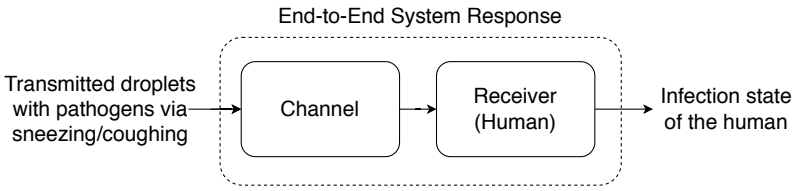


Figure 6.1. Block diagram of the end-to-end system model.

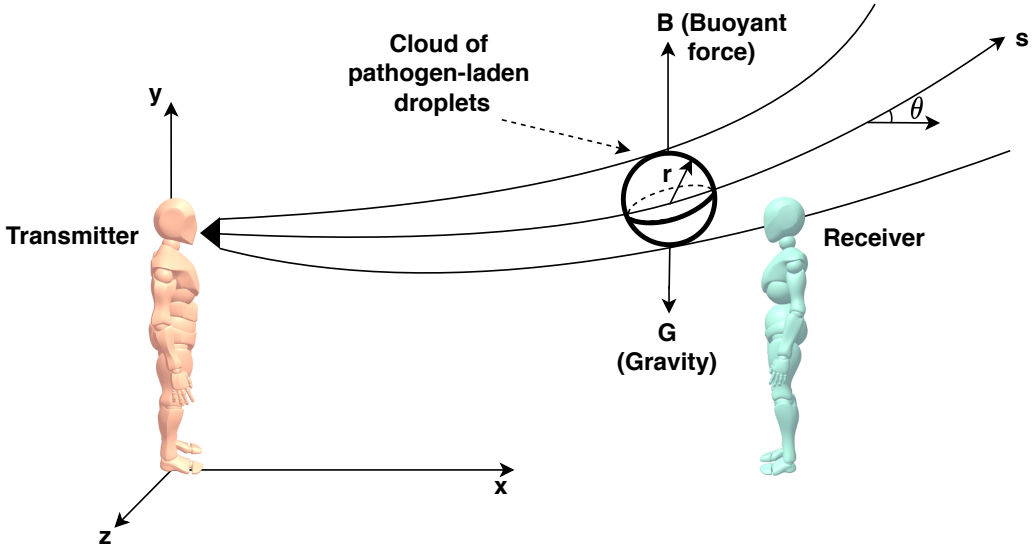


Figure 6.2. Trajectory of the cloud between the TX and RX.

6.2.1. Trajectory of the Transmitted Cloud

In our scenario, the TX emits the cloud with an initial velocity on the x-axis by sneezing or coughing. It is assumed that the emitted cloud propagates in still air, i.e., with no ventilation or wind. Due to the warmer air in the mouth (with density ρ_f at $34°C$) with respect to ambient air (with density ρ_a at $23°C$), where $\rho_f < \rho_a$ (Duguid, 1946), the emitted cloud is subject to buoyancy on y-axis. As illustrated in Figure 6.2, buoyancy, gravity and initial velocity of the cloud affect the trajectory of the cloud. Therefore, the trajectory is defined with the curvilinear s-axis and θ which shows the angle between the s and x axes. In order to find the position of the cloud in 3-D space, it is essential to derive the time-dependent density of the cloud ($\rho_c(t)$). At the initial state ($t = 0$), the initial cloud mass ($m_c(0)$) can be represented by the addition of the initial droplet mass

($m_d(0)$) and initial air mass in the cloud ($m_a(0)$) as given by (Munson et al., 2009b)

$$m_c(0) = m_d(0) + m_a(0) \tag{6.1}$$

$$\rho_c(0)V(0) = \rho_d V_d(0) + \rho_f(V(0) - V_d(0)). \tag{6.2}$$

where $V(t)$ is the cloud volume, ρ_d is the droplet density, ρ_f is the air density in the mouth. The volume of droplets in the cloud ($V_d(t)$) consists of different sized droplets with diameter d_k and can be defined in terms of the volume fraction of droplets ($\phi_k(t)$) in the cloud as $V_d(0) = \sum_{k=1}^{K} \phi_k(0)V(0)$. Here, K is the number of the different droplet sizes, $\phi_k(0) = N_k(0)V_k/V(0)$, $N_k(t)$ and V_k is the number and volume of the spherical droplets of diameter d_k, respectively. $N_k(t)$ changes during the propagation of the cloud due to the settling of droplets to the ground as explained in Section 6.2.2. By substituting $V_d(0)$ into (6.2) and solving for $\rho_c(0)$, the initial density of the cloud is derived as

$$\rho_c(0) = \sum_{k=1}^{K}(\rho_d - \rho_f)\phi_k(0) + \rho_f. \tag{6.3}$$

As the cloud moves, it entrains the ambient air with density ρ_a and its volume becomes $V(t) = V(0) + V_a(t)$ where $V_a(t)$ is the acquired air volume. Since the initial volume fraction of the air in the mouth with density ρ_f is relatively small in the moving cloud (10^{-5}) (Duguid, 1946), it is assumed as $V(t) \approx V_a(t)$. After the emission of droplets, $\rho_c(t)$ can be derived via the conservation of mass (Munson et al., 2009b). Hence, the mass of droplets at the time instance t ($m_d(t)$) is equal to the initial mass of droplets ($m_d(0)$) and droplet mass can be expressed by the difference of cloud mass and air mass according to (6.1) as given by

$$m_d(t) = m_d(0) \tag{6.4}$$

$$\begin{pmatrix}\text{Cloud mass}\\ \text{at time } t\end{pmatrix} - \begin{pmatrix}\text{Air mass}\\ \text{at time } t\end{pmatrix} = \begin{pmatrix}\text{Initial}\\ \text{cloud mass}\end{pmatrix} - \begin{pmatrix}\text{Initial}\\ \text{air mass}\end{pmatrix} \tag{6.5}$$

$$\rho_c(t)V(t) - \rho_a V(t) = \rho_c(0)V(0) - \rho_f V(0). \tag{6.6}$$

Via the substitution of (6.3) into (6.6) and some algebraic manipulation, the cloud density is derived as

$$\rho_c(t) = \sum_{k=1}^{K}(\rho_d - \rho_f)\phi_k(t) + \rho_a. \tag{6.7}$$

As illustrated in Figure 6.2, there are two acting forces on the cloud which stem from the gravity and buoyancy on y-axis. Since $\rho_f < \rho_a$, the buoyant force ($B(t)$) affects

the movement of the cloud upwards. The net buoyant force acting on the cloud ($F_B(t)$) on y-axis is given by the difference of $B(t)$ and the gravitational force ($G(t)$) as given by (Munson et al., 2009b)

$$F_B(t) = B(t) - G(t) = V(t)\rho_a g - V(t)\rho_c(t)g \tag{6.8}$$

$$= V(t)(\rho_a - \rho_c(t))g \tag{6.9}$$

where g is the gravitational acceleration.

On x-axis, the movement of the cloud is driven by the momentum (I) which is defined as the multiplication of the mass and velocity (Munson et al., 2009b). The momentum is not effective on y-axis and also there is not any acting force on the cloud for $z-$axis. I is defined on s-axis and is decomposed into two components on x ($I_x = |I|\cos(\theta)$) and y axes ($I_y = |I|\sin(\theta)$). Since the force can be represented as the derivative of the momentum (Munson et al., 2009b) and there is not any acting force on the x-axis during the propagation, the net force (F_x) on x-axis is given by

$$F_x = \frac{dI_x}{dt} = \frac{d|I|\cos(\theta)}{dt} = 0. \tag{6.10}$$

Furthermore, the net force on y-axis, i.e., $F_B(t)$ is given by

$$F_B(t) = \frac{dI_y}{dt} = \frac{d|I|\sin(\theta)}{dt}. \tag{6.11}$$

Since the initial buoyancy is conserved (Bourouiba et al., 2014), we have $F_B(t) = F_0$ where F_0 is the net initial buoyant force. With the initial conditions which are $|I(0)| = I_0$, $\theta(0) = \theta_0$, $I_x(0) = I_0\cos(\theta_0)$ and $I_y(0) = I_0\sin(\theta_0)$, I_x and I_y can be given as the solutions of (6.10) and (6.11) as

$$I_x = I_0\cos(\theta_0), \tag{6.12}$$

$$I_y = F_0 t + I_0\sin(\theta_0). \tag{6.13}$$

Since there is not any acting force on z-axis, the momentum can be expressed as $|I| = \sqrt{I_x^2 + I_y^2}$. Due to its definition, the momentum can be written as (Munson et al., 2009b)

$$|I| = m_c v_c(t) = \rho_c(t)V(t)v_c(t), \tag{6.14}$$

where we can express the cloud velocity ($v_c(t)$) as the displacement on s-axis ($s(t)$) in an infinitesimal time interval, i.e., $v_c(t) = ds(t)/dt$. The cloud volume is defined as

$V(t) = \eta r(t)^3$ where $\eta = 4\pi/3$ for a spherical cloud. Furthermore, the radius of the cloud ($r(t)$) is linearly related with the distance such that $r(t) = \alpha_e s(t)$ where α_e is the entrainment coefficient and is empirically determined (Morton et al., 1956). Hence, (6.14) can be rewritten as

$$|I| = \rho_c(t)\eta\alpha_e^3 s(t)^3 \frac{ds(t)}{dt}. \tag{6.15}$$

When $|I| = \sqrt{I_x^2 + I_y^2}$ is incorporated into (6.15), we have

$$\frac{ds(t)}{dt} = \frac{\sqrt{I_x^2 + I_y^2}}{\rho_c(t)\eta\alpha_e^3 s(t)^3}. \tag{6.16}$$

Here, (6.7) and (6.12)-(6.13) are substituted into (6.16) as given by

$$\frac{ds(t)}{dt} = \frac{\sqrt{F_0^2 t^2 + 2F_0 I_0 sin(\theta_0)t + I_0^2}}{\left(\sum_{k=1}^{K}(\rho_d - \rho_f)\phi_k(t) + \rho_a\right)\eta\alpha_e^3 s(t)^3}. \tag{6.17}$$

Remembering that $V(t) = \eta\alpha_e^3 s(t)^3$ and $\phi_k(t) = \frac{N_k(t)V_k}{V(t)}$, the denominator part of (6.17) is simplified as

$$\frac{ds}{dt} = \frac{\sqrt{F_0^2 t^2 + 2F_0 I_0 sin(\theta_0)t + I_0^2}}{Z + \rho_a\eta\alpha_e^3 s^3}, \tag{6.18}$$

where $Z = \sum_{k=1}^{K}(\rho_d - \rho_f)V_k N_k(t)$.

For convenience, θ_0 is chosen as 0 in (Bourouiba et al., 2014). However, $\theta_0 \neq 0$ is also considered in order to observe the effect of the initial cough/sneeze angle in our study. In addition to this, the initial conditions which are $s(0) = 0$ and $t(0) = 0$ are taken into account in the integration of (6.18) to obtain the quartic equation as given by

$$\frac{\eta\alpha^3\rho_a}{4}s(t)^4 + Zs(t) - \left(\frac{F_0 t + I_0 sin(\theta_0)}{2F_0}\right)\sqrt{F_0^2 t^2 + 2F_0 I_0 sin(\theta_0)t + I_0^2}$$
$$-\left(\frac{I_0^2(sin(\theta_0)^2 - 1)}{2F_0}\right)\ln\left(2F_0\sqrt{F_0^2 t^2 + 2F_0 I_0 sin(\theta_0)t + I_0^2} + 2F_0^2 t + 2F_0 I_0 sin(\theta_0)\right) = 0. \tag{6.19}$$

Since the discriminant of the quartic equation (6.19) is less than zero for physically meaningful parameter values, two of the roots are complex and one of the roots is a real and negative number. Therefore, there is only one possible positive real root which is used as the solution. However, this solution is a very long expression to write in this chapter and (6.19) is solved numerically as detailed later in this section.

In order to determine the trajectory of the cloud, θ needs to be derived. Via the substitution of $I_x = |I|\cos(\theta)$ and $I_y = |I|\sin(\theta)$ into (6.12) and (6.13), two expressions

are obtained. Then, by solving these two expressions for $|I|$ and equating them gives the equation below

$$|I| = \frac{I_0 \cos(\theta_0)}{\cos(\theta)} = \frac{F_0 t + I_0 \sin(\theta_0)}{\sin(\theta)}. \tag{6.20}$$

which can be solved for θ as given by

$$\theta = \tan^{-1}\left(\frac{F_0 t}{I_0 \cos(\theta_0)} + \tan(\theta_0)\right). \tag{6.21}$$

6.2.2. Number of Propagating Droplets in the Cloud

After the emission of droplets, some of the droplets settle to the ground due to gravity and their interaction with the air (Gulec and Atakan, 2020a; De Cock et al., 2017). Therefore, the number of droplets decreases during the propagation due to the settling of droplets. Moreover, it is assumed that the droplets are homogeneously distributed within the cloud and the settling droplets move out of the cloud instantaneously. In addition, since the movement of each droplet in the cloud is assumed to be independent of each other at each time instance, the number of droplets in the cloud can be modeled as a Poisson process with an intensity function $\lambda(t)$ (Shaw, 2003). Here, $\lambda(t)$ which is the mean number of droplets in the cloud can be derived by using the flow rate of the droplets (Bourouiba et al., 2014). Moreover, it is assumed that the droplets are homogeneously distributed within the cloud. The flow rate of the droplets (J), which is the derivative of the number of droplets and gives the number of droplets flowing through a surface in unit time (number of droplets/s), is defined as (Munson et al., 2009b)

$$J = \frac{d\lambda(t)}{dt} = v_c A \bar{\rho}_c(t) \tag{6.22}$$

where $\bar{\rho}_c(t) = \lambda(t)/V(t)$ and A is the cross-sectional area that droplets are flowing through. Since droplets settle through the lower half of the cloud due to the observations in (Bourouiba et al., 2014), A in (6.22) is substituted with $A(t)/2$ where $A(t)$ is the surface area of the cloud. Hence, (6.22) becomes

$$\frac{d\lambda(t)}{dt} = -v_s \frac{A(t)}{2} \frac{\lambda(t)}{V(t)}, \tag{6.23}$$

where v_s is the settling velocity of droplets and $(-)$ sign represents the decrease in the number of droplets due to the settling. For a spherical cloud, when the substitutions

$A(t) = 4\pi r(t)^2$, $V(t) = 4\pi r(t)^3/3$ and $r(t) = \alpha_e s(t)$ are made into (6.23), the rate of change of the mean number of droplets in the cloud is given by

$$\frac{d\lambda(t)}{dt} = \frac{-3v_s\lambda(t)}{2r(t)} = \frac{-3v_s\lambda(t)}{2\alpha_e s(t)}. \tag{6.24}$$

Since the solution of (6.19) is very long and makes the solution of (6.24) very complicated, the trajectory and number of droplets can be found numerically for each time instance. Furthermore, the number of droplets can be different for each droplet size. Thus, (6.24) is manipulated to derive the change of the mean number of droplets at each time instance ($\Delta\lambda$) as given by

$$\Delta\lambda = \frac{-3v_{s_{k,i}}\lambda_{k,i}\Delta t}{2\alpha s_{k,i}}, \tag{6.25}$$

where Δt is the time step, the subscripts i and k show the corresponding variables at $t = t_i$ for the droplets of diameter d_k. At each time step, t_i is increased by Δt and the mean number of droplets of diameter d_k in the cloud is increased via $\lambda_{k,i+1} = \lambda_{k,i} + \Delta\lambda$. For each time step and droplet size, the number of droplets in the cloud ($N_{k,i}$) follows a Poisson distribution with the rate $\lambda_{k,i}$ (Shaw, 2003). Due to the large number of emitted droplets, this Poisson distribution can be approximated as a Gaussian distribution with a mean and variance $\lambda_{k,i}$ (Papoulis and Pillai, 2002). Hence, $N_{k,i}$ is represented as $N_{k,i} \sim \mathcal{N}(\lambda_{k,i}, \lambda_{k,i})$.

Settling velocities of droplets during the propagation are defined according to the flow regimes which are Newton's (turbulent) flow, intermediate flow and Stokes (laminar) flow regimes (Reuter et al., 2005). These regimes are determined according to Reynolds number (Re) which is a dimensionless coefficient showing the flow type of the fluid as defined by (Munson et al., 2009b)

$$Re_{k,i} = \frac{d_{k,i}\rho_a v_{c_{k,i}}}{\mu_a}. \tag{6.26}$$

(6.26) shows that $Re_{k,i}$ depends on the changing cloud velocity and droplet diameter at each time step. The settling velocities according to the aforementioned flow regimes are derived as follows.

For a settling droplet, the downwards net force for a spherical droplet at the i^{th} time step with the diameter d_k is given as

$$F_{down_i} = G_i - B_i = V_d\rho_d g - V_d\rho_a g = \frac{\pi d_{k,i}^3}{6}(\rho_d - \rho_a)g \tag{6.27}$$

Furthermore, an upward drag force acts in the opposite direction of gravity due to the interaction of the droplet with the air. This upward drag force is given as (Munson et al., 2009b)

$$F_{up_i} = \frac{1}{2}\rho_d v_s^2 \pi \frac{d_{k,i}^2}{4} C_D \tag{6.28}$$

where C_D is the drag coefficient. For the settling condition, these upward and downward forces are in equilibrium. Hence, we can obtain the settling velocity by equating these two forces and pulling out v_s as given by

$$v_{s_{k,i}} = \sqrt{\frac{4 d_{k,i} g(\rho_d - \rho_a)}{3\rho_d C_D}}. \tag{6.29}$$

Here, C_D changes according to Re as given by (Reuter et al., 2005)

$$C_D = \begin{cases} \dfrac{24}{Re} & , Re < 2 \text{ (Stokes flow)} \quad (6.30) \\ \dfrac{18.5}{Re^{3/5}} & , 2 \leq Re \leq 500 \text{ (Intermediate flow)} \quad (6.31) \\ 0.44 & , 500 < Re \leq 2 \times 10^5 \text{ (Newton's flow)} \quad (6.32) \end{cases}$$

When the drag coefficients in (6.30)-(6.32) and Re in (6.26) are substituted into (6.29), the settling velocities can be obtained as given by (Reuter et al., 2005)

$$v_{s_{k,i}} = \begin{cases} \dfrac{g d_{k,i}^2 (\rho_d - \rho_a)}{18\mu_a} & , Re < 2 \text{ (Stokes flow)} \quad (6.33) \\ \dfrac{g d_{k,i}^{8/5} (\rho_d - \rho_a)}{13.875 \rho_d^{2/5} \mu_a^{3/5}} & , 2 \leq Re \leq 500 \text{ (Intermediate flow)} \quad (6.34) \\ \dfrac{3.03 g d_{k,i} (\rho_d - \rho_a)}{\rho_d} & , 500 < Re \leq 2 \times 10^5 \text{ (Newton's flow).} \quad (6.35) \\ & \quad (6.36) \end{cases}$$

The effective factors for the cloud are gravity, buoyancy, and different flow regimes generated due to the initial velocity and different droplet sizes as given above. In Newton's and intermediate flow regime, the trajectory of the cloud is dominated by their high horizontal velocity with respect to gravity and buoyancy. However, in Stokes flow, the buoyancy and gravity is much more effective, since droplets lose most of their initial momentum in the horizontal axis. The effect of the turbulent flows is only considered to calculate the settling velocities of the droplets, not for the velocity distribution of them.

6.2.3. Receiver Model

Airborne pathogen transmission via droplets is infectious, since emitted pathogen-laden droplets can be sensed by nose, mouth and eyes (Ai and Melikov, 2018; Peiris et al., 2003). Hence, the human face is where the sensing of the infectious pathogens mostly occurs. Even if the pathogens are not directly received via the facial sensory organs, it is possible to become infected by directly touching the face and sensory organs consecutively. With this motivation, the cross-sectional area of the central part of the human face is considered as the RX cross-section as shown in Figure 6.3. Moreover, a receiver model is proposed for the reception of droplets as shown in Figure 6.4.

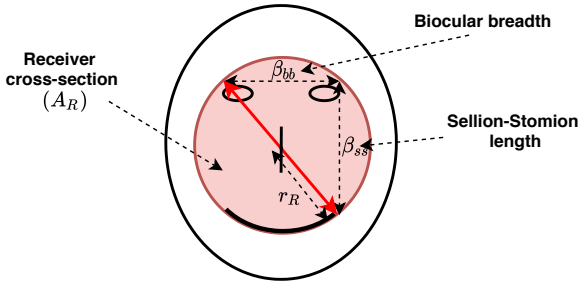


Figure 6.3. Receiver cross-section in the human face.

As the first step of the reception, the droplets in the vicinity of the RX is sensed by the human, which is defined as the signal reconstruction. Different signal reconstruction models for a sensor in macroscale and a nanomachine in microscale are proposed in (Gulec and Atakan, 2021a) and (Atakan and Gulec, 2019), respectively. As given in Figure 6.3, the RX is assumed to be the cross-section of the human face. In order to determine a circular cross-section area (A_R) for the RX by encompassing the eyes, mouth

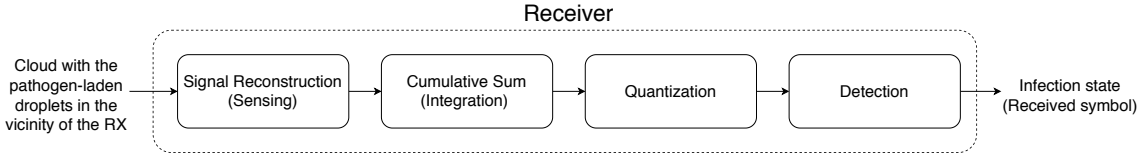


Figure 6.4. Block diagram of the receiver model.

and nose, a right-angled triangle whose sides are biocular (biectocanthus) breadth (β_{bb}), Sellion-Stomion length (β_{ss}) and the diameter of the receiver cross-section ($2r_R$) as the hypotenuse side is formed as depicted in Figure 6.3. Here, β_{bb} is the length of the line connecting the outer end points of the left and right eyes (eyelid junctions) and β_{ss} is the vertical distance between the eye and mouth (Young, 1993). Hence, the radius of the cross-sectional area of the RX (r_R) is given as $r_R = (\sqrt{\beta_{bb}^2 + \beta_{ss}^2})/2$.

When the cloud is transmitted via sneezing/coughing, there are three cases for the interaction of the RX and the cloud of droplets with diameter d_k whose centers are at the positions for the i^{th} time step (x_R,y_R,z_R) and ($x_{k,i}$,$y_{k,i}$,$z_{k,i}$), respectively. The relation between the 3-D Cartesian coordinates and curvilinear s-axis is detailed in the next subsection. In the first case, the cloud and RX do not coincide and there is no reception. The other two cases include the reception of droplets. As illustrated in Figure 6.5, when the cloud and RX coincide, the reception of droplets is related with the cross-sectional area of the cloud at $x_{k,i} = x_R$ ($A_{CS_{k,i}}$), A_R and their intersection area ($A_{RC_{k,i}}$). The second case occurs when the intersection area is less than or equal to the cross-sectional area of the RX, i.e., $A_{RC_{k,i}} < A_R$. In the last case, the cloud encompasses the RX, i.e., $A_{CS_{k,i}} \geq A_R = A_{RC_{k,i}}$.

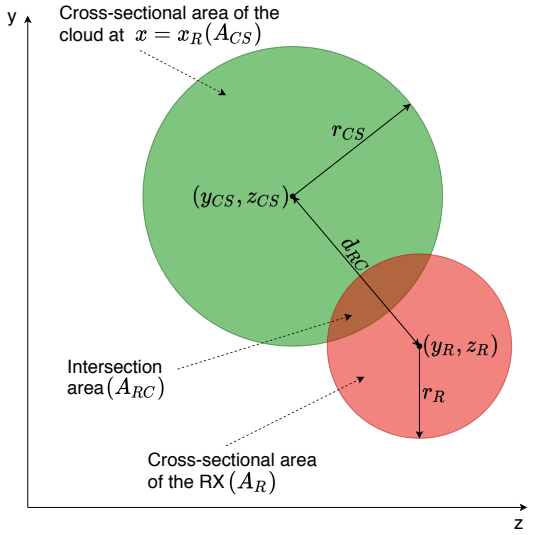


Figure 6.5. Intersection of the RX and the cloud cross-section.

For the case shown in Figure 6.5, $A_{RC_{k,i}}$ can be derived by calculating the inter-

section area of two circles as given by (6.37) (Weisstein, 2003)

$$A_{RC_{k,i}} = r_R^2 cos^{-1}\left(\frac{d_{RC_{k,i}}^2 + r_R^2 - r_{k,i}^2}{2d_{RC_{k,i}} r_R}\right) + r_{k,i}^2 cos^{-1}\left(\frac{d_{RC_{k,i}}^2 + r_{k,i}^2 - r_R^2}{2d_{RC_{k,i}} r_{k,i}}\right)$$
$$-\frac{1}{2}\sqrt{(-d_{RC_{k,i}} + r_R + r_{k,i})(d_{RC_{k,i}} + r_R - r_{k,i})(d_{RC_{k,i}} - r_R + r_{k,i})(d_{RC_{k,i}} + r_R - r_{k,i})}. \tag{6.37}$$

where the distance between the centers of $A_{RC_{k,i}}$ and A_R ($d_{RC_{k,i}}$) is defined as given below.

$$d_{RC_{k,i}} = \sqrt{(y_R - y_{k,i})^2 + (z_R - z_{k,i})^2}. \tag{6.38}$$

The received number of droplets can be derived by multiplying the time step (Δt) with the flow rate of the droplets (J in number of droplets/s) at each time step. Since the RX senses the droplets proportional to $A_{RC_{k,i}}$, the received signal after the signal reconstruction step for the aforementioned cases at each time step by recalling (6.22) is expressed as

$$\tilde{N}_{R_i} = \begin{cases} \sum_{k=1}^{K} v_{c_{k,i}} A_{RC_{k,i}} \frac{N_{k,i}}{\eta r_{k,i}^3} \Delta t \ , A_{RC_{k,i}} < A_R & (6.39) \\ \sum_{k=1}^{K} v_{c_{k,i}} A_R \frac{N_{k,i}}{\eta r_{k,i}^3} \Delta t \quad , A_{RC_{k,i}} = A_R & (6.40) \\ 0 \quad , \text{otherwise}, & (6.41) \end{cases}$$

where $A_{RC_{k,i}} = \pi r_{RC_{k,i}}^2$, $A_R = \pi r_R^2$, $\eta = 4\pi/3$ and the volume is $\eta r_{k,i}^3$ for the spherical cloud.

As the time elapses, the cumulative exposure to the pathogen-laden droplets at the RX can be modeled by the cumulative sum (integration) of droplets with respect to time subsequent to the signal reconstruction step as given in Figure 6.4. Afterwards, the signal is quantized, since the rate of the change in the mean number of droplets may not be an integer. The number of droplets after the cumulative sum is rounded to the nearest integer in the quantization step. The received signal after the quantization step is given by

$$N_{R_i} = \begin{cases} \sum_{k=1}^{K} \bar{N}_{R_{k,i}} \ , A_{RC_{k,i}} < A_R & (6.42) \\ \sum_{k=1}^{K} \bar{N}_{R_{k,i}} \ , A_{RC_{k,i}} = A_R & (6.43) \\ 0 \quad , \text{otherwise}. & (6.44) \end{cases}$$

Here, the mean received number of droplets after the quantization step for the k^{th} droplet diameter at the i^{th} time step ($\bar{N}_{R_{k,i}}$) is defined as

$$\bar{N}_{R_{k,i}} = \begin{cases} \left\lfloor \dfrac{v_{c_{k,i}} A_{RC_{k,i}} \Delta t}{\eta r_{k,i}^3} \displaystyle\sum_{m=0}^{i} N_{k,i-m} + \dfrac{1}{2} \right\rfloor & , A_{RC_{k,i}} < A_R \quad (6.45) \\ \left\lfloor \dfrac{v_{c_{k,i}} A_R \Delta t}{\eta r_{k,i}^3} \displaystyle\sum_{m=0}^{i} N_{k,i-m} + \dfrac{1}{2} \right\rfloor & , A_{RC_{k,i}} = A_R \quad (6.46) \end{cases}$$

where $\lfloor . \rfloor$ shows the floor function which maps a variable to the integer less than or equal to this variable. The addition with $\frac{1}{2}$ within the floor function in (6.45)-(6.46) provides the quantization by rounding the number of received droplets to the nearest integer.

Subsequent to quantization, the infection state of the human needs to be determined as the output of the system as shown by Figures 6.1 and 6.4. Therefore, the detection is essential according to a threshold value (γ) as the last step of the reception. Physically, γ corresponds to the quantity of pathogen-laden droplets that suffice to make a human infected. Furthermore, γ depends on the immune system of a human. Thus, detection via the threshold γ enables to quantify the strength of the human immune system and to handle the determination of the infection state as a detection problem. Hence, the infection state (or the received symbol) can be expressed by binary hypothesis testing which is given as

$$N_{R_i} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \quad (6.47)$$

where the hypotheses H_0 and H_1 are defined as the situations of no infection as the received symbol 0 and infection as the received symbol 1, respectively. Here, the received symbol sequence with M samples is represented as $\mathbf{h} = [h_0, h_1, ..., h_M]$ which also gives the end-to-end system response.

6.2.4. Algorithm of the End-to-End System Model

In this part, the way to obtain the output of the end-to-end system model by employing the derivations made up to here is clarified in the proposed Algorithm 2. Before the procedure starts, the initial parameters such as the number of droplet sizes (K), total simulation time (t_s), initial number of droplets for each droplet size ($N_{(1,..K),0}$), initial momentum (I_0), initial net buoyant force (F_0), initial velocity of the cloud (v_{c_0}), RX dimensions (β_{bb}, β_{ss}) and initial positions of the TX and RX are taken as the input.

In the algorithm, the propagation, changing number of droplets in the cloud and their interaction with the RX are handled separately at each time step for each droplet size d_k. As the first step of the algorithm, the trajectory of the spherical cloud is calculated for each time step, i.e., the distance on the s-axis and θ values for $t = t_i$ are calculated by (6.19) and (6.21). Then, the updated position of the cloud on the s-axis is utilized to find its step length (Δs). As shown in Figure 6.2, Δs can be employed to express the step lengths on x (Δx) and y axes (Δy) as given by

$$\Delta x = \Delta s \cos(\theta_{k,i}), \tag{6.48}$$

$$\Delta y = \Delta s \sin(\theta_{k,i}), \tag{6.49}$$

where Δx and Δy values are used to update the cloud position on the corresponding axis. Since there is not any acting force on z-axis, the center of the cloud maintains its position on this axis. However, the cloud expands on x, y and z axes linearly due to the relation $r_{k,i} = \alpha_e s_{k,i}$ as given in Section 6.2.1.

In the second step of Algorithm 2, the mean number of droplets in the propagating cloud is calculated. To this end, the cloud velocity ($v_{c_{k,i}}$) at the corresponding time step is calculated by the displacement on s-axis. Then, settling velocity ($v_{s_{k,i}}$) is determined by (6.33)-(6.35) according to $Re_{k,i}$ which is calculated by using $v_{c_{k,i}}$ and droplet size. $v_{s_{k,i}}$ is exploited to calculate the change in the mean number of droplets ($\Delta\lambda$) and thus, the mean number of droplets is updated according to this change. Here, the flow type of the cloud found by employing the velocity of droplets affects the number of droplets in the cloud.

The third step of Algorithm 2 describes the reception via the interaction of the cloud with the RX. When the cloud comes to a sufficient distance to interact with the RX, the radius of the cloud's circular cross-section ($r_{CS_{k,i}}$) is determined by the geometrical relation with the the radius of the cloud and the positions of the TX and RX on the x-axis as given by

$$r_{CS_{k,i}} = \sqrt{r_{k,i}^2 - (x_R - x_{k,i})^2}, \tag{6.50}$$

which allows us to calculate the circular area of the cloud's cross-section ($A_{CS_{k,i}}$). During the reception, the case that $A_{CS_{k,i}} \leq A_R$ can also be represented in terms of radii of the circles and the distance between them such that $(r_{CS_{k,i}} - r_R) < d_{RC_{k,i}} < (r_{CS_{k,i}} + r_R)$. In addition, the case for the cloud encompassing the RX ($A_{CS_{k,i}} > A_R = A_{RC_{k,i}}$) can be expressed as $d_{RC_{k,i}} < (r_{CS_{k,i}} - r_R)$. Using these conditions, the mean number of droplets

given in (6.42)-(6.44) is calculated. Then, the detection is made according to the threshold γ in order to determine the infection state. Next, the probability of infection is derived by using the end-to-end system model.

Algorithm 2 Algorithm of the End-to-End System Model

$t = 0 : \Delta t : t_s$
for $k = 1 : 1 : K$ **do**
  for $i = 1 : 1 : \text{length}(t)$ **do**
▷ **Step 1: Trajectory**
    Calculate $s_{k,i}$ by the real positive root of (6.19)
    $r_{k,i} = \alpha_e s_{k,i}$
    Calculate $\theta_{k,i}$ by (6.21)
    $\Delta s = s_{k,i} - s_{k,i-1}$
    $\Delta x = \Delta s \cos(\theta_{k,i})$; $\Delta y = \Delta s \sin(\theta_{k,i})$
    $x_{k,i} = x_{k,i-1} + \Delta x$; $y_{k,i} = y_{k,i-1} + \Delta y$
▷ **Step 2: Number of Droplets in the Cloud**
    $v_{c_{k,i}} = \Delta s / \Delta t$
    $Re_{k,i} = d_{k,i} v_{c_{k,i}} \rho_a / \mu_a$
    Calculate $v_{s_{k,i}}$ according to $Re_{k,i}$ by (6.33)-(6.35)
    Calculate $\Delta\lambda$ by (6.25)
    $\lambda_{k,i} = \lambda_{k,i-1} + \Delta\lambda$
    Generate $N_{k,i} \sim \mathcal{N}(\lambda_{k,i}, \lambda_{k,i})$
▷ **Step 3: Reception**
    if $(x_R - r_{k,i}) < x_{k,i} < (x_R + r_{k,i}))$ **then**
      $r_{CS_{k,i}} = \sqrt{r_{k,i}^2 - (x_R - x_{k,i})^2}$;
      $A_{CS_{k,i}} = \pi r_{CS_{k,i}}^2$
      Calculate $d_{RC_{k,i}}$ by (6.38)
      if $(r_{CS_{k,i}} - r_R) < d_{RC_{k,i}} < (r_{CS_{k,i}} + r_R)$ **then**
        Calculate $\bar{N}_{R_{k,i}}$ by (6.45)
      else if $d_{RC_{k,i}} < (r_{CS_{k,i}} - r_R)$ **then**
        Calculate $\bar{N}_{R_{k,i}}$ by (6.46)
      else
        $\bar{N}_{R_{k,i}} = 0$
      end if
      $N_{k,i} = N_{k,i} - \bar{N}_{R_{k,i}}$
    end if
  end for
end for
Calculate N_{R_i} by (6.42)-(6.44)
Determine the infection state ($\mathbf{h}$) by applying (6.47)

6.3. Probability of Infection

The probabilistic approach which is considered in the system model enables the derivation of the probability of infection of a human exposed to a sneeze or cough. To this end, it is essential to derive the probability density function (pdf) of the received number of droplets before the detection. As given in Section 6.2.2, the number of droplets in the cloud is a Gaussian random variable for the droplet diameter of d_k and i^{th} time step, i.e., $N_{k,i} \sim \mathcal{N}(\lambda_{k,i}, \lambda_{k,i})$. Hence, its pdf ($f_N(N_{k,i})$) is given by

$$f_N(N_{k,i}) = \frac{1}{\sqrt{2\pi\lambda_{k,i}}} e^{-\frac{(N_{k,i}-\lambda_{k,i})^2}{2\lambda_{k,i}}}. \tag{6.51}$$

Since the received number of droplets before the detection (N_{R_i}) is a function of $N_{k,i}$ as given in (6.42)-(6.46), its mean and variance for the reception cases can be given as

$$E(N_{R_i}) = \begin{cases} \mu_1 = \sum_{k=1}^{K} \left\lfloor \sum_{m=0}^{i} a_{k,i}\lambda_{k,i-m} + \frac{1}{2} \right\rfloor, & A_{RC_{k,i}} < A_R \quad (6.52) \\ \mu_2 = \sum_{k=1}^{K} \left\lfloor \sum_{m=0}^{i} b_{k,i}\lambda_{k,i-m} + \frac{1}{2} \right\rfloor, & A_{RC_{k,i}} = A_R \quad (6.53) \end{cases}$$

$$Var(N_{R_i}) = \begin{cases} \sigma_1^2 = \sum_{k=1}^{K} \left\lfloor \sum_{m=0}^{i} a_{k,i}^2\lambda_{k,i-m} + \frac{1}{2} \right\rfloor, & A_{RC_{k,i}} < A_R \quad (6.54) \\ \sigma_2^2 = \sum_{k=1}^{K} \left\lfloor \sum_{m=0}^{i} b_{k,i}^2\lambda_{k,i-m} + \frac{1}{2} \right\rfloor, & A_{RC_{k,i}} = A_R, \quad (6.55) \end{cases}$$

where $a_{k,i} = \frac{v_{c_{k,i}} A_{RC_{k,i}} \Delta t}{\eta r_{k,i}^3}$, $b_{k,i} = \frac{v_{c_{k,i}} A_R \Delta t}{\eta r_{k,i}^3}$, $E(.)$ and $Var(.)$ are expectation and variance operators, respectively. By using (6.52)-(6.55), the pdf of N_{R_i} ($f_{N_R}(N_{R_i})$) can be written as

$$f_{N_R}(N_{R_i}) = \begin{cases} \frac{\exp\left(-\frac{(N_{R_i}-\mu_1)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi\sigma_1^2}}, & A_{RC_{k,i}} < A_R \quad (6.56) \\ \frac{\exp\left(-\frac{(N_{R_i}-\mu_2)^2}{2\sigma_2^2}\right)}{\sqrt{2\pi\sigma_2^2}}, & A_{RC_{k,i}} = A_R \quad (6.57) \\ 0, & \text{otherwise}, \quad (6.58) \end{cases}$$

where μ_1, μ_2, σ_1^2 and σ_2^2 are defined in (6.52)-(6.55). The probability of infection corresponds to the situation where $N_{R_i} > \gamma$ as given by

$$P(N_{R_i} > \gamma) = \int_{\gamma}^{\infty} f_{N_R}(u)du. \tag{6.59}$$

The solution for (6.59) can be derived for the pdf given in (6.56)-(6.58) in terms of Q-function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{\frac{-u^2}{2}} du$ as given by

$$P(N_{R_i} > \gamma) = \begin{cases} Q\left(\dfrac{\gamma - \mu_1}{\sigma_1}\right) & , A_{RC_{k,i}} < A_R \qquad (6.60) \\ Q\left(\dfrac{\gamma - \mu_2}{\sigma_2}\right) & , A_{RC_{k,i}} = A_R \qquad (6.61) \\ 0 & , \text{otherwise.} \qquad (6.62) \end{cases}$$

The derived probability of infection and the system model given in the previous section can be employed to analyze the dynamics of the pathogen transmission as given with the numerical results in the next section.

Table 6.1. Experimental parameters

Parameter	**Value**	**Parameter**	**Value**
Δt	0.1 s	t_s	10 s
TX's position	(0,1.7,0) m (x,y,z)	$v_{c_{(1,..K)},0}$	11.2 m/s (cough) (Zhu et al., 2006)
RX's position	(1.7,0) m (y,z)	g	9.81 m/s^2
I_0	0.0131 kg m/s (Bourouiba et al., 2014)	α_e	0.2116 (Tang et al., 2009)
F_0	0.0023 kg m/s^2 (Bourouiba et al., 2014)	β_{bb} (female)	8.853 cm (Young, 1993)
ρ_d	993 kg/m^3 (Nicas et al., 2005)	β_{bb} (male)	9.131 cm (Young, 1993)
ρ_f (at 34°C)	0.98 kg/m^3 (Picard et al., 2008)	β_{ss} (female)	6.901 cm (Young, 1993)
ρ_a (at 23°C)	1.172 kg/m^3 (Bourouiba et al., 2014)	β_{ss} (male)	7.57 cm (Young, 1993)
μ_a	19 × 10^{-6} kg/(m s) (Bourouiba et al., 2014)		

6.4. Numerical Results

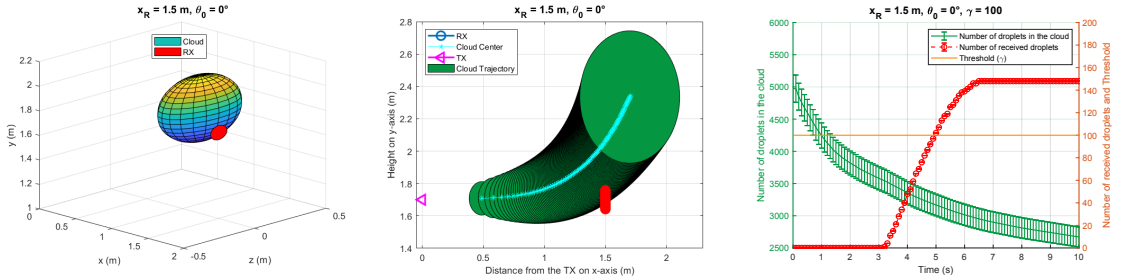
In this section, numerical results using the algorithmic end-to-end system model and derived probability of infection are given. The values of the experimental parameters are given in Table 6.1. Except the simulation parameters such as Δt, t_s and the positions of the TX and RX, this table includes measured values which are obtained by empirical studies (Bourouiba et al., 2014; Young, 1993; Zhu et al., 2006; Nicas et al., 2005; Picard et al., 2008; Tang et al., 2009). Furthermore, the initial number of droplets according to

their diameters for a cough is given in Table 6.2. For sneezing, there are not sufficient empirical data in the literature to obtain the parameter values given in Table 6.1 such as α_e and initial velocity (Ai and Melikov, 2018). Therefore, although our proposed model is applicable to a sneezing scenario, we only consider a coughing scenario for two static humans where one of them is the TX and the other is RX.

Table 6.2. Initial number of droplets (Duguid, 1946)

d (μm)	**Quantity (cough)**	d (μm)	**Quantity (cough)**	d (μm)	**Quantity (cough)**
2	50	40	240	200	35
4	290	50	110	250	29
8	970	75	140	500	34
16	1600	100	85	1000	12
24	870	125	48	2000	2
32	420	150	38	Total	4973

For the results in Figures 6.6-6.7, the dimensions of the RX are applied by using the average values of male and female humans. Namely, β_{bb} and β_{ss} are obtained by calculating the average of the female and male values given in Table 6.1.



(a) (b) (c)

Figure 6.6. The trajectory of the cloud and its interaction with the RX in (a) 3-D (b) 2-D. (c) The number of droplets.

For $x_R = 1.5$ m and $\theta_0 = 0°$, Figures 6.6 (a) and (b) depict the trajectories of the cloud with respect to RX in 3-D and 2-D, respectively. Due to initial momentum of coughing and net buoyant force, the cloud propagates in the horizontal and vertical directions, respectively. As also observed in these trajectories, the 3-D cloud can encompass (or intersect with) the RX and the RX is exposed to droplets during the passage of the cloud. Figure 6.6 (c) shows the interaction of a human who can be infected with a

disease-spreading human by quantifying the number of droplets received from the cloud. The number of droplets are given by their mean and their variations (vertical bars) due to the Gaussian distribution. These variations are calculated as three times the standard deviation (99.73% confidence interval) for each sample. As observed in Figure 6.6 (c), the number of droplets in cloud decreases, since large-sized droplets settle due to the gravity.

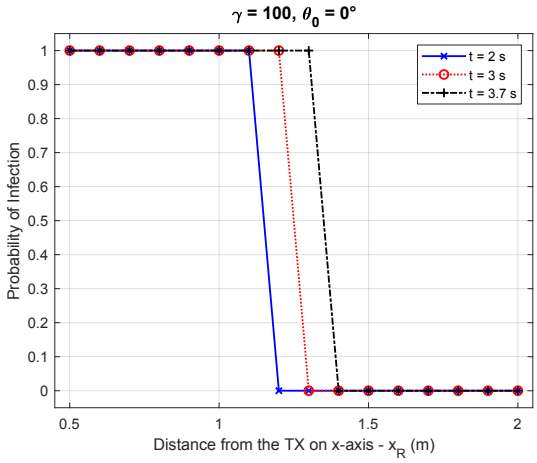
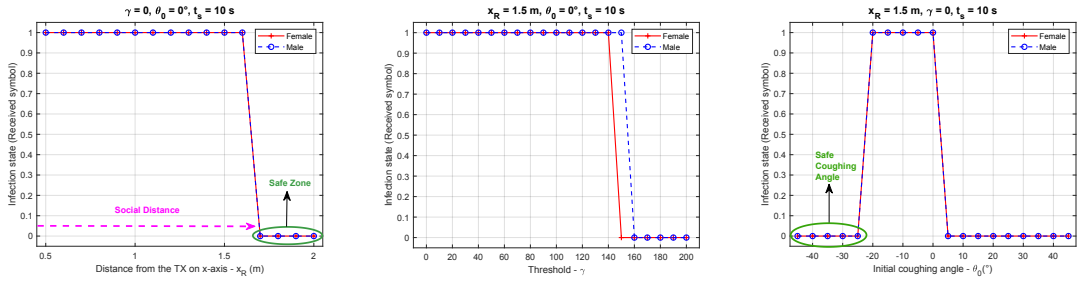


Figure 6.7. Probability of infection according to distance for different time values.

As mentioned in Section 6.2, the RX gets infected, when the received number of droplets is above γ as plotted in Figure 6.6 (c). This figure shows the importance of the exposure time which is the interval of the changing zone in the received signal. If the RX is exposed to the cloud less, it is possible not to be infected, since the received number of droplets can be below the threshold. This interaction is also clarified in Figure 6.7 by showing the relation of the infection probability with the distance of the RX to the TX for different propagation time instances. Actually, this figure reveals that if the RX is exposed to the cloud for a longer period, the RX is more likely to be infected.

Different scenarios can be analyzed by employing the infection state, which is the output of the end-to-end system model, for various x_R, γ, θ_0 values and male/female receivers. In Figure 6.8 (a), the threshold is set to zero to determine the safe zone where there is no possibility of infection. This safe zone starts at $x_R = 1.7$ m which also shows the minimum social distance. Figure 6.8 (c) shows that initial coughing angle affects the infection state severely. For θ_0 values between 0 and -25 degrees, it is more likely to infect someone due to the buoyant forces affecting the propagation of the cough cloud.

Therefore, it is safer to cough with an initial angle $\theta_0 \leq -25^\circ$ which is depicted as safe coughing angle in Figure 6.8 (c).



(a) (b) (c)

Figure 6.8. Infection state of the RX with respect to (a) x_R (b) γ (c) θ_0.

Coughing with $\theta_0 \geq 0^\circ$ may not be safe, since small droplets (aerosols) can suspend in the air and settle eventually due to gravity or drift due to the indoor air currents in the long term. In Figures 6.8 (a) and (c), the results are indistinguishable for male and female receivers. However, Figure 6.8 (b) shows that the infection states of female and male humans can be affected differently for the same γ values. Actually, this figure reveals that female humans are less likely to get infected, even if their immune system's strength are the same with male humans due to the slight difference in the face dimensions.

In this chapter, numerical results are based on our model which focuses on the propagation and reception of droplets in the outer layer of the human face. For a more rigorous model, the effect of turbulent flows and the survival rate of pathogens can be taken into account. The reception part can be improved by considering the interaction of pathogens with the human cells. In addition, γ can be estimated by considering several effects such as age, chronic diseases and genetic factors. These improvements are left as open research issues.

CHAPTER 7

MOBILE HUMAN AD HOC NETWORKS: A COMMUNICATION ENGINEERING VIEWPOINT ON INTERHUMAN AIRBORNE PATHOGEN TRANSMISSION

7.1. Introduction

Throughout the history, epidemics caused by infectious diseases have been a major threat to human life. Epidemic diseases such as black plague, smallpox, Spanish flu and recent coronavirus disease 2019 (COVID-19) gave rise to millions of human deaths. In addition, epidemics can induce mental disorders in humans and recessions in the world economy due to prevention and control measures such as lockdown. Owing to these facts, it is essential to understand and accurately model the spread of infectious diseases among humans.

The interhuman spread of infectious diseases occur via direct contact and airborne transmission[1] where pathogens are transferred from an infectious human to a susceptible one. In airborne transmission, these pathogens (viruses, bacteria, fungi, and so on) are carried by large droplets and aerosols (droplet nuclei) which are emitted via breathing, speaking, coughing and sneezing (Ai and Melikov, 2018). Throughout this chapter, we use the term *droplet* to refer to both large droplets and aerosols together.

As for the airborne pathogen transmission, it is not fully unraveled how its mechanisms operate between two humans, for example, it is still a matter of debate whether large droplets or aerosols are more infectious. In addition, the mobility and interplay of people during their daily life makes the problem of modeling infectious disease spread in an epidemic more chaotic. As people displace, there exist dynamic human groups exchanging pathogens among each other. Due to their mobility, humans form different

[1]Here, *transmission* is employed synonymously with *contagion* rather than its usage in communication engineering.

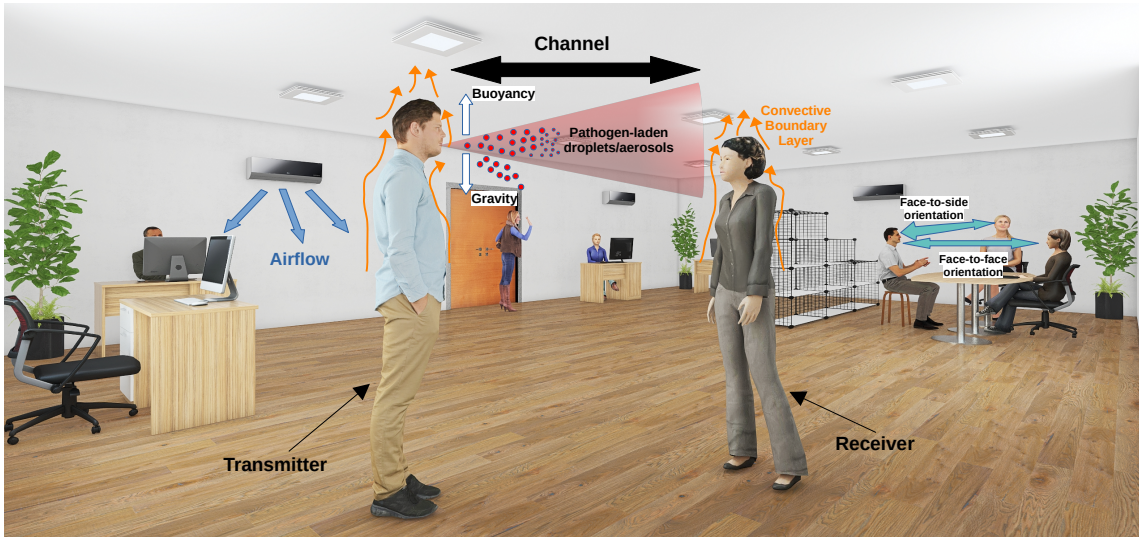


Figure 7.1. The spread of an infectious disease through airborne pathogen transmission with communication engineering perspective and effective issues for an indoor sneezing/coughing scenario.

groups in an ad hoc fashion as their smart phones do in a mobile telecommunication network. Actually, a human emitting expiratory droplets is an information source (Khalid et al., 2019). When these emitted information carrying droplets are received by another human through sensory organs, we can consider there exists a communication path between them. Hence, an analogy between human groups and mobile telecommunication networks can be established, since they both possess an intermittent connectivity which is detailed later. By utilizing this analogy, we propose an approach to modeling interhuman airborne pathogen transmission with communication engineering perspective where mobile humans forming a group are considered as a mobile human ad hoc network (MoHANET). In a MoHANET, the infectious human is the transmitter (TX), the susceptible human is the receiver (RX) and pathogen-laden droplets are information carriers propagating in the communication channel, that is, air. Here, molecular communication (MC) employing chemical signals instead of electrical signals emerges as an enabler paradigm for the communication among humans due to its biocompatibility with the human body and multiscale applicability.

On the other hand, researchers from many disciplines work separately in different scales to reveal the mechanisms of airborne pathogen transmission and model the behavior of epidemics. In fluid dynamics literature, researchers focus on the propagation

of pathogen-laden droplets and their interactions with air (Ai and Melikov, 2018). Biologists deal with the survival of airborne pathogens in macroscale (Schaffer et al., 1976) and their interactions with the human cells in microscale (Cohen, 2016). Furthermore, the medical literature conducts researches in cellular level to discover new drugs which cure the infectious diseases. In a larger scale, epidemiology literature focuses on the epidemic data to develop mathematical models for the spread of epidemics in time and space (Rock et al., 2014). However, the epidemiological models for the spread of infectious diseases do not consider the information from fluid dynamics, biology and medicine. These models generally make estimations by fitting statistical data and use unrealistic assumptions such that the population is homogeneous and becomes infectious at a constant rate. The fluid dynamics of droplets, the geometries and air distribution of indoor environments, the pathogen-human interaction, the medical efficacy of the drugs and locations of mobile humans are essential to be taken into account for accurate models. Thus, there is a need to merge all of these research efforts in a unified framework. Communication engineering approach can provide this framework by combining micro- and macroscale modeling issues. With this approach, a MoHANET is partitioned into layers where each layer is associated with a research area in different scales such as fluid dynamics, biology, medicine or epidemiology. As in the conventional networks, each layer sends its outputs to an upper layer. In this way, the spread of infectious diseases can be modeled more accurately by considering all parameters from various disciplines. In addition, researchers will be able to utilize theoretical tools of communication theory in order to model the complicated nature of airborne pathogen transmission. As a proof-of-concept study, we show that the number of infected people during an epidemic can be estimated by taking into account the propagation and reception of pathogen-laden droplets and the mobility of humans.

In the remainder of this chapter, we first review the airborne pathogen transmission mechanisms and the motivation to use MC as the enabler communication paradigm. Then, the communication engineering approach which merges different disciplines. In this approach, the layered architecture of MoHANET is presented in detail and open research issues are discussed. Finally, we give the existing and possible experimental techniques. The study given in this chapter was first introduced in (Gulec and Atakan, 2020c).

7.2. Airborne Pathogen Transmission Mechanisms and Molecular Communication

This section provides a brief overview for the main issues of the airborne pathogen transmission mechanisms. Then, the roles of molecular signals in the transfer of pathogens among humans are discussed.

7.2.1. Overview of Main Issues on Airborne Pathogen Transmission

In this part, main issues related to fluid dynamics and biology on airborne pathogen transmission are given.

7.2.1.1. Respiratory Activity, Droplet Size and Evaporation

Pathogen-laden droplets are emitted to the air from an infected human via respiratory activities such as coughing, sneezing, speaking and breathing. These activities have different initial droplet velocities allowing different propagation distances. For instance, the initial velocities for coughing and breathing are about 10 m/s and 2.67 m/s, respectively (Ai and Melikov, 2018). Therefore, a cough can infect people at a greater distance than breathing in still air in a short time interval. Furthermore, the expiratory droplets are defined according to their diameters where aerosols and large droplets are assumed to have smaller and larger diameters than 10 μm, respectively (Mittal et al., 2020). While speaking, sneezing, and coughing release more large droplets into the air, breathing mostly contains aerosols. In addition, larger droplets settle to the ground due to gravity before evaporation and smaller droplets can become aerosols via evaporation depending on the temperature and relative humidity (RH) (Seminara et al., 2020). For long durations, aerosols can be more infectious than large droplets, since they can remain suspended in the air and be drifted by airflows.

7.2.1.2. Air Distribution

In addition to the initial velocity, emitted droplets are influenced by the airflows, similar to a MC channel with drift. In outdoor environments, winds carry the droplets and dilute the concentration of pathogens via dispersion. Therefore, it is less probable to get infected in outdoor environments. However, in indoor environments such as hospitals, offices or residential buildings, airflows generated by ventilation systems are critical for the spread of pathogens due to the circulation of air in bounded conditions. Furthermore, personalized ventilation and exhaust systems are proposed as advanced ventilation systems to diminish the infection risk (Ai and Melikov, 2018). These air distributions are required to be considered for realistic indoor airborne transmission models.

7.2.1.3. Posture, Relative Orientation, Distance and Movement of the Human

For short distances, the posture, that is, standing, sitting or lying position, and the relative orientation of the infected and susceptible persons are important for the infection risk as shown in Figure 7.1. For instance, a doctor can reduce the exposure from an infected lying patient in a hospital ward via a standing posture and sideways orientation instead of face-to-face orientation (Ai and Melikov, 2018). Furthermore, a walking person can increase the infection risk in a closed and ventilated room by increasing the dispersion of the droplets (Halvoňová and Melikov, 2010). Another important factor that influences the infection risk is the relative distance of the humans which is also referred as the social distance. Surely, the infection risk decreases, as the relative distance between two people increases.

7.2.1.4. Thermofluid Boundary Conditions

The temperature difference between the human body surface and the surrounding air generates a thermal plume which is a buoyancy-driven upward flow of the surrounding air. As illustrated in Figure 7.1, this thermal plume leads to a convective boundary layer

(CBL) around the human body, which should be taken into account for the movement of the droplets in the breathing zone (Licina et al., 2015). This upward flow can change the channel impulse response via generating an upward drift for the pathogens during the reception into the human body.

7.2.1.5. Survival of Pathogens

Subsequent to a respiratory activity, all of the emitted pathogens may not survive. In (Schaffer et al., 1976), it is shown that more than 80 percent of the influenza viruses cannot survive within one minute. However, these survival rates are severely influenced by environmental factors such as temperature and relative humidity (RH). While increasing temperature decreases survival rates of the pathogens due to its effect at molecular levels, increasing RH results in decreasing evaporation of droplets (Marr et al., 2019). The decreasing number of pathogens results in a time-varying channel due to the dependence on the previous number of pathogens.

7.2.2. Do Humans Communicate via Molecular Signals?

Via the aforementioned respiratory activities, a human can transfer pathogen-laden droplets to another human. This type of transfer (or communication) among humans is investigated in the medical literature where pheromone-based molecular signals are studied for the interaction of humans. In (Stern and McClintock, 1998), it is proposed that pheromones secreted from the axillary apocrine glands of women living in close proximity provides a synchronization in their menstrual cycle. Hence, molecular signals may give rise to some biological responses in human organism. As given in the next section, the transfer of the pathogen-laden droplets which cause infection can be considered in the context of MC.

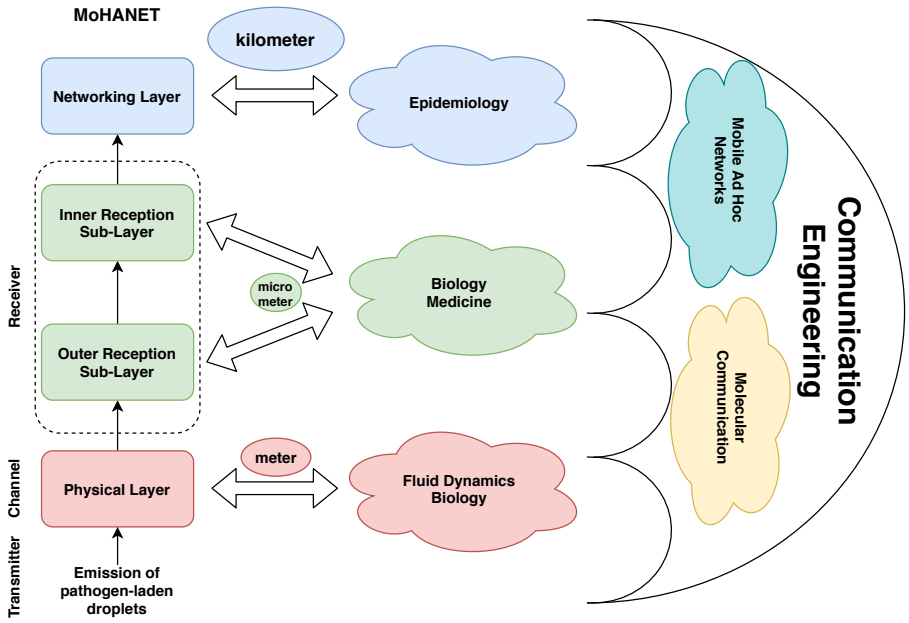


Figure 7.2. Communication engineering framework to model the spread of infectious diseases through airborne pathogen transmission and the layered MoHANET architecture.

7.3. Communication Engineering Approach to Interhuman Airborne Pathogen Transmission

In this section, we present a framework with communication engineering perspective to model the spread of infectious diseases through airborne pathogen transmission. Furthermore, open research issues are given.

As shown in Figure 7.2, the proposed framework merges all of the multiscale research efforts in various disciplines such as fluid dynamics, biology, medicine, and epidemiology under the umbrella of communication engineering. MC emerges as the key paradigm that connects the studies among different disciplines in macro- and microscales. First, the MoHANET is introduced through a layered architecture as depicted in Figure 7.2. Layers are associated with different disciplines from μm to km scale in this architecture where each layer sends its output to a upper layer. The first layer is defined as the physical layer where the infectious human (TX) emits pathogen-laden droplets through the communication channel (air) as illustrated in Figure 7.1. The next layer is the reception layer which takes place at the susceptible human (RX) and includes two sub-layers,

that is, outer and inner reception sub-layers. The outer reception sub-layer comprises the interactions of the facial sensory organs with the droplets and inner reception sub-layer provides the details about the interactions of pathogens with the biological cells in the human body. The networking layer where infectious diseases spread among different people is given at the top of the MoHANET architecture. Here, methods from mobile telecommunication networks literature are exploited and the outputs of the lower layers are employed rather. The details of this layered architecture are introduced as follows.

7.3.1. Physical Layer

In this part, the details of the transmitter and channel for still and windy air conditions are given.

7.3.1.1. Transmitter

In a MoHANET, an infected person is considered as a TX and her/his respiratory activities determine the TX parameters such as initial droplet velocities and droplet size distribution (Khalid et al., 2019). The respiratory activities which are mentioned earlier can be classified as impulsive (sneezing and coughing) and continuous (breathing and speaking) emission signals. For continuous emissions, the respiration rate is an influential factor for the transmission models. However, it is crucial to characterize speaking, since it is not always periodic and has more complex patterns than breathing. In addition, the respiratory organs such as nose or mouth affect the direction of the emitted signals. For example, the infection risk increases, when the TX uses the mouth instead of nose (Ai and Melikov, 2018). Furthermore, the convective boundary layer (CBL) of the human body, posture and relative orientation should be taken into account for accurate TX models. As mentioned earlier, the upward flow stemming from the CBL can affect the direction of the emitted pathogens in a TX model.

7.3.1.2. Channel

From the viewpoint of communication engineering, the channel is the physical medium between the TX and RX including the boundary conditions. As shown in Figure 7.2, channel modeling in the physical layer requires knowledge from fluid dynamics and biology due to the air-droplet interaction and survival of pathogens, respectively. The propagation dynamics of droplets can be examined under two subheadings depending on whether there is an external airflow or not.

Still Air: In indoor environments such as residential buildings, it is generally assumed that there is no airflow, if there is not any ventilation system. After the emission of pathogen-laden droplets with an initial velocity, they are subject to Newtonian mechanics during their interaction with the air. Emitted droplets can be modeled as a cloud consisting of droplets and air particles. The movement of this cloud can be defined as a two-phase flow where these phases represent the gaseous state of air and liquid state of droplets (Gulec and Atakan, 2021c). Due to gravity, large droplets may fall earlier to the ground with respect to aerosols and evaporation can shrink the size of the droplets. As mentioned earlier, the temperature of the air and evaporation influence the survival rates of the pathogens. For continuous emissions, this fact can affect the channel memory, which is crucial for channel modeling. Furthermore, initial velocities of droplets determined by respiratory activities can give rise to short-term laminar and turbulent flows. These flows fade out as the distance between the TX and RX increases.

Windy Air: For windy outdoor environments and indoor environments with airflows such as ventilation or wind arising from the open doors and windows, airflows dominate the propagation of droplets rather than other factors given for still air environments. The airflow which carries the pathogen-laden droplets can be examined by advection and dispersion (turbulent diffusion) mechanisms. Briefly, advection results from the airflow velocity and dispersion depends on the turbulent eddies during the mass transfer (De Visscher, 2013). It should be noted that molecular diffusion related with the thermal energy of molecules is negligible in macroscale. In order to calculate the concentration of droplets in time and space, deterministic and stochastic approaches which are based on differential Navier-Stokes and continuity equations are employed. For certain initial and boundary

conditions, the solutions of these equations for deterministic concentration are known as Gaussian Plume for steady-state and Gaussian Puff Model for transient analysis (De Visscher, 2013). Actually, the concentration and velocity of droplets are random processes whose mean values are represented by these deterministic solutions. Thus, stochastic differential equations are obtained which are non-trivial to solve as a closed form expression. Therefore, these equations are mostly solved by numerical methods using Eulerian and Langrangian approaches (De Visscher, 2013). In addition, indoor ventilation types such as under floor air distribution, mixing, displacement, and downward ventilation should be incorporated into these airflow models. For example, downward ventilation can reduce the infection risk by diluting the dispersion of droplets (Ai and Melikov, 2018). By using the models at physical and reception layers, the infection rate can be derived to be used in the networking layer as given in the next part.

7.3.2. Reception Layer

A human gets infected, when the transmitted pathogens are received into the body. As shown in Figure 7.2, the reception layer covers the issues related to biology and medicine in microscale where MC is utilized for the interactions of pathogens with the human body. The reception of these pathogens by the exposed human (RX) have not been well investigated, although there are myriads of theoretical, experimental and clinical studies for the propagation of pathogens. To this end, we propose a two-layered RX as shown in Figure 7.3 and detailed below.

7.3.2.1. Outer Reception Layer

The reception of pathogen-laden droplets occur in the eyes (Peiris et al., 2003), mouth and nose for many pathogens such as influenza virus (Ai and Melikov, 2018). Hence, we define the first step of reception as the outer layer sensing for the reception via facial sensory organs as illustrated in Figure 7.3. The whole surface of the human face is also important for the reception, since an infection may occur by touching the face contaminated with pathogens and these organs consecutively.

Pathogen-laden droplets emitted via a respiratory activity propagate as a mixture of droplets and air particles, which can be represented as a cloud (Gulec and Atakan, 2021c). This cloud is affected by the momentum due to the initial velocity of droplets, gravity and buoyancy stemming from the temperature difference of the mouth and ambient air. According to this model, as also detailed in Chapter 6, Fig. 7.4 gives the change of the number of droplets in the cloud by taking settling and reception of droplets into account for a coughing TX in still air as illustrated in Fig. 7.1. The cross-section of the RX is assumed to cover a circular area including eyes, mouth and nose at the outer layer as illustrated in Figure 7.3. At this point, an analogy with the communication systems can be established by considering the *infected state* of the RX as symbol *1* and *no infection* as symbol *0*. This reception is accomplished by a detection according to a threshold value ($\gamma = 80$) indicating the number of droplets required to become infected, as given in Figure 7.4. γ is a critical parameter in the airborne tranmission model, since it depends on the strength of human's immune system. To this end, biomedical data of humans such as body mass index, glucose level and whether or not having chronic diseases can be employed to estimate γ. Moreover, γ can be effective to determine the number of infected people in an epidemic as given in Section 7.3.3.

In addition to these issues in the outer layer, the posture, relative orientation and CBL of the RX should be taken into account for an accurate receiver model as considered for the TX. Furthermore, the reception of pathogen-laden droplets at the outer layer with different types of masks is an open issue to be investigated.

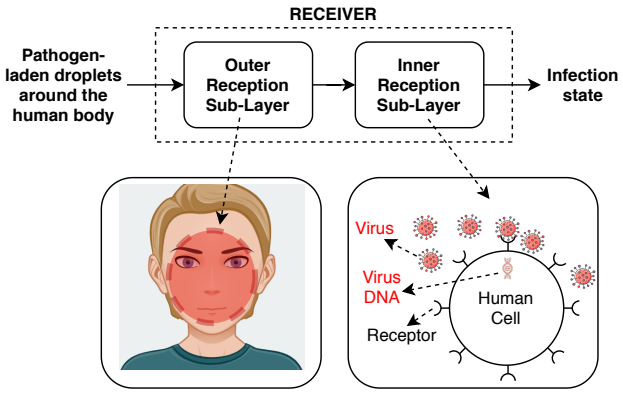


Figure 7.3. Two-layered Receiver.

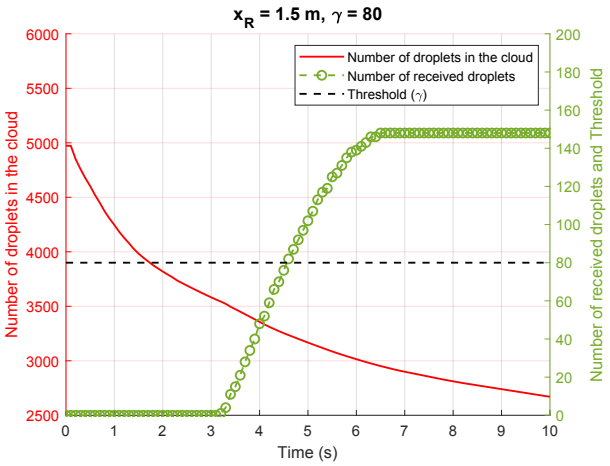


Figure 7.4. The mean number of droplets in the cloud and their reception by the RX.

7.3.2.2. Inner Reception Layer

As shown in Figure 7.3, pathogens actually *enter* human body at the cellular level and increase their population. For example, viruses replicate themselves by inserting their genetic material (DNA or RNA) into human cells in two ways: They can bind their fusion (or spike) protein on specific receptor sites on the human cell or they can enter by using endosomes like a Trojan horse (Cohen, 2016). Their binding sites can have different concentrations in different parts of the body. For instance, severe acute respiratory syndrome coronavirus-2, which causes COVID-19, binds to angiotensin converting enzyme-2 receptors which are mostly found at upper respiratory tract (Zhou et al., 2020). While large droplets are effective in upper respiratory tract, aerosols can reach down to alveoli in lower respiratory tract. Hence, the droplet size can be effective to determine the infection risk according to the type of the disease. Moreover, the viruses diffuse among human cells, bind to receptors and copy their genetic material in a random way. All of these issues at the inter- and intracellular level need to be modeled for an accurate transmission model for the spread of infectious diseases in MoHANETs. These modeling efforts can also contribute to drug and vaccine developments.

7.3.3. Networking Layer

What we examine up to here in lower layers of the MoHANET architecture is about the transmission of infectious diseases between two humans. However, these transmissions occur many times in an epidemic, which requires a perspective to handle the population as a connected group, that is, a network. In the networking layer, the details of the MoHANET architecture are presented in order to model the spread of infectious diseases in a large scale (km) within the communication engineering framework as shown in Figure 7.2.

7.3.3.1. Mobile Human Ad Hoc Networks

In epidemiology literature, each human, that is, a node, can be represented as susceptible (S), exposed (E), infectious (I) or recovered (R) according to the SEIR-based models in the infectious disease modeling approaches (Rock et al., 2014). According to the disease type, different combinations of these node types can be employed for the models such as SIR or SIRS. For example, COVID-19 is suitable to use all the node types due to a non-infectious incubation period. In the literature, the number of these node types are modeled by ordinary differential equations where the number of the nodes can be deterministic or a stochastic process. The transition among different types of nodes (S,E,I,R) are defined with certain rates which are obtained by fitting statistical epidemic data. In experimental studies, these data are obtained by oral surveys or exploiting wireless sensor network technology (Rock et al., 2014). It is noteworthy that very few studies model the spatial change of the epidemic rather than its temporal change.

By utilizing the widespread SIR model, a MoHANET is given in Figure 7.5 which gives both the spatial and temporal changes. As the time elapses, the number of nodes may alter and the nodes can make transitions between states such as S, I or R. For example, a susceptible node can become infected, if it is in the transmission range of an infectious node or an infectious node can recover after a certain period.

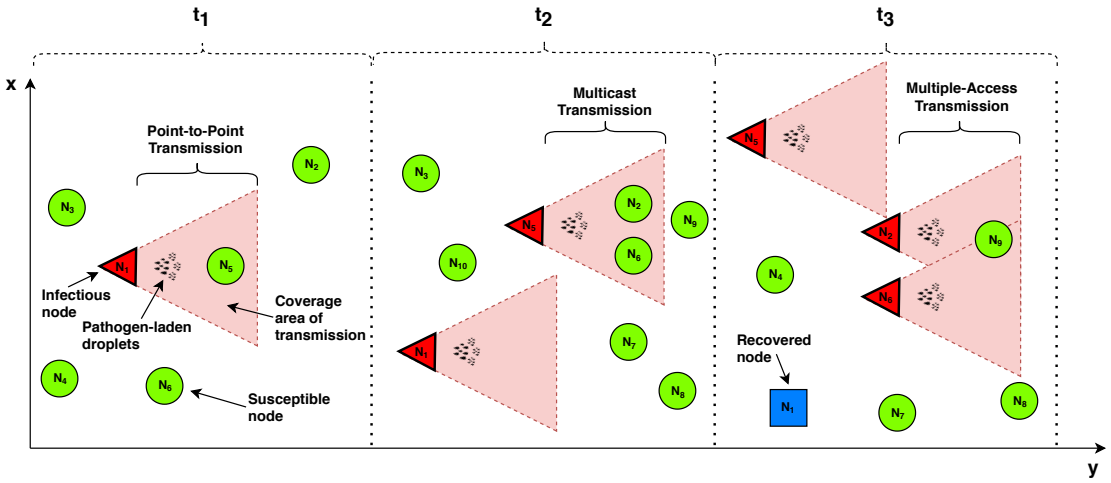


Figure 7.5. The spread of an infectious disease in a MoHANET through 2-D space for three different time instances. As time progresses, the number of nodes changes with mobility and the nodes change their state according to their exposure to pathogen-laden droplets.

7.3.3.2. Transmission Types in MoHANETs

As illustrated in Figure 7.5, three transmission types are defined for the propagation of pathogen-laden droplets from the infectious nodes to the susceptible nodes as follows:

- **Point-to-Point Transmission** includes the communication between two nodes where the infectious and susceptible nodes are the TX and RX, respectively.
- **Multicast Transmission** is the scheme that one infectious node spreads the disease to more than one node within its communication range.
- **Multiple-Access Transmission** comprises the scenario where a susceptible node is exposed to pathogen-laden droplets from multiple infectious nodes.

7.3.3.3. Routing and Mobility in MoHANETs

Humans are susceptible to infectious diseases in indoor places such as public transportation vehicles, shopping malls or offices. However, this is not the case that is encountered continuously. Instead, the risk to get infected is intermittent due to the mobility

of humans. As people displace, their smart phones can communicate opportunistically with each other as they are in the communication range. The same type of networking is also used in many applications such as wireless sensor, vehicular, and flying ad hoc networks. These dynamically changing structures defined as mobile ad hoc networks (MANETs) enable communication using the infrastructure at their location without a dedicated router. Therefore, a MoHANET can be resembled as a specific type of MANET, that is, a delay tolerant network (DTN) in which an end-to-end link among the nodes may not always exist. The nodes in a DTN store their data and wait until they find a suitable connection. By considering this waiting delay, the routing algorithms in DTNs provide the path to the desired user. Similarly, an infected human can store its pathogens until finding a susceptible human to infect via airborne transmission. Hence, we propose that opportunistic routing protocols such as *epidemic* or *spray and wait* can be adopted to model the spread of the infectious diseases. Interestingly, *epidemic routing protocol* which is a reference method for routing in MANETs was already inspired by the mechanism of infectious disease spread during an epidemic (Vahdat and Becker, 2000).

In the SIR model, the effective contact rate (β) is employed to find the rate of transitions from state S to state I which is generally estimated by epidemic data (Vynnycky and White, 2010). By using mobility models as applied in MANETs, the average contact rate of humans ($\bar{N}_c$) can be determined. In addition, the average probability of infection $\bar{P}_{inf}$ can be derived by considering the propagation and reception of pathogen-laden droplets in physical and reception layers of the MoHANET. Here, $\bar{P}_{inf} = P(\bar{N}_R > \gamma) = Q((\gamma - \mu_R)/\sigma_R)$ where $\bar{N}_R$ is the received number of droplets with $\mathcal{N}(\mu_R, \sigma_R^2)$ and γ is the detection threshold as derived in (Gulec and Atakan, 2021c). Hence, we propose that β can be derived by $\beta = \bar{N}_c\bar{P}_{inf}$ to be employed in the SIR model. The results given in Fig. 7.6 show that the rate of an epidemic reduces as the average strength of the humans' immune system increases. $\bar{N}_R$ depends on the parameters such as the velocity and size distribution of droplets, exposure time and receiver geometry. Thus, this modeling approach gives the opportunity to include the parameters in the physical and reception layers in the networking layer of the MoHANET. For convenience, random waypoint model is used to determine $\bar{N}_c$ and the average recovery rate (α) to be used for the transitions from state I to state R is taken as a constant value. However, α can be estimated at the reception layer by using human's immune system response or

drug-human interaction at the cellular level.

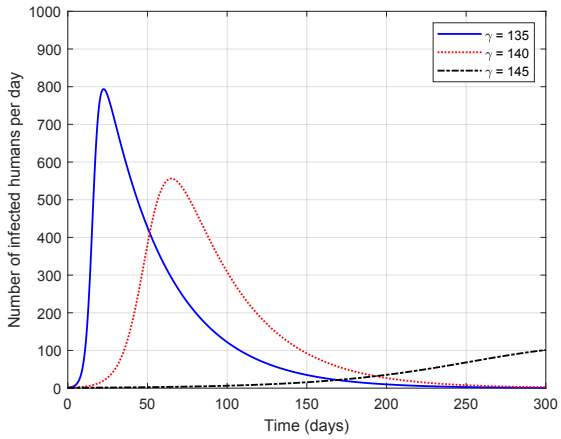


Figure 7.6. Number of infected humans (nodes) in a MoHANET according to the random waypoint model for mobility and SIR model with parameters from physical and reception layers. Number of nodes = 1000, initial number of infected nodes $I(0) = 1$, initial number of susceptible nodes $S(0) = 999$, maximum node velocity = 3 m/s, simulation area = 2000 m $\times$ 2000 m, infection (transmission) range = 1 m, $\alpha = 0.025$, $\bar{N}_R \sim \mathcal{N}(120, 100)$.

7.4. Experimental Techniques and Simulations

In order to observe and model the airborne transmission mechanisms among humans, experimental setups and computer simulations can be employed. In this section, we present and discuss how the performance of the proposed methods in different layers of the MoHANET architecture can be evaluated.

In physical and reception layers, the emulation of breathing, coughing and sneezing in experimental setups are realized by respiratory machines or thermal manikins which can be heated to change their temperature. These devices emit tracer gases including droplets. The concentration of droplets is measured by air samplers or via imaging techniques such as particle image velocimetry which gives the velocity and directions of droplets (Ai and Melikov, 2018). Moreover, sprayer-based MC systems can also be used instead of respiratory machines, manikins and air samplers.

Albeit reliable results can be obtained by the physical experiments regarding the

consideration of droplet-air interaction and airflows, the collected data have a low-resolution in space and time and experimental devices are expensive. Therefore, computational fluid dynamics (CFD) simulations are employed to evaluate the airborne transmission mechanisms with a high spatiotemporal resolution and less cost (Ai and Melikov, 2018). However, the simulation software programs are based on Navier-Stokes equations which lack the capability to model all of the effects during the transmission realistically.

These experimental techniques and CFD simulations can be employed to model the airborne pathogen transmission with communication engineering perspective for various scenarios between two humans. In a larger scale, for example, in a crowded city, it is essential to model the spread of infectious diseases with an approach that takes into account the interaction of people and their mobility in both time and space. The movement patterns of humans can be simulated by synthetic models such as random waypoint model or trace-based models which rely on real mobility data of mobile nodes as applied in MANETs. The adapted routing protocols for MoHANETs can also be evaluated in time and space by employing these mobility models according to the scenario via network simulation software. With a holistic perspective, new software is needed to model all of the issues at different layers of the MoHANET in a single platform.

CHAPTER 8

LOCALIZATION OF A PASSIVE MOLECULAR TRANSMITTER WITH A SENSOR NETWORK

8.1. Introduction

All of the macroscale platforms proposed in the molecular communication (MC) literature and covered in Chapters 3-5 focus on active transmission of molecules such as sprayers or pumps. However, there is not any platform to understand the dynamics of macroscale MC with passive transmission such as evaporating toxic molecules from a threatening source through the air. Moreover, it can be difficult to determine the exact location of the transmitter (TX) with only the distance information. Therefore, localization methods are needed for multi-dimensional practical scenarios. In the literature, a localization algorithm is proposed by using a mobile search robot as the receiver (RX) moving towards the source according to molecule concentration gradient for a long range underwater scenario (Qiu et al., 2015). However, the performance of this algorithm is not known for a practical scenario. Hence, there is not any experimentally validated localization method for practical macroscale scenarios.

Within this context, we propose a novel experimental platform for macroscale MC applications and a novel localization algorithm by using this platform, which was first presented in (Gulec and Atakan, 2020b). Firstly, our experimental platform consists of a passive source which include freely evaporating ethanol molecules. The experimental platform is placed in a fume hood which is a closed box to provide controlled conditions. Evaporating molecules are detected by a sensor network (SN) which includes 24 MQ-3 alcohol sensor nodes in a rectangular order. The novelty of our experimental platform lies in the usage of the SN which can pave the way to novel methods by adapting techniques from the SN literature. Moreover, the concept of employing a SN can be applied for different practical scenarios such as the localization of an underwater molecular TX.

Secondly, the Sensor Network-based Clustered Localization Algorithm (SNCLA) is proposed for the localization of a passive molecular TX as a proof-of-concept application employing our experimental platform. The SN is divided into four clusters. Primarily, the Gaussian plume model which is employed widely in the meteorology literature to model the movement of the pollutant particles in the air is given as the system model. As for the SNCLA, the location estimator is derived for the sensor node pairs in each cluster. In order to use the location estimator, some experimental parameters such as the actual concentration, transmitted mass and the wind velocity are required to be estimated or calculated. To this end, the measured sensor voltage is smoothed by using a moving average filter and a detection is made according to a predetermined detection threshold voltage. The measured sensor voltages at the chosen threshold voltage are converted to actual concentration values via the sensitivity response of the sensors. The detection time of the SN is employed to estimate the velocity of the wind in the medium for four directions on the $x - y$ plane. The estimated wind velocity is taken as the input for the mass calculation of the evaporated molecules. The location estimator employs all these estimated/calculated values as the input. Finally, SNCLA determines two clusters according to the magnitude of the wind velocities estimated for the four directions and makes the location estimation for the sensor nodes in these clusters. The numerical results show that SNCLA performs better, when the wind velocity is higher. Furthermore, the average detection times for all of the sensor nodes are given to show the propagation pattern of the evaporating molecules. Surprisingly, the evaporating molecules do not propagate in an isotropic fashion. It is observed that there is always a wind in the medium that affects the propagation of molecules. In addition, cluster error is defined as an error metric to evaluate the performance of the clusters in the SN. It is shown that cluster errors decrease and more stable results can be obtained for higher detection thresholds.

The remainder of the paper is organized as follows. In Section 8.2, the experimental platform is given in detail. Section 8.3 introduces the system model on which the SNCLA is based. The SNCLA is presented in Section 8.4. Finally, the numerical results are shown and analyzed in Section 8.5.

8.2. Experimental Platform

In this section, the experimental platform which is employed for the localization of a molecular transmitter using a SN is introduced. As shown in Figure 8.1, this platform consists of a TX and a SN placed inside a fume hood, which is a closed cabinet to conduct chemical experiments at controlled conditions without being exposed to chemicals. The TX includes a pipette pump, two pipettes, a rubber hose and a circular petri dish. The pipette connected to the tip of the pipette pump is filled with liquid ethanol before the transmission. When liquid ethanol is pumped through the rubber hose, it fills the petri dish which has a radius of 2.25 cm. The petri dish is deployed at the midpoint of the SN. The transmission is realized by the evaporation of ethanol molecules in the petri dish at room temperature $(25°C)$. After the transmission, evaporated ethanol molecules propagate in the air.

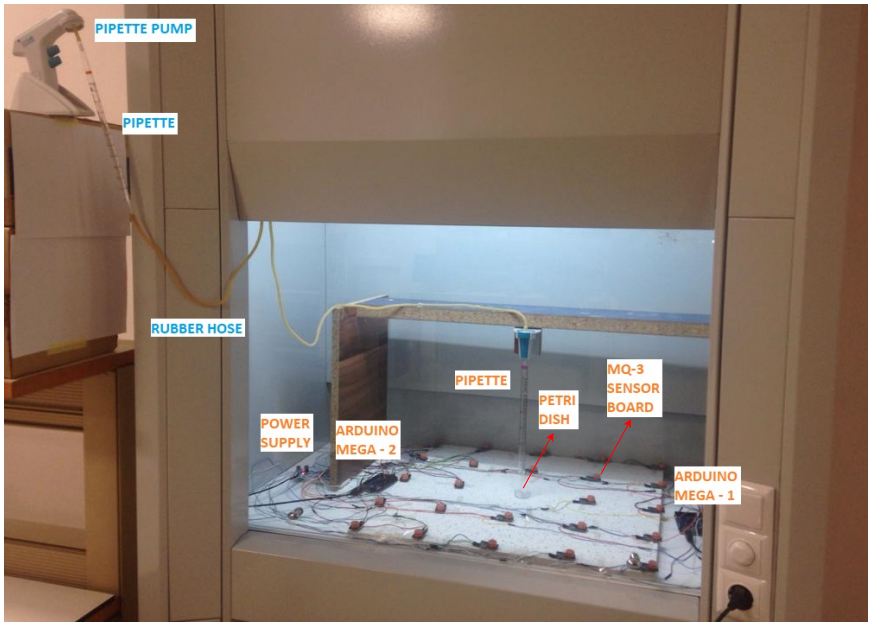


Figure 8.1. Experimental platform.

The SN consists of 24 MQ-3 alcohol sensor boards (or nodes), a power supply and two Arduino Mega microcontroller boards which are connected to a computer. The sensor nodes are placed on a rectangular surface of 60×60 cm. The distance between two adjacent nodes on the horizontal and vertical axis is 15 cm. Each sensor board has

a 1 kΩ load resistor on it in order to generate a voltage to be an analog input signal for the microcontroller board. As shown in Figure 8.1, while fourteen of the sensor nodes are wired to the first Arduino microcontroller board, ten of them are connected to the second Arduino board. In order to synchronize the TX and SN, the nodes start to receive signals as soon as the petri dish is filled with 5 ml of liquid ethanol. Next, the system model to explain the propagation of evaporated molecules employed in the proposed experimental platform is given.

8.3. System Model

This section details the system model on which the localization algorithm is based. In macroscale MC, diffusion-based models are employed to explain the propagation of molecules through the air (Farsad et al., 2014; McGuiness et al., 2018; Zhai et al., 2018). Unlike macroscale experimental studies in the literature, molecules are released by evaporation at room temperature in our scenario. Since there is not any applied force for the emission, we can classify it as a passive transmission. The absence of the force applied to this emission makes molecules susceptible to the effects of wind or flows in the air, even at low velocities. Actually, there is almost always a slight wind in the air (Hanna et al., 1982). Considering the passive transmission of molecules and winds in the air, Gaussian plume model, which is widely used in the meteorology literature for the dispersion of air pollutants, can be applied for our scenario to model the propagation of evaporated molecules. By using the conservation of mass, the equation below can be written as (Stockie, 2011)

$$\frac{\partial C}{\partial t} + \nabla \cdot \vec{J} = S, \tag{8.1}$$

where S is the source term, $\vec{J}$ and C represent the mass flux and concentration of evaporated molecules, respectively. Here, the mass flux can be given as the summation of diffusive flux $\vec{J_D}$, which stems from the turbulent diffusivity in the atmosphere, and the advective flux $\vec{J_a}$ stemming from the wind velocity ($\vec{u}$). Hence, the mass flux is given as

$$\vec{J} = \vec{J_D} + \vec{J_a} = -\vec{K}\nabla C + C\vec{u}, \tag{8.2}$$

where $\vec{K} = diag(K_x, K_y, K_z)$ is a diagonal matrix showing the turbulent diffusivities in three dimensions. Thus, (8.1) takes the form of the equation which is known as the

atmospheric diffusion (or dispersion) equation as given by (Stockie, 2011)

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\vec{u}) = \nabla \cdot (\vec{K}\nabla C) + S. \tag{8.3}$$

In our scenario, we define the TX at the position (x_T, y_T, z_T) as an instantaneous source to have a time-dependent solution in (8.3). In fact, our experimental platform is in a sufficiently small scale so that the TX can be considered as a source releasing molecules in an instantaneous puff. Furthermore, the wind velocity is defined with two components in x and y axes, i.e., u_x and u_y. It is assumed that the plane at $z = 0$ is a reflective plane and there is not any other boundaries. For the source term which is defined as $S = \frac{m_T}{\vec{u}}\delta(x - x_T)\delta(y - y_T)\delta(z - z_T)\delta(t)$ where m_T is the transmitted mass and $\delta(.)$ is the Dirac delta function, the solution of (8.3) is given as (De Visscher, 2013)

$$C(x, y, z, t) = \frac{(\pi t)^{-3/2} m_T}{8(K_x K_y K_z)^{1/2}} \exp\left(-\frac{(x - x_T - u_x t)^2}{4K_x t} - \frac{(y - y_T - u_y t)^2}{4K_y t}\right) \times \left[\exp\left(-\frac{(z - z_T)^2}{4K_z t}\right) + \exp\left(-\frac{(z + z_T)^2}{4K_z t}\right)\right], \tag{8.4}$$

which is known as the Gaussian puff solution. Here, $e^{-\frac{(z+z_T)^2}{4K_z t}}$ represents the reflection of the plume from the ground. In the literature of atmospheric dispersion, the turbulent diffusivities are defined in terms of dispersion parameters such that $\sigma_x^2 = 2K_x t$, $\sigma_y^2 = 2K_y t$, $\sigma_z^2 = 2K_z t$ (Seinfeld and Pandis, 2016). Hence, (8.4) is rearranged as

$$C(x, y, z, t) = \frac{m_T}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \exp\left(-\frac{(x - x_T - u_x t)^2}{2\sigma_x^2} - \frac{(y - y_T - u_y t)^2}{2\sigma_y^2}\right) \times \left[\exp\left(-\frac{(z - z_T)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z + z_T)^2}{2\sigma_z^2}\right)\right]. \tag{8.5}$$

The advantage of this conversion is to determine the dispersion parameters (σ_x, σ_y, σ_z) by using empirically derived models which depend on the distance between the TX and RX. According to the model given in (Briggs, 1973) which is widely used in the meteorology literature, σ_y and σ_z for stable air conditions as in our case are calculated by

$$\sigma_y = \frac{0.04r}{(1 + 0.0001r)^{0.5}} \tag{8.6}$$

$$\sigma_z = \frac{0.016r}{(1 + 0.0003r)}, \tag{8.7}$$

where r is the distance to the source in meters and σ_x can be approximated as $\sigma_x \approx \sigma_y$ (De Visscher, 2013). Regarding these empirical models in (8.6) and (8.7), the effect of

the dispersion parameters on the system model are negligible, since the scale of our SN is small (60×60 cm). Therefore, the dispersion parameters are defined as constant values. In our scenario, the SN and TX are all deployed at $z = 0$. Accordingly for each sensor, the concentration is given as

$$C_{i,j} = \frac{m_T}{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_z} \exp\left(-\frac{(x_{i,j} - x_T - u_x t_{i,j})^2}{2\sigma_x^2} - \frac{(y_{i,j} - y_T - u_y t_{i,j})^2}{2\sigma_y^2}\right), \tag{8.8}$$

where $(x_{i,j}, y_{i,j})$ and $t_{i,j}$ show the location and detection time of the node $N_{i,j}$ which is in the i^{th} row and j^{th} column of the SN, respectively. Here, $i = 1, ..., M_r$ and $j = 1, ..., M_c$.

In addition, the deployment of the sensor nodes are illustrated in Figure 8.2. It is assumed that each sensor node knows its location in the Cartesian coordinate system. As shown in this figure, the SN is divided into four clusters. These clusters are employed for the localization algorithm of the TX in MC as given in the next section.

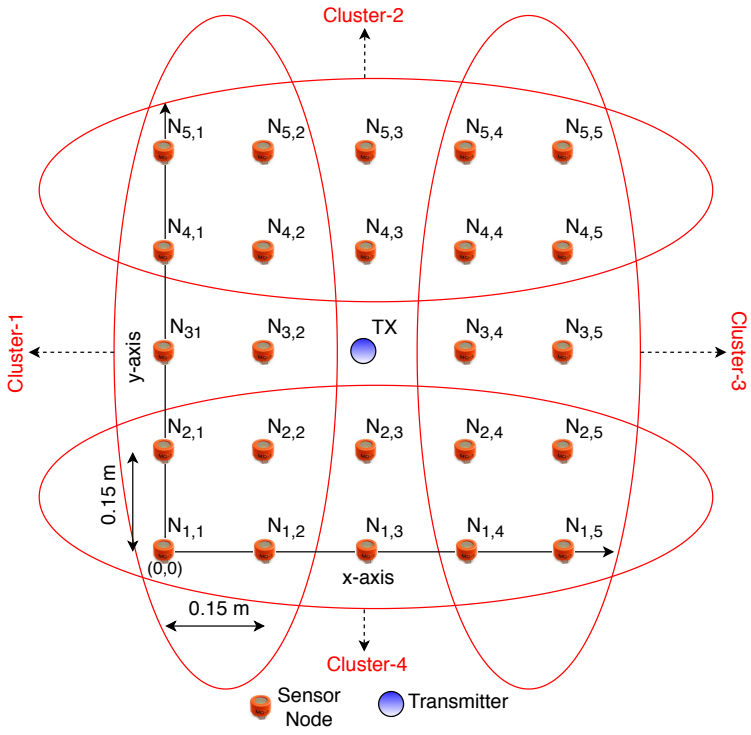


Figure 8.2. The deployment of the sensor nodes and TX.

8.4. Sensor Network-Based Clustered Localization Algorithm

In this section, Sensor Network-Based Clustered Localization Algorithm (SNCLA) whose block diagram is given in Figure 8.3 is proposed. First, the location estimator is derived using the system model given in Section 8.3. Then, the estimation and calculation of the required parameters for the location estimator is detailed. At the end of this section, SNCLA is detailed by employing all the estimated and calculated parameters.

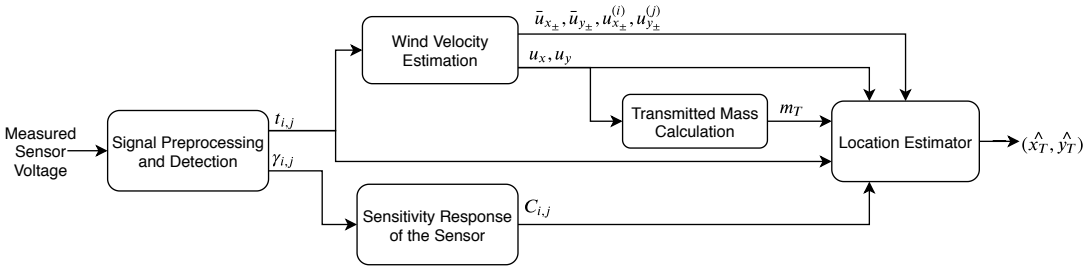


Figure 8.3. Block diagram of the SNCLA.

8.4.1. Derivation of the Location Estimator

In order to derive the location estimator, (8.8) can be written as

$$\frac{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_z C_{i,j}}{m_T} = \exp\left(-\frac{(x_{i,j}-x_T-u_x t_{i,j})^2}{2\sigma_x^2} - \frac{(y_{i,j}-y_T-u_y t_{i,j})^2}{2\sigma_y^2}\right). \tag{8.9}$$

When the natural logarithm, i.e., ln(.), of both sides is taken, then (8.9) is given by

$$\ln\left(\frac{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_z C_{i,j}}{m_T}\right) = -\frac{(x_{i,j}-x_T-u_x t_{i,j})^2}{2\sigma_x^2} - \frac{(y_{i,j}-y_T-u_y t_{i,j})^2}{2\sigma_y^2}. \tag{8.10}$$

For convenience, let $n_{i,j} = \ln\left(\frac{\sqrt{2}(\pi)^{3/2}\sigma_x\sigma_y\sigma_z C_{i,j}}{m_T}\right)$. Hence, the final equation for the location estimator is given by

$$\frac{(x_{i,j}-x_T-u_x t_{i,j})^2}{2\sigma_x^2} + \frac{(y_{i,j}-y_T-u_y t_{i,j})^2}{2\sigma_y^2} + n_{i,j} = 0. \tag{8.11}$$

For two sensor nodes, a system of nonlinear equations can be generated using (8.11) where x_T and y_T are the variables and the other parameters are constant. Since the solution of this system is not easily tractable, numerical methods can be used to obtain the solution as detailed later in this section. In order to solve these equations, parameters such as $C_{i,j}$, $t_{i,j}$, m_T and wind velocity values are required to be estimated or calculated.

8.4.2. Signal Preprocessing and Detection

In our experimental platform, the concentration is measured as a voltage value from the sensor nodes. Due to the random movements of molecules, there are fluctuations on the measured sensor voltage. In order to detect the signals more accurately, the received signal by the sensor is needed to be smoothed via removing the fluctuations. Therefore, a moving average filter is employed as defined by (Oppenheim, 1999)

$$y[n] = \frac{1}{L} \sum_{k=0}^{L} C[n-k]. \tag{8.12}$$

where $C[n]$ is the measured sensor voltage, $y[n]$ and L are the output and window size of the filter, respectively.

When there is no transmission from the TX, the sensors still output a positive voltage value which is defined as the offset level, i.e., $A_{o_{i,j}}$. Since the offset levels of the sensors can be different, the threshold voltage ($\gamma_{i,j}$) is defined for each sensor node by employing a constant detection threshold amplitude (A_T) as given by

$$\gamma_{i,j} = A_{o_{i,j}} + A_T. \tag{8.13}$$

In order to calculate $A_{o_{i,j}}$, the first p samples of the received signal is averaged before the moving average filter. After $\gamma_{i,j}$ is determined for each sensor node, the time instances that each sensor reaches the $\gamma_{i,j}$ value in $y[n]$ is recorded as $t_{i,j}$. These $t_{i,j}$ and $\gamma_{i,j}$ values are employed as input for velocity estimation and the sensitivity response of the sensor, respectively.

8.4.3. Sensitivity Response of the Sensor

Sensor voltage is obtained via a sensor measurement circuit on the MQ-3 sensor boards which is shown in Figure 5.4. As the concentration around the sensor changes, its resistance (R_s) changes. Hence, the molecule concentration is converted to an electrical signal via the circuit in Figure 8.4 where V_{out} gives the output voltage. Using this circuit, R_s is derived as

$$R_s = \left(\frac{V_{in}}{V_{out}} - 1 \right) R_l, \tag{8.14}$$

where V_{in} shows the DC input voltage and R_l is the load resistance.

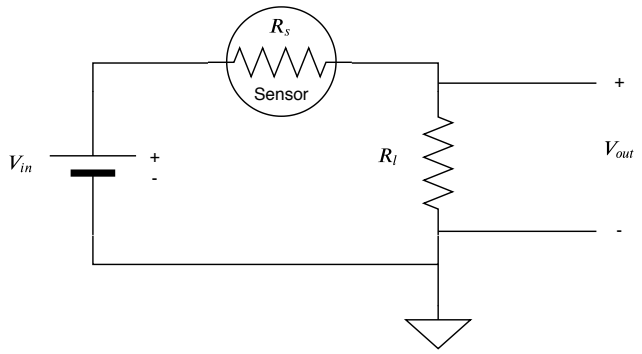


Figure 8.4. Measurement circuit of the sensor board.

For each concentration value, the sensor has a different R_s value. The sensor resistance can be normalized by dividing R_s to R_o where R_o is the sensor resistance measured at the concentration value of 0.0004 kg/m^3 which is the minimum concentration level MQ-3 sensor can measure (Hanwei-Electronics, 2018). According to its datasheet, MQ-3 sensor has a sensitivity characteristic which maps each concentration value to the normalized resistance value (R_s/R_o) (Hanwei-Electronics, 2018). This sensitivity characteristic can be expressed as a sensitivity function

$$f\left(C_{i,j}\right) = \frac{R_s}{R_o} = \left(\frac{V_{in}}{V_{out}} - 1 \right) \frac{R_l}{R_o}, \tag{8.15}$$

where $C_{i,j}$ is the actual molecule concentration around the sensor and R_s/R_o is given by substituting (8.14) into (8.15). By employing the values in its datasheet, $f(C_{i,j})$ can be obtained via curve fitting technique. Nonlinear least squares method that minimizes the

sum of the square errors is employed to fit the datasheet values of the MQ-3 sensitvity characteristic. As the result of the curve fitting, $f(C_{i,j})$ is given by

$$f\left(C_{i,j}\right) = a_1 \left(C_{i,j}\right)^{b_1} + d_1, \tag{8.16}$$

where a_1, b_1 and d_1 are the curve fitting parameters. By employing Levenberg-Marquardt algorithm, these parameters are estimated as $a_1 = 0.0116$, $b = -0.5855$ and $d_1 = -0.0743$ with a Root Mean Square Error (RMSE) value of 0.0371 (Hagan and Menhaj, 1994). The sensitivity response of the MQ-3 sensor is also employed in Chapter 5 which is a part of the signal reconstruction approach of the RX in macroscale. The signal reconstruction of the RX is first proposed in (Atakan and Gulec, 2019) in order to investigate how the actual concentration around the RX is sensed in microscale.

In order to find the molecule concentration for the given detection threshold voltage, V_{out} is set as $\gamma_{i,j}$ in (8.15). (8.15) and (8.16) are combined to obtain the equation as given by

$$\left(\frac{V_{in}}{V_{out}} - 1\right)\frac{R_l}{R_o} = a_1 \left(C_{i,j}\right)^{b_1} + d_1. \tag{8.17}$$

As the result of the sensitivity response of the sensor, (8.17) is manipulated to obtain $C_{i,j}$ which is given by

$$C_{i,j} = \left(\frac{V_{in}R_l - \gamma_{i,j}R_l - d_1\gamma_{i,j}R_o}{\gamma_{i,j}R_o a_1}\right)^{(1/b_1)}. \tag{8.18}$$

8.4.4. Wind Velocity Estimation

After the threshold voltages ($\gamma_{i,j}$) and detection times ($t_{i,j}$) are obtained for each sensor, the wind velocity flowing over two sensor nodes can be estimated in x and y directions as given by

$$u_x = \frac{|x_2 - x_1|}{t_2 - t_1}, u_y = \frac{|y_2 - y_1|}{t_2 - t_1}, \tag{8.19}$$

where (x_1, y_1) and (x_2, y_2) are the coordinates for the first and second sensor node, respectively and t_1 and t_2 are the detection times for the first and second sensor node, respectively. For our scenario, (8.19) is generalized by averaging the wind velocities estimated by the sensor node pairs in the corresponding cluster defined in Figure 8.3 for four

directions according to the formulas given below

$$\bar{u}_{x_-} = \frac{1}{M_r}\sum_{i=1}^{M_r} u_{x_-}^{(i)} = \frac{1}{M_r}\sum_{i=1}^{M_r} \frac{|x_{i,1} - x_{i,2}|}{(t_{i,1} - t_{i,2})}, \text{ Cluster-1 (-x direction)} \tag{8.20}$$

$$\bar{u}_{x_+} = \frac{1}{M_r}\sum_{i=1}^{M_r} u_{x_+}^{(i)} = \frac{1}{M_r}\sum_{i=1}^{M_r} \frac{|x_{i,5} - x_{i,4}|}{(t_{i,5} - t_{i,4})}, \text{ Cluster-3 (+x direction)} \tag{8.21}$$

$$\bar{u}_{y_+} = \frac{1}{M_c}\sum_{j=1}^{M_c} u_{y_+}^{(j)} = \frac{1}{M_c}\sum_{j=1}^{M_c} \frac{|y_{5,j} - y_{4,j}|}{(t_{5,j} - t_{4,j})}, \text{ Cluster-2 (+y direction)} \tag{8.22}$$

$$\bar{u}_{y_-} = \frac{1}{M_c}\sum_{j=1}^{M_c} u_{y_-}^{(j)} = \frac{1}{M_c}\sum_{j=1}^{M_c} \frac{|y_{1,j} - y_{2,j}|}{(t_{1,j} - t_{2,j})}, \text{ Cluster-4 (-y direction)} \tag{8.23}$$

where $u_{x_\pm}^{(i)}$ and $u_{y_\pm}^{(j)}$ show the instantaneous wind velocity of the i^{th} and j^{th} sensor node pair in the corresponding direction (or cluster), respectively, $\bar{u}_{x_\pm}$ and $\bar{u}_{y_\pm}$ show the average of these sensor node pair velocities in the corresponding direction (or cluster), respectively, $x_{i,j}$ and $y_{i,j}$ indicate the horizontal and vertical position of the sensor node $N_{i,j}$ given in Figure 8.2, respectively and M_r and M_c are the total number of rows and columns of the SN, respectively. Here, the instantaneous velocities whose values are negative are not considered for the velocity estimation. During the experiments it is observed that the wind blows stronger in one direction which means that it can only have at most two velocity components among the estimated velocities in four directions. Therefore, u_x and u_y are defined as:

$$u_x = \max(\bar{u}_{x_-}, \bar{u}_{x_+}), \tag{8.24}$$

$$u_y = \max(\bar{u}_{y_-}, \bar{u}_{y_+}). \tag{8.25}$$

8.4.5. Transmitted Mass Calculation

The estimated values of u_x and u_y are used to calculate the evaporation rate of ethanol in the air (Q_e). For the wind blowing over a surface with a velocity u at room temperature ($25°C$), Q_e (kg/m^{2}s) is given by (Lyulin et al., 2015)

$$Q_e = h_1 u^{0.54}, \tag{8.26}$$

where $h_1 = 4 \times 10^{-3}\ kg/m^3$ and $u = \sqrt{u_x^2 + u_y^2}$. In order to find the mass flow rate of evaporated molecules, i.e., Q (kg/s), which is defined as the mass flowing through a

surface per unit time, $Q = Q_e A$ where A is the surface area of evaporated molecules, i.e., the surface area of the petri dish for our case. Here, the instantaneous puff of the TX which is represented by $\delta(t)$ in the system model is approximated by a pulse with a short emission time (T_e). Hence, the transmitted mass can be calculated as (Munson et al., 2009a)

$$m_T = QT_e = Q_e A T_e. \tag{8.27}$$

Algorithm 3 SNCLA

input: $\bar{u}_{x_\pm}$, $\bar{u}_{y_\pm}$, $u^{(i)}_{x_\pm}$, $u^{(j)}_{y_\pm}$, m_T, $C_{i,j}$, $t_{i,j}$ for all $i = 1, ..., M_r$, $j = 1, ..., M_c$
if ($u_x == u_{x_-}$) and ($u_y == u_{y_+}$) **then**
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.28) for Cluster 1
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.29) for Cluster 2
else if ($u_x == u_{x_-}$) and ($u_{y_-} == u_{y_+}$) **then**
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.28) for Cluster 1
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.31) for Cluster 4
else if ($u_x == u_{x_+}$) and ($u_y == u_{y_+}$) **then**
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.29) for Cluster 2
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.30) for Cluster 3
else
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.30) for Cluster 3
    Calculate $(\hat{x_T}, \hat{y_T})$ by (8.31) for Cluster 4
end if

8.4.6. Operation of the SNCLA

Thus far, the required input parameters for the location estimator (see Figure 8.3) are obtained. By using these parameters, Algorithm 3 is proposed for the localization of the TX. In this algorithm, two clusters are chosen according to the direction of the wind velocity on x and y axes. For instance, if the wind blows stronger in the $+x$ direction on the x-axis and $+y$ direction on the y axis, then the node pairs in Cluster-3 and Cluster-2 are chosen for the location estimation. Similar to the wind velocity estimation, the node pairs whose absolute instantaneous wind velocity value is negative are not considered for the location estimation. Afterwards, the equation pairs given in (8.28)-(8.31) are solved for x_T and y_T according to the chosen two clusters. The solution for each of two equations gives the estimated coordinates of the TX, i.e., $\hat{x}_T$ and $\hat{y}_T$. The solutions of (8.28)-(8.31) are too long to write in this paper. Instead, these equations are solved numerically as given

in the numerical results.

$$\left.\begin{aligned}\frac{(x_{i,1}-x_T-u_x t_{i,1})^2}{2\sigma_x^2}+\frac{(y_{i,1}-y_T-u_y t_{i,1})^2}{2\sigma_y^2}+n_{i,1}=0\\ \frac{(x_{i,2}-x_T-u_y t_{i,2})^2}{2\sigma_x^2}+\frac{(y_{i,2}-y_T-u_y t_{i,2})^2}{2\sigma_y^2}+n_{i,2}=0\end{aligned}\right\} \begin{matrix} i=1,...,M_r. \\ \text{(Cluster-1)}\end{matrix} \tag{8.28}$$

$$\left.\begin{aligned}\frac{(x_{5,j}-x_T-u_x t_{5,j})^2}{2\sigma_x^2}+\frac{(y_{5,j}-y_T-u_y t_{5,j})^2}{2\sigma_y^2}+n_{5,j}=0\\ \frac{(x_{4,j}-x_T-u_x t_{4,j})^2}{2\sigma_x^2}+\frac{(y_{4,j}-y_T-u_y t_{4,j})^2}{2\sigma_y^2}+n_{4,j}=0\end{aligned}\right\} \begin{matrix} j=1,...,M_c. \\ \text{(Cluster-2)}\end{matrix} \tag{8.29}$$

$$\left.\begin{aligned}\frac{(x_{i,5}-x_T-u_x t_{i,5})^2}{2\sigma_x^2}+\frac{(y_{i,5}-y_T-u_y t_{i,5})^2}{2\sigma_y^2}+n_{i,5}=0\\ \frac{(x_{i,4}-x_T-u_y t_{i,4})^2}{2\sigma_x^2}+\frac{(y_{i,4}-y_T-u_y t_{i,4})^2}{2\sigma_y^2}+n_{i,4}=0\end{aligned}\right\} \begin{matrix} i=1,...,M_r. \\ \text{(Cluster-3)}\end{matrix} \tag{8.30}$$

$$\left.\begin{aligned}\frac{(x_{1,j}-x_T-u_x t_{1,j})^2}{2\sigma_x^2}+\frac{(y_{1,j}-y_T-u_y t_{1,j})^2}{2\sigma_y^2}+n_{1,j}=0\\ \frac{(x_{2,j}-x_T-u_x t_{2,j})^2}{2\sigma_x^2}+\frac{(y_{2,j}-y_T-u_y t_{2,j})^2}{2\sigma_y^2}+n_{2,j}=0\end{aligned}\right\} \begin{matrix} j=1,...,M_c. \\ \text{(Cluster-4)}\end{matrix} \tag{8.31}$$

Table 8.1. Experimental parameters.

Parameter	**Value**
Number of measurements (M_m)	25
Detection threshold amplitude (A_T)	0.055 V
Emission time (T_e)	0.1 s
Actual TX location (x_T, y_T)	$(0.3, 0.3)$ m
Area of the petri dish (A)	0.0024 m^2
Input voltage of the sensor board (V_{in})	5 V
Load resistance (R_l)	1 kΩ
Sensor resistance at 0.0004 kg/m^3 (R_o)	24 kΩ
Standard deviation of the Gaussian concentration distribution on the x, y and z axis ($\sigma_x, \sigma_y, \sigma_z$)	0.0115 m, 0.0115 m, 0.0046 m
Number of samples to be averaged for the offset level $A_{o_{i,j}}$ of the sensors (p)	3
Window size of the moving average filter (L)	7

8.5. Numerical Results

In this section, numerical results of the SNCLA is given. 25 measurements each lasting 180 s were performed with the experimental platform. There were at least 30 min-

utes left among adjacent measurements in order to decrease the concentration level with the ventilation of the fume hood. The ventilation was not used during the measurements.

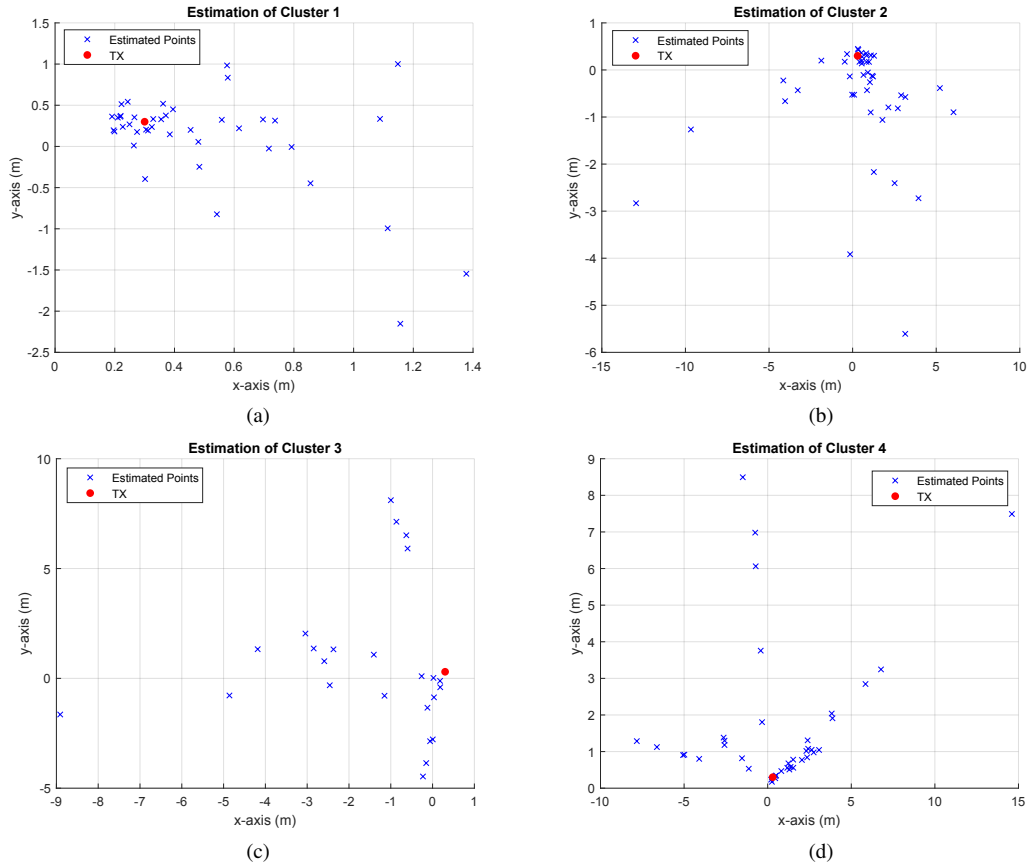


Figure 8.5. Estimated points using SNCLA for each cluster.

The experimental parameters are given in Table 8.1. Among these parameters, R_o is calculated by employing output voltage (V_{out}) of the sensor measurement circuit. According to the MQ-3 sensor datasheet, its detection scope is between 5×10^{-5} and 10^{-2} kg/m^3 (Hanwei-Electronics, 2018). This detection scope is scaled for V_{out} values between 0 and 5 V. Thus, 0.0004 kg/m^3 corresponds to $V_{out} = 0.2$ V which is used to calculate the sensor resistance value via (8.14). As mentioned in Section 8.3, the dispersion parameters $(\sigma_x, \sigma_y, \sigma_z)$ are assumed as constant values. For our experimental scenario and according to (8.6)-(8.7), the ranges of σ_y and σ_z are $0.006 - 0.017$ m and $0.0024 - 0.0068$ m. Therefore, σ_y and σ_z are chosen as the average values of these ranges. σ_x is also taken as equal to σ_y (De Visscher, 2013). The window size of the moving average filter (L) and number of samples to be averaged to determine the offset level ($A_{o_{i,j}}$) of the sensor node $N_{i,j}$, i.e., p, is determined empirically for our experimental scenario. In addition, the

detection threshold amplitude (A_T) is also chosen as an empirical value in order to have accurate estimations. However, the error values for a range of A_T values are also given.

The estimation results for each cluster are shown in Figure 8.5. For the experimental values, $\hat{x}_T$ and $\hat{y}_T$ have two complex conjugate roots for each. Therefore, only real parts of the solutions are considered for the numerical results. As shown in Figure 8.5, there are more results for Cluster-1 and Cluster-2, since the wind velocity is mostly in $-x$ and $+y$ direction for our measurements. The results of these two clusters are also more accurate than the other clusters, since the wind blows stronger in $-x$ and $+y$ direction than the other directions. The accuracy of these clusters are more clearly depicted in Figure 8.6. This figure shows the results of the best ten measurements in the same scale for a better visual perception of the figure. When the wind velocity is higher, SNCLA gives better results, since the effect of the dispersion of evaporated molecules decreases.

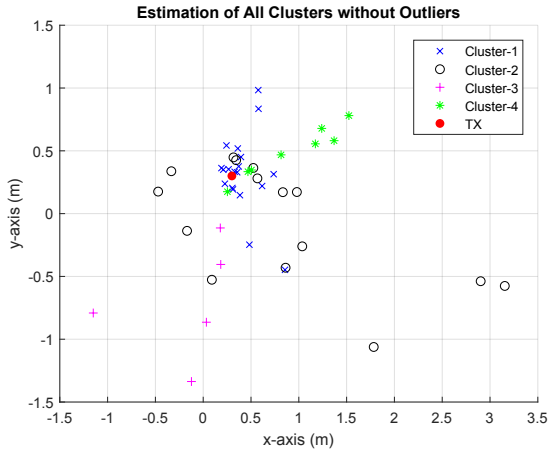


Figure 8.6. Estimated points using SNCLA for all clusters without outliers.

In Figure 8.7, the average of the detection times for each sensor node is given as a heatmap. This figure verifies the direction of the wind by the detection times. Interestingly, there is no detection for the given threshold for the sensor node $N_{5,3}$. Actually, there are also very few detections for the other sensor nodes in the third column of the SN. According to these observations, some of the evaporated molecules move in the same direction with the wind whereas the rest of the molecules move mostly in the opposite horizontal direction of the wind due to the initial puff of the TX.

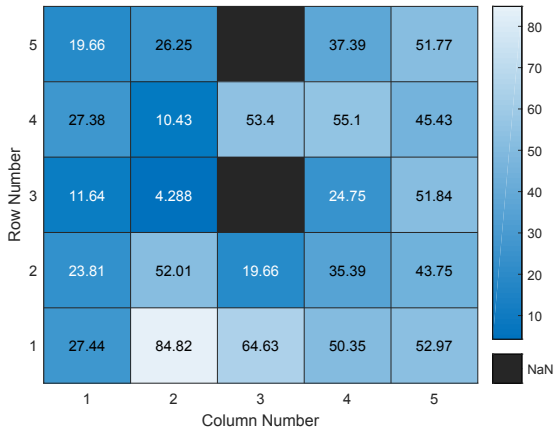


Figure 8.7. Average of the detection times for each sensor node.

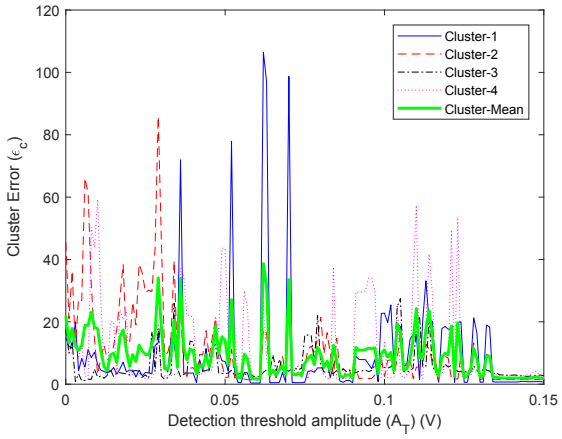


Figure 8.8. Cluster errors for different detection threshold amplitudes.

For the last part of the numerical results, an error metric, which is called Cluster Error (ϵ_c), is defined for each cluster by

$$\epsilon_c = \begin{cases} \dfrac{1}{M_r}\displaystyle\sum_{i=1}^{M_r}\dfrac{1}{M_m}\sum_{k=1}^{M_m}\sqrt{(x_T-\hat{x}_{T_k}^{(i)})^2+(y_T-\hat{y}_{T_k}^{(i)})^2}, & \text{Cluster 1,3} \quad (8.32) \\ \dfrac{1}{M_c}\displaystyle\sum_{j=1}^{M_c}\dfrac{1}{M_m}\sum_{k=1}^{M_m}\sqrt{(x_T-\hat{x}_{T_k}^{(j)})^2+(y_T-\hat{y}_{T_k}^{(j)})^2}, & \text{Cluster 2,4} \quad (8.33) \end{cases}$$

where M_m is the number of the measurements and $(\hat{x}_{T_k}^{(i)}, \hat{y}_{T_k}^{(i)})$ show the estimated points for the i^{th} node pair in the corresponding cluster at the k^{th} measurement. In our case, there are $M_r = M_c = 5$ node pairs for each cluster. First, the Euclidean distance between the actual and estimated points are calculated for each node pair in the cluster. Then,

these distances for all the measurements are averaged. This process is repeated for each node pair in the cluster. The results of ϵ_c for A_T values between $0 - 0.15$ V with 0.001 V steps are given in Figure 8.8. For higher threshold values, Cluster-1 and 2 outperform the other clusters due to the higher wind velocities. For lower thresholds, Cluster-3 has better results due to the lower number of detections. Figure 8.8 also shows that the choice of the detection threshold for lower error is significant which is left as an open research issue.

CHAPTER 9

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this chapter, the thesis is concluded and the future research topics are discussed. First, the concluding remarks are reviewed in Section 9.1. Based on these conclusions, the future research directions are identified in Section 9.2.

9.1. Conclusions

In the first part of the thesis, the studies in microscale molecular communication (MC) are given. In Chapter 2, a new concept about how accurate the molecular signal is sensed and reconstructed by the receiver nanomachines (RNs) is proposed. The RN is assumed as a perfect absorbing molecule counting machine and the reconstructed signal is modeled as a counting process. The molecule concentration is treated as a molecular signal and a signal distortion is defined as the mean square error between the existing molecular signal outside the RN and the reconstructed signal. Instead of the deterministic approach, Gaussian random process is used for the molecular signal outside the RN in a more realistic way and DSPP is obtained to model the reconstructed signal. The derived signal distortion function is validated by means of the random walk simulations. Numerical results are given to highlight the effect of the system parameters such as the diffusion coefficient, sampling period and RN radius, on the signal distortion. By the minimization of the signal distortion, the optimum RN design parameters are derived. The analysis about the effect of the signal distortion on the RN design parameters shows that the RN can reconstruct signals with a small distortion, when the RN design parameters are properly configured.

The signal reconstruction distortion, which is proposed as a novel performance parameter, can be employed to design an efficient MC system from the signal reconstruction perspective. For example, a smaller diffusion coefficient can be chosen for a faster signal reconstruction with minimum signal distortion. Furthermore, this perspective can be utilized to show the trade-off among the RN design parameters such that the optimum

sampling period decreases and the minimum signal distortion increases, as the radius of the RN decreases.

In the second part of the thesis, the studies in macroscale MC with active sources (transmitters) are presented. In Chapter 3, five distance estimation methods are proposed for a practical macroscale MC system. The existing two machine learning (ML) methods, which are multivariate linear regression (MLR) and neural network regression (NNR), are applied in distance estimation for the first time. In order to use these methods, an experimental setup was established and received signals were recorded. A novel feature extraction algorithm is proposed to generate training and test data from the measured signals for the ML methods. By analyzing these generated data, three novel methods for distance estimation are proposed and compared with the applied ML methods. The ML methods result better than data analysis based methods, with NNR being the best method. However, MLR has a very close performance to NNR, which shows that the input features such as peak time, gradient of the received signal during the transition from the noise floor up to the peak time and the received energy of the signal, have a linear relation with the distance. MLR can be an efficient way for distance estimation with high accuracy and low complexity as a ML method. Furthermore, data analysis based methods perform worse, but are not as complex as ML methods. Especially, the peak time based estimation performs close to ML methods, as the distance increases.

As a result of the experiments, the phenomena that cannot be explained with only the diffusion of the molecules are revealed. The liquid ethanol is emitted as droplets from the transmitter (TX) and these droplets are affected by the initial drift of the TX, Brownian motion and gravity. Moreover, a possible trajectory of the droplets in which there are two propagation zones as the unsteady flow zone and the Brownian motion zone is given. It is analyzed that the transmitted droplets are affected by unsteady flows due to the induced air as soon as they are emitted from the TX. Furthermore, our analysis shows that the non-linear characteristic of the sensor can cause measurement errors. The joint effect of the non-linearity of the sensor and unsteady flows complicates the estimation of the distance at the receiver (RX) side as revealed by the experimental data. Hence, it can be concluded that a fluid dynamics perspective is needed for a precise distance and any other channel parameter estimation. It is also necessary to consider the imperfection of the TX and RX to improve parameter estimation in macroscale practical MC system models.

Based on the results and analysis in Chapter 3, the the Fluid Dynamics-based Distance Estimation (FDDE) algorithm which uses a fluid dynamics perspective is proposed for distance estimation in practical macroscale MC scenarios. For this algorithm, the propagation of transmitted droplets is modeled by using fluid dynamics where it is assumed that droplets and air particles move together as a mixture, i.e., two-phase flow. Here, emitted droplets are modeled as evaporating spherical particles in the MC channel. The traveling distance of this mixture derived using this model is employed in the proposed algorithm to estimate the distance between the TX and RX. Afterwards, the FDDE algorithm is validated by experimental data. It is revealed that the distance depends on the initial velocity of droplets, air-droplet interaction in the two-phase flow, nozzle orifice, beamwidth of the TX and densities of droplet and air. In addition, the effect of droplet evaporation is shown with the numerical results. Hence, it is concluded that fluid dynamics approach can be employed to model the movements of droplets and estimate the parameters of a macroscale MC channel. In particular, the proposed method can be applied to detect the distance to a pathogen emitting source using biosensors in a scenario where pathogen-laden droplets spread through the air as a result of sneezing/coughing.

In Chapter 5, an end-to-end system model is proposed for macroscale sprayer-based MC systems. Our model takes the liquid droplets as information carriers into account with a signal reconstruction approach. The signal reconstruction of the RX considers the adhesion/detachment of droplets and the sensor's sensitivity. This study reveals the physical meanings of the channel parameters for sprayer-based MC systems.

In Chapter 6, an algorithmic end-to-end system model is proposed for droplet-based communication via coughing/sneezing between two static humans for an indoor scenario. The TX emits a cloud which is a mixture of droplets and air and it propagates under the influence of the initial momentum, gravity and buoyancy. A receiver model which defines the central part of the human face as the RX cross-section is proposed for the reception of droplets to give an output of infection state of the RX. The transmitted and received number of droplets are modeled as random processes which lead us to derive the probability of infection. Numerical results show that the safe zone for the RX starts at 1.7 m for a horizontally coughing TX. It is also revealed that the initial coughing angle of the TX, the distance between the TX and RX and the detection threshold which actually corresponds to the strength of the human immune system are significant parameters to

model the airborne pathogen transmission. Furthermore, the reception of pathogens can be affected by the sex of the human.

Chapter 7 presents a framework to model airborne pathogen transmission with a communication engineering perspective. First, airborne pathogen transmission mechanisms are reviewed and MC is utilized to model the propagation and reception of this transmission. The concept of MoHANET is proposed to handle the infectious disease spread modeling problem by using a layered structure in macro- and microscales. Furthermore, simulation techniques and experimental methods to model airborne pathogen transmission are reviewed and discussed. Throughout the article, open research issues possessing the potential for development opportunities are given. The efforts to model the infectious disease spread via airborne pathogen transmission with a novel approach given in this article has the potential for a holistic viewpoint. This communication engineering viewpoint can bring different disciplines such as fluid dynamics, medicine, biology and epidemiology together for accurate predictions about the spread of infectious diseases. Hence, the most proper intervention method (lockdown, social distancing, wearing masks, and so on) can be chosen and how it will be applied can be determined to stop the epidemics in an effective way.

Chapter 8 presents a novel experimental platform for macroscale MC and a novel algorithm for the localization of a molecular TX with a sensor network of four clusters, i.e., Sensor Network-based Clustered Localization Algorithm (SNCLA). In our experimental platform, the molecular TX emits molecules by evaporation at room temperature and the signals are received with the sensor network. First, Gaussian plume model is given as the system model for our scenario. Based on this system model, a location estimator is derived. Then, estimation/calculation methods for the unknown parameters in the location estimator such as detection time, transmitted mass, wind velocity and the actual concentration are proposed. Finally, SNCLA is explained by combining all these estimated/calculated parameters. In SNCLA, the estimated location of the TX is based on the estimated wind velocity direction and the derived location estimator. SNCLA gives more accurate results for the clusters in the same direction with the wind for higher detection threshold values. Since the Gaussian plume model on which the SNCLA is based is employed for longer distances in the meteorology domain, it is anticipated to have more accurate results on larger scales with the proposed SNCLA.

9.2. Future Research Directions

MC is studied in communication engineering literature since 2005 and there is still a long way to accomplish breakthroughs in this area. In this section, the future research directions are highlighted by giving open research issues as the extension of this thesis.

- **Microscale MC:** The studies in microscale MC literature does not take into account the signal reconstruction (or sensing) of the molecular signals by the RNs for the design of the communication systems. For example, in concentration shift keying, it is essential to consider the concentration levels for different information symbols according to the signal reconstruction of the RN. As the future work, MC methods to transmit and receive information efficiently according to the signal reconstruction of the RNs can be developed. Moreover, the future works should include the modeling and analysis of the signal reconstruction by the RNs with receptors.

- **Macroscale MC with Active Transmitters:** The experimental platforms proposed in the macroscale MC literature get sophisticated by incorporating the usage of magnetic nanoparticles as the information carriers and optical measurement techniques for the reception. However, simple sprayer-based experimental MC systems start to attract more researchers due to their role in modeling airborne pathogen transmission.

  The studies given in Chapter 3 and 4 can be extended to be applied in direction estimation by using multiple sensors or sensor networks. Moreover, estimation methods for MC channel parameters except the distance can be developed by using our fluid dynamics approach. As for the channel model proposed in Chapter 5, there is a need to investigate what the proposed novel channel parameters depend on. Furthermore, it is essential to highlight that these practical MC systems can be implemented for the research in airborne pathogen transmission with a MC perspective. In a communication system where the TX is a person emitting pathogen-laden droplets via sneezing/coughing and the RX is a biological sensor, the channel model and distance estimation methods proposed in Chapters 3-5 can be applied.

  In Chapter 6, the airborne pathogen transmission between two humans is modeled by considering the propagation of droplets and their reception by their face via

droplets. However, there are more factors affecting the transmission mechanism as given in detail in Chapter 7. The convective boundary layer stemming from the temperature difference of the human body and air, airflows of the ventilation systems, and survival rate of the pathogens during their propagation in the air should be included in the system models. In addition, the spread of infectious diseases can be modeled by using the mobile human ad hoc network (MoHANET) concept given in Chapter 7. The similarity of the human groups exchanging pathogens and mobile telecommunication networks can be exploited for modeling the spread of infectious diseases in a large scale. There is a need for further research to adopt and implement techniques from telecommunication networks for modeling the spread of the infectious disease through airborne pathogen transmission.

- **Macroscale MC with Passive Transmitters:** In the last part of the thesis, we focus on the passive sources such as evaporating alcohol molecules. Passive sources are not studied in macroscale MC literature, although they have the potential for practical applications. For example, the propagation of toxic gases causing air pollution can be analyzed with MC perspective. In our study given in Chapter 8, sensor networks are employed to find the location of a passive source for a similar scenario in small scale. As the future work, this study can be improved in larger scales via different localization algorithms from the sensor network literature with MC perspective. With the usage of the sensor networks, novel methods can be developed for the problems in macroscale MC by utilizing the wide literature of sensor networks.

REFERENCES

Abbaszadeh, M., P. J. Thomas, and W. Guo (2018). Toward high capacity molecular communications using sequential vortex rings. *IEEE Trans. on Mol., Biol. and Multi-Scale Commun. 4*(1), 39–42.

Abbaszadeh et.al., M. (2019). Mutual information and noise distributions of molecular signals using laser induced fluorescence. In *2019 IEEE Global Commun. Conf.*, pp. 1–6. IEEE.

Aghababaiyan, K., V. Shah-Mansouri, and B. Maham (2018a). Axonal channel capacity in neuro-spike communication. *IEEE transactions on nanobioscience 17*(1), 78–87.

Aghababaiyan, K., V. Shah-Mansouri, and B. Maham (2018b). Joint optimization of input spike rate and receiver decision threshold to maximize achievable bit rate of neuro-spike communication channel. *IEEE transactions on nanobioscience*.

Ai, Z. and A. K. Melikov (2018). Airborne spread of expiratory droplet nuclei between the occupants of indoor environments: A review. *Indoor air 28*(4), 500–524.

Akyildiz, I. F., F. Brunetti, and C. Blázquez (2008). Nanonetworks: A new communication paradigm. *Computer Networks 52*(12), 2260–2279.

Akyildiz, I. F., J. M. Jornet, and M. Pierobon (2011). Nanonetworks: A new frontier in communications. *Communications of the ACM 54*(11), 84–89.

Al Heidary, M., J. Douzals, C. Sinfort, and A. Vallet (2014). Influence of spray characteristics on potential spray drift of field crop sprayers: A literature review. *Crop protection 63*, 120–130.

Anderson Jr, J. D. (2016). *Fundamentals of aerodynamics*. Tata McGraw-Hill Education.

Aquino, G., N. S. Wingreen, and R. G. Endres (2016). Know the single-receptor sensing limit? think again. *Journal of statistical physics 162*, 1353.

Atakan, B. (2014). *Molecular communications and nanonetworks: from nature to practical systems*. Springer Science & Business Media.

Atakan, B. and O. B. Akan (2007). An information theoretical approach for molecular communication. In *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007. 2nd*, pp. 33–40. IEEE.

Atakan, B. and O. B. Akan (2010). Deterministic capacity of information flow in molecular nanonetworks. *Nano Communication Networks 1*(1), 31–42.

Atakan, B., O. B. Akan, and S. Balasubramaniam (2012). Body area nanonetworks with molecular communications in nanomedicine. *IEEE Communications Magazine 50*(1).

Atakan, B. and F. Gulec (2019). Signal reconstruction in diffusion-based molecular communication. *Transactions on Emerging Telecommunications Technologies 30*(12), e3699. e3699 ett.3699.

Atkins, P. and J. De Paula (2010). *Physical chemistry*. W.H. Freeman and Company.

Avila, K., D. Moxey, A. de Lozar, M. Avila, D. Barkley, and B. Hof (2011). The onset of turbulence in pipe flow. *Science 333*(6039), 192–196.

Bai, C., M. Leeson, M. D. Higgins, and Y. Lu (2016). Throughput and energy efficiency-based packet size optimisation of arq protocols in bacterial quorum communications. *Transactions on Emerging Telecommunications Technologies 27*(8), 1128–1143.

Barati, R., S. A. A. S. Neyshabouri, and G. Ahmadi (2014). Development of empirical models with high accuracy for estimation of drag coefficient of flow around a smooth sphere: An evolutionary approach. *Powder Technology 257*, 11–19.

Begg, S., F. Kaplanski, S. Sazhin, M. Hindle, and M. Heikal (2009). Vortex ring-like structures in gasoline fuel sprays under cold-start conditions. *International Journal*

of Engine Research 10(4), 195–214.

Berg, H. C. (1993). *Random walks in biology*. Princeton University Press.

Berg, H. C. and E. M. Purcell (1977). Physics of chemoreception. *Biophysical journal 20*(2), 193–219.

Bossert, W. H. and E. O. Wilson (1963). The analysis of olfactory communication among animals. *Journal of theoretical biology 5*(3), 443–469.

Bourouiba, L., E. Dehandschoewercker, and J. W. Bush (2014). Violent expiratory events: on coughing and sneezing. *Journal of Fluid Mechanics 745*, 537–563.

Briggs, G. A. (1973). Diffusion estimation for small emissions. *Atmospheric Turbulence and Diffusion Laboratory*, 83.

Busco, G., S. R. Yang, J. Seo, and Y. A. Hassan (2020). Sneezing and asymptomatic virus transmission. *Phys. of Fluids 32*(7), 073309.

Carmo, M. and J. Gubulin (1997). Ethanol-water adsorption on commercial 3a zeolites: kinetic and thermodynamic data. *Brazilian Journal of Chemical Engineering 14*, 00–00.

Christopher, M. B. (2016). *PATTERN RECOGNITION AND MACHINE LEARNING*. Springer-Verlag New York.

Cohen, F. S. (2016). How viruses invade cells. *Biophysical journal 110*(5), 1028–1032.

Coppola, M., P. Cascone, V. Madonna, I. Di Lelio, F. Esposito, C. Avitabile, A. Romanelli, E. Guerrieri, A. Vitiello, F. Pennacchio, et al. (2017). Plant-to-plant communication triggered by systemin primes anti-herbivore resistance in tomato. *Scientific reports 7*(1), 15522.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 129–164.

Dbouk, T. and D. Drikakis (2020). On coughing and airborne droplet transmission to humans. *Physics of Fluids 32*(5), 053310.

De Cock, N., M. Massinon, S. O. Salah, and F. Lebeau (2017). Investigation on optimal spray properties for ground based agricultural applications using deposition and retention models. *Biosystems engineering 162*, 99–111.

De Visscher, A. (2013). *Air dispersion modeling: foundations and applications*. John Wiley & Sons.

Duguid, J. (1946). The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. *Epidemiology & Infection 44*(6), 471–479.

Eckford, A. W. (2007). Achievable information rates for molecular communication with distinct molecules. In *Bio-Inspired Models of Network, Information and Computing Systems, 2007. Bionetics 2007. 2nd*, pp. 313–315. IEEE.

Endres, R. G. and N. S. Wingreen (2008). Accuracy of direct gradient sensing by single cells. *Proceedings of the National Academy of Sciences 105*(41), 15749–15754.

Endres, R. G. and N. S. Wingreen (2009). Maximum likelihood and the single receptor. *Physical review letters 103*(15), 158101.

Farsad, N., W. Guo, and A. W. Eckford (2013). Tabletop molecular communication: Text messages through chemical signals. *PloS one 8*(12), e82935.

Farsad, N., N.-R. Kim, A. W. Eckford, and C.-B. Chae (2014). Channel and noise models for nonlinear molecular communication systems. *IEEE Journal on Selected Areas in Communications 32*(12), 2392–2401.

Farsad, N., D. Pan, and A. Goldsmith (2017). A novel experimental platform for in-vessel multi-chemical molecular communications. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6. IEEE.

Farsad, N., H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo (2016). A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys & Tutorials 18*(3), 1887–1919.

Friedman, J., T. Hastie, and R. Tibshirani (2008). *The elements of statistical learning*. Springer series in statistics New York.

Ghosh, S. and J. C. R. Hunt (1994). Induced air velocity within droplet driven sprays. *Proc. R. Soc. Lond. A 444*(1920), 105–127.

Giannoukos, S., A. Marshall, S. Taylor, and J. Smith (2017). Molecular communication over gas stream channels using portable mass spectrometry. *Journal of The American Society for Mass Spectrometry 28*(11), 2371–2383.

Giné, L. P. and I. F. Akyildiz (2009). Molecular communication options for long range nanonetworks. *Computer Networks 53*(16), 2753–2766.

Gombrich, E. H. (1995). *The story of art*, Volume 12. Phaidon London.

Gulec, F. and B. Atakan (2020a). Distance estimation methods for a practical macroscale molecular communication system. *Nano Communication Networks 24*, 100300.

Gulec, F. and B. Atakan (2020b). Localization of a passive molecular transmitter with a sensor network. In *Int. Conf. on Bio-inspired Inf. and Commun.*, pp. 317–335. Springer.

Gulec, F. and B. Atakan (2020c). Mobile human ad hoc networks: A communication engineering viewpoint on interhuman airborne pathogen transmission. *arXiv preprint arXiv:2011.00884, submitted to IEEE Network*.

Gulec, F. and B. Atakan (2021a). A droplet-based signal reconstruction approach to channel modeling in molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications 7*(1), 64–68.

Gulec, F. and B. Atakan (2021b). Fluid dynamics-based distance estimation algorithm for macroscale molecular communication. *Nano Communication Networks 28*, 100351.

Gulec, F. and B. Atakan (2021c). A molecular communication perspective on airborne pathogen transmission and reception via droplets generated by coughing and sneezing. *to appear in IEEE Transactions on Molecular, Biological and Multi-Scale Communications*.

Guo, W., I. Atthanayake, and P. Thomas (2020). Vertical underwater molecular communications via buoyancy: Gaussian velocity distribution of signal. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6. IEEE.

Guo, W., C. Mias, N. Farsad, and J.-L. Wu (2015). Molecular versus electromagnetic wave propagation loss in macro-scale environments. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications 1*(1), 18–25.

Hagan, M. T. and M. B. Menhaj (1994). Training feedforward networks with the marquardt algorithm. *IEEE transactions on Neural Networks 5*(6), 989–993.

Halvoňová, B. and A. K. Melikov (2010). Performance of ductless personalized ventilation in conjunction with displacement ventilation: Impact of disturbances due to walking person(s). *Building and Environment 45*(2), 427–436.

Hamidović, M., U. Marta, G. Fink, R. Wille, A. Springer, and W. Haselmayr (2019). Information encoding in droplet-based microfluidic systems: First practical study. In *ACM Int. Conf. on Nanoscale Comput. and Commun.*, pp. 1–6.

Hanna, S. R., G. A. Briggs, and R. P. Hosker Jr (1982). Handbook on atmospheric diffusion. Technical report, National Oceanic and Atmospheric Administration, Oak Ridge, TN (USA).

Hansen, P. C., V. Pereyra, and G. Scherer (2013). *Least squares data fitting with*

applications. JHU Press.

Hanwei-Electronics (2018). *Technical Data of MQ-3 Gas Sensor*.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.

Hiyama, S., Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore, and T. Nakano (2005). Molecular communication. In *Proc. of the 2005 NSTI Nanotechnology Conference*, pp. 391–394.

Huang, J.-T., H.-Y. Lai, Y.-C. Lee, C.-H. Lee, and P.-C. Yeh (2013). Distance estimation in concentration-based molecular communications. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 2587–2591. IEEE.

Ishii, M. and T. Hibiki (2010). *Thermo-fluid dynamics of two-phase flow*. Springer Science & Business Media.

Khalid, M., O. Amin, S. Ahmed, B. Shihada, and M.-S. Alouini (2019). Communication through breath: Aerosol transmission. *IEEE Communications Magazine 57*(2), 33–39.

Khalid, M., O. Amin, S. Ahmed, B. Shihada, and M.-S. Alouini (2020). Modeling of viral aerosol transmission and detection. *IEEE Trans. on Commun.*.

Khaloopour et.al., L. (2019). An experimental platform for macro-scale fluidic medium molecular communication. *IEEE Trans. on Mol., Biol. and Multi-Scale Commun. 5*(3), 163–175.

Kilinc, D. and O. B. Akan (2013). Receiver design for molecular communication. *IEEE Journal on Selected Areas in Communications 31*(12), 705–714.

Killingley, B. and J. Nguyen-Van-Tam (2013). Routes of influenza transmission. *Influenza and other respiratory viruses 7*, 42–51.

Kim, N.-R., N. Farsad, C.-B. Chae, and A. W. Eckford (2015). A universal channel

model for molecular communication systems with metal-oxide detectors. In *Communications (ICC), 2015 IEEE International Conference on*, pp. 1054–1059. IEEE.

Kleppner, D. and R. Kolenkow (2013). *An introduction to mechanics*. Cambridge University Press.

Koo, B.-H., C. Lee, H. B. Yilmaz, N. Farsad, A. Eckford, and C.-B. Chae (2016). Molecular mimo: From theory to prototype. *IEEE Journal on Selected Areas in Communications 34*(3), 600–614.

Lee, C., B. Koo, N.-R. Kim, B. Yilmaz, N. Farsad, A. Eckford, and C.-B. Chae (2015). Molecular mimo communication link. In *Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on*, pp. 13–14. IEEE.

Li, B., M. Sun, S. Wang, W. Guo, and C. Zhao (2016a). Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications. *IEEE Transactions on Communications 64*(5), 2079–2091.

Li, B., M. Sun, S. Wang, W. Guo, and C. Zhao (2016b). Low-complexity noncoherent signal detection for nanoscale molecular communications. *IEEE transactions on nanobioscience 15*(1), 3–10.

Li, B., C. Zhao, and W. Guo (2017). Non-linear signal detection for molecular communications. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6. IEEE.

Li, J., W. Zhang, X. Bao, M. Abbaszadeh, and W. Guo (2020). Inference in turbulent molecular information channels using support vector machine. *IEEE Trans. on Mol., Biol. and Multi-Scale Commun. 6*(1), 25–35.

Licina, D. et al. (2015). Human convective boundary layer and its interaction with room ventilation flow. *Indoor Air 25*(1), 21–35.

Lin, L., Z. Luo, L. Huang, C. Luo, Q. Wu, and H. Yan (2019). High-accuracy distance estimation for molecular communication systems via diffusion. *Nano Communication Networks 19*, 47–53.

Lin, L., Q. Wu, F. Liu, and H. Yan (2018). Mutual information and maximum achievable rate for mobile molecular communication systems. *IEEE transactions on nanobioscience 17*(4), 507–517.

Lin, L., Q. Wu, M. Ma, and H. Yan (2019). Concentration-based demodulation scheme for mobile receiver in molecular communication. *Nano Communication Networks*.

Liu, L., J. Wei, Y. Li, and A. Ooi (2017). Evaporation and dispersion of respiratory droplets from coughing. *Indoor Air 27*(1), 179–190.

Liu, S., Z. Wei, B. Li, W. Guo, and C. Zhao (2019). Metric combinations in non-coherent signal detection for molecular communication. *Nano Communication Networks 20*, 1–10.

Lugg, G. (1968). Diffusion coefficients of some organic and other vapors in air. *Analytical Chemistry 40*(7), 1072–1077.

Lyulin, Y. V., D. V. Feoktistov, I. Afanasev, E. Chachilo, O. A. Kabov, and G. V. Kuznetsov (2015). Measuring the rate of local evaporation from the liquid surface under the action of gas flow. *Technical Physics Letters 41*(7), 665–667.

Maham, B. and R. C. Kizilirmak (2018). Neuro-spike communications with multiple synapses under inter-neuron interference. *IEEE Access 6*, 39962–39968.

Mařík, J., M. Pexa, M. Kotek, V. Hönig, et al. (2014). Comparison of the effect of gasoline-ethanol e85-butanol on the performance and emission characteristics of the engine saab 9-5 2.3 l turbo. *Agronomy Research 12*(2), 359–366.

Marr, L. C. et al. (2019). Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J. of the Roy. Soc.*

Interface 16(150), 20180298.

Maxey, M. (1987). The gravitational settling of aerosol particles in homogeneous turbulence and random flow fields. *Journal of Fluid Mechanics 174*, 441–465.

McGuiness, D. T., S. Giannoukos, A. Marshall, and S. Taylor (2018). Parameter analysis in macro-scale molecular communications using advection-diffusion. *IEEE Access 6*, 46706–46717.

McGuiness, D. T., S. Giannoukos, S. Taylor, and A. Marshall (2019). Experimental and analytical analysis of macro-scale molecular communications within closed boundaries. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*.

Miller, M. B. and B. L. Bassler (2001). Quorum sensing in bacteria. *Annual Reviews in Microbiology 55*(1), 165–199.

Mittal, R., R. Ni, and J.-H. Seo (2020). The flow physics of covid-19. *Journal of fluid Mechanics 894*.

Mokeba, M., D. Salt, B. Lee, and M. Ford (1997). Simulating the dynamics of spray droplets in the atmosphere using ballistic and random-walk models combined. *Journal of Wind Engineering and Industrial Aerodynamics 67*, 923–933.

Moore, M., T. Nakano, A. Enomoto, and T. Suda (2010). Measuring distance with molecular communication feedback protocols. *Proc. ICST BIONETICS*, 1–13.

Moore, M. J. and T. Nakano (2012). Comparing transmission, propagation, and receiving options for nanomachines to measure distance by molecular communication. In *Communications (ICC), 2012 IEEE International Conference on*, pp. 6132–6136. IEEE.

Moore, M. J., T. Nakano, A. Enomoto, and T. Suda (2012). Measuring distance from single spike feedback signals in molecular communication. *IEEE Transactions on*

Signal Processing 60(7), 3576–3587.

Morton, B., G. I. Taylor, and J. S. Turner (1956). Turbulent gravitational convection from maintained and instantaneous sources. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 234*(1196), 1–23.

Mosayebi, R., A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari (2018). Type-based sign modulation and its application for isi mitigation in molecular communication. *IEEE Transactions on Communications 66*(1), 180–193.

Munson, B. R., D. F. Young, T. H. Okiishi, and W. W. Huebsch (2009a). *Fundamentals of fluid mechanics*. John Wiley & Sons.

Munson, B. R., D. F. Young, T. H. Okiishi, and W. W. Huebsch (2009b). *Fundamentals of fluid mechanics*. John Wiley & Sons, Inc.

Mustam, S. M., S. K. Syed Yusof, and S. Nejatian (2017). Multilayer diffusion-based molecular communication. *Transactions on Emerging Telecommunications Technologies 28*(1), e2935.

Nakano, T., Y. Okaie, and A. V. Vasilakos (2013). Transmission rate control for molecular communication among biological nanomachines. *IEEE Journal on Selected Areas in Communications 31*(12), 835–846.

Nicas, M., W. W. Nazaroff, and A. Hubbard (2005). Toward understanding the risk of secondary airborne infection: emission of respirable pathogens. *Journal of occupational and environmental hygiene 2*(3), 143–154.

Nilsson, J. W. and S. Riedel (2010). *Electric circuits*. Prentice Hall Press.

Noel, A., K. C. Cheung, and R. Schober (2014a). Bounds on distance estimation via diffusive molecular communication. In *Global Communications Conference (GLOBECOM), 2014 IEEE*, pp. 2813–2819. IEEE.

Noel, A., K. C. Cheung, and R. Schober (2014b). Optimal receiver design for diffusive molecular communication with flow and additive noise. *IEEE transactions on nanobioscience 13*(3), 350–362.

Noel, A., K. C. Cheung, and R. Schober (2015). Joint channel parameter estimation via diffusive molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications 1*(1), 4–17.

Oppenheim, A. V. (1999). *Discrete-time signal processing*. Pearson Education India.

Ozmen, M., E. Kennedy, J. Rose, P. Shakya, J. K. Rosenstein, and C. Rose (2018). High speed chemical vapor communication using photoionization detectors in turbulent flow. *IEEE Trans. on Mol., Biol. and Multi-Scale Commun. 4*(3), 160–170.

Papoulis, A. and S. U. Pillai (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Peiris, J. S., K. Y. Yuen, A. D. Osterhaus, and K. Stöhr (2003). The severe acute respiratory syndrome. *New England J. of Medicine 349*(25), 2431–2441.

Pendar, M.-R. and J. C. Páscoa (2020). Numerical modeling of the distribution of virus carrying saliva droplets during sneeze and cough. *Physics of Fluids 32*(8), 083305.

Picard, A., R. Davis, M. Gläser, and K. Fujii (2008). Revised formula for the density of moist air (cipm-2007). *Metrologia 45*(2), 149.

Pierobon, M. and I. F. Akyildiz (2010). A physical end-to-end model for molecular communication in nanonetworks. *IEEE Journal on Selected Areas in Communications 28*(4).

Pierobon, M. and I. F. Akyildiz (2011a). Diffusion-based noise analysis for molecular communication in nanonetworks. *IEEE Transactions on Signal Processing 59*(6), 2532–2547.

Pierobon, M. and I. F. Akyildiz (2011b). Noise analysis in ligand-binding reception for molecular communication in nanonetworks. *IEEE Transactions on Signal Processing 59*(9), 4168–4182.

Pierobon, M. and I. F. Akyildiz (2013). Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. *IEEE Transactions on Information Theory 59*(2), 942–954.

Prather, K. A., C. C. Wang, and R. T. Schooley (2020). Reducing transmission of sars-cov-2. *Science*.

PubChem Database, N. C. B. I. (2019). Ethanol, cid=702. Accessed: December 18, 2019. https://pubchem.ncbi.nlm.nih.gov/compound/Ethanol.

Qiu, S., W. Guo, B. Li, Y. Wu, X. Chu, S. Wang, and Y. Y. Dong (2015). Long range and long duration underwater localization using molecular messaging. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications 1*(4), 363–370.

Qiu, S., W. Guo, S. Wang, N. Farsad, and A. Eckford (2014). A molecular communication link for monitoring in confined environments. In *Communications Workshops (ICC), 2014 IEEE International Conference on*, pp. 718–723. IEEE.

Reuter, M., U. Boin, E. Verhoef, K. Heiskanen, Y. Yang, G. Georgalli, et al. (2005). *The metrics of material and metal ecology: harmonizing the resource, technology and environmental cycles*. Elsevier.

Rock, K. et al. (2014). Dynamics of infectious diseases. *Rep. on Prog. in Phys. 77*(2), 026602.

Rothe, P. and J. Block (1977). Aerodynamic behavior of liquid sprays. *International Journal of Multiphase Flow 3*(3), 263–272.

Sazhin, S., G. Feng, and M. Heikal (2001). A model for fuel spray penetration.

Fuel 80(15), 2171–2180.

Schaffer, F., M. Soergel, and D. Straube (1976). Survival of airborne influenza virus: effects of propagating host, relative humidity, and composition of spray fluids. *Archives of virology 51*(4), 263–273.

Seinfeld, J. H. and S. N. Pandis (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.

Seminara, G., B. Carli, G. Forni, S. Fuzzi, A. Mazzino, and A. Rinaldo (2020). Biological fluid dynamics of airborne covid-19 infection. *Rendiconti Lincei. Scienze fisiche e naturali*, 1–33.

Shaw, R. A. (2003). Particle-turbulence interactions in atmospheric clouds. *Annual Review of Fluid Mechanics 35*(1), 183–227.

Srinivas, K., A. W. Eckford, and R. S. Adve (2012). Molecular communication in fluid media: The additive inverse gaussian noise channel. *IEEE Transactions on Information Theory 58*(7), 4678–4692.

Stern, K. and M. K. McClintock (1998). Regulation of ovulation by human pheromones. *Nature 392*(6672), 177–179.

Stockie, J. M. (2011). The mathematics of atmospheric dispersion modeling. *Siam Review 53*(2), 349–372.

Tang, J. W., T. J. Liebner, B. A. Craven, and G. S. Settles (2009). A schlieren optical study of the human cough with and without wearing masks for aerosol infection control. *Journal of the Royal Society Interface 6*(suppl_6), S727–S736.

Tsay, W.-J., C. J. Huang, T.-T. Fu, and I.-L. Ho (2013). A simple closed-form approximation for the cumulative distribution function of the composite error of stochastic frontier models. *Journal of Productivity Analysis 39*(3), 259–269.

Unterweger, H., J. Kirchner, W. Wicket, A. Ahmadzadeh, D. Ahmed, V. Jamali, C. Alexiou, G. Fischer, and R. Schober (2018). Experimental molecular communication testbed based on magnetic nanoparticles in duct flow. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5. IEEE.

Vahdat, A. and D. Becker (2000). Epidemic routing for partially connected ad hoc networks.

Valderrama, M. J., F. Jimenez, R. Gutierrez, and A. Martinez-Almecija (1995). Estimation and filtering on a doubly stochastic poisson process. *Applied Stochastic Models in Business and Industry 11*(1), 13–24.

Vuorinen et.al, V. (2020). Modelling aerosol transport and virus exposure with numerical simulations in relation to sars-cov-2 transmission by inhalation indoors. *Safety Science 130*, 104866.

Vynnycky, E. and R. White (2010). *An introduction to infectious disease modelling*. OUP oxford.

Wang, B., H. Wu, and X.-F. Wan (2020). Transport and fate of human expiratory droplets-a modeling approach. *Phys. of Fluids 32*(8), 083307.

Wang, L., N. Farsad, W. Guo, S. Magierowski, and A. W. Eckford (2015). Molecular barcodes: Information transmission via persistent chemical tags. In *2015 IEEE International Conference on Communications (ICC)*, pp. 1097–1102. IEEE.

Wang, X., M. D. Higgins, and M. S. Leeson (2015a). An algorithmic distance estimation scheme for diffusion based molecular communication systems. In *Communications (ICC), 2015 IEEE International Conference on*, pp. 1134–1139. IEEE.

Wang, X., M. D. Higgins, and M. S. Leeson (2015b). Distance estimation schemes for diffusion based molecular communication systems. *IEEE Communications Letters 19*(3), 399–402.

Wei, J. and Y. Li (2015). Enhanced spread of expiratory droplets by turbulence in a cough jet. *Building and Environment 93*, 86–96.

Wei, J. and Y. Li (2017). Human cough as a two-stage jet and its role in particle transport. *PloS one 12*(1), e0169235.

Weisstein, E. W. (2003). Circle-circle intersection. Accessed: 2020-06-15.

Xie, X., Y. Li, A. Chwang, P. Ho, and W. Seto (2007). How far droplets can move in indoor environments-revisiting the wells evaporation-falling curve. *Indoor air 17*(3), 211–225.

Yilmaz, H. B., A. C. Heren, T. Tugcu, and C.-B. Chae (2014). Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Communications Letters 18*(6), 929 – 932.

Young, J. W. (1993). Head and face anthropometry of adult us civilians. Technical report, Federal Aviation Administration Washington DC Office Of Aviation Medicine.

Zhai, H., Q. Liu, A. V. Vasilakos, and K. Yang (2018). Anti-isi demodulation scheme and its experiment-based evaluation for diffusion-based molecular communication. *IEEE transactions on nanobioscience 17*(2), 126–133.

Zhou, P. et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature 579*(7798), 270–273.

Zhu, S., S. Kato, and J.-H. Yang (2006). Study on transport characteristics of saliva droplets produced by coughing in a calm indoor environment. *Building and environment 41*(12), 1691–1702.

Ziff, R. M., S. N. Majumdar, and A. Comtet (2009). Capture of particles undergoing discrete random walks. *The Journal of chemical physics 130*(20), 204104.

VITA

EDUCATION

2021, Doctor of Philosophy in Electronics and Communications Engineering
Graduate School of Engineering and Sciences, Izmir Institute of Technology
2015, Master of Science in Electrical and Electronics Engineering
Graduate School of Natural and Applied Sciences, Gazi University
2007, Bachelor of Science in Electrical and Electronics Engineering
Faculty of Engineering, Gazi University

PUBLICATIONS

Gulec, F. and B. Atakan (2021). A molecular communication perspective on airborne pathogen transmission and reception via droplets generated by coughing and sneezing. *to appear in IEEE Transactions on Molecular, Biological and Multi-Scale Communications*.

Gulec, F. and B. Atakan (2021). Fluid dynamics-based distance estimation algorithm for macroscale molecular communication. *Nano Communication Networks 28*, 100351.

Gulec, F. and B. Atakan (2021). A droplet-based signal reconstruction approach to channel modeling in molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 7(1), 64-68.

Gulec, F. and B. Atakan (2020). Mobile human ad hoc networks: A communication engineering viewpoint on interhuman airborne pathogen transmission. *arXiv preprint arXiv:2011.00884, submitted to IEEE Network*.

Gulec, F. and B. Atakan (2020). Distance estimation methods for a practical macroscale molecular communication system. *Nano Communication Networks 24*, 100300.

Gulec, F. and B. Atakan (2020). Localization of a passive molecular transmitter with a sensor network. In *Int. Conf. on Bio-inspired Inf. and Commun.*, pp. 317-335. Springer.

Atakan, B. and F. Gulec (2019). Signal reconstruction in diffusion-based molecular communication. *Transactions on Emerging Telecommun. Tech.* 30(12), e3699.