



HAL
open science

Advanced Detection and Synchronization Schemes for Molecular Communications

Xuewen Qian

► **To cite this version:**

Xuewen Qian. Advanced Detection and Synchronization Schemes for Molecular Communications. Signal and Image processing. Université Paris-Saclay, 2020. English. NNT : 2020UPASG042 . tel-03251316

HAL Id: tel-03251316

<https://theses.hal.science/tel-03251316>

Submitted on 7 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Advanced Detection and Synchronization Schemes for Molecular Communications

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580 Sciences et Technologies de
l'Information et de la Communication (STIC)
Spécialité de doctorat: Réseaux, Information et Communications
Unité de recherche: Université Paris-Saclay, CNRS, CentraleSupélec,
Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
Référent: Faculté des Sciences d'Orsay

**Thèse présentée et soutenue en visioconférence totale,
le 16 Décembre 2020, par**

Xuewen QIAN

Composition du jury:

Jalel Ben Othman Professeur, Université Paris 13	Président
Valeria Loscri Chargée de recherche (HDR), Inria Lille-Nord Europe	Rapportrice & Examinatrice
Lin Chen Professeur, Sun Yat-sen University, China	Rapporteur & Examineur
Yansha Deng Maître de conférences, King's College London, UK	Examinatrice
Trung Q. Duong Professeur, Queen's University Belfast, UK	Examineur
Marco Di Renzo Directeur de recherche, CNRS	Directeur de thèse

To my family...

Abstract

Today modern communication networks are inherently based on the laws/properties of electromagnetism. A plethora of important applications exist, however, in which it is difficult or expensive to use traditional electromagnetic waves. This includes emerging applications, e.g., mining industries, smart cities, nano-robotics for industrial and nano-medicine applications, remote healthcare provisioning, etc., which are expected to play a fundamental role towards improving the quality of life and the safety of millions of people.

The challenging and ambitious objective is to re-think the way communication networks are designed and engineered for application to the aforementioned scenarios. The breakthrough innovation lies in enabling communication by replacing electromagnetic signals with chemical signals. Chemical signals are, in fact, inherently suitable for the above-mentioned applications, where radio signals fail to be reliable. They have, in addition, the desirable bonus of being environmentally-friendly and bio-compatible/degradable. Facilitated by the development of the biological techniques, these chemical signals are basically stimulated by releasing information-carried molecules, which gives this kind of communication mechanism the name, molecular communication. To pave the way to realize the reliable molecular communication systems, the fundamental problem must be solved under some realistic assumptions. Thus, in this thesis, we focus on the receiver design for realistic scenarios in molecular communications, i.e., without the explicit knowledge of channel information. In particular, we concentrate our attention on detection and synchronization scheme design based on either pilot symbols or no pilot symbol at all (only received signals).

First, we propose the artificial neural networks based detectors and the detection performance improvement is witnessed and analyzed compared with the conventional multi-memory-bit-assisted threshold-based detectors, which in turn inspires the design of the optimal multi-memory-bit-assisted threshold-based detectors as well as the performance evaluation frameworks for the threshold-based detectors in molecular communications.

Then, we reformulate the multi-memory-bit-assisted threshold whose key parameters are shown to be able to extracted explicitly only from the received signals. Afterwards, with some illustrations, we depict the detailed steps to retrieve these parameters from the received signals and reveal that by constructing multi-dimensional data, these parameters can be obtained with less errors. In particular, with the predefined initial centroids, we propose two

approaches that apply K-Means clustering algorithm to the constructed data to obtain these needed parameters. The first approach uses the maximum received signal to construct the initial centroids while the second approach works in an iterative manner that can give some initial centroids that are close to the theoretical ones, thus provides better performance.

Third, in order to find the optimal start point which may yield the lowest error probability, we incorporate the start point into the error probability in order to obtain the theoretical optimal start point. By approximating the error probability in large signal-noise-ratio regime, we are able to obtain the analytical condition the asymptotic optimal start point follows. Based on this condition, a data-aided synchronization algorithm is proposed as well as its theoretical upper bounds for the absolute mean and variance of the synchronization errors. In addition, a blind algorithm based on the K-Means clustering is proposed inspired by the data-aided scheme. Besides, the Cramer-Rao lower bounds for the data-aided algorithm are obtained to compare the synchronization performance.

Key words: Molecular communications, artificial neural networks, detector, K-Means, synchronization, Cramer-Rao lower bound.

Résumé

Aujourd'hui, les réseaux de communication modernes sont intrinsèquement basés sur les lois / propriétés de l'électromagnétisme. Il existe cependant une pléthore d'applications importantes dans lesquelles il est difficile ou coûteux d'utiliser les ondes électromagnétiques traditionnelles. Cela inclut les applications émergentes, par exemple les industries minières, les villes intelligentes, la nano-robotique pour les applications industrielles et de nano-médecine, la fourniture de soins de santé à distance, etc., qui devraient jouer un rôle fondamental dans l'amélioration de la qualité de vie et de la sécurité de millions de personnes.

L'objectif ambitieux est de repenser la façon dont les réseaux de communication sont conçus et conçus pour être appliqués aux scénarios susmentionnés. L'innovation de rupture consiste à permettre la communication en remplaçant les signaux électromagnétiques par des signaux chimiques. Les signaux chimiques sont, en fait, intrinsèquement appropriés pour les applications mentionnées ci-dessus, où les signaux radio ne sont pas fiables. Ils ont, en outre, l'avantage souhaitable d'être écologiques et biocompatibles / dégradables. Facilités par le développement des techniques biologiques, ces signaux chimiques sont essentiellement stimulés par la libération de molécules porteuses d'informations, ce qui donne à ce type de mécanisme de communication le nom de communication moléculaire. Pour ouvrir la voie à la réalisation de systèmes de communication moléculaire fiables, le problème fondamental doit être résolu sous certaines hypothèses réalistes. Ainsi, dans cette thèse, nous nous concentrons sur la conception du récepteur pour des scénarios réalistes dans les communications moléculaires, c'est-à-dire sans la connaissance explicite des informations de canal. En particulier, nous concentrons notre attention sur la conception de schémas de détection et de synchronisation basés soit sur des symboles pilotes, soit sur aucun symbole pilote (uniquement les signaux reçus).

Tout d'abord, nous proposons les détecteurs basés sur les réseaux de neurones artificiels et l'amélioration des performances de détection est observée et analysée par rapport aux détecteurs conventionnels basés sur des seuils assistés par bits multi-mémoire, qui à leur tour inspirent la conception du bit multi-mémoire optimal. détecteurs assistés basés sur des seuils ainsi que les cadres d'évaluation des performances des détecteurs basés sur des seuils dans les communications moléculaires.

Ensuite, nous reformulons le seuil assisté par bits multi-mémoire dont il est démontré que les

paramètres clés ne peuvent être extraits explicitement que des signaux reçus. Ensuite, avec quelques illustrations, nous décrivons les étapes détaillées pour récupérer ces paramètres à partir des signaux reçus et révélons qu'en construisant des données multidimensionnelles, ces paramètres peuvent être obtenus avec moins d'erreurs. En particulier, avec les centroïdes initiaux prédéfinis, nous proposons deux approches qui appliquent l'algorithme de clustering K-Means aux données construites pour obtenir ces paramètres nécessaires. La première approche utilise le signal reçu maximum pour construire les centroïdes initiaux tandis que la seconde approche fonctionne de manière itérative qui peut donner des centroïdes initiaux proches des centres théoriques, offrant ainsi de meilleures performances.

Troisièmement, afin de trouver le point de départ optimal qui peut produire la probabilité d'erreur la plus faible, nous incorporons le point de départ dans la probabilité d'erreur afin d'obtenir le point de départ optimal théorique. En approchant la probabilité d'erreur dans un régime de rapport signal-bruit élevé, nous pouvons obtenir la condition analytique suivie par le point de départ optimal asymptotique. Sur la base de cette condition, un algorithme de synchronisation assistée par données est proposé ainsi que ses limites supérieures théoriques pour la moyenne et la variance absolues des erreurs de synchronisation. De plus, un algorithme aveugle basé sur le clustering K-Means est proposé inspiré du schéma assisté par les données. En outre, les limites inférieures de Cramer-Rao pour l'algorithme assisté par données sont obtenues pour comparer les performances de synchronisation.

Mots clés: Communications moléculaires, réseaux neuronaux artificiels, détecteur, K-Means, synchronisation, borne inférieure de Crammer-Rao.

Contents

Abstract (English/Français)	i
Acronyms	ix
Notation	xi
List of figures	xiii
List of tables	xvii
1 Introduction	1
1.1 What is Molecular Communications	3
1.2 Background	3
1.2.1 Modelling the Realistic MC Systems	4
1.2.2 MC System Design	5
1.2.3 Simulation and Experiment Verifications	6
1.3 Thesis Overview and Major Contributions	7
1.4 Publications	9
2 Fundamentals of MC Systems	15
2.1 Channel Modelling	17
2.1.1 Diffusion-based Propagation	17
2.1.2 Flow-aided Propagation	21
2.1.3 Bacteria-assisted Propagation	22
2.1.4 Motor-Protein-based Propagation	23
2.1.5 Received Signal and Noise Modelling	23
2.2 Modulation Schemes	25
2.2.1 Basic Modulation	26
2.2.2 Composite Modulation Scheme	27
2.3 Machine Learning for Molecular Communication Systems	28
2.3.1 Feedforward Neural Networks	29
2.3.2 K-Means Clustering Algorithm	33
2.4 Detection	35
2.4.1 Basics of the Detection	35
2.4.2 Single-Sample/Multiple-Sample Detectors	35

Contents

2.4.3	Single-Symbol/Multiple-Symbol Detectors	36
2.4.4	Coherent/Non-Coherent Detectors	37
2.5	Synchronization	38
2.5.1	Synchronization with Channel Information	38
2.5.2	Synchronization without Channel Information	40
3	Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization	41
3.1	Introduction	43
3.2	System model	45
3.3	Model-Based Receivers Design in Molecular Communications	48
3.3.1	Optimal Zero-Bit Memory Receiver	48
3.3.2	Optimal One-Bit Memory Receiver	51
3.3.3	Optimal K -Bit Memory Receiver	52
3.4	Data-Driven Receiver Design in Molecular Communications	53
3.4.1	Data-Driven Design of Zero-Bit Memory Receiver	54
3.4.2	Data-Driven Design of One-Bit Memory Receiver	55
3.4.3	Data-Driven Design of K -Bit Memory Receiver	56
3.5	Numerical Results	56
3.5.1	Zero-Bit Memory Receivers	58
3.5.2	One-Bit Memory Receivers	60
3.5.3	K -Bit Memory Receivers	60
3.6	Conclusion	62
3.7	Appendix	63
3.7.1	Proof of the BER of the one-bit memory receiver	63
3.7.2	Proof of the BER of the multi-bit memory receiver	64
4	K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications	67
4.1	Introduction	69
4.1.1	Motivation	69
4.1.2	Related Works	70
4.1.3	Contributions	70
4.1.4	Chapter Organization	71
4.2	System model	71
4.3	Multi-memory-bit Threshold Reformulation	72
4.3.1	Estimation of $\hat{r}_i _{s_i=j, 0 \leq j \leq \mathcal{L}}$ from the Received Data	75
4.4	Clustering-based Non-coherent Detection	77
4.4.1	K -Means Clustering Algorithm	77
4.4.2	Single-Dimensional Clustering: Challenges and Limitations	78
4.4.3	Multi-Dimensional Clustering	79
4.4.4	Setup of the Initial Centroids	80
4.4.5	Direct Clustering-Based Inference and Clustering-plus-Threshold Detection	82

4.5	Iterative Algorithm for Computing the Initial Centroids	86
4.6	Numerical Results	89
4.7	Conclusion	90
5	Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems	91
5.1	Introduction	93
5.2	System Model	95
5.3	The Optimal Start Point	97
5.3.1	Formulation of $P_e(\rho, \tau)$	97
5.3.2	Asymptotic Optimal Start Point	100
5.4	The Proposed Synchronization Schemes	102
5.4.1	Data-Aided Synchronization Algorithm	103
5.4.2	K-Means-based Blind Synchronization Algorithm	109
5.5	The Cramer-Rao Lower Bounds with known and unknown CSI	110
5.6	Simulation Results	114
5.7	Conclusions	116
5.8	Appendix	118
6	Conclusions and Future Work	121
6.1	Conclusion	122
6.2	Future Work	123
6.2.1	Machine-Learning-Aided Detection	123
6.2.2	Non-Coherent Detection	124
6.2.3	Synchronization	125
	Reference	127

Acronyms

AI	Artificial Intelligence
ASK	Amplitude Shift Keying
AcCoRD	Actor-based Communication via Reaction-Diffusion
ATP	Adenosine Triphosphate
ANN	Artificial Neural Network
BER	Bit-Error-Ratio
CE	Channel Estimation
CSI	Channel State Information
CSK	Concentration Shift Keying
CLT	Central Limit Theorem
CRLB	Crammer-Rao Lower Bound
DL	Deep Learning
DNN	Deep Neural Network
EM	Eletromagnetic
ETM	Emission Time-based Modulation
FSK	Frequency Shift Keying
ISI	Inter-Symbol-Interference
mmWave	Milimeter Wave
MC	Molecular Communications
MT	Microtubule
ML	Machine Learning

Acronyms

MIMO	Multiple Input Multiple Output
MU	Multi-User
MoSK	Molecular Shift Keying
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
OOK	On-Off Keying
PDE	Phase Differential Equation
PSK	Phase Shift Keying
PPM	Pulse Phase Modulation
PDF	Probability Distribution Function
PMF	Probability Mass Function
RF	Radio Frequency
RNN	Recurrent Neural Network
ReLU	Rectified Linear Unit
SSK	Space Shift Keying
SNR	Signal-to-Noise Ratio
SIMO	Single Input Multiple Output
THz	Terahertz

Notation

The following notation is used throughout this thesis.

\mathbf{x}	a vector
\mathbf{x}_n	the n th vector
$\ \mathbf{x}\ $	the norm of \mathbf{x}
\mathbf{X}	a matrix
\mathbf{X}_n	the n th matrix
s_i	the i th symbol
$r_i(m)$	the received signal in the m th sample of the i th symbol slot
$\bar{r}_i(m)$	the average received signal in the m th sample of the i th symbol slot
$\bar{r}_i(m, \rho, s_{i+1}, \dots, s_{i-L})$	the average received signal in the m th sample of the i th symbol slot given the start point ρ and the transmitted bits $[s_{i+1}, \dots, s_{i-L}]$
r_i	the received signal in the i th symbol slot
$\mathcal{B}(n, p)$	Binomial random variable for n trials and success probability p
$\mathcal{N}(\mu, \sigma^2)$	Normal random variable with mean μ and variance σ^2
$\text{Poisson}(\lambda)$	Poisson random variable with mean λ
$P(A B), \Pr(A B)$	the probability of event A conditioned upon B
∇	gradient of a function ($\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}]^T$)
∇^2	Laplace operator ($\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$)
$\delta(\cdot)$	Dirac delta function
$\exp\{\cdot\}$	the exponential function
$\text{erf}\{\cdot\}$	the error function
$\text{erfc}\{\cdot\}$	the complementary error function
$\ln(\cdot)$	the natural logarithm function
$Q(\cdot, \cdot)$	the incomplete Gamma function
$\max\{\cdot, \cdot\}$	the maximum function
$\text{diag}\{\mathbf{x}\}$	diagonal matrix with the \mathbf{x} on its main diagonal

List of Figures

1.1	Experimental verifications	8
2.1	Diffusion-based propagation	18
2.2	Channel response comparison ($D = 4.265 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500\text{nm}$ and $N_{TX} = 2000$)	20
2.3	Flow-aided propagation	21
2.4	Bacteria-assisted propagation	22
2.5	Motor-protein-based propagation	23
2.6	Three basic modulation schemes and the received signal comparison (the transmitted symbols are [1, 1, 0, 1]) in MC systems. ($D = 4.265 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500\text{nm}$, $N_{TX} = 100,000$, $\Delta T = 5\text{us}$ and $T = 100\Delta T$ [1])	26
2.7	The composite modulation scheme and the illustration of the received signals. The transmitted symbols are [1, 1, 0, 1]. ($D = 4.365 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500 \text{ nm}$, $N_{TX} = 100,000$, $\Delta T = 5 \text{ us}$ and $T = 100\Delta T$)	28
2.8	Typical structure of a feed-forward NN with fully-connected layers.	29
2.9	Basic structure of a neuron	30
2.10	A non-linear function $g(x)$	30
2.11	Plot of $x_2(1)$, $x_2(2)$ and $x_2(3)$	31
2.12	Typical behaviors of the training and testing processes	33
2.13	Illustration of the K-Means clustering (The magenta circles denote the initial centroids for current iteration)	34
3.1	Block diagram of a typical MC system.	45
3.2	The 3D unbounded molecular channel model without flow including a point transmitter and a spherical absorbing receiver.	46
3.3	BER in (3.14) as a function of τ (the SNR is 30 dB).	50
3.4	Typical structure of a feed-forward ANN with fully-connected layers.	54
3.5	Data-driven one-bit memory receiver.	55
3.6	Data-driven K -bit memory receiver.	56
3.7	BER of optimal vs. conventional zero-bit memory receiver - $T = 30\Delta T$	57
3.8	Approximated distributions of the received bits from (3.11) (the SNR is 25 dB).	57
3.9	Empirical distributions of the received bits (the SNR is 25 dB).	57

List of Figures

3.10	The ANN-based receiver achieves the same BER performance as the optimal zero-bit receiver - $T = 30\Delta T$	58
3.11	Optimal detection thresholds of model-based and data-driven schemes. The ANN-based receiver automatically emulates the optimal zero-bit memory receiver. - $T = 30\Delta T$	59
3.12	System model assuming an observing receiver. BER of the optimal vs. sub-optimal (i.e., based on the sub-optimal threshold) zero-bit memory receiver - $T = 30\Delta T$	59
3.13	System model assuming an observing receiver. The data-driven receiver achieves the same BER performance as the optimal (i.e., based on the optimal threshold) zero-bit receiver - $T = 30\Delta T$	60
3.14	BER of optimal vs. sub-optimal one-bit memory receiver. The performance gap between optimal and sub-optimal receiver gets smaller since the ISI is modeled more accurately - $T = 30\Delta T$	61
3.15	The ANN-based receiver achieves the same BER performance as the optimal one-bit receiver - $T = 30\Delta T$	61
3.16	Two-bit memory detector: Comparison between the optimal and sub-optimal setups of the demodulation thresholds. The more the number of memory bits, the better the ISI is modeled. As the memory length approaches the channel length, thus, the optimal threshold converges towards the conventional threshold - $T = 30\Delta T$	62
3.17	The ANN-based two-bit memory receiver achieves the same BER performance as the optimal two-bit receiver - $T = 30\Delta T$	62
3.18	The BER performance of ANN-based and model-based L-bit memory detectors overlap with each other - $T = 30\Delta T$	63
4.1	The positions of $\bar{r}_i _{s_i, s_{i-1}}$ (calculated via (4.26)) and $\hat{r}_i _{s_i, s_{i-1}}$ (corresponding to the centroids calculated from (4.24) and (4.25) and assuming that the labels $[s_i, s_{i-1}]$ are known) for $T = 30\Delta T$	79
4.2	Theoretical and estimated centroids obtained from clustering the data vector $[r_n, r_{n-1}]$ ($T = 30\Delta T$)	81
4.3	The customized initial centroids and the corresponding clusters ($T = 30\Delta T$)	82
4.4	Distribution of $[r_n, r_{n-1}] _{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1$ and $T = 30\Delta T$).	84
4.5	Distribution of $[r_n, r_{n-1}] _{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$ and $T = 30\Delta T$).	84
4.6	Distribution of $[r_n, r_{n-1}] _{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1$, $T = 20\Delta T$).	85
4.7	Distribution of $[r_n, r_{n-1}] _{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$, $T = 20\Delta T$).	85
4.8	Distribution of $[r_n, r_{n-1}] _{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$, $T = 20\Delta T$).	88
4.9	BER comparison of the proposed algorithms (mild ISI)	89
4.10	BER comparison of the proposed algorithms (severe ISI)	90

5.1	The average received molecule numbers at each time instant due to $s_{i+1} = 1$, $s_i = 1$ and $s_{i-1} = 1$. If $\rho > 0$, $\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})$ should take s_{i+1} into account.	96
5.2	Comparison of $P_e(\rho, \tau)$ (shaded area) for ρ in and out of the range $[\rho_0, \rho_1]$.	99
5.3	$P_e(\rho, \tau)$ as a function of ρ and τ (SNR is 25 dB)	100
5.4	Comparison of $\frac{E[F(\rho)]}{K/2}, \frac{F(\rho)}{K/2}$ and $C_0(\rho)$ (SNR is 30 dB and $K = 128$)	105
5.5	Comparison of $E[F(\hat{\rho}_s^*)]$ and $E[F(\rho_s^*)]$ (SNR is 30 dB and $K = 128$)	107
5.6	Comparison of $E[G(\rho)]$ and $C_0(\rho)$ when SNR is 30 dB	110
5.7	The absolute mean synchronization error comparison $ E_s(\hat{\rho}_s^* - \rho^*) $ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 64$).	115
5.8	The synchronization variance comparison $E \hat{\rho}_s^* - \rho^* ^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 64$).	115
5.9	The absolute mean synchronization error comparison $ E_s(\hat{\rho}_s^* - \rho^*) $ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 128$).	116
5.10	The synchronization variance comparison $E \hat{\rho}_s^* - \rho^* ^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 128$).	116
5.11	The absolute mean synchronization error comparison $ E_s(\hat{\rho}_s^* - \rho^*) $ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 256$).	117
5.12	The synchronization variance comparison $E \hat{\rho}_s^* - \rho^* ^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 256$).	117

List of Tables

3.1	Simulation parameters	48
4.1	Simulation parameters	89
4.2	Computational complexity	89
5.1	Simulation parameters	114

1 Introduction

Chapter 1. Introduction

This chapter begins with *Section 1.1* which introduces the generic definition of the molecular communication. *Section 1.2* provides the general background introduction of the three main tasks in the molecular communications. *Section 1.3* highlights the major contributions in this thesis work and the organization of the thesis. *Section 1.4* provides lists of publications produced during my Ph.D. candidature.

1.1	What is Molecular Communications	3
1.2	Background	3
1.2.1	Modelling the Realistic MC Systems	4
1.2.2	MC System Design	5
1.2.3	Simulation and Experiment Verifications	6
1.3	Thesis Overview and Major Contributions	7
1.4	Publications	9

1.1 What is Molecular Communications

Disseminating information from one place to another place has been investigated and deployed for hundreds of years and the modern electromagnetic (EM) wave [2] based technology has enabled a wide range of communication scenarios since Marconi invented the modern telecommunication system. The communication distance of the modern telecommunication systems spans a wide range from thousands of kilometers to several meters while communications within the microscale remain undeveloped deeply. Although existing techniques, e.g., optical communications [3] and the terahertz (THz) techniques [4] can be applied to the microscale communications, these techniques will suffer from the path loss [5, 6] due to the conductive medium and the size of the components is non-negligible, forcing researchers to find another way to implement the communication systems in microscale.

The potential applications that necessitate the microscale communications include biosensing [5], nano-networks or nano-machines [7] etc. The scientists in biology have been working on figuring out how to fabricate nano-machines [7] in order to invent some robots that can be injected into blood vessels, identify the target organs and deliver specified drugs [8] to the organs. Meanwhile, environmental control and better understanding the biology are among the ultimate objectives of the microscale communications.

In nature, there are countless number of cases of communications that are not based on the EM wave. A typical example is in our bodies, e.g., dopamine molecules excite our nerve cells to make us feel happy after we hear some jokes. In this regard, the dopamine molecules convey the signals of molecular communications. The mechanism that the information is transferred by some molecules is referred to as molecular communications (MC) [9]. From the definition, the MC is not restricted to the microscale communication, it can be employed in the macroscale communications that conventional EM-based techniques cannot play its role in.

1.2 Background

The concept of the MC was first introduced in [10]. Since then, the MC has been a hot topic that attracts considerably many scientists. They mainly focus on three issues: modelling the realistic MC examples with the simplified components (the transmitters, the receivers and the channels), designing the communication system based on the simplified components and implementing the corresponding experiment verifications.

1.2.1 Modelling the Realistic MC Systems

As aforementioned, there exist many MC systems in nature. Currently, one of scientists' aims is to characterize the transmitter, the receiver [11], the channel and the molecules in reality in order to design and optimize the communication systems by applying the existing techniques in the wireless communication systems to the MC systems.

Concerning the transmitter where the molecules are produced and stored, many factors need to be taken into account. For instance, the production process needs the extra energy and some fundamental ingredients which are associated with the resource management [12]. In addition, the number of the transmitted molecules may not be stable during the transmission and the molecules may not be released in a short time. The effect of the unstable release time would impose some negative effects on [13] the signal synchronization etc.

It is even more complicated at the receiver side, since the mechanisms of receiving molecules are diverse and are associated with the type of the molecules. There exist two types of receivers, i.e., the absorbing receiver [14] and the passive receiver [15]. The absorbing receiver takes in the transmitted molecules into the cells and takes a record of the number of received molecules in each time slot or record the time when the number of received molecules reaching to a certain value. No molecule may be let out of the absorbing receivers. The passive receiver does the same task while the molecules may propagate out of the passive receiver, which is the main difference between the two types of receiver. Meanwhile, the mechanism of taking molecules is related to the molecules reaching to the receiver. For example, some molecules may be hold in the bubbles and the cell membrane of the receiver may directly merge the bubbles as if there is no barrier to the molecules. While for some types of molecules, they are taken into the receiver by some protein structure named receptors which can be regarded as a tunnel. Once a tunnel is occupied, other molecules cannot go through it [16]. Apart from the mechanism of collecting molecules, the shape [17] and the size of receivers have slight effects on the modelling.

In addition to the transmitter and the receiver, the channel that the molecules pass through is essential to the MC systems. The channel determines the capacity of the MC systems which in turn restricts the techniques borrowed from the wireless communication domain. It's worth noting that not only the molecules can propagate through the channel, some carriers (bubbles or bacteria [9]) that load the information molecules can go through the channel and disseminate signals, have effects on the channel modelling. There are several types of molecule/loader propagations: 1) diffusion based propagation [18, 19], 2) flow assisted propagation [20], 3) active transport using molecular motors and cytoskeletal filaments [21], 4) bacterial assisted propagation [22, 23], and 5) kinesin molecular motors moving over immobilized microtubule (MT) tracks [24].

Some mathematical models (channel response) can be derived for the above stated propagations [14]. Take the diffusion-based propagation for instance. Although the inter-symbol-interference (ISI) is inevitable as the molecules propagate via Brownian motion and not all molecules reach to the receiver within a given period, the channel response can be obtained analytically by solving a partial differential equation (PDE) [14]. To mitigate the ISI, enzymes [25] are exploited that can react with and decompose the molecules to reshape the channel response into the desired form. The channel response can be obtained analytically by solving a PDE [26]. While for some propagation types, the channel response can only be acquired via numerical simulations, or even worse, we can only resort to machine learning (ML) techniques [27] to model the input and output relations. For instance, when applying the multiple-input-multiple-output (MIMO) or multi-user (MU) [28] techniques to the diffusion-based MC systems that has the close-form channel response, it would be impossible to provide an analytical formula for the channel response [28]. We refer readers to a tutorial [29] for more details on the channel modelling.

1.2.2 MC System Design

The objective of the researchers is to design the MC systems based on the simplified models, including the transmitters, receivers and channels such that the reliable transmission can proceed in the natural scenarios. Considering the characteristics of the MC channels, i.e., the inevitable ISI and the signal-dependent noise, the systematic performance is determined by the modulation scheme [30, 31, 22], the detection scheme [16, 32, 33, 34, 35, 36], synchronization [37, 13, 38] and the channel estimation (CE) [39] etc. As stated in the above subsection, the performance needs to be evaluated and optimized. In the following text, we will introduce the modulation scheme, detection techniques, performance evaluations and the synchronization in the MC systems.

First, the modulation scheme is fundamental for the following tasks, e.g., encoding/decoding and the CE since it decides what kind of complete solutions in wireless communications can be employed in the MC system. In particular, the information can be modulated based on the following properties: the number of the released molecules [30], the type of the released molecules [22] and the release time [31] which are similar to the amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK) in radio frequency (RF) systems [2] and are referred to as concentration shift keying (CSK) [30], molecular shift keying (MoSK) [40] and emission time-based modulation (ETM) [31] or pulse-phase modulation (PPM). These modulation schemes can be mixed in order to further mitigate the effect of the inevitable ISI.

Once the molecules are released at the transmitter, the receiver would record the number of received molecules at each time instant. The receiver need to obtain the start point of

the received signals in order to perform the detection etc. Usually, the transmitter sends a segment of pilot symbols [41], i.e., preamble, and the receiver can rely on the preamble to synchronize signals which is referred to as data-aided synchronization. On the other hand, the synchronization without knowing the transmitted symbols (blind synchronization) [42] is necessary, since a long preamble is a waste of energy to the transmitter. Moreover, the aforementioned schemes rely on the complete knowledge of channel state information (CSI) [13] which is not always known in advance. Specifically, before the link between the transmitter and the receiver is established, neither the channel response is known, nor the channel can be estimated since the CE is based on the synchronized signals. Observing the channel responses of the common MC channels, the main feature is that the number of received signals will first go up and then descend gradually after the transmitter release the molecules. Based on this feature, synchronization can be performed without preamble and CSI. In chapter 5, we will present our work on synchronization without CSI and show the start point is optimal.

After the modulation scheme is decided and the synchronization is done, at the receiver side, the detection is performed to convert the received signals (the number of received molecules in each slot) into the estimated symbols. Concerning the signal-dependent noise in the MC channels, the detection schemes in the scattering light communications [3] can be employed in the MC systems since the received signals can be modelled by using Poisson channel. For instance, for the MC system using binary CSK, a binary one is detected if the received signal is greater than the threshold [43]. The optimal detection performance can be achieved if the error probability associated with the threshold is optimized [1]. Likewise, detection using other modulation schemes can be optimized by the same means. Besides, the auxiliary information [44], e.g., the previously detected symbols, can be exploited to optimized the threshold. Whilst, detection without CSI is more realistic which can be fulfilled by deep learning (DL). From the perspective of DL [27], the detection is to associate the received signals with certain outputs (estimated symbols). However, the detection performance of the DL-based detector is unclear. One may wonder, to what extent, the DL-based detectors are good? Can the DL-based detectors be optimized? In terms of the detection performance, does the DL-based detector achieve the same performance as existing detectors or does the DL-based detectors motivate the better detector design? These questions will be answered in chapter 3. In addition, the training of the DL-based detectors will waste a large amount of energy which is unacceptable in some scenarios. This necessitates the non-coherent detection, i.e., detection without CSI [33, 35]. In chapter 4, the practical non-coherent detector design is introduced.

1.2.3 Simulation and Experiment Verifications

In order to identify and bridge the gap between the theoretical MC systems and the realistic MC systems, the first action to take is the molecule-level simulation. As a fundamental step, the molecule-level simulation can provide prior information on the difficulties of experiments. Among all the molecule-level simulation platforms, the Actor-Based Communication

via Reaction-Diffusion (AcCoRD) [45] has the versatile capabilities of modelling the realistic transmitting and receiving processes. Specifically, AcCoRD can accommodate reactions such as molecule degradation, enzyme kinetics and the reception process like reversible or irreversible surface binding, ligand-receptor binding, transitions across boundary membranes etc. Apart from the AcCoRD, N3Sim [46] is a simulator for nano-networks using diffusion-based molecular communication which supports modelling MC systems with several transmitter, receiver and harvester nodes. N3Sim is believed to be capable of designing and validating novel protocols for future nano-networks. In [47], a custom end-to-end MolecUlar CommunicatIoN (MUCIN) simulator for MCvD systems is presented, which especially designed for investigating the first hitting process for the signal reception and supports 1-D to 3-D channels, sending consecutive symbols, imperfect molecule reception, extendable modulation, and filtering modules. we refer readers to [48, 49] for other MC simulators.

Currently, the experiments mainly are implemented in macroscale. The first MC testbed was based on spraying alcohol into free space [50] which, to some extent, is similar to the diffusion-based MC systems with flows. Furthermore, the platform based on the acids and bases in vessels was proposed [51] where the DL-based detectors are exploited and the corresponding performances are evaluated. The two testbeds were then extended to the MIMO systems in order to verify the performances of the MIMO MC system. In [52], researchers in the Friedrich-Alexander University (FAU) Erlangen-Nurnberg developed a MC experimental system using the magnetic nano-particles, aiming at imitating the MC in the blood vessels. Besides the aforementioned testbeds with the generic functionality, in [53], a tabletop MC platform was built to verify the effectiveness of the anti-ISI demodulation scheme for the concentration-encoded diffusion-based MC systems. In [54], an underwater prototype of the 4×4 MIMO MC systems was implemented to evaluate the performance of the space shift keying (SSK). Unlike the previously stated platforms that are easy to control, the promising biological signal conversion interface [55] was proposed in order to implement the microscale MC experiments by translating light signals into chemical signals in microscale. The realistic experiments are ongoing with more advanced devices.

1.3 Thesis Overview and Major Contributions

Based on the existing simplified component models and the most common system settings, we are tackling the realistic challenges in the MC systems in order to put the theoretical MC systems forward. In particular, we focus on the simplest MC system consisting of a point transmitter, a spherical absorbing receiver and the diffusion-based unbounded channel and in future we can extend the work to more realistic systems. The main contributions of this thesis are the following:

i) this thesis provides the detailed introduction of the channel modelling as well as the signal

Chapter 1. Introduction

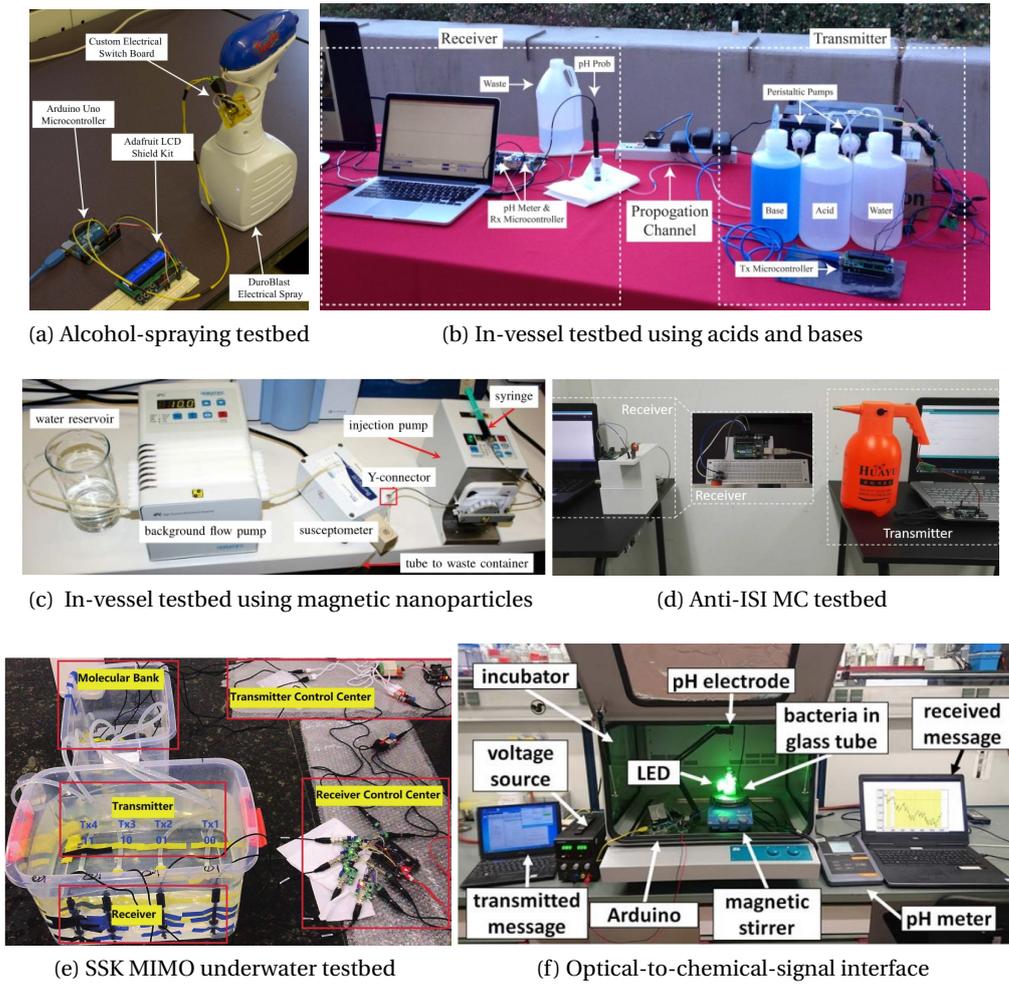


Figure 1.1: Experimental verifications

and noise modelling, and presents the modulation, detection and synchronization schemes for the MC systems (chapter 2) as well as the ML techniques used in this thesis.

ii) this thesis studies the general detection performance evaluation framework for the threshold-based detectors, proposes several DL-based detectors aided by previously detected symbols and in turn optimizes the threshold-based detectors based on the insights acquired from the proposed DL-based detectors (chapter 3).

iii) this thesis reformulates the multi-memory-bit-aided threshold using the intermediate variables that can be obtained by applying K-means clustering algorithm to the multi-dimension data constructed using the received signals yielding the practical non-coherent detectors (chapter 4).

iv) this thesis studies and optimizes the error probability as a function of the constant threshold and the start point of the transmission in the large SNR regime and proposes two synchronization metric functions based on the condition the asymptotic optimal start point follows, as well as the Cramer-Rao lower bounds for the known transmitted bits case and unknown transmitted bits case(chapter 5).

1.4 Publications

Journal Papers The following is a list of publications in refereed journals produced from the research outcomes of this thesis. These journal papers are used as the basis for this thesis.

- (J1) X. Qian, M. Di Renzo, A. Eckford, “Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization”, *IEEE Access*, April, 2019. [1]

Abstract: In this paper, we consider a molecular communication system that is made of a 3D unbounded diffusion channel model without flow, a point transmitter, and a spherical absorbing receiver. In particular, we study the impact of inter-symbol interference and analyze the performance of different threshold-based receiver schemes. The aim of this paper is to analyze and optimize the receivers by using the conventional model-based approach, which relies on an accurate model of the system, and the emerging data-driven approach, which, on the other hand, does not need any a priori information about the system model and exploits deep learning tools. We develop a general analytical framework for analyzing the performance of threshold-based receiver schemes, which are suitable to optimize the detection threshold. In addition, we show that data-driven receiver designs yield the same performance as receivers that have perfect knowledge of the underlying channel model.

- (J2) X. Qian, M. Di Renzo, A. Eckford, “K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications”, *IEEE Transactions on Communications*, submitted, August, 2020. [56]

Abstract: In this paper, we consider non-coherent detection for molecular communication systems in the presence of inter-symbol-interference. In particular, we study non-coherent detectors based on memory-bits-based thresholds in order to achieve low bit-error-ratio (BER) transmission. The main challenge of realizing detectors based on memory-bits-based thresholds is to obtain the channel state information based only on the received signals. We tackle this issue by reformulating the thresholds through intermediate variables, which can be obtained by clustering multi-dimensional data from the received signals, and by using the K-means clustering algorithm. In addition

to estimating the thresholds, we show that the transmitted bits can be retrieved from the clustered data. To reduce clustering errors, we propose iterative clustering methods from one-dimensional to multi-dimensional data, which are shown to reduce the BER. Simulation results are presented to verify the effectiveness of the proposed methods.

- (J3) X. Qian, M. Di Renzo, A. Eckford, "Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems", *IEEE Transactions on Communications*, in Submission, October, 2020. [57]

Abstract: In this paper, we consider the synchronization problem in the diffusive molecular communication system consisting of a point transmitter and a spherical absorbing receiver using the constant threshold. Basically, the synchronization is to estimate the start point of a symbol duration in the received signals. In particular, we focus on the synchronization with unknown channel information. We formulate the error probability as a function of the constant threshold and the start point in order to compute the theoretical optimal start point. Asymptotically, we obtain the condition where the optimal start point satisfies. Based on this condition, we devise a data-aided synchronization metric function. The metric results follow the Skellam distribution allowing us to compute the upper bounds for the absolute mean and variance of the synchronization error assuming that the length of the known symbols is long. Furthermore, the intermediate variables in the metric function can be obtained by applying the unsupervised K-means clustering algorithm to partition the received signals into two groups, yielding a blind synchronization scheme. To verify the synchronization performance, we derive the Cramer-Rao lower bounds for the cases of the known channel information and entirely unknown channel information. Simulation results and the theoretical upper bounds verify the effectiveness of the proposed methods.

Conference Papers The following is the refereed conference that originated from the main findings of this thesis.

- (C1) X. Qian, M. Di Renzo, "Receiver Design in Molecular Communications: An Approach Based on Artificial Neural Networks", *IEEE International Symposium on Wireless Communication Systems*, pp. 1-5, Lisbon, Portugal, 2018. [58]

Abstract: The design of communication systems typically relies on the development of mathematical models that describe the underlying communication channel. In many communication systems, however, accurate channel models may not be known, or the models may not be accurate enough or even not available for efficient system design. In these scenarios, a completely new approach to communication system design and analysis is required. An example of such situations arises in the emerging research field of molecular communications, for which it is very difficult to develop accurate

analytical models for several operating scenarios. In this context, the use of data-driven techniques based on artificial neural networks may provide an alternative and suitable solution towards the design and analysis of molecular communication systems. In this paper, we explore the potential of artificial neural networks for application to the design of robust receiver schemes. We study a molecular communication system in the presence of inter-symbol interference and show that a receiver based on artificial neural networks can be trained by using only empirical (raw) data and can provide the same performance as a receiver that has perfect knowledge of the underlying channel model.

Journal Papers The following is a list of publications the author has contributed to and is not related to the topic of this thesis.

- (J1) X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M. S. Alouini, “Beamforming Through Reconfigurable Intelligent Surfaces in Single-User MIMO Systems: SNR Distribution and Scaling Laws in the Presence of Channel Fading and Phase Noise”, *IEEE Wireless Communications Letters*, accepted, Sep., 2020.

Abstract: We consider a fading channel in which a multi-antenna transmitter communicates with a multi-antenna receiver through a reconfigurable intelligent surface (RIS) that is made of N reconfigurable passive scatterers impaired by phase noise. The beamforming vector at the transmitter, the combining vector at the receiver, and the phase shifts of the N scatterers are optimized in order to maximize the signal-to-noise-ratio (SNR) at the receiver. By assuming Rayleigh fading (or line-of-sight propagation) on the transmitter-RIS link and Rayleigh fading on the RIS-receiver link, we prove that the SNR is a random variable that is equivalent in distribution to the product of three (or two) independent random variables whose distributions are approximated by two (or one) gamma random variables and the sum of two scaled non-central chi-square random variables. The proposed analytical framework allows us to quantify the robustness of RIS-aided transmission to fading channels. For example, we prove that the amount of fading experienced on the transmitter-RIS-receiver channel linearly decreases with N . This proves that RISs of large size can be effectively employed to make fading less severe and wireless channels more reliable.

- (J2) Zappone, A., Di Renzo, M., Farshad Shams, Qian, X, Debbah, M., “Overhead aware design of reconfigurable intelligent surfaces in smart radio environments”, *IEEE Transactions on Wireless Communications*, accepted, Sep., 2020.

Abstract: Reconfigurable intelligent surfaces have emerged as a promising technology for future wireless networks. Given that a large number of reflecting elements is typically used, and that the surface has no signal processing capabilities, a major challenge is to cope with the overhead that is required to estimate the channel state information and

to report the optimized phase shifts to the surface. This issue has not been addressed by previous works, which do not explicitly consider the overhead during the resource allocation phase. This work aims at filling this gap, developing an overhead-aware resource allocation framework for wireless networks where reconfigurable intelligent surfaces are used to improve the communication performance. An overhead model is developed and incorporated in the expressions of the system rate and energy efficiencies, which are then optimized with respect to the phase shifts of the reconfigurable intelligent surface, the transmit and receive filters, and the power and bandwidth used for the communication and feedback phases. The bi-objective maximization of the rate and energy efficiency is carried out as well. The proposed framework allows characterizing the trade-off between optimized radio resources and the related overhead in networks with reconfigurable intelligent surfaces.

- (J3) Di Renzo, M., Ntontin, K., Song, J., Danufane, F.H., Qian, X., Lazarakis, F., De Rosny, J., Phan-Huy, D.T., Simeone, O., Zhang, R. and Debbah, M., “Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison”, *IEEE Open Journal of the Communications Society*, accepted, Jun., 2020.

Abstract: Reconfigurable intelligent surfaces have emerged as a promising technology for future wireless networks. Given that a large number of reflecting elements is typically used, and that the surface has no signal processing capabilities, a major challenge is to cope with the overhead that is required to estimate the channel state information and to report the optimized phase shifts to the surface. This issue has not been addressed by previous works, which do not explicitly consider the overhead during the resource allocation phase. This work aims at filling this gap, developing an overhead-aware resource allocation framework for wireless networks where reconfigurable intelligent surfaces are used to improve the communication performance. An overhead model is developed and incorporated in the expressions of the system rate and energy efficiencies, which are then optimized with respect to the phase shifts of the reconfigurable intelligent surface, the transmit and receive filters, and the power and bandwidth used for the communication and feedback phases. The bi-objective maximization of the rate and energy efficiency is carried out as well. The proposed framework allows characterizing the trade-off between optimized radio resources and the related overhead in networks with reconfigurable intelligent surfaces.

- (J4) Zappone, A., Di Renzo, M., Debbah, M., Lam, T. T., Qian, X., “Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization”, *IEEE Vehicular Technology Magazine*, accepted, July, 2019.

Abstract: Deep learning based on artificial neural networks is a powerful machine learning method that, in the last few years, has been successfully used to realize tasks, e.g., image classification, speech recognition, translation of languages, etc., that are usually simple to execute by human beings but extremely difficult to perform by ma-

chines. This is one of the reasons why deep learning is considered to be one of the main enablers to realize the notion of artificial intelligence. In order to identify the best architecture of an artificial neural network that allows one to fit input-output data pairs, the current methodology in deep learning methods consists of employing a data-driven approach. Once the artificial neural network is trained, it is capable of responding to never-observed inputs by providing one with the optimum output based on past acquired knowledge. In this context, a recent trend in the deep learning community is to complement pure data-driven approaches with prior information based on expert knowledge. In this work, we describe two methods that implement this strategy, which aim at optimizing wireless communication networks. In addition, we illustrate numerical results in order to assess the performance of the proposed approaches compared with pure data-driven implementations.

2 Fundamentals of MC Systems

Chapter 2. Fundamentals of MC Systems

This chapter begins with *Section 2.1* which introduces several typical propagations and mainly takes the diffusion-based propagation as an example to elucidate how the channel response is modelled in addition to the signal and noise modelling. *Section 2.2* presents the introduction and analysis of three basic modulation schemes, coupled with the composite modulation scheme that stresses the importance of the concentration shift keying. *Section 2.3* presents the machine learning techniques exploited in this thesis. *Section 2.4* introduces the detectors used in the MC systems. *Section 2.5* introduces the basic synchronization methods in the MC systems.

2.1	Channel Modelling	17
2.1.1	Diffusion-based Propagation	17
2.1.2	Flow-aided Propagation	21
2.1.3	Bacteria-assisted Propagation	22
2.1.4	Motor-Protein-based Propagation	23
2.1.5	Received Signal and Noise Modelling	23
2.2	Modulation Schemes	25
2.2.1	Basic Modulation	26
2.2.2	Composite Modulation Scheme	27
2.3	Machine Learning for Molecular Communication Systems	28
2.3.1	Feedforward Neural Networks	29
2.3.2	K-Means Clustering Algorithm	33
2.4	Detection	35
2.4.1	Basics of the Detection	35
2.4.2	Single-Sample/Multiple-Sample Detectors	35
2.4.3	Single-Symbol/Multiple-Symbol Detectors	36
2.4.4	Coherent/Non-Coherent Detectors	37
2.5	Synchronization	38
2.5.1	Synchronization with Channel Information	38
2.5.2	Synchronization without Channel Information	40

2.1 Channel Modelling

As stated in the aforementioned chapter, there are several kinds of propagation mechanisms in the MC that are associated with the type of molecules. In addition, each type of propagation has the specified transmitter and receiver. Thus, in this section, we directly include the introduction of the transmitter and the receiver in the statement of four typical propagation mechanisms: diffusion-based propagation, flow-aided propagation, bacteria-assisted propagation and motor-protein-based propagation and introduce how to model the channel.

2.1.1 Diffusion-based Propagation

The diffusion-based MC system is common in the fluid medium, e.g., a liquid or a gas environment, and serves as the most energy-efficient MC system. It is mainly based on the random motion of the particles in the medium, which is also known as the Brownian motion using the thermal energy in the medium. Thus, no extra energy is needed for the diffusion-based propagation.

The Brownian motion can be accurately modelled by the Wiener process and simulated by the Monte Carlo simulation. Particularly, let (x_0, y_0, z_0) be the initial position of the information particle at time $t = 0$ in the three-dimensional Cartesian coordinates, for a discrete time period ΔT and any integer $i > 0$, the position of the information particle at time $t = i\Delta T$, expressed as (x_i, y_i, z_i) , has the relationship as follows:

$$(x_i, y_i, z_i) = (x_{i-1}, y_{i-1}, z_{i-1}) + (\Delta x_i, \Delta y_i, \Delta z_i) \quad (2.1)$$

where the Δx_i , Δy_i and Δz_i represent the random displacement during the i -th ΔT period and are quantified by Gaussian random variables as follows:

$$\Delta x_i \sim \mathcal{N}(0, 2D\Delta T) \quad (2.2)$$

$$\Delta y_i \sim \mathcal{N}(0, 2D\Delta T) \quad (2.3)$$

$$\Delta z_i \sim \mathcal{N}(0, 2D\Delta T) \quad (2.4)$$

where D denotes the free diffusion coefficient. The Brownian motion in one-dimensional and two-dimensional coordinates can be simulated in a similar fashion. This coefficient D is affected by the sizes of the information molecule (S_m) and the particle (S_p) already existing in the medium. In practice, D is given by:

$$D = \begin{cases} \frac{k_B T}{6\pi\eta R_H}, & S_m \gg S_p \\ \frac{k_B T}{4\pi\eta R_H}, & S_m \simeq S_p \end{cases} \quad (2.5)$$

where $k_B = 1.38 \cdot 10^{-23} J/K$ denotes the Boltzman constant, T denotes the temperature in

K (here T represents the temperature and we redefine it where used differently as period time), η denotes the dynamic viscosity of the fluid and the R_H denotes the hydraulic radius of the information molecule. In realistic scenarios, the environment is bounded and heterogeneous. However, considering the size of the MC system, the environment can be deemed to be homogeneous and unbounded such that D is constant anywhere in the medium.

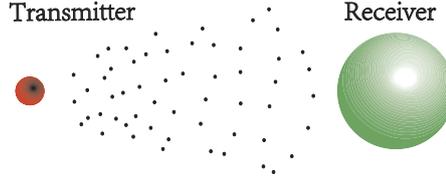


Figure 2.1: Diffusion-based propagation

In most scenarios, once the information particle reaches to the receiver, it will be absorbed at once and will not bounce back into the medium, i.e., the information molecule only hits the receiver once. In this regard, the receiver is referred to as a absorbing receiver. In order to quantify the probability for an information molecule reaching to the receiver from the transmitter, we resort to the microscopic diffusion theory based on the assumption that molecules move down according to its concentration gradient. We assume that the receiver is spherical with radius r_r (depicted in Fig. 2.1) and is located at the origin. We also assume that an ideal point transmitter locates at some points with the distance r_0 from the center of the receiver. Since the information molecule moves randomly and independently, for any molecule released on the surface of a large sphere, the probability for this molecule reaching to the spherical receiver of radius r_r is the same. Equivalently, it is the same as the case where a sphere of molecules of radius r_0 are released and are absorbed by a spherical receiver of radius r_r .

In the 3D system, the probability $P(r, t)$ for a molecule existing at the surface of a sphere with radius r at the time t after the release at the spherical surface of radius r_0 follows the Fick's second law [14] as follows:

$$\frac{\partial P(r, t|r_0)}{\partial t} = D\nabla^2 P(r, t|r_0) \quad (2.6)$$

where ∇^2 denotes the Laplacian operator. To simplify the system, we assume that the transmitter is ideally a point transmitter and the environment is unbounded, yielding the initial condition as follows:

$$P(r, t = 0|r_0) = \frac{1}{4\pi r_0^2} \delta(r - r_0) \quad (2.7)$$

as and the first boundary condition as follows:

$$\lim_{r \rightarrow \infty} P(r, t = 0|r_0) = 0 \quad (2.8)$$

Subsequently, the absorbing receiver will have a second boundary condition as follows:

$$D \frac{\partial P(r, t|r_0)}{\partial r} = wP(r, t|r_0), r = r_r \quad (2.9)$$

where w denotes the rate of absorption, which is set to be infinitely large. With the Laplacian transform, we can obtain the following results:

$$P(r, t|r_0) = \frac{1}{4\pi r r_0} \frac{1}{\sqrt{4\pi D t}} \left(e^{-\frac{(r-r_0)^2}{4Dt}} - e^{-\frac{(r+r_0-2r_r)^2}{4Dt}} \right) \quad (2.10)$$

Inserting $r = r_r$ into (2.10) and integrating over a sphere (the surface of the receiver), we obtain the hitting rate of a molecule reaching to the receiver at time t as follows:

$$f_{hit}(t) = 4\pi r_r^2 w P(r, t|r_0) \quad (2.11)$$

by the second boundary condition, we have

$$f_{hit}(t) = \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D t}} \frac{r_0 - r_r}{t} e^{-\frac{(r_r-r_0)^2}{4Dt}} \quad (2.12)$$

In addition, the probability for a molecule reaching to the receiver until time t is:

$$P(t) = \int_0^t f_{hit}(t') dt' = \frac{r_r}{r_0} \operatorname{erfc} \left[\frac{r_0 - r_r}{\sqrt{4Dt}} \right] \quad (2.13)$$

If the transmitter releases N_{TX} molecules, the channel response is

$$H(t) = N_{TX} f_{hit}(t) \quad (2.14)$$

Concerning the one-dimensional and two-dimensional diffusion-based MC systems, the channel response can be modelled in the same manner.

Apart from the absorbing receiver, there exists another type of receiver that just counts the received molecules, referred to as the passive receiver. First, we assume that the point transmitter locates at the origin, after the transmitter releases N_{TX} molecules, the concentration C follows the Fick's second law as follows:

$$\frac{\partial C}{\partial t} = D \nabla^2 C \quad (2.15)$$

The initial condition is subsequently given by:

$$C(x, y, z, t = 0) = N_{TX} \delta(x) \delta(y) \delta(z) \quad (2.16)$$

where $\delta(x)$ denotes the Dirac delta function that $\delta(x) = 0$ for $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x) dx = 1$. The

solution to the above problem is:

$$C(x, y, z, t) = \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{x^2+y^2+z^2}{4Dt}} \quad (2.17)$$

If we assume that the volume of the receiver is not so large that within the spherical receiver, the concentration can be regarded as a constant, subsequently the channel response can be written as follows:

$$H(t) = \frac{4}{3}\pi r_r^3 C(x, y, z, t) = \frac{4}{3}\pi r_r^3 \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{x^2+y^2+z^2}{4Dt}} \quad (2.18)$$

Concerning the one-dimensional and two-dimensional cases, the channel response can be modelled in a similar fashion. The concentrations can be written as follows:

$$C(x, t) = \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{x^2}{4Dt}} \quad (2.19)$$

$$C(x, y, t) = \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{x^2}{4Dt}} \quad (2.20)$$

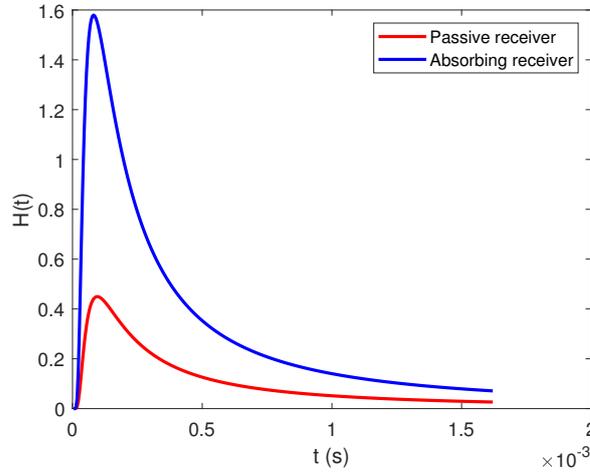


Figure 2.2: Channel response comparison ($D = 4.265 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500\text{nm}$ and $N_{TX} = 2000$)

In Fig. 2.2, we compare the difference between the two types of channel response modelling by considering a simple MC system where the systematic parameters are from the literature [1]. From this figure, we evince that the absorbing receiver has larger magnitude in channel response while the main trends are the same which firstly goes up to a maximum point and then decreases gradually. This is of great importance as some approaches exploit the trend to design synchronization schemes. We also benefit from this which will be clarified in the following chapter.

As mentioned in the previous chapter, the enzymes can be exploited in order to reshape the channel response. However, due to the complexity of the reaction of the enzyme and the information, the analytical channel response is hard to obtain. Instead, researchers resort to the molecule-level simulators to model the reaction mechanisms and the realistic transmitter and the realistic receiver. The channel response is averaged after the simulations.

2.1.2 Flow-aided Propagation

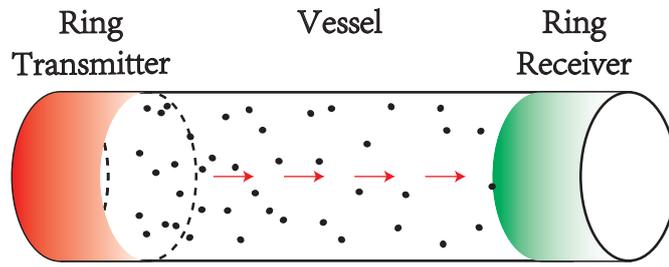


Figure 2.3: Flow-aided propagation

Although no external energy is needed in the diffusion-based propagation, if we want to convey information to a longer distance, the diffusion-based propagation is incapable due to the slow speed. A direct way to accelerate the propagation speed is to make use of the flow [20] in the environment. An example is our blood vessels as plotted in Fig. 2.3. With the aid of the flow, the movement of the molecules can be modelled in a similar manner like the diffusion-based propagation as follows:

$$\Delta x_i \sim N(v_x \Delta T, 2D\Delta T) \quad (2.21)$$

$$\Delta y_i \sim N(v_y \Delta T, 2D\Delta T) \quad (2.22)$$

$$\Delta z_i \sim N(v_z \Delta T, 2D\Delta T) \quad (2.23)$$

where v_x , v_y and v_z are the flow velocities in the x , y , z directions. Meanwhile, the flow velocities can be a function of the time or position since the velocity may be non-homogeneous in the medium.

The passive receiver (ring receiver) is usually employed, the concentration at the receiver side (take the one-dimensional MC system as an example) can be written as follows:

$$C(x, t) = \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{(x-v_x t)^2}{4Dt}} \quad (2.24)$$

where v_x denotes the flow velocity in the positive x direction. And similarly, for 2D system,

$$C(x, t) = \frac{N_{TX}}{(\sqrt{4\pi Dt})^3} e^{-\frac{(x-v_x t)^2 + (y-v_y t)^2}{4Dt}} \quad (2.25)$$

where where v_y denotes the flow velocity in the positive y direction. If more factors, e.g., degradation (the information molecules are automatically decomposed into other molecules with a certain probability), are considered, the way to derive the expected concentration in space and time is similar. We refer readers to the advection-reaction-diffusion equation [59] for more details.

2.1.3 Bacteria-assisted Propagation

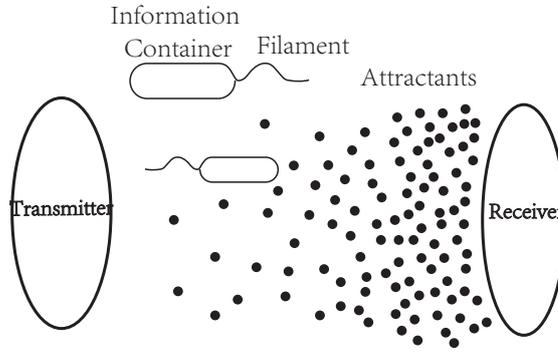


Figure 2.4: Bacteria-assisted propagation

Different from the above mentioned passive MC systems, the bacteria-assisted propagation [23] exploits the active bacteria to transport the information molecules to the receiver. The bacteria act as a postman and can be directly customized by using the modern biological techniques. The transmitter lets out the bacteria and the bacteria move randomly in the same fashion as the diffusion-based propagation. The position of the bacteria can be modelled similarly as follows:

$$(x_i, y_i, z_i) = (x_{i-1}, y_{i-1}, z_{i-1}) + (\Delta x_i, \Delta y_i, \Delta z_i) \quad (2.26)$$

Thus, a similar result as the (2.12) can be obtained if we assume that the bacteria will be collected once reaching to the receiver. In practice, the receiver may constantly let out some specified particles and the bacteria may be guided by these particles as illustrated in Fig. 2.4. In this context, we can model this channel as the flow-aided propagation and the channel response is similar to (2.24). This type of propagation is essential to the application of drug delivery, as the target organs may malfunction and produce some particles such that the bacteria can distinguish the target organ from the normal running organs.

2.1.4 Motor-Protein-based Propagation

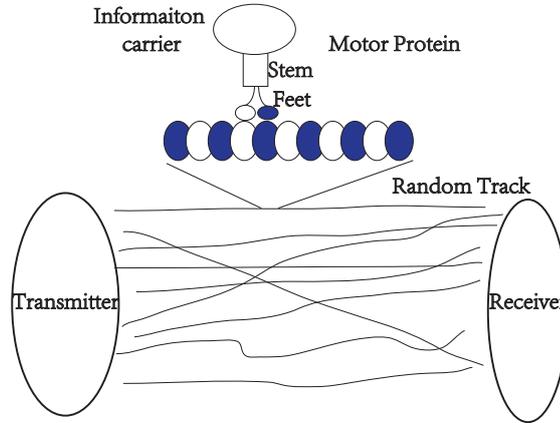


Figure 2.5: Motor-protein-based propagation

In biology, the information molecules can also be transported via motor proteins [21]. Similar to the bacteria-assisted propagation, the motor proteins serve as the postmen carrying molecules and walk along the microtubule that can be regarded as the track connecting the transmitter and the receiver. The motor protein does have the feet, depicted in Fig. 2.5 and it takes a step if the energy molecules in the bodies, the adenosine triphosphate (ATP), decompose and pass energy to it. In particular, the movement of the motor protein can be modelled as one-dimensional movement. Denote the l_i as the position at the i th time instant on the track, subsequently, l_i is given by:

$$l_i = l_{i-1} + v\Delta T \quad (2.27)$$

where v denotes the speed of the motor protein and is decided by the energy the motor protein takes.

There exist many tracks from the transmitter and the receiver and the motor protein randomly selects the track after released by the transmitter. It is worth noting that the track is not a straight line and the tracks are not of the same length. Therefore, even though we model the movement as a one-dimensional movement, the position of the motor protein in the three-dimensional space does not change linearly [24]. On the other hand, if we consider the probability that the motor proteins detach from the track, it is in practice similar to the degradation in the diffusion-based propagation, thus, similar results can be exploited.

2.1.5 Received Signal and Noise Modelling

As we mentioned in the previous subsections, the probability for a transmitted molecule collected by the receiver can be obtained with analytical expressions which are continuous

with respect to the time t . Practically, the receiver does not count the received molecule number in a continuous manner. The receiver just reports the received molecules within a short duration ΔT (known as the sampling duration), at the end of this sampling duration. Consequently, the number of the received molecules in the i th sample duration after releasing the molecules is expressed as follows:

$$C_i = \int_{t=i\Delta T}^{(i+1)\Delta T} H(t) dt \quad (2.28)$$

In terms of the noise, it occurs as there are molecules from other MC systems and we assume that the distance between the other MC system and the MC system we are interested in is long such that on average the number of noise molecule is stable.

As we have pointed, the motion of the molecules is normally modelled by the Brownian motion, the realistic received molecule number is a random variable. After releasing N_{TX} molecules at the transmitter, in the i th sample duration, the actual number of received molecules is not simply C_i . In the following, we will introduce three models for modelling the actual received molecule number.

Binomial Model

Considering an unbounded environment in the MC system, we assume that the movement of each molecule is independent of each other. Thus, for one molecule, it is either collected by the receiver or moving in the medium, i.e., the number of observed molecules c_i follows the Binomial distribution $c_i \sim \mathcal{B}(N_{TX}, \frac{C_i}{N_{TX}})$ with the probability mass function (PMF) as follows:

$$f_B(c_i) = \binom{N_{TX}}{c_i} \left[\frac{C_i}{N_{TX}} \right]^{c_i} \left[1 - \frac{C_i}{N_{TX}} \right]^{N_{TX}-c_i} \quad (2.29)$$

for all $c_i \in \{0, 1, \dots, N_{TX}\}$. However, this model only suits for simple MC systems as this model does not take the noise and the ISI into consideration. Realistically, the received number may exceed N_{TX} . In the following, we present other two approaches for simplicity and better tractability.

Gaussian Model

If the transmitted molecule number N_{TX} is large enough, by the central limit theorem (CLT), we can approximate the c_i as a Gaussian random variable with the probability density function (PDF) as follows:

$$f_N(c_i) = \frac{1}{2\pi C_i \left(1 - \frac{C_i}{N_{TX}}\right)} \exp\left(-\frac{(c_i - C_i)^2}{2C_i \left(1 - \frac{C_i}{N_{TX}}\right)}\right) \quad (2.30)$$

We assume that the MC systems we are interested in is far away from the other MC systems such that the probability of receiving molecules from other MC systems P_n is a constant, i.e., both the mean and variance of the noise are constant. Consequently, we can integrate the noise in this model as follows:

$$f_N(c_i) = \frac{1}{2\pi[\mathbb{C}_i(1 - \frac{\mathbb{C}_i}{N_{TX}}) + \sigma_n^2]} \exp\left(-\frac{(c_i - \mathbb{C}_i - \lambda)^2}{2[\mathbb{C}_i(1 - \frac{\mathbb{C}_i}{N_{TX}}) + \sigma_n^2]}\right) \quad (2.31)$$

where $\lambda = N_{TX}P_n$ and $\sigma_n^2 = N_{TX}P_n(1 - P_n)$ are the mean and variance of the noise. However, even N_{TX} is sufficiently large, the mean received molecule number \mathbb{C}_i may be small and the Gaussian model still may not be applicable as the c_i is non-negative and discrete.

Poisson Model

A better signal modelling may be the Poisson distribution when the received number can be sufficiently large and we want to compute the probability of each discrete received particle number rather than the overall probability of received signal less than a threshold where the Gaussian model is the better choice. Consequently the PMF is given by:

$$f_P(c_i) = \frac{(\mathbb{C}_i + \lambda)^{c_i} \exp(-[\mathbb{C}_i + \lambda])}{c_i!} \quad (2.32)$$

In summary, there are many propagation mechanisms in nature and based on the diffusion-based propagation, we can model other propagation mechanisms under the universal framework in order to obtain the channel response. Based the signal and noise modelling, we are able to obtain the probability of receiving a certain number of molecules at each sample duration such that we can employ the modulation scheme and detecting the received signals.

2.2 Modulation Schemes

As mentioned in the previous section, once the molecules are released, the molecules will randomly propagate in the medium, resulting in the intrinsic ISI. Traditional wireless communication systems exploit the modulation techniques to mitigate the ISI and improve the transmission performance. In MC systems, some modulation schemes that are similar to the modulation schemes in the wireless communications are employed in order to combat the ISI. In the following, we will review three basic modulation schemes and will introduce a variant based on the common modulation schemes.

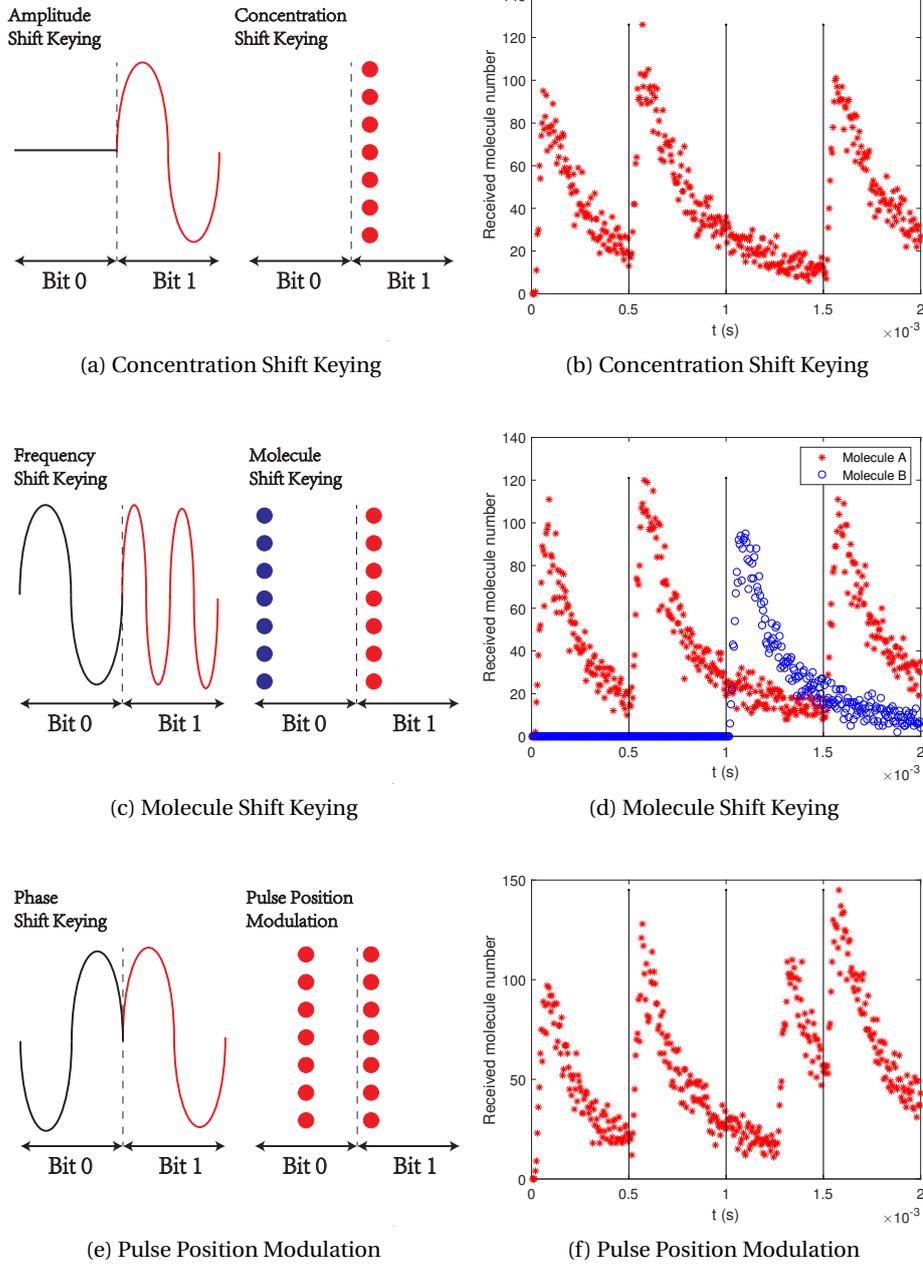


Figure 2.6: Three basic modulation schemes and the received signal comparison (the transmitted symbols are [1, 1, 0, 1]) in MC systems. ($D = 4.265 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500\text{nm}$, $N_{TX} = 100,000$, $\Delta T = 5\mu\text{s}$ and $T = 100\Delta T$ [1])

2.2.1 Basic Modulation

In wireless communications, the information is conveyed based on the manipulation of the sinusoidal waves over a long period. Specifically, information bits are modulated onto the

amplitude, frequency and phase of a segment of sinusoidal wave. These modulation schemes are known as amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK). While in MC systems, molecules are released in a ultra-short time at the beginning of the symbol slot. The characteristics that can be used to transmit information are the number of molecules, the type of released molecules and the release time. These modulation schemes are referred to as concentration shift keying (CSK), molecule shift keying (MoSK) and pulse position modulation (PPM). The comparison between CSK, MoSK and PPM with the counterparts in wireless communications is depicted in the Fig. 2.6.

After passing through the non-sparse memory channel, the received molecule number at each sampling instant is recorded. For the binary CSK, at the beginning of each symbol slot, N_{TX} molecules are released if the symbol is binary one otherwise no molecule is transmitted. As observed in Fig. 2.6b, even the transmitted symbol is binary zero, the received signals are not zero due to the ISI. For the multiple amplitude CSK, the received signals are similar. However, the ISI will accordingly increase and will impose more constraints on the demodulation. Meanwhile, the binary CSK is more commonly seen. Thus, without specific notation, we refer the binary CSK to CSK. As for the MoSK, two kinds of molecules are exploited, if the current symbol is binary one, molecule A is released otherwise molecule B is transmitted. From Fig. 2.6d, we observe that the received number of the molecule A is similar to the Fig. 2.6b as we can regard the MoSK as the mixture of two CSKs. By comparing the numbers of the received molecule A and molecule B, we can easily infer the transmitted symbols, thus leading to better transmission performance. In terms of the PPM, if the current symbol is binary one, the molecules are released at the beginning of the symbol or if the symbol is zero, the molecules are released at the middle of the symbol slot. Among the three basic modulation schemes, the binary CSK is the most energy-efficient and easiest to implement as if the current symbol is zero, no molecule is released.

2.2.2 Composite Modulation Scheme

As shown in the Fig. 2.6, the unwanted ISI will degrade the transmission performance. As the ISI is related to the length of the symbol slot, if the symbol slot is long, the ISI will be mild. Based on this intuition, by combining the CSK and the MoSK, a composite modulation scheme depicted in Fig. 2.7 is proposed by alternatively releasing the molecule A and the molecule B according to the transmitted symbols. The effective symbol length for the same molecule is doubled, thus the receiver experiences a mild ISI. Meanwhile, the detection techniques used for CSK can be exploited directly in the composite modulation scheme. In this case, the composite modulation scheme combines the advantages of the CSK and the MoSK, thus, the CSK is of great importance to the MC systems. In the following, we mainly focus on the CSK-based detection algorithms before we start to exploit the composite modulation scheme and analyse its transmission performance.

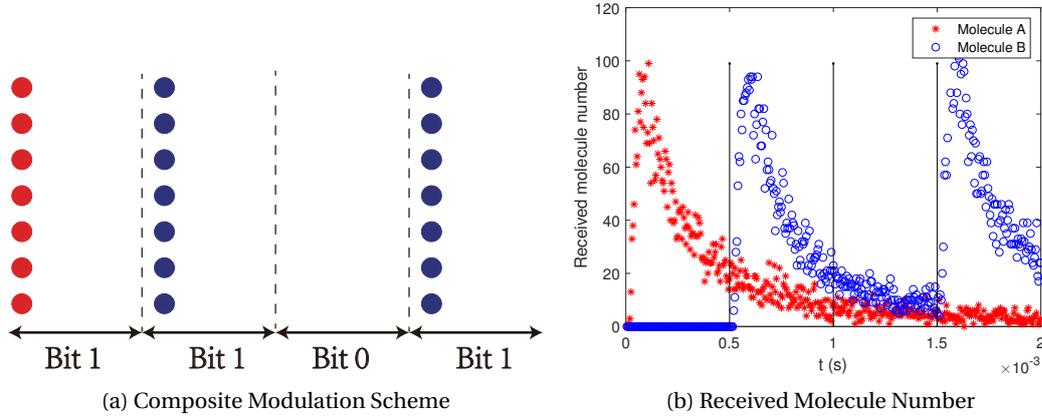


Figure 2.7: The composite modulation scheme and the illustration of the received signals. The transmitted symbols are [1, 1, 0, 1]. ($D = 4.365 \times 10^{-10} \text{ m}^2/\text{s}$, $r_r = 45\text{nm}$, $r_0 = 500 \text{ nm}$, $N_{TX} = 100,000$, $\Delta T = 5 \text{ us}$ and $T = 100\Delta T$)

2.3 Machine Learning for Molecular Communication Systems

As in the aforementioned chapter, under some severe cases, the channels may not be modelled with the explicit expressions. The only information about the channel is the input-output relations, i.e., the MC system is totally a black box to the researchers. Among all techniques than can handle this problem is the neural networks (NNs) in the machine learning (ML). In particular, NN is deemed to be capable of solving almost every problem as long as trained with enough data. We are interested in applying the NN to the MC systems, specifically in terms of the detection, observing how the NN-based detection schemes behave and we are keen on figuring out the relationship between the NN-based detectors and their counterparts, threshold-based detectors.

In addition to the supervised NN-based scheme, we borrow an unsupervised learning technique in ML in order to fulfil the detection without known channel information, which is referred to as the non-coherent detection. As from the perspective of ML, the non-coherent detection (observing the received signal and inferring the transmitted bit) is in practice an unsupervised assigning problem, assigning a point (the received signal) with a label (the information containing the transmitted bit). As there are a limited number of labels in the detection, this problem in fact is converted into a clustering problem, separating the received signals into several groups and assigning each group of signals with a predefined label that can be done by the K-means clustering algorithm.

There are many other ML techniques, e.g., multi-arm-bandit, particle swarm optimization, and we only select the two techniques (NN and K-means) among all techniques. It's not because they are the only choices, but we just try to employ the two and investigate what the two techniques can give. The future work can be exploiting other technologies. In the following, we will present the ML techniques we employed in the chapter 3 and chapter 4 before we introduce the detection schemes in the MC systems.

2.3.1 Feedforward Neural Networks

The NN is comprised of several layers of elementary processing units, that are referred to as neurons. Basically, the NN has an input layer, an output layer and several hidden layers. In each layer, the number of the neurons is not necessarily equal. If more than one hidden layers are deployed, the NN is also named as deep NN (DNN). Specifically, there are two types of NN: 1) feedforward NN which can only handle the input-output-like problems, 2) recurrent NN (RNN) that is generalized from the feedforward NN and is capable of processing time-series-based problems. We mainly introduce the feedforward NN in this chapter as we deploy it in the following chapter and it is easy to understand.

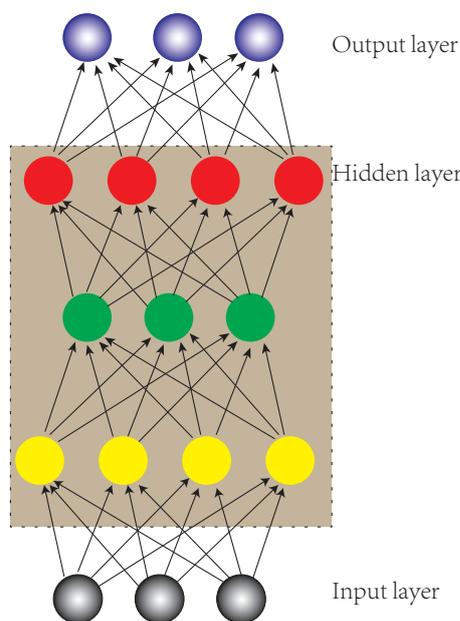


Figure 2.8: Typical structure of a feed-forward NN with fully-connected layers.

A typical feedforward NN structure is depicted in Fig. 2.8. We observe that the output of one layer is actually the input of the next layer. In addition, each neuron has the structure as shown in Fig. 2.9. After the N_0 -dimensional datum (vector) \mathbf{x}_0 is fed into the input layer, the output of the input layer would go through the hidden layers and finally is processed by the output layer.

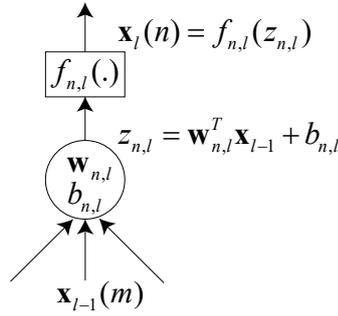


Figure 2.9: Basic structure of a neuron

It's worth noting that each neuron only has one output and this output will be connected to every neurons of the next layer. Take the l th layer that is comprised of N_l neurons for example and let us denote the output of the $(l - 1)$ th layer as \mathbf{x}_{l-1} . Consequently, the output of the n th neuron in the l th layer, if denoted as $\mathbf{x}_l(n)$, is given by:

$$\mathbf{x}_l(n) = f_{n,l}(z_{n,l}) \tag{2.33}$$

$$z_{n,l} = \mathbf{w}_{n,l}^T \mathbf{x}_{l-1} + b_{n,l} \tag{2.34}$$

where \mathbf{x}_{l-1} denotes the outputs of all neurons in the $(l - 1)$ th layer, $\mathbf{w}_{n,l}$ and $b_{n,l}$ denote the weights and bias of the n th neuron in the l th layer and $f_{n,l}(\cdot)$ is the activation function.

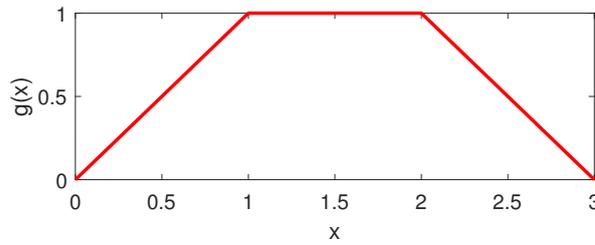


Figure 2.10: A non-linear function $g(x)$

The activation function $f_{n,l}(\cdot)$ is of great importance to the NN and we will elaborate how a non-linear function is emulated by the NN from an example, $g(x)$, depicted in Fig. 2.10.

The NN for this function has three layers. The input layer and output layer both have one neuron and there is only one hidden layer consisting of three neurons. The corresponding activation function exploited in the NN is the Rectified Linear Unit (ReLU) function, given by:

$$\text{ReLU}(x) = \max(0, x) \tag{2.35}$$

Denote the input as x , subsequently, the output of the input layer is give by:

$$x_1 = f_{1,1}(x) = \max(0, x) \quad (2.36)$$

In the hidden layer, the weights of the neurons are unit one while the biases of neurons are different, which are $b_{1,2} = 0$, $b_{2,2} = -1$ and $b_{3,2} = -2$, respectively. In this case, the corresponding outputs are depicted in Fig. 2.11 and given by:

$$x_2(1) = f_{1,2}(x_0) = \max(0, x_1) = \max(0, x) \quad (2.37)$$

$$x_2(2) = f_{2,2}(x_0 - 1) = \max(0, x_1 - 1) = \max(0, x - 1) \quad (2.38)$$

$$x_2(3) = f_{3,2}(x_0 - 2) = \max(0, x_1 - 2) = \max(0, x - 2) \quad (2.39)$$

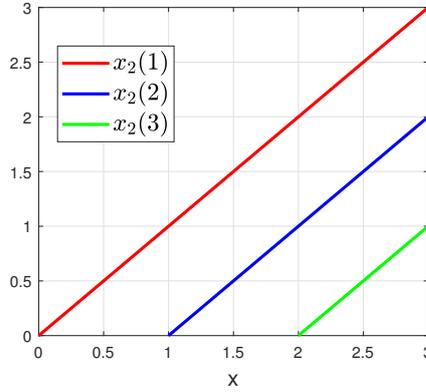


Figure 2.11: Plot of $x_2(1)$, $x_2(2)$ and $x_2(3)$.

As for the output layer, $b_{1,3} = 0$, and the weight vector is $\boldsymbol{w}_{1,3} = [1, -1, -1]^T$ such that the output can be written as follows:

$$x_3 = f_{1,3}(x_2(1) - x_2(2) - x_2(3)) = \max(\max(0, x) - \max(0, x - 1) - \max(0, x - 2)) \quad (2.40)$$

The (2.40) corresponds to Fig. 2.10. This example shows how a piece-wise (non-linear) function can be interpreted by the NN using the ReLU function. However, exploiting the ReLU function has two main drawbacks: cannot processing problems with negative inputs and hard to optimize as the ReLU function does not have a continuous derivative. Having no continuous derivative leads to the difficulty in training as the NN mainly relies on the gradient descent-like algorithms to update the inner parameters (weights and biases). In pursuit of emulating more complex non-linear functions, more neurons are needed and the activation functions innate with the continuous derivatives are necessary.

Over the past decades, several functions have been taken into account. Among these function, the sigmoid function $\sigma(x)$ and the hyperbolic tangent function $\tanh(x)$ are essential which are

given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.41)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.42)$$

However, due to the saturation in large x in these activation functions, the convergence speed of the corresponding NN will slow down, known as vanishing gradient problem. To cope with this issue, the training data (including input and output data) are usually normalized within a predefined small range, e.g., $[-1, 1]$ or $[0, 1]$.

As elucidated in (2.40), the output of a NN with respect to the input to some extent is a function containing some parameters. Generally, we can formulate the NN as a function $\hat{\mathbf{x}}(\{\mathbf{W}_i, \mathbf{b}_i\}_{i=1}^L, \mathbf{x}_0)$ as follows:

$$\hat{\mathbf{x}}(\{\mathbf{W}_i, \mathbf{b}_i\}_{i=1}^L, \mathbf{x}_0) = f_L(\mathbf{b}_L + \mathbf{W}_L^T f_{L-1}(\mathbf{b}_{L-1} + \mathbf{W}_{L-1}^T (\dots f_1(\mathbf{b}_1 + \mathbf{W}_1^T \mathbf{x}_0)))) \quad (2.43)$$

where \mathbf{W}_i and \mathbf{b}_i are the matrix and the vector containing the weights and the biases of the neurons in the i th layer and \mathbf{x}_0 denotes the input vector and $f_i(\cdot)$ for $1 \leq i \leq L$ denotes the activation function that works in an element-wise manner. Our objective is to optimize these matrices and vectors in order to minimize difference between the realistic results and the NN output results, quantified by the loss function \mathcal{L} . For instance, in terms of the mean square error, $\mathcal{L} = \frac{1}{N} \sum_{n=1}^N [\mathbf{x}(n) - \hat{\mathbf{x}}(\{\mathbf{W}_i, \mathbf{b}_i\}_{i=1}^L, \mathbf{x}_0)]^2$, where N denotes the size of the training data. Of course, these parameters can be optimized via gradient descent based algorithms from the \mathbf{W}_L and \mathbf{b}_L to the \mathbf{W}_1 and \mathbf{b}_1 , which is called back-propagation (BP) algorithm.

The parameters \mathbf{W}_i and \mathbf{b}_i in the function (2.43) determine the learning ability of the NN. If more layers are deployed, i.e., more \mathbf{W}_i and \mathbf{b}_i in the NN, by applying BP algorithm, the loss function \mathcal{L} may reduce to a satisfactory level. However, this does not mean that the more layers in the NN, the better performance this NN will achieve. By training the NN with enough time, the loss function will decrease with respect to the amount of parameters (degrees of freedom) in the NN, while the testing error may not necessarily reduce along with the training error, as depicted in the Fig. 2.12 [60].

If the NN is structured inappropriately, the NN does not have enough degrees of freedom to capture all the features of the training data, which is named as underfitting. On the other hand, if too many layers are deployed, the NN of course may learn the desired features in the training data, however, with more degrees of freedom, the specific properties in the training data may be stored, leading to more testing errors, which is called overfitting. Therefore, the appropriate NN structure is in particular important in order to avoid underfitting and

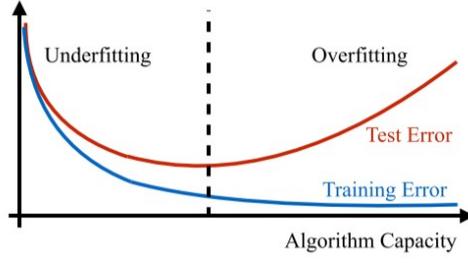


Figure 2.12: Typical behaviors of the training and testing processes

overfitting. In summary, the key point in applying NN to the MC systems lies in designing the appropriate NN structure. In this thesis, the NN structures we employ are appropriate so that there is no underfitting or overfitting.

2.3.2 K-Means Clustering Algorithm

As we have mentioned, the non-coherent detection is in fact an unsupervised clustering problem [61]. Suppose we have a set of points $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ of size N , in which each point \mathbf{x}_i is a D -dimensional vector. Our objective is to separate the points into K groups (the group of points is referred to as a cluster). Usually, we assume the number of clusters is known.

Realistically, there are many rules in terms of the points separation and the specific rule must be selected according to the distribution of the points. In most cases, there is no overlapping between clusters or the overlapping area is negligible such that we can attribute a point to a cluster if the distance between this point to the assigned cluster's centroid (the arithmetic mean of all the points in a cluster) is smaller than the distance from this point to other cluster's centroid.

For convenience, we denote the centroid of the k th cluster as $\boldsymbol{\mu}_k$ and we introduce the binary indicator variable $\kappa_{n,k}$ such that if the n th point belongs to the k th cluster, $\kappa_{n,k} = 1$, and otherwise $\kappa_{n,k} = 0$. In this context, the objective function is:

$$J = \sum_{n=1}^N \sum_{k=1}^K \kappa_{n,k} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 \quad (2.44)$$

Optimizing this function involves two steps to iteratively update the $\kappa_{n,k}$ and $\boldsymbol{\mu}_k$. First, given the centroids, $\boldsymbol{\mu}_k$, we minimize J with respect to $\kappa_{n,k}$, subsequently, we fix $\kappa_{n,k}$ and minimize J with respect to $\boldsymbol{\mu}_k$. The two steps can be formulated as follow:

Step I: Assign \mathbf{x}_n to the closest cluster (initial, if this is the first iteration) centroid:

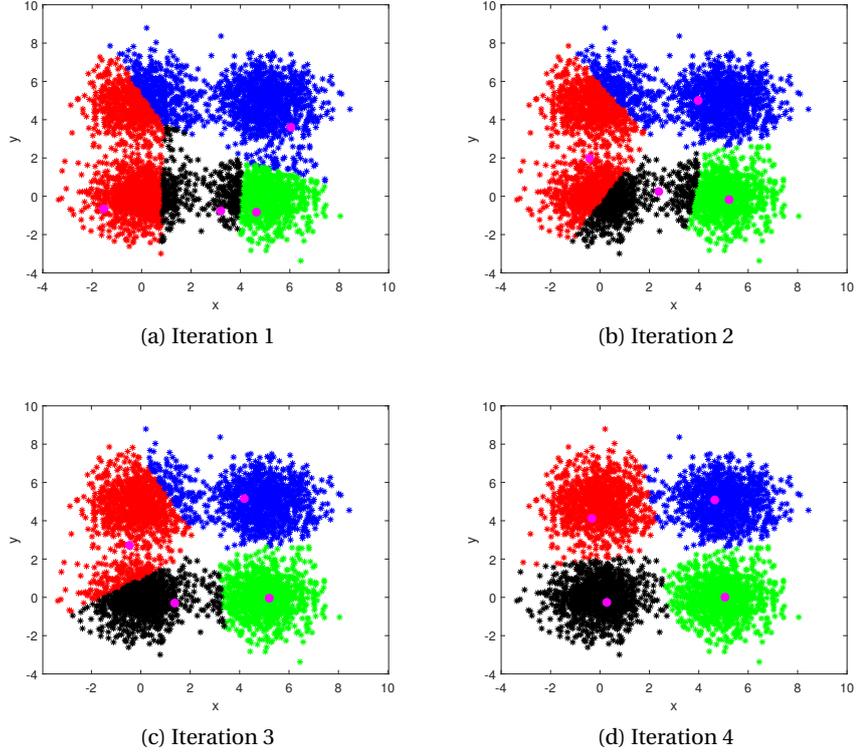


Figure 2.13: Illustration of the K-Means clustering (The magenta circles denote the initial centroids for current iteration)

$$\kappa_{n,k} = \begin{cases} 1, & \text{if } k = \underset{j}{\operatorname{argmin}} \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (2.45)$$

Step II: Update the cluster centroid:

$$\boldsymbol{\mu}_k = \frac{\sum_n \kappa_{n,k} \mathbf{x}_n}{\sum_n \kappa_{n,k}} \quad (2.46)$$

where $\sum_n \kappa_{n,k}$ is the number of points in the k th cluster. To start the iterations, we either select several point as the initial centroids randomly or just construct the initial centroids according the information extracted from the systems.

In the following, we will illustrate how the K-Means algorithm iterates from an example. Assume that there are four clusters of randomly distributed points whose centres lie in $[0, 0]$, $[0, 5]$, $[5, 0]$ and $[5, 5]$, respectively. We randomly select four points as the initial centroids and

the average mean values of clustered points serve as the initial centroids for the next iteration. As depicted in Fig. 2.13, with only four iterations, the points can be separated into four clusters with satisfactory precision.

2.4 Detection

As stated in the previous sections, the diffusion-based propagation is the most fundamental propagation in the MC and the CSK is the most energy-efficient and common in nature. In the following, we mainly present several detectors for the diffusion-based MC systems that employ binary CSK scheme.

2.4.1 Basics of the Detection

From the Fig. 2.6b, we observe that once the molecules are released, at the receiver, the received signals at each sampling instant are associated with the previously transmitted molecules (ISI), i.e., the MC channel is a memory channel. In order to quantify the probability of receiving $r_i(m)$ molecules in the m th sampling duration of the i th symbol slot, we need to take the previously transmitted symbols into consideration, and for simplicity, we only consider previous L symbols (s_{i-j} for $1 \leq j \leq L$). Assume the sampling duration is ΔT and there are M samples in each symbol slot. We also assume that the noise is constant at each sampling duration and is denoted as C_n . In this case, the $r_i(m)$ is a Poisson random variable whose mean $\bar{r}_i(m)$ is expressed as follows:

$$\bar{r}_i(m) = \sum_{j=0}^L s_{i-j} C_{jM+m} + C_n \quad (2.47)$$

Consequently, the probability for receiving $r_i(m)$ molecules is written as follow:

$$\Pr(r_i(m)) = \frac{e^{-\bar{r}_i(m)} [\bar{r}_i(m)]^{r_i(m)}}{r_i(m)!} \quad (2.48)$$

2.4.2 Single-Sample/Multiple-Sample Detectors

If we know the theoretical channel information C_j for $0 \leq j \leq LM$ and the background noise C_n , based on (2.48), we can obtain the maximum likelihood (ML) function to devise the ML-based detector which is expressed as follows:

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} \prod_{i=1}^N \prod_{m=0}^{M-1} \Pr(r_i(m)) \quad (2.49)$$

where $\mathbf{s} = [s_1, \dots, s_N]$ denotes the transmitted symbols, $\hat{\mathbf{s}} = [\hat{s}_1, \dots, \hat{s}_N]$ denotes the estimated symbols and N denotes the length of transmitted symbols. This detector is referred to as

the multiple-sample detector. It enjoys the low BER while suffers from the high calculation complexity.

In order to reduce the computational complexity of the above ML function, based on the fact that the superposition of independent Poisson random variables also follows the Poisson distribution, we can sum up all the samples in the i th symbol slot, i.e., $r_i = \sum_{m=0}^{M-1} r_i(m)$. As the i th average received signal is expressed as:

$$\bar{r}_i = MC_n + \sum_{j=0}^L s_{i-j} \sum_{m=0}^{M-1} C_{jM+m} \quad (2.50)$$

subsequently, the single-sample ML detector is given by:

$$\hat{s} = \arg \max_s \prod_{i=1}^N \frac{e^{-\bar{r}_i} \bar{r}_i^{r_i}}{r_i!} \quad (2.51)$$

2.4.3 Single-Symbol/Multiple-Symbol Detectors

The optimizations in (2.49) and (2.51) over a long sequence are in practice the multi-symbol detectors. This kind of detector is still computationally complex which may not be suitable for the MC systems, especially for the nano-machines due to the limited computation capacity. A direct solution is to detect the symbols one by one. The corresponding ML function is as follows:

$$\hat{s}_i = \arg \max_{s_i} \frac{e^{-\bar{r}_i} \bar{r}_i^{r_i}}{r_i!} = \arg \max_{s_i} \frac{e^{-(MC_n + \sum_{j=0}^L s_{i-j} \sum_{m=0}^{M-1} C_{jM+m})} [MC_n + \sum_{j=0}^L s_{i-j} \sum_{m=0}^{M-1} C_{jM+m}]^{r_i}}{r_i!} \quad (2.52)$$

To estimate the i th symbol s_i , the exact previous L symbols, i.e., s_{i-j} for $1 \leq j \leq L$, need to be known in advance which are practically unknown.

To avoid using the exact symbols, we instead exploit the previously estimated symbols, i.e., \hat{s}_{i-j} for $1 \leq j \leq L$, to replace the exact symbols in the ML function which yields the following detector:

$$\hat{s}_i = \arg \max_{s_i} \text{Pr}_{\text{appro}}(r_i | s_i) \quad (2.53)$$

where

$$\text{Pr}_{\text{appro}}(r_i | s_i) = \frac{e^{-\bar{r}_i | \hat{s}_{i-j}, 1 \leq j \leq L} [\bar{r}_i | \hat{s}_{i-j}, 1 \leq j \leq L]^{r_i}}{r_i!} \quad (2.54)$$

and

$$\bar{r}_i|_{\hat{s}_{i-j}, 1 \leq j \leq L} = M C_n + \sum_{j=1}^L \hat{s}_{i-j} \sum_{m=0}^{M-1} C_{jM+m} + s_i \sum_{m=0}^{M-1} C_m \quad (2.55)$$

In particular, we note that if fixing \hat{s}_{i-j} for $1 \leq j \leq L$, the the above ML detector is equivalent to the following:

$$\hat{s}_i = \begin{cases} 1 & r_i > \tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} \\ 0 & r_i \leq \tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} \end{cases} \quad (2.56)$$

where

$$\tau|_{\hat{s}_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln\left(1 + \frac{C_0}{\lambda_0 T + \sum_{j=1}^L \hat{s}_{i-j} C_j}\right)} \quad (2.57)$$

since if $\Pr_{\text{appro}}(r_i|s_i = 1) = \Pr_{\text{appro}}(r_i|s_i = 0)$, either binary one or binary zero can be the estimated symbol and the condition $\Pr_{\text{appro}}(r_i|s_i = 1) = \Pr_{\text{appro}}(r_i|s_i = 0)$ yields the threshold $\tau|_{\hat{s}_{i-j}, 1 \leq j \leq L}$.

2.4.4 Coherent/Non-Coherent Detectors

Based on relying on the known channel information or not, we can classify the detectors into two categories: 1) coherent detectors that need the direct channel information or have some training data for estimating the channel information or learning the input-output relationship; 2) non-coherent detectors that avoid using the channel information or can retrieve the channel information directly only from the received signals and subsequently fulfil the detection task.

In the aforementioned subsections, we have introduced some coherent detectors using the direct channel information. In chapter. 3, we exploit the NN to learn the input-output (inferring the transmitted symbols from the received signal relationship). Our aim is to figure out which traditional detector has the same BER performance as the the NN-based detector, and in turn if the NN-based detectors can provide some insights for optimizing the traditional threshold-based detectors. In the chapter. 4, we present some non-coherent detectors that can extract information only from the received signals and take the advantage of the multi-memory-bit thresholds with analytical expressions in the presence of the serve ISI. In reality, the non-coherent detectors are essential to MC systems as estimating channel is avoided, thus, saving energy.

2.5 Synchronization

The synchronization is crucial to every communication system. Imperfect synchronization would degrade the detection performance. In wireless communications, the synchronization is usually performed based on the preamble, i.e., the known transmitted signals [62, 63] before the following tasks, e.g., channel estimation (CE) [64, 41]. The channel of wireless communications is sparse such that the synchronization and CE can be performed jointly. In the MC systems, however, the channel is non-sparse [20] and is infinitely long. Therefore, it's impossible to jointly fulfil CE and synchronization tasks. As depicted in the Fig. 2.2, we see that the $H(t)$ first goes up to a maximum point and then drops gradually. Thus, we can divide the synchronization into two categories: 1) synchronization with channel information (either full CSI or the channel model) and 2) synchronization without any channel information. Basically, the synchronization is to estimate the start point of a symbol duration in the received signals.

2.5.1 Synchronization with Channel Information

There are many synchronization algorithms that are based on the full channel information, either data-aided or blind synchronization. The core idea is the same and we will present in the following.

As mentioned in the previous sections, the received signals at each time instant follow the Poisson distribution. We elucidate the rationale of the synchronization with channel information from an example. Assume that the transmitter only transmit one symbol and the corresponding received signal is $r_i(m)$ at the m th sample instant. We know that, after the release of N_{TX} molecules, the average received signal at k th sample instant is \mathbb{C}_k . The ML metric function is subsequently written as follows:

$$\Lambda = \prod_{k=0}^{LM} \frac{e^{-\mathbb{C}_k} \mathbb{C}_k^{r_i(t+k)}}{[r_i(t+k)]!} \quad (2.58)$$

and the synchronization is performed by solving the ML problem as follows:

$$\hat{t} = \operatorname{argmax}_t \Lambda = \operatorname{argmax}_t \ln \Lambda \quad (2.59)$$

if we set the release time as the start point of the transmission. Particularly, this ML metric function is maximized when $t = 0$.

It's worth pointing out that this ML problem is versatile as if we know the full channel information \mathbb{C}_k . For instance, in some MC systems, one type of molecules is employed just for transmission and another kind of molecules is used only for synchronization. Or the two types

of molecules are exploited for detection and synchronization alternatively. The ML metric function is always similar. For the MC system using the MoSK, only one kind of molecules is released at a time, the corresponding molecules can be jointly used for detection and synchronization. On the other hand, we can extend (2.58) to synchronization using multiple symbols as follows:

$$\Lambda = \prod_{i=0}^N \prod_{m=0}^M \frac{e^{-\bar{r}_i(m)} [\bar{r}_i(m)]^{r_i(t+m)}}{[r_i(t+m)]!} \quad (2.60)$$

and this ML function is maximized when $t = 0$.

Compared with the aforementioned synchronization with full channel information, the synchronization without full channel information is more realistic as the nano-machine is weak at storing too much information and has limited computation capacity. In this case, only the information extracted from the full channel information could be exploited.

The first type of extracted channel information is the maximum point of the channel response. Recall the channel model if it is known, e.g., $H(t) = N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D t}} \frac{r_0 - r_r}{t} e^{-\frac{(r_r - r_0)^2}{4Dt}}$. Assume that the MC system has K independent receivers and these receivers experience the same channel response as long as the size of the receivers are negligible. For each receiver, the received signal is given by:

$$r(t) = N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t-t_0)}} \frac{r_0 - r_r}{t-t_0} e^{-\frac{(r_r - r_0)^2}{4D(t-t_0)}} \quad (2.61)$$

after information molecules are released at time t_0 . Let the maximum received signal of the k th receiver is $r_k(t_k)$ and t_k corresponds to the time, we have

$$r_k(t_k) = N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t_k - t_0)}} \frac{r_0 - r_r}{t_k - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_k - t_0)}} \quad (2.62)$$

Collecting equations for all the receivers, we have the following equations:

$$\begin{aligned} r_1(t_1) &= N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t_1 - t_0)}} \frac{r_0 - r_r}{t_1 - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_1 - t_0)}} \\ r_2(t_2) &= N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t_2 - t_0)}} \frac{r_0 - r_r}{t_2 - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_2 - t_0)}} \\ r_3(t_3) &= N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t_3 - t_0)}} \frac{r_0 - r_r}{t_3 - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_3 - t_0)}} \\ r_4(t_4) &= N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D(t_4 - t_0)}} \frac{r_0 - r_r}{t_4 - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_4 - t_0)}} \end{aligned} \quad (2.63)$$

we can estimate t_0 using the least square (LS) method as follows:

$$\hat{t}_0 = \operatorname{argmin}_{t_0} \sum_{k=1}^M \left(r_k(t_k) - N_{TX} \frac{r_r}{r_0} \frac{1}{\sqrt{4\pi D t}} \frac{r_0 - r_r}{t_k - t_0} e^{-\frac{(r_r - r_0)^2}{4D(t_k - t_0)}} \right)^2 \quad (2.64)$$

Even though some parameters, e.g., N_{TX} may be unknown, these can still be estimated as long as there are enough receivers.

2.5.2 Synchronization without Channel Information

Despite of the existing synchronization schemes achieving good performances, these schemes rely on the full knowledge of the channel information or the specific channel model with analytical expression must exist. In order to fulfil the synchronization without channel information, some useful information generated from the **general channel model** must be obtained to avoid using the information of specific channels. Meanwhile, one challenging issue still remains, i.e., what is the optimal start point. The previous methods all assume that the release time of the molecules is the start point. In chapter 5, we provide a solution to this issue as we associate the error probability with the start point and the threshold so as to optimize jointly.

3 Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization

Chapter 3. Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization

In this chapter, we consider a molecular communication system that is made of a 3D unbounded diffusion channel model without flow, a point transmitter, and a spherical absorbing receiver. In particular, we study the impact of inter-symbol interference, and analyze the performance of different threshold-based receiver schemes. The aim of this chapter is to analyze and optimize the receivers by using the conventional model-based approach, which relies on an accurate model of the system, and the emerging data-driven approach, which, on the other hand, does not need any a priori information about the system model and exploits deep learning tools. We develop a general analytical framework for analyzing the performance of threshold-based receiver schemes, which are suitable to optimize the detection threshold. In addition, we show that data-driven receiver designs yield the same performance as receivers that have perfect knowledge of the underlying channel model.

3.1	Introduction	43
3.2	System model	45
3.3	Model-Based Receivers Design in Molecular Communications	48
3.3.1	Optimal Zero-Bit Memory Receiver	48
3.3.2	Optimal One-Bit Memory Receiver	51
3.3.3	Optimal K -Bit Memory Receiver	52
3.4	Data-Driven Receiver Design in Molecular Communications	53
3.4.1	Data-Driven Design of Zero-Bit Memory Receiver	54
3.4.2	Data-Driven Design of One-Bit Memory Receiver	55
3.4.3	Data-Driven Design of K -Bit Memory Receiver	56
3.5	Numerical Results	56
3.5.1	Zero-Bit Memory Receivers	58
3.5.2	One-Bit Memory Receivers	60
3.5.3	K -Bit Memory Receivers	60
3.6	Conclusion	62
3.7	Appendix	63
3.7.1	Proof of the BER of the one-bit memory receiver	63
3.7.2	Proof of the BER of the multi-bit memory receiver	64

3.1 Introduction

Traditional electromagnetic-based transmission techniques may not be appropriate to enable the communication among nano-devices [65]. Molecular Communications (MC) are, on the other hand, a more suitable and emerging option [9]. In a MC system, the information is transmitted via the release of information particles [9]. If the information is encoded onto the number of particles that are released, the corresponding modulation scheme is referred to as Concentration Shift Keying (CSK) modulation.

In MC systems, diffusion [66] is the easiest option to enable information particles propagate from the transmitter to the receiver. Due to the intrinsic characteristics of diffusion, the resulting transmission channel is usually affected by non-negligible Inter-Symbol Interference (ISI) which, if not taken into account for system optimization, may severely degrade the system performance [67, 39, 34, 36, 35]. The enzyme-based MC system [25] is one of the available schemes to mitigate the intrinsic ISI in MC systems. If the data rate is high, however, the ISI may not be negligible, and the approach in [25] may not provide satisfactory performance. For this reason, we focus our attention on optimizing MC systems in the presence of ISI. Developing solutions to reduce the impact of ISI is an important research topic in MC systems. For example, approaches based on modulation [68], channel coding [69], and receiver design [70] are available in the literature. In the present chapter, we focus our attention on developing robust receiver schemes.

In MC systems, a simple approach [25] to demodulate the binary symbol is to compare the number of received particles r_i with a fixed threshold τ : if $r_i \geq \tau$, the symbol is detected as 1, otherwise it is detected as 0. The threshold of this threshold-based detector is relatively simple to be optimized in the absence of ISI or if the ISI is negligible. In general, on the other hand, the threshold needs to be optimized by taking the ISI into account in order to minimize the error probability and obtain good communication performance. In [43], the authors have proposed a scheme that uses the number of particles received in the previous time-slot, i.e., r_{i-1} , as the detection threshold in a given time-slot. In [70], the authors have designed an adaptive receiver that combines a channel estimator and a decision-feedback equalizer. The channel estimator updates the channel parameters and detects the symbols constantly. Further results are available in [44]. Therein, the authors propose a new decoder that divides each slot into sub-slots. According to the number of received particles in each sub-slots, an associated decision rule is adopted and the whole scheme improves the detection performance. Similar results are available in [35].

The aforementioned approaches rely on the knowledge of the channel and system models. This, however, may not always be possible either due to the complexity of modeling the entire system in an accurate manner or due to the complexity of optimizing the resulting system

Chapter 3. Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization

model. These issues can be solved by using Machine Learning (ML) methods. With the help of ML, several schemes have been proposed [71, 72], e.g., for application to Orthogonal Frequency-Division Multiplexing (OFDM) [41], to circumvent these issues. In MC systems, ML-based schemes have been proposed in, e.g., [73, 74], to address the issue of accurate system modeling. Furthermore, the authors of [27] have recently proposed a sequence detection scheme based on sliding bidirectional recurrent neural networks that does not need channel information. Compared with existing ISI mitigation schemes, with the exception of the enzyme-based approach, ML-based schemes are less complex and easier to implement. With the aid of deep learning methods, the authors of [27] have shown that their proposed scheme is capable of automatically learning the whole system from empirical data and of performing data detection without using complex channel estimation and data equalization techniques.

In this chapter, motivated by the promising results obtained in [27], we study the possibility of optimizing the receiver design of MC systems in the presence of ISI by using Artificial Neural Networks (ANNs). In particular, our implementation is based on feed-forward ANNs with fully-connected layers. Our study shows that ANNs without prior knowledge of the system model are capable of providing the same performance as conventional detection methods that rely on the perfect knowledge of the underlying system model. In particular, the novelty and contribution of this chapter can be summarized as follows.

- We compute the Bit Error Rate (BER) of many threshold-based detection schemes. Compared with other frameworks, the proposed approach takes the background noise and the ISI into account in an accurate manner. In particular, we show that the proposed ANN-based implementation is capable of estimating the optimal threshold that minimizes the BER, as opposed to sub-optimal solutions where the dynamic nature of the ISI is not taken into account. Our approach is, therefore, particularly suitable for optimal threshold-based receiver implementations without prior knowledge of the system model.
- We develop different receiver schemes, both based on conventional detection theory and by applying recent results on data-driven optimization based on ANNs. In particular, we consider receiver structures that account for one-bit, two-bit, or all-bit (genie-aided) prior knowledge. We show that both model-based and data-driven approaches yield similar performance, with the latter approach having the benefits of not requiring any a priori information about the system model.

Compared with the companion conference version [75], this chapter is largely expanded, since it encompasses the modeling, analysis, and optimization of different types of receivers whose major difference is the number of previously-detected bits that they use for demodulation. In [75], in fact, only the zero-bit memory receiver was considered. In [75], in addition, no general framework to compute the BER was proposed. The approach proposed in this chapter, on

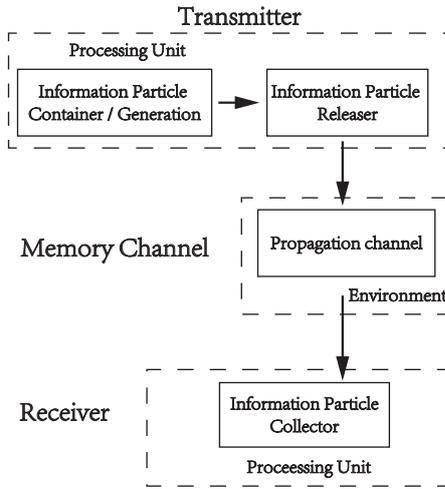


Figure 3.1: Block diagram of a typical MC system.

the other hand, can be applied to receivers with an arbitrary number of past bits used for detection. The corresponding architectures of the ANNs are proposed as well.

The remainder of this chapter is organized as follows. In Section 3.2, the system model is introduced. In Section 3.3, model-based detection schemes are proposed. In Section 3.4, data-driven detection schemes are developed and optimized. In Section 3.5, the proposed schemes are validated via numerical simulations and illustrations are provided. Finally, Section 3.6 concludes this chapter.

3.2 System model

Figure 3.1 depicts the main components of a MC system. The transmitter generates the information particles, which are released into the channel. The transmitter is assumed to be small enough to be considered as a point. We assume that the information particles diffuse randomly and independently of each other through the medium (Brownian motion). Even though a large number of information particles are emitted, not all of them reach the receiver in the considered time-slot. Some information particles remain in the channel and reach the receiver in subsequent time-slots: this causes the ISI. If not appropriately taken into account, the ISI may severely degrade the performance of MC systems. As an example, we consider a spherical absorbing-type receiver [15, 14].

We assume that the temperature is constant and the viscosity η remains unchanged during the whole transmission duration. The diffusion coefficient D [65], thus, remains constant as well. In the considered system model, no extra energy is needed since particles diffuse freely.

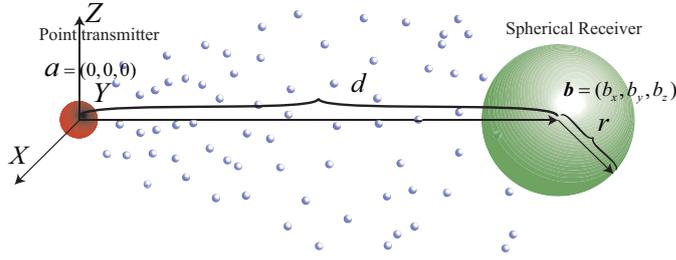


Figure 3.2: The 3D unbounded molecular channel model without flow including a point transmitter and a spherical absorbing receiver.

We consider a 3D unbounded diffusion channel model without flow, as illustrated in Fig. 3.2. By assuming the transmitter located at $\mathbf{a} = (0, 0, 0)$ and the receiver at $\mathbf{b} = (b_x, b_y, b_z)$, the hitting rate of each information particle can be expressed as follows [43, 14]:

$$f_{hit}^{3D}(t) = \frac{r(d-r)}{d\sqrt{4\pi Dt^3}} e^{-\frac{(d-r)^2}{4Dt}} \quad (3.1)$$

where $\|\mathbf{a} - \mathbf{b}\| = d$ is the distance between the transmitter and the center of the receiver, and r is the radius of the receiver that is assumed to have a spherical shape.

For ease of illustration, an On-Off Keying (OOK) modulation scheme is considered. At the i th slot, the transmitter releases N_{TX} information particles into the environment when the symbol is $s_i = 1$, otherwise the transmitter does not release any particles. We assume that the transmitter can release the N_{TX} information particles in a very short time so that the release time effect of the transmitter on the received signal is negligible.

The hitting probability of an absorbing receiver, i.e., the probability to absorb one particle after t seconds that the information particle is released, can be expressed as follows:

$$P_{hit}(t) = \int_0^t f_{hit}(t) dt \quad (3.2)$$

From (3.1) and (3.2), we have:

$$P_{hit}(t) = \int_0^t f_{hit}(t) dt = \frac{r}{d} \operatorname{erfc}\left(\frac{d-r}{\sqrt{4Dt}}\right) \quad (3.3)$$

where: $\operatorname{erf}(y) = \int_0^y \frac{2}{\sqrt{\pi}} e^{-x^2} dx$ and $\operatorname{erfc}(y) = 1 - \operatorname{erf}(y)$.

Therefore, during the $(i-1)$ th time slot after releasing the particle, the probability that one

particle hits the receiver is:

$$P_{i-1} = \int_{(i-1)T}^{iT} f_{hit}(t) dt \quad (3.4)$$

Then, we obtain:

$$P_{i-1} = \frac{r}{d} \left\{ \operatorname{erfc}\left(\frac{d-r}{\sqrt{4DiT}}\right) - \operatorname{erfc}\left(\frac{d-r}{\sqrt{4D(i-1)T}}\right) \right\} \quad (3.5)$$

Let $C_j = N_{TX} P_j$ denote the average received particles at the j th time-slot if N_{TX} particles are released. Then, the number of received particles [44] at the i th time-slot follows the Poisson distribution as follows:

$$r_i \sim \text{Poisson}(I_i + s_i C_0) \quad (3.6)$$

where $I_i = \lambda_0 T + \sum_{j=1}^{\infty} s_{i-j} C_j$ is the sum of ISI and background noise, and λ_0 is the background noise power per unit time.

More precisely, the probability of receiving r_i information particles is:

$$P(r_i | I_i + s_i C_0) = \frac{e^{-(I_i + s_i C_0)} (I_i + s_i C_0)^{r_i}}{r_i!} \quad (3.7)$$

For ease of tractability, we assume that C_i for $i > L$ is small enough to be integrated into the background noise, and we denote by L the length of the Poisson channel. Therefore, C_i for $1 \leq i \leq L$ contributes, in practice, to the ISI. The signal-to-noise ratio (SNR) can be defined as follows:

$$\text{SNR} = 10 \log_{10} \frac{C_0}{2\lambda_0 T} \quad (3.8)$$

where the information bits are assumed to be equiprobable.

Accordingly, given a certain SNR value, the number of released particles, N_{TX} , is:

$$N_{TX} = 2\lambda_0 T 10^{\frac{\text{SNR}}{10}} / P_0 \quad (3.9)$$

Table 3.1: Simulation parameters

Parameter	Value
λ_0	$100s^{-1}$
Receiver radius r	45 nm
Distance d	500 nm
Diffusion coefficient D	$4.265 * 10^{-10} m^2/s$
Discrete time length ΔT	9 us
Slot length T	$30\Delta T$
Channel length L	5

For future reference, the system parameters of a typical MC system are listed in Table 3.1.

3.3 Model-Based Receivers Design in Molecular Communications

In this section, we study some receivers in the presence of ISI with the objective of computing their bit error rate (BER) performance and optimizing their parameters in order to minimize the BER. For all cases, the system model is the same as in the previous section. We consider different types of threshold-based detectors, whose main difference consists of the amount of prior information, i.e., the number of previous bits that they use for demodulation.

3.3.1 Optimal Zero-Bit Memory Receiver

We study a threshold-based zero-bit memory receiver. The demodulation threshold is denoted by τ . Let \bar{s}_i be the estimate of symbol s_i at time-slot i . The demodulation rule can be formulated as follows:

$$\bar{s}_i = \begin{cases} 0, & r_i \leq \tau \\ 1, & r_i > \tau \end{cases} \quad (3.10)$$

The traditional approach to determine the threshold τ is obtained by imposing $P(r_i = \tau | s_i = 0) = P(r_i = \tau | s_i = 1)$. The rationale behind this approach is that the values of C_i for $1 \leq i \leq L$ are unknown, but the averaged ISI, equal to $\sum_{i=1}^L C_i/2$, is known. Under these assumptions, the probability of receiving r_i particles conditioned upon s_i can be written as follows:

$$P_{app}(r_i | s_i) = \frac{e^{-\lambda_{|s_i}} (\lambda_{|s_i})^{r_i}}{r_i!} \quad (3.11)$$

where $\lambda_{|s_i} = C_0 s_i + \frac{\sum_{j=1}^L C_j}{2} + \lambda_0 T$.

By imposing the equality $P_{app}(r_i|s_i = 0) = P_{app}(r_i|s_i = 1)$, we obtain:

$$\tau = \frac{C_0}{\ln\left(1 + \frac{C_0}{\sum_{i=1}^L C_i/2 + \lambda_0 T}\right)} \quad (3.12)$$

This approach is, however, sub-optimal. If, in fact, the slot length is long enough, then the term $\sum_{i=1}^L C_i/2$ is a good approximation for the ISI. If the slot length is short, on the other hand, the ISI changes according to the previously transmitted symbols and $\sum_{i=1}^L C_i/2$ is not a good approximation any more. Thus, the obtained demodulation threshold is not the optimal choice any more.

In the following proposition, we develop the optimal demodulation threshold that minimizes the BER. To this end, we propose also a new accurate analytical formulation of the BER. For ease of notation, we denote by $\mathbf{s}_{i-1} = \{s_{i-1}, s_{i-2}, \dots, s_{i-L}\}$ the vector of bits that are transmitted in the L time-slots preceding the i th time-slot of interest.

Proposition 3.1 *The optimal threshold that minimizes the BER of the zero-bit memory receiver is as follows:*

$$(\tau^*, P_e^*) = \underset{\tau}{\operatorname{argmin}} P_e(\tau) \quad (3.13)$$

where $P_e(\tau)$ is the BER as a function of τ :

$$P_e(\tau) = \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} P_e(\mathbf{s}_{i-1}, \tau) \quad (3.14)$$

and:

$$P_e(\mathbf{s}_{i-1}, \tau) = \frac{1}{2} \left[Q(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j, \lceil \tau \rceil) + 1 - Q(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0, \lceil \tau \rceil) \right] \quad (3.15)$$

Proof 3.1 *The BER is defined as follows:*

$$P_e(\mathbf{s}_{i-1}, \tau) = \frac{1}{2} [P(r_i \geq \tau | s_i = 0, \mathbf{s}_{i-1}) + P(r_i < \tau | s_i = 1, \mathbf{s}_{i-1})] \quad (3.16)$$

Chapter 3. Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization

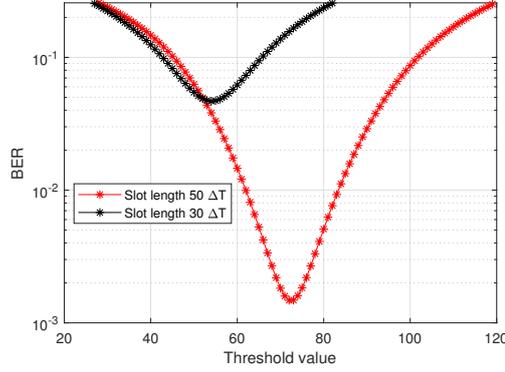


Figure 3.3: BER in (3.14) as a function of τ (the SNR is 30 dB).

where:

$$\begin{aligned}
 P(r_i \geq \tau | s_i = 0, \mathbf{s}_{i-1}) &= P(r_i \geq \tau | \lambda_0 T + \sum_{j=1}^L s_{i-j} C_j) = \sum_{k=\lceil \tau \rceil}^{\infty} \frac{e^{-(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j)} (\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j)^k}{k!} \\
 &= Q(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j, \lceil \tau \rceil)
 \end{aligned} \tag{3.17}$$

where $Q(\lambda, n) = \sum_{k=n}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!}$ is the incomplete Gamma function and $Q(\lambda, 0) = 1$. Similarly, we have:

$$\begin{aligned}
 P(r_i < \tau | s_i = 1, \mathbf{s}_{i-1}) &= P(r_i < \tau | \lambda_0 T + \sum_{j=0}^L s_{i-j} C_j) = \sum_{k=0}^{\lceil \tau \rceil - 1} \frac{e^{-(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0)} (\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0)^k}{k!} \\
 &= 1 - \sum_{k=\lceil \tau \rceil}^{\infty} \frac{e^{-(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0)} (\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0)^k}{k!} = 1 - Q(\lambda_0 T + \sum_{j=1}^L s_{i-j} C_j + C_0, \lceil \tau \rceil)
 \end{aligned} \tag{3.18}$$

From (3.17) and (3.18), we obtain (3.15). Finally, the BER is obtained by averaging (3.14) with respect to the vector \mathbf{s}_{i-1} .

The optimal detection threshold, τ , is obtained by minimizing the BER (see (3.13)). In Fig. 3.3, we depict (3.14) as a function of τ . We observe that an optimal value of τ exists that minimizes the BER and that it depends on the time slot duration T , i.e., the amount of ISI.

3.3.2 Optimal One-Bit Memory Receiver

In this section, we study and optimize the performance of a one-bit memory receiver, which has more prior information than the zero-bit memory receiver. The receiver can be formulated as follows:

$$\bar{s}_i = \begin{cases} 0, & r_i \leq \tau|_{s_{i-1}} \\ 1, & r_i > \tau|_{s_{i-1}} \end{cases} \quad (3.19)$$

where $\tau|_{s_{i-1}}$ denotes the threshold for the i th symbol when the previously transmitted symbol is s_{i-1} . Since the exact value of s_{i-1} is unknown, the estimate \bar{s}_{i-1} is employed instead, i.e., $\tau|_{\bar{s}_{i-1}}$ is used.

In simple terms, in contrast to the zero-bit memory receiver that accounts only for the number of received particles in the time-slot of interest, the one-bit memory detector adapts the detection threshold as a function of the previously transmitted bit (that is, in practice, replaced by its estimate). Therefore, the detection threshold changes from time-slot to time-slot, but a better approximation of the ISI is obtained.

The following proposition yields the optimal value of the detection threshold, by using the same line of thought as for the zero-bit memory detector.

Proposition 3.2 *The optimal detection threshold of the one-bit memory receiver can be formulated as follows:*

$$\tau^*|_{s_{i-1}} = \underset{\tau}{\operatorname{argmin}} P_e(\tau, s_{i-1}) \quad (3.20)$$

where the BER is as follows:

$$P_e(\tau, s_{i-1}) = \frac{1}{2^{L-1}} \sum_{s_{i-2}, \dots, s_{i-L}} P_e(\mathbf{s}_{i-1}, \tau) \quad (3.21)$$

In order to compute the optimal threshold for the i th time-slot, the previously transmitted symbol s_{i-1} is assumed to be known. In practice, this is not possible, since only its estimates is available, as discussed already. Therefore, the BER needs to take this into account. The BER of the one-bit memory receiver is given in the following theorem.

Chapter 3. Molecular Communications: Model-Based and Data-Driven Receiver Design and Optimization

Theorem 3.1 *The BER of the one-bit memory detector can be formulated as follows:*

$$P_e = \frac{m+n}{2} \quad (3.22)$$

where m and n are the solutions of the following equations:

$$m = \frac{1}{2^L} \sum_{s_{i-1}} \sum_{\bar{s}_{i-1}} Q(\lambda|_{s_{i-1}, s_i=0}, \lceil \tau|_{\bar{s}_i} \rceil) \Psi(s_{i-1}, \bar{s}_{i-1}, m, n) \quad (3.23)$$

$$n = \frac{1}{2^L} \sum_{s_{i-1}} \sum_{\bar{s}_{i-1}} (1 - Q(\lambda|_{s_{i-1}, s_i=1}, \lceil \tau|_{\bar{s}_i} \rceil)) \Psi(s_{i-1}, \bar{s}_{i-1}, m, n) \quad (3.24)$$

where $\tau|_{\bar{s}_{i-1}}$ is the optimal threshold that corresponds to the previously detected bit \bar{s}_{i-1} , $\lambda|_{s_{i-1}, s_i=0}$ is the average number of received particles by conditioning on the current symbol being 0 and the previous L symbols being s_{i-1} , i.e., $\lambda|_{s_{i-1}, s_i=0} = \sum_{j=1}^L C_j s_{i-j} + \lambda_0 T$. Similarly, $\lambda|_{s_{i-1}, s_i=1} = \sum_{j=1}^L C_j s_{i-j} + \lambda_0 T + C_0$. The function $\Psi(s_{i-1}, \bar{s}_{i-1}, m, n)$ is defined as follows:

$$\Psi(s_{i-1}, \bar{s}_{i-1}, m, n) = \begin{cases} m, & (s_{i-1} = 0, \bar{s}_{i-1} = 1) \\ 1 - m, & (s_{i-1} = 0, \bar{s}_{i-1} = 0) \\ n, & (s_{i-1} = 1, \bar{s}_{i-1} = 0) \\ 1 - n, & (s_{i-1} = 1, \bar{s}_{i-1} = 1) \end{cases} \quad (3.25)$$

Proof 3.2 *See Appendix 3.7.1.*

3.3.3 Optimal K -Bit Memory Receiver

Inspired by the one-bit memory detector in Section 3.3.2, we generalize this receiver design by considering a generic K -bit memory receiver. It is worth nothing that K may be set equal to L , which is the actual length of the ISI channel. This setup yields the optimal performance but needs more a priori information on the previously detected bits, which increases the complexity of the receiver.

The optimal detection threshold and BER are given in the following proposition and theorem, respectively.

Proposition 3.3 *The optimal detection threshold of the K -bit memory receiver can be formu-*

3.4. Data-Driven Receiver Design in Molecular Communications

lated as follows:

$$\tau^*|_{s_{i-1}, \dots, s_{i-K}} = \underset{\tau}{\operatorname{argmin}} P_e(\tau, s_{i-1}, \dots, s_{i-K}) \quad (3.26)$$

where the BER is as follows:

$$P_e(\tau, s_{i-1}, \dots, s_{i-K}) = \frac{1}{2^{L-K}} \sum_{s_{i-K+1}, \dots, s_{i-L}} P_e(\mathbf{s}_{i-1}, \tau) \quad (3.27)$$

Since the exact symbols s_{i-j} , for $1 \leq j$ are unknown, the estimates $\bar{s}_{i-1}, \dots, \bar{s}_{i-K}$ are used to perform the detection:

$$\bar{s}_i = \begin{cases} 0, & r_i \leq \tau|_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} \\ 1, & r_i > \tau|_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} \end{cases} \quad (3.28)$$

Theorem 3.2 *The BER of the K -bit memory receiver can be approximated by using (3.22), where m and n are the solution of the equations:*

$$m = \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} \sum_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} Q(\lambda|_{\mathbf{s}_{i-1}, s_i=0, \lceil \tau|_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} \rceil}) \prod_{j=1}^K \Psi(s_{i-j}, \bar{s}_{i-j}, m, n)$$

$$n = \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} \sum_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} (1 - Q(\lambda|_{\mathbf{s}_{i-1}, s_i=1, \lceil \tau|_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} \rceil})) \prod_{j=1}^K \Psi(s_{i-j}, \bar{s}_{i-j}, m, n)$$

Proof 3.3 *See Appendix 3.7.2. It is worth mentioning that the obtained expression of the BER is an approximation for general values of K . The details of the approximation are available in the appendix.*

3.4 Data-Driven Receiver Design in Molecular Communications

In the previous section, we have optimized the operation of receivers by assuming that the underlying system model is perfectly known. In this section, we do not rely on this assumption anymore, and take advantage of feed-forward ANNs with fully-connected layers and deep learning [76] to optimize the design of molecular receivers. The architecture of a typical feed-forward ANN with fully-connected layers is depicted in Fig 3.4, and it consists of an input layer, several hidden layers, and an output layer. The nodes of the hidden layers are referred to as neurons.

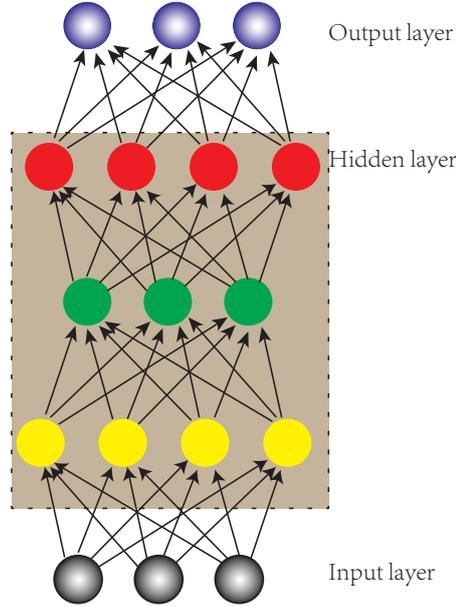


Figure 3.4: Typical structure of a feed-forward ANN with fully-connected layers.

The objective of this section is to describe how to design and optimize receivers by using ANNs and by training an ANN by using only empirical data, i.e., the number of received particles in the presence of ISI. We describe each receiver studied in the previous section in the following sub-sections.

3.4.1 Data-Driven Design of Zero-Bit Memory Receiver

The objective of an ANN-based design is to identify an ANN structure that demodulates the transmitted data by minimizing the BER. An ANN-based zero-bit memory demodulator is a system whose input consists of the received information particles r_i at the i th time-slot, and the outputs are the probabilities that the transmitted bit is 0 or 1, i.e., $P_i(s_i = 0|r_i)$ and $P_i(s_i = 1|r_i)$, respectively. Since, $P_i(s_i = 1|r_i) + P_i(s_i = 0|r_i) = 1$, only one of the two probabilities is needed. In the sequel, we use the notation $P_i = P_i(s_i = 1|r_i)$. Based on the inputs, the ANN demodulate the received data as follows:

$$\bar{s}_i = \begin{cases} 0, & P_i \leq 0.5 \\ 1, & P_i > 0.5 \end{cases} \quad (3.29)$$

where the threshold 0.5 accounts for the fact that the bits are equiprobable.

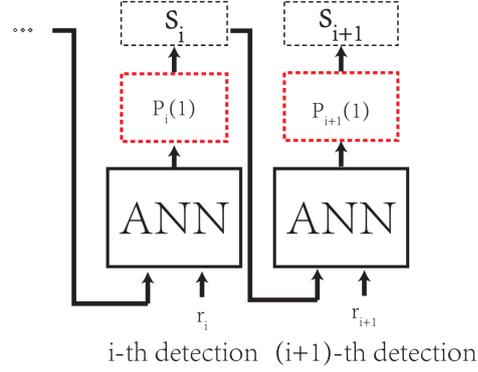


Figure 3.5: Data-driven one-bit memory receiver.

In order to train the ANN, we consider a supervised learning approach, i.e., we compute the parameters (e.g., the bias factors and the weights) of the ANN by using a known sequence of transmitted bits. In particular, we use the Bayesian regularization back propagation technique, which updates the weights and biases by using the Levenberg-Marquardt optimization algorithm. A hyperbolic tangent sigmoid activation function is employed in all neurons. The set of parameters that are used to train and operate the ANN are the following: The number of hidden layers is 10, the number of neurons per layer is 5, the learning rate is 0.01, the training epoch is 200, the total number of training bits is 50000, and the number of test bits is 100000. In particular, the training is performed in a batch mode, and the number of bits in each batch is 1000. This setup is used to obtain the numerical results in the next section.

3.4.2 Data-Driven Design of One-Bit Memory Receiver

If the one-bit memory receiver is considered, the input of the ANN is not just the number of received particles at the i th time-slot, r_i , but also the estimated symbol at the $(i - 1)$ th time-slot, \bar{s}_{i-1} . In mathematical terms, the output estimate of the ANN can be formulated as follows:

$$\bar{s}_i = \begin{cases} 0, & P(s_i = 1 | r_i, \bar{s}_{i-1}) \leq 0.5 \\ 1, & P(s_i = 1 | r_i, \bar{s}_{i-1}) > 0.5 \end{cases}$$

A block diagram representation of the ANN-based architecture is depicted in Fig. 3.5.

As far as the ANN architecture is concerned, the same system setup as for the zero-bit memory receiver is considered with the only exception that the number of hidden layers is equal to 5 and the number of neurons per layer is 4.

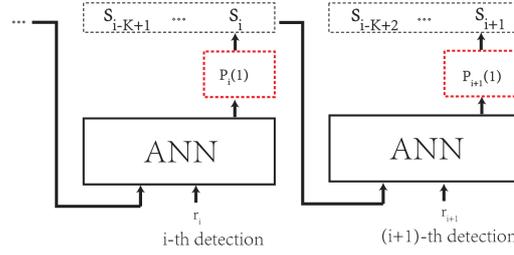


Figure 3.6: Data-driven K -bit memory receiver.

3.4.3 Data-Driven Design of K -Bit Memory Receiver

By using the same line of thought as for the one-bit memory receiver, the decision rule of the K -bit memory receiver can be formulated as follows:

$$\bar{s}_i = \begin{cases} 0, & P(s_i = 1 | r_i, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) \leq 0.5 \\ 1, & P(s_i = 1 | r_i, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) > 0.5 \end{cases}$$

The corresponding block diagram is illustrated in Fig. 3.6.

In this case, in particular, the training data-unit is constituted by the vector $\{r_i, s_{i-1}, \dots, s_{i-K}; s_i\}$. The same system setup as for the one-bit memory receiver is considered to obtain the numerical results illustrated in the next section.

3.5 Numerical Results

In this section, we report and describe some simulation results in order to validate the analysis, design, and optimization of the proposed receivers for application to molecular communications. In addition, we compare both model-based and data-driven designs.

As far as Monte Carlo simulations are concerned, the MC system is assumed to be perfectly synchronized. Accordingly, the hitting rate at each ΔT can be obtained directly from (3.1), and the number of received particles can be, thus, computed from (3.6) without the need of implementing particle-based Monte Carlo simulations. This approach reduces the simulation time without compromising, under the considered system model, the accuracy of the results.

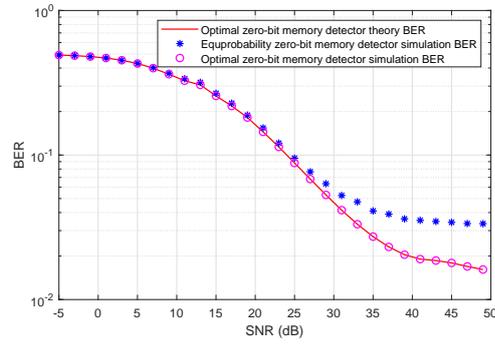


Figure 3.7: BER of optimal vs. conventional zero-bit memory receiver - $T = 30\Delta T$.

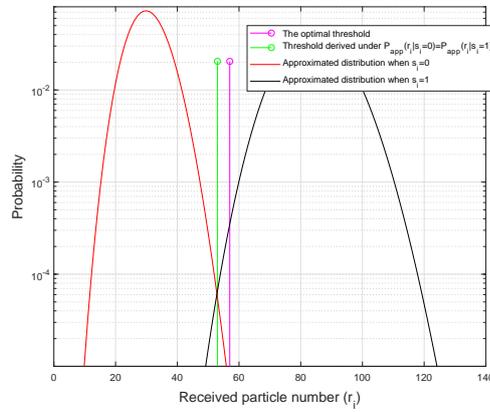


Figure 3.8: Approximated distributions of the received bits from (3.11) (the SNR is 25 dB).

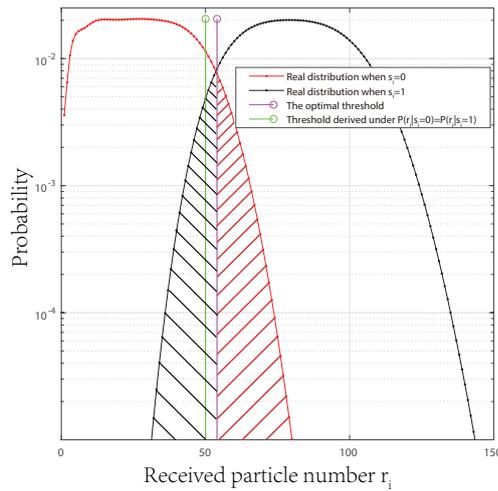


Figure 3.9: Empirical distributions of the received bits (the SNR is 25 dB).

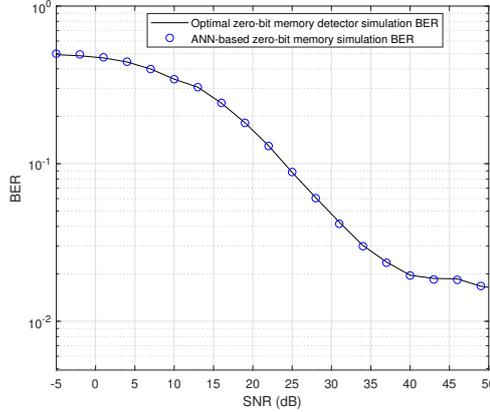


Figure 3.10: The ANN-based receiver achieves the same BER performance as the optimal zero-bit receiver - $T = 30\Delta T$.

3.5.1 Zero-Bit Memory Receivers

In Fig. 3.7, we observe that the proposed design based on optimizing the detection threshold that minimizes the BER provides us with better performance than the sub-optimal design. We note, in particular, that for each SNR the optimal detection threshold is used. We observe, in addition, a very good accuracy of the proposed analytical framework. Notably, Fig. 3.8 and Fig. 3.9 provide us with a simple representation of the receiver sub-optimality discussed in Sec. 3.3.1. First of all, we observe that the theoretical and empirical distributions of the received number of particles are different. More importantly, we observe that the empirical distributions cross each other in correspondence of the estimated optimal detection threshold, while the approximated ones cross each other in a different point. This justifies the reason why our approach yields the optimum and a better BER.

The results shown in Fig. 3.8 and Fig. 3.9 are, therefore, very important in order to highlight the sub-optimality of the demodulation thresholds that have been used in the literature to date. The results in Fig. 3.7, in addition, highlight the advantages of the proposed optimal thresholds in the context of MC systems design and optimization.

In Fig. 3.10, we compare the BER of the ANN-based demodulator against the model-based receiver design. We observe a good accuracy, which confirms the correct optimization of the ANN, and, at the same time, the correct calculation of the analytical framework. In Fig. 3.11, we compare the optimal threshold computed numerically from (3.13) as a function of the SNR, and the demodulation threshold that is learned by the ANN-based demodulator. In the latter case, the threshold is obtained, after completing the training of the ANN, and identifying the input, i.e., the number of information particles, for which the output probability is equal to 0.5. We observe that the ANN-based implementation is capable of learning the demodulation

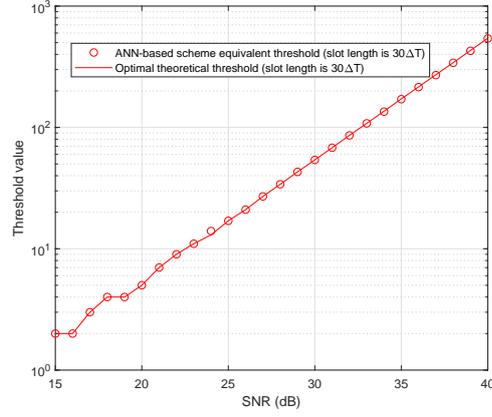


Figure 3.11: Optimal detection thresholds of model-based and data-driven schemes. The ANN-based receiver automatically emulates the optimal zero-bit memory receiver. $-T = 30\Delta T$.

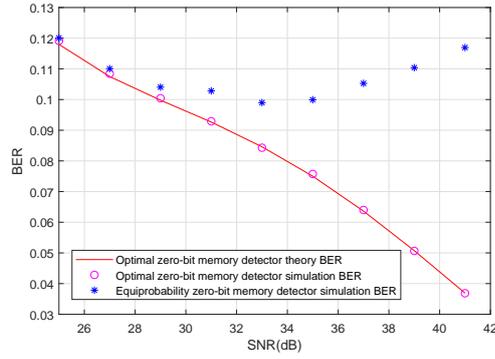


Figure 3.12: System model assuming an observing receiver. BER of the optimal vs. sub-optimal (i.e., based on the sub-optimal threshold) zero-bit memory receiver - $T = 30\Delta T$.

threshold in a very accurate manner. This result is very interesting, as it allows us to unveil the hidden behavior of the optimized ANN. It highlights, in particular, that the optimized ANN is, indeed, a threshold-based demodulator.

In order to further test the robustness of our ANN-based design, we consider another case study where an observing receiver is considered [77]. In this case, the hitting rate can be formulated as follows:

$$f_{\text{observing}}(t) = \frac{1}{2} [\text{erf}(\tau_1) + \text{erf}(\tau_2)] - \frac{\sqrt{Dt}}{d\sqrt{\pi}} [e^{-\tau_1^2} - e^{-\tau_2^2}]$$

where $\tau_1 = \frac{r+d}{\sqrt{4Dt}}$ and $\tau_2 = \frac{r-d}{\sqrt{4Dt}}$.

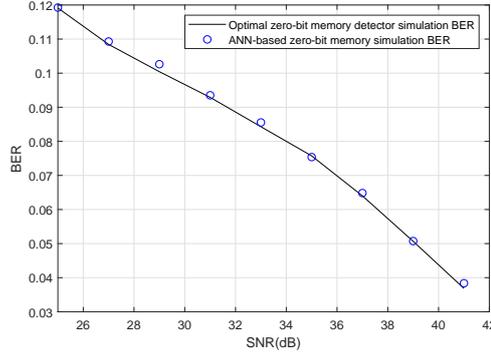


Figure 3.13: System model assuming an observing receiver. The data-driven receiver achieves the same BER performance as the optimal (i.e., based on the optimal threshold) zero-bit receiver - $T = 30\Delta T$.

The rest of the equations can be obtained from this new expression of the hitting rate. As for the simulation parameters, they are the same as those in Table 3.1, with the exception of the distance $d = 100nm$ and the channel length $L = 2$.

The corresponding results are illustrated in Fig. 3.12 and Fig. 3.13. We observe that similar performance trends as for the first case study are obtained. There exists a gap between the BER of optimal and sub-optimal zero-bit receivers. In the following sections, therefore, we will consider only the channel where the distance between the transmitter and the receiver is large.

3.5.2 One-Bit Memory Receivers

In Fig. 3.14, we compare the BER of the optimal and sub-optimal one-bit threshold receivers. Also in this case, we observe that better performance is obtained by using the proposed optimal design. In addition, the numerical results confirm the correctness of our analytical framework. In Fig.3.15, we observe that the proposed ANN-based design yields the same results as the model-based approach, which, however, assumes perfect knowledge of the system model.

3.5.3 K-Bit Memory Receivers

In this section, finally, we consider the design of receivers that exploit more than one bit for improving the performance.

In Fig. 3.16, we compare the BER of the optimal and sub-optimal receivers by assuming $K = 2$.

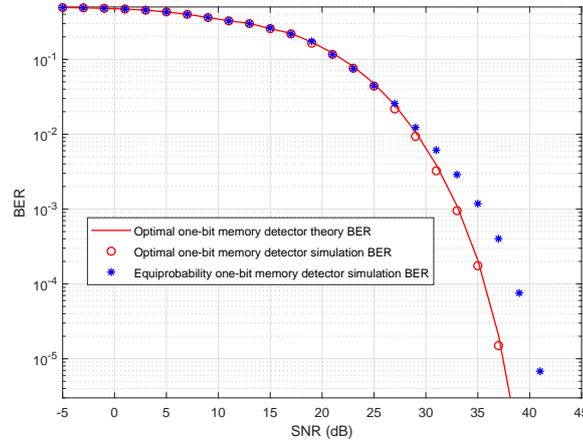


Figure 3.14: BER of optimal vs. sub-optimal one-bit memory receiver. The performance gap between optimal and sub-optimal receiver gets smaller since the ISI is modeled more accurately - $T = 30\Delta T$.

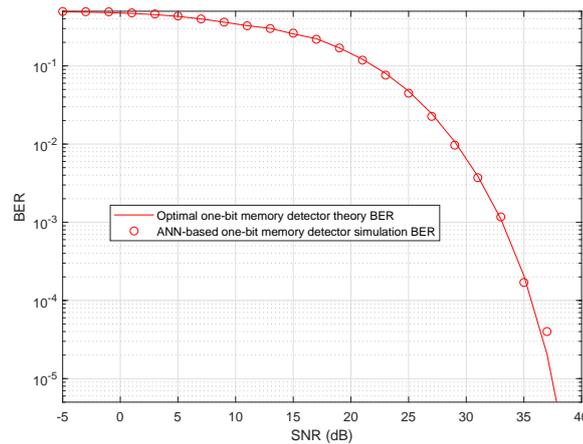


Figure 3.15: The ANN-based receiver achieves the same BER performance as the optimal one-bit receiver - $T = 30\Delta T$.

We observe that, in this case, the sub-optimal receiver is closer to the optimal one, if compared to the case studies with $K = 0$ and $K = 1$. We have observed, in general, that the larger the number of bits of memory is, the closer the BER of optimal and sub-optimal receivers are.

In Fig. 3.17 and Fig. 3.18, we compare model-based and ANN-based receiver designs, and we observe a good agreement. In particular, Fig. 3.18 highlights the improved performance that is obtained by increasing the number of bits of memory, which, for the considered setup, is $K = L$, i.e., the actual length of the ISI.

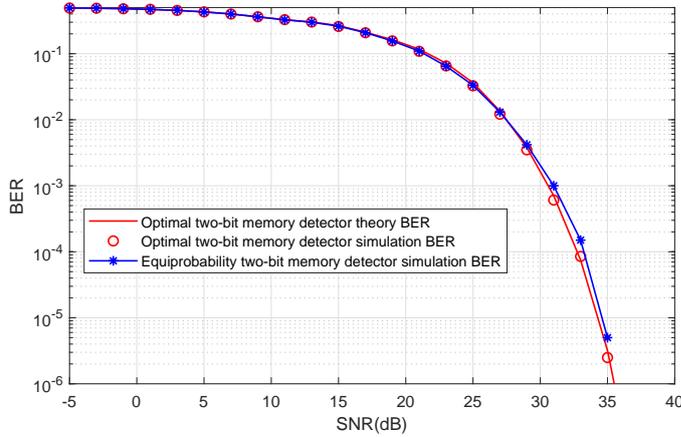


Figure 3.16: Two-bit memory detector: Comparison between the optimal and sub-optimal setups of the demodulation thresholds. The more the number of memory bits, the better the ISI is modeled. As the memory length approaches the channel length, thus, the optimal threshold converges towards the conventional threshold - $T = 30\Delta T$.

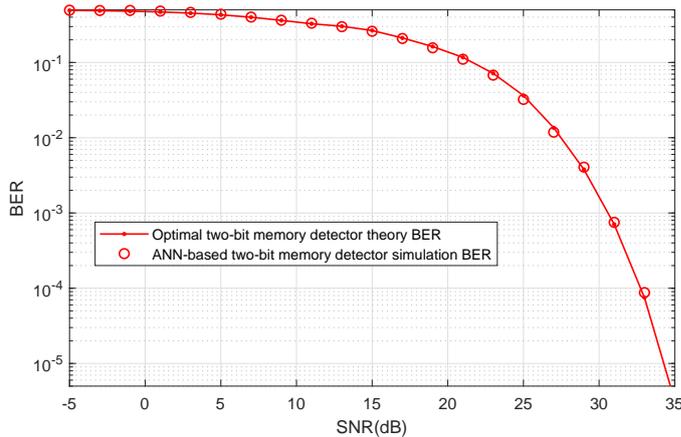


Figure 3.17: The ANN-based two-bit memory receiver achieves the same BER performance as the optimal two-bit receiver - $T = 30\Delta T$.

3.6 Conclusion

In this chapter, we have introduced a new analytical framework to compute the BER of a MC system that uses threshold-based demodulators. By modeling the receiver as an ANN, in addition, we have proved that data-driven receivers provide similar performance as those that are optimized based on the exact knowledge of the channel model. From the considered ANN-based receiver design, in addition, we have shown that the resulting ANN architecture results in a threshold-based receiver whose threshold coincides with that predicted theoretically. This is an interesting result for better optimizing and further understanding MC systems.

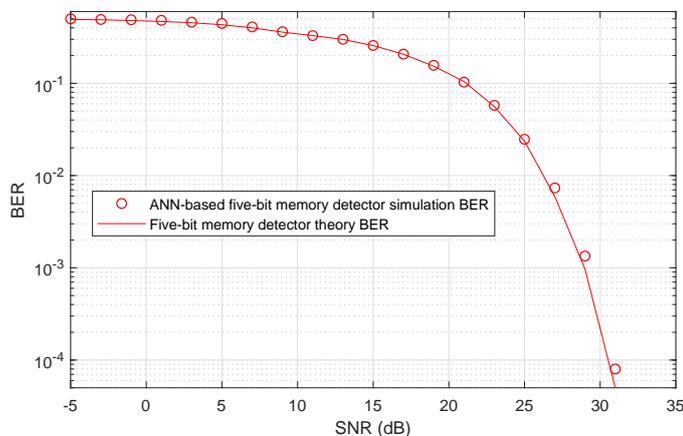


Figure 3.18: The BER performance of ANN-based and model-based L-bit memory detectors overlap with each other - $T = 30\Delta T$.

3.7 Appendix

3.7.1 Proof of the BER of the one-bit memory receiver

The BER is defined as follows:

$$P_e = \frac{1}{2} [P(\bar{s}_i = 1|s_i = 0) + P(\bar{s}_i = 0|s_i = 1)] \quad (3.30)$$

where $P(\bar{s}_i = 1|s_i = 0)$ is the probability of detecting s_i equal to 1 when the transmitted symbol is 0. We have the following:

$$\begin{aligned} P(\bar{s}_i = 1|s_i = 0) &= \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(\bar{s}_i = 1|s_i = 0, \mathbf{s}_{i-1}, \bar{s}_{i-1})P(\bar{s}_{i-1}, s_{i-1}, \dots, s_{i-L}) \\ &= \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(r_i \geq \tau|\bar{s}_{i-1}|s_i = 0, \mathbf{s}_{i-1}, \bar{s}_{i-1})P(\bar{s}_{i-1}|s_{i-1})P(s_{i-1})P(s_{i-2}, \dots, s_{i-L}) \\ &= \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(\bar{s}_{i-1}|s_{i-1})Q(\lambda|_{s_{i-1}, s_i=0}, \lceil \tau|\bar{s}_{i-1} \rceil) \end{aligned} \quad (3.31)$$

where $\lambda|_{s_{i-1}, s_i=0} = \lambda_0 T + \sum_{j=1}^L s_{i-j} C_j$. By using similar steps, we obtain:

$$\begin{aligned} P(\bar{s}_i = 0|s_i = 1) &= \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(\bar{s}_i = 0|s_i = 1, \mathbf{s}_{i-1}, \bar{s}_{i-1})P(\bar{s}_{i-1}, s_{i-1}, \dots, s_{i-L}) \\ &= \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(r_i < \tau|\bar{s}_{i-1}|s_i = 1, \mathbf{s}_{i-1}, \bar{s}_{i-1})P(\bar{s}_{i-1}|s_{i-1})P(s_{i-1})P(s_{i-2}, \dots, s_{i-L}) \\ &= \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}, \bar{s}_{i-1}} P(\bar{s}_{i-1}|s_{i-1})(1 - Q(\lambda|_{s_{i-1}, s_i=1}, \lceil \tau|\bar{s}_{i-1} \rceil)) \end{aligned} \quad (3.32)$$

The proof follows.

3.7.2 Proof of the BER of the multi-bit memory receiver

From (3.30), we have the following:

$$\begin{aligned}
 & P(\bar{s}_i = 1 | s_i = 0) \\
 &= \sum_{s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_i = 1 | s_i = 0, s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) P(s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) \\
 &= \sum_{s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_i = 1 | s_i = 0, s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) P(\bar{s}_{i-1} | s_{i-1}, \dots, s_{i-L}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K}) P(s_{i-1}) \\
 &\quad P(s_{i-2}, \dots, s_{i-L}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K}) \tag{3.33} \\
 &= \sum_{s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_i = 1 | s_i = 0, s_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) P(\bar{s}_{i-1} | s_{i-1}, \dots, s_{i-L}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K}) P(s_{i-1}) \\
 &\quad P(\bar{s}_{i-2} | s_{i-2}, \dots, s_{i-L}, \bar{s}_{i-3}, \dots, \bar{s}_{i-K}) P(s_{i-2}) \cdots P(\bar{s}_{i-K} | s_{i-K}, \dots, s_{i-L}) P(s_{i-K}) P(s_{i-K+1}, \dots, s_{i-L})
 \end{aligned}$$

The term $P(\bar{s}_{i-1} | s_{i-1}, \dots, s_{i-L}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K})$ can be calculated as follows:

$$\begin{aligned}
 & P(\bar{s}_{i-1} | s_{i-1}, \dots, s_{i-L}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K}) \\
 &= \sum_{s_{i-L-1}, \bar{s}_{i-K-1}} P(\bar{s}_{i-K-1} | s_{i-K-1}) P(s_{i-L-1}) P(\bar{s}_{i-1} | s_{i-1}, \dots, s_{i-L-1}, \bar{s}_{i-2}, \dots, \bar{s}_{i-K-1}) \tag{3.34}
 \end{aligned}$$

Also, $P(\bar{s}_{i-2} | s_{i-2}, \dots, s_{i-L}, \bar{s}_{i-3}, \dots, \bar{s}_{i-K})$ can be obtained as follows:

$$\begin{aligned}
 & P(\bar{s}_{i-2} | s_{i-2}, \dots, s_{i-L}, \bar{s}_{i-3}, \dots, \bar{s}_{i-K}) \\
 &= \sum_{s_{i-L-1}, s_{i-L-2}, \bar{s}_{i-K-1}, \bar{s}_{i-K-2}} P(\bar{s}_{i-K-2} | s_{i-K-2}) P(\bar{s}_{i-K-1} | s_{i-K-1}, s_{i-K-2}, \bar{s}_{i-K-2}) P(s_{i-L-1}) P(s_{i-L-2}) \tag{3.35} \\
 & P(\bar{s}_{i-2} | s_{i-2}, \dots, s_{i-L-2}, \bar{s}_{i-3}, \dots, \bar{s}_{i-K-2})
 \end{aligned}$$

To obtain a tractable closed-form expression, we use the following approximation to compute, recursively, $P(\bar{s}_i = 1 | s_i = 0)$:

$$\begin{aligned}
 & P(\bar{s}_i = 1 | s_i = 0) \\
 &= \frac{1}{2} \sum_{s_{i-1}, \bar{s}_{i-1}} P(\bar{s}_{i-1} | s_{i-1}) P(\bar{s}_i = 1 | s_i = 0, s_{i-1}, \bar{s}_{i-1}) \\
 &\approx \frac{1}{2^2} \sum_{s_{i-1}, \bar{s}_{i-1}} P(\bar{s}_{i-1} | s_{i-1}) \sum_{s_{i-2}, \bar{s}_{i-2}} P(\bar{s}_{i-2} | s_{i-2}) P(\bar{s}_i = 1 | s_i = 0, s_{i-1}, \bar{s}_{i-1}, s_{i-2}, \bar{s}_{i-2})
 \end{aligned}$$

$$\approx \frac{1}{2^L} \sum_{s_{i-1}, \bar{s}_{i-1}} P(\bar{s}_{i-1}|s_{i-1}) \dots \sum_{s_{i-K}, \bar{s}_{i-K}} P(\bar{s}_{i-K}|s_{i-K}) P(\bar{s}_i = 1 | s_i = 0, \mathbf{s}_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) \quad (3.36)$$

The calculation of $P(\bar{s}_i = 1 | s_i = 0, \mathbf{s}_i, \bar{s}_{i-1}, \dots, \bar{s}_{i-K})$ can be done for any threshold, $\tau |_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}}$, and from the average number of received particles as follows:

$$\lambda_{\mathbf{s}_{i-1}, s_i=0} = \lambda_0 T + \sum_{j=1}^L s_{i-j} C_j.$$

From (5.8), we obtain:

$$P(\bar{s}_i = 1 | s_i = 0) = \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} \sum_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_{i-1}|s_{i-1}) \dots P(\bar{s}_{i-K}|s_{i-K}) Q(\lambda_{\mathbf{s}_{i-1}, s_i=0}, [\tau |_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}}]) \quad (3.37)$$

$P(\bar{s}_i = 0 | s_i = 1)$ can be computed by using similar steps and assumptions:

$$P(\bar{s}_i = 0 | s_i = 1) \approx \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} \sum_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_{i-1}|s_{i-1}) \dots P(\bar{s}_{i-K}|s_{i-K}) P(\bar{s}_i = 0 | s_i = 1, \mathbf{s}_{i-1}, \bar{s}_{i-1}, \dots, \bar{s}_{i-K}) \quad (3.38)$$

where:

$$\lambda_{\mathbf{s}_i, s_i=1} = C_0 + \lambda_0 T + \sum_{j=1}^L s_{i-j} C_j$$

Finally, we obtain the following:

$$P(\bar{s}_i = 0 | s_i = 1) = \frac{1}{2^L} \sum_{\mathbf{s}_{i-1}} \sum_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}} P(\bar{s}_{i-1}|s_{i-1}) \dots P(\bar{s}_{i-K}|s_{i-K}) (1 - Q(\lambda_{\mathbf{s}_{i-1}, s_i=1}, [\tau |_{\bar{s}_{i-1}, \dots, \bar{s}_{i-K}}])) \quad (3.39)$$

This concludes the proof.

4 K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

Chapter 4. K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

In this chapter, we consider non-coherent detection for molecular communication systems in the presence of inter-symbol-interference. In particular, we study non-coherent detectors based on memory-bits-based thresholds in order to achieve low bit-error-ratio (BER) transmission. The main challenge of realizing detectors based on memory-bits-based thresholds is to obtain the channel state information based only on the received signals. We tackle this issue by reformulating the thresholds through intermediate variables, which can be obtained by clustering multi-dimensional data from the received signals, and by using the K-means clustering algorithm. In addition to estimating the thresholds, we show that the transmitted bits can be retrieved from the clustered data. To reduce clustering errors, we propose iterative clustering methods from one-dimensional to multi-dimensional data, which are shown to reduce the BER. Simulation results are presented to verify the effectiveness of the proposed methods.

4.1	Introduction	69
4.1.1	Motivation	69
4.1.2	Related Works	70
4.1.3	Contributions	70
4.1.4	Chapter Organization	71
4.2	System model	71
4.3	Multi-memory-bit Threshold Reformulation	72
4.3.1	Estimation of $\hat{r}_i _{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the Received Data	75
4.4	Clustering-based Non-coherent Detection	77
4.4.1	K-Means Clustering Algorithm	77
4.4.2	Single-Dimensional Clustering: Challenges and Limitations	78
4.4.3	Multi-Dimensional Clustering	79
4.4.4	Setup of the Initial Centroids	80
4.4.5	Direct Clustering-Based Inference and Clustering-plus-Threshold Detection	82
4.5	Iterative Algorithm for Computing the Initial Centroids	86
4.6	Numerical Results	89
4.7	Conclusion	90

4.1 Introduction

Recent developments in biology and nano-technology have gained attention in micro-scale communications in light of emerging potential applications, e.g., bio-robots [78]. In this context, molecular communication (MC) is regarded as a promising solution [9] because of the difficulty of using traditional electromagnetic-based techniques [65]. In a MC system, the information is disseminated via small particles [9]. Among the different modulation schemes that can be used in MC systems [79], binary concentration shift keying (CSK) modulation is the simplest method for encoding information onto the number of released particles. In addition, diffusion via Brownian motion [66] is the most common solution for allowing information particles propagate from a transmitter to a receiver.

4.1.1 Motivation

In diffusion-based MC systems, there exist several issues to be solved. A major issue is the non-negligible inter-symbol interference (ISI) that is caused by the intrinsic characteristics of channels with memory. If not considered appropriately, the ISI can severely degrade the bit-error-ratio (BER) performance [39, 36, 35]. Even though the use of enzymes [25] appropriately injected in the propagation environment can help mitigate the ISI, this approach cannot completely eliminate the ISI. Motivated by these considerations, we consider an MC system in the presence of ISI. Solutions to mitigate the ISI exist and include methods based on the design of the modulation [68], the code [69], and the detector [70]. In this chapter, we focus our attention on the design of detectors that use binary CSK modulation.

In MC systems, different types of detectors exist, including 1) no memory threshold-based schemes [25, 43]; 2) optimal constant-threshold-based schemes [75]; 3) multi-memory-bit threshold-based schemes [1]; 4) adaptive decision-feedback equalizers [70]; and 5) multi-sub-slot threshold-based schemes [44]. Among these approaches, multi-memory-bit threshold-based schemes constitute a promising solution because they can achieve low BER performance at a low computational complexity. However, perfect CSI needs to be known at the receiver in order to calculate the optimal thresholds. In this regard, channel estimation (CE) [39, 80] methods based on predefined preambles are often utilized in order to ensure the reliable detection of data. In realistic scenarios, however, some parameters of MC communication systems may not remain constant during the CE and data detection phases, thus leading to variations in the channel response and, hence, to CE inaccuracies. In some cases, it may be difficult to model the channel theoretically, e.g. for multiple absorbing receivers [28]. To counteract these issues, some authors have used machine learning (ML) methods to design the receiver [27] and [75]. However, ML-based receivers must be trained with long sequences of symbols. The performance of ML-based receivers, in addition, highly depends on the quality of the data used for training [81].

4.1.2 Related Works

Non-coherent detection methods for MC systems are discussed in [35], where the authors obtain closed-form expressions of the optimal detection metric for a multi-symbol detector that uses an approximated probability mass function for the Poisson channel. In addition, the authors propose a blind single-symbol detector based on a constant threshold. However, the authors assume no ISI in the system model which may hold true when the symbol length is long enough and the symbol rate is low. In [82], the authors consider the impact of ISI and exploit the local convexity of the received signals in order to detect the symbols from the difference of molecular concentration. To capitalize on the local convexity of the received signals, the transmitter needs, however, to release a sufficiently large number of molecules. In [32], a non-linear detector with an adaptive threshold is analyzed. The approach is based on the quick-rising and slow-decaying trends of the received signals after passing through a filter. In [83], the authors exploit the unsupervised fuzzy C-mean approach to detect symbols from the quick-rising and slow-decaying of the processed signals. In [33], the authors propose an approach based on the energy difference of the received signals. This method requires large numbers of molecules in order to exploit the energy difference. Non-coherent detection is investigated in Poisson channels for application to optical wireless communications in [3]. However, current non-coherent detection schemes assume negligible ISI. In summary, the aforementioned approaches either do not consider ISI, or require large numbers of molecules and, in general, cannot achieve low BER performance.

4.1.3 Contributions

In this chapter, motivated by the results in [35] and [1], we devise multi-memory-bit threshold-based schemes without prior CSI for the reliable transmission of data in the presence of ISI. The specific novelty and contributions made by this chapter can be summarized as follows.

- We reformulate the multi-memory-bit threshold-based detectors via intermediate variables, i.e., the average number of received particles as a function of the memory bits. Unlike threshold-based methods that use the prior knowledge of the CSI, the considered intermediate variables can be obtained by applying clustering methods to the received signals.
- In order to reduce the clustering error, we construct multi-dimensional data from the received signals and designate the initial centroids of the clusters using the largest received signal. To the best of the authors' knowledge, it is the first time in the communication literature that the main features are extracted by increasing the dimension for clustering data. Conventional methods, on the other hand, are based on decreasing the dimension

for clustering data [84]. In addition, we can infer the transmitted symbols from the clustered data. We design and analyze two approaches, which are referred to as direct clustering-based inference and clustering-plus-threshold detection.

- We devise an iterative method to further improve the BER performance and to avoid clustering failures in high data rate systems, i.e., in the presence of non-negligible ISI. The essence of the proposed method is to iteratively apply clustering from one-dimensional data to high-dimensional data. In each iteration, we construct the initial centroids based on the estimated centroids in the previous iteration. By using the proposed iterative method, we show that the clustering errors can be reduced.

4.1.4 Chapter Organization

The remainder of this chapter is organized as follows. In Section 4.2, the system model is introduced. In Section 4.3, the multi-memory-bit threshold is reformulated via intermediate variables. In Section 4.4, the proposed non-coherent detection method based on multi-dimensional clustering is introduced. In Section 4.5, an iterative method for estimating the initial centroids for application to multi-dimensional clustering is introduced. In Section 4.6, numerical results and the computational complexity of the proposed schemes are discussed. Finally, Section 4.7 concludes this chapter.

4.2 System model

We consider a three-dimensional unbounded MC system without flow that consists of a point transmitter and a spherical absorbing receiver, as depicted in Fig. 3.2. We assume that each information particle diffuses randomly and independently through the medium. The particles are assumed not to degrade rapidly, thus resulting in ISI.

We assume that the temperature is constant and the viscosity remains unchanged during the whole transmission. Thus, the diffusion coefficient D [65] is assumed to be constant. Assuming that the transmitter is located in $\mathbf{a} = (0, 0, 0)$ and the receiver is located in $\mathbf{b} = (b_x, b_y, b_z)$, the hitting rate is expressed as follows [43, 14]:

$$f_{hit}(t) = \frac{r(d-r)}{d\sqrt{4\pi Dt^3}} e^{-\frac{(d-r)^2}{4Dt}} \quad (4.1)$$

where $\|\mathbf{a} - \mathbf{b}\| = d$ is the distance between the center of the transmitter and the center of the receiver, and r denotes the radius of the receiver. A binary CSK modulation scheme is assumed, and the bit transmitted in the i th time slot is denoted by s_i . During the i th slot, the transmitter releases N_{TX} information particles if $s_i = 1$, otherwise the transmitter does not release any particles. We assume that the N_{TX} information particles are released in a very short time such that we can ignore the release time effect on the received signal. The hitting

probability of an absorbing receiver in the i th time slot of duration T is:

$$P_{i-1} = \int_{(i-1)T}^{iT} f_{hit}(t) dt = \frac{r}{d} \left\{ \operatorname{erfc}\left(\frac{d-r}{\sqrt{4DiT}}\right) - \operatorname{erfc}\left(\frac{d-r}{\sqrt{4D(i-1)T}}\right) \right\} \quad (4.2)$$

where $\operatorname{erf}(y) = \int_0^y \frac{2}{\sqrt{\pi}} e^{-x^2} dx$ and $\operatorname{erfc}(y) = 1 - \operatorname{erf}(y)$.

Let $C_j = N_{TX} P_j$ denote the average number of received particles in the j th time-slot after the release of N_{TX} particles. Thus, the number of received particles [44] in the i th time-slot follows a Poisson distribution:

$$r_i \sim \text{Poisson}(I_i + s_i C_0) \quad (4.3)$$

where $I_i = \bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j$ is the sum of the ISI and background noise, $\bar{\lambda}_0$ is the background noise power per unit time, and L denotes the length (memory) of the Poisson channel.

We define the signal-noise-ratio (SNR) as follows:

$$\text{SNR} = 10 \log_{10} \frac{C_0}{2\bar{\lambda}_0 T} \quad (4.4)$$

since the information bits are assumed to be equiprobable. Thus,

$$N_{TX} = \frac{2\bar{\lambda}_0 T 10^{\frac{\text{SNR}}{10}}}{P_0} \quad (4.5)$$

Therefore, the probability of receiving r_i information particles is:

$$\Pr(r_i | I_i + s_i C_0) = \frac{e^{-(I_i + s_i C_0)} (I_i + s_i C_0)^{r_i}}{r_i!} \quad (4.6)$$

4.3 Multi-memory-bit Threshold Reformulation

To detect the symbols, existing methods are based on calculating thresholds based on previously detected bits, which are referred to as memory bits. The main idea is to exploit previously detected bits, even erroneously estimated, in order to detect new transmitted bits based on the number of received particles. If the number of particles is below the threshold, a binary zero is estimated. Otherwise, a binary one is estimated. We first discuss the case of knowing the exact channel length L , and then we consider the case of limited channel information. The analytical threshold based on the knowledge of L memory bits can be obtained from the

4.3. Multi-memory-bit Threshold Reformulation

following equation:

$$\Pr(r_i | s_i = 0, s_{i-j}, 1 \leq j \leq L) = \Pr(r_i | s_i = 1, s_{i-j}, 1 \leq j \leq L) \quad (4.7)$$

where $\Pr(r_i | s_i = 0, s_{i-j}, 1 \leq j \leq L)$ and $\Pr(r_i | s_i = 1, s_{i-j}, 1 \leq j \leq L)$ denote the probabilities of receiving r_i information particles in the i th time slot conditioned upon the previously transmitted symbols s_{i-j} for $1 \leq j \leq L$ and the current symbol $s_i = 0$ or $s_i = 1$, respectively. These probabilities correspond to (5.8).

By solving (4.7) based on (5.8), the optimal threshold $\tau |_{s_{i-j}, 1 \leq j \leq L}$ conditioned on the previously transmitted symbols s_{i-j} for $1 \leq j \leq L$ is as follows:

$$\tau |_{s_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{I_i})} = \frac{C_0}{\ln(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j})} \quad (4.8)$$

Since the symbols s_{i-j} for $0 \leq j \leq L$ are unknown, the previously estimated symbols \hat{s}_{i-j} for $0 \leq j \leq L$ are used to compute the threshold in (4.8), i.e., $\tau |_{\hat{s}_{i-j}, 1 \leq j \leq L} = \frac{C_0}{\ln(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L \hat{s}_{i-j} C_j})}$. Based on the computed threshold, \hat{s}_i is demodulated as follows:

$$\hat{s}_i = \begin{cases} 1 & r_i > \tau |_{\hat{s}_{i-j}, 1 \leq j \leq L} \\ 0 & r_i \leq \tau |_{\hat{s}_{i-j}, 1 \leq j \leq L} \end{cases} \quad (4.9)$$

Besides necessitating the bits detected in the previous time slots, the computation of the threshold in (4.8) relies upon the prior knowledge of the CSI, i.e., the coefficients C_j are assumed to be known. To avoid using prior information about the channel, i.e., C_j for $0 \leq j \leq L$, when calculating the threshold in (4.8), we resort to intermediate variables, i.e., the average number of received particles $\bar{r}_i |_{s_{i-j}, 0 \leq j \leq L}$ that is defined as follows:

$$\bar{r}_i |_{s_{i-j}, 0 \leq j \leq L} = I_i + s_i C_0 = \bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j \quad (4.10)$$

Equation (4.10) corresponds to the theoretical average number of particles, and, therefore, it depends on the variables C_j . In the next sub-section, we show how the average number of particles $\bar{r}_i |_{s_{i-j}, 0 \leq j \leq L}$ can be obtained directly from the received data without any prior knowledge. In order to formulate the problem analytically and understand the rationale of the proposed approach, we consider, just for this section, that $\bar{r}_i |_{s_{i-j}, 0 \leq j \leq L}$ is given by its analytical expression in (4.10).

Chapter 4. K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

The threshold in (4.8) can be rewritten, in an equivalent form, only in terms of $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq L}$:

$$\begin{aligned} \tau|_{s_{i-j}, 1 \leq j \leq L} &= \frac{C_0}{\ln\left(1 + \frac{C_0}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j}\right)} = \frac{C_0 + \bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j - (\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j)}{\ln\left(\frac{\bar{\lambda}_0 T + C_0 + \sum_{j=1}^L s_{i-j} C_j}{\bar{\lambda}_0 T + \sum_{j=1}^L s_{i-j} C_j}\right)} \\ &= \frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq L} - \bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq L}}{\ln\left(\frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq L}}{\bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq L}}\right)} \end{aligned} \quad (4.11)$$

In practice, the memory length used by the receiver may be limited and may be smaller than L . By assuming that $\mathcal{L} \leq L$ memory bits are used, the threshold can be obtained by solving the equation $\Pr(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$ where:

$$\Pr(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L}) = \frac{1}{2^{L-\mathcal{L}}} \sum_{s_{i-j}, \mathcal{L}+1 \leq j \leq L} \Pr(r_i|s_{i-j}, 0 \leq j \leq L) \quad (4.12)$$

However, it is barely possible to obtain an analytical formula for the threshold with \mathcal{L} memory bits by imposing $\Pr(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$. Instead of using $\Pr(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$, we approximate it as $\Pr_{\text{appro}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$ as follows:

$$\Pr_{\text{appro}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L}) = \frac{e^{-\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}} (\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}})^{r_i}}{r_i!} \quad (4.13)$$

and

$$\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}} = \sum_{j=0}^{\mathcal{L}} s_{i-j} C_j + \sum_{j=\mathcal{L}+1}^L C_j / 2 + \bar{\lambda}_0 T \quad (4.14)$$

Since $\Pr_{\text{appro}}(r_i|s_{i-j}, 0 \leq j \leq \mathcal{L})$ is not a sum of probabilities, by solving $\Pr_{\text{appro}}(r_i|s_i = 0, s_{i-j}, 1 \leq j \leq \mathcal{L}) = \Pr_{\text{appro}}(r_i|s_i = 1, s_{i-j}, 1 \leq j \leq \mathcal{L})$, the analytical threshold can be obtained. By using the intermediate variables $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$, the detection threshold can be written as follows:

$$\tau|_{s_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}} - \bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\bar{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\bar{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (4.15)$$

As anticipated, the threshold in (4.15) is a theoretical formulation for the time being. In practice, it needs to be obtained (estimated) only from the received data without any prior information. To this end, we replace $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ with its estimate $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$, which is introduced and defined in the next sub-section. Therefore, the detection threshold is expressed

4.3. Multi-memory-bit Threshold Reformulation

as follows:

$$\hat{\tau}|_{s_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\hat{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}} - \hat{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\hat{r}_i|_{s_i=1, s_{i-j}, 1 \leq j \leq \mathcal{L}}}{\hat{r}_i|_{s_i=0, s_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (4.16)$$

Since s_{i-j} for $0 \leq j \leq \mathcal{L}$ are unknown, the estimates \hat{s}_{i-j} are used instead, which results in the following formulation for the threshold:

$$\hat{\tau}|_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} = \frac{\hat{r}_i|_{\hat{s}_i=1, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} - \hat{r}_i|_{\hat{s}_i=0, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}{\ln\left(\frac{\hat{r}_i|_{\hat{s}_i=1, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}{\hat{r}_i|_{\hat{s}_i=0, \hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}}}\right)} \quad (4.17)$$

Based on this reformulation, each symbol is demodulated as follows:

$$\hat{s}_i = \begin{cases} 0, & r_i \leq \hat{\tau}|_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} \\ 1, & r_i > \hat{\tau}|_{\hat{s}_{i-j}, 1 \leq j \leq \mathcal{L}} \end{cases} \quad (4.18)$$

4.3.1 Estimation of $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the Received Data

As anticipated, our objective is to develop non-coherent detection schemes that do not need CSI to operate. For example, we cannot rely on the knowledge of C_j for $j = 0, \dots, \mathcal{L}$ to compute the detection thresholds. In this sub-section, we show an example of how $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ can be obtained by only using the received data, which implies that the threshold in (4.17) can be obtained without prior CSI information.

As a case study, we show how the intermediate variables $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ are obtained for $\mathcal{L} = 1$. From (4.12), the exact probability mass function of receiving r_i particles given s_i and s_{i-1} is

$$\Pr(r_i|s_i, s_{i-1}) = \frac{1}{2^{L-1}} \sum_{s_{i-j}, 2 \leq j \leq L} \frac{e^{-\bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j} (\bar{\lambda}_0 T + \sum_{j=0}^L s_{i-j} C_j)^{r_i}}{r_i!} \quad (4.19)$$

Theoretically, $\bar{r}_i|_{s_i, s_{i-1}}$ is defined and can be computed as follows:

$$\bar{r}_i|_{s_i, s_{i-1}} = E[r_i|s_i, s_{i-1}] = \sum_{r_i=0}^{\infty} r_i \Pr(r_i|s_i, s_{i-1}) \quad (4.20)$$

Let us consider a sequence of received particles $\{r_1, \dots, r_n, \dots, r_K\}$ of length K . The theoretical average number of received particles in (4.20) can be estimated empirically from the sequence of K observations. In practice (empirically), the probability $\Pr(r_i|s_i, s_{i-1})$ can be interpreted as

Chapter 4. K-Means Clustering-Aided Non-Coherent Detection for Molecular Communications

a ratio:

$$\Pr(r_i | s_i, s_{i-1}) = \lim_{N_{s_i, s_{i-1}} \rightarrow \infty} \frac{N_{r_i, s_i, s_{i-1}}}{N_{s_i, s_{i-1}}} \quad (4.21)$$

where $N_{r_i, s_i, s_{i-1}}$ denotes the number of elements within the sequence of K observations that are equal to r_i and for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$, and $N_{s_i, s_{i-1}} = \sum_{r_i=0}^{\infty} N_{r_i, s_i, s_{i-1}}$ denotes the number of elements within the sequence of K observations for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$ (but the number of particles is not necessarily equal to r_i). Therefore, (4.20) can be rewritten as follows:

$$\bar{r}_i |_{s_i, s_{i-1}} = \lim_{N_{s_i, s_{i-1}} \rightarrow \infty} \frac{\sum_{r_i=0}^{\infty} r_i N_{r_i, s_i, s_{i-1}}}{N_{s_i, s_{i-1}}} = \lim_{K \rightarrow \infty} \frac{\sum_{n=0}^K r_n \kappa_{n, s_i, s_{i-1}}}{\sum_{n=0}^K \kappa_{n, s_i, s_{i-1}}} \quad (4.22)$$

where $\kappa_{n, s_i, s_{i-1}} = 1$ if $[s_n, s_{n-1}] = [s_i, s_{i-1}]$ and $\kappa_{n, s_i, s_{i-1}} = 0$ otherwise. The variable $\kappa_{n, s_i, s_{i-1}}$ can be regarded as an indicator variable that provides information on whether the observation r_n belongs to a group of observations for which $[s_n, s_{n-1}] = [s_i, s_{i-1}]$.

In the literature [85], a group of observations is referred to as a cluster, a single observation of the cluster is referred to as a point of the cluster, and the condition that defines the cluster, i.e., the condition $[s_n, s_{n-1}] = [s_i, s_{i-1}]$, is referred to as the label of the cluster. In addition, the arithmetic mean of all the points in a subset X of \mathbb{R}^l is referred to as the centroid of the cluster [86] and is defined, in general terms, as follows:

$$\boldsymbol{\mu} = \frac{\int \mathbf{x} g(\mathbf{x}) d\mathbf{x}}{\int g(\mathbf{x}) d\mathbf{x}} \quad (4.23)$$

where \mathbf{x} is a vector in \mathbb{R}^l , the integrals are computed over the whole space \mathbb{R}^l , and $g(\mathbf{x})$ is a characteristic function that can be defined for different purposes. Based on (4.22) and (4.23), $\bar{r}_i |_{s_i, s_{i-1}}$ can be viewed as the centroid of the cluster of observations whose label is $[s_i, s_{i-1}]$. Thus $\bar{r}_i |_{s_i, s_{i-1}}$ is obtained by computing the centroid of the cluster with label $[s_i, s_{i-1}]$ that is obtained from the set of received signals (the observations) $\{r_1, \dots, r_n, \dots, r_K\}$.

Based on (4.22), we evince that the thresholds of interest can be computed by calculating the centroids in (4.23). However, (4.22) necessitates the knowledge of the symbols s_n , which are usually unknown. Our objective is to estimate $\bar{r}_i |_{s_i, s_{i-1}}$ from the received signals without knowing the symbols s_n . Estimating the centroids in (4.22) or (4.23) without any prior information can be obtained by applying clustering methods, which partition the received signals into several groups, each one corresponding to a given label $[s_i, s_{i-1}]$, and then obtaining the corresponding centroids.

In summary, the thresholds for data detection can be obtained by applying clustering methods directly to the received signals without any prior information on the CSI and without knowledge of the transmitted symbols. In particular, the thresholds can be retrieved from the centroids of the clusters. In the next section, we detail how to use clustering for non-coherent detection in MC systems, and, in particular, we introduce our approach based on multi-dimensional clustering.

4.4 Clustering-based Non-coherent Detection

With the aid of the reformulated thresholds in Section 4.3, we show that data detection can be realized without prior CSI. In particular, we introduce non-coherent detection methods based on the K-means clustering algorithm. To this end, we first introduce some background information on clustering in general and the K-means algorithm in particular. To elucidate the operating principle of the proposed methods, we report some illustrations that are obtained by using the simulation setup in Table 4.1 (see Section 4.6) in the presence of mild ISI ($T = 30\Delta T$) and severe ISI ($T = 20\Delta T$).

4.4.1 K-Means Clustering Algorithm

Assume that we have a data set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ of N observations and each element is a \mathbb{D} -dimensional vector \mathbf{x}_n . The objective is to partition the data set into N_c clusters whose centroids are \mathbb{D} -dimensional vectors denoted by $\boldsymbol{\mu}_k$. The centroid $\boldsymbol{\mu}_k$ is the mean of its clustered points. In clustering, $\kappa_{n,k}$ is a variable that represents if the distance between the n th observation and the k th centroid is smaller than the distance between that observation and the other centroids. The K-means algorithm is a clustering method that works iteratively in order to compute or estimate the centroids $\boldsymbol{\mu}_k$ and the indicator variables $\kappa_{n,k}$ from a set of observations. In particular, the K-means algorithm encompasses two steps that iteratively compute $\kappa_{n,k}$ and $\boldsymbol{\mu}_k$ at each step. The K-means algorithm needs an initial estimate of the centroids to operate. The initial centroids $\boldsymbol{\mu}_k$ can be either randomly selected from the data set or other methods can be employed. The selection of the initial centroids is discussed in further text. The K-means clustering algorithm can be summarized as follows.

Step I: Assign \mathbf{x}_n to the closest cluster (initial, if this is the first iteration) centroid:

$$\kappa_{n,k} = \begin{cases} 1, & \text{if } k = \underset{j}{\operatorname{arg\,min}} \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (4.24)$$

Step II: Update the cluster centroid:

$$\boldsymbol{\mu}_k = \frac{\sum_n \kappa_{n,k} \mathbf{x}_n}{\sum_n \kappa_{n,k}} \quad (4.25)$$

where $\sum_n \kappa_{n,k}$ is the number of points in the k th cluster.

The two steps are repeated until $\kappa_{n,k}$ does not change. In the next sub-section, we show how to apply this algorithm for non-coherent detection in MC systems.

4.4.2 Single-Dimensional Clustering: Challenges and Limitations

In (4.16), the empirical average number of received particles $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ need to be as close as possible to the theoretical average number of received particles $\bar{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ in order to obtain a low BER. In order to elucidate the complexity of the problem at hand, let us consider the example in (4.19) by assuming $\mathcal{L} = 1$.

The objective is to estimate $\hat{r}_i|_{s_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the set of received observations, $\{r_1, \dots, r_n, \dots, r_K\}$, that are obtained from (4.3) given the transmitted symbols $\{s_1, \dots, s_n, \dots, s_K\}$. Since $\mathcal{L} = 1$, we are interested in identifying four clusters in the received signals, which correspond to the labels $[s_i, s_{i-1}] = [0, 0]$, $[s_i, s_{i-1}] = [0, 1]$, $[s_i, s_{i-1}] = [1, 0]$, and $[s_i, s_{i-1}] = [1, 1]$. From (4.20) or equivalently (4.14), we obtain:

$$\bar{r}_i|_{s_i, s_{i-1}} = s_i C_0 + s_{i-1} C_1 + \sum_{j=2}^L C_j / 2 + \bar{\lambda}_0 T \quad (4.26)$$

The average number of received particles in (4.26) is the theoretical one that can be obtained from analysis. In order to understand the difficulty of obtaining (4.26) from the empirical data, we illustrate an example in Fig. 4.1. Based on a large set of empirical data ($K = 2^{16}$ samples), we calculate the empirical probability mass function (PMF) of the number of particles, by assuming that the labels $[s_i, s_{i-1}]$ are known. How the labels can be estimated from the data is discussed in further text. From the empirical data, we apply the K-means clustering algorithm in (4.24) and (4.25) in order to estimate the centroids. The estimated average number of received particles $\hat{r}_i|_{s_i, s_{i-1}}$ is set equal to the estimated centroids. From Fig. 4.1, we observe that there is a non-negligible difference between the theoretical values obtained from (4.26) and the empirical values estimated by using the K-means clustering algorithm, even if the labels are assumed to be known, i.e., the association between the centroids and the labels is error-free. The differences between $\hat{r}_i|_{s_i, s_{i-1}}$ and $\bar{r}_i|_{s_i, s_{i-1}}$ mainly originate from the overlapping areas of the conditional probabilities $\Pr(r_i|s_i, s_{i-1})$.

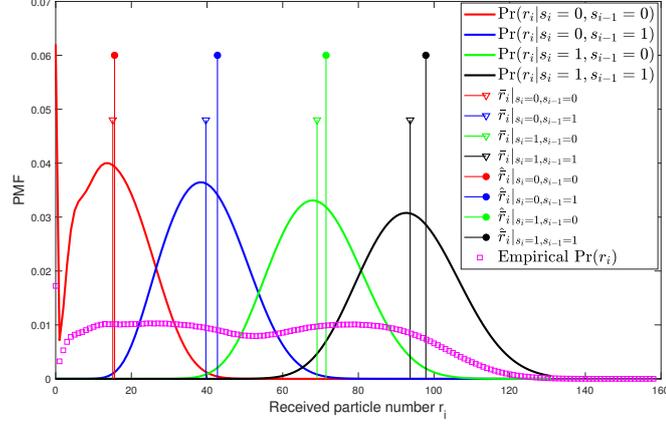


Figure 4.1: The positions of $\bar{r}_i|_{s_i, s_{i-1}}$ (calculated via (4.26)) and $\hat{\bar{r}}_i|_{s_i, s_{i-1}}$ corresponding to the centroids calculated from (4.24) and (4.25) and assuming that the labels $[s_i, s_{i-1}]$ are known for $T = 30\Delta T$

The simple example illustrated in Fig. 4.1 highlights the challenges of applying clustering for estimating the average number of received particles in MC systems and, therefore, the detection thresholds. In the following sub-section, we introduce our proposed approach for improving the estimation accuracy based on multi-dimensional clustering methods.

4.4.3 Multi-Dimensional Clustering

The proposed solution to increase the accuracy of estimating the average number of received particles from the empirical data relies on multi-dimensional clustering. The main idea is not to apply the K-means clustering algorithm in (4.24) and (4.25) to individual observations r_n but to vectors of observations:

$$\mathbf{r}_n = [r_n, r_{n-1}, \dots, r_{n-\mathcal{L}}] \quad (4.27)$$

where \mathcal{L} is the number of memory bits that we can use for detection and for multi-dimensional clustering. In order to illustrate the advantages of the proposed approach, let us consider the same example as in Fig. 4.1, but by assuming that the K-means clustering algorithm in (4.24) and (4.25) is applied to the vector $[r_n, r_{n-1}]$, rather than to r_n only.

The results are illustrated in Fig. 4.2, where the following notation is used. The cluster of points whose label is, e.g., $[s_i = 0, s_{i-1} = 1]$ is denoted by $[r_n, r_{n-1}]|_{s_i=0, s_{i-1}=1}$. A similar notation is used for the other clusters. By definition, $\bar{r}_i|_{s_i, s_{i-1}} = s_i C_0 + s_{i-1} C_1 + \sum_{j=2}^L C_j / 2 + \bar{\lambda}_0 T$ and $\bar{r}_{i-1}|_{s_{i-1}} = s_{i-1} C_0 + \sum_{j=1}^L C_j / 2 + \bar{\lambda}_0 T$ given the label $[s_i, s_{i-1}]$. Therefore, the corresponding

theoretical centroids are computed from (4.14) as follows:

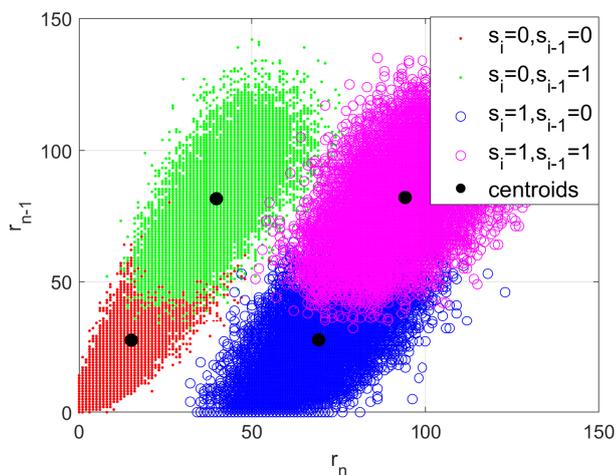
$$\begin{aligned}
 [\bar{r}_i, \bar{r}_{i-1}]|_{s_i=0, s_{i-1}=0} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [\bar{\lambda}_0 T, \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]|_{s_i=0, s_{i-1}=1} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_1 + \bar{\lambda}_0 T, C_0 + \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]|_{s_i=1, s_{i-1}=0} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_0 + \bar{\lambda}_0 T, \bar{\lambda}_0 T] \\
 [\bar{r}_i, \bar{r}_{i-1}]|_{s_i=1, s_{i-1}=1} &= \left[\sum_{j=2}^L C_j/2, \sum_{j=1}^L C_j/2 \right] + [C_0 + C_1 + \bar{\lambda}_0 T, C_0 + \bar{\lambda}_0 T]
 \end{aligned} \tag{4.28}$$

Given a sequence of received signals (observations), $\{r_1, \dots, r_n, \dots, r_K\}$, the vectors of observations are $\{[r_2, r_1], \dots, [r_n, r_{n-1}], \dots, [r_K, r_{K-1}]\}$. The empirical distributions of each cluster of points and the theoretical centroids (computed from (4.28)) are depicted in Fig. 4.2a. We apply the K-means clustering algorithm in (4.24) and (4.25) to the empirical vectors using some initial centroids $\boldsymbol{\mu}_k$. The setup of the initial centroids is described in the next sub-section. In general, the exact labels $[s_n, s_{n-1}]$ are unknown, which implies that the association between an estimated centroid and the corresponding label $[s_i, s_{i-1}]$ is not known a priori. We solve this issue by appropriately selecting the initial centroids, and by deciding the association between centroids and labels at the beginning of the K-means algorithm. Based on our approach, the estimated final centroids inherit the labels associated to the initial centroids. If $[s_i, s_{i-1}]$ is the label of the initial centroid $\boldsymbol{\mu}_k$, then $[s_i, s_{i-1}]$ will be the label of the estimated (final) centroid $\hat{\boldsymbol{\mu}}_k$. Therefore, the choice of the initial centroids is important to ensure a correct labelling. By comparing Fig. 4.2b and Fig. 4.1, we observe that $\hat{r}_i|_{s_i, s_{i-1}}$ is closer to $\bar{r}_i|_{s_i, s_{i-1}}$ in Fig. 4.2b, which results in better performance. Therefore, multi-dimensional clustering is shown to be more accurate than one-dimensional clustering.

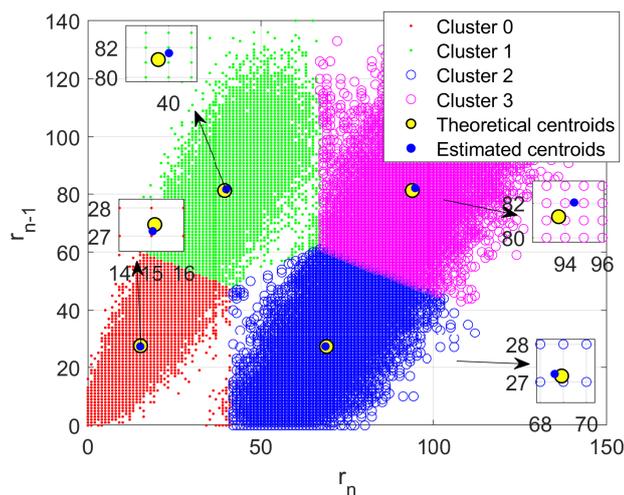
4.4.4 Setup of the Initial Centroids

As mentioned, an important issue is to assign the correct labels to the initial centroids and to ensure that the association between the estimated centroids and the correct labels does not change when applying the K-means clustering algorithm. In order to motivate the proposed approach for the setup of the initial centroids and the corresponding labels, let us consider Fig. 4.3a. In the figure, the four black dots correspond to the points $[0, 0]$, $[0, \max(\mathbf{r})]$, $[\max(\mathbf{r}), 0]$, and $[\max(\mathbf{r}), \max(\mathbf{r})]$, where $\max(\mathbf{r})$ returns that largest value of the observations $\{r_1, \dots, r_n, \dots, r_K\}$. Our approach consists of choosing as initial centroids the four black dots in Fig. 4.3a, which, by direct inspection of the figure, are shown to provide us with the correct labels as well, i.e., $[0, 0]$ corresponds to the label $[s_i, s_{i-1}] = [0, 0]$, $[0, \max(\mathbf{r})]$ corresponds to the label $[s_i, s_{i-1}] = [0, 1]$, etc. If K is sufficiently large, the proposed association between centroids and labels is expected not to change, with high probability, during the application of the K-means

4.4. Clustering-based Non-coherent Detection



(a) Empirical distributions of $[r_n, r_{n-1}]|_{s_i, s_{i-1}}$ and theoretical centroids obtained from (4.28). In this case, the labels are assumed to be known a priori



(b) Theoretical centroids obtained by using (4.28) (yellow circles) vs. estimated centroids obtained by using the K-means algorithms in (4.24) and (4.25) (blue circles). The labels are not known a priori but are obtained from the estimated centroids (the updated initial centroids with assigned labels)

Figure 4.2: Theoretical and estimated centroids obtained from clustering the data vector $[r_n, r_{n-1}]$ ($T = 30\Delta T$)

algorithm.

If the number of memory-bits \mathcal{L} is greater than one, the initial centroids are assigned in a similar fashion. If $\mathcal{L} = 2$, for example, the initial centroids are $[0, 0, 0]$ whose label is $[s_i, s_{i-1}, s_{i-2}] = [0, 0, 0]$; $[0, 0, \max(\mathbf{r})]$ whose label is $[s_i, s_{i-1}, s_{i-2}] = [0, 0, 1]$; $[0, \max(\mathbf{r}), \max(\mathbf{r})]$

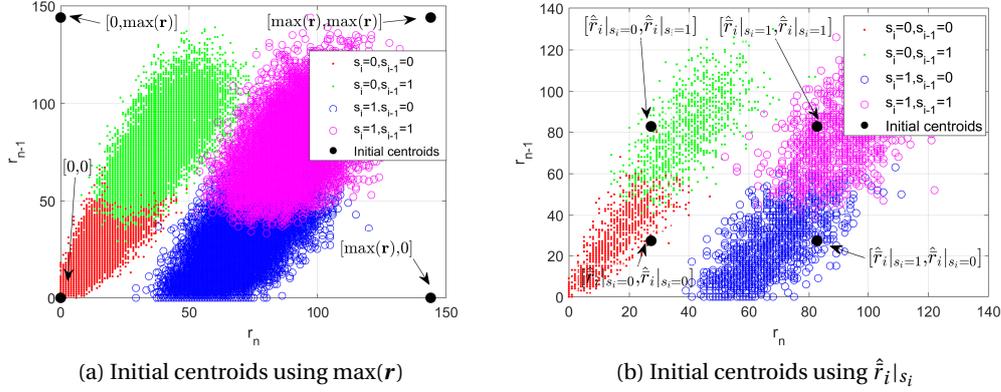


Figure 4.3: The customized initial centroids and the corresponding clusters ($T = 30\Delta T$)

whole label is $[s_i, s_{i-1}, s_{i-2}] = [0, 1, 1]$; etc.

Based on these proposed initial centroids and the corresponding labels, the K-means clustering algorithm is applied according to (4.24) and (4.25), which returns the final estimated centroids ($\hat{\mu}_k$) and the indicator variables $\kappa_{n,k}$. In particular, $\kappa_{n,k}$ allows us to implicitly perform data detection at the end of the clustering process, since it informs us, by definition, if a vector of points belongs or not to a given cluster. Based on this remark, we propose two algorithms for data detection that are referred to as (1) direct clustering-based inference and (2) clustering-plus-threshold detection.

4.4.5 Direct Clustering-Based Inference and Clustering-plus-Threshold Detection

The proposed *direct clustering-based inference* algorithm is based only on clustering methods and does not exploit the thresholds in (4.17) for data detection. The algorithm is given in Algorithm 1 and works as follows (assuming, e.g., $\mathcal{L} = 1$). If the observation vector $[r_n, r_{n-1}]$ belongs to the cluster with label $[s_i, s_{i-1}]$, i.e., the corresponding indicators variables are $\kappa_{n,k} = 1$, then the estimated bits are those of the corresponding label.

The proposed *clustering-plus-threshold detection* algorithm, on the other hand, combines together clustering methods and the estimated thresholds in (4.17) for data detection. The algorithm is given in Algorithm 2 and it has one main difference compared with Algorithm 1: After finalizing the clustering process based on the K-means algorithm, the bits are not detected by using the indicator variables $\kappa_{n,k}$, but the thresholds in (4.17) are computed from the estimated centroids, which are then used for data detection by using (4.18).

Algorithm 1 Direct clustering-based inference

First Step: Clustering

- 1: Set \mathcal{L}
- 2: Construct the data \mathbf{r}_n via (4.27)
- 3: Construct the initial centroids $\boldsymbol{\mu}_k$
- 4: Cluster the \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$

Second Step: Inference

- 5: Infer \hat{s}_n from the indicator variables $\kappa_{n,k}$
-

Algorithm 2 Clustering-plus-threshold detection

First Step: Clustering

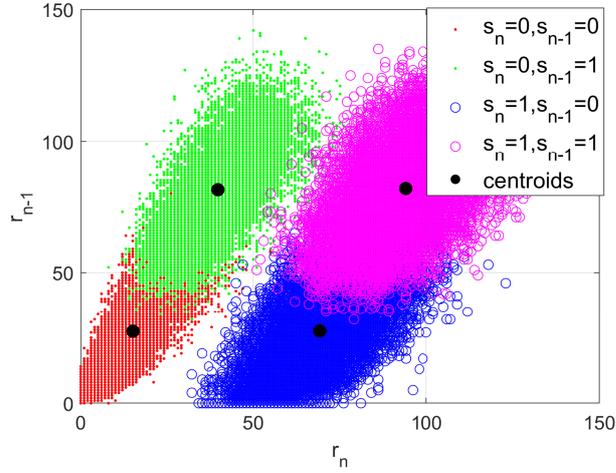
- 1: Set \mathcal{L}
- 2: Construct the data \mathbf{r}_n via (4.27)
- 3: Construct the initial centroids $\boldsymbol{\mu}_k$
- 4: Cluster the \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$

Second Step: Detection

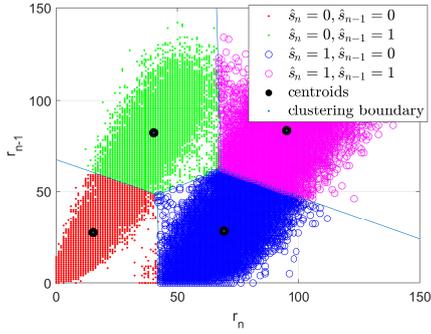
- 5: Obtain $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the estimated centroids $\hat{\boldsymbol{\mu}}_k$
- 6: Compute the detection thresholds using (4.17)
- 7: Detect the symbols using (4.18)

In order to appreciate the difference between the two algorithms and the advantages and limitations of clustering and using or not using the detection thresholds in (4.17), we provide some results in Figs. 4.4-4.7. These figures are obtained as follows. Consider, for example, the case study $\mathcal{L} = 1$ for ease of exposition. After estimating the received data \hat{s}_n , all the pairs of observations $[r_n, r_{n-1}]$ that are detected as $[\hat{s}_n, \hat{s}_{n-1}] = [0, 0]$, $[\hat{s}_n, \hat{s}_{n-1}] = [0, 1]$, $[\hat{s}_n, \hat{s}_{n-1}] = [1, 0]$, and $[\hat{s}_n, \hat{s}_{n-1}] = [1, 1]$ are depicted in red, green, blue, and magenta colors, respectively. The estimated centroids are depicted as black dots. The figures report the empirical decision boundaries that are obtained after clustering (Algorithm 1) and after applying the empirical thresholds estimated from the centroids (Algorithm 2). We note that the two algorithms provide different results, and their advantages and limitations depend on the severity of the ISI and number of memory bits. It is worth mentioning that in Figs. 4.5 and 4.7, due to the many memory bits used, there exist more than four clusters and the dimension of the clusters is greater than two. To be able to illustrate the results, we merge together the clusters whose \mathcal{L} -tuples $[r_n, r_{n-1}, \dots, r_{n-\mathcal{L}}]$ have the same pair $[\hat{s}_n, \hat{s}_{n-1}]$. By using this approach, we obtain again four clusters that can be readily represented. The merged clusters are visualized by using the same color-based code as for $\mathcal{L} = 1$.

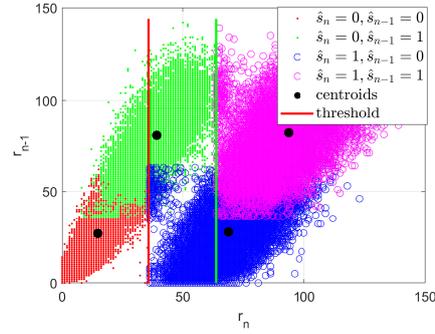
By direct inspection of the figures, the following conclusions can be drawn. From Fig. 4.4, i.e., for mild ISI, we observe that Algorithm 1 outperforms Algorithm 2 if $\mathcal{L} = 1$. In particular, Fig. 4.4b yields decision regions that are closer, as compared with Fig. 4.4c, to the theoretical ones in Fig. 4.4a. From Fig. 4.5, on the other hand, we obtain an opposite trend if $\mathcal{L} = 4$. Therefore,



(a) Transmitted bits and exact labels

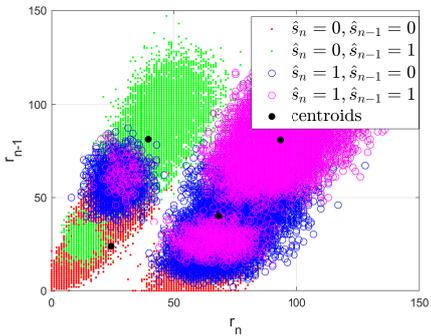


(b) Direct clustering-based inference

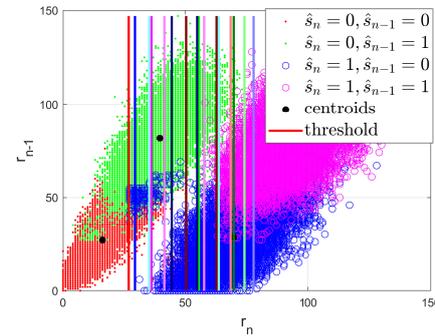


(c) Clustering-plus-threshold detection

Figure 4.4: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1$ and $T = 30 \Delta T$).



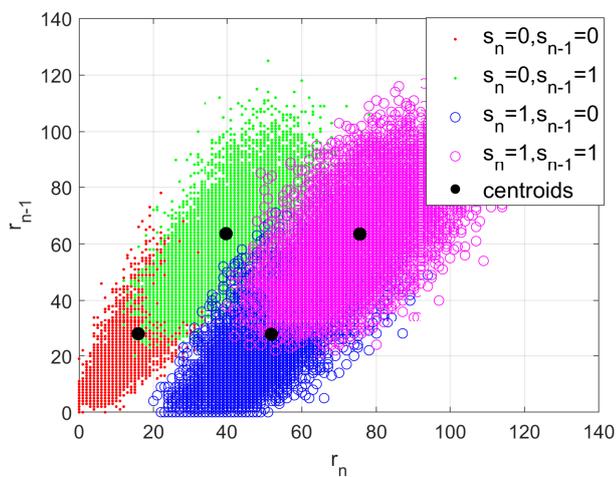
(a) Direct clustering-based inference



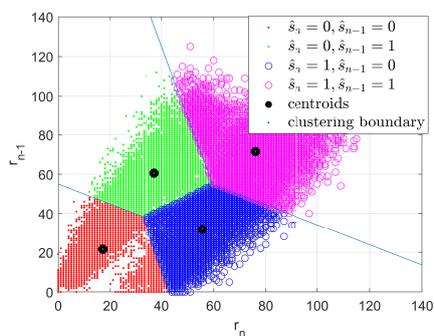
(b) Clustering-plus-threshold detection

Figure 4.5: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4$ and $T = 30 \Delta T$).

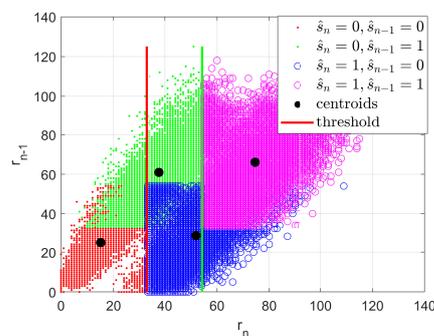
4.4. Clustering-based Non-coherent Detection



(a) Transmitted bits and exact labels

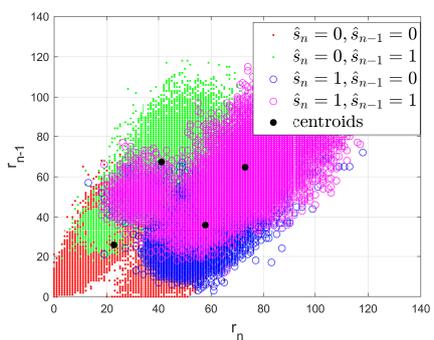


(b) Direct clustering-based inference

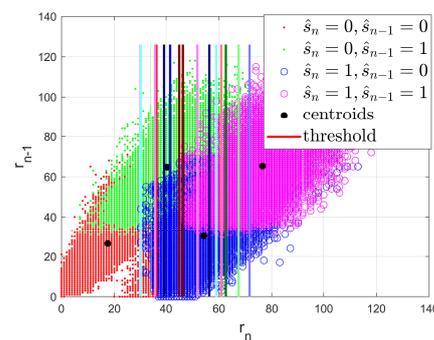


(c) Clustering-plus-threshold detection

Figure 4.6: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$, estimated centroids and clustering boundaries/thresholds ($\mathcal{L} = 1, T = 20 \Delta T$).



(a) Direct clustering-based inference



(b) Clustering-plus-threshold detection

Figure 4.7: Distribution of $[r_n, r_{n-1}]|_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4, T = 20 \Delta T$).

we evince that Algorithm 1 with few memory bits may be a sufficiently good approach for non-coherent data detection under mild ISI. Under severe ISI, on the other hand, the performance trends are different, as illustrated in Figs. 4.6 and 4.7. We observe, in particular, that the best estimation performance is obtained by using Algorithm 2 with a large number of memory bits, i.e., $\mathcal{L} = 4$ in Fig. 4.7b. Errors are still clearly visible, but the estimated clusters of points are closer to the expected ones, i.e., those shown in Fig. 4.6a. It is worth mentioning that in the presence of severe ISI the combination of clustering methods and detection thresholds provides better performance than using only clustering. The use of detection thresholds, in particular, allows us to correct the increased number of mis-detections that originate from increasing the dimension of clustering. These *qualitative* conclusions that are drawn from the direct inspection of the estimated clusters from the detected bits are corroborated in Section 4.6 with the aid of BER simulations.

4.5 Iterative Algorithm for Computing the Initial Centroids

In the previous section, we have illustrated the core ideas of the proposed clustering-based algorithms for non-coherent detection in MC systems, which may or may not use detection thresholds. The comparison of, e.g., Fig. 4.6a and Fig. 4.7b reveals, however, that detection errors still exist, especially in the presence of non-negligible ISI. The sources of error include the initial estimates of the centroids and the resulting estimates of the detection thresholds. In this section, motivated by these considerations, we introduce improved, iterative-based, clustering-based methods in order to enhance the detection performance in the presence of severe ISI.

The initial centroids obtained from $\max(\mathbf{r})$, e.g., $[0, 0]$, $[0, \max(\mathbf{r})]$, $[\max(\mathbf{r}), 0]$ and $[\max(\mathbf{r}), \max(\mathbf{r})]$ if $\mathcal{L} = 1$, are, in general, not always sufficiently close to the theoretical centroids. This may result in detection errors. In this section, therefore, we introduce a more accurate iterative-based approach for estimating the initial centroids.

In order to illustrate the proposed approach, let us consider $\mathcal{L} = 1$ (two-dimensional clustering) and the cluster whose label is $[s_i, s_{i-1}] = [0, 0]$. Based on (4.28), the theoretical centroid that corresponds to $[s_i, s_{i-1}] = [0, 0]$ is $[\bar{r}_i|_{s_i=0, s_{i-1}=0}, \bar{r}_{i-1}|_{s_i=0, s_{i-1}=0}]$, where:

$$\bar{r}_{i-1}|_{s_i=0, s_{i-1}=0} = \bar{r}_{i-1}|_{s_{i-1}=0} = \sum_{j=1}^L C_j/2 + \bar{\lambda}_0 T \quad (4.29)$$

$$\bar{r}_i|_{s_i=0, s_{i-1}=0} = \sum_{j=2}^L C_j/2 + \bar{\lambda}_0 T \quad (4.30)$$

4.5. Iterative Algorithm for Computing the Initial Centroids

From (4.29) and (4.30), we observe that a good estimate for the initial centroid is $[a, a]$ with $a = \sum_{j=1}^L C_j / 2 + \bar{\lambda}_0 T$, since this estimate would be close to the theoretical centroid, and, in general, it is closer than the initial centroid $[0, 0]$ that was used in the previous section. Even though a depends on C_j for $j = 1, 2, \dots, L$, an estimate for it can be obtained from one-dimensional clustering without any prior CSI. In particular, a is, by definition, approximately equal to $a = \bar{r}_i|_{s_i=0} \simeq \hat{r}_i|_{s_i=0}$, where $\hat{r}_i|_{s_i=0}$ is the estimated centroid obtained from one-dimensional clustering. A similar approach can be used to estimate the centroids of the other clusters.

In general, the proposed approach lies in setting the initial centroids for \mathcal{L} -dimensional clustering from the estimated centroids obtained by applying $(\mathcal{L} - 1)$ -dimensional clustering. In turn, the initial centroids for $(\mathcal{L} - 1)$ -dimensional clustering are obtained from the estimated centroids obtained by applying $(\mathcal{L} - 2)$ -dimensional clustering. This procedure can be iterated until one-dimensional clustering, whose two initial centroids can be initialized to 0 and $\max(\mathbf{r})$.

The proposed approach for setting the initial centroids in an iterative-based manner is summarized in Algorithm 3 and Algorithm 4 for application to iterative clustering-based inference and iterative clustering-plus-threshold detection, respectively. In particular, the proposed approach for setting the initial centroids works in an iterative manner, as follows: (i) one-dimensional clustering is applied to the observations $\{r_1, \dots, r_K\}$ by setting 0 and $\max(\mathbf{r})$ as the initial centroids; (ii) the K-means clustering algorithm is applied and the estimated centroids $\hat{r}_i|_{s_i=0}$ and $\hat{r}_i|_{s_i=1}$ are obtained; (iii) the estimated centroids $\hat{r}_i|_{s_i=0}$ and $\hat{r}_i|_{s_i=1}$ are used to construct new initial centroids for application to two-dimensional clustering. In particular, the four initial centroids are set to $[\hat{r}_i|_{s_i=0}, \hat{r}_i|_{s_i=0}]$, $[\hat{r}_i|_{s_i=0}, \hat{r}_i|_{s_i=1}]$, $[\hat{r}_i|_{s_i=1}, \hat{r}_i|_{s_i=0}]$, and $[\hat{r}_i|_{s_i=1}, \hat{r}_i|_{s_i=1}]$; (iv) two-dimensional clustering is applied to the vectors of observations $\{[r_2, r_1], \dots, [r_K, r_{K-1}]\}$ by using the initial estimated centroids; (v) the K-means (two-dimensional) clustering algorithm is applied again and the estimated two-dimensional centroids $[\hat{r}_i, \hat{r}_{i-1}]|_{s_i, s_{i-1}}$ are obtained; (vi) the procedure is iterated until the clustering dimension $\mathcal{L} + 1$.

In general terms, the initial centroids for application to the $(l + 1)$ -dimensional clustering can be constructed from the estimated centroids obtained from l -dimensional clustering. In mathematical terms, let us denote the estimated centroids from l -dimensional clustering by $[\hat{r}_m, \dots, \hat{r}_{m-l+1}]|_{s_{m-j}, 0 \leq j \leq l-1}$. The initial centroid $\boldsymbol{\mu}_k$ that corresponds to the label $[s_i, \dots, s_{i-l+1}, s_{i-l}]$ for $(l + 1)$ -dimensional clustering can be obtained as follows:

$$\boldsymbol{\mu}_k = [\hat{r}_m|_{s_{m-j}=s_{i-j}, 0 \leq j \leq l-1}, \hat{r}_m|_{s_{m-j}=s_{i-1-j}, 0 \leq j \leq l-1}, \hat{r}_{m-1}|_{s_{m-j}=s_{i-1-j}, 1 \leq j \leq l-1}, \dots, \hat{r}_{m-l+1}|_{s_{m-l+1}=s_{i-l}}] \quad (4.31)$$

which does not require any prior CSI information.

Algorithm 3 Iterative clustering-based inference

- 1: Set the initial centroids $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\mu}_1 = \max(\mathbf{r})$
- 2: **for** $l = 1$ to $\mathcal{L} + 1$ **do**
- 3: Construct the data $\mathbf{r}_n = [r_n, \dots, r_{n-l+1}]$
- 4: Cluster \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$
- 5: Set the new initial centroids to $\boldsymbol{\mu}_k$ by using (4.31)
- 6: **end for**
- 7: Infer \hat{s}_n from the indicator variables $\kappa_{n,k}$

Algorithm 4 Iterative clustering-plus-threshold detection

- 1: Set the initial centroids $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\mu}_1 = \max(\mathbf{r})$
- 2: **for** $l = 1$ to $\mathcal{L} + 1$ **do**
- 3: Construct the data $\mathbf{r}_n = [r_n, \dots, r_{n-l+1}]$
- 4: Cluster \mathbf{r}_n using the K-means algorithm with the initial centroids $\boldsymbol{\mu}_k$
- 5: Set the new initial centroids to $\boldsymbol{\mu}_k$ by using (4.31)
- 6: **end for**
- 7: Obtain $\hat{r}_i |_{\hat{s}_{i-j}, 0 \leq j \leq \mathcal{L}}$ from the estimated centroids $\hat{\boldsymbol{\mu}}_k$
- 8: Compute the detection thresholds using (4.17)
- 9: Detect the symbols using (4.18)

By considering the same case study as in Fig. 4.7, we illustrate the performance obtained by employing the proposed iterative approach in Algorithms 3 and 4 in Fig. 4.8. We observe that, even in the presence of severe ISI, much better estimation performance is obtained by a more accurate initial estimate of the initial centroids. These results confirm the effectiveness of the proposed iterative-based approach.

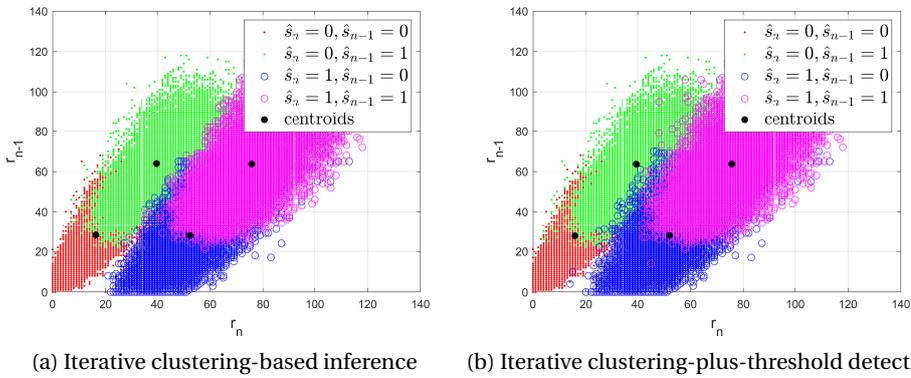


Figure 4.8: Distribution of $[r_n, r_{n-1}] |_{\hat{s}_n, \hat{s}_{n-1}}$ and estimated centroids ($\mathcal{L} = 4, T = 20\Delta T$).

4.6 Numerical Results

In this section, we report some simulation results in order to quantitatively analyze the performance of the proposed non-coherent detectors. The simulation parameters are listed in Table 4.1. For completeness, we summarize the computational complexity of the proposed schemes in Table 4.2, by assuming that the K-means clustering algorithms is repeated P times.

Table 4.1: Simulation parameters

Parameter	Value
λ_0	$100s^{-1}$
Receiver radius r	45 nm
Distance d	500 nm
Diffusion coefficient D	$4.265 \times 10^{-10} m^2/s$
Discrete time length ΔT	9 us
Channel length L	5
K	2^{12}

Table 4.2: Computational complexity

Algorithm	Additions	Multiplications
Algorithm 1, 2	$2^{\mathcal{L}+1}PK(2\mathcal{L}+1)$	$2^{\mathcal{L}+1}PK(\mathcal{L}+1)$
Algorithm 3, 4	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i-1)$	$PK \sum_{i=1}^{\mathcal{L}+1} 2^i(2i-1)$

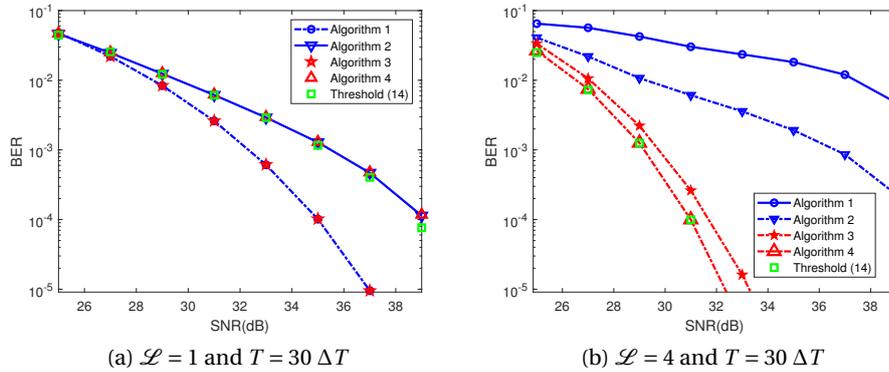


Figure 4.9: BER comparison of the proposed algorithms (mild ISI)

Figure 4.9a depicts the BER for $\mathcal{L} = 1$. As qualitatively illustrated in Fig. 4.4, we observe that, under mild ISI, the algorithms that exploit the detection thresholds yield a BER that is closer to the BER obtained by using the theoretical threshold in (4.15). However, the algorithms that exploit only the indicator variables $\kappa_{n,k}$ achieve better BER performance. As shown in Fig. 4.9b, if more memory bits are used ($\mathcal{L} = 4$), on the other hand, Algorithm 1 yields poor BER performance, as anticipated in Fig. 4.5a. By combining clustering methods with the

detection thresholds, Algorithm 2 yields better BER performance, even though it is worse than Algorithms 3 and 4. In addition, as shown in Fig. 4.8a, Algorithms 3 and 4 yield good BER performance since they iteratively update the centroids.

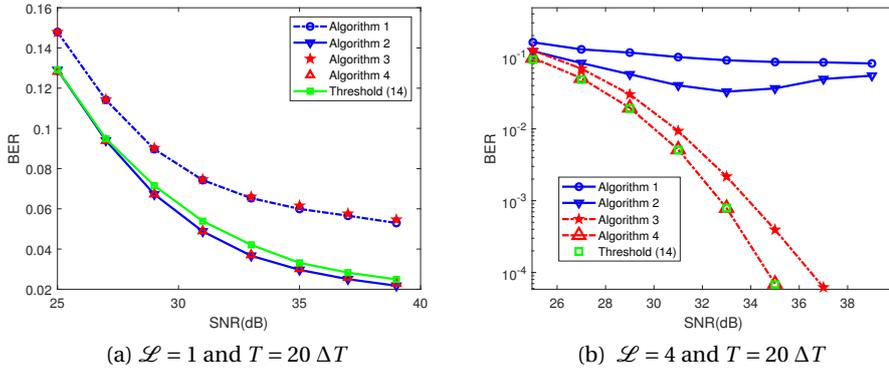


Figure 4.10: BER comparison of the proposed algorithms (severe ISI)

Under mild ISI, therefore, we conclude that Algorithm 1 with a few memory bits can be considered to be a sufficiently good solution. The performance trends, however, are different in the presence of severe ISI, which unveils the gain offered by the proposed algorithms that combine clustering and detection thresholds together. This is illustrated in Fig. 4.10. We observe, in particular, that not using the detection thresholds may lead to wrong clustering estimates if $\mathcal{L} = 1$, which results in poor BER performance. In the presence of severe ISI, in particular, we observe that the proposed Algorithms 3 and 4 yield good BER performance. In particular, Algorithm 4 offers the best BER performance, and it is, therefore, the most suitable choice for non-coherent detection in MC systems.

4.7 Conclusion

In this chapter, we have introduced non-coherent detection schemes for application to MC systems in the presence of ISI. The proposed algorithms are based on combining together clustering methods and empirical estimates of the detection thresholds that are employed in memory-bits detection methods. In order to apply the proposed clustering-based algorithms, different methods for initializing the centroids of the clusters directly from the empirical data are proposed and analyzed. Simulation results show that, in the presence of severe ISI, the proposed algorithms that combine multi-dimensional clustering methods and detection thresholds together yield the best BER performance.

5 Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

In this chapter, we consider the synchronization problem in the diffusive molecular communication system consisting of a point transmitter and a spherical absorbing receiver using the constant threshold. Basically, the synchronization is to estimate the start point of a symbol duration in the received signals. In particular, we focus on the synchronization with unknown channel information. We formulate the error probability as a function of the constant threshold and the start point in order to compute the theoretical optimal start point. Asymptotically, we obtain the condition where the optimal start point satisfies. Based on this condition, we devise a data-aided synchronization metric function. The metric results follow the Skellam distribution allowing us to compute the upper bounds for the absolute mean and variance of the synchronization error assuming that the length of the known symbols is long. Furthermore, the intermediate variables in the metric function can be obtained by applying the unsupervised K-means clustering algorithm to partition the received signals into two groups, yielding a blind synchronization scheme. To verify the synchronization performance, we derive the Cramer-Rao lower bounds for the cases of the known channel information and entirely unknown channel information. Simulation results and the theoretical upper bounds verify the effectiveness of the proposed methods.

5.1	Introduction	93
5.2	System Model	95
5.3	The Optimal Start Point	97
5.3.1	Formulation of $P_e(\rho, \tau)$	97
5.3.2	Asymptotic Optimal Start Point	100
5.4	The Proposed Synchronization Schemes	102
5.4.1	Data-Aided Synchronization Algorithm	103
5.4.2	K-Means-based Blind Synchronization Algorithm	109
5.5	The Cramer-Rao Lower Bounds with known and unknown CSI	110
5.6	Simulation Results	114
5.7	Conclusions	116
5.8	Appendix	118

5.1 Introduction

Recently, the molecular communication (MC) has attracted considerable attentions and is deemed to be a promising solution [9] to the micro-scale communications [78] which show the potential of the the emerging applications, e.g., the nano-machine and bio-robots [65]. The MC system is based on the information nano-particles to transmit signals in order to cope with the difficulty of wireless communications in micro-scale, i.e., the non-negligible component size and the strong channel loss.

Before detecting signals, the synchronization is crucial to every communication system. Imperfect synchronization would degrade the detection performance. In wireless communications, the synchronization is usually performed based on the preamble, i.e., the known transmitted signals [62, 63] before the following tasks, e.g., channel estimation (CE) [64, 41]. The channel of wireless communications is sparse such that the synchronization and CE can be performed jointly. In the MC systems, however, the channel is non-sparse [20] and is infinitely long. Thus, approaches in wireless communications cannot apply directly. In addition, the inter-symbol-interference (ISI) [15, 34], the intrinsic characteristic of the MC systems, deteriorates the synchronization performances of the algorithms that are extended to the MC systems. Although the severe ISI can be mitigated by enzymes [25], the ISI still cannot be eliminated entirely. All the mentioned issues make the synchronization in the MC systems a tough problem. Therefore in this chapter, we consider the synchronization problem in the MC system in the presence of the non-negligible ISI.

The aim of the synchronization is to estimate the start point of a symbol within the received signals. In [87, 38, 88, 89], Lin et al. proposed some methods that are based on a two-way message exchange mechanism to estimate the clock offset and the clock skew such that the transmitter and the receiver are able to synchronize the systematic time delay and the symbol interval. However, the proposed approaches need too many rounds of information exchanges, which, however, would be impossible if there exist flows in the channel. While in [42], without requiring the long preamble, a blind synchronization scheme is proposed for a molecule shift keying (MoSK) based MC system with several receivers that detect different types of molecules. During each time slot, the released particles serve as not only the information carrier particles but also the synchronization particles. However, this scheme also needs the prior channel information. In [37], the authors proposed a single input multiple output (SIMO) blind clock synchronization scheme. By inserting the peak concentration and the corresponding time instant into the channel model, several equations are obtained. The start point can be estimated by solving these equations. Although only one symbol is required, this scheme depends on the analytical channel model which may not be possible to obtain in the realistic systems. In [13], the authors employ two types of molecules and propose two approaches, 1) using type A molecule for detection and type B molecule for synchronization, and 2) alternatively exploiting the two types of molecules for synchronization and detection

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

aiming to mitigate the unstable symbol slot length caused by the imperfect internal clock in the nano-machines. However, the proposed schemes need the channel model to construct the maximum likelihood metrics.

Even though the aforementioned schemes solve several synchronization issues in different scenarios, one challenging problem still remains, i.e., synchronization without the channel state information (CSI). Before the reliable transmission is built, neither the start point nor the CSI is known. Thus the aforementioned CSI-dependent synchronization schemes are impossible to apply and the common CE algorithms that rely on the known start point cannot be exploited to provide CSI for the synchronization schemes. In this chapter, we assume that the most energy-efficient modulation scheme is adopted, i.e., binary concentration shift keying (CSK) modulation [79] is employed and the symbol length is known to the receiver. Inspired by the interesting results [13, 42], we study the synchronization problem in the static molecular communication systems using the constant-threshold-based detector. The novelty and contributions of this chapter can be summarized as follows.

- We formulate the error probability as a function of the start point and the constant threshold based on the large SNR assumption in order to compute the theoretical optimal start point. In addition, the optimal threshold is acquired asymptotically and the start point can accordingly be optimized. In particular, the optimal asymptotic start point is obtained when the channel gain for the current symbol is the largest.
- Based on the condition that the asymptotic start point follows, we design a data-aided synchronization metric function. The optimal start point corresponds to the largest metric result. In addition, the results of the metric function follow the Skellam distribution. Based on the metric results, the upper bounds for the absolute mean and variance of the synchronization error can be obtained. Meanwhile, the intermediate variables in the data-aided synchronization metric function can be obtained without knowing the transmitted symbols by applying the clustering method [84] to the received signals. In this regard, a blind synchronization algorithm is devised.
- The synchronization performance is quantified by deriving the Cramer-Rao lower bounds (CRLB) for two cases, i.e., the known CSI case and the unknown CSI case. The proposed synchronization schemes can achieve superior performances as the synchronization error variances and the upper bound of the synchronization variances are lower than the CRLB without CSI.

The remainder of this chapter is organized as follows. In Section 5.2, the system model is introduced. In Section 5.3, the error probability of the constant threshold and the start point is derived and the condition that the asymptotic optimal start point follows is obtained. In Section 5.4, the proposed data-aided synchronization scheme is analysed and the upper bounds for the synchronization error mean and variance are provided. In the same section,

the K-means-based blind synchronization scheme is proposed. In Section 5.5, the Cramer-Rao lower bounds are analysed for both known and unknown channel information cases. In Section 5.6, the simulation is implemented and analysed. Finally, Section 5.7 concludes this chapter.

5.2 System Model

We consider a static three-dimensional unbounded molecular communication system consisting of a point transmitter and a spherical absorbing receiver as depicted in Fig. 3.2. We ignore the process of generating particles at the transmitter and model the receiving process ideally such that particles hitting the spherical surface of the receiver are collected at once. In addition, we assume that the receiver records the received particle numbers at each time instant before performing synchronization and detection. We assume that the information particles diffuse randomly and independently through the medium via Brownian motion, thus no extra energy is needed for propagation. The ISI will always exist since the particles are assumed not to degrade quickly and will not reach to the receiver within one symbol slot.

We assume that the temperature is constant and the viscosity remains unchanged during the whole transmission, i.e., the diffusion coefficient D [65] remains unchanged. Assuming that the transmitter is located in $\mathbf{a} = (0, 0, 0)$ and the receiver is located in $\mathbf{b} = (b_x, b_y, b_z)$, the hitting rate [43, 14] is expressed as follows:

$$f_{hit}(t) = \frac{r(d-r)}{d\sqrt{4\pi Dt^3}} e^{-\frac{(d-r)^2}{4Dt}} \quad (5.1)$$

where $\|\mathbf{a} - \mathbf{b}\| = d$ is the distance between the transmitter center and the receiver center, and r denotes the radius of the receiver. We assume that information particles are released in such a short time that the release process is neglected. Denote the sampling time resolution as ϱ , then the hitting probability of an absorbing receiver at the k -th time instant is

$$\mathbb{P}_k = \int_{k\varrho}^{(k+1)\varrho} f_{hit}(t) dt = \frac{r}{d} \left\{ \operatorname{erfc}\left(\frac{d-r}{\sqrt{4D(k+1)\varrho}}\right) - \operatorname{erfc}\left(\frac{d-r}{\sqrt{4Dk\varrho}}\right) \right\}. \quad (5.2)$$

where $\operatorname{erf}(y) = \int_0^y \frac{2}{\sqrt{\pi}} e^{-x^2} dx$ and $\operatorname{erfc}(y) = 1 - \operatorname{erf}(y)$. For the sake of continuity, we set $\mathbb{P}_k = 0$ for all $k < 0$. The average received particle number in the k th time instant after the release of N_{TX} particles is $\mathbb{C}_k = N_{TX}\mathbb{P}_k$.

In our system, the binary CSK modulation scheme is adopted and the symbol slot length is assumed to be $T\varrho$ where T is an integer. Once the synchronization is performed, i.e., the start point ρ is determined, the average received particle number at the t -th time instant in the i -th

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

symbol slot is expressed as follows:

$$\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L}) = \sum_{j=-1}^L s_{i-j} \mathbb{C}_{jT+\rho+t} + \mathbb{C}_n \quad (5.3)$$

where $\mathbb{C}_n = \lambda_0 \rho$ denotes the background noise at each instant with property $C_n = T\mathbb{C}_n$ and λ_0 is the background noise power per unit time. The start point ρ is an integer and the $(i+1)$ th symbol may have effect on the received signal if $\rho > 0$ as illustrated in Fig. 5.1.

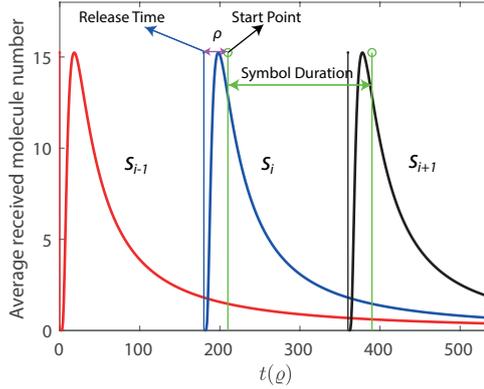


Figure 5.1: The average received molecule numbers at each time instant due to $s_{i+1} = 1$, $s_i = 1$ and $s_{i-1} = 1$. If $\rho > 0$, $\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})$ should take s_{i+1} into account.

The channel response is considered to last for $(L+1)T\rho$ for simplicity, i.e., for any $k > (L+1)T$, $\mathbb{C}_k = 0$. During the i th symbol slot, the average received particle number is expressed as follows:

$$\begin{aligned} \bar{r}_i(\rho, s_{i+1}, s_i, \dots, s_{i-L}) &= \sum_{t=0}^{T-1} \bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L}) = \sum_{t=0}^{T-1} \sum_{j=-1}^L s_{i-j} \mathbb{C}_{jT+\rho+t} + \mathbb{C}_n \\ &= \mathbb{C}_n + \sum_{j=-1}^L s_{i-j} \sum_{t=0}^{T-1} \mathbb{C}_{jT+\rho+t} = \mathbb{C}_n + \sum_{j=-1}^L s_{i-j} C_j(\rho) \end{aligned} \quad (5.4)$$

where $C_j(\rho) = \sum_{t=0}^{T-1} \mathbb{C}_{jT+\rho+t}$.

We define the signal-to-noise ratio (SNR) as

$$\text{SNR} = 10 \log_{10} \frac{\sum_{k=0}^{T-1} \mathbb{C}_k}{2\mathbb{C}_n} \quad (5.5)$$

Accordingly, the N_{TX} is computed as follows:

$$N_{TX} = \frac{2\lambda_0 T 10^{\frac{\text{SNR}}{10}}}{\sum_{k=0}^{T-1} \mathbb{P}_k} \quad (5.6)$$

as the information bits are assumed to be equiprobable.

Let us denote the received information particle number at the t th instant in the i th symbol slot given the start point ρ as $r_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})$, and without specific statement, we let $r_i(t, \rho) = r_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})$. Then $r_i(t, \rho)$ follows the Poisson distribution with the probability as:

$$\Pr(r_i(t, \rho) | \bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})) = \frac{e^{-\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})} [\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})]^{r_i(t, \rho)}}{r_i(t, \rho)!} \quad (5.7)$$

Accordingly, the i th signal $r_i(\rho) = \sum_{t=0}^{T-1} r_i(t, \rho)$ follows

$$\Pr(r_i(\rho) | \bar{r}_i(\rho, s_{i+1}, s_i, \dots, s_{i-L})) = \frac{e^{-\bar{r}_i(\rho, s_{i+1}, s_i, \dots, s_{i-L})} [\bar{r}_i(\rho, s_{i+1}, s_i, \dots, s_{i-L})]^{r_i(\rho)}}{r_i(\rho)!} \quad (5.8)$$

5.3 The Optimal Start Point

Once the information particles are released, it takes some time for each particle to propagate from the transmitter to the receiver. Therefore, our intuition is that the start point ρ should not be set to zero and needs to be optimised. In order to achieve low BER, we combine the start point and the threshold into the error probability. In this case, the optimal start point can be optimised as follows:

$$(P_e^*, \rho^*) = \min_{\rho, \tau} P_e(\rho, \tau) \quad (5.9)$$

where $P_e(\rho, \tau)$ denotes the error probability, a function of the start point ρ and the threshold τ . In the following, we present the formulation of the $P_e(\rho, \tau)$ and we provide some illustrations for the assumption we made. Moreover, we analyse the high SNR asymptotic regime to approximate and simplify the $P_e(\rho, \tau)$ and from this we obtain the asymptotic threshold and the condition the optimal start point follows.

5.3.1 Formulation of $P_e(\rho, \tau)$

The SNR is assumed to be large such that the received signal $r_i(\rho)$ can be modelled as a Gaussian distributed variable with the same mean and variance $\bar{r}_i(\rho, s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1})$ and the $P_e(\rho)$ can be formulated in an analytical form. Let us define

$$\mu_i(\rho, k) = \bar{r}_i(\rho, s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1}) \quad (5.10)$$

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

where $[s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1}]$ is the binary expression of k . The probability density function of receiving $r_i(\rho)$ information particles in the i th symbol slot can be written as follows:

$$\Pr(r_i(\rho)|\bar{r}_i(\rho, s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1})) = \Pr(r_i(\rho)|\mu_i(\rho, k)) = \frac{1}{\sqrt{2\pi\mu_i(\rho, k)}} e^{-\frac{(r_i(\rho)-\mu_i(\rho, k))^2}{2\mu_i(\rho, k)}} \quad (5.11)$$

Then the error probability is expressed as:

$$P_e(\rho, \tau) = \frac{1}{2} \left(\int_{\tau}^{+\infty} \Pr(r_i(\rho)|s_i = 0) dr_i(\rho) + \int_{-\infty}^{\tau} \Pr(r_i(\rho)|s_i = 1) dr_i(\rho) \right)$$

where

$$\Pr(r_i(\rho)|s_i = 0) = \frac{1}{2^{L+1}} \sum_{s_i=0} \Pr(r_i(\rho)|\bar{r}_i(\rho, s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1})) = \sum_{k=0}^{2^{L+1}-1} \frac{\Pr(r_i(\rho)|\mu_i(\rho, k))}{2^{L+1}}$$

$$\Pr(r_i(\rho)|s_i = 1) = \frac{1}{2^{L+1}} \sum_{s_i=1} \Pr(r_i(\rho)|\bar{r}_i(\rho, s_i, s_{i-1}, \dots, s_{i-L}, s_{i+1})) = \sum_{k=2^{L+1}}^{2^{L+2}-1} \frac{\Pr(r_i(\rho)|\mu_i(\rho, k))}{2^{L+1}}$$

are the probabilities of receiving $r_i(\rho)$ particles conditioned upon $s_i = 0$ and $s_i = 1$, respectively. The $\frac{1}{2} \int_{\tau}^{+\infty} \Pr(r_i(\rho)|s_i = 0) dr_i(\rho)$ and $\frac{1}{2} \int_{-\infty}^{\tau} \Pr(r_i(\rho)|s_i = 1) dr_i(\rho)$ represent the false alarm probability and the miss detection probability, respectively. Thus, the error probability can be formulated as follows:

$$\begin{aligned} P_e(\rho, \tau) &= \frac{1}{2^{L+2}} \sum_{k=0}^{2^{L+1}-1} \int_{\tau}^{+\infty} \Pr(r_i(\rho)|\mu_i(\rho, k)) dr_i(\rho) + \sum_{k=2^{L+1}}^{2^{L+2}-1} \int_{-\infty}^{\tau} \Pr(r_i(\rho)|\mu_i(\rho, k)) dr_i(\rho) \\ &= \frac{1}{2^{L+2}} \sum_{k=0}^{2^{L+1}-1} \int_{\tau}^{+\infty} \Pr(r_i(\rho)|\mu_i(\rho, k)) dr_i(\rho) + \int_{-\infty}^{\tau} \Pr(r_i(\rho)|\mu_i(\rho, 2^{L+2} - 1 - k)) dr_i(\rho) \end{aligned} \quad (5.12)$$

We assume that the symbol duration $T\rho$ is not too short such that there exists a range $[\rho_0, \rho_1]$ for $\rho_0 \leq \rho \leq \rho_1$, $\bar{r}_i(\rho, s_i = 0, s_{i-1} = 1, \dots, s_{i-L} = 1, s_{i+1} = 1) < \bar{r}_i(\rho, s_i = 1, s_{i-1} = 0, \dots, s_{i-L} = 0, s_{i+1} = 1)$ i.e., $\mu_i(\rho, 2^{L+1} - 1) < \mu_i(\rho, 2^{L+1})$ holds true. Since out of the range, no matter what threshold is selected, the error probability is large as depicted in the shaded areas in the Fig.5.2. It's worth noting that the ρ_0 can be negative as the negative start point means that the receiver start to count the received particle number even before the transmitter starts to release the particles. In this regard, the optimal threshold should be within the range $[\mu_i(\rho, 2^{L+1} - 1), \mu_i(\rho, 2^{L+1})]$ to ensure the low error probability as illustrated in the Fig. 5.2a.

We have the property for $\mu_i(\rho, k)$ in (5.13) which will facilitate the analysis of $P_e(\rho, \tau)$ in the

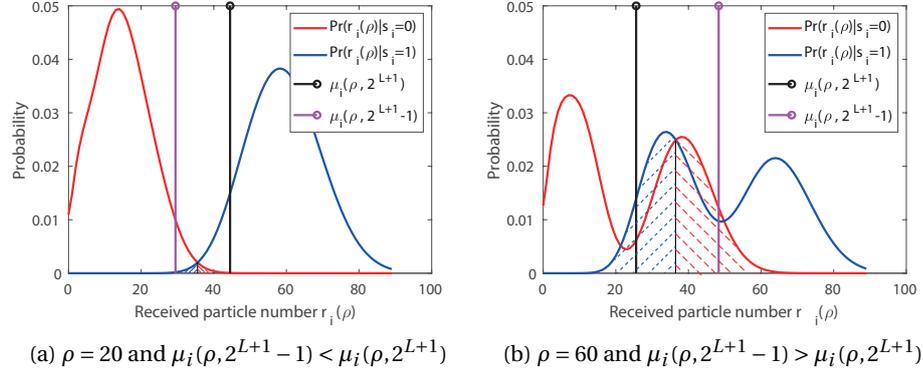


Figure 5.2: Comparison of $P_e(\rho, \tau)$ (shaded area) for ρ in and out of the range $[\rho_0, \rho_1]$.

following results.

$$\begin{aligned} \mu_i(\rho, k) + \mu_i(\rho, 2^{L+2} - 1 - k) &= \bar{r}_i(\rho, s_i, \dots, s_{i-L}, s_{i+1}) + \bar{r}_i(\rho, 1 - s_i, \dots, 1 - s_{i-L}, 1 - s_{i+1}) \\ &= 2C_n + \sum_{t=0}^{(L+1)T-1} C_j \end{aligned} \quad (5.13)$$

Since for the case where $\tau > \mu$, we have the following result.

$$\int_{\tau}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{\frac{\tau-\mu}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \operatorname{erfc}\left(\frac{\tau-\mu}{\sqrt{2}\sigma}\right) \quad (5.14)$$

and for the case where $\tau < \mu$,

$$\begin{aligned} \int_{-\infty}^{\tau} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= \int_{-\infty}^{\tau-\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx = \int_{-\infty}^{\frac{\tau-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{\frac{|\tau-\mu|}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \operatorname{erfc}\left(\frac{\mu-\tau}{\sqrt{2}\sigma}\right) \end{aligned} \quad (5.15)$$

The (5.12) can be written as follows:

$$P_e(\rho, \tau) = \frac{1}{2^{L+2}} \sum_{k=0}^{2^{L+1}-1} \operatorname{erfc}\left(\frac{\tau - \mu_i(\rho, k)}{\sqrt{2\mu_i(\rho, k)}}\right) + \operatorname{erfc}\left(\frac{\mu_i(\rho, 2^{L+2} - 1 - k) - \tau}{\sqrt{2\mu_i(\rho, 2^{L+2} - 1 - k)}}\right) \quad (5.16)$$

Based on (5.16), we evaluate the $P_e(\rho, \tau)$ for all possible ρ and τ in Fig. 5.3a and evaluate the $P_e(\rho, \tau)$ over the range $[\rho_0, \rho_1] \times [\mu_i(\rho, 2^{L+1}-1), \mu_i(\rho, 2^{L+1})]$ in Fig. 5.3b. Comparing the two figures, we confirm that optimizing $P_e(\rho, \tau)$ in the range $[\rho_0, \rho_1] \times [\mu_i(\rho, 2^{L+1}-1), \mu_i(\rho, 2^{L+1})]$ is reasonable. The theoretical optimal start point can be obtained by searching the minimum

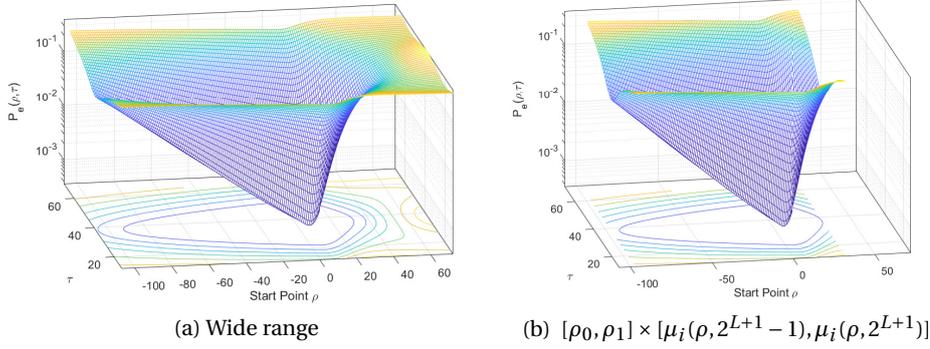


Figure 5.3: $P_e(\rho, \tau)$ as a function of ρ and τ (SNR is 25 dB)

error probability over this effective range and is denoted as ρ^* .

5.3.2 Asymptotic Optimal Start Point

In order to gain more insights from the aforementioned $P_e(\rho, \tau)$, based on the large SNR assumption, i.e., N_{TX} is large, and according to the searching range $[\mu_i(\rho, 2^{L+1} - 1), \mu_i(\rho, 2^{L+1})]$ of the optimal threshold, we have the following results.

Let us denote $\mu_i(\rho, k) = N_{TX}P_i(\rho, k)$, then $P_e(\rho, \tau)$ can reformulated as follows:

$$\begin{aligned}
 P_e(\rho, \tau) &= \frac{1}{2^{L+2}} \sum_{k=0}^{2^{L+1}-1} \operatorname{erfc}\left(\frac{\tau - N_{TX}P_i(\rho, k)}{\sqrt{2N_{TX}P_i(\rho, k)}}\right) + \operatorname{erfc}\left(\frac{N_{TX}P_i(\rho, 2^{L+2}-1-k) - \tau}{\sqrt{2N_{TX}P_i(\rho, 2^{L+2}-1-k)}}\right) \\
 &= \frac{1}{2^{L+2}} \sum_{k=0}^{2^{L+1}-1} \operatorname{erfc}\left(\frac{\frac{\tau}{N_{TX}} - P_i(\rho, k)}{\sqrt{2P_i(\rho, k)}}\sqrt{N_{TX}}\right) + \operatorname{erfc}\left(\frac{P_i(\rho, 2^{L+2}-1-k) - \frac{\tau}{N_{TX}}}{\sqrt{2P_i(\rho, 2^{L+2}-1-k)}}\sqrt{N_{TX}}\right) \quad (5.17)
 \end{aligned}$$

Based on (5.10), $\mu_i(\rho, 2^{L+1} - 1) > \mu_i(\rho, k)$ and $\mu_i(\rho, 2^{L+1}) < \mu_i(\rho, 2^{L+2} - 1 - k)$ for $0 \leq k \leq 2^{L+1} - 2$, then $P_i(\rho, 2^{L+1} - 1) > P_i(\rho, k)$ and $P_i(\rho, 2^{L+1}) < P_i(\rho, 2^{L+2} - 1 - k)$ for $0 \leq k \leq 2^{L+1} - 2$. Meanwhile $\frac{\tau}{N_{TX}} > P_i(\rho, k)$ and $P_i(\rho, 2^{L+2} - 1 - k) > \frac{\tau}{N_{TX}}$ for $0 \leq k \leq 2^{L+1} - 1$ as $\mu_i(\rho, 2^{L+1} - 1) \leq \tau \leq \mu_i(\rho, 2^{L+1})$. Since if $N \rightarrow \infty$ and $0 < \alpha_1 < \alpha_2$, $\operatorname{erfc}(\alpha_1 N) \gg \operatorname{erfc}(\alpha_2 N)$. In this context, the error probability can be approximated as follows:

$$P_e(\rho, \tau) \approx \frac{1}{2^{L+2}} \left(\operatorname{erfc}\left(\frac{\tau - N_{TX}P_i(\rho, 2^{L+1} - 1)}{\sqrt{2N_{TX}P_i(\rho, 2^{L+1} - 1)}}\right) + \operatorname{erfc}\left(\frac{N_{TX}P_i(\rho, 2^{L+1}) - \tau}{\sqrt{2N_{TX}P_i(\rho, 2^{L+1})}}\right) \right)$$

$$= \frac{1}{2^{L+2}} \left(\operatorname{erfc}\left(\frac{\tau - \mu_i(\rho, 2^{L+1} - 1)}{\sqrt{2\mu_i(\rho, 2^{L+1} - 1)}}\right) + \operatorname{erfc}\left(\frac{\mu_i(\rho, 2^{L+1}) - \tau}{\sqrt{2\mu_i(\rho, 2^{L+1})}}\right) \right) \quad (5.18)$$

We can obtain the asymptotic threshold at the start point ρ in the following lemma.

Lemma 5.1 *The asymptotic threshold when SNR approaches to infinity is*

$$\begin{aligned} \tau^*(\rho) &= \sqrt{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1) \left[1 + \frac{\ln \frac{\mu_i(\rho, 2^{L+1})}{\mu_i(\rho, 2^{L+1} - 1)}}{\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)} \right]} \\ &\simeq \sqrt{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1)} \end{aligned} \quad (5.19)$$

Proof 5.1 *The asymptotic threshold is derived from $\frac{dP_e(\rho, \tau)}{d\tau} = 0$ of (5.18), i.e.,*

$$\begin{aligned} \frac{1}{\sqrt{2\pi\mu_i(\rho, 2^{L+1} - 1)}} e^{-\frac{[\tau - \mu_i(\rho, 2^{L+1} - 1)]^2}{2\mu_i(\rho, 2^{L+1} - 1)}} &= \frac{1}{\sqrt{2\pi\mu_i(\rho, 2^{L+1})}} e^{-\frac{[\tau - \mu_i(\rho, 2^{L+1})]^2}{2\mu_i(\rho, 2^{L+1})}} \\ \frac{e^{\frac{[\tau - \mu_i(\rho, 2^{L+1})]^2}{2\mu_i(\rho, 2^{L+1})}}}{e^{\frac{[\tau - \mu_i(\rho, 2^{L+1} - 1)]^2}{2\mu_i(\rho, 2^{L+1} - 1)}}} &= \frac{\sqrt{\mu_i(\rho, 2^{L+1} - 1)}}{\sqrt{\mu_i(\rho, 2^{L+1})}} \\ \frac{e^{\frac{\tau^2[\mu_i(\rho, 2^{L+1} - 1) - \mu_i(\rho, 2^{L+1})] + \mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1) - [\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)]}{2\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1)}}}{e^{\frac{\tau^2[\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)] - [\mu_i(\rho, 2^{L+1} - 1) - \mu_i(\rho, 2^{L+1})]}{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1)}}} &= \frac{\sqrt{\mu_i(\rho, 2^{L+1} - 1)}}{\sqrt{\mu_i(\rho, 2^{L+1})}} \\ e^{\left[\frac{\tau^2}{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1)} - 1\right][\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)]} &= \frac{\mu_i(\rho, 2^{L+1} - 1)}{\mu_i(\rho, 2^{L+1})} \\ e^{\left[\frac{\tau^2}{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1} - 1)} - 1\right][\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)]} &= \frac{\mu_i(\rho, 2^{L+1})}{\mu_i(\rho, 2^{L+1} - 1)} \end{aligned}$$

Calculating the natural logarithm of both sides and with some mathematical manipulations, the analytical asymptotic threshold is obtained. Since SNR is large and $\mu_i(\rho, 2^{L+1}) >$

$\mu_i(\rho, 2^{L+1} - 1)$, $\frac{\ln\left(\frac{\mu_i(\rho, 2^{L+1})}{\mu_i(\rho, 2^{L+1} - 1)}\right)}{\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)}$ can be regarded as the approximation of the slope of $\ln(x)$ at $x = \mu_i(\rho, 2^{L+1})$, i.e.,

$$\frac{\ln\left(\frac{\mu_i(\rho, 2^{L+1})}{\mu_i(\rho, 2^{L+1} - 1)}\right)}{\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)} = \frac{\ln(\mu_i(\rho, 2^{L+1})) - \ln(\mu_i(\rho, 2^{L+1} - 1))}{\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 2^{L+1} - 1)} \simeq \frac{1}{\mu_i(\rho, 2^{L+1})}$$

which tends to be zero if the SNR is large. The approximation of the threshold is obtained and this concludes the proof.

According to Lemma.5.1, we have the following result by inserting the approximated asymptotic threshold into (5.18).

Lemma 5.2 *The optimal start point ρ^* maximizes $C_0(\rho^*)$ when SNR approaches to infinity.*

Proof 5.2 *According to the approximation $\operatorname{erfc}(x) \simeq \frac{1}{6}e^{-x^2} + \frac{1}{2}e^{-\frac{3}{4}x^2}$ for $x > 0$, we have the following results.*

$$P_e(\rho, \tau^*) \simeq \frac{1}{2^{L+2}} \left(\frac{1}{6} e^{-\frac{(\mu_i(\rho, 2^{L+1}-1)-\tau^*)^2}{2\mu_i(\rho, 2^{L+1}-1)}} + \frac{1}{2} e^{-\frac{3}{4} \frac{(\mu_i(\rho, 2^{L+1}-1)-\tau^*)^2}{2\mu_i(\rho, 2^{L+1}-1)}} + \frac{1}{6} e^{-\frac{(\mu_i(\rho, 2^{L+1})-\tau^*)^2}{2\mu_i(\rho, 2^{L+1})}} + \frac{1}{2} e^{-\frac{3}{4} \frac{(\mu_i(\rho, 2^{L+1})-\tau^*)^2}{2\mu_i(\rho, 2^{L+1})}} \right) \quad (5.20)$$

Observing from the (5.20), we evince that the $P_e(\rho, \tau^*)$ is approximately composed of a sum in terms of $[e^{-a \frac{(\mu_i(\rho, 2^{L+1}-1)-\tau^*)^2}{\mu_i(\rho, 2^{L+1}-1)}} + e^{-a \frac{(\mu_i(\rho, 2^{L+1})-\tau^*)^2}{\mu_i(\rho, 2^{L+1})}}]$ for $a > 0$. Meanwhile, the $e^{-a \frac{(\mu_i(\rho, 2^{L+1}-1)-\tau^*)^2}{\mu_i(\rho, 2^{L+1}-1)}} + e^{-a \frac{(\mu_i(\rho, 2^{L+1})-\tau^*)^2}{\mu_i(\rho, 2^{L+1})}}$ can be written as follows:

$$\begin{aligned} & e^{-a \frac{(\mu_i(\rho, 2^{L+1}-1)-\tau^*)^2}{\mu_i(\rho, 2^{L+1}-1)}} + e^{-a \frac{(\mu_i(\rho, 2^{L+1})-\tau^*)^2}{\mu_i(\rho, 2^{L+1})}} \\ = & e^{-a \frac{(\mu_i(\rho, 2^{L+1}-1) - \sqrt{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1}-1)})^2}{\mu_i(\rho, 2^{L+1}-1)}} + e^{-a \frac{(\mu_i(\rho, 2^{L+1}) - \sqrt{\mu_i(\rho, 2^{L+1})\mu_i(\rho, 2^{L+1}-1)})^2}{\mu_i(\rho, 2^{L+1})}} \\ = & 2e^{-a(\sqrt{\mu_i(\rho, 2^{L+1})} - \sqrt{\mu_i(\rho, 2^{L+1}-1)})^2} = 2e^{-a \left[\sqrt{\mu_i(\rho, 2^{L+1})} - \sqrt{2C_n + \sum_{i=0}^{(L+1)T-1} C_j} - \mu_i(\rho, 2^{L+1}) \right]^2} \quad (5.21) \end{aligned}$$

Since $\mu_i(\rho, 2^{L+1}) > \mu_i(\rho, 2^{L+1}-1)$ and $\mu_i(\rho, 2^{L+1}) + \mu_i(\rho, 2^{L+1}-1) = 2C_n + \sum_{i=0}^{(L+1)T-1} C_j$, the (5.21) is least when $[\sqrt{\mu_i(\rho, 2^{L+1})} - \sqrt{\mu_i(\rho, 2^{L+1}-1)}]$ is the largest or equivalently $\mu_i(\rho, 2^{L+1})$ is the largest which concludes the proof.

5.4 The Proposed Synchronization Schemes

According to the aforementioned asymptotic analysis, if the SNR is large, the minimum BER is achieved when the $C_0(\rho)$, i.e., $\mu_i(\rho, 2^{L+1})$ is the largest. To estimate this value, we need to estimate $\bar{r}_i(\rho, s_{i+1} = 0, s_i = 1, s_{i-1} = 0, \dots, s_{i-L} = 0)$, i.e., provided the candidate start point ρ , all the signals $r_n(\rho, s_{n+1} = 0, s_n = 1, s_{n-1} = 0, \dots, s_{n-L} = 0)$ whose current symbol $s_n = 1$ and $s_{n-j} = 0$ for $-1 \leq j \leq L$ and $j \neq 0$, are averaged. In this case, most of the received signals would not be used, which is a waste of energy. To this end, we propose a data-aided synchronization metric function that can exploit all the received signals.

5.4.1 Data-Aided Synchronization Algorithm

Note that $\mu_i(\rho, 2^{L+1}) - \mu_i(\rho, 0) = \mu_i(\rho, 2^{L+1} + k) - \mu_i(\rho, k) = C_0(\rho)$ for $0 \leq k \leq 2^{L+1} - 1$ from the (5.10), thus

$$\begin{aligned} & E_{s_{i-j}, -1 \leq j \leq L} [\bar{r}_i(\rho, s_i = 1, \dots, s_{i-L}, s_{i+1}) - \bar{r}_i(\rho, s_i = 0, \dots, s_{i-L}, s_{i+1})] \\ &= \bar{r}_i(\rho, s_i = 1) - \bar{r}_i(\rho, s_i = 0) = \frac{1}{2^{L+1}} \sum_{k=0}^{2^{L+1}-1} \mu_i(\rho, 2^{L+1} + k) - \mu_i(\rho, k) = C_0(\rho) \end{aligned} \quad (5.22)$$

where $\bar{r}_i(\rho, s_i = 1)$ and $\bar{r}_i(\rho, s_i = 0)$ denote the average received signals when $s_i = 1$ and $s_i = 0$, respectively. In theory, $\bar{r}_i(\rho, s_i = 1)$ and $\bar{r}_i(\rho, s_i = 0)$ should be computed as follows:

$$\begin{aligned} \bar{r}_i(\rho, s_i = 1) &= \lim_{K \rightarrow \infty} \frac{1}{N_{s_i=1}} \sum_{m=0}^{K-1} r_m(\rho, s_m = 1, \dots, s_{m-L}, s_{m+1}) \\ \bar{r}_i(\rho, s_i = 0) &= \lim_{K \rightarrow \infty} \frac{1}{N_{s_i=0}} \sum_{m=0}^{K-1} r_m(\rho, s_m = 0, \dots, s_{m-L}, s_{m+1}) \end{aligned} \quad (5.23)$$

where K denotes the length of the transmitted symbols, $N_{s_i=1}$ and $N_{s_i=0}$ represent the numbers of transmitted binary one and binary zero, respectively. The (5.22) and (5.23) provide us an option to design the synchronization metric function as follows:

$$F(\rho) = \frac{1}{N_{s_i=1}} \sum_{m=0}^{K-1} r_m(\rho, s_m = 1, \dots, s_{m-L}, s_{m+1}) - \frac{1}{N_{s_i=0}} \sum_{m=0}^{K-1} r_m(\rho, s_m = 0, \dots, s_{m-L}, s_{m+1}) \quad (5.24)$$

and $\lim_{K \rightarrow \infty} F(\rho) = C_0(\rho)$ such that this metric function emulates the $C_0(\rho)$.

However, this metric does not provide enough insights for performance analysis. Since $E[N_{s_i=1}] = E[N_{s_i=0}] = K/2$,

$$\begin{aligned} \lim_{K \rightarrow \infty} \sum_{m=0}^{K-1} r_m(\rho, s_m = 1, \dots, s_{m-L}, s_{m+1}) &= \frac{K}{2} \bar{r}_i(\rho, s_i = 1) \\ \lim_{K \rightarrow \infty} \sum_{m=0}^{K-1} r_m(\rho, s_m = 0, \dots, s_{m-L}, s_{m+1}) &= \frac{K}{2} \bar{r}_i(\rho, s_i = 0) \end{aligned} \quad (5.25)$$

then, a new metric function is proposed as follows:

$$\begin{aligned} F(\rho) &= \sum_{m=0}^{K-1} r_m(\rho, s_m = 1, \dots, s_{m-L}, s_{m+1}) - \sum_{m=0}^{K-1} r_m(\rho, s_m = 0, \dots, s_{m-L}, s_{m+1}) \\ &= \sum_{m=0}^{K-1} r_m(\rho, s_m, \dots, s_{m-L}, s_{m+1}) (2s_m - 1) \end{aligned} \quad (5.26)$$

In practice, this metric $F(\rho)$ is associated with the transmitted symbols $\mathbf{s} = [s_1, \dots, s_K]$ as K is

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

limited. The corresponding estimated start point is obtained using the following formula:

$$\hat{\rho}_{\mathbf{s}}^* = \max_{\rho} F(\rho) \quad (5.27)$$

Denote $\rho_{\mathbf{s}}^*$ as the theoretical optimal start point associated with the \mathbf{s} . It is computed as follows:

$$\rho_{\mathbf{s}}^* = \max_{\rho} E_{\tilde{\mathbf{r}}} [F(\rho)] = \max_{\rho} \sum_{m=0}^{K-1} \tilde{r}_m(\rho, s_m, \dots, s_{m-L}, s_{m+1}) (2s_m - 1) \quad (5.28)$$

where $\tilde{\mathbf{r}} = [r_1(\rho, s_1, s_0, \dots, s_{-L}), \dots, r_K(\rho, s_{K+1}, s_K, \dots, s_{K-L})]$ and we set $s_{-j} = 0$ for $1 \leq j \leq L$.

We have the following results to discuss the distribution of the $F(\rho)$ based on the Skellam distribution. A Skellam distributed variable [90] is the difference z of two independent Poisson variables x and y whose mean values are μ_x and μ_y , respectively. Let us denote $z = x - y \sim \text{Skellam}(\mu_x, \mu_y)$. It's worth mentioning that the metric function has the following result as the sum of independent Skellam distributed variables still follows the Skellam distribution.

Lemma 5.3 *The $F(\rho)$ follows the Skellam distribution, i.e., $F(\rho) \sim \text{Skellam}(\mu_1(\rho), \mu_0(\rho))$ where $\mu_1(\rho) = C_n \sum_{m=0}^{K-1} s_m + \sum_{j=-1}^L C_j(\rho) \sum_{m=0}^{K-1} s_{m-j} s_m$ and $\mu_0(\rho) = KC_n + \sum_{j=-1}^L C_j(\rho) \sum_{m=0}^{K-1} s_m - \mu_1(\rho)$ with properties $E_s[\mu_1(\rho)] = \frac{K}{2}C_n + \frac{K}{4}C_0(\rho) + \frac{K}{4} \sum_{j=-1}^L C_j(\rho)$ and $E_s[\mu_0(\rho)] = \frac{K}{2}C_n - \frac{K}{4}C_0(\rho) + \frac{K}{4} \sum_{j=-1}^L C_j(\rho)$.*

Proof 5.3 *From the definition of Skellam distribution, we have*

$$\begin{aligned} \mu_1(\rho) &= E_r \left\{ \sum_{m=0}^{K-1} r_m(\rho, s_{m+1}, s_m = 1, \dots, s_{m-L}) \right\} = \sum_{m=0}^{K-1} \tilde{r}_m(\rho, s_{m+1}, s_m = 1, \dots, s_{m-L}) \\ &= \sum_{m=0}^{K-1} (C_n + \sum_{j=-1}^L s_{m-j} C_j(\rho)) s_m = C_n \sum_{m=0}^{K-1} s_m + \sum_{j=-1}^L C_j(\rho) \sum_{m=0}^{K-1} s_{m-j} s_m \end{aligned}$$

$$\begin{aligned} \mu_0(\rho) &= E_r \left\{ \sum_{m=0}^{K-1} r_m(\rho, s_{m+1}, s_m = 0, \dots, s_{m-L}) \right\} = \sum_{m=0}^{K-1} (C_n + \sum_{j=-1}^L s_{m-j} C_j(\rho)) (1 - s_m) \\ &= C_n (K - \sum_{m=0}^{K-1} s_m) + \sum_{j=-1}^L C_j(\rho) \sum_{m=0}^{K-1} s_{m-j} (1 - s_m) = KC_n + \sum_{j=-1}^L C_j(\rho) \sum_{m=0}^{K-1} s_m - \mu_1(\rho) \end{aligned}$$

Since $E[s_m] = E[s_m s_m] = \frac{1}{2}$ and $E[s_m s_{m-j}] = \frac{1}{4}$ for $j \neq 0$, we can obtain $E_s[\mu_1(\rho)]$ and $E_s[\mu_0(\rho)]$ in the lemma, which concludes the proof.

Based on this lemma, we have the following proposition.

Proposition 5.1 $\lim_{K \rightarrow \infty} \rho_s^* = \rho^*$

Proof 5.4 Since $\lim_{K \rightarrow \infty} \frac{\sum_{m=1}^K s_m}{K} = \frac{1}{2}$ and $\lim_{K \rightarrow \infty} \frac{\sum_{m=1}^K s_m s_{m-j}}{K} = \frac{1}{4}$ for $j \neq 0$, then $\lim_{K \rightarrow \infty} \mu_1(\rho) = E_s[\mu_1(\rho)]$ and $\lim_{K \rightarrow \infty} \mu_0(\rho) = E_s[\mu_0(\rho)]$. Based on (5.28), we have

$$\begin{aligned} \lim_{K \rightarrow \infty} \rho_s^* &= \lim_{K \rightarrow \infty} \max_{\rho} E_{\mathbf{r}}[F(\rho)] = \max_{\rho} \lim_{K \rightarrow \infty} E_{\mathbf{r}}[F(\rho)] = \max_{\rho} \lim_{K \rightarrow \infty} \mu_1(\rho) - \mu_0(\rho) \\ &= \max_{\rho} E_s[\mu_1(\rho)] - E_s[\mu_0(\rho)] = \max_{\rho} \lim_{K \rightarrow \infty} \frac{K}{2} C_0(\rho) = \rho^* \end{aligned}$$

This concludes the proof.

Therefore, to let ρ_s^* be as close to ρ^* as possible, the length K needs to be large.

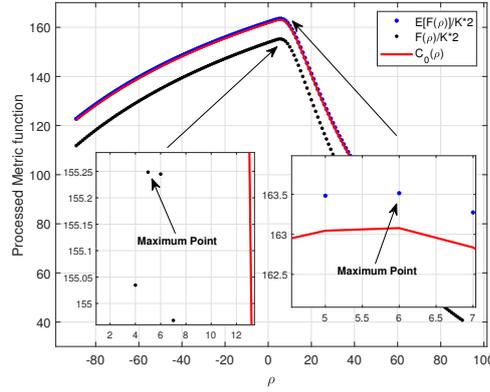


Figure 5.4: Comparison of $\frac{E[F(\rho)]}{K/2}$, $\frac{F(\rho)}{K/2}$ and $C_0(\rho)$ (SNR is 30 dB and $K = 128$)

In order to show that the average metric results divided by $K/2$, i.e., $\frac{E(F(\rho))}{K/2}$ is similar to the $C_0(\rho)$, we need to compute the expectation of $F(\rho)$ over all possible cases of \mathbf{s} and \mathbf{r} . Instead of calculating the $E(F(\rho))$ over all cases of \mathbf{s} (the length of \mathbf{s} is K) which is an impossible task, we just calculate $E(F(\rho))$ by averaging $F(\rho)$ over a limited number of cases of \mathbf{s} (\mathbf{s} is generated for K_s times and constitutes a set Θ). As shown in Fig. 5.4, on average, $F(\rho)$ does emulate the $C_0(\rho)$ and one realization of $F(\rho)$ has the similar shape as $C_0(\rho)$. We also observe that in one realization, the $\hat{\rho}_s^*$ is close to ρ^* .

To evaluate the synchronization performance, e.g., the mean and variance of the synchronization errors, we need the probabilities for the estimated optimal start point $\hat{\rho}_s^*$, i.e., $\Pr(\hat{\rho}_s^* = \max_{\rho} F(\rho))$. For simplicity, we denote $\Pr(\hat{\rho}_s^*) = \Pr(\hat{\rho}_s^* = \max_{\rho} F(\rho))$. Since the exact $\Pr(\hat{\rho}_s^*)$ are

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

not easy to obtain, what can be computed are the upper bounds of $\Pr(\hat{\rho}_s^*)$ that allow us to compute the upper bounds for the mean and variance of the synchronization errors. Given the transmitted symbols \mathbf{s} , we have the following relation:

$$\Pr(\hat{\rho}_s^*) = \Pr(F(\hat{\rho}_s^*) \geq F(\rho), \rho \neq \hat{\rho}_s^*) \leq \Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0) \quad (5.29)$$

Based on the Lemma. 5.3, we first introduce the following results for computing $\Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0)$.

Lemma 5.4 *The $F(\rho) - F(\rho + k) \sim \text{Skellam}(\mu'_1(k, \rho), \mu'_0(k, \rho))$ given the transmitted sequence \mathbf{s} and $k > 0$ where*

$$\mu'_1(k, \rho) = 2 \frac{kC_n}{T} \sum_{m=0}^{K-1} (s_m - s_{m-1})^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{k-1} \mathbb{C}_{jT+\rho+t} \sum_{m=0}^{K-1} s_{m-j} (s_m - s_{m-1})^+ \quad (5.30)$$

$$\mu'_0(k, \rho) = 2 \frac{kC_n}{T} \sum_{m=0}^{K-1} (s_{m-1} - s_m)^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{k-1} \mathbb{C}_{jT+\rho+t} \sum_{m=0}^{K-1} s_{m-j} (s_{m-1} - s_m)^+ \quad (5.31)$$

and $(x)^+ = 1$ if $x > 0$ else $(x)^+ = 0$. There exist the following equations:

$$E_s[\mu'_1(k, \rho)] = \frac{K}{2} \left(\frac{kC_n}{T} + \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+jT} + \frac{1}{2} \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t-T} \right) \quad (5.32)$$

$$E_s[\mu'_0(k, \rho)] = \frac{K}{2} \left(\frac{kC_n}{T} + \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+T} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+jT} + \frac{1}{2} \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t-T} \right) \quad (5.33)$$

Proof 5.5 *See Appendix. 5.8.*

Practically, either $\hat{\rho}_s^* \leq \rho_s^*$ or $\hat{\rho}_s^* \geq \rho_s^*$. Thus we need to calculate $\Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0 | \hat{\rho}_s^* \leq \rho_s^*)$ and $\Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0 | \hat{\rho}_s^* \geq \rho_s^*)$. If $\hat{\rho}_s^* \leq \rho_s^*$, based on Lemma. 5.4, we have $F(\hat{\rho}_s^*) - F(\rho_s^*) \sim \text{Skellam}(\mu''_1(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*), \mu''_0(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*))$ where

$$\mu''_1(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) = 2 \frac{(\rho_s^* - \hat{\rho}_s^*)C_n}{T} \sum_{m=0}^{K-1} (s_m - s_{m-1})^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{\rho_s^* - \hat{\rho}_s^* - 1} \mathbb{C}_{jT+\hat{\rho}_s^*+t} \sum_{m=0}^{K-1} s_{m-j} (s_m - s_{m-1})^+ \quad (5.34)$$

5.4. The Proposed Synchronization Schemes

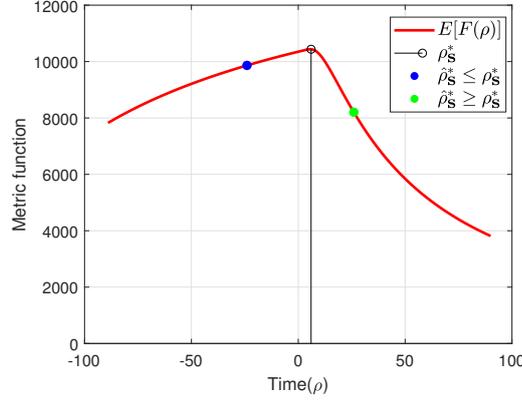


Figure 5.5: Comparison of $E[F(\hat{\rho}_s^*)]$ and $E[F(\rho_s^*)]$ (SNR is 30 dB and $K = 128$)

$$\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) = 2 \frac{(\rho_s^* - \hat{\rho}_s^*) C_n}{T} \sum_{m=0}^{K-1} (s_{m-1} - s_m)^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{\rho_s^* - \hat{\rho}_s^* - 1} C_{jT + \hat{\rho}_s^* + t} \sum_{m=0}^{K-1} s_{m-j} (s_{m-1} - s_m)^+ \quad (5.35)$$

As depicted in the Fig. 5.5, if $\hat{\rho}_s^* \leq \rho_s^*$, we have $E[F(\hat{\rho}_s^*) - F(\rho_s^*)] = \mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) - \mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) \leq 0$, i.e., $\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) \leq \mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$. According to the property of the Skellam distribution [91], if $X \sim \text{Skellam}(a, b)$ and $a < b$,

$$\Pr(X \geq 0) \leq e^{-(\sqrt{a} - \sqrt{b})^2} \quad (5.36)$$

then we have

$$\Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0 | \hat{\rho}_s^* \leq \rho_s^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)})^2} \quad (5.37)$$

If $\hat{\rho}_s^* \geq \rho_s^*$, we have $F(\hat{\rho}_s^*) - F(\rho_s^*) \sim \text{Skellam}(\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*), \mu_0''(\hat{\rho}_s^*) | \hat{\rho}_s^* \geq \rho_s^*)$ where

$$\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*) = 2 \frac{(\hat{\rho}_s^* - \rho_s^*) C_n}{T} \sum_{i=0}^{K-1} (s_i - s_{i-1})^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{\hat{\rho}_s^* - \rho_s^* - 1} C_{\rho_s^* + t + jT} \sum_{i=0}^{K-1} s_{i-j} (s_i - s_{i-1})^+ \quad (5.38)$$

$$\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*) = 2 \frac{(\hat{\rho}_s^* - \rho_s^*) C_n}{T} \sum_{i=0}^{K-1} (s_{i-1} - s_i)^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{\hat{\rho}_s^* - \rho_s^* - 1} C_{\rho_s^* + t + jT} \sum_{i=0}^{K-1} s_{i-j} (s_{i-1} - s_i)^+ \quad (5.39)$$

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

$\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*) > \mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$ as illustrated in Fig. 5.5. Accordingly, we have

$$\Pr(F(\hat{\rho}_s^*) - F(\rho_s^*) \geq 0 | \hat{\rho}_s^* \geq \rho_s^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)})^2} \quad (5.40)$$

Thus, we have

$$\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)})^2} \quad (5.41)$$

$$\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)})^2} \quad (5.42)$$

Let $[\rho_{\min}, \rho_{\max}]$ be the effective range for $\hat{\rho}_s^*$ and we set $\rho^* - \rho_{\min} = \rho_{\max} - \rho^* = 10$, i.e., we are interested in the estimated start point $\hat{\rho}_s^*$ that follows $\rho_{\min} < \hat{\rho}_s^* < \rho_{\max}$, since out of this range, $\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$ and $\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)$ tend to be zero exponentially.

As the $(s_m - s_{m-1})^+$ and $s_{m-j}(s_m - s_{m-1})^+$ etc. are random variables, the $\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$, $\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$, $\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)$ and $\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)$ are also random variables. In this context, it is impossible to average $e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)})^2}$ and $e^{-(\sqrt{\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)} - \sqrt{\mu_0''(\hat{\rho}_s^* | \hat{\rho}_s^* \geq \rho_s^*)})^2}$ over all cases of \mathbf{s} . To gain more insights, we assume that the length of the transmitted symbols K is large such that the probability $\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$ tends to be the probability for the estimated start point $\hat{\rho}^*$ conditioned upon $\hat{\rho}^* \leq \rho^*$, i.e., $\Pr(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*) \rightarrow \Pr(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*)$. In this case, the $\mu_1''(\hat{\rho}_s^* | \hat{\rho}_s^* \leq \rho_s^*)$ turns into $\mu_1''(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*)$ which is expressed as follows:

$$\mu_1''(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) = \frac{K}{2} \left(\frac{(\rho^* - \hat{\rho}^*) C_n}{T} + \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\hat{\rho}^* + t} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{jT + \hat{\rho}^* + t} + \frac{1}{2} \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\hat{\rho}^* + t - T} \right)$$

Similarly,

$$\mu_0''(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) = \frac{K}{2} \left(\frac{(\rho^* - \hat{\rho}^*) C_n}{T} + \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\hat{\rho}^* + t + T} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{jT + \hat{\rho}^* + t} + \frac{1}{2} \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\hat{\rho}^* + t - T} \right)$$

$$\mu_1''(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) = \frac{K}{2} \left(\frac{(\rho^* - \hat{\rho}^*) C_n}{T} + \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\rho^* + t} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{\rho^* - \rho^* - 1} \mathbb{C}_{jT + \rho^* + t} + \frac{1}{2} \sum_{t=0}^{\rho^* - \rho^* - 1} \mathbb{C}_{\rho^* + t - T} \right)$$

$$\mu_0''(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) = \frac{K}{2} \left(\frac{(\rho^* - \hat{\rho}^*) C_n}{T} + \sum_{t=0}^{\rho^* - \hat{\rho}^* - 1} \mathbb{C}_{\rho^* + t + T} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{\rho^* - \rho^* - 1} \mathbb{C}_{jT + \rho^* + t} + \frac{1}{2} \sum_{t=0}^{\rho^* - \rho^* - 1} \mathbb{C}_{\rho^* + t - T} \right)$$

Then we can compute the

$$\Pr(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*)} - \sqrt{\mu_0''(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*)})^2} = \Pr_U(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) \quad (5.43)$$

$$\Pr(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) \leq e^{-(\sqrt{\mu_1''(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*)} - \sqrt{\mu_0''(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*)})^2} = \Pr_U(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) \quad (5.44)$$

5.4. The Proposed Synchronization Schemes

As we obtain the upper bounds $\Pr_U(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*)$ and $\Pr_U(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*)$, we can compute the upper bound ϵ_U for the absolute mean synchronization error $E_s \|\hat{\rho}_s^* - \rho^*\|$ as follows:

$$\epsilon_U = \max(E_U \|\hat{\rho}^* - \rho^*\|_{\hat{\rho}^* < \rho^*}, E_U \|\hat{\rho}^* - \rho^*\|_{\hat{\rho}^* > \rho^*}) \quad (5.45)$$

where

$$E_U \|\hat{\rho}^* - \rho^*\|_{\hat{\rho}^* \leq \rho^*} = \sum_{\hat{\rho}^* = \rho_{\min}}^{\rho^*} \|\hat{\rho}^* - \rho^*\| \Pr_U(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) \quad (5.46)$$

$$E_U \|\hat{\rho}^* - \rho^*\|_{\hat{\rho}^* \geq \rho^*} = \sum_{\hat{\rho}^* = \rho^*}^{\rho_{\max}} \|\hat{\rho}^* - \rho^*\| \Pr_U(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) \quad (5.47)$$

In addition, the corresponding upper bound σ_U^2 for the variance of the synchronization error $E_s |\hat{\rho}_s^* - \rho^*|^2$ is computed as follows:

$$\sigma_U^2 = \sum_{\hat{\rho}^* = \rho_{\min}}^{\rho^*} \|\hat{\rho}^* - \rho^*\|^2 \Pr_U(\hat{\rho}^* | \hat{\rho}^* \leq \rho^*) + \sum_{\hat{\rho}^* = \rho^*}^{\rho_{\max}} \|\hat{\rho}^* - \rho^*\|^2 \Pr_U(\hat{\rho}^* | \hat{\rho}^* \geq \rho^*) \quad (5.48)$$

5.4.2 K-Means-based Blind Synchronization Algorithm

The aforementioned scheme relies on the preamble which may not be known at the receiver side. In order to fulfil the synchronization without preamble, i.e., the blind synchronization, we focus our attention on the metric function (5.26). Since what the (5.26) does is partitioning the received signals $r_m(\rho, s_{m+1}, \dots, s_{m-L})$ in to two groups with the aid of the s_m for $0 \leq m \leq K-1$ and compute the difference of the estimates of $\bar{r}_i(\rho, s_i = 1)$ and $\bar{r}_i(\rho, s_i = 0)$, i.e., $\hat{r}_i(\rho, s_i = 1)$ and $\hat{r}_i(\rho, s_i = 0)$. In practice, even though the s_m for $1 \leq m \leq K$ are unknown, the two groups of signals in $r_m(\rho, s_{m+1}, \dots, s_{m-L})$ can still be separated which can be done by the clustering [84]. As $\bar{r}_i(\rho, s_i = 1) > \bar{r}_i(\rho, s_i = 0)$, $\hat{r}_i(\rho, s_i = 1) > \hat{r}_i(\rho, s_i = 0)$ holds true. By applying the simplest clustering method, K-means clustering to the received signals, we can partition the received signals $r_m(\rho, s_{m+1}, \dots, s_{m-L})$ into two groups whose mean values are regarded as $\hat{r}_i(\rho, s_i = 1)$ and $\hat{r}_i(\rho, s_i = 0)$. The metric function using K-means clustering is then

$$G(\rho) = \hat{r}_i(\rho, s_i = 1) - \hat{r}_i(\rho, s_i = 0) \quad (5.49)$$

The comparison of $E[G(\rho)]$ and $C_0(\rho)$ is depicted in Fig. 5.6.

The corresponding start point of each transmission is obtained by:

$$\tilde{\rho}^* = \max_{\rho} G(\rho). \quad (5.50)$$

Since the $E[G(\rho)]$ is in accord with $C_0(\rho)$ around the ρ^* , the estimated optimal start point $\tilde{\rho}^*$ is close to ρ^* .

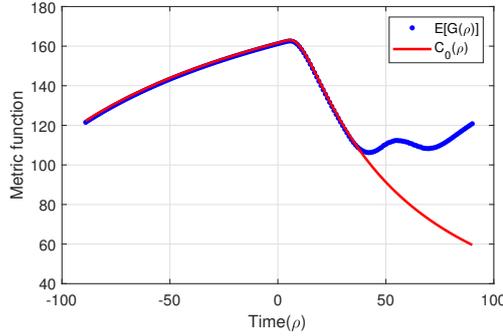


Figure 5.6: Comparison of $E[G(\rho)]$ and $C_0(\rho)$ when SNR is 30 dB

5.5 The Cramer-Rao Lower Bounds with known and unknown CSI

Before the link between the transmitter and the receiver is established, the CSI is unknown to the receiver. The unknown CSI becomes unwanted parameters to the whole system which may degrade the synchronization performance compared with the case where the CSI is known. In this section, we investigate the Cramer-Rao lower bounds for the synchronization with unknown and known CSI.

Let us denote $\bar{r}_i(t, \rho, s_{i+1}, s_i, \dots, s_{i-L})$ as $\bar{r}_i(t, \rho)$ for simplicity. At each time instant, the probability of receiving $r_i(t, \rho)$ information particles is expressed as follows:

$$\Pr(r_i(t, \rho) | \bar{r}_i(t, \rho)) = \frac{e^{-\bar{r}_i(t, \rho)} [\bar{r}_i(t, \rho)]^{r_i(t, \rho)}}{r_i(t, \rho)!} \quad (5.51)$$

Let us denote $\mathbb{C} = [\mathbb{C}_0, \dots, \mathbb{C}_{T-1}, \dots, \mathbb{C}_{LT-1}, \mathbb{C}_n]$, $\mathbf{r} = [r_0(0, \rho), \dots, r_0(T-1, \rho), \dots, r_K(0, \rho), \dots, r_K(T-1, \rho)]$ and $\bar{\mathbf{r}}(\rho) = [\bar{r}_0(0, \rho), \dots, \bar{r}_0(T-1, \rho), \dots, \bar{r}_K(T-1, \rho)]$. The likelihood function of the receiving signals \mathbf{r} conditioned upon the transmitted symbols \mathbf{s} is then:

$$\Pr(\mathbf{r} | \rho, \mathbb{C}, \mathbf{s}) = \prod_{i=L}^{K-1} \prod_{t=0}^{T-1} \frac{e^{-\bar{r}_i(t, \rho)} [\bar{r}_i(t, \rho)]^{r_i(t, \rho)}}{r_i(t, \rho)!} \quad (5.52)$$

The natural logarithm likelihood function is accordingly:

$$\begin{aligned} \ln(\Pr(\mathbf{r} | \rho, \mathbb{C}, \mathbf{s})) &= \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} \ln[\Pr(r_i(t, \rho) | \rho, \mathbb{C}, s_i, \dots, s_{i-L})] \\ &= \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\bar{r}_i(t, \rho) + r_i(t, \rho) \ln[\bar{r}_i(t, \rho)] - \ln[r_i(t, \rho)!] \end{aligned} \quad (5.53)$$

5.5. The Cramer-Rao Lower Bounds with known and unknown CSI

The first order derivatives of the natural logarithm likelihood function are

$$\frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} \quad (5.54)$$

$$\frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_{kT+l}} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_{kT+l}} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_{kT+l}} \quad (5.55)$$

$$\frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_n} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_n} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_n} \quad (5.56)$$

Since $\bar{r}_i(t, \rho) = \sum_{j=-1}^L s_{i-j} \mathbf{C}_{jT+\rho+t} + \mathbf{C}_n$,

$$\frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_{kT+l}} = \delta(\rho + t - l) s_{i-k} \quad (5.57)$$

where $\delta(t)$ denotes the Dirac delta function. Meanwhile,

$$\frac{\partial \bar{r}_i(t, \rho)}{\partial \mathbf{C}_n} = 1 \quad (5.58)$$

and we define

$$\frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} = \bar{r}_i(t, \rho + 1) - \bar{r}_i(t, \rho) \quad (5.59)$$

As the conditions $E_{\mathbf{r}} \frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho} = 0$, $E_{\mathbf{r}} \frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_{kT+l}} = 0$ and $E_{\mathbf{r}} \frac{\partial \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_n} = 0$ are satisfied, the Fisher information matrix needs to be computed by the second order derivatives:

$$E_{\mathbf{r}} \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho^2} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \left[\frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} \right]^2 \quad (5.60)$$

$$E_{\mathbf{r}} \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_{kT+l}^2} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \delta(t - l) s_{i-k} \quad (5.61)$$

$$E_{\mathbf{r}} \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \mathbf{C}_n^2} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \quad (5.62)$$

Chapter 5. Synchronization for Constant-Threshold-based Receivers in Diffusive Molecular Communication Systems

$$\begin{aligned}
 E_r \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho \partial C_{kT+l}} &= E_r \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial^2 \bar{r}_i(t, \rho)}{\partial \rho \partial C_{kT+l}} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial^2 \bar{r}_i(t, \rho)}{\partial \rho \partial C_{kT+l}} - \frac{r_i(t, \rho)}{\bar{r}_i^2(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial C_{kT+l}} \frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} \\
 &= \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \delta(t-l) s_{i-k} \frac{\partial \bar{r}_i(t, \rho)}{\partial \rho}
 \end{aligned} \tag{5.63}$$

$$\begin{aligned}
 E_r \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho \partial C_n} &= E_r \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial^2 \bar{r}_i(t, \rho)}{\partial \rho \partial C_n} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial^2 \bar{r}_i(t, \rho)}{\partial \rho \partial C_n} - \frac{r_i(t, \rho)}{\bar{r}_i^2(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial C_n} \frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} \\
 &= \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial \rho}
 \end{aligned} \tag{5.64}$$

$$\begin{aligned}
 E_r \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_{kT+l} \partial C_n} &= E_r \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{\partial^2 \bar{r}_i(t, \rho)}{\partial C_{kT+l} \partial C_n} + \frac{r_i(t, \rho)}{\bar{r}_i(t, \rho)} \frac{\partial^2 \bar{r}_i(t, \rho)}{\partial C_{kT+l} \partial C_n} - \frac{r_i(t, \rho)}{\bar{r}_i^2(t, \rho)} \frac{\partial \bar{r}_i(t, \rho)}{\partial C_n} \frac{\partial \bar{r}_i(t, \rho)}{\partial C_{kT+l}} \\
 &= \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} -\frac{1}{\bar{r}_i(t, \rho)} \delta(t-l) s_{i-k}
 \end{aligned} \tag{5.65}$$

The Fisher matrix can be expressed as:

$$\begin{aligned}
 \mathbf{J} &= -E_r \begin{pmatrix} \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_0 \partial C_0} & \dots & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_0 \partial C_{LT-l}} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_0 \partial C_n} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_0 \partial \rho} \\ \vdots & \dots & \dots & \ddots & \vdots \\ \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_{LT-l} \partial C_0} & \dots & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_{LT-l} \partial C_{LT-l}} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_{LT-l} \partial C_n} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_{LT-l} \partial \rho} \\ \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_n \partial C_0} & \dots & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_n \partial C_{LT-l}} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_n \partial C_n} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial C_n \partial \rho} \\ \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho \partial C_0} & \dots & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho \partial C_{LT-l}} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho \partial C_n} & \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbf{C}, \mathbf{s}))}{\partial \rho^2} \end{pmatrix} \\
 &= \mathbf{P}^T \mathbf{\Lambda}^T \mathbf{\Lambda} \mathbf{P}
 \end{aligned} \tag{5.66}$$

$$\mathbf{\Lambda} = \text{diag}\left(\left[\frac{1}{\sqrt{\bar{r}_L(0, \rho)}}, \dots, \frac{1}{\sqrt{\bar{r}_L(T-1, \rho)}}, \dots, \frac{1}{\sqrt{\bar{r}_{K-1}(T-1, \rho)}}\right]\right) \tag{5.67}$$

$$\mathbf{P} = \begin{pmatrix} \mathbf{U} & \boldsymbol{\psi} & \boldsymbol{\eta} \end{pmatrix} \quad (5.68)$$

$$\mathbf{U} = \begin{pmatrix} s_L \mathbf{I} & s_{L-1} \mathbf{I} & \cdots & s_0 \mathbf{I} \\ s_{L+1} \mathbf{I} & s_L \mathbf{I} & \cdots & s_1 \mathbf{I} \\ \vdots & \cdots & \cdots & \vdots \\ s_{K-1} \mathbf{I} & s_{K-2} \mathbf{I} & \cdots & s_{K-L-1} \mathbf{I} \end{pmatrix} \quad (5.69)$$

$$\boldsymbol{\eta} = \left[\frac{\partial \bar{r}_L(0, \rho)}{\partial \rho}, \dots, \frac{\partial \bar{r}_L(T-1, \rho)}{\partial \rho}, \dots, \frac{\partial \bar{r}_{K-1}(T-1, \rho)}{\partial \rho} \right]^T \quad (5.70)$$

where \mathbf{I} is an identical matrix of size $T \times T$, $\boldsymbol{\psi}$ denotes a column vector of size $T(K-L)$ whose element is 1 and $\text{diag}(\cdot)$ converts a vector into a diagonal matrix.

To compute the $(LT+2, LT+2)$ th entry in the inverse of the matrix \mathbf{J} , we resort to the rank-1 update of inverse of inner product in the matrix theory, i.e.,

$$(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} = \begin{pmatrix} \mathbf{A} + \frac{\mathbf{A}\mathbf{X}^T \mathbf{v}\mathbf{v}^T \mathbf{X}\mathbf{A}^T}{\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{X}\mathbf{A}\mathbf{X}^T \mathbf{v}} & \frac{-\mathbf{A}\mathbf{X}^T \mathbf{v}}{\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{X}\mathbf{A}\mathbf{X}^T \mathbf{v}} \\ \frac{-\mathbf{v}^T \mathbf{X}\mathbf{A}^T}{\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{X}\mathbf{A}\mathbf{X}^T \mathbf{v}} & \frac{1}{\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{X}\mathbf{A}\mathbf{X}^T \mathbf{v}} \end{pmatrix}$$

if $\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1}$ and $\tilde{\mathbf{X}} = [\mathbf{X}\mathbf{v}]$.

In our case,

$$\mathbf{A} = \left(\begin{pmatrix} \mathbf{U} & \boldsymbol{\psi} \end{pmatrix}^T \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} \begin{pmatrix} \mathbf{U} & \boldsymbol{\psi} \end{pmatrix} \right)^{-1} \quad (5.71)$$

and the Cramer-Rao lower bound for estimating $\hat{\rho}_s^*$ without CSI given the known transmitted symbols \mathbf{s} is

$$\xi_s(\rho) = \{ \mathbf{J}^{-1} \}_{LT+2, LT+2} = \frac{1}{\boldsymbol{\eta}^T \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} \boldsymbol{\eta} - \boldsymbol{\eta}^T \boldsymbol{\Lambda} \begin{pmatrix} \mathbf{U} & \boldsymbol{\psi} \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{U} & \boldsymbol{\psi} \end{pmatrix}^T \boldsymbol{\Lambda}^T \boldsymbol{\eta}} \quad (5.72)$$

where $\{ \mathbf{J}^{-1} \}_{LT+2, LT+2}$ denotes the $(LT+2, LT+2)$ th entry in the matrix \mathbf{J}^{-1} . And the Cramer-Rao lower bound without CSI is computed as follows:

$$\xi(\rho) = \frac{\sum_{s \in \Theta} \xi_s(\rho)}{K_s} \quad (5.73)$$

If we know the complete CSI, then the corresponding Fisher information given the \mathbf{s} is

$$J' = -E_{\mathbf{r}} \left\{ \frac{\partial^2 \ln(\Pr(\mathbf{r}|\rho, \mathbb{C}, \mathbf{s}))}{\partial \rho^2} \right\} = \sum_{i=L}^{K-1} \sum_{t=0}^{T-1} \frac{1}{\bar{r}_i(t, \rho)} \left[\frac{\partial \bar{r}_i(t, \rho)}{\partial \rho} \right]^2 = \boldsymbol{\eta}^T \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} \boldsymbol{\eta} \quad (5.74)$$

i.e., the Cramer-Rao lower bound with known CSI given the \mathbf{s} is

$$\xi'_s(\rho) = \frac{1}{\boldsymbol{\eta}^T \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} \boldsymbol{\eta}} \quad (5.75)$$

then the Cramer-Rao lower bound with known CSI is computed as follows:

$$\xi'(\rho) = \frac{1}{K_s} \sum_{\mathbf{s} \in \Theta} \xi'_s(\rho) \quad (5.76)$$

From (5.72) and (5.75), we know that $\xi_s(\rho) > \xi'_s(\rho)$, thus $\xi(\rho) > \xi'(\rho)$.

5.6 Simulation Results

In this section, we report and describe some simulation results in order to validate the effectiveness of the proposed synchronization schemes. The system parameters of the typical MC system in this chapter are listed in Table. 5.1.

Table 5.1: Simulation parameters

Parameter	Value
λ_0	$100s^{-1}$
Receiver radius r	45 nm
Distance d	500 nm
Diffusion coefficient D	$4.265 \times 10^{-10} m^2/s$
Discrete time length ρ	4.5 us
Slot length	180ρ
Channel length L	7

We assume that the receiver can record the received signals for a long sequence to synchronize signals. In this chapter, we set the SNR range as [20 dB, 40 dB]. In order to evaluate the synchronization performance for all possible \mathbf{s} , we generate the set κ of length $K_s = 1000$ containing the randomly generated \mathbf{s} whose entry is either 1 or 0 with equiprobability. At each SNR value and for each \mathbf{s} in the set κ , 100 trials are simulated to obtain the estimated optimal start point. As the optimal start point ρ^* is constant within the SNR range [20 dB, 40 dB] which is obtained by optimizing $P_e(\rho, \tau)$ and is the same as the one obtained by Lemma.5.2. Thus, in the following simulations, we just compare the estimated optimal start point with this constant ρ^* .

In Fig.5.7, we compare the absolute mean synchronization errors of the proposed data-aided

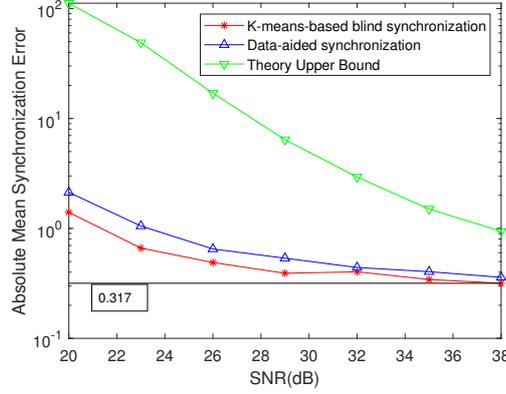


Figure 5.7: The absolute mean synchronization error comparison $|E_s(\hat{\rho}_s^* - \rho^*)|$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 64$).

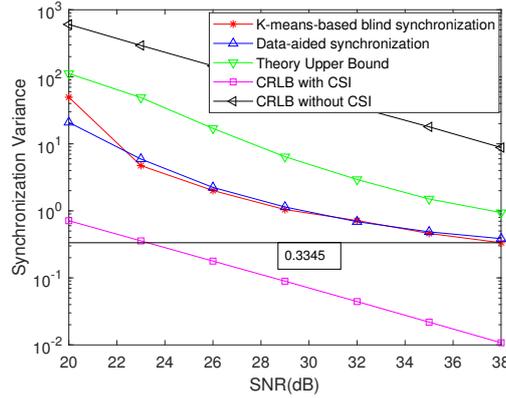


Figure 5.8: The synchronization variance comparison $E|\hat{\rho}_s^* - \rho^*|^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 64$).

synchronization algorithm and the K-means-based blind synchronization algorithm with the theoretical upper bound ϵ_U . We observe that the absolute mean of the synchronization errors of the proposed schemes both decrease to the same low level and are lower than the upper bound derived in (5.45). In Fig.5.8, we observe that the synchronization variances both reduce gradually to a low level which is lower than the corresponding theoretical upper bound and is greater than the CRLB with full CSI. We observe that using partial channel knowledge as in the Lemma. 5.1 does improve the synchronization performance of proposed schemes drastically as the synchronization variances of proposed schemes are smaller than CRLB without CSI.

In Fig.5.9, Fig.5.10, Fig.5.11 and Fig.5.12, we compare the absolute mean of the synchronization error and synchronization variance of the proposed algorithms with the theoretical results.

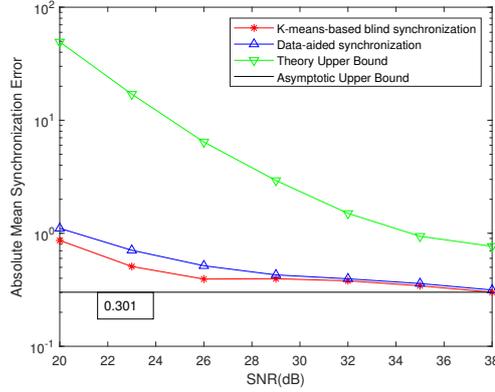


Figure 5.9: The absolute mean synchronization error comparison $|E_s(\hat{\rho}_s^* - \rho^*)|$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 128$).

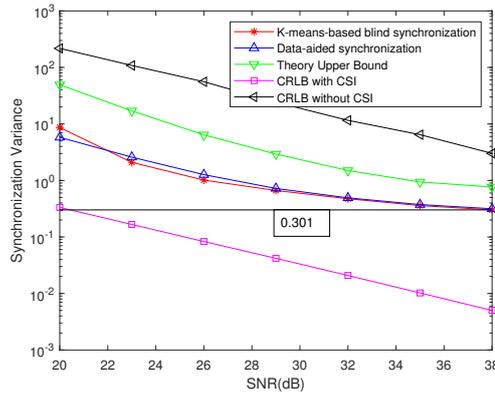


Figure 5.10: The synchronization variance comparison $E|\hat{\rho}_s^* - \rho^*|^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 128$).

We observe that as the number of symbols increase, the synchronization performances of the proposed algorithm are improved. In addition, the theoretical upper bounds are tighter. Though CRLB without CSI is large, it is greater than the theoretical upper bound. In general, the proposed algorithms achieve superior synchronization performance.

5.7 Conclusions

In this chapter, we investigate the synchronization without prior CSI for the MC systems that employ the constant-threshold-based detectors. Based on the formulated error probability associated with the threshold and the start point, the condition that the asymptotic optimal start point follows is presented. Inspired by this condition, a data-aided synchronization

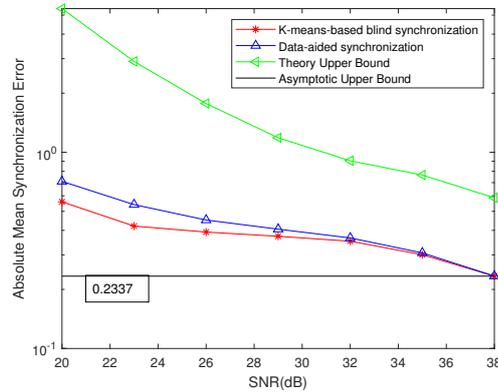


Figure 5.11: The absolute mean synchronization error comparison $|E_s(\hat{\rho}_s^* - \rho^*)|$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the upper bound ϵ_U ($T = 180$ and $K = 256$).

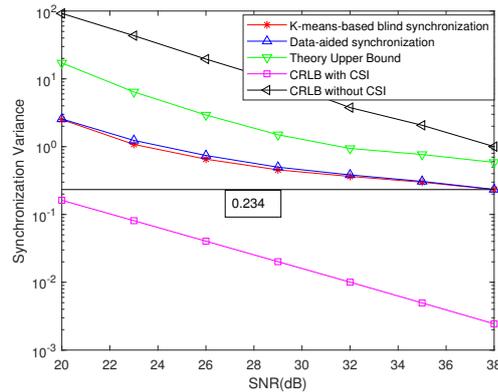


Figure 5.12: The synchronization variance comparison $E|\hat{\rho}_s^* - \rho^*|^2$ of the proposed data-aided synchronization algorithm and K-means-based blind synchronization algorithm compared with the theoretical upper bound σ_U^2 and the CRLB with/without CSI ($T = 180$ and $K = 256$).

metric function is proposed. As the corresponding metric results follows the Skellam distribution, the upper bounds for the absolute mean and variance of the synchronization errors are derived. Motivated by the data-aided metric function, a blind synchronization algorithm is proposed based on the intermediate variables in the metric function that can be obtained by applying the K-means clustering to the received signals. In order to quantify the performance, the CRLBs with and without CSI are derived. Simulation results validate the effectiveness of proposed schemes.

5.8 Appendix

Rewrite $F(\rho)$ as

$$\begin{aligned} F(\rho) &= \sum_{m=0}^{K-1} r_m(\rho, s_{m+1}, \dots, s_{m-L})(2s_m - 1) \\ &= \sum_{m=0}^{K-1} \sum_{t=0}^{T-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L})(2s_m - 1) \end{aligned} \quad (5.77)$$

Thus, for $0 < k < T$

$$\begin{aligned} F(\rho) - F(\rho + k) &= 2 \sum_{m=0}^{K-1} s_m \sum_{t=0}^{T-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L}) - r_m(t, \rho + k, s_{m+1}, \dots, s_{m-L}) \\ &= 2 \sum_{m=0}^{K-1} s_m \sum_{t=0}^{T-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L}) - r_m(t + k, \rho, s_{m+1}, \dots, s_{m-L}) \\ &= 2 \sum_{m=0}^{K-1} s_m \left(\sum_{t=0}^{k-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L}) - \sum_{t=T}^{T-1+k} r_m(t, \rho, s_{m+1}, \dots, s_{m-L}) \right) \\ &= 2 \sum_{m=0}^{K-1} s_m \left(\sum_{t=0}^{k-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L}) - \sum_{t=0}^{k-1} r_{m+1}(t, \rho, s_{m+2}, \dots, s_{m-L+1}) \right) \\ &= 2 \sum_{t=0}^{k-1} \sum_{m=0}^{K-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L})(s_m - s_{m-1}) \end{aligned} \quad (5.78)$$

$F(\rho) - F(\rho + k) \sim \text{Skellam}(\mu'_1(k, \rho), \mu'_0(k, \rho))$, since $r_m(t, \rho, s_{m+1}, \dots, s_{m-L})$ and $r_{m+1}(t, \rho, s_{m+2}, \dots, s_{m-L+1})$ are Poisson random variables. According to (5.78), note that when $s_m - s_{m-1}$ is not zero, $r_m(t, \rho, s_{m+1}, \dots, s_{m-L})$ may contribute to $F(\rho) - F(\rho + k)$. If $s_m - s_{m-1} > 0$, $r_m(t, \rho, s_{m+1}, \dots, s_{m-L})$ will contribute to $\mu'_1(k, \rho)$. If $s_m - s_{m-1} < 0$, $r_m(t, \rho, s_{m+1}, \dots, s_{m-L})$ will contribute to $\mu'_0(k, \rho)$. From (5.3), we have

$$\begin{aligned} \mu'_1(k, \rho) &= E_r \left\{ 2 \sum_{t=0}^{k-1} \sum_{m=0}^{K-1} r_m(t, \rho, s_{m+1}, \dots, s_{m-L})(s_m - s_{m-1}) \right\}_{s_m - s_{m-1} > 0} \\ &= \left\{ 2 \sum_{t=0}^{k-1} \sum_{m=0}^{K-1} \bar{r}_m(t, \rho, s_{m+1}, \dots, s_{m-L})(s_m - s_{m-1}) \right\}_{s_m - s_{m-1} > 0} \\ &= \left\{ 2 \sum_{t=0}^{k-1} \sum_{m=0}^{K-1} \left(\sum_{j=-1}^L s_{m-j} \mathbb{C}_{jT+\rho+t} + \mathbb{C}_n \right) (s_m - s_{m-1}) \right\}_{s_m - s_{m-1} > 0} \\ &= \left\{ 2 \mathbb{C}_n \sum_{t=0}^{k-1} \sum_{m=0}^{K-1} s_m - s_{m-1} \right\}_{s_m - s_{m-1} > 0} + \end{aligned}$$

$$\begin{aligned}
 & \left\{ \sum_{j=-1}^L 2 \sum_{t=0}^{k-1} \mathbb{C}_{jT+\rho+t} \sum_{m=0}^{K-1} s_{m-j}(s_m - s_{m-1}) \right\}_{s_m - s_{m-1} > 0} \\
 &= 2 \frac{kC_n}{T} \sum_{m=0}^{K-1} (s_m - s_{m-1})^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{k-1} \mathbb{C}_{jT+\rho+t} \sum_{m=0}^{K-1} s_{m-j}(s_m - s_{m-1})^+
 \end{aligned}$$

Thus, considering the transmitted symbols \mathbf{s} , $\mu'_1(k, \rho)$ can be calculated via (5.30). Similarly,

$$\mu'_0(k, \rho) = 2 \frac{kC_n}{T} \sum_{m=0}^{K-1} (s_{m-1} - s_m)^+ + 2 \sum_{j=-1}^L \sum_{t=0}^{k-1} \mathbb{C}_{jT+\rho+t} \sum_{m=0}^{K-1} s_{m-j}(s_{m-1} - s_m)^+$$

On average, we know that

$$E_{\mathbf{s}} \left\{ \sum_{m=0}^{K-1} (s_m - s_{m-1})^+ \right\} = \frac{K}{4} \quad (5.79)$$

$$E_{\mathbf{s}} \left\{ \sum_{m=0}^{K-1} s_m (s_m - s_{m-1})^+ \right\} = \frac{K}{4} \quad (5.80)$$

$$E_{\mathbf{s}} \left\{ \sum_{m=0}^{K-1} s_{m-1} (s_m - s_{m-1})^+ \right\} = 0 \quad (5.81)$$

$$E_{\mathbf{s}} \left\{ \sum_{m=0}^{K-1} s_{m-j} (s_m - s_{m-1})^+ \right\} = \frac{K}{8}; j \neq 0, 1 \quad (5.82)$$

Thus, we have

$$E_{\mathbf{s}}[\mu'_1(k, \rho)] = \frac{K}{2} \left(\frac{kC_n}{T} + \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+jT} + \frac{1}{2} \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t-T} \right)$$

Similarly,

$$E_{\mathbf{s}}[\mu'_0(k, \rho)] = \frac{K}{2} \left(\frac{kC_n}{T} + \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+T} + \frac{1}{2} \sum_{j=2}^L \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t+jT} + \frac{1}{2} \sum_{t=0}^{k-1} \mathbb{C}_{\rho+t-T} \right)$$

6 Conclusions and Future Work

This chapter closes the thesis with general conclusions in *Section 6.1* and summaries of future work in *Section 6.2*.

6.1	Conclusion	122
6.2	Future Work	123
6.2.1	Machine-Learning-Aided Detection	123
6.2.2	Non-Coherent Detection	124
6.2.3	Synchronization	125

6.1 Conclusion

In this thesis, new contributions on evaluating and designing new detectors for the MC systems have been reported based on the insights acquired from the proposed NN-based detectors. Besides, the non-coherent detectors based on the K-Means clustering are introduced as a promising enabler for the integrated MC systems in reality. In addition, synchronization without channel information has been investigated and the synchronization schemes have been introduced and inspected. More specifically, the contribution of this thesis can be summarized as follows.

As a first step, the performance evaluation is ultra essential as it provides the theoretical guarantee on comparing the different detector schemes and identifying the best detector. Meanwhile, the threshold-based detectors are of great importance as this kind of detectors are easy to implement considering the limited computation capacity of the nano-machines. Hence, in chapter 3, we provide the BER performance evaluation frameworks for threshold-based detectors. Beside, inspired by the recent application of deep learning to the MC systems, e.g., the feed-forward NN, we find some questions are not answered: 1) is the NN-based scheme the same as the threshold-based detectors? 2) If not, what can we learn from the NN-based scheme? By implementing the NN-based scheme using the same input and output as the threshold-based detectors, i.e., the received signal and the previously detected symbols (memory bits) as the input and the current bit as the output, surprisingly, better BER performances are achieved. The conventional threshold-based detectors are not optimal, which motivates us to revisit the threshold-based detectors. Based on the analysis on the optimal constant-threshold-detector we have proposed, the optimal memory-bit-assisted threshold-based detectors are devised and they have the consistent performances as the NN-based detectors.

In realistic MC systems, detection without prior channel information is especially essential as the channel may vary from time to time and the CE may be not accurate enough for the following detection. To solve this issue and to take the low BER advantage brought by the multi-memory-bit-aided thresholds, we work on the detection using the multi-memory-bit-aided threshold by retrieving the channel information in the threshold only from the received signals. By reformulating the thresholds with some intermediate parameters, i.e., the average received signal corresponding to the \mathcal{L} transmitted bits, we evince that estimating the intermediate parameters has been the key point of the non-coherent detection. And this is done by assigning the received signal with a label (the \mathcal{L} transmitted bits) and averaging the signals with the same label. From the perspective of ML, assigning data to their corresponding labels without any assistance information is a clustering problem and should be performed via the clustering algorithms (K-means clustering algorithm is adopted in the thesis). To reduce the clustering errors, we construct the multi-dimensional data using the received signals. In addition, two methods for initializing the centroids of the clusters directly from the empirical

data are proposed. By applying the clustering algorithm to the multi-dimensional data, we can obtain the clustered results and extract information used in the multi-memory-bit-aided thresholds, yielding the non-coherent detectors. Simulation results show that, in the presence of mild ISI, the proposed algorithms can achieve satisfactory BER performance. Even under severe ISI, the non-coherent detector using the iterative centroid scheme can still work quite well.

In light of the realistic MC system that has no prior channel information, the MC system still faces a challenging issue as most synchronization schemes assume known CSI or known channel model. Thus, in chapter 5, we consider the synchronization without CSI via incorporating the start point into the error probability that uses the general channel information such that the optimization of this error probability is not related to a specific channel. According to the asymptotic analysis in the high SNR regime, we conclude that the optimal start point yields the largest channel gain for current transmitted bit. By imitating this condition, a synchronization metric function is proposed using the received signals with the aid of the known transmitted bits. Since the received signals follow the Poisson distribution, the metric results follow the Skellam distribution, providing us a way to derive the upper bounds for the absolute mean and variance of the synchronization errors. Apart from the data-aided metric function, inspired by the K-means clustering, we apply the K-mean clustering to the received signals to compute the parameters used in the data-aided metric function, yielding a blind synchronization algorithm. To quantify the synchronization performance, we also derive the Cramer-Rao Lower bounds for the cases of known CSI and unknown CSI. Simulation results and the theoretical upper bounds verify the effectiveness of the proposed methods.

6.2 Future Work

Many interesting topics in MC, especially the fundamental works, remain open and are motivating researchers to pursuit realizing the medical applications, e.g., drug delivery. In the continuity of our results, we will discuss some possible future research directions in the following.

6.2.1 Machine-Learning-Aided Detection

As discussed in the chapter 3, we proposed some feed-forward NN-based detectors. This kind of NN does not provide the interpretable and intuitive understanding. Specifically, we do not know the structure of a perfect NN-based detector, i.e., the number of layers and the number of neurons in each individual layer are unknown. In the domain of deep learning, there exists another type of NN that is capable of providing the desired interpretation as this type of NN is based on the Bayesian theory, which is known as probabilistic graphical models [92]. In practice, the threshold-based detector are derived based on the Bayesian theory. In this

context, we can map the threshold-based detector in the form of probabilistic graphical model and by training this kind of NN, some essential parameters, e.g., the CSI, can be estimated accordingly. Thus, applying the probabilistic graphical model may have twofold effects.

Currently, we only take the static channel into consideration and the channel parameters in the thresholds are constants, while in the mobile MC systems [93], the channel parameters changes with time. By inserting the time-dependent channel parameters in the original thresholds may fulfil the detection task. However, it's not clear that if the direct substitution of variable channel parameters yields the optimal performance. Meanwhile, the time-dependent channel parameters are not easy to obtain either because a long sequence of known data are needed for the channel estimation or the mathematical channel model is unclear. Proper solution to this difficulty may be the RNN [27] that can handle time-series problems. Although some RNN-based detectors have been proposed, the performance and proper structure design are not well investigated. The link between the analytical threshold-based detectors and the RNN-based detectors must be built, and in turn, the RNN-based detectors will provide more insights for the design of the analytical detectors.

6.2.2 Non-Coherent Detection

In chapter 4, we proposed two algorithms for initializing the centroids for clustering the multi-dimensional data constructed using the received signals. The first one is simple but may lead to clustering errors. Though the second one has better cluster performance, the computational complexity is a heavy burden to the MC system. In MC systems, the coding techniques [94] are essential as it may simplify the detection, channel estimation schemes and improve the transmission performance. Our direct intuition is to design a codebook such that, when using our non-coherent detectors, there are both less clustering errors and in turn less detection errors by using the initial centroids constructed by $\max(\mathbf{r})$.

As pointed in the previous subsection, the mobile MC systems are not well investigated, not to mention the non-coherent detection in mobile MC systems. Practically, as time goes, the shapes of the clusters and the positions of the corresponding centroids as shown in the Fig. 4.3 will change accordingly. If we treat the centroids of the clusters as the state of the MC systems, the state changes as time goes. As long as we can figure out how the state evolves, we can predict its future state by applying some algorithm in the control theory [95] to the state estimation. Based on the learned evolution pattern, the parameters used in the threshold may be estimated for future time such that the non-coherent detection can still be performed in the mobile MC systems.

Except from detection in the mobile MC systems, the modulation scheme information can

also be inferred from the received signals, still based on the constructed multi-dimensional data. For example, if we exploit the binary CSK, there will be four clusters as in Fig. 4.3. If four-amplitude CSK is employed, there will be sixteen clusters. If the PPM is used, the shape of the clusters will be different from the one in Fig. 4.3. This constitutes an interesting topic for future research.

6.2.3 Synchronization

In chapter 5, we consider the synchronization for the MC systems exploiting the constant-threshold-based receiver. Therefore, one of the promising directions may be the investigation of the synchronization for memory-bit-assisted threshold-based receivers. We may figure out if the optimal start point for the constant-threshold receiver is the same for the receiver using the threshold with memory bits. Of course, more transmitted bits need to be known in advance. Thus, one possible direction may be designing the synchronization metric function while reducing the preamble length.

Meanwhile, we can extend the analysis framework in chapter 5 to the MC system using the MoSK or PPM, still associating the start point with the error probability, deriving the (asymptotic) optimal start point and devising the synchronization metric function. The best case may be that some useful results in the chapter 5 can be exploited directly.

Concerning the realistic MC systems, it takes time and energy to produce the molecules to be release. The number of released molecule may not be a constant or the release time may have a delay. In the received signals, each symbol has its individual start point. Until now, some CSI-dependent synchronization schemes have been proposed based on the maximum likelihood metric. However, the CSI is almost impossible to estimate with satisfactory precision if the release time is unstable. Thus, generalizing our results to grasp the overall start point of the transmission and based on received signals in each slot, the one-symbol synchronization may be performed. It's more realistic and will be an interesting research topic for future research.

Bibliography

- [1] Xuewen Qian, Marco Di Renzo, and Andrew Eckford. Molecular communications: Model-based and data-driven receiver design and optimization. *IEEE Access*, 7:53555–53565, 2019.
- [2] Yg Li and GJ Stuber. *Orthogonal frequency division multiplexing for wireless communications*. 2006.
- [3] Chen Gong and Zhengyuan Xu. Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference. *IEEE Transactions on Wireless Communications*, 14(10):5326–5337, 2015.
- [4] Josep Miquel Jornet and Ian F Akyildiz. Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band. *IEEE Transactions on Wireless Communications*, 10(10):3211–3221, 2011.
- [5] Sarah Hussein, Youmni Ziade, and Raed M Shubair. On propagation losses due to in-vivo electromagnetic nanoscale communication. In *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, pages 1309–1310. IEEE, 2019.
- [6] Weisi Guo, Christos Mias, Nariman Farsad, and Jiang-Lun Wu. Molecular versus electromagnetic wave propagation loss in macro-scale environments. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 1(1):18–25, 2015.
- [7] George M Whitesides. The once and future nanomachine. *Scientific American*, 285(3):78–83, 2001.
- [8] Uche AK Chude-Okonkwo, Reza Malekian, Bodhaswar T Maharaj, and Athanasios V Vasilakos. Molecular communication and nanonetwork for targeted drug delivery: A survey. *IEEE Communications Surveys & Tutorials*, 19(4):3046–3096, 2017.
- [9] Nariman Farsad, H Birkan Yilmaz, Andrew Eckford, Chan-Byoung Chae, and Weisi Guo. A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys & Tutorials*, 18(3):1887–1919, 2016.

Bibliography

- [10] Tatsuya Suda, Michael Moore, Tadashi Nakano, Ryota Egashira, Akihiro Enomoto, Satoshi Hiyama, and Yuki Moritani. Exploratory research on molecular communication between nanomachines. In *Genetic and Evolutionary Computation Conference (GECCO), Late Breaking Papers*, volume 25, page 29, 2005.
- [11] Murat Kuscü, Ergin Dinc, Bilgesu A Bilgin, Hamideh Ramezani, and Ozgur B Akan. Transmitter and receiver architectures for molecular communications: A survey on physical design with modulation, coding, and detection techniques. *Proceedings of the IEEE*, 107(7):1302–1341, 2019.
- [12] Weisi Guo, Yansha Deng, H Birkan Yilmaz, Nariman Farsad, Maged ElKashlan, Andrew Eckford, Arumugam Nallanathan, and Chan-Byoung Chae. Smiet: Simultaneous molecular information and energy transfer. *IEEE Wireless Communications*, 25(1):106–113, 2017.
- [13] Vahid Jamali, Arman Ahmadzadeh, and Robert Schober. Symbol synchronization for diffusion-based molecular communications. *IEEE transactions on nanobioscience*, 16(8):873–887, 2017.
- [14] H Birkan Yilmaz, Akif Cem Heren, Tuna Tugcu, and Chan-Byoung Chae. Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Communications Letters*, 18(6):929–932, 2014.
- [15] Yansha Deng, Adam Noel, Weisi Guo, Arumugam Nallanathan, and Maged ElKashlan. Analyzing large-scale multiuser molecular communication via 3-d stochastic geometry. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 3(2):118–133, 2017.
- [16] Murat Kuscü and Ozgur B Akan. Maximum likelihood detection with ligand receptors for diffusion-based molecular communications in internet of bio-nano things. *IEEE transactions on nanobioscience*, 17(1):44–54, 2018.
- [17] Meriç Turan, Bayram Cevdet Akdeniz, Mehmet Şükrü Kuran, H Birkan Yilmaz, Ilker Demirkol, Ali E Pusane, and Tuna Tugcu. Transmitter localization in vessel-like diffusive channels using ring-shaped molecular receivers. *IEEE Communications Letters*, 22(12):2511–2514, 2018.
- [18] Tadashi Nakano, Tatsuya Suda, Takako Koujin, Tokuko Haraguchi, and Yasushi Hiraoka. Molecular communication through gap junction channels: System design, experiments and modeling. In *2007 2nd Bio-Inspired Models of Network, Information and Computing Systems*, pages 139–146. IEEE, 2007.
- [19] Massimiliano Pierobon and Ian F Akyildiz. A physical end-to-end model for molecular communication in nanonetworks. *IEEE Journal on Selected Areas in Communications*, 28(4):602–611, 2010.

-
- [20] KV Srinivas, Andrew W Eckford, and Raviraj S Adve. Molecular communication in fluid media: The additive inverse gaussian noise channel. *IEEE transactions on information theory*, 58(7):4678–4692, 2012.
- [21] Satoshi Hiyama, Yuki Moritani, Riho Gojo, Shoji Takeuchi, and Kazuo Sutoh. Biomolecular-motor-based autonomous delivery of lipid vesicles as nano-or microscale reactors on a chip. *Lab on a Chip*, 10(20):2741–2748, 2010.
- [22] Maria Gregori and Ian F Akyildiz. A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors. *IEEE Journal on selected areas in communications*, 28(4):612–619, 2010.
- [23] Luis C Cobo and Ian F Akyildiz. Bacteria-based communication in nanonetworks. *Nano Communication Networks*, 1(4):244–256, 2010.
- [24] Akihiro Enomoto, Michael J Moore, Tatsuya Suda, and Kazuhiro Oiwa. Design of self-organizing microtubule networks for molecular communication. *Nano Communication Networks*, 2(1):16–24, 2011.
- [25] Adam Noel, Karen C Cheung, and Robert Schober. Improving receiver performance of diffusive molecular communication with enzymes. *IEEE Transactions on NanoBioscience*, 13(1):31–43, 2014.
- [26] Steven S Andrews and Dennis Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical biology*, 1(3):137, 2004.
- [27] Nariman Farsad and Andrea Goldsmith. Sliding bidirectional recurrent neural networks for sequence detection in communication systems. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.
- [28] Xu Bao, Jie Lin, and Wence Zhang. Channel modeling of molecular communication via diffusion with multiple absorbing receivers. *IEEE Wireless Communications Letters*, 8(3):809–812, 2019.
- [29] Vahid Jamali, Arman Ahmadzadeh, Wayan Wicke, Adam Noel, and Robert Schober. Channel modeling for diffusive molecular communication—a tutorial review. *Proceedings of the IEEE*, 107(7):1256–1301, 2019.
- [30] Baris Atakan. Optimal transmission probability in binary molecular communication. *IEEE communications letters*, 17(6):1152–1155, 2013.
- [31] Andrew W Eckford. Nanoscale communication with brownian motion. In *2007 41st Annual Conference on Information Sciences and Systems*, pages 160–165. IEEE, 2007.
- [32] Bin Li, Weisi Guo, Xiang Wang, Yansha Deng, Yueheng Lan, Chenglin Zhao, and Arumugam Nallanathan. Csi-independent non-linear signal detection in molecular communications. *IEEE Transactions on Signal Processing*, 68:97–112, 2019.

Bibliography

- [33] Chunfu Luo, Xia Wu, Lin Lin, Chao Wang, and Fuqiang Liu. Non-coherent signal detection technique for mobile molecular communication at high data rates. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.
- [34] Ge Chang, Lin Lin, and Hao Yan. Adaptive detection and isi mitigation for mobile molecular communication. *IEEE transactions on nanobioscience*, 17(1):21–35, 2018.
- [35] Vahid Jamali, Nariman Farsad, Robert Schober, and Andrea Goldsmith. Non-coherent detection for diffusive molecular communication systems. *IEEE Transactions on Communications*, 66(6):2515–2531, 2018.
- [36] Yuting Fang, Adam Noel, Nan Yang, Andrew W Eckford, and Rodney A Kennedy. Symbol-by-symbol maximum likelihood detection for cooperative molecular communication. *IEEE Transactions on Communications*, 67(7):4885–4899, 2019.
- [37] Zhan Luo, Lin Lin, Weisi Guo, Siyi Wang, Fuqiang Liu, and Hao Yan. One symbol blind synchronization in simo molecular communication systems. *IEEE Wireless Communications Letters*, 2018.
- [38] Lin Lin, Chengfeng Yang, Maode Ma, Shiwei Ma, and Hao Yan. A clock synchronization method for molecular nanomachines in bionanosensor networks. *IEEE Sensors Journal*, 16(19):7194–7203, 2016.
- [39] Vahid Jamali, Arman Ahmadzadeh, Christophe Jardin, Heinrich Sticht, and Robert Schober. Channel estimation for diffusive molecular communications. *IEEE Transactions on Communications*, 64(10):4238–4252, 2016.
- [40] Md Humaun Kabir, SM Riazul Islam, and Kyung Sup Kwak. D-mosk modulation in molecular communications. *IEEE transactions on nanobioscience*, 14(6):680–683, 2015.
- [41] X. Qian, H. Deng, and H. He. Joint synchronization and channel estimation of aco-ofdm systems with simplified transceiver. *IEEE Photonics Technology Letters*, 30(4):383–386, Feb 2018.
- [42] Hoda ShahMohammadian, Geoffrey G Messier, and Sebastian Magierowski. Blind synchronization in diffusion-based molecular communication channels. *IEEE communications letters*, 17(11):2156–2159, 2013.
- [43] Martin Damrath and Peter Adam Hoeher. Low-complexity adaptive threshold detection for molecular communication. *IEEE transactions on nanobioscience*, 15(3):200–208, 2016.
- [44] Reza Mosayebi, Hamidreza Arjmandi, Amin Gohari, Masoumeh Nasiri-Kenari, and Urbashi Mitra. Receivers for diffusion-based molecular communication: Exploiting memory and sampling rate. *IEEE Journal on Selected Areas in Communications*, 32(12):2368–2380, 2014.

-
- [45] Adam Noel, Karen C Cheung, Robert Schober, Dimitrios Makrakis, and Abdelhakim Hafid. Simulating with accord: Actor-based communication via reaction–diffusion. *Nano Communication Networks*, 11:44–75, 2017.
- [46] Ignacio Llatser, Deniz Demiray, Albert Cabellos-Aparicio, D Turgay Altılar, and Eduard Alarcón. N3sim: Simulation framework for diffusion-based molecular communication nanonetworks. *Simulation Modelling Practice and Theory*, 42:210–222, 2014.
- [47] H Birkan Yilmaz and Chan-Byoung Chae. Simulation study of molecular communication systems with an absorbing receiver: Modulation and isi mitigation techniques. *Simulation Modelling Practice and Theory*, 49:136–150, 2014.
- [48] Benjamin FrantzDale, Steven J Plimpton, and Mark S Shephard. Software components for parallel multiscale simulation: an example with lammps. *Engineering with Computers*, 26(2):205–211, 2010.
- [49] Zahedeh Bashardanesh and Per Lötstedt. Efficient green’s function reaction dynamics (gfrd) simulations for diffusion-limited, reversible reactions. *Journal of Computational Physics*, 357:78–99, 2018.
- [50] Nariman Farsad, Weisi Guo, and Andrew W Eckford. Tabletop molecular communication: Text messages through chemical signals. *PloS one*, 8(12):e82935, 2013.
- [51] Nariman Farsad, David Pan, and Andrea Goldsmith. A novel experimental platform for in-vessel multi-chemical molecular communications. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.
- [52] Harald Unterweger, Jens Kirchner, Wayan Wicke, Arman Ahmadzadeh, Doaa Ahmed, Vahid Jamali, Christoph Alexiou, Georg Fischer, and Robert Schober. 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)*, pages 1–5. IEEE, 2018.
- [53] Haoyang Zhai, Qiang Liu, Athanasios V Vasilakos, and Kun Yang. Anti-isi demodulation scheme and its experiment-based evaluation for diffusion-based molecular communication. *IEEE transactions on nanobioscience*, 17(2):126–133, 2018.
- [54] Yu Huang, Miaowen Wen, Lie-Liang Yang, Chan-Byoung Chae, Xuan Chen, and Yuankun Tang. Space shift keying for molecular communication: Theory and experiment. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.
- [55] Laura Grebenstein, Jens Kirchner, Renata Stavracakis Peixoto, Wiebke Zimmermann, Florian Irnstorfer, Wayan Wicke, Arman Ahmadzadeh, Vahid Jamali, Georg Fischer, Robert Weigel, et al. Biological optical-to-chemical signal conversion interface: A small-scale modulator for molecular communications. *IEEE transactions on nanobioscience*, 18(1):31–42, 2018.

Bibliography

- [56] Xuewen Qian, Marco Di Renzo, and Andrew Eckford. K-means clustering-aided non-coherent detection for molecular communications. *IEEE Transactions on Communications*, 2020.
- [57] Xuewen Qian, Marco Di Renzo, and Andrew Eckford. Synchronization for constant-threshold-based receivers in diffusive molecular communication systems. *IEEE Transactions on Wireless Communications*, 2020.
- [58] Xuewen Qian and Marco Di Renzo. Receiver design in molecular communications: An approach based on artificial neural networks. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–5. IEEE, 2018.
- [59] Rodolfo Araya, Edwin Behrens, and Rodolfo Rodríguez. An adaptive stabilized finite element scheme for the advection–reaction–diffusion equation. *Applied Numerical Mathematics*, 54(3-4):491–503, 2005.
- [60] Alessio Zappone, Marco Di Renzo, and Mérouane Debbah. Wireless networks design in the era of deep learning: Model-based, ai-based, or both? *IEEE Transactions on Communications*, 67(10):7331–7376, 2019.
- [61] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [62] Xuewen Qian, Yansha Deng, Honggui Deng, Yue Hu, Chaoyang Zhang, and Jie Du. Synchronisation algorithm based on zero correlation code pair for ofdm-based vlc systems. *Iet Communications*, 11, 2016.
- [63] Xuewen Qian, Yansha Deng, Honggui Deng, Chaoyang Zhang, Jie Du, et al. Synchronization algorithm for ofdm/oqam systems based on zero autocorrelation code. *IET Communications*, 2017.
- [64] Xuewen Qian, Honggui Deng, and Hailang He. Pilot-based parametric channel estimation algorithm for DCO-OFDM-based visual light communications. *Opt. Commun.*, 400(April):150–155, 2017.
- [65] S Hiyama, Y Moritani, Tatsuya Suda, R Egashira, A Enomoto, M Moore, and T Nakano. Molecular communication. *Journal-Institute of Electronics Information and Communication Engineers*, 89(2):162, 2006.
- [66] Massimiliano Pierobon and Ian F Akyildiz. Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. *IEEE Transactions on Information Theory*, 59(2):942–954, 2013.
- [67] S Mohammadreza Rouzegar and Umberto Spagnolini. Channel estimation for diffusive mimo molecular communications. In *Networks and Communications (EuCNC), 2017 European Conference on*, pages 1–5. IEEE, 2017.

-
- [68] Reza Mosayebi, Amin Gohari, Mahtab Mirmohseni, and Masoumeh Nasiri-Kenari. Type-based sign modulation and its application for isi mitigation in molecular communication. *IEEE Transactions on Communications*, 66(1):180–193, 2018.
- [69] Po-Jen Shih, Chia-han Lee, and Ping-Cheng Yeh. Channel codes for mitigating intersymbol interference in diffusion-based molecular communications. In *2012 IEEE global communications conference (GLOBECOM)*, pages 4228–4232. IEEE, 2012.
- [70] Deniz Kilinc and Ozgur B Akan. Receiver design for molecular communication. *IEEE Journal on Selected Areas in Communications*, 31(12):705–714, 2013.
- [71] Hao Ye, Geoffrey Ye Li, and Bing-Hwang Juang. Power of deep learning for channel estimation and signal detection in ofdm systems. *IEEE Wireless Communications Letters*, 7(1):114–117, 2018.
- [72] P. Dong, H. Zhang, and G. Y. Li. Machine learning prediction based csi acquisition for fdd massive mimo downlink. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2018.
- [73] H Birkan Yilmaz, Changmin Lee, Yae Jee Cho, and Chan-Byoung Chae. A machine learning approach to model the received signal in molecular communications. In *2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pages 1–5. IEEE, 2017.
- [74] Changmin Lee, H Birkan Yilmaz, Chan-Byoung Chae, Nariman Farsad, and Andrea Goldsmith. Machine learning based channel modeling for molecular mimo communications. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2017.
- [75] X. Qian and M. Di Renzo. Receiver design in molecular communications: An approach based on artificial neural networks. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–5, Aug 2018.
- [76] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [77] Adam Noel, Karen C Cheung, and Robert Schober. Using dimensional analysis to assess scalability and accuracy in molecular communication. In *2013 IEEE International Conference on Communications Workshops (ICC)*, pages 818–823. IEEE, 2013.
- [78] Philip Nelson and Sebastian Doniach. *Biological physics: Energy, information life*. WH Freeman New York, 2004.
- [79] Mehmet S Kuran, Huseyin Birkan Yilmaz, Tuna Tugcu, and Ian F Akyildiz. Modulation techniques for communication via diffusion in nanonetworks. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5. IEEE, 2011.

Bibliography

- [80] Adam Noel, Karen C Cheung, and Robert Schober. Joint channel parameter estimation via diffusive molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 1(1):4–17, 2015.
- [81] Alessio Zappone, Marco Di Renzo, Mérouane Debbah, Thanh Tu Lam, and Xuewen Qian. Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization. *IEEE Vehicular Technology Magazine*, 14(3):60–69, 2019.
- [82] Bin Li, Mengwei Sun, Siyi Wang, Weisi Guo, and Chenglin Zhao. Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications. *IEEE Transactions on Communications*, 64(5):2079–2091, 2016.
- [83] S. Liu, Z. Wei, B. Li, and C. Zhao. Unsupervised clustering-based non-coherent detection for molecular communications. *IEEE Communications Letters*, 24(8):1687–1690, 2020.
- [84] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.
- [85] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.
- [86] Alan C Bovik. *Handbook of image and video processing*. Academic press, 2010.
- [87] Lin Lin, Feiyan Li, Maode Ma, and Hao Yan. Synchronization of bio-nanomachines based on molecular diffusion. *IEEE Sensors Journal*, 16(19):7267–7277, 2016.
- [88] Lin Lin, Jiali Zhang, Maode Ma, and Hao Yan. Time synchronization for molecular communication with drift. *IEEE Communications Letters*, 21(3):476–479, 2016.
- [89] Lin Lin, Chengfeng Yang, Maode Ma, and Shiwei Ma. Diffusion-based clock synchronization for molecular communication under inverse gaussian distribution. *IEEE Sensors Journal*, 15(9):4866–4874, 2015.
- [90] Youngbae Hwang, Jun-Sik Kim, and In-So Kweon. Sensor noise modeling using the skellam distribution: Application to the color edge detection. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007.
- [91] Govinda M Kamath, Eren Şaşoğlu, and David Tse. Optimal haplotype assembly from high-throughput mate-pair reads. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 914–918. IEEE, 2015.
- [92] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- [93] Ge Chang, Lin Lin, and Hao Yan. Adaptive detection and isi mitigation for mobile molecular communication. *IEEE Transactions on nanobioscience*, 17(1):21–35, 2017.

- [94] A Oguz Kislal, H Birkan Yilmaz, Ali E Pusane, and Tuna Tugcu. Isi-aware channel code design for molecular communication via diffusion. *IEEE transactions on nanobioscience*, 18(2):205–213, 2019.
- [95] Richard C Dorf and Robert H Bishop. *Modern control systems*. 1998.

Titre: Schémas avancés de détection et de synchronisation pour les communications moléculaires

Mots clés: Communications moléculaires, réseaux de neurones artificiels, détecteur, K-Means, synchronisation, borne inférieure de Cramer-Rao.

Résumé: Avec une capacité améliorée à manipuler la matière aux échelles nano et micro via des techniques biologiques et chimiques synthétiques, il existe de nouvelles opportunités pour relever des défis allant du diagnostic et du traitement des maladies à la protection de l'environnement. Un cadre clé pour développer ces outils est le nano-réseau, où les réseaux sont constitués de dispositifs à l'échelle nanométrique qui fonctionnent dans des environnements à l'échelle nanométrique à micrométrique et effectuent des tâches simples telles que la détection et l'actionnement. L'efficacité du nano-réseau dépend cependant fortement de la capacité des appareils à se coordonner entre eux. Les communications moléculaires ont été proposées comme moyen prometteur de coordination dans les nano réseaux. Dans les communications moléculaires, les informations entre les appareils sont échangées via des molécules émises et absorbées respectivement par l'émetteur et le récepteur. Les principes de base des communications moléculaires sont

basés sur des aspects fondamentaux et interdisciplinaires qui englobent la physique, la chimie, la biologie, ainsi que d'autres domaines de recherche tels que la pharmacologie, la microfluidique et la médecine. Cependant, il existe plusieurs défis dans le développement de techniques de traitement du signal et de communication pour le codage et le décodage d'informations, ainsi que dans le développement d'implémentations pratiques pour les réseaux à l'échelle nanométrique et micro basés sur les communications moléculaires.

Dans la présente thèse, nous abordons deux questions fondamentales liées à la conception et à l'optimisation de la conception du récepteur, qui englobent la détection des données et la synchronisation temporelle. Plus précisément, nous proposons des algorithmes innovants qui ne nécessitent pas d'informations sur l'état des canaux et tiennent compte des interférences entre symboles. De plus, nous caractérisons leurs performances par des analyses mathématiques et des simulations numériques.

Title: Advanced Detection and Synchronization Schemes for Molecular Communications

Keywords: Molecular communications, artificial neural networks, detector, K-Means, synchronization, Cramer-Rao lower bound.

Abstract: With an improved ability to manipulate the matter at the nano and micro scales via synthetic biological and chemical techniques, there exist new opportunities to address challenges ranging from disease diagnosis and treatment to environmental protection. A key framework to develop these tools is nano networking, where networks are made of nanoscale devices that operate in nano to micrometer scale environments and perform simple tasks such as sensing and actuation. The effectiveness of nano networking strongly depends, however, on the ability of the devices to coordinate among themselves.

Molecular communications have been proposed as a promising means of coordination in nano networks. In molecular communications, the information between devices is exchanged via molecules that are emitted and absorbed by the transmitter and receiver, respectively. The ba-

sic principles of molecular communications are based on fundamental and interdisciplinary aspects that encompass physics, chemistry, biology, as well as other research areas including pharmacology, microfluidics, and medicine. However, there are several challenges in developing signal processing and communication techniques for encoding and decoding information, as well as developing practical implementations for nano and micro scale networks based on molecular communications.

In the present thesis, we tackle two fundamental issues that are concerned with the design and optimization of the receiver design, which encompass data detection and time synchronization. More precisely, we propose innovative algorithms that do not necessitate channel state information and account for inter-symbol interference. In addition, we characterize their performance through mathematical analysis and numerical simulations.