

Energy and Spectrum-Efficient Communication Techniques for Next-Generation Internet of Things

by

Ecehan Berk Pehlivanoglu

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Electrical and Electronics Engineering



January 9, 2018

**Energy and Spectrum-Efficient Communication Techniques for
Next-Generation Internet of Things**

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Ecehan Berk Pehlivanoglu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Prof. Dr. Özgür B. Akan (Advisor)

Prof. Dr. A. Murat Tekalp

Assoc. Prof. Dr. Berk Canberk

Assoc. Prof. Dr. Sinem Çöleri Ergen

Assoc. Prof. Dr. Özgür Gürbüz

Date: _____



*To my fiancée Merve Altınoğlu, my parents Şebnem and Ertan Pehlivanoglu and my
best friends Paşa, Mustafa and Oktay...*

ABSTRACT

Internet of Things (IoT) is a recent paradigm, rapidly gaining traction across a variety of disciplines, including but not limited to communications, electronics, computer engineering and social sciences. In its fundamentals, IoT provides a link between cyber and physical world by connecting devices over the Internet. Recent advances in hardware systems, signal processing and wireless communications have made it possible to manufacture low cost, low complexity wireless communicating things; which are the key to full scale adoption of the IoT paradigm. On the other hand, with the recent progress in nanotechnology, things are also quickly moving into to microscales and nanoscales. As the offspring of nanotechnology, nanomachines are strong candidates for realization of Internet of Bio-Nano Things (IoBNT).

Full realization of IoT and IoBNT, however, will require surmounting of additional challenges. One of the upcoming difficulties is spectrum scarcity, which is a result of the inefficiency of the fixed spectrum assignment paradigm. Additionally, most of the communicating entities are traditionally batter-powered and run out of batteries at fairly random instants after their deployment, which makes battery replenishments an unfeasible operation. With the scale game, rising need for self-sustainability and coordination is a clear reality. Whilst IoBNT opens doors for advanced applications such as nanoscale sensing and intelligent drug delivery, they depend on nanomachines with scarce processing, memory, and networking capabilities. There is clearly a need for energy and medium-efficient nanonetworking techniques, that would help nanomachines carry out more complex tasks, towards the IoBNT vision.

To that end, in this thesis, we focus on energy and spectrum-efficient communication techniques for next-generation networks that are likely to constitute the backbones of IoT and IoBNT. We address key challenges brought by IoT and IoBNT ap-

plications, and solutions provided by next-generation networks and techniques thereof in two main parts:

In Part I, we consider next-generation ad hoc networks, which are enhanced with energy and spectrum-efficient communication techniques to satisfy demanding future IoT functions. More specifically, architectures and techniques in Cognitive Radio Ad Hoc Networks (CRAHN) and Cognitive Radio Sensor Networks (CRSN) are studied to optimize performance in critical metrics such as sensing coverage, connectivity, spectrum- and energy-efficient communication duration and goodput-meters-per-Joule toward the fusion center. For future IoT applications in which demanding data rates cannot be achieved with today's traditional spectrum, a self-organizing capacity optimization procedure for low THZ band is developed. Lastly, a novel electric-field base energy harvesting method is proposed, tested and studied in detail to enable self-sustaining networks for real-life IoT applications.

In Part II, we consider architectures that will help realization of nanonetworks, composed of nanomachines. Different communication types, namely; acoustic, electromagnetic, nanomechanical and molecular communications could be viable alternatives nanonetworking. However, acoustic and electromagnetic communications require entities that can carry out acoustic or electromagnetic operations at nanoscale. On the other hand, nanomechanical communications require direct contact between communicating parties and is not suitable for distant applications. In this context, we focused on Molecular Communications (MC) in this thesis, as it is a viable method for nanonetworking, backed by the fact that similar mechanisms are already present in many living organisms and environments. We extensively review in detail the existing modulation schemes in MC, and propose a novel messenger-based modulation scheme to increase channel performance in terms of achievable channel capacity without the need for synchronization.

ÖZETÇE

Nesnelerin İnterneti (Nİ), son dönemde haberleşme, elektronik, bilgisayar mühendisliği ve sosyal bilimler de dahil olmak üzere birçok alanda ivme kazanmış, yeni bir paradigmadır. Temelde Nİ, nesneleri çevrimiçi hale getirerek, siber ve fiziksel dünyalar arasında bir bağlantı sağlamaktadır. Elektronik donanımlar, sinyal işleme ve kablosuz haberleşme alanlarındaki son dönemdeki gelişmeler, düşük maliyetli ve düşük karmaşıklığa sahip kablosuz haberleşebilen nesnelerin üretimine olanak sağlamıştır. Bu durum, Nİ paradigmasının geniş çapta benimsenmesinde anahtar rol oynamaktadır. Öte yandan, nanoteknolojideki son gelişmelerle birlikte, nesneler hızla mikro ve hatta nano ölçeğe doğru inmektedirler. Nanoteknoloji ürünü olan nanomakinalar, Biyo Nano Nesnelerin İnterneti (BNNİ) paradigmasının gerçekleşmesi için en önemli adaylardır.

Fakat, Nİ ve BNNİ'nin geniş çapta gerçekleşmesi için, bir takım ilave zorlukların üstesinden gelinmelidir. Baş göstermekte olan zorluklardan biri tayf kısıtlılığıdır, ve bu kısıtlılık günümüzde nesnelere verimsiz şekilde sabit tayf atama yaklaşımından kaynaklanmaktadır. Bununla birlikte, nesnelerin birçoğu geleneksel olarak bataryalardan güç almaktadır, ve konuşlanmaları sonrasında nesnelerin oldukça farklı zamanlarda enerji depoları tükenmektedir. Bu durum, günümüzde nesnelerin yoğun konuşlanması da göz önünde bulundurulduğunda, batarya değişimini de olanaksız kılmaktadır. Büyüyen ağ ölçekleri sebebiyle, kendini idame ettiren ve koordine eden ağlara ve tekniklere ihtiyacın arttığı da bir gerçektir. Öte yandan, BNNİ, nano-ölçekte algılama ve akıllı ilaç taşıma uygulamalarını mümkün kılrsa da, çoğunlukla kısıtlı işleme, hafıza ve haberleşme yeteneklerine sahip nanomakinalara bağlı kalmaktadır. Dolayısıyla, BNNİ vizyonunun başarılması adına, nanomakinaların daha zorlu görevleri yerine getirmesini sağlayabilecek enerji ve ortam-verimli nano-ağ oluşturma tekniklerine ihtiyaç

duyulduğu aşıkardır.

Bu bağlamda, bu tezde, gelecekteki Nİ ve BNNİ sistemlerinin belkemiği olması beklenen gelecek nesil ağlar için, enerji ve tayf-verimli haberleşme teknikleri geliştirilmesine odaklanılmıştır. Gelecek Nİ ve BNNİ uygulamalarının getirdiği zorluklara, bunlara gelecek nesil haberleşme ağlarının nasıl cevap verebileceğine ve ne gibi teknikleri kullanabileceklerine, iki bölümde değinilmektedir:

İlk bölümde, gelecek-nesil tasarsız ağlar üzerine yoğunlaşmış, bu ağların Nİ uygulama gereksinimlerine cevap verebilmesi adına enerji ve tayf-verimli teknikler geliştirilmiştir. Bilişsel Radyolu Tasarsız Ağlar (BRTA) ve Bilişsel Radyolu Algılayıcı Ağlar (BRAA) üzerine yoğunlaşarak, bu ağlarda algılama kapsama alanı, bağlanırlık, enerji ve tayf-verimli iletişim süresi uzunluğu ve harcanan Joule başına başarılan veri teslimi ve katedilen yol gibi performans göstergeleri eniyilenmiştir. Günümüz geleneksel tayfı kullanılarak erişilmesi mümkün görünmeyen seviyelerde veri hızı talebi olabilecek Nİ uygulamaları için; düşük THz bandını kullanan, kendini örgütleyebilen ve kapasiteyi eniyilen bir prosedür geliştirilmiştir. Son olarak, Nİ uygulamalarında ağların kendini idame ettirebilmesi adına elektrik alan bazlı yenilikçi bir enerji hasadı metodu önerilmiş, test edilmiş ve performansı ortaya konmuştur.

İkinci bölümde, nanomakinalardan oluşan nanoağların gerçekleşmesine yardımcı olacak mimariler üzerine yoğunlaşmıştır. Buna göre, nanoağların gerçekleşmesi için; akustik, elektromanyetik, nanomekanik ve moleküler olmak üzere dört tür haberleşme yöntemi izlenebilir. Fakat nano ölçekte akustik veya elektromanyetik operasyonları yürütebilecek nesneler henüz geliştirilmemiştir. Öte yandan nanomekanik haberleşme, iletişim kurmak isteyen nesneler arasında direk fiziksel teması zorunlu kılmaktadır; dolayısıyla uzak uygulamalar için uygun değildir. Bu nedenle, bu bölümde nanoağlar için Moleküler Haberleşme (MH) seçeneği üzerinde durulmuştur. MH benzeri haberleşme mekanizmaları, doğada canlı birçok organizma tarafından hali hazırda da kullanılmaktadır. Bu tezde, MH için mevcut tüm kipleme yöntemleri detaylı şekilde araştırılmış ve karşılaştırılmıştır. Bunun sonucunda, MH için senkronizasyon gerek-

tirmeyen ve moleküler kanalın kullanımı başına yüksek kapasiteler sağlayabilen bir kipleme yöntemi geliştirilmiştir.



ACKNOWLEDGMENTS

I would like to gratefully acknowledge the exemplary advisership of Dr. Özgür Barış Akan throughout the entire development process of this thesis. He has been a role model for me not only in academia but also in other professional activities that I conduct in the corporate world. I will always admire his mentoring style, which pushes you to realize your full potential, and will always feel very proud and fortunate to be one of his students.

I would like to wholeheartedly thank Dr. A. Murat Tekalp and Dr. Berk Canberk for taking part in my thesis committee, accepting to supervise this thesis and to drive it further with their valuable comments and suggestions throughout the process. I am also grateful for the time and dedication of the thesis jury members Dr. Sinem Çöleri Ergen and Dr. Özgür Gürbüz.

I want to thank the Next-generation and Wireless Communications Laboratory members for their continued friendship and collaboration. I want to specifically thank Mustafa Özger, Oktay Çetinkaya, Bige Deniz Ünlütürk and Ahmet Ozan Biçen for their extreme support and co-authorship in numerous research studies that we conducted.

I would like to acknowledge the magnificent support of TÜBİTAK and Koç University, which made the completion of this thesis possible.

Last but not the least, I would like to thank my fiancée Merve, my parents Şebnem and Ertan, and my best friend Paşa for supporting me at all times, good and bad. I feel very fortunate to be a part of such a loving circle.

TABLE OF CONTENTS

List of Tables	xvi
List of Figures	xvii
Nomenclature	xxi
Chapter 1: Introduction	1
1.1 Energy and Spectrum-Efficient Communication Techniques for IoT . . .	1
1.2 Contributions	4
1.3 Detailed Thesis Outline	6

I Energy and Spectrum-Efficient Communication Techniques in Next-Generation Ad Hoc and Sensor Networks for IoT 11

Chapter 2: Sensing Coverage and Connectivity in Cognitive Radio Sensor Networks	12
2.1 Introduction	12
2.2 Motivation for Sensing Coverage and Connectivity Research in CRSN	18
2.3 Sensing Coverage in Ad Hoc Networks	21
2.3.1 Energy-Efficient Sensing Coverage by Node Scheduling in WSN	22
2.3.2 Sensing Coverage Quality and Network Lifetime Trade-off in WSN	24
2.3.3 Sensing Coverage Quality in Randomly Deployed WSN	25
2.3.4 Node Density for Energy-Efficient Monitoring in WSN	26
2.3.5 Deployment Strategies for Sensing Coverage in WSN	27

2.3.6	k -coverage in WSN	28
2.3.7	Coverage in Cognitive Radio Networks	28
2.4	Connectivity in Ad Hoc Networks	31
2.4.1	Connectivity Derived from Node Isolation Probability, Boolean Model	31
2.4.2	Connectivity Derived from Node Isolation Probability, Under Fading	34
2.4.3	Connectivity Derived From Asymptotic Behavior	35
2.4.4	Connectivity in Cognitive Radio Ad Hoc Networks	35
2.5	Model Recommendations for Sensing Coverage and Connectivity in CRSN	38
2.5.1	Joint Sensing Coverage and Connectivity Model for CRSN	38
2.5.2	Connectivity Model for CRSN Based on Clustering	44
2.5.3	Connectivity Assessment Models for CRSN	45
2.6	Open Topics for Sensing Coverage and Connectivity in CRSN	46

Chapter 3: Energy-Efficient Transmission Range and Duration for CRSN 48

3.1	Introduction	48
3.2	Related Work and Contributions	50
3.3	System Model	52
3.4	Analysis Framework	54
3.4.1	Spectrum Sensing Events and Their Probabilities	54
3.4.2	Energy Consumption Dynamics	55
3.4.3	Spectrum Utilization Scenarios	58
3.5	Energy-Efficient r_s and τ_t for CRSN	62
3.5.1	Expected Hop Progress	63
3.5.2	Expected Hop Distance	64

3.5.3	Problem Formulation to Reveal Optimal r_s^* and τ_t^*	66
3.6	Numerical Results	66
3.6.1	Preliminaries on τ_t and τ_s	67
3.6.2	Behavior of SFT Probability P_{S6}	69
3.6.3	Reward Function Λ With Respect to r_s and τ_t	71
3.6.4	Sensitivity of Optimal r_s^* , τ_t^* and Λ^*	74
3.6.5	Sensitivity of r_s^* , τ_t^* , Λ^* to κ , β and α	78
Chapter 4:	Common Control Interface in Cognitive Radio Networks:	
	When to Use It?	81
4.1	Introduction	81
4.2	Related Work and Contributions	83
4.3	System Model	85
4.4	Analysis Framework	86
4.4.1	Spectrum Access (SA) Delay	87
4.4.2	Successful Frame Transmission (SFT) Delay Without CCI . .	90
4.4.3	Successful Frame Transmission (SFT) Delay With CCI	91
4.4.4	Energy-Efficiency During SFT With and Without CCI	93
4.4.5	Spectrum-Efficiency During SFT With and Without CCI . . .	94
4.4.6	Event Sensing Performance in CRSN With and Without CCI	94
4.5	Numerical Results	96
4.5.1	Network Setup	96
4.5.2	Evaluation of SFT Delay	97
4.5.3	Evaluation of Energy-Efficiency With and Without CCI	101
4.5.4	Evaluation of Spectrum-Efficiency With and Without CCI . .	102
4.5.5	Evaluation of Event Sensing Performance With and Without CCI	103
Chapter 5:	Self-Optimization for Capacity in Low Terahertz Band	
	Indoor Wireless Networks	105

5.1	Introduction	105
5.2	System Model	107
5.2.1	Network Model	107
5.2.2	Antenna Radiation Pattern	108
5.2.3	Channel and Interference Model	109
5.3	Self-Optimization for Capacity in Low THz Band Indoor Wireless Net- works	110
5.3.1	Analysis Framework	110
5.3.2	Proposed Approximate Water Filling (AWF) Protocol	112
5.4	Numerical Results	112
Chapter 6: Harvesting-Throughput Trade-off for Wireless-powered Smart Grid IoT Applications: An Experimental Study		116
6.1	Introduction	116
6.2	System Model	119
6.2.1	Experimental Set-up for the Proposed EFEH Method	119
6.2.2	Nodal Operation and Communication Model	121
6.3	Harvesting-Throughput Trade-off Problem Formulation	122
6.4	Numerical Results	126
6.4.1	Theoretical and Practical Battery-less Operating Regions	126
6.4.2	Behavior of Throughput Function θ With Respect to T_H , T_T	128
6.4.3	Sensitivity of Optimal T_H^* , T_T^* to System Parameters	128
II Energy and Spectrum-Efficient Communication Tech- niques in Nanonetworks for IoBNT		131
Chapter 7: Modulation in Molecular Communications: A Look on Methodologies		132
7.1	Introduction	132

7.2	Motivation for Modulation Research in Molecular Communications . . .	134
7.3	Modulation in Molecular Communications	137
7.3.1	Concentration Based Modulation Schemes	137
7.3.2	Molecule Type Based Modulation Schemes	143
7.3.3	Release Time Based Modulation Schemes	146
7.3.4	Mixed Type Modulation Schemes	150
7.3.5	Modulation and Nanomachines in Molecular Communication . .	151
7.4	Segmentation of Related Work on Key Criteria	152
Chapter 8:	Polyethylene: A Promising Messenger for High Perfor-	
	mance Molecular Communications	155
8.1	Introduction	155
8.2	Related Work and Contributions	156
8.3	System Model	157
8.3.1	Encoding Information on Polyethylene Molecules	158
8.3.2	Transmission of Polyethylene Messenger Molecules	161
8.3.3	Propagation of Polyethylene Messenger Molecules	162
8.3.4	Reception of Polyethylene Messenger Molecules	162
8.3.5	Decoding of Polyethylene Messenger Molecules	163
8.4	Molecular Communications as an Erasure Channel	164
8.5	Numerical Results	167
8.5.1	Impact of Viscosity of the Medium on Performance	167
8.5.2	Impact of Molecule Lifetime τ on Performance	167
8.5.3	Impact of Number of Redundant Molecules N_M on Performance	168
8.5.4	Closed Form Expression for Optimal Molecule Size	169
Chapter 9:	Conclusions	172

Bibliography	178
Bibliography	178



LIST OF TABLES

2.1	Key related work on sensing coverage in wireless sensor and ad hoc networks.	30
3.1	Spectrum utilization scenarios, probabilities and energy consumption thereof.	57
3.2	Probability mass function of random variable T.	61
3.3	Parameters and their assumed values to reveal optimal r_s^* and τ_t^* . . .	67
3.4	Changes to terms of the reward function Λ when individual parameters are increased, all else being equal.	72
3.5	Sensitivity of optimal r_s^* , τ_t^* and Λ^* values with respect to changes in critical network parameters.	73
4.1	Illustration of Spectrum Access (SA) delay for a channel list composed of 3 channels.	89
5.1	System parameters used in simulations for capacity optimization in low THz band.	114
6.1	System parameters used for throughput optimization in proposed system [119].	126
7.1	Comparison of traditional vs. molecular communications	134
7.2	Research efforts in modulation in molecular communications.	154

LIST OF FIGURES

1.1	Domains covered in this thesis.	2
1.2	Contributions of the thesis.	4
2.1	(a) Unidirectional link from node A to node B ($R_a > R_d > R_b$); (b) Bidirectional link between node A and node B ($R_a > R_d$ and $R_b > R_d$).	17
2.2	Actual and assumed forwarding regions for a source CRSN node.	41
2.3	Guardrings of radius R_g centered at transmitter and receiver SU nodes.	42
2.4	Illustration of cluster formation in [31] by iterations.	44
3.1	Assumed time slot structure for SU spectrum access and frame transmission.	54
3.2	Illustration of guardring concept.	59
3.3	Illustration of Least Remaining Distance (LRD) forwarding approach.	63
3.4	Relation between hop progress and hop distance.	65
3.5	(a) Required spectrum sensing duration τ_s vs. τ_t for different P_{col} cases; (b) Comparison of sensing efficiency (defined as $\Theta = \frac{\tau_s}{\tau_s + \tau_t}$) and SFT efficiency (defined as $\Omega = \frac{\tau_t}{E\{\tau_{sft}\}}$ [52] for different P_{col} cases.	68
3.6	Probabilities of scenario sets A, B and C vs. τ_t	69
3.7	(a) Probability of success scenario P_{S6} vs. τ_t for different r_s cases; (b) Probability of success scenario P_{S6} vs. τ_t for different P_{col} cases.	70
3.8	Illustration of the reward function Λ with respect to r_s and τ_t	71

3.9	Optimal $(r_s^*, \tau_t^*, \Lambda^*)$ with respect to (a) $\alpha = 10, \beta = 10$; (b) $\alpha = 10, \beta = 20$; (c) $\alpha = 10, \beta = 30$; (d) $\alpha = 20, \beta = 10$; (e) $\alpha = 20, \beta = 20$; (f) $\alpha = 20, \beta = 30$; (g) $\alpha = 30, \beta = 10$; (h) $\alpha = 30, \beta = 20$; (i) $\alpha = 30, \beta = 30$, when κ changes from 2.0 to 3.4 with 0.1 steps, and PU guardring radius set as $r_p = 100$ m.	79
4.1	Illustration of Successful Frame Transmission (SFT) delay τ_{sft} with CCI.	92
4.2	Comparison of analytical and simulation based results on SFT delay.	98
4.3	SFT delay with and without CCI, with respect to $\frac{\tau_f}{E\{\tau_s\}}$ ratio, for different PU arrival rates.	99
4.4	SFT delay with respect to λ_{SU} , with and without CCI. $E_b/N_0 = 10$ dB at $\lambda_{\text{SU}} = 0.005$, λ_{PU} is assumed to be 0.0125.	100
4.5	Equivalent node cost, to achieve a given SFT delay requirement, for different CCI cost to full node cost ratios.	100
4.6	Energy-efficiency with and without CCI, for two transmission and sensing power level pairs under low (a) and high (b) PU traffic.	101
4.7	Goodput/(SFT energy consumption) ratio (in bits/(second·Joule)), under different PU traffic and CCI power levels.	102
4.8	Distributed sensing distortion performance with and without CCI, under different PU arrival rates.	103
5.1	Assumed network model, with FD-MIMO capable BS at the center of the cell, UEs are deployed randomly	107
5.2	(a) Radiation pattern of a single antenna with ideal <i>sinc</i> function, directed at $\theta = 0$; (b) Realistic radiation pattern of a single antenna with modified <i>sinc</i> function, directed at $\theta = 0$	109
5.3	(a) Total network capacity with respect to N for equal power allocation vs. AWF; (b) Average user capacity with respect to N for equal power allocation vs. AWF	114

5.4	Average number of iterations and algorithm run-time with respect to N	115
6.1	(a) Representative diagram of the proposed EFEH method; (b) Experimental set-up of the proposed EFEH method.	120
6.2	Operation cycle of the proposed EFEH method.	121
6.3	Harvest-then-transmit cycle of the EFEH-powered node.	123
6.4	(a) Accumulated voltage V_s and energy E_H on capacitor C with respect to T_H ; (b) Accumulated energy E_H and consumed energy for sleep mode and nodal operations $E_N + E_Z$ with respect to T_H	127
6.5	Value of objective function θ with respect to T_H and T_T	128
6.6	T_H^* and T_T^* values that maximize θ , for path loss exponents of (a) $\nu = 2$, (b) $\nu = 2.5$ and (c) $\nu = 3$, respectively.	129
6.7	T_H^* and T_T^* values that maximize θ , for path loss exponents of $\nu = 2 : 0.5 : 3$, where T_{max} is swept from 160 to 300 sec. Points with same T_{max} are connected with dashed lines.	130
7.1	Diffusion based molecular communication [140]	133
7.2	Detection by thresholds for Binary CSK and Quadruple CSK	141
7.3	Constellation diagram for QMoSK [134]	144
7.4	Isomers of aldohexoses	145
7.5	Illustration of TEC signaling at the sender and receiver sides [141].	147
7.6	Molecular array modulation, when only two molecule types are considered.	149
8.1	Illustration of Quadruple MoSK modulation, for molecule size $n = 2$	160
8.2	Illustration of transmitter and receiver nanomachines, messenger molecule emission and reception.	161
8.3	Block erasure channel representation.	165

8.4	Molecular erasure channel capacity with respect to messenger molecule size n , for varying viscosity levels, η	168
8.5	Molecular erasure channel capacity with respect to messenger molecule size n , for varying molecule lifetimes, τ	169
8.6	Molecular erasure channel capacity with respect to messenger molecule size n , for varying number of identical copies per emission, N_M	169
8.7	The fitted surface for the optimal number of bits, n	170

NOMENCLATURE

5G	Fifth Generation
AM	Amplitude Modulation
AP	Access Point
AWGN	Additive White Gaussian Noise
B4G	Beyond Fourth Generation
BER	Bit Error Rate
BLUE	Best Linear Unbiased Estimator
BS	Base Station
CCI	Common Control Interface
CR	Cognitive Radio
CRAHN	Cognitive Radio Ad Hoc Networks
CRN	Cognitive Radio Networks
CRSN	Cognitive Radio Sensor Networks
DSA	Dynamic Spectrum Access
EFEH	Electric Field Energy Harvesting
EH	Energy Harvesting
FA	False Alarm
FCC	Federal Communications Commission
FD-MIMO	Full Dimensional MIMO
FM	Frequency Modulation
FSK	Frequency Shift Keying
GHz	Gigahertz
ICW	Intercellular Calcium Waves

IoT	Internet of Things
ISI	Inter Symbol Interference
ISM	Industrial, Scientific and Medical
LoS	Line of Sight
LRD	Least Remaining Distance
M-AM	Multilevel Amplitude Modulation
MC	Molecular Communications
MDMA	Molecule Division Multiple Access
MoSK	Molecule Shift Keying
MSE	Mean Square Error
OOK	On-Off Keying
PPP	Poisson Point Process
PU	Primary User
PVD	Poisson Voronoi Diagram
QoS	Quality of Service
SA	Spectrum Access
SFT	Successful Frame Transmission
SG	Smart Grid
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SO	Spectrum Opportunity
SSA	Successful Spectrum Access
SU	Secondary User
TEC	Time Elapse Communications
THz	Terahertz
UE	User Equipment
WSN	Wireless Sensor Network

Chapter 1

INTRODUCTION

1.1 Energy and Spectrum-Efficient Communication Techniques for IoT

Internet of Things (IoT) provides a link between cyber and physical world by connecting devices over the Internet. Recent advances in micro-electromechanical systems, signal processing and wireless communications have made it possible to manufacture low cost, low complexity wireless communicating entities, which are the key to full scale take-up of the IoT paradigm. This proliferation of wireless entities, however, brings additional challenges including but not limited to spectrum scarcity, rising need for self-sustainability and coordination. On the other hand, with the recent advancements in nanotechnology, things are also quickly moving into to micro-scales and nanoscales. Offspring of nanotechnology, nanomachines are strong candidates for realization of IoT at nanoscale, yet they have scarce processing, memory, and networking capabilities. There is clearly a need for energy and medium-efficient nanonetworking techniques, that would help nanomachines to carry out more complex tasks, to enable Internet of Bio-Nano Things (IoBNT).

In this thesis, we focus on energy and spectrum-efficient communication techniques for next-generation networks that are likely to constitute the backbones of IoT and IoBNT. We address key challenges brought by IoT and IoBNT applications, and solutions provided by next-generation networks and techniques thereof in two main parts:

In **Part I**, we consider next-generation ad hoc networks, which are enhanced with energy and spectrum-efficient communication techniques to satisfy demanding

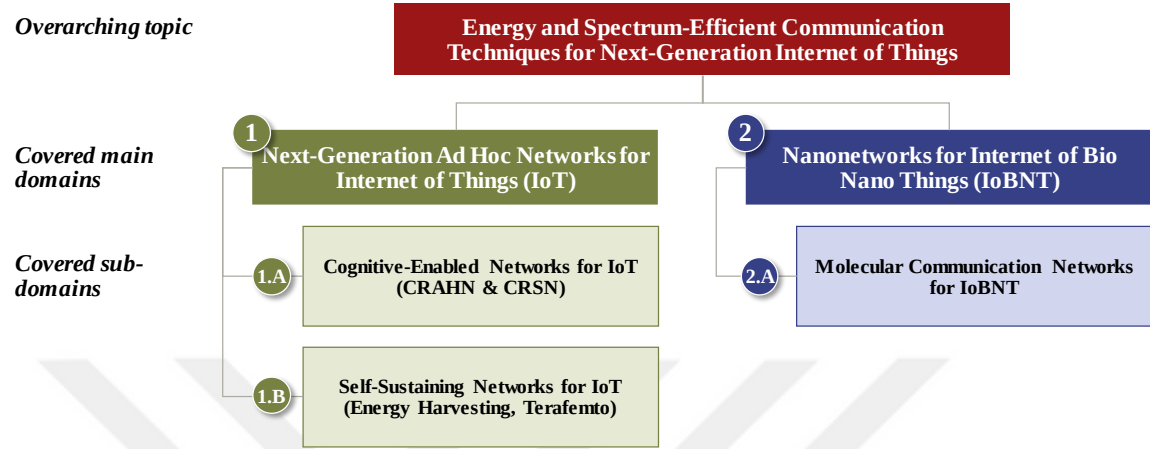


Figure 1.1: Domains covered in this thesis.

future IoT functions. IoT applications require reliability, higher data rates and self-sustainability from their underlying communicating entities and networks. Main challenges faced by traditional communication networks under these applications, include but are not limited to:

- Spectrum scarcity, driven by proliferation in wireless devices and low spectrum efficiency of the current fixed spectrum assignment paradigm,
- Limited lifetime in traditional battery mode due to energy constraints,
- Current spectral efficiency methods being already at their limits, not being able to provide desired data rate levels on the order of 10 Gbps with currently occupied spectrum,
- Random run out time for batteries, coupled with high densities, making it very challenging to replenish batteries,
- Deployment in usually high densities, increasing need for diligence in design for performance for pre-deployment and in self-coordinating and self-sustaining approaches for post-deployment.

On the other hand, several research fields are rising as solution frameworks for such difficulties. We build on these emerging fields, developing innovative architectures and algorithms thereof to address these challenges. More specifically, as seen in Fig. 1.1, we consider:

- **Part I.A** - Cognitive Radio Ad Hoc Networks (CRAHN) and Cognitive Radio Sensor Networks (CRSN) for IoT, which can remedy the spectrum scarcity problem in an energy-efficient manner,
- **Part I.B** - Self-Sustaining Networks for IoT with different functionalities, such as:
 - Innovative Energy Harvesting (EH) methods that enable self-sustainable operations,
 - Self-optimizing terafemto architectures, which leverage vast vacant low THz spectrum to achieve desired data rates on the order of 10 Gbps.

In **Part II**, we consider architectures that will enable realization of IoBNT through nanonetworking. Field of nanotechnology envisages development of very small scale entities, called nanomachines, that are able to carry out tasks such as communication, sensing and computation. Capabilities of a nanomachine on its own is very limited. Nanonetworking aims to overcome these limitations, by defining a set of rules for communication and collaboration between a collection of nanomachines, so as to achieve bigger tasks. Different communication types, namely; acoustic, electromagnetic, nanomechanical and molecular communications could be viable alternatives nanonetworking. However, acoustic and electromagnetic communications require entities that can carry out acoustic or electromagnetic operations at nanoscale. On the other hand, nanomechanical communications require direct contact between communicating parties and is not suitable for distant applications. In this context, we focused on Molecular Communications (MC) in this thesis, as it is the most viable method for nanonetworking, backed by the fact that similar mechanisms are already present in

many living organisms in nature. To that end, we develop energy-efficient modulation schemes for MC, that would efficiently utilize the spectrum in consideration, i.e. an aqueous medium, for nanonetworking.

1.2 Contributions

The contributions of this thesis on energy and spectrum-efficient communication techniques in next-generation IoT is illustrated in Fig. 1.2. Accordingly, 7 novel research studies have been incorporated into this thesis across 2 main domains in consideration. Highlights on key contributions are listed below:

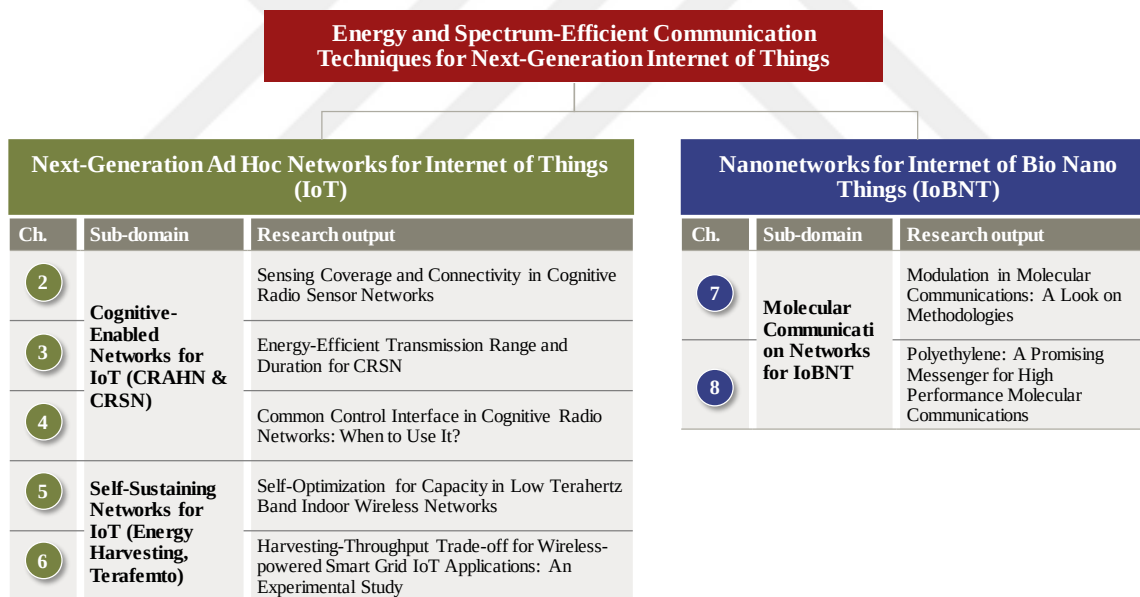


Figure 1.2: Contributions of the thesis.

Part I: Next-Generation Ad Hoc Networks for Internet of Things

- **Chapter 2:** In depth review of existing sensing coverage and communication connectivity assessment models from ad hoc and sensor networks, their applicability to CRSN, with a detailed recommendation on a spatio-temporal distortion based joint sensing coverage and connectivity model for CRSN. I have lead-authored a book chapter, published by IGI, with this contribution.

- **Chapter 3:** Revealing of joint energy-efficient transmission range and duration for CRSN for the first time in the literature. I have lead-authored a journal article, submitted to and currently under review in IEEE Transactions on Wireless Communications, with this contribution.
- **Chapter 4:** Impact of Common Control Interface (CCI) incorporation into CRAHN and CRSN for spectrum handoff, in terms of achievable delay, energy consumption, spectrum utilization and event estimation performance. I have co-authored a journal article, published in IEEE Transactions on Wireless Communications, with this contribution.
- **Chapter 5:** Development of a self-optimization protocol for capacity in realities of the low THz band for the first time in the literature, assessment of its complexity vs. utility in terms of additional capacity in low THz environment. I have lead-authored a journal article, submitted to and currently under review in IEEE Wireless Communications Letters, with this contribution.
- **Chapter 6:** Investigation of actual harvestable energy through E-field Energy Harvesting (EFEH) to be used in an IoT application setting, optimization of achievable throughput with a real-life “harvest-then-transmit” type system model. I have lead-authored a conference paper, submitted to and currently under review in IEEE International Conference on Communications, with this contribution.

Part II: Nanonetworks for Internet of Bio-Nano Things

- **Chapter 7:** Thorough assessment of all existing modulation schemes in molecular communications, their categorization and comparison, discussions on their synchronization requirements and their bio-compatibility for applications inside human body. I have lead-authored a book chapter, published by Springer, with this contribution.

- **Chapter 8:** Development of a novel modulation scheme using polyethylene molecules as the messenger, modeling of this approach as an erasure channel model, formulation of the channel capacity and its optimization with respect to critical network parameters. I have co-authored a conference paper, published in IEEE International Black Sea Conference on Communications and Networking with this contribution.

1.3 Detailed Thesis Outline

The rest of this thesis is organized as follows:

Part I: Next-Generation Ad Hoc Networks for Internet of Things

Chapter 2: Sensing Coverage and Connectivity in Cognitive Radio Sensor Networks

- Sensing coverage of a field of interest and connectivity are two very important performance measures in Wireless Sensor Networks (WSNs). Existing design methodologies and protocols for enhanced field sensing coverage and connectivity in WSNs are not directly applicable to Cognitive Radio Sensor Networks (CRSNs) due to their cognitive nature. In this chapter, we first review sensing coverage and connectivity models for traditional WSNs. Then, we propose novel approaches for sensing coverage and connectivity establishment in CRSN, benefiting from useful existing models from WSN and Cognitive Radio Ad Hoc Networks (CRAHNs). Proposed approaches span a wide variety of CRSN requirements, and also point out open research problems in the field to guarantee sufficient sensing coverage quality and connectivity in CRSN. Given its suitability for CRSN context, a spatio-temporal distortion based detailed model has been proposed to jointly assess sensing coverage and connectivity in CRSN.

Chapter 3: Energy-Efficient Transmission Range and Duration for CRSN

- Cognitive Radio (CR) promises more efficient utilization of the radio spectrum by enabling dynamic spectrum access. Cognitive Radio Sensor Networks

(CRSN) are Wireless Sensor Networks (WSN) equipped with CR capabilities. CRSN nodes need to operate energy-efficiently to extend network lifetime. In this chapter, for the first time in literature, we formulate the problem of finding energy-efficient transmission range and transmission duration for CRSN deployment that would minimize the energy consumed per goodput per meter towards the sink in a greedy forwarding scenario. Results reveal non-trivial relations for energy-efficient CRSN transmission range and duration as a function of 9 critical network parameters. These relations provide valuable insights for detailed CRSN design prior to deployment.

Chapter 4: Common Control Interface in Cognitive Radio Networks: When to Use It?

- To perform spectrum handoff, Cognitive Radio (CR) nodes communicating with each other need to exchange licensed user detection information, i.e., perform spectrum coordination, over a common control channel. The spectrum coordination can be fulfilled either via existing cognitive radio interface with time division or via a separate dedicated radio, i.e., a Common Control Interface (CCI), continuously. CR nodes with CCI can instantly exchange licensed user detection information and cease frame transmission, while spectrum coordination can only be performed after the frame transmission period without CCI. Nevertheless, the impact of CCI incorporation into CR nodes in terms of common performance metrics must be thoroughly assessed to evaluate the worthiness of additional radio cost. In this chapter, an analytical framework is presented to assess the impact of CCI incorporation into CR nodes for spectrum handoff. The developed framework enables analyzing potential benefits and disadvantages of employing CCI in terms of achievable delay, energy consumption, spectrum utilization and event estimation performance. Extensive performance evaluations are presented to illustrate the impact of CCI utilization on efficiency of spectrum handoff. The network and communication regimes that would yield

having CCI favorable are characterized in terms of spectrum conditions and CR parameters.

Chapter 5: Self-Optimization for Capacity in Low Terahertz Band Indoor Wireless Networks

- Achieving data rates required by 5G applications on the order of 10 Gbps is very challenging within the highly congested sub 6 gigahertz (GHz) bands. With its wide vacant bandwidth, low terahertz (THz) band could be suitable for such applications. In this chapter, we consider an indoor wireless network with Full Dimensional MIMO (FD-MIMO) capabilities, operating in the low THz band with a realistic and highly directional antenna radiation pattern. We propose a self-optimization protocol, called Approximate Water Filling (AWF), which enables the Base Station (BS) of a cell to allocate optimal power to its antennas, in order to optimize network downlink capacity. Our proposed protocol is shown to bring 20%+ capacity gains when the number of receivers are in the range of $10 \leq N \leq 30$. However, protocol run time increases exponentially with N . Total run-time is acceptable as per potential 5G applications requirements when $N \leq 30$.

Chapter 6: Harvesting-Throughput Trade-off for Wireless-powered Smart Grid IoT Applications: An Experimental Study

- Sensor nodes, one of the most crucial elements of Internet of Things (IoT), sense the environment and send their observations to a remote Access Point (AP). One drawback of sensor nodes in an IoT setting is their limited battery supply. Hereby, Energy Harvesting (EH) stands as a promising solution to reduce or even completely eliminate lifetime constraints of sensors with exploitation of available resources. In this chapter, we propose an Electric-Field EH (EFEH) method to enable battery-less execution of sensor-based IoT services for Smart Grid (SG) context. For this purpose, for the first time in the literature, harvestable energy through EFEH method is investigated with a

transformer room experimental set-up. Our experiments reveal that 40 mJ of energy can be harvested in a period of 900 sec with the proposed EFEH method. Building on this energy profile, we define a throughput objective function θ for a “harvest-then-transmit” type system model, to shed light on the harvesting-throughput trade-off specific to IoT-assisted SG applications. Numerical results disclose non-trivial relationships between optimal harvesting period T_H , optimal transmission period T_T and critical network parameters such as node-AP hop distance, path loss exponent and minimum reporting frequency requirement.

Part II: Nanonetworks for Internet of Bio-Nano Things

Chapter 7: Modulation in Molecular Communications: A Look on Methodologies

- Nanonetworking is a recently proposed paradigm that aims to achieve collaboration between nanomachines to carry out complex tasks. Molecular communications has been the most vibrant area of research for nanonetworking, mostly because of its feasibility and existence of analogous communication schemes in nature. In molecular communications, two nanomachines communicate with each other via propagation of molecules from the transmitter to the receiver nanomachines through the medium they reside in. How and where to encode the message, i.e. modulation, plays a key role in molecular communications since it greatly affects the communication performance at nanoscale. To this end, in this chapter, we thoroughly examine the landscape of modulation in molecular communications, categorize the modulation schemes in molecular communications by methodology and discuss how convenient they are in terms of synchronization requirements in a nanoscale environment and their bio-compatibility for applications inside human body.

Chapter 8: Polyethylene: A Promising Messenger for High Performance Molecular Communications

- Molecular communications is a bio-inspired paradigm, proposed to communicate nanomachines via diffusion of molecules through an aqueous medium. The

type and structure of the molecules to be propagated bear great importance, since they directly affect the modulation structure and the overall performance of molecular communications. With this regard, in this chapter, we propose a messenger-based molecular communication model where information is encoded on the atoms of polyethylene molecules in the form of $CH_3(CHX)_nCH_2F$, where X is replaced either by an H or F atom, representing 0 and 1 bits, respectively. The encoded polyethylene molecules are released from the transmitter nanomachine, and their propagation towards the receiver is modeled as a Brownian Motion. Using an erasure channel model, our analysis focuses on formulating the capacity of this channel and revealing the parameters affecting it such as molecule size and number of redundant molecules for one transmission.

Part I

Energy and Spectrum-Efficient Communication Techniques in Next-Generation Ad Hoc and Sensor Networks for IoT

Chapter 2

SENSING COVERAGE AND CONNECTIVITY IN COGNITIVE RADIO SENSOR NETWORKS

2.1 Introduction

Recent advances in micro-electromechanical systems, signal processing and wireless communications have made it possible to manufacture low cost, low complexity wireless sensor nodes. These nodes, called wireless sensor nodes, are capable of sensing their environment, processing the sensed data if needed, and transmitting the sensing data toward the data fusion center (also known as the sink). A collection of wireless sensor nodes, which sense a field of interest and report their data to the sink are called a Wireless Sensor Network (WSN).

A WSN consist of a large number of nodes deployed in the area of interest. The locations of nodes within this area may be predetermined and engineered, or may be totally random. These deployed nodes are composed of four basic units: Sensing unit, processing unit, transceiver unit and power unit. A wireless sensor node monitors its environment thanks to its sensing unit. The transceiver allows a wireless sensor node to communicate its sensed and possibly processed data to other nodes. The power unit provides the power required for sensing, communication and processing tasks. The sensor node may decide to process the sensed data to obtain a useful piece of information and transmit it to the intended node or to the sink, instead of sending the raw sensing data. Application areas of WSNs cover a wide range of topics, including but not limited to military applications, environmental monitoring, health care and industrial applications. For instance, in a military setting, a WSN deployed as a thin strip to a country's border can serve to detect intruders, whereas a WSN deployed

over the large area of a forest can do periodic monitoring and provide early warning when fires outbreak. A wide range of applications, as well as factors affecting WSN design are thoroughly investigated in [1].

Two main challenges faced by WSNs are communication of sensing data to the sink in a timely manner, and to do this communication under the resource constraints, especially the limited power. WSN nodes need to carry out sensing, communication and data processing tasks under the constraint of limited battery power. Since WSN nodes are both data generators and forwarders of other nodes' data to the sink, communication related power consumption has to be kept at an optimal level during network design, for enhanced network lifetime. In WSN, nodes are communicating to each other via wireless medium. Although infrared and optical communications are also possible, WSN nodes usually benefit from wireless radios for communication. WSNs often operate on the Industrial, Scientific and Medical (ISM) bands for communication purposes. However, these bands are becoming fast overcrowded due to many networks of diverse set of applications going live on them. To this end, WSNs need to compete with other networks that operate on these bands for access to frequency medium, creating spectrum scarcity issues.

WSNs suffer from communication and constrained power source challenges in the following manner: In case of event detection within the field of WSN, nodes that detect the event consequently generate packets and race among them and also with other networks they are overlaid with for access to frequency spectrum. Due to collisions, nodes may have to retransmit their packets several times, causing excessive energy consumption. It may even be the case that packets related to the event may not make it to the sink in a time interval that would make their contents significant and valuable for consideration.

The spectrum scarcity problem is envisaged to be addressed by Dynamic Spectrum Access (DSA) schemes. Cognitive Radio (CR) is the key enabling technology to provide dynamic, i.e. opportunistic, spectrum access. With CR capabilities, communicating parties can sense the spectrum, determine the vacant spectrum bands and use

them for their operations. In that sense, CR devices can also use the licensed bands opportunistically in addition to unlicensed bands, hence ameliorating the spectrum inefficiency problem of traditional fixed spectrum access schemes. CR users, called Secondary Users (SUs) can use the licensed bands as long as they do not interfere with the transmission of licensed band users, i.e. Primary Users (PUs).

Governmental agencies regulate the spectrum usage and licenses. The fixed spectrum portions are assigned to the license holders over certain geographical regions. In this type of fixed spectrum assignment approach, none of the unlicensed users can utilize licensed bands. According to the Federal Communications Commission (FCC), these licensed bands are underutilized temporally and geographically. On the other hand, unlicensed users can access only unlicensed bands which are ISM bands. These bands are crowded due to excessive usage by unlicensed users. Cognitive radio technology can enable them to use licensed bands in an opportunistic manner [2]. Nodes with cognitive capability can have real time interactions with the environment. These interactions can be achieved with cognitive cycle operations through adaptations in a dynamic radio environment. General set of communication networks equipped with CR capabilities and cognitive cycle operations are called Cognitive Radio Networks (CRNs).

DSA necessitates three cognitive cycle operations, which are detailed below:

1. In ***spectrum sensing***, available spectrum bands are detected by monitoring the licensed spectrum bands. This sensing available white bands and their characteristics.
2. ***Spectrum decision*** ensures that, most appropriate vacant spectrum band is selected according to the communication requirements (e.g. data rate, link quality, etc.).
3. ***Spectrum handoff*** ensures that the SU node leaves the portion of the spectrum when a PU arrives on it and/or communication quality is not meeting

required standards, and restarts spectrum sensing to access another portion of the spectrum.

One of the most important requirements in opportunistic communication is limiting the interference to PUs. SU activity may interrupt the communication between PUs. Hence, SUs are expected to change their communication parameters to limit their interference to licensed users. Spectrum mobility is the ability to change operation frequency. Sudden appearance of PUs trigger spectrum handoff. Furthermore, spectrum sharing is another important challenge in cognitive radio networks. Spectrum sharing issues include the coordination of communication, medium access control and spectrum allocation among numerous unlicensed entities.

Cognitive Radio Sensor Networks (CRSNs) have been proposed as a new paradigm, envisaging WSNs equipped with CR capabilities, to overcome these spectrum scarcity problems [34]. Wireless sensor networks use ISM bands, and hence they suffer from the spectrum scarcity problem. CR capable wireless sensor nodes can dynamically access licensed bands and overcome the spectrum scarcity challenge. Employing CR capable sensor nodes brings additional advantages over WSN. Firstly, the event detection cases create bursty traffic, resulting in potentially multiple event detecting nodes racing to access the same spectrum band to send their event readings. This will create collisions and packet losses. However, accessing licensed channels opportunistically can decrease the probability of collisions. Secondly, adaptability to the radio environment decreases retransmission due to degradations in channel quality. These advantages of cognitive radio technology can result in power consumption reduction, which is highly desirable for resource constrained sensor nodes.

CRSN is a distributed event-based system such that event detecting nodes send their event readings towards the sink in multi-hop manner over available channels. Although the realization of CRSN brings mentioned advantages, it also poses significant challenges which are managing extra CR operations with resource constrained sensor nodes, additional processing and communication overhead due to cognitive capability and coordinating a spectrum-aware multi-hop communication over licensed

and unlicensed bands. In order to overcome these challenges, energy-efficient and spectrum-aware solutions should be considered. The existing communication protocols for Cognitive Radio Networks are not best fits for CRSN since they are not considering energy-efficiency from a sensor network perspective. The protocols for WSNs inherently do not have the capability for the adaption to the radio environment. These types of protocols do not fulfill the unique requirements of CRSNs.

Cognitive radio sensor nodes can have different architectures depending on their type of operation, such as totally ad-hoc, clustered, hierarchical or mobile architectures. Irrespective of architecture, the main duty of the cognitive radio sensor nodes is to perform sensing on the environment in which they are located. Hence, they can be applied for various application areas. Potential advantages of CRSN may be used in multimedia, indoor sensing, multi-class heterogeneous sensing and real-time surveillance applications.

Although WSN has been a widely studied field during the last decade, CRSN is a relatively new paradigm, with many additional work needed to be done until practical feasibility is achieved. Sensing coverage and connectivity are two of those important metrics for ensuring high quality operations in a sensor network, that have been heavily researched in traditional WSN but requiring new studies for CRSN applications.

Sensing coverage in the context of sensor networks can be broadly defined as the ability of the network to monitor a certain field or detect a certain event on that field. Depending on the application type, sensing coverage has numerous interpretations. For instance, it can be measured by the maximal breach path (worst case coverage, path with the maximal length that is also at the greatest distance from each sensor node at any given point on the path), by the maximal support path (best case coverage), by the ratio of the area covered with at least one node to the total field area, by the k -coverage ratio (ratio of the area covered with at least k sensor nodes to the total area), and many other metrics. Irrespective of assessment metric, sensing coverage is a Quality of Service (QoS) measure for a sensor network.

Connectivity, on the other hand, has a broadly accepted definition. Two nodes are topologically connected if they are within the communication range of each other. Communication range of a node is defined as the radius of the Boolean disc centered on the exact position of the node. In such Boolean models, the node is assumed to be able to communicate with any point within this disc, and not be able to communicate at all with any point outside of this disc. The term connectivity, on the other hand, can perhaps be best explained by an example. Let two nodes A and B have a distance R_d between them, and have communication radii of R_a and R_b , respectively. As shown in Figure 2.1(a), when $R_a > R_d > R_b$, the connectivity is unidirectional and is from A to B . Figure 2.1(b) depicts the bilateral connectivity. When both $R_a > R_d$ and $R_b > R_d$, the connectivity between these two nodes is bilateral since they can theoretically reach out to each other. Although connectivity, in the topological sense, is based on this Boolean model, actually connectivity between nodes is also affected by channel conditions.

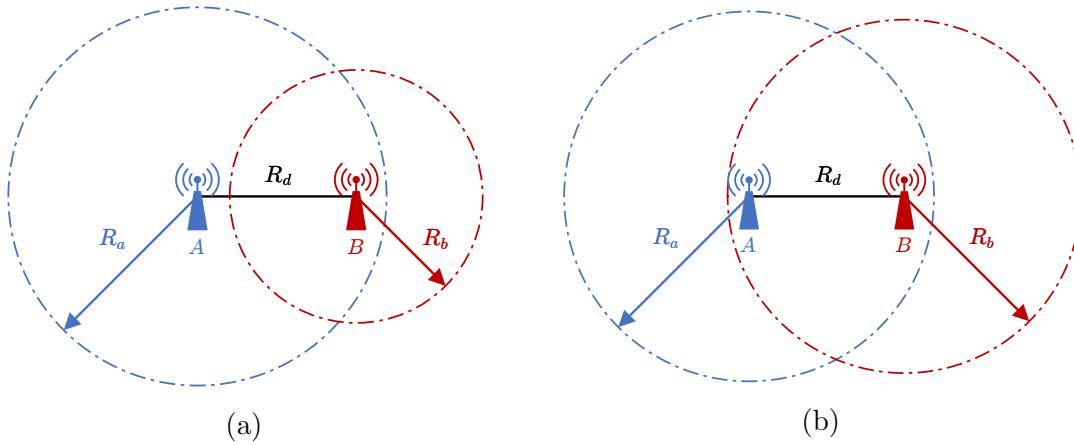


Figure 2.1: (a) Unidirectional link from node A to node B ($R_a > R_d > R_b$); (b) Bidirectional link between node A and node B ($R_a > R_d$ and $R_b > R_d$).

Although the definition of connectivity is highly consolidated, how it is assessed for random networks varies greatly. One possibility is to assess the probability that a node is isolated (i.e. a node has no neighbors within its communication range) and to approximate the probability that all nodes in a network are connected to each other

by the probability that no node in the network is isolated. This approximation clearly misses the probability of having no isolated nodes but having isolated clusters containing two nodes or more, but still can be a very good approximation for sufficiently high node densities. Another alternative is to employ percolation theory, and to assess above which node density an infinitely large connected component exists with high probability. Although sensing coverage and connectivity have been studied for WSNs, unique challenges to satisfy them in the CRSN context call for additional studies. Inherently, cognitive radio technology changes the topological connectivity conditions of two neighbor nodes to being in the communication range of each other and having a common available channel for communication. This new definition brings in additional challenges in assessing connectivity for CRSN, but also has implications on the quality of the sensed data and how it changes until it is conveyed to the sink.

The rest of this chapter is organized as follows. In Section 2.2, motivations to carry out research in sensing coverage and connectivity in CRSN are laid out. Section 2.3 categorizes and extensively reviews the related work on sensing coverage in WSNs, while Section 2.4 summarizes the literature on connectivity. Building on these foundations, we recommend a joint sensing coverage and connectivity model and analytically characterize it in detail in Section 2.5. Finally, open research directions in sensing coverage and connectivity and CRSN are revealed in Section 2.6.

2.2 Motivation for Sensing Coverage and Connectivity Research in CRSN

Sensing coverage and connectivity are mutually inclusive issues in sensor networks, since they are two vital requirements to be fulfilled concurrently for communication between sensor nodes [3]. Sensing coverage area of a sensor network is usually defined as ratio of the area that can be reliably monitored by sensor nodes to the total area of interest. Monitoring the area drives packet generation by sensor nodes, either periodically or upon event detection. Packets generated by sensor nodes can be relayed to the sink if the network is connected. If the connectivity among sensor nodes is not achieved, the sensing coverage and hence the sensed data are not available to the sink.

Even if the packets can be delivered to the sink under the condition that the network is connected, the delivery time of the packets could be crucial for the sensor network. Under such conditions, the sensing coverage quality is affected by both the spatial sufficiency of sensor nodes to cover the area of interest and by the timely delivery of field sensing related packets to the sink. Consequently, metrics on sensing quality are more insightful for assessing sensing coverage.

Apart from WSN challenges, connectivity and hence sensing coverage quality are affected by activities of licensed users in CRSN. Hence, these performance measures should be investigated in the context of dynamic radio environment. Sensing coverage should be engineered in terms of not only node placement and node scheduling for traditional coverage, but also spectrum availability between active sensor nodes. Without such considerations, due to the PU channel usage pattern, sensed data may not be relayed to the sink timely manner. The packets may become obsolete and the energy for relaying and sensing could be wasted.

Another perspective on how sensing coverage and its quality and connectivity are interlinked in CRSN is coming from energy-efficiency considerations. CRSN nodes carry out field sensing to detect events. In addition to that, they perform spectrum sensing to determine available spectrum bands and to utilize them for communication towards the sink. To extend the lifetime of CRSN, field sensing, communication and spectrum sensing devices in a node may preferably go to sleep mode periodically. The ON/OFF duty cycles of transceiver (including spectrum sensing unit) and the sensor will ultimately determine the lifetime of the CRSN. In some settings, spatio-temporal variation of the PU communication can be low, hence less energy would be allocated to spectrum sensing event (since now less spectrum sensing operations are required) and more energy could be allocated to field sensing for the sensor device. This would ultimately enable the sensor to monitor its area with a higher duty-cycle, or would increase network lifetime if the duty cycle of the sensor is not to be varied. The other point is that, due to the different ON/OFF duty cycling rates of sensor and the transceiver in energy-efficient applications, the transceiver might go to sleep by

the time the sensor generates a data packet, hence would have to wait for the next cycle of the transceiver to start transmitting the packet. Since the quality of sensing coverage is not only about covering the whole region but also about conveying the packets from the field to be monitored to the sink, packets not reaching the sink on time become obsolete and degrade sensing coverage.

In the light of the above discussions, the literature on sensing coverage and connectivity in WSNs and other wireless ad hoc networks cannot be directly applied to CRSN for several reasons, which can be listed as follows:

- In WSNs, a network graph is connected with high probability if node density is higher than critical node density, namely, λ_c . In this case, all nodes have a communication path to each other with high probability. CRSN nodes communicate opportunistically, unlike WSN nodes. Hence, even if there is a topological path between two CRSN nodes, due to non-availability of channels on this path, they may not have actual connectivity. Even if there is actual connectivity, according to the cognitive communication type (underlay or overlay), routing approach (i.e. greedy forwarding, nearest neighbor forwarding, or else) and communication radii, CRSN communication links would bear varying PU traffic. Since CRSN nodes have to abide by the rules that limit their interference on PUs, all these factors affect the delay of transmitting sensing information, which in turn would have significant implications on the sensing quality due to timeliness of gathered data by the sink.
- Numerous works on coverage and connectivity in WSN also include aspects for energy-efficient operation. Measures such as duty cycling, random probing, configuration for ensured 1-coverage with minimal energy consumption and alike have been proposed to this end. However, these existing protocols for energy-efficiency do not take into account the fact that some nodes in the network may be unreachable due to the DSA nature of CRSN, and hence are not directly applicable. Besides, literature on connectivity of Cognitive Radio Ad Hoc Net-

works (CRAHN) does not consider energy-efficiency as a constraint and hence cannot be applied to sensor networks.

- Many works on coverage and connectivity in WSN assume a Boolean disc model for coverage and connectivity computations on a geometric graph where the area inside the sensing range disc is perfectly covered (for sensing coverage consideration) and nodes inside the communication range are assumed to be perfectly connected (for connectivity consideration). Although useful and simple, these models are not realistic for CRSN. What is critical for CRSN is the joint distortion on the sensed data, both in space and time, since it ultimately dictates the sensing coverage quality. The joint distortion is important for WSN, however, CRSN has unique properties which make joint distortion model different from existing literature such as [4]. To this end, Voronoi cells and exposure problem are more suitable to CRSN than traditional Boolean disc model. Reliability of communication must also be considered for a realistic model while studying connectivity.

The following sections aim to extensively review the literature on sensing coverage and connectivity, point out which works can be adopted in CRSN works for ensuring sensing coverage and connectivity and emphasize the remaining open issues for more detailed analysis of sensing coverage and connectivity in CRSN.

2.3 Sensing Coverage in Ad Hoc Networks

In this section, we review the significant works on sensing coverage in wireless networks. We categorize them according to the fundamental problems they address and point out how they can be adopted for CRSN. These works investigate the sensing coverage problem from different perspectives, such as achieving energy-efficient sensing coverage, deployment strategies for cost efficient coverage, node scheduling, lifetime trade-offs and k -coverage. Despite the vast amount of work on WSN sensing coverage, intermittent connectivity nature of CRSN call for additional research to assess how it

could affect the sensing coverage and its quality. Sensing coverage and connectivity is highly coupled in CRSN since connectivity challenges can hamper and in some cases even totally prevent the transmission of sensing data to the sink, where it is ultimately intended to be sent. Hence, although the main sensing coverage functions are similar in WSN and CRSN, intermittent connectivity in CRSN has important implications on how the quality of sensed data is affected, until it can hopefully be conveyed to the sink. To this end, spatial and temporal variations of spectrum channels, spectrum mobility and spectrum management issues should be taken into account to propose sensing coverage solutions for CRSN.

Sensing coverage studies in the literature mainly focus on the reduction of the deployed active nodes via scheduling, redundancy and lifetime trade-offs, and duty cycle and lifetime trade-offs. Ultimate goal is to provide sufficient QoS in addition to extending network lifetime. Below, existing works are presented and discussed in detail based on the type of the problems they target to solve.

2.3.1 Energy-Efficient Sensing Coverage by Node Scheduling in WSN

Sensor nodes are deployed with high density. Hence, select nodes may become redundant in the operation of sensor networks, as long as the application requirements are met without them. Keeping all sensor nodes active drains the battery sources earlier. Redundancy and scheduling of the nodes must be implemented to achieve higher lifetime. Hence, many studies focus on finding redundant nodes and organizing scheduling accordingly. [5] aims to have longer lifetime with proper sensing coverage in large-scale, energy-constrained and unreliable sensor networks. The energy consumption can be reduced by turning off the redundant nodes in the network without harming the sensing coverage area. Nodes are scheduled to be turned off under certain circumstances to conserve energy by a self-configuring and distributed protocol. The protocol's purpose is to preserve the initial sensing coverage area while turning off maximum number of nodes possible. Since it is an extension of the clustering protocol LEACH, it also employs data aggregation and caching for energy-efficiency.

Such a node scheduling problem is divided into two sub-problems which are:

- Determination of nodes to be turned off and to be kept on.
- Determination of the scheduling of this activity pattern.

To be a better fit for CRSN applications, modifications on this protocol, such as taking also into account the spectrum availability of each node and its close neighbors while deciding on which nodes to turn off and scheduling their active periods according to PU activity, could be needed. Such modifications would help to achieve sensing coverage with energy-efficiency in CRSN.

Selecting the redundant nodes for energy-efficiency is an NP-complete problem. To able to solve the selection problem in polynomial time, heuristic algorithms must be offered. [6] offers a heuristic scheme that determines the exclusive sets of sensor nodes to completely cover the monitored area. Only one set is active at a time among others. Using the subset of sensor nodes provides energy-efficiency in addition to full coverage preservation. Since the number of exclusive sets is related to the network energy-efficiency, the heuristic scheme maximizes this number. Exclusive node set may provide sensing coverage over the monitoring region, however intermittent communication in CRSN is an obstacle for this protocol. To overcome those difficulties, the coverage can be guaranteed via utilizing more than one exclusive set. This modification would reduce energy-efficiency at the expense of more opportunities in spectrum access, with the condition that a MAC protocol runs efficiently to manage contentions.

Applying heuristic schemes do not guarantee the optimal sensing coverage quality. Therefore, there are optimization efforts on node scheduling for providing sensing coverage. [7] uses Integer Linear Programming (ILP) for optimal and heuristic algorithm for sub-optimal solutions for sensing coverage problem. Main motivation is the coverage with power-aware operation. It presents the coverage of the sensor networks with minimum nodes as an NP-hard problem. When ILP formulations are solved, they provide certain level of event sensing coverage with minimal subset of

sensor nodes. [6] and [7] have solved the sensing coverage problem by selecting subset of nodes in different time slices. However, the PU traffic is not considered in these papers since they are targeting WSN. By redefining sensing coverage metrics according to spectrum availability and taking into account PU traffic, these works can be leveraged for CRSN.

2.3.2 Sensing Coverage Quality and Network Lifetime Trade-off in WSN

Sensing coverage problem is also being studied in the context of coverage level and network lifetime trade-offs. The lifetime maximization is one of the key research areas in sensor networks due to their energy-constrained nature. One of the existing works, [8], proposes pCover that provides partial coverage but higher lifetime compared with the protocols providing full coverage. Full coverage brings better surveillance; however, partial coverage results in higher coverage duration. This study is mainly focused on coverage level-lifetime trade-off rather than partial coverage protocol presentation. The proposed protocol organizes sleep-wake schedule of the nodes while providing some degree of coverage. Although [8] considers the challenges posed by sensor networks, once again taking into account the spatial and temporal variation of channel availability is needed as an additional constraint to achieve meaningful sleep-wake schedules in CRSN and coverage level-lifetime trade-off.

Sensing coverage and connectivity are coupled since sensed data has to reach the sink with acceptable delays. Finding subset of nodes that provides partial coverage with established connectivity between them achieves certain coverage level and network connectivity beside energy-efficiency [9]. This problem is referred to as the connected, partial coverage problem. Since this problem is NP-hard, a heuristic algorithm is proposed for the connectivity and the partial coverage simultaneously. In some applications, the full coverage is not necessary since the sensed data may be satisfactory for estimating the gathered data at the sink.

2.3.3 Sensing Coverage Quality in Randomly Deployed WSN

Boolean disc model represents coverage area as a disk in the center of which the sensor node is located. Under this model, sensing coverage quality is another topic to be investigated to provide higher detection resolution. If a region is observed by a higher number of nodes, the spatial resolution of sensing in this region increases. Hence, studies in this direction investigate k -coverage. If a region is covered by k nodes, then it is said to be k -covered. [10] investigates the relationship between the probability of k -coverage and the sensing radius and the number of nodes in sensor network where the nodes are distributed according to Poisson Point Process (PPP) and uniform point process. Unlike the existing sensing coverage literature, this work considers boundary effects, which become significant when the sensing radius of nodes is not very small compared to the area to be monitored. In such a case, each node may not be able to utilize its full potential sensing area. A highly beneficial feature of the work in [10] is the study on the challenge of the boundary effect.

The sensing coverage quality can be quantified by the asymptotic behavior of sensor networks. The determination of asymptotic behavior of sensing coverage is a challenging research topic in large-scale sensor networks. [11] investigates the sensing coverage area and the fraction of nodes and coverage under Boolean and other general sensing models. Furthermore, the detectability of intruding objects in sensor networks is another sensing performance metric in [11]. Percolation theory has been employed in this work to introduce the critical node density above which there is no path across the network without being detected by any sensor. However, implications of intermittent connectivity on both area coverage and intruder detection must be included for further similar studies in CRSN.

Graph theory and computational geometry are also used to increase the sensing coverage quality. [12] employs graph theory and computational geometry for proposing an optimal polynomial time algorithm that solves for the best and worst coverage cases. For best and worst coverage, this work employs maximal breach path algorithm and maximal support algorithm, respectively. Voronoi diagrams and Delaunay

triangulation are used for these algorithms. Voronoi diagrams can also be very useful for coverage assessment in CRSN, and in the following recommended CRSN sensing coverage and connectivity toolkit, we propose to use them in a coverage distortion minimization setting.

2.3.4 Node Density for Energy-Efficient Monitoring in WSN

The main task of a sensor network is to monitor the field of interest. Energy constraint of sensor nodes necessitates the energy-efficient sensing coverage for the detection of events. [13] considers the energy-constrained distributed detection in wireless multi-hop wireless networks. In this problem, the node density of a randomly distributed network with a PPP is under consideration. The main objective is to find the optimal node density λ^* that minimizes the Neyman-Pearson error exponent D_λ under a constraint E on the average per node energy consumption $C(\lambda)$, assuming a spatially correlated measurement model at sensor nodes. This problem has been formulated as:

$$\begin{aligned} \Lambda^* = & \underset{\lambda}{\operatorname{argmax}} D_\lambda \\ \text{sub. to } & C(\lambda) < E \end{aligned} \tag{2.1}$$

Another energy-efficient field monitoring study, [4], considers the spatio-temporal correlation between the measurement data generated by the wireless nodes. Objective of the sensor network here the estimation of the whole data in the network with minimum distortion at sink. According to delay and energy-efficiency, the effect of node density on estimation distortion is investigated. Measured data has the aspects of spatial and temporal distortion due to the deployment and multi-hopping, respectively. The optimal number of nodes in the network exists, which minimizes total spatio-temporal distortion for the entire data gathered in the network. However, in dynamic access networks such as CRSN, nodes also suffer from the delay caused by PU activities. Time distortion model in [4] cannot be directly applied to CRSN since the communication links are subject to PU activities, hence adjustments and

additional models are needed to capture those effects.

2.3.5 Deployment Strategies for Sensing Coverage in WSN

Deployment strategies are also important for coverage in sensor networks. However, they must be reconsidered if they are applied to CRSN since the sensing coverage quality is affected by the PU activities, which can hamper timely delivery of the detection information from sensor nodes to the sink. PU activities may even restrict some CRSN nodes altogether from relaying the sensed data towards the sink. The existing literature on deployment strategies does not consider this effect. [14] investigates the level of target detection with the cost of the deployment. Performance of the detection is dependent on the deployment since local observations of the sensors rely on their positions. Path exposure, the likelihood of detection of an object while traversing the network, is a performance measure of deployment. It finds the number of nodes that need to be deployed for target detection considering the trade-offs between network performances, the cost of the sensors deployed, and the cost of deploying sensors. In the cost strategy, a number of nodes are deployed at a time until achieving minimum exposure with cost function depending on the sensor deployment and the deployment strategy.

Sensing coverage can be increased by employing mobile nodes in the sensor networks as a new deployment strategy. [15] proposes the redeployment of the mobile nodes for covering the holes in the network according to a particle swarm optimization. It employs robots as mobile wireless sensor nodes to fix the broken part of the network by locating them. A modified particle swarm optimization, called Particle Swarm Genetic Optimization (PSGO), is proposed to employ mobile robots in the target detection region by changing the node density. This technique finds the position of coverage holes. However, the temporal activities of PUs necessitate consideration, and could lead to a case where there is a coverage hole but unlicensed communication at this coverage hole is not likely due to heavy PU traffic, therefore mobile sensor dispatching to this area would not be beneficial.

2.3.6 k -coverage in WSN

The reliability of event detection is higher if they are detected by more than one node. The reliability improvement requires that the network should be covered by at least one sensor node. As a coverage performance metric, [16] investigates the expected area in the network that is covered by at least k nodes. This area is said to be k -covered. It considers effects of the stochastic position of the nodes and the border effects on the deployment region. A higher degree of coverage means higher QoS and elevated accuracy and reliability. [16] also studies the scheduling problem of WSN nodes to achieve energy-efficiency without affecting the QoS in the context of network coordination.

On the other hand, node failure probabilities must be considered for k -coverage, since nodes are error prone. [17] investigates the number of nodes that need to be deployed in the network, the failure probability of nodes, and their sensing range in order to cover the whole field of interest by at least k nodes. k -coverage is studied under different deployment strategies such as grid, random, and Poisson. The nodes are in sleep mode with probability $(1 - p)$ and the sensors sleep most of their lifetime. The critical conditions for k -coverage are derived for the deployment strategies. On the other hand, the sleeping schedules of the nodes in CRSN should be consistent with PU activities for proper operation.

2.3.7 Coverage in Cognitive Radio Networks

Coverage in CRNs has been inherently studied from the perspective of communication coverage, which is affected by primary network activities as well as communication range and deployment of nodes. [18] tries to maximize the coverage area of a CRN while the interference to the PUs is under a tolerable threshold. They consider a simple network consisting of one primary radio link and one secondary radio link. Secondary transmitter and secondary receiver constitute a simple cognitive radio network in this setting. A secondary user can access the spectrum if the interference is limited to a predefined threshold. In this network setup, upper bound on the dis-

tance between secondary transmitter and secondary receiver relation is investigated with respect to distance between secondary transmitter and primary receiver, interference level at the primary receiver, and workable threshold at cognitive radio network. This study considers only one cognitive radio communication link. However, in CRSN there are inherently many nodes and communication links. Additionally, CRSN nodes have stricter energy constraints compared to CRN entities, requiring a revisit of these schemes with an efficiency mindset.

The coverage in wireless networks is mainly studied in the context of node scheduling, lifetime trade-off and coverage quality. Most of the existing studies considers the fixed frequency assignment approach and does not take into account DSA schemes. The proposed coverage schemes are energy-efficient which fulfill the requirement of energy-constrained nature of sensor networks. Table 2.1 outlines the existing research studies on sensing coverage.

Table 2.1: Key related work on sensing coverage in wireless sensor and ad hoc networks.

Problem type	Related Work	Main topics covered	CR ¹	EE ²	SD ³
Energy-efficient sensing coverage by node scheduling in WSN	[5], [6], [7]	Schemes which turn off redundant nodes, while preserving sensing coverage under Boolean models		✓	
Sensing coverage quality and life-time trade-off in WSN	[8], [9]	Studies on the tradeoff between level of sensing coverage and expected network lifetime		✓	
Sensing coverage quality in randomly deployed WSN	[10], [11], [12]	Assessments on coverage with Boolean models, maximal breach and support paths and Voronoi cells in random deployments			✓
Node density for energy-efficient monitoring in WSN	[13], [4]	Minimization of spatio-temporal distortion on sensing data by adjusting node density under energy constraints		✓	✓
Deployment strategies for sensing coverage in WSN	[14], [15]	Node density adjustment by location for better coverage, mobile nodes, deterministic vs. random deployment			✓
k -coverage in WSN	[16], [17]	Studies on percentage of area covered by k sensors, with respect to node density, duty cycle and other parameters			
Coverage in Cognitive Radio Networks	[18]	Achieving secondary network coverage while maintaining a minimum interference to primary network	✓		

Legend 1: Cognitive Radio consideration || 2: Energy-efficiency consideration || 3: Sensing distortion metric consideration

2.4 Connectivity in Ad Hoc Networks

CRSNs are most often ad hoc networks, i.e. networks with a flat hierarchy of only regular sensor nodes without any supporting infrastructure, or at most a clustered network composed of only cluster heads and regular sensor nodes. Topological connectivity of an ad hoc network is established if any node is connected to every other node in the network, either in a single hop or through multiple hops. How connectivity is assessed in wireless sensor networks can be categorized under three broad groups, as follows:

1. Connectivity derived from node isolation probability, under Boolean communication range of a node.
2. Connectivity derived from node isolation probability, under realistic channel models (e.g. with fading).
3. Connectivity derived from percolation theory and asymptotic behavior.

2.4.1 Connectivity Derived from Node Isolation Probability, Boolean Model

Deriving connectivity from node isolation probability is simple yet could be very insightful depending on the node density of the network to be designed. Assuming a random deployment of the CRSN and that deployment is carried out as a homogeneous PPP of density λ , the isolation probability is the likelihood that a given node has no neighbor within its communication range R_c and is formulated as in Eq. (2.2).

$$P_{iso} = e^{-\lambda\pi R_c^2} \quad (2.2)$$

In this simplistic model, a wireless node's topological connectivity area is Boolean. The node is assumed to be topologically connected to any other node that is within a distance R_c , and is certainly not directly topologically connected to nodes outside of this zone, although it can be connected to such nodes through other intermediary nodes.

The expression in Eq. (2.2) for arbitrary deployment schemes has been rigorously derived by [19], by first considering a network of n nodes, then defining the probability that a given node at location $\vec{x}$ has d neighbors under the binomial distribution as

$$P(D = d|\vec{x}) = \binom{n-1}{d} (P_0(\vec{x}))^d (1 - P_0(\vec{x}))^{n-1-d} \quad (2.3)$$

where $P_0(\vec{x})$ is the probability that a second node is placed within a radius R_c of the given node with location $\vec{x}$, under a general deployment probability density function $f_{\vec{x}}(\vec{x})$ applied to each new node to be deployed. As governed by the Poisson Limit Theorem, when n is large and probability $P_0(\vec{x})$ in the binomial distribution is small such that $nP_0(\vec{x}) \rightarrow \mu_0(\vec{x})$, the isolation probability can be approximated as the Poisson expression in Eq. (2.4).

$$P_{iso} = P(D = 0|\vec{x}) = e^{-\mu_0(\vec{x})} \quad (2.4)$$

Getting from node isolation probability to network connectivity probability P_{con} is through Eq. (2.5).

$$P_{con} \approx (1 - P_{iso})^n \quad (2.5)$$

Eq. (2.5) is tighter when P_{iso} is closer to 0. The network connectivity probability P_{con} shows a transition behavior, i.e.

- Given a fixed node density λ , below a critical communication range $R_{c,crit}$, the network is connected with a very low probability; and above $R_{c,crit}$, the network is connected with a very high probability.
- Similarly, given a fixed communication range R_c , below a critical node density λ_{crit} , the network is connected with a very low probability; and above λ_{crit} , the network is connected with a very high probability.

Another valuable work on the connectivity in networks with finite number of nodes

was carried out in [20]. Authors begin by considering a one-dimensional network over a length z . A total of n nodes, each having a communication range r , are distributed uniformly over the strip of length z . Exact connectivity probability of this one dimensional network is then found to be

$$P_{1,con} = \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{(z - kr)^n}{z^n} Heaviside(z - kr) \quad (2.6)$$

and this result is extended to the two dimensional networks with bounds as $P_{2,no\ con} > (1 - P_{1,con})^2$ and $P_{2,con} < P_{1,con}^2$.

To relax the assumption of homogeneous communication range, [21] uses a case study where two types of nodes are employed. One type of deployed nodes are with communication range r_1 , and the other type are with communication range r_2 . n_1 number of nodes for type r_1 and n_2 number of nodes for type r_2 are assumed to be deployed. Two nodes of different types are assumed connected if they have a bidirectional link only (i.e. they are both in the range of each other). Assuming $n_1 = n_2$ first, authors reveal the following:

1. Both node types have their own certain threshold range below which they become disconnected, i.e. isolated nodes or node clusters appear.
2. If the communication range for one of the types of the nodes is below its critical range to be connected but the other node type's communication range is above its critical range, the overall connectivity of the network is dictated by the node type that has a communication range below its critical range.
3. Increasing the communication range for one type of node helps only with a limited extent if the other type of node has a communication range below its critical range.

Relaxing $n_1 = n_2 = n/2$ assumption to $n_1 = n - n_2$, it is observed that to ensure connectivity, both node types still need to fulfill a critical communication range, but

now the critical communication range for the node type with higher number of nodes is bigger than the type with lower number of nodes.

[19], [20] and [21] provide approximations to connectivity probabilities in two dimensional ad hoc flat networks and are applicable to CRSN in a topological sense, i.e. they can provide insights on whether CRSN nodes have topological connectivity. Nevertheless, for actual connectivity, two CRSN nodes also need to have a common available channel, in addition to being in the communication range of each other, which is not a considered fact in these works for WSN.

2.4.2 Connectivity Derived from Node Isolation Probability, Under Fading

Connectivity studies in Section 2.4.1 depend on the Boolean communication model, where two nodes are assumed to be able to communicate without any errors if they are within each other's maximum communication range, and are not able to communicate at all if they are not within that range. Unlike those works, [22] considers the effect of fading on connectivity. To take fading effect into account, authors first define a metric called *Power Attenuation*, which is simply the ratio of received power to transmitted power. By assessing this random variable, they define an *Effective Communication Range* R . The effective node degree is then $E\{D\} = \lambda\pi E\{R^2\}$ and the node isolation probability under fading is then

$$P_{iso,fading} = e^{-E\{D\}}. \quad (2.7)$$

Results in [22] show that

- Fading increases the probability of node isolation $P_{iso,fading}$ when $P_{iso,boolean}$ for no fading case is above 0.65 and vice versa,
- Yet interestingly, fading increases the path probability, as a measure of overall network connectivity, for all node density values.

2.4.3 Connectivity Derived From Asymptotic Behavior

Seminal work on the assessment of connectivity by analyzing asymptotic behavior is without doubt [23]. Authors have shown that, assuming a homogeneous range assignment for each node, n nodes will be all connected with high probability if and only if each node covers an area $\pi r^2 = (\log(n) + c(n))/n$ and $c(n) \rightarrow \infty$.

[24] is another very important work not only because it assesses the connectivity of an ad hoc network but also because it investigates how transmission delay scales under different phases of connectivity. Authors consider that links may be unreliable between two neighbor nodes, and hence their connectivity may be affected. It is shown that there is still a critical node density threshold above which this network with dynamically unreliable links becomes connected with high probability, and that transmission delay scales linearly with distance when the network is in sub-critical node density phase, and that it scales sub-linearly with the distance when the network is in the supercritical phase. The subcritical and supercritical phases refer to the states of having a density smaller or larger than the critical node density for asymptotic connectivity, respectively.

2.4.4 Connectivity in Cognitive Radio Ad Hoc Networks

Connectivity studies in Sections 2.4.1, 2.4.2 and 2.4.3 consider connectivity in WSNs. However, CRSN has its unique challenges in connectivity, since nodes have to watch for PU transmissions, and hence two CRSN nodes are connected to each other when they are within the communication range of each other and they have an available common channel to communicate. Additionally, a realistic communication model may also be considered, which also accounts for effects of shadowing and/or fading on connectivity. In this context, works in the literature regarding connectivity in general CRAHN provide very useful insights for connectivity assessment of CRSN, and can be divided into two broad categories:

- Connectivity in underlay CRAHN.

- Connectivity in overlay CRAHN.

Connectivity in Underlay CRAHN

In underlay mode of cognitive radio, CRSN nodes communicate as long as the interference they cause upon PUs is not higher than a given threshold. In that sense, for a given PU network, a secondary network designer has to ensure that the PU Signal-to-Interference Ratio (SIR) is not below this required threshold.

Both [25] and [26] have studied the underlay transmission mechanisms and their resulting capacity expressions, however [26] also considers connectivity under the constraints of this communication mechanism. In this work, authors assume that primary and secondary networks are overlaid on the same area with homogeneous PPPs of intensities λ_P and λ_S , respectively. It is also assumed that each of the randomly deployed PU (SU) nodes is a transmitter PU (SU) node with probability p_p (p_s), and is a receiver with probability $1 - p_p$ ($1 - p_s$). Given that each PU (SU) transmitter transmits with the same power P_P (P_S), and that channel gain for a communication link of length r is given by $g(r) = r^{-\alpha}u$ (where α is the path loss exponent and u is a random variable for Rayleigh fading); then the SIR ratio for a primary transmitter-receiver $SIR_P(r)$ pair is defined as

$$SIR_P(r) = \frac{P_P g(r)}{I_P + I_{SP}} \quad (2.8)$$

where I_P is the interference caused on this link by other primary communicating parties and I_{SP} is the interference caused on that particular primary link by communicating SU nodes. $SIR_S(r)$ is defined in a similar way. Then, authors put the constraint on the outage probability of the primary network to be less than ϵ_P , expressed as

$$P(SIR_P(r) < T_P) < \epsilon_P \quad (2.9)$$

where T_P is the minimum threshold for $SIR_P(r)$ to evaluate that the communication

in this PU link as successful.

Taking this constraint into account, and also interpreting $SIR_S(r)$, authors find an expression for the random variable r , which renders the corresponding $SIR_S(r)$ larger than the minimum SIR threshold T_S for the secondary network. This random variable r has the interpretation of the effective communication range under the outage constraint of primary network, and can be leveraged to give a node isolation probability for the CRN nodes. From node isolation probability, one can approximate network connectivity, as previously summarized in Section 2.4.1.

Connectivity in Overlay CRAHN

In overlay CRAHN, secondary nodes cease their communication if they have a PU communication nearby. In select related works, overlay CRAHN is also called as interweave CRAHN. If a CRAHN node's communication range according to the Boolean model is R_c , then if there is at least an active PU within the circle of radius R_c centered at that node, that CRAHN node has to keep silent as long as the PU(s) are communicating.

A pioneering work on assessing connectivity in underlay CRAHN is [27]. Authors consider a primary network, overlaid with a secondary network. The primary network transmitters are assumed to be distributed according to a homogeneous PPP, with intensity λ_{PT} , and each primary transmitter is expected to have its own receiver within its communication range R_P , and the receiver's location within the communication range is determined in a uniformly random way. Secondary network, on the other hand, is distributed according to a homogeneous PPP, with intensity λ_S and each SU is assumed to have the same transmission range R_S . Based on this model, for the connectivity of SU nodes A (transmitter) and B (receiver), two conditions need to be satisfied:

- In a distance of R_S to A , there should not be any active PU receivers.
- In a distance of R_P to B , there should not be any active PU transmitters.

By considering this overlay activity model, authors have shown two very important results:

1. Leaving aside the connection interruptions due to PU activity, there still exists a critical node density for the secondary network to be topologically connected (i.e. each node has a topological connection to other node).
2. There is a critical PU node density above which SU network cannot be actually connected, although it can be topologically connected (i.e. under such a PU node density, SU network has topological connections but is left without spectrum opportunities to achieve actual communication connectivity).

2.5 Model Recommendations for Sensing Coverage and Connectivity in CRSN

In the next subsections, we discuss potential models for sensing coverage and connectivity issues in CRSN, building upon the reviewed literature in the previous sections, and also point out relevant open issues that require further research.

2.5.1 Joint Sensing Coverage and Connectivity Model for CRSN

One of the most important features of sensor networks is the estimation of actual field data with gathered data at the sink node with minimum distortion. The sink decision depends on the observations by nodes distributed in the network. The estimation of event has spatial distortion due to the spatial distribution of sensor nodes in the event region. Instead of the direct communication between energy-constrained nodes and the sink, multi-hopping provides energy-efficiency in sensor networks. Hence, data generated by a node is conveyed in multiple hops to the sink node. However, at each hop data packet is exposed to delays due to channel conditions of the relay nodes. As the arrival time of this data packet becomes greater, it becomes less useful to the sink for estimation, and the distortion increases. This delay causes temporal distortion.

The total distortion is the spatio-temporal distortion in the estimation of the event data. The design aim of a CRSN in this realistic setting should be that the sink estimates the whole field with minimum total distortion. As we increase the number of nodes in the network, spatial distortion decreases but time distortion may increase due to lesser channel availability and the smaller spectrum access probability for SU nodes. On the other hand, decrease in the number of nodes increases the spatial distortion due to greater inter-node distance. [4] investigates this issue by finding the optimal node number for 1D and 2D grid networks that minimizes the total distortion without taking into account DSA. We believe this work can be an invaluable guide for CRSN design by considering intermittent connectivity of CRSN and energy-efficiency constraints to model spatio-temporal distortion.

Assume that each node in the network observes a Gaussian process $S(x, y, t)$ with zero mean and variance $\sigma_s^2(x, y, t) = 1$ and it is correlated both in time and space with the rest of the network. The correlation of two observations in different times and locations is then

$$\begin{aligned}
 E[S(x_1, y_1, t_1)S(x_2, y_2, t_2)] &= \sigma_{S(x_1, y_1, t_1)S(x_2, y_2, t_2)} \\
 &= \sigma(|(x_1, y_1) - (x_2, y_2)|, |t_1 - t_2|) \\
 &= e^{-\alpha(|(x_1, y_1) - (x_2, y_2)|^2 + \beta^2(t_1 - t_2)^2)^K}
 \end{aligned} \tag{2.10}$$

where β is the scaling factor for time, $K = 0.5$ corresponds to a Gauss-Markov field, $K = 1$ corresponds to a squared distance correlated model field and α is a constant indicating correlation intensity. The type and impact of spatial and temporal correlations can be increased by leveraging the correlation models in [28].

If a one-dimensional network is assumed, the process $S(x, y, t_0)$ can be observed at the nearest sensor node at (x_0, y_0) as $S(x_0, y_0, t_0)$. However, this observation is sent to the sink via multiple hops, which causes a delay in CRSN due to PU activities.

The mean-square error of $S(x, y, t_0)$ at the sink can be given as

$$\begin{aligned}
D_{(x,y,t_0),(x_0,y_0,t_0+\tau)} &= E[(S(x, y, t_0) - S(x_0, y_0, t_0 + \tau))^2] \\
&= E[S(x, y, t_0)^2] + E[S(x_0, y_0, t_0 + \tau)^2] + \dots \\
&\quad + 2E[S(x, y, t_0)S(x_0, y_0, t_0 + \tau)] \\
&= \sigma_S^2(x, y, t_0) + \sigma_S^2(x_0, y_0, t_0 + \tau) - \dots \\
&\quad - 2\sigma_{S(x,y,t_0)S(x_0,y_0,t_0+\tau)} \\
&= 2 - 2e^{-\alpha(|(x,y)-(x_0,y_0)|^2 + \beta^2\tau^2)^K}
\end{aligned}$$

where τ is the random variable for total delay due to multi-hopping.

In CRSN, packets suffer from the random delay due to PU activities per each hop on the route to the sink. If nodes employ greedy forwarding technique, the neighbor within the communication range R_c with the smallest distance to the sink is chosen as the next hop for the packet. Furthermore, given the SU node density λ_S , the number of eligible nodes for greedy forwarding can be approximated to be Poisson with mean $\lambda_g = (\lambda_S \pi R_c^2)/2$. Hop progress, denoted as Φ , is then defined as the progress that the packet makes towards the sink at each hop. [29] reveals the probability density function $f_\Phi(\phi)$ of hop progress as given in Eq. (2.11), where l is the distance between the source and the sink. Then, the hop count h_c can be expressed as $\left\lceil \frac{l}{E\{\Phi\}} \right\rceil$. Forwarding region that includes eligible nodes for the next hop, which have been approximated to be the half circle closer to the sink of the full circle centered at the source node, has been illustrated in Figure 2.2. Here, shaded gray area represents the real forwarding region for possible relay node locations under greedy forwarding. However, whole semicircle is assumed to be the forwarding region, for analytical tractability. Source to sink distance is the random variable L , remaining

$$f_\Phi(\phi) = \begin{cases} \sum_{n=0}^{\infty} e^{-\lambda_g} \frac{(\lambda_g)^n}{n!} 2n \left(\frac{2}{\pi R_c^2} \right)^n (l - \phi) \left[\frac{\pi}{2} - \arcsin \left(1 + \frac{\phi^2 - R_c^2}{2l(l - \phi)} \right) \right] \left[(l - \phi)^2 \arcsin \left(1 + \frac{\phi^2 - R_c^2}{2l(l - \phi)} \right) \right. \\ \left. + \frac{1}{2} \sqrt{4R_c^2 l^2 - (\phi^2 - R_c^2 - 2l\phi)^2} - R_c^2 \arcsin \left(\frac{\phi^2 - R_c^2 - 2l\phi}{2lR_c} \right) - \frac{\pi(l - \phi)^2}{2} \right]^{n-1} & 0 \leq \phi \leq R_c \\ 0, & \text{otherwise} \end{cases} \quad (2.11)$$

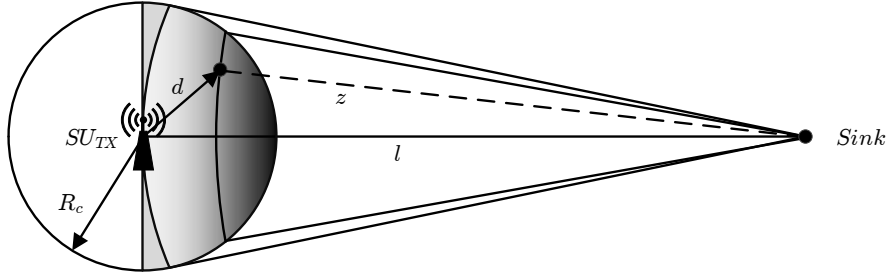


Figure 2.2: Actual and assumed forwarding regions for a source CRSN node.

distance after relaying is Z , then hop progress Φ is defined as $\Phi = L - Z$. For successful operation in CRSN, the frame must be received successfully. We have revealed in [52] that the mean delay until the Successful Frame Transmission (SFT) is achieved without a common control interface can be given as

$$E\{\tau_{sft}\} = \sum_{k=1}^{\infty} k(E\{\tau_{sa}\} + \tau_f)e^{-\gamma_k\tau_f} \prod_{c=1}^{k-1} (1 - e^{-\gamma_c\tau_f}) \quad (2.12)$$

where k is the number of transmissions until successful frame transmission, τ_f is the deterministic duration for transmission of a single frame, γ_c is the PU arrival rate of the channel in consideration, and $E\{\tau_{sa}\}$ is the mean spectrum access delay, i.e. the average time spent until finding a vacant channel from a sequential channel list. $E\{\tau_{sa}\}$ is highly dependent on the number of receiver PUs that can be found in the guardring zone centered at the transmitter CRSN node and on the number of transmitter PUs that can be found in the guardring zone centered at the receiver CRSN node. Those guardring areas can be a scaled version of the communication range of CRSN nodes, and the exact radius R_g for guardring zones would depend on the bearable interference between PU-SU parties. The guardrings centered at sender and receiver CRSN nodes, regions in which no PU should be active and the hop distance d are illustrated in Figure 2.3. From the results of [29] and [52], the delay of packet between source and the sink can be estimated as $\tau(h_c) = h_c E\{\tau_{sft}\}$, on average.

In randomly deployed sensor networks, Voronoi cell of a CRSN node is the set

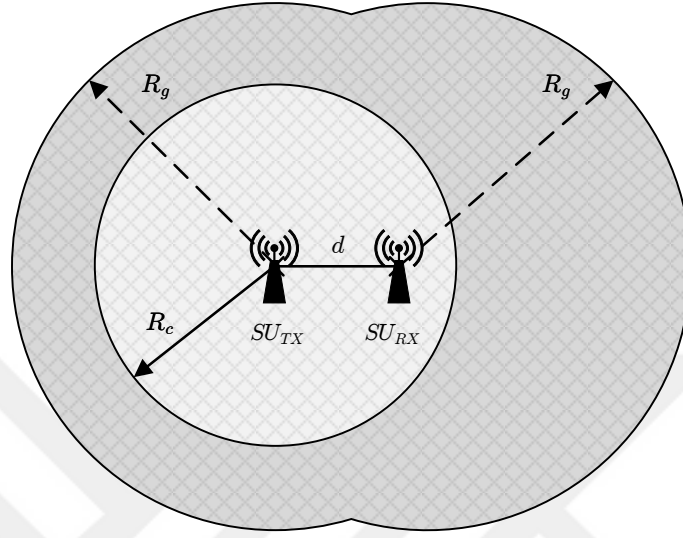


Figure 2.3: Guardrings of radius R_g centered at transmitter and receiver SU nodes.

of all (x, y) points that are closer to this node than any other node. Each CRSN node is responsible for sensing its respective Voronoi cell. An observation made by the node in a Voronoi cell is distorted spatially due to the distance between node and the observation point. In CRSN, Poisson Voronoi Diagram (PVD) is generated by the Voronoi diagrams of the node points. Since nodes are usually assumed to be distributed in an area A according to a PPP, Voronoi cell shapes are irregular and differ in size.

Assume that A is a circular field with radius L , and the sink resides at the center of A . If the mean of PPP is λ_S , then the number of Poisson Voronoi cells is $N_{cell} \approx \lambda_S |A|$ and the expected number of edges of Voronoi cell is $E\{N_{edge}\} = 6$. The area A_{cell} of Poisson Voronoi cells is fitted to a 3-parameter generalized Gamma distribution

$$f_{A_{cell}}(a_{cell}) = \frac{ab^{\frac{c}{a}}}{\lambda_S^c \Gamma(\frac{c}{a})} (a_{cell})^{c-1} e^{-b(\frac{a_{cell}}{\lambda_S})^a} \quad (2.13)$$

where $a = 1.07950$, $b = 3.03226$ and $c = 3.31122$, as shown in [30]. From this probability distribution function, a mean value $E\{A_{cell}\}$ can be easily computed. The exact analytical analysis of the spatial distortion is infeasible due to the random nature of Voronoi cell. However, the cell shape can be approximated as circle of

equivalent area which is centered at that node's position. It gives an approximate distortion expression although it does not fit the non-overlapping nature of PVD.

Hence, from above results and assumptions, the total spatio-temporal distortion due to a node that is k hops away from the sink can be estimated as

$$D_k = \int_{\theta=0}^{2\pi} \int_{x=0}^r (2 - 2e^{-\alpha(x^2 + \beta^2(\tau(k))^2)^K}) 2\pi x dx d\theta \quad (2.14)$$

where r is the estimated radius of a circularly-approximated Voronoi cell. The total distortion due to all N nodes can be formulated as

$$D(N) = \sum_{k=1}^{\lceil \frac{L}{E\{\Phi\}} \rceil} N(k) \int_{\theta=0}^{2\pi} \int_{x=0}^r (2 - 2e^{-\alpha(x^2 + \beta^2(\tau(k))^2)^K}) 2\pi x dx d\theta \quad (2.15)$$

where $N(k)$ are the number of nodes that are k hops away from the sink and that reside within a radius of $[(k-1)E\{\Phi\}, kE\{\Phi\}]$ from the sink, $E\{\Phi\}$ being the mean hop progress computed from Eq. (2.11).

In delay constrained networks, the delay's contribution to total distortion would be far more important than spatial distortion that is caused by sparseness of the CRSN nodes. In such cases, the distortion minimization problem would transform into a delay minimization problem. Heuristically, assuming channel access opportunities are present under the PU activity model, a CRSN design with a rather low node density λ_S and a rather high communication range R_c , up to the point where any additional increase in the R_c would violate interference constraint on PUs would be preferable. In other networks where spatial correlation is very low compared to the temporal correlation of sensed data, a high CRSN node density λ_S with low communication range R_c would be the preference. Exact optimal node density-communication range pairs are analytically infeasible to characterize, but can be found by numerical evaluations.

2.5.2 Connectivity Model for CRSN Based on Clustering

Previous sections laid out the popular literature on energy-efficient sensing coverage protocols for WSNs, which aimed at turning off redundant nodes while preserving significant coverage. [6] is one of the pioneering works in this area, aiming to find exclusive subsets which would cover the field of interest each on their own.

However, in the context of CRSNs, subsetting the whole network and making one subset of nodes active at each turn could cause problems since the nodes in a subset are not guaranteed to have a common channel to communicate on, in case an event occurs and they detect it. [31] has proposed a heuristic algorithm which clusters nodes based on proximity such that topological connectivity is guaranteed, and also maximizes the overlap between available channels intra-cluster. Figure 2.4 illustrates the iterative cluster formation with maximal available channel overlap. Tree structure in Figure 2.4 shows how each node can use which channel to communicate with each other. Node 1 chooses its cluster members, in such a way to scan all channel subsets, and chooses the channel subsets that maximizes the following metric: number of channels $\times$ number of one hop neighbors $\times$ number of two hop neighbors.

It is possible to enhance the work in [6] by also incorporating the maximization of channel availability overlap in [31], such that when a subset is ON, all the nodes

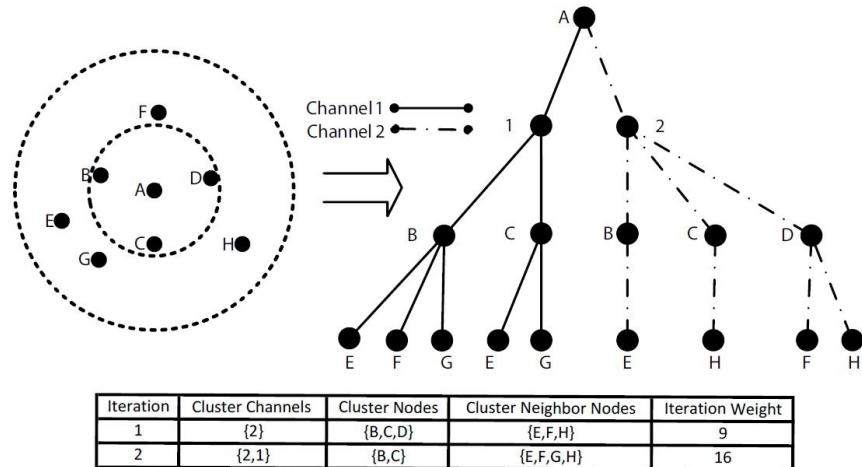


Figure 2.4: Illustration of cluster formation in [31] by iterations.

would have a common channel to quickly convey event detection information within the cluster. The clustering is done not periodically but when an event is detected, hence saving energy in reasonable event occurrence frequencies. Nevertheless, for periodic monitoring applications, the frequency of re-clustering is still not clear, and an adaptive algorithmic enhancement, which takes into account the recent historical temporal variation of PU communications while deciding on frequency of re-clustering is needed.

On the other hand, this proposed model would require choosing of cluster heads for coordination intra-subset. Although the cluster head duties may be switched between cluster nodes at each clustering operation, one may need a total distributed ON/OFF scheme for CRSN nodes. Distributed duty cycling and wake/sleep schedules are still open research problems in CRSN, where additional studies are needed to carry out distributed sensing coverage energy-efficiently.

2.5.3 Connectivity Assessment Models for CRSN

As noted in Section 2.4, the connectivity assessments in CRAHN depend on whether the transmission type is overlay or underlay. In this subsection, we would like to emphasize two very useful models for connectivity assessment in each of those transmission categories.

[32] considers a very interesting work on connectivity of wireless nodes based on communication delay. More specifically, two nodes are said to be connected if the delay incurred until transmitting a frame correctly from one of them to the other is below a delay threshold τ . The work considers a general wireless network setting, where delay is caused by interference coming from other transmitting nodes. In that sense, it can be applied to underlay CRSN (where interference would be caused by other CRSN nodes and PUs) by incorporating a CR related delay model, as in [52]. However, two questions still remain open:

- From an asymptotic sense, whether there is a critical delay threshold τ_c above which a CRSN becomes connected with high probability could be studied.

- The work can be tailored for CRSN and other sensor networks in general to assess node to sink connectivity, since in its current form it assesses connectivity among nodes that are one-hop neighbors.

On the overlay side, [27] is a very comprehensive work that assesses what is the needed secondary network density for it to be connected given the primary network density, and also finds out the critical primary node density above which a secondary network cannot be connected, irrespective of its own node density. This work is restrictive in the sense that PU communication is assumed to be static; hence whenever there is a PU at a particular location, an SU node overlaid very close to that location cannot be connected throughout the whole secondary network lifetime. Authors extended their work to include PU temporal traffic variations and clearly show that PU traffic variations in time increase the secondary network connectivity possibilities [33]. Still, the connectivity is assumed to be established when the multihop communication delay is finite; in that sense it could be incorporated in tailored case studies for CRSN to numerically assess how the source to sink delay varies under different node densities, transmission ranges and MAC protocols.

2.6 Open Topics for Sensing Coverage and Connectivity in CRSN

Although models from WSN and CRAHN are very useful for sensing coverage and connectivity in CRSN, we believe the following topics could be of great interest for the realization of CRSN:

1. How a CRSN node's sensor and transceiver duty cycling could be optimized is still an under-emphasized area. The sensor and transceiver duty cycling would affect how frequently a node could monitor its field and how frequently it could convey the sensing data towards the sink. An interesting application would be to minimize sensing data distortion while increasing network lifetime by appropriate duty cycling of sensor and the transceiver separately.

2. Mobile sensor can drastically change the landscape in CRSN. Mobility would certainly affect what percentage of an area is covered at a snapshot in time and through which path the data is forwarded to the sink. Works in the literature do not cover how to increase sensing coverage quality in mobile CRSN scenarios, neither through analytical nor heuristic approaches.
3. Delay based connectivity, first proposed in [32], could be an extremely useful model for CRSN, considering the intermittent connectivity nature of CRSN. However, its extension to sensor node to sink connectivity, rather than sensor node to sensor node connectivity, could be insightful for better engineering of CRSN.
4. Since frequency handoff is possible in CRSN, the selection of frequency could be based on energy-efficiency considerations. Although a lower frequency spectrum portion could be more energy-efficient, if its temporal variation in PU traffic is larger than a higher frequency spectrum portion, it may not be the best choice for CRSN operation since it could lead to higher number of intermittencies in connectivity of CRSN, with respect to the higher frequency spectrum portion. A spectrum decision protocol which considers distortion on field sensing data and tries to minimize energy expenditure could be an interesting application.

Chapter 3

ENERGY-EFFICIENT TRANSMISSION RANGE AND DURATION FOR CRSN

3.1 Introduction

Proliferation of wireless devices has caused spectrum scarcity, since available spectrum is not efficiently utilized with current fixed spectrum allocation approach. Dynamic Spectrum Access (DSA) is the paradigm to overcome such inefficiencies, by allowing unlicensed users to leverage the idle licensed spectrum. Cognitive Radio (CR) has been proposed as the key enabling technology for realization of DSA. CR is applied to Wireless Sensor Networks (WSNs) to form Cognitive Radio Sensor Networks (CRSNs), making most use of available spectrum for sensor networks without the burden of spectrum licensing costs [34].

Given the energy constraints of unlicensed CRSN nodes, called Secondary Users (SUs), communication between a CRSN source node and the remote sink takes place through multiple hops under DSA. Successful frame transmission under DSA in each hop depends on certain requirements. Firstly, SU transmitter and receiver nodes need to be in the transmission range of each other (i.e. they need to be topologically connected one-hop neighbors), and both need to correctly identify the same portion of the spectrum as idle through spectrum sensing. Secondly, no licensed user, called Primary User (PU), should arrive on that portion of spectrum during the SU transmission. In case of PU arrival, the SU pair need to restart sensing the spectrum in the next time slot and attempt retransmission if they can seize some portion of the spectrum. In this context, transmission range and duration have a non-trivial impact on energy-efficient operations of a CRSN. A larger transmission range results in less

number of hops towards the remote sink, yet faces higher energy consumption at each hop and a bigger potential number of PUs that may interrupt the SU communications. On the other hand, a smaller transmission range decreases energy consumption at each hop and reduces number of interfering PUs but increases the number of hops to reach the remote sink, hence might potentially increase overall source-to-sink energy consumption. Additionally, a longer transmission duration can deliver more bits in one transmission slot for a given bit rate, yet is more susceptible to incoming PUs on the accessed part of the spectrum during transmission.

Transmission range is often assumed to be a set parameter before sensor network deployment as lean sensor nodes may not have complex power control capabilities [35]-[36] or power control may not be cost-effective due to limited energy saving potential and added protocol complexity [37]. Energy-efficient transmission range is even more important in the CRSN context, given its impact on successful frame transmission under DSA schemes. Existing works that shed light on transmission range for ad hoc wireless networks and WSNs prior to their deployment [38]- [39] are not directly applicable to CRSN context given the intermittent nature of communication under DSA. Additionally, while it could be possible for CRSN to tune its transmission duration, finding an energy-efficient duration for different scenarios prior to deployment is key. In fact, our previous work considers this from the perspective of energy-efficient packet size [43]. To that end, energy-efficient transmission range and duration can be studied with a composite metric on energy consumed per goodput per meter progress towards the sink at each hop.

The remainder of this chapter is organized as follows. Related work on and motivations to carry out research on energy-efficient transmission range and duration is provided in Section 3.2. In Section 3.3, the system model to investigate energy-efficient transmission range and duration for CRSN is introduced. Derivations related to DSA characteristics, energy consumption dynamics of each node, successful frame transmission probability under DSA scheme and energy consumption thereof are presented in Section 3.4. For a given homogeneous transmission range, we revisit the

expected hop progress towards the sink as well as expected hop distance in a greedy forwarding scheme in Section 3.5. Using this result, we formulate the optimization problem to obtain energy-efficient homogeneous transmission range and duration for CRSN in the same section. The numerical results are presented and discussed in Section 3.6.

3.2 Related Work and Contributions

Recent research on CRSNs focused mostly on spectrum sensing [40]- [41], spectrum decision and handoff [42], physical layer studies on adaptive modulation [44] and channel bonding for CRSN [45] and first attempts on MAC layer [46]- [47], network layer [48] and transport layer [49]- [85] protocols and solutions. There is clearly a need to reveal energy-efficient transmission range and duration for CRSNs, which would serve as valuable guidelines for designers prior to deployment.

In this chapter, for the first time in literature, we formulate the problem of finding energy-efficient homogeneous transmission range and duration for CRSN deployment. To that end, we define two clear objectives, which then translate into the network model and consequent analyses to shed light on energy-efficient transmission range and duration for CRSNs:

- **Objective 1: Spectrum-aware transmissions:** Transmission range and duration should enable efficient transmission in each hop. Success of frame transmission on each hop of a CRSN depends on:
 - Presence of a spectrum opportunity and its correct sensing (after false alarms and mis-detections are accounted for) both by the transmitter and the receiver
 - No PU arrival during the entire transmission duration on the accessed spectrum portion
 - Correct reception of all bits on the receiver side under the unreliable radio link model

- **Objective 2: Energy-efficient operations:** Transmission range and duration should ensure overall energy-efficiency of the CRSN. This depends on:
 - Minimization of energy consumed per goodput per meter progress towards sink at each hop, hence increasing overall network energy-efficiency

Following these two core objectives, we first clearly lay our non-binding assumptions on the system model. Afterwards, we characterize the transmissions in CRSN under DSA; deriving the probability of successful frame transmission at each attempt between one-hop neighbor CRSN nodes, extending our previous work in [51]. We later revisit the strong foundations on expected hop progress concept from source towards sink at each hop (i.e. meters taken towards the sink at each hop) [38]. Given a homogeneous transmission range, we lay out the expected hop progress (i.e. the expected progress towards the sink at each hop) as well as the expected hop distance (i.e. the expected distance between one-hop neighbors) in a greedy forwarding scheme. From expected hop distance, we reveal the expected energy consumption until successful frame transmission under different spectrum utilization scenario probabilities, building on our work in [52]. Finally, we derive the energy consumed per goodput per hop progress meter towards sink and reveal optimal transmission range and transmission duration values thereof. The main contribution of this chapter is to reveal energy-efficient homogeneous transmission range and transmission duration for CRSN deployment and to investigate its relationship with the following critical network parameters:

- Path loss exponent (κ)
- PU death and birth rates (α and β)
- PU node density (ρ_p)
- Radius of PU guardring protection zone (r_p)

- Reference Signal-to-Noise Ratio (SNR) at SU receivers for correct demodulation (γ_0)
- Power consumed by SU during spectrum sensing (P_s)
- Maximum SU-PU collision probability that a PU can tolerate (P_{col})
- Ratio of PU signal variance (σ_p^2) to noise variance (σ_n^2)

3.3 System Model

In the system model adopted in our work, there are two types of network devices. The PUs have exclusive access to the licensed spectrum. On the other hand, the SUs access these licensed bands opportunistically, i.e., whenever there are no PU transmissions on them. The following features are incorporated in our analysis:

- PUs coexists with SUs on a 2-dimensional circular area (called $\mathcal{F}$, of radius Γ) to be monitored.
- Both PUs and SUs are randomly deployed over $\mathcal{F}$ with a 2-D homogeneous Poisson Point Processes (PPP) with means ρ_p and ρ_s , respectively.
- CRSN has a flat structure. Each SU node senses its environment and transmits event information to the sink in multiple hops. The sink resides at the center of $\mathcal{F}$.
- Simple greedy forwarding is assumed as routing mechanism [54]. Hence, an SU's next hop neighbor to transmit its field sensing information is assumed to be the neighbor within a range of r_s that is closest to the remote sink.
- SU nodes are assumed to know their neighbors' locations with respect to the remote sink, through a geographical location service [55].

- SUs have a common predetermined channel list, which they follow during periodic spectrum sensing operations [52]. This list may be provided by a centralized spectrum sharing mechanism. Formation of the optimal channel list is beyond the scope of this chapter, and its effect on our analysis is neutralized assuming all channels have the same PU density, PU activity model and bandwidth.
- Time is slotted for SU-SU communication. In each slot, two one-hop neighbors that are willing to communicate sense a licensed channel from their channel list, for Spectrum Access (SA). If they run out of channels in their list, they restart the list in a circular fashion.
- Two SU neighbors have a Spectrum Opportunity (SO) if i) they are not within the guardring of any PU at all times OR ii) the PUs in whose guardring they reside are all inactive during spectrum sensing. The guardring of a PU is a protection zone, centered at that PU's location, with radius r_p .
- Spectrum sensing might be erroneous. In case both SUs correctly sense the SO, that event is called Successful Spectrum Access (SSA).
- In case any of the two SU neighbors sense (either correctly or erroneously) a PU presence during spectrum sensing on the licensed channel, they perform spectrum handoff, moving to the next channel in their channel list, in a circular fashion.
- When two SU neighbors carry out SA (whether successful or not), *Frame Transmission (FT)* is started. In case no PU arrives during frame transmission, Successful Frame Transmission (SFT) takes place. Otherwise, the SU packet is not correctly delivered, and both SUs perform spectrum handoff to the next channel at the end of the transmission. We assume there is no dedicated radio to notify the SUs that a PU has arrived on the channel they communicate [52], hence SUs perform spectrum handoff in such a case at the of the transmission slot τ_f .

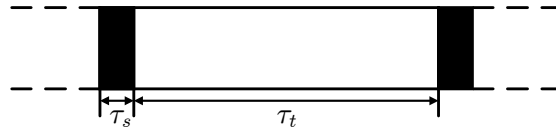


Figure 3.1: Assumed time slot structure for SU spectrum access and frame transmission.

- The PU activity is modeled as exponentially distributed interarrivals, with a two state birth-death process with death rate α and birth rate β . As a result of this birth-death process, the licensed spectrum portion under scrutiny can have ON and OFF states. An ON (Busy) state represents the period where this spectrum portion is occupied by PUs, whereas an OFF (Idle) state represents the period where it is unused. The length of the ON and OFF periods are exponentially distributed [43, 56].

3.4 Analysis Framework

3.4.1 Spectrum Sensing Events and Their Probabilities

The time is slotted, with two periods in each slot. The first period is reserved for spectrum sensing (τ_s), and the subsequent period is for the frame transmission (τ_t). Total duration of one slot is $\tau_f = \tau_s + \tau_t$, as depicted in Fig. 3.1.

T_{idle} and T_{busy} are the random variables that define the idle and busy time periods of each licensed channel with probability distribution functions $f_{T_{idle}}(t) = \beta e^{-t\beta}$ and $f_{T_{busy}}(t) = \alpha e^{-t\alpha}$, as per PU activity model explained in Section 3.3 [56]. Resultant channel probabilities are $P_{idle} = \alpha/(\alpha + \beta)$ and $P_{busy} = \beta/(\alpha + \beta)$, respectively.

Due its practicality, energy detection based spectrum sensing is assumed, with non-zero false alarm and mis-detection probabilities. Let the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ define idle and busy channel conditions, respectively. Hence, the probability of false alarm is

$$P_{fa} = Pr[Busy, \mathcal{H}_0] = P_{idle} \mathcal{Q}\left(\frac{\delta - 2\tau_s B \sigma_n^2}{\sqrt{4\tau_s B \sigma_n^4}}\right), \quad (3.1)$$

where B is the bandwidth of the channel, δ is the detection threshold, σ_n^2 is the noise variance, and $\mathcal{Q}(\cdot)$ is Q-function [56]. The probability of correct busy detection is given as

$$P_d = Pr[Busy, \mathcal{H}_1] = P_{busy} \mathcal{Q}\left(\frac{\delta - 2\tau_s B(\sigma_p^2 + \sigma_n^2)}{\sqrt{4\tau_s B(\sigma_p^2 + \sigma_n^2)^2}}\right), \quad (3.2)$$

where σ_p^2 is the PU signal variance [56]. Probability of correct vacant detection of idle spectrum is

$$P_v = P_{idle} - P_{fa} \quad (3.3)$$

and mis-detection probability is

$$P_{md} = P_{busy} - P_d. \quad (3.4)$$

3.4.2 Energy Consumption Dynamics

Energy consumed in each hop has three components. The first component is E_s , consumed for spectrum sensing during τ_s at the beginning of each slot. Both the transmitter SU and the receiver SU in a one-hop transmission consume $E_s = P_s \tau_s$ amount of energy for each spectrum sensing operation.

The second component is E_t , energy consumed for transmission which can be formulated as [39, 57]

$$E_t = (q_1 r_s^\kappa + q_2) \tau_t, \quad (3.5)$$

where κ is the path loss exponent, r_s is the transmission range and τ_t is the transmission period in a single time slot. q_1 and q_2 can be defined as [39]

$$q_1 = \frac{\gamma_0 N_{rx} N_0 B \left(\frac{4\pi}{\lambda}\right)^\kappa 10^\kappa}{G_a \eta_{amp}}, \quad (3.6)$$

$$q_2 = P_{elec}, \quad (3.7)$$

where γ_0 is the minimum required signal-to-noise (SNR) ratio at the receiver's demodulator, N_{rx} is the noise figure of the receiver, N_0 is the thermal noise in a 1 Hz bandwidth, B is the channel bandwidth, λ is the signal wavelength, η_{amp} is the transmitter amplifier efficiency, G_a is the antenna gain and P_{elec} is the power consumed by the device.

The third component is E_r , energy consumed on receiver side, formulated as $E_r = P_{rx}\tau_t$, where P_{rx} is the receiver power consumption.

Table 3.1: Spectrum utilization scenarios, probabilities and energy consumption thereof.

		Spectrum Access (SA)				Frame Transmission (FT)					
		TX sensing		RX sensing		PU arrival in τ_t		Reliability in τ_t			
S #	Init. state	Event	Prob.	Event	Prob.	Event	Prob.	Event	Prob.	Total probability	Cons. energy
S1	P_{idle}	FA ¹	$\overline{P_{fa}}$	FA ¹	$\overline{P_{fa}}$	-	1	-	1	$P_{idle}\overline{P_{fa}}^2$	$2E_s$
S2	P_{idle}	FA ¹	$\overline{P_{fa}}$	CVD ²	$\overline{P_v}$	-	1	-	1	$P_{idle}\overline{P_{fa}}\overline{P_v}$	$2E_s$
S3	P_{idle}	CVD ²	$\overline{P_v}$	FA ¹	$\overline{P_{fa}}$	-	1	-	1	$P_{idle}\overline{P_{fa}}\overline{P_v}$	$2E_s$
S4	P_{idle}	CVD ²	$\overline{P_v}$	CVD ²	$\overline{P_v}$	≥ 1 PU ⁵	$1 - P_{np}$	-	1	$P_{idle}\overline{P_v}^2(1 - P_{np})$	$2E_s + E_t + E_r$
S5	P_{idle}	CVD ²	$\overline{P_v}$	CVD ²	$\overline{P_v}$	No PU	P_{np}	≥ 1 error	$1 - P_r$	$P_{idle}\overline{P_v}^2 P_{np}(1 - P_r)$	$2E_s + E_t + E_r$
S6	P_{idle}	CVD ²	$\overline{P_v}$	CVD ²	$\overline{P_v}$	No PU	P_{np}	No error	P_r	$P_{idle}\overline{P_v}^2 P_{np}P_r$	$2E_s + E_t + E_r$
S7	P_{busy}	CBD ³	$\overline{P_d}$	CBD ³	$\overline{P_d}$	-	1	-	1	$P_{busy}\overline{P_d}^2$	$2E_s$
S8	P_{busy}	CBD ³	$\overline{P_d}$	MD ⁴	$\overline{P_{md}}$	-	1	-	1	$P_{busy}\overline{P_d}\overline{P_{md}}$	$2E_s$
S9	P_{busy}	MD ⁴	$\overline{P_{md}}$	CBD ³	$\overline{P_d}$	-	1	-	1	$P_{busy}\overline{P_d}\overline{P_{md}}$	$2E_s$
S10	P_{busy}	MD ⁴	$\overline{P_{md}}$	MD ⁴	$\overline{P_{md}}$	-	1	-	1	$P_{busy}\overline{P_{md}}^2$	$2E_s + E_t + E_r$

1: False Alarm — 2: Correct Vacant Detection — 3: Correct Busy Detection — 4: Mis-Detection — 5: At least 1 PU arrives within duration τ_t

3.4.3 Spectrum Utilization Scenarios

The spectrum holes are utilized opportunistically by SUs. Without loss of generality, MAC level scheduling between SUs is assumed to be collision-free, which can be enabled for a communicating SU pair via Guaranteed Time Slots through MAC layer protocols such as IEEE 802.15.4 [58]. Additionally, as mentioned in Section 3.3, a central entity or cluster heads can ensure SUs receive customized licensed channel lists for access.

For a successful transmission to take place among two one-hop neighbors, four conditions need to be met in total:

- The first two conditions are on finding a Spectrum Opportunity (SO) through reliable spectrum sensing on both the transmitter and the receiver sides, and are referred to as conditions for Successful Spectrum Access (SSA).
- The second two conditions cover the cases that throughout the transmission of the frame in τ_t , no PUs arrive on the channel and no bit errors occur on the transmitted frame, and referred to as conditions for Successful Frame Transmission (SFT).

Different events in these four conditions bring about 10 scenarios in total, which are summarized in Table 3.1. Within notations from Table 3.1, it should be noted that $\overline{P_{fa}} = P_{fa}/P_{idle}$, $\overline{P_v} = P_v/P_{idle} = (P_{idle} - P_{fa})/P_{idle}$, $\overline{P_d} = P_d/P_{busy}$ and $\overline{P_{md}} = P_{md}/P_{busy} = (P_{busy} - P_d)/P_{busy}$.

In the scenarios S1-S6 from Table 3.1, the licensed channel (that the SU transmitter SU_{TX} and receiver SU_{RX} are both sensing) is idle during spectrum sensing, presenting an SO. In this context, the SU_{TX} and SU_{RX} could correctly sense this channel as idle and access it subsequently. However, given imperfections in energy detection based spectrum sensing, there might be cases where they decide not to access this channel despite it being idle, summarized as scenarios S1-S3. These 3 scenarios cover the cases that at least one of the SU_{TX} and SU_{RX} decide the channel

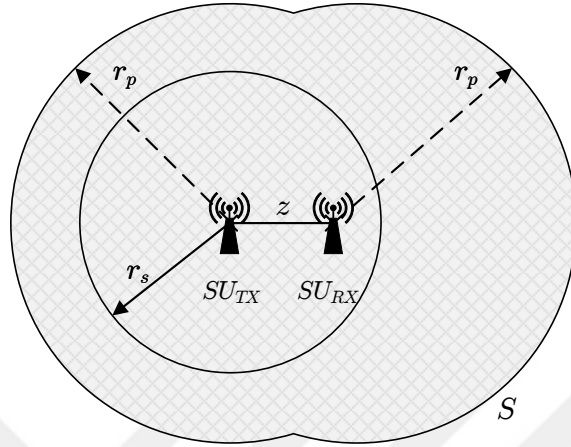


Figure 3.2: Illustration of guardring concept.

is busy, despite the SO, referred to as a False Alarm (FA) case. Consequently, SSA does not take place in scenarios S1-S3, and total energy consumed by the SU_{TX} and SU_{RX} is $2E_s$, for spectrum sensing operations.

In scenarios S4-S6, the SU_{TX} and SU_{RX} correctly identify the SO and achieve SSA. What differs in between these 3 cases are the fulfillment of two conditions for SFT. First condition is regarding the potential arrival of PU(s) during a time interval of τ_t on the channel that the SU_{TX} and SU_{RX} previously accessed. Any potential PU in the shaded area around the SU_{TX} and SU_{RX} illustrated in Fig. 3.2 should remain silent during τ_t so that no PU-SU collision occurs. With the underlying assumption that PU transmission range r_p is greater than SU transmission range r_s , no PU arrival probability during τ_t is expressed as

$$P_{np} = \sum_{k=0}^{\infty} \frac{e^{-\rho_p S} (\rho_p S)^k}{k!} (e^{-\tau_t \beta})^k, \quad (3.8)$$

$$= e^{-\rho_p S} \left(\sum_{k=0}^{\infty} \frac{(\rho_p S e^{-\tau_t \beta})^k}{k!} \right), \quad (3.9)$$

$$= e^{-\rho_p S (1 - e^{-\tau_t \beta})}, \quad (3.10)$$

where the first term in Eq. (3.8) is the probability of having k PUs in the area S illustrated in Fig. 3.2 according to Poisson distribution, and the second term is the probability that all of these k PUs remain silent at least for a duration of τ_t as per

the PU activity model from Section 3.3. The area S between the SU_{TX} and SU_{RX} that are z apart can be defined as

$$S(z, r_p) = 2\pi r_p^2 - 2r_p \cos^{-1}\left(\frac{z}{2r_p}\right) + \frac{z}{2} \sqrt{4r_p^2 - z^2}. \quad (3.11)$$

The second condition to achieve SFT once SSA is achieved is that all bits of the transmitted frame are received correctly on SU_{RX} side. Under unreliable radio link model [60], the probability that all the bits in the frame are received successfully can be written as

$$P_{rel} = (1 - BER(\gamma))^\zeta, \quad (3.12)$$

where BER is the bit error rate (BER) to evaluate the unreliability of the link, γ is the signal-to-noise ratio in the receiver, $\zeta = \tau_t R = \tau_t B \log_2(1 + \gamma)$ is the packet size sent during τ_t in bits with theoretically achievable rate R . The BER can be written as $BER(\gamma) = 0.5 \operatorname{erfc}(\sqrt{k\gamma})$ where $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{\sqrt{x}}^{\infty} \exp(-u^2) du$ is the complementary error function [59]. The channel is assumed to be an AWGN channel, and k depends on the modulation type used. If the modulation type is BPSK, then $k = 1$.

Under these conditions, as can be followed from Table 3.1, only scenario 6 refers to SSA followed by SFT. The probability of this scenario can be calculated as follows:

$$P_{S6} = P_v^2 P_{np} P_{rel}, \quad (3.13)$$

$$= (P_{idle} - P_{fa})^2 (e^{-\rho_p S(1 - e^{-\tau_t \beta})}) (1 - BER(\gamma))^\zeta \quad (3.14)$$

In scenarios S7-S10, the sensed licensed channel is busy when the SU_{TX} and SU_{RX} start their spectrum sensing, hence there is no SO available. In scenarios S7-S9, at least one of the SU_{TX} and SU_{RX} correctly detects this channel as busy. Consequently, both nodes carry out spectrum handoff to the next channel in their list and total energy consumed by them in that slot is $2E_s$, for spectrum sensing operations. However, in scenario S10, the SU_{TX} and SU_{RX} both incorrectly sense the

Table 3.2: Probability mass function of random variable T.

Number of trials (t)	P(T=t)
1	P_C
2	$P_C(P_A + P_B)$
3	$P_C(P_A^2 + 2P_AP_B + P_B^2)$
4	$P_C(P_A^3 + 3P_A^2P_B + 3P_AP_B^2 + P_B^3)$
...	...

channel as idle, which in turn triggers the SU frame transmission that will not result in an SFT. The probability of this scenario is

$$P_{S10} = P_{md}^2 = \left(P_{busy} \mathcal{Q} \left(\frac{2\tau_s B(\sigma_p^2 + \sigma_n^2) - \delta}{\sqrt{4\tau_s B(\sigma_p^2 + \sigma_n^2)^2}} \right) \right)^2, \quad (3.15)$$

where P_{md} is the mis-detection probability by an SU. Total energy consumed in that slot for S10 is $2E_s + E_t + E_r$, as SU_{TX} and SU_{RX} try to communicate, despite no SO being available, due to double mis-detection on both sides.

From a joint SFT and energy consumption perspective, there are three sets of scenarios:

- Set $A : \{S1, S2, S3, S7, S8, S9\}$, consisting of scenarios with no SFT and an energy consumption of $2E_s$ only.
- Set $B : \{S4, S5, S10\}$, consisting of scenarios with no SFT and an energy consumption of $2E_s + E_t + E_r$ due to failed transmission.
- Set $C : \{S6\}$, consisting of the only scenario with SFT and an energy consumption of $2E_s + E_t + E_r$.

As part of finding an energy-efficient transmission range r_s and transmission duration τ_t for SUs, it is of interest to compute the expected energy consumption among an

SU_{TX} and SU_{RX} that are z apart. We are specifically interested in the probability mass function (PMF) for the random variable T that denotes number of time slot trials until event C (i.e. scenario $S6$) resulting in an SFT takes place. Let us denote the probabilities of three independent events A , B , C as P_A , P_B , P_C respectively. Then the probability mass function for the random variable T , which represents the number of trials until SFT is achieved, can be illustrated as in Table 3.2 and expressed as follows:

$$P(T = t) = \sum_{k=0}^{t-1} P_C \left(\binom{t-1}{k} P_A^k P_B^{t-1-k} \right), \quad (3.16)$$

The expected energy consumption until SFT between SU_{TX} and SU_{RX} that are z apart can be computed as

$$E\{E_{sft}\} = \sum_{t=1}^{\infty} \sum_{k=0}^{t-1} P_C \left(\binom{t-1}{k} P_A^k P_B^{t-1-k} \right) (E_C + kE_A + (t-1-k)E_B), \quad (3.17)$$

where $E_A = 2E_s$ and $E_B = E_C = 2E_s + E_t + E_r$. Similarly, expected time until SFT can be expressed as

$$E\{\tau_{sft}\} = \sum_{t=1}^{\infty} \sum_{k=0}^{t-1} P_C \left(\binom{t-1}{k} P_A^k P_B^{t-1-k} \right) ((t-k)(\tau_s + \tau_t) + k\tau_s). \quad (3.18)$$

3.5 Energy-Efficient r_s and τ_t for CRSN

Our aim in this section is to lay the foundations to investigate energy-efficient homogeneous transmission range r_s and transmissions duration τ_t numerically for CRSNs in Section 3.6. To that end, we first revisit the strong foundations on expected hop progress, denoted as $E\{W\}$, from source towards sink at each hop [38]. Assumed routing approach is a greedy forwarding scheme called Least Remaining Distance (LRD), which would guarantee to minimize the remaining distance to sink at each hop. Secondly, based on expected hop progress towards the sink, we derive the expected hop distance $E\{Z\}$ under LRD scheme. Using $E\{Z\}$, the expected energy

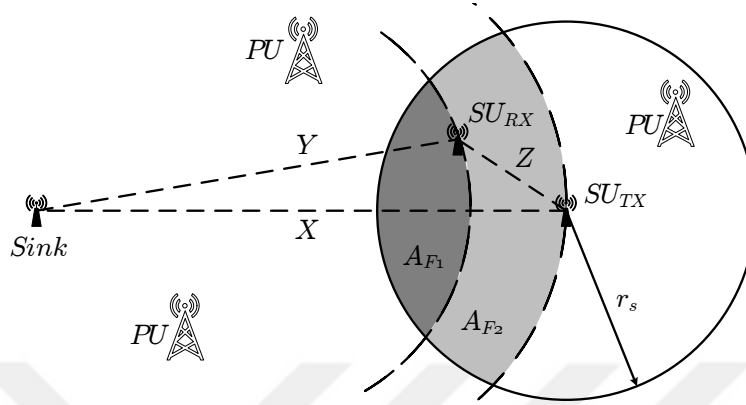


Figure 3.3: Illustration of Least Remaining Distance (LRD) forwarding approach.

consumption $E\{E_{sft}\}$ and time $E\{\tau_{sft}\}$ until SFT is achieved can be computed using scenario probabilities as in Eq. (3.17-3.18). We later use $E\{W\}$, $E\{Z\}$ and $E\{E_{sft}\}$ to numerically compute optimal homogeneous transmission range r_s^* and transmission duration τ_t^* in Section 3.6.

3.5.1 Expected Hop Progress

To study the expected hop progress towards the sink at each hop, we consider the setup illustrated in Fig. 3.3. Let SU_{TX} be the CRSN node with relevant information that needs to be conveyed to the sink. SU_{TX} has a transmission radius of r_s and would need to trigger multi-hopping towards the sink, unless $r_s < X$, in which case a single hop would suffice. SU_{TX} employs the LRD as the greedy forwarding scheme, choosing its neighbor in the forwarding region that is closest to the sink as the next hop and attempts SSA and SFT subsequently. The forwarding region for SU_{TX} is defined by $A_F = A_{F1} \cup A_{F2}$.

Given DSA and successful communication probabilities derived in detail in Section 3.4.3, SU_{TX} continues its attempts until SFT is achieved. At each hop, the same process is repeated, hence, each hop can be regarded as the first step of the remaining forwarding process towards the sink [37, 38].

For the setup in Fig. 3.3, assume that SU_{TX} chooses node SU_{RX} as its next hop neighbor according to LRD greedy forwarding [54]. Per system model characteristics,

both SU and PU locations are assumed to be random. Therefore, all the distances between SU_{TX} , SU_{RX} and the sink can be characterized by random variables. Let the distances between SU_{TX} and the sink, SU_{RX} and the sink and SU_{TX} and SU_{RX} be represented by the random variables X , Y and Z , respectively. In this context, $W = X - Y$ is the random variable that represents hop progress. LRD forwarding chooses the neighbor of SU_{TX} in the region $A_F = A_{F1} \cup A_{F2}$ with minimum Y , which in turn maximizes W . In the example in Fig. 3.3, this neighbor is SU_{RX} , with no other neighbor closer to the sink being present in the region A_{F1} .

W also depends on the distance between SU_{TX} and the sink. When $X \leq r_s$, SU_{TX} and the sink are one-hop neighbors, the hop progress is then automatically equal to X . If $X > r_s$, there must be a neighbor node in the forwarding region for hop progress towards the sink. This condition can be represented by the following random variable:

$$F = \begin{cases} 1 & , \text{if there is a neighbor in the forwarding region,} \\ 0 & , \text{otherwise.} \end{cases} \quad (3.19)$$

The expected hop progress is described and calculated as in Eq. (3.22), which is studied in detail in [38]. Since CRSNs have usually high and even ultra-high densities, the following approximations,

$$1 - e^{-\rho_s A_{F1}} \approx 1 \text{ and } 1 - e^{-\rho_s A_F} \approx 1, \quad (3.20)$$

can be made. Then, $E\{W\}$ expression simplifies to

$$E\{W\} = r_s - \frac{r_s^3}{3\Gamma^2}. \quad (3.21)$$

3.5.2 Expected Hop Distance

Although $r_p > r_s$ is strictly assumed, it is also considered that transmission ranges of PUs and SUs are comparable, so that channel availability for SUs has both spatial and

$$\begin{aligned}
E\{W|((X \leq r_s) \cup (X > r_s \cap F = 1))\} &= \int_0^\infty Pr\{W > w|((X \leq r_s) \cup (X > r_s \cap F = 1))\}dw \\
&= \frac{3\Gamma^2 r_s - 6 \int_0^{r_s} \int_{r_s}^\Gamma x e^{-\rho_s A_{F1}} dx dw}{3(x^2 - 2 \int_{r_s}^\Gamma x e^{-\rho_s A_F} dx)}
\end{aligned} \tag{3.22}$$

temporal variations, as in our previous work [51]. Hence, SO depends on the distance between SU_{TX} and SU_{RX} . This distance, depicted as Z in Fig. 3.4, determines the area of the guardring (as per Eq. (3.11)) where there should be no active PU for a duration of τ_t after SSA. To that end, Z has a direct impact on the expected energy $E\{E_{sft}\}$ and time $E\{\tau_{sft}\}$ required to achieve SFT. For a given $E\{W\} = w$, the next hop node SU_{RX} can be uniformly anywhere on the black solid arc that is from point K to point L as seen in Fig. 3.4. We assume linear approximation for the KL arc to decrease the computational burden, which is shown by the red line in Fig. 3.4. Then, the angle Φ between the edges that are X and Z long is uniformly distributed in $[-\Phi_{max}, \Phi_{max}]$. With the aforementioned approximation, for a given $E\{W\} = w$, it is clear that $\Phi_{max} = \cos^{-1}(\frac{E\{W\}}{r_s}) = \cos^{-1}(\frac{w}{r_s})$. Additionally, the hop distance can be written as $Z = \frac{w}{\cos(\Phi)}$ for a random point on the red line, with its maximum being equal to r_s . The expected hop distance for a given $E\{W\} = w$ is then

$$\begin{aligned}
E\{Z|w\} &= \int_{-\Phi_{max}}^{\Phi_{max}} \frac{w}{2\Phi_{max}\cos(\Phi)} d\Phi, \\
&= \frac{w}{2\Phi_{max}} \left(\ln(\tan(\Phi_{max}) + \sec(\Phi_{max})) - \ln(\tan(-\Phi_{max}) + \sec(-\Phi_{max})) \right)
\end{aligned} \tag{3.23}$$

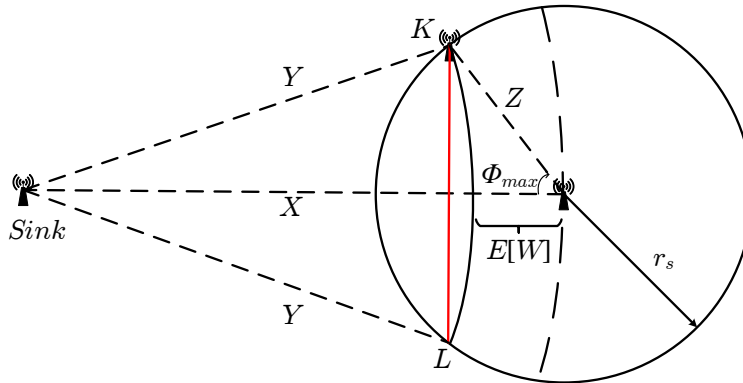


Figure 3.4: Relation between hop progress and hop distance.

The expected value given in Eq. (3.23) is evaluated and plugged into Eq. (3.11) to calculate guardring sizes for different network scenarios in Section 3.6.

3.5.3 Problem Formulation to Reveal Optimal r_s^* and τ_t^*

Based on the derivations for SSA and SFT probabilities from Section 3.4.3 and study of expected hop progress towards the sink and hop distance in a greedy forwarding scheme from Sections 3.5.1 and 3.5.2, we are ready to formulate the optimization problem. The aim of the optimization is to find optimal (r_s^*, τ_t^*) pair that would maximize the goodput-meter per Joule achieved in the provided CRSN system model. Hence, optimization problem can be formulated as a reward function Λ as

$$\max_{r_s, \tau_t} \Lambda = \frac{\frac{\tau_t R}{E\{\tau_{sft}\}} \cdot E\{W\}}{E\{E_{sft}\}} \quad (3.24)$$

$$\text{subject to } 0 < r_s < r_p,$$

$$0 < \tau_t < \tau_{max},$$

where the unit for Λ is (bits/s)×m per Joule. The maximum transmission duration constraint is $\tau_{max} = -\frac{1}{\mu} \log \left(1 - \frac{P_{col}}{P_{idle}} \right)$, where $\mu = \max\{\alpha, \beta\}$ and P_{col} is the maximum SU-PU collision probability that a PU can tolerate [56].

3.6 Numerical Results

In this section, first of all, we provide the preliminaries on required spectrum sensing duration τ_s given the desired transmission duration τ_t and the maximum SU-PU collision probability that a PU can tolerate P_{col} [56]. Consequently, behavior of the SFT probability P_{S6} as the underlying driver of the reward function Λ given in Eq. (3.24), is studied for different (r_s, τ_t) scenarios. Moving forward, reward function Λ is evaluated and examined in detail. As part of that, sensitivities of optimal $(r_s^*, \tau_t^*, \Lambda^*)$ values with respect to changes in 9 critical network parameters are tabulated in

detail. As a last step, how optimal $(r_s^*, \tau_t^*, \Lambda^*)$ values respond to combined changes in path loss exponent κ , PU death rate α and birth rate β are examined. Employed parameters and their values throughout all following analyses, unless otherwise stated, are provided in Table 3.3.

3.6.1 Preliminaries on τ_t and τ_s

Transmission duration τ_t is one of the two decision variables (r_s, τ_t) for the reward function Λ in Eq. (3.24) that is intended to be maximized. Sensing duration τ_s also directly affects SFT probability P_{S6} in a slot for the given system model, yet cannot be chosen arbitrarily. In fact, τ_s is dependent on the desired τ_t value, driven by the maximum SU-PU collision probability that a PU can tolerate P_{col} , hence is not a decision variable in our problem formulation.

Table 3.3: Parameters and their assumed values to reveal optimal r_s^* and τ_t^* .

Parameter	Symbol	Value
PU birth rate	β	3 / s
PU death rate	α	3 / s
PU guardring radius	r_p	200 m
Path loss exponent	κ	2.5
Reference SU SNR	γ_0	20 dB
PU signal vs. noise variance	$\gamma_p = \sigma_p^2 / \sigma_n^2$	10
Receiver noise figure	F_{rx}	12.589
Thermal noise	N_0	4.17×10^{-21} W/Hz
Bandwidth	B	10 KHz
Signal wavelength	λ	0.125 m
Amplifier efficiency	η_{amp}	0.2
Antenna gain	G_a	0.01
TX circuit power consumption	P_{elec}	3.63 mW
Spectrum sensing power cons.	P_s	700 mW
RX total power consumption	P_{rx}	11.13 mW
Network diameter	Γ	1000 m

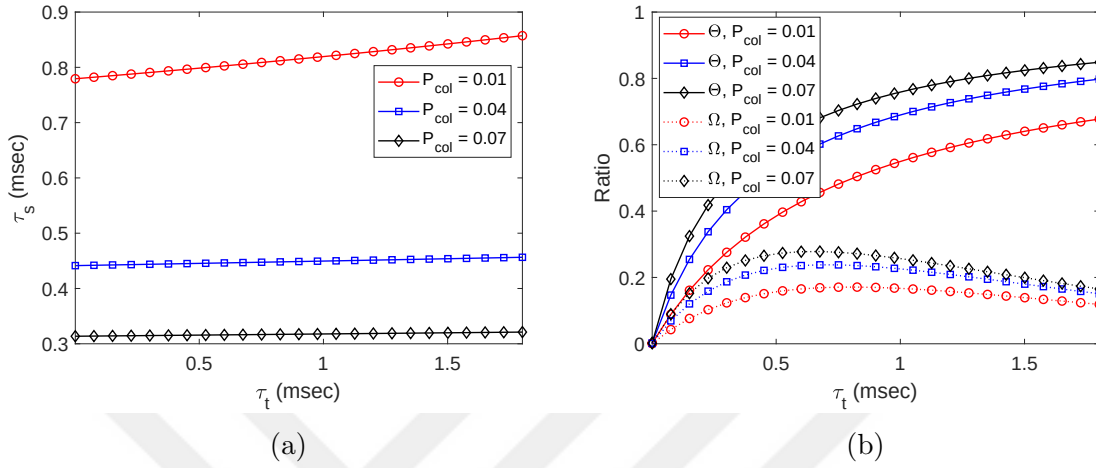


Figure 3.5: (a) Required spectrum sensing duration τ_s vs. τ_t for different P_{col} cases; (b) Comparison of sensing efficiency (defined as $\Theta = \frac{\tau_s}{\tau_s + \tau_t}$) and SFT efficiency (defined as $\Omega = \frac{\tau_t}{E\{\tau_{sft}\}}$ [52]) for different P_{col} cases.

For the optimization purposes in this chapter, τ_s - τ_t relation from [56] has been adopted. More specifically, for given α , β , τ_t and P_{col} values, the maximum value of the false alarm probability allowed as per P_{col} requirement is

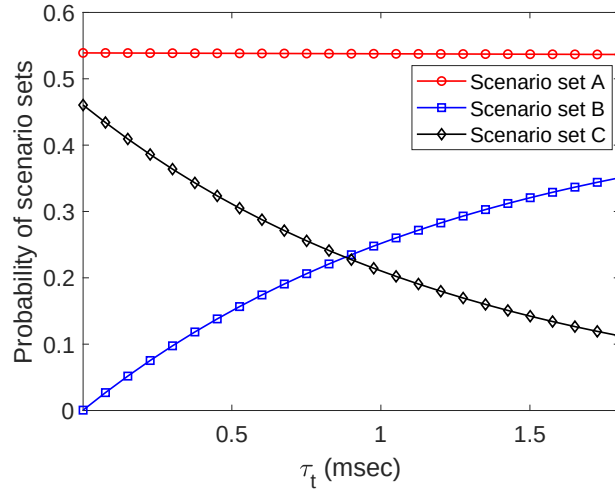
$$P_{fa} = P_{idle}P_{busy} - P_{idle}P_{busy} \left(1 - \frac{P_{col}}{P_{idle}}\right) e^{\mu\tau_t} \quad (3.25)$$

and τ_s is picked such that mis-detection and false alarm probabilities are equal, i.e. $P_{fa} = P_{md}$. This serves to balance interference caused upon PUs through mis-detection and lost spectrum opportunities through false alarm. With this approach, τ_s is given by

$$\tau_s = \frac{1}{B\gamma_p^2} \left(Q^{-1}\left(\frac{P_{fa}}{P_{idle}}\right) + (\gamma_p + 1)Q^{-1}\left(\frac{P_{fa}}{P_{busy}}\right) \right)^2. \quad (3.26)$$

Required τ_s for different P_{col} constraints are depicted in Fig. 3.5(a). Inherently, required τ_s that satisfies $P_{fa} = P_{md}$ is prolonged as P_{col} constraint gets stricter. Additionally, with higher P_{col} , slope of the τ_t vs. τ_s increases.

As per this τ_s - τ_t relation, it is of interest to reveal how much of the spectrum is intended and used for communications across time. Two metrics can be employed

Figure 3.6: Probabilities of scenario sets A, B and C vs. τ_t .

for that purpose. First one is the spectrum efficiency, defined as the intended ratio for communications in a slot, i.e. $\Theta = \frac{\tau_s}{\tau_s + \tau_t}$ [56]. Second one is the SFT efficiency, defined as the ratio of time used for communications to actual time spent until SFT is achieved, i.e. $\Omega = \frac{\tau_t}{E\{\tau_{sft}\}}$. Fig. 3.5(b) illustrates both metrics for different P_{col} cases. Accordingly, Θ monotonically increases in the given τ_t interval, whereas Ω peaks around 0.6-0.8 msec interval, depending on the case for P_{col} . Although neither Θ nor Ω considers energy-efficiency, Ω provides a better indication for likely optimal τ_t^* values for the problem here, given the $(\tau_t R)/E\{\tau_{sft}\}$ term in Eq. (3.24). Optimal τ_t values from Ω perspective, i.e. $\tau_t^{*,\Omega}$, for $P_{col} = 0.07$ is $\tau_t^{*,\Omega} = 0.63$ msec, for $P_{col} = 0.04$ is $\tau_t^{*,\Omega} = 0.72$ msec and for $P_{col} = 0.01$ is $\tau_t^{*,\Omega} = 0.78$ msec. As stricter P_{col} constraints drive τ_s significantly, it is logical that $\tau_t^{*,\Omega}$ follows this upward trend.

Consequently, in the following analyses, SFT efficiency Ω and suitable values for τ_s - τ_t thereof will be considered to optimize the reward function in Eq. (3.24).

3.6.2 Behavior of SFT Probability P_{S6}

As part of the optimization process of the reward function Λ in Eq. (3.24), it is important to reveal how probability of SFT, i.e. P_{S6} behaves and varies with critical parameters. Fig. 3.6 depicts the probabilities of scenario sets A, B and C (which were

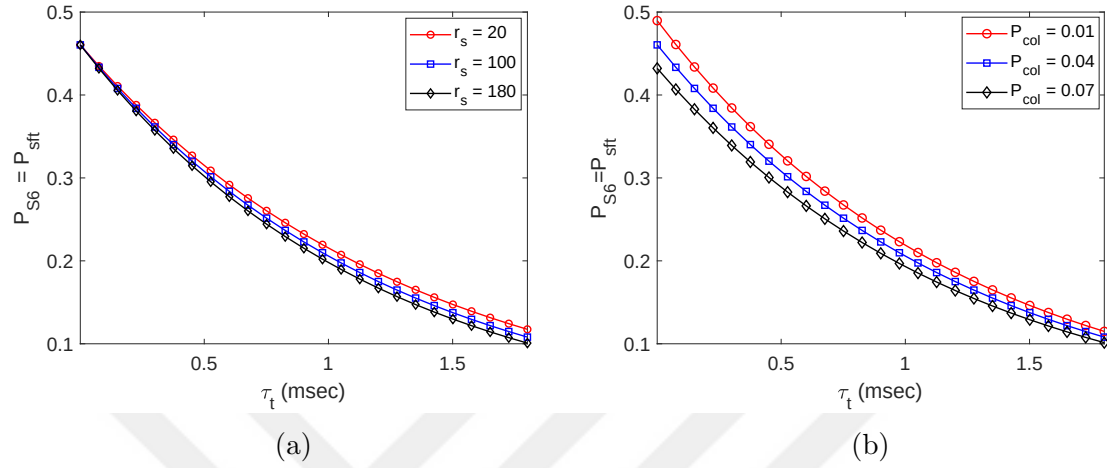


Figure 3.7: (a) Probability of success scenario P_{S6} vs. τ_t for different r_s cases; (b) Probability of success scenario P_{S6} vs. τ_t for different P_{col} cases.

previously defined in Section 3.4.3) with respect to τ_t , using parameter values defined in Table 3.3. Accordingly, probability of set A , which includes scenarios related to no SSA and is denoted by P_A , is relatively flat with changing τ_t . This result is natural, given the fact that P_A includes two types of events:

- Channel is actually idle, yet at least one of the SU_{TX} and SU_{RX} sense the channel as occupied.
- Channel is actually busy, and at least one of the SU_{TX} and SU_{RX} sense the channel as occupied.

Probabilities of both of these event types depend on spectrum sensing duration τ_s , which is also relatively flat with changing τ_t , as previously illustrated in Fig. 3.5(a). On the other hand, probability of SFT, i.e. $P_C = P_{S6}$ decays quickly towards null with increasing τ_t . To that end, P_{S6} and hence reward function Λ in Eq. (3.24) are significantly sensitive to changes in τ_t .

Fig. 3.7(a) breaks-down the same P_{S6} further by different r_s cases with respect to τ_t . Consequently, the effect of varying r_s on P_{S6} is insignificant when τ_t is low, yet suddenly becomes material with increasing τ_t . In fact, increasing r_s from 20 m

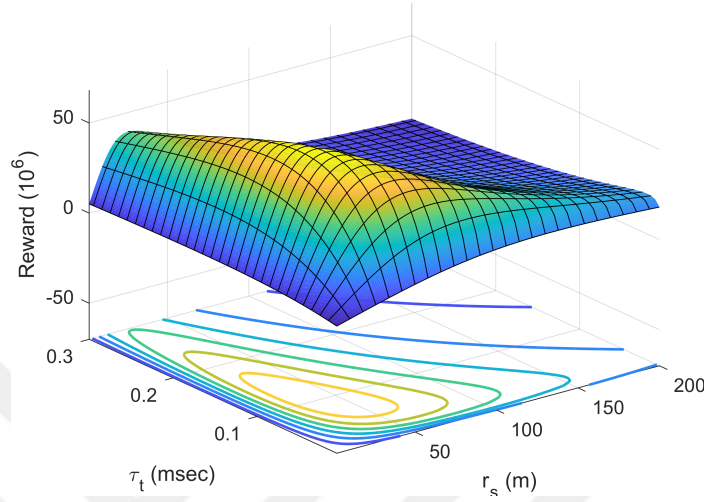


Figure 3.8: Illustration of the reward function Λ with respect to r_s and τ_t .

to 180 m reduces P_{S6} only by 1% when $\tau_t = 0.5$ msec; yet same r_s change reduces P_{S6} by 14% when $\tau_t = 1.8$ msec. To that end, effect of r_s and τ_t changes on P_{S6} are correlated and requires further deep-dive.

Fig. 3.7(b) details the combined effect of P_{col} and τ_t on P_{S6} . A stricter P_{col} requirement brings a lower $P_{fa} = P_{md}$, hence SFT probability P_{S6} is higher for $P_{col} = 0.01$ compared to when $P_{col} = 0.07$ as in Fig. 3.7(b). P_{S6} reduces by $\sim 10\%$ as P_{col} is relaxed from 0.01 to 0.07, with very little variation as τ_t is increased. To that end, effect of P_{col} and τ_t changes on P_{S6} are uncorrelated. Additionally, although a stricter P_{col} brings a higher P_{S6} in each slot, it still results in a lower SFT efficiency Ω as previously revealed in Fig. 3.5(b).

3.6.3 Reward Function Λ With Respect to r_s and τ_t

After the preliminary on τ_s - τ_t relation, and how P_{S6} reacts to changes in τ_s , τ_t and P_{col} ; it is time to construct the reward function Λ . Fig. 3.8 provides a sample illustration of Λ for the parameter values given in Table 3.3. The reward function Λ is smooth and well defined for all $r_s < r_p$ and $\tau_t < \tau_{max}$, as seen in Fig. 3.8. For this sample illustration, optimal $(r_s^*, \tau_t^*, \Lambda^*)$ trio is (40 m, 0.1 msec, 6.8×10^7 (bps×m)/Joule). We examine in detail how $(r_s^*, \tau_t^*, \Lambda^*)$ trio responds to changes in critical network parameters in Section 3.6.4.

Table 3.4: Changes to terms of the reward function Λ when individual parameters are increased, all else being equal.

Event	$\mathbf{E}\{\tau_{\text{sft}}\}$	$\mathbf{E}\{\mathbf{E}_{\text{sft}}\}$	$\mathbf{E}\{\mathbf{W}\}$	$\mathbf{R}$
$\kappa \uparrow$	$\leftrightarrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$
$\beta \uparrow$	$\uparrow$	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$
$\alpha \uparrow$	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
$\rho_p \uparrow$	$\uparrow$	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$
$r_p \uparrow$	$\uparrow$	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$
$\gamma_0 \uparrow$	$\leftrightarrow$	$\uparrow$	$\leftrightarrow$	$\uparrow$
$P_s \uparrow$	$\leftrightarrow$	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$
$P_{col} \uparrow$	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
$\sigma_p^2/\sigma_n^2 \uparrow$	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$

$\uparrow$: Increases || $\downarrow$: Decreases || $\leftrightarrow$: No change observed

Table 3.5: Sensitivity of optimal r_s^* , τ_t^* and Λ^* values with respect to changes in critical network parameters.

Parameter	Base value	Variable	Sensitivity	Change in r_s^* , τ_t^* and Λ^* when parameter is changed by...							
				-50%	-25%	-10%	-5%	+5%	+10%	+25%	+50%
κ	2.5	r_s^*	●	—	—	+242%	+74%	-42%	-63%	-88%	-97%
		τ_t^*	●	—	—	-11%	-5%	+5%	+9%	+17%	+32%
		Λ^*	●	—	—	+199%	+64%	-39%	-59%	-86%	-96%
β	3.0	r_s^*	○	-14%	-7%	-3%	-1%	+1%	+3%	+7%	+15%
		τ_t^*	○	+9%	+5%	+2%	+1%	-1%	-2%	-5%	-10%
		Λ^*	●	+59%	+25%	+9%	+4%	-4%	-8%	-19%	-34%
α	3.0	r_s^*	○	+9%	+3%	+1%	+1%	—	-1%	-2%	-3%
		τ_t^*	●	+25%	+9%	+3%	+1%	-1%	-3%	-6%	-11%
		Λ^*	●	-43%	-19%	-7%	-3%	+3%	+6%	+14%	+25%
ρ_p	0.001	r_s^*	○	-5%	-5%	-2%	-1%	+1%	+2%	+4%	+8%
		τ_t^*	○	+15%	+14%	+5%	+2%	-2%	-5%	-10%	-19%
		Λ^*	○	+13%	+6%	+2%	+1%	-1%	-2%	-5%	-9%
r_p	100	r_s^*	○	-18%	-9%	-3%	-2%	+2%	+4%	+9%	+19%
		τ_t^*	●	+64%	+27%	+10%	+5%	-5%	-9%	-20%	-36%
		Λ^*	○	+23%	+11%	+4%	+2%	-2%	-4%	-10%	-19%
γ_0	20 dB	r_s^*	●	+32%	+12%	+4%	+2%	-2%	-4%	-8%	-12%
		τ_t^*	○	—	—	—	—	—	—	—	—
		Λ^*	○	+12%	+5%	+2%	+1%	-1%	-2%	-4%	-7%
P_s	700 mW	r_s^*	○	-24%	-11%	-4%	-2%	+2%	+4%	+9%	+18%
		τ_t^*	○	-1%	—	—	—	—	—	—	—
		Λ^*	●	+51%	+19%	+7%	+3%	-3%	-6%	-13%	-22%
P_{col}	0.04	r_s^*	○	+7%	+3%	+1%	—	—	-1%	-2%	-4%
		τ_t^*	○	+13%	+6%	+2%	+1%	-1%	-2%	-5%	-9%
		Λ^*	●	-30%	-14%	-6%	-3%	+3%	+6%	+14%	+27%
σ_p^2/σ_n^2	10	r_s^*	○	+7%	+2%	+1%	—	—	—	-1%	-2%
		τ_t^*	○	+16%	+6%	+2%	+1%	-1%	-2%	-4%	-6%
		Λ^*	●	+32%	-12%	-4%	-2%	+2%	+4%	+9%	+15%

Legend ○: Not sensitive at all || ○: Slightly sensitive || ○: Mildly sensitive || ●: Significantly sensitive || ●: Very sensitive

3.6.4 Sensitivity of Optimal r_s^* , τ_t^* and Λ^*

Assuming parameter base values and (r_s^*, τ_t^*) are kept, Table 3.4 reveals how individual changes in parameters, all else being equal, impact each term on the reward function Λ^* .

On the other hand, Table 3.5 presents the sensitivity of optimal $(r_s^*, \tau_t^*, \Lambda^*)$ to individual changes in 9 critical network parameters, *ceteris paribus*. The base values for each parameter, from which the sensitivities are tested, are provided in second column of Table 3.5. Third and fourth columns together indicate how sensitive optimal $(r_s^*, \tau_t^*, \Lambda^*)$ trio are to changes in the corresponding network parameter. Next paragraphs detail out the impact of each parameter on network design and provides the reasoning thereof.

Sensitivity to Path Loss Exponent κ

Path loss exponent κ is the parameter that optimal $(r_s^*, \tau_t^*, \Lambda^*)$ are most sensitive to. When κ is raised, all else being equal (incl. base value (r_s^*, τ_t^*)), $E\{E_{sft}\}$ increases and achievable rate R decreases, as shown in Table 3.4. Increase in $E\{E_{sft}\}$ is very significant, driven by the fact that consumed energy in transmission is $E_t = (q_1 r_s^\kappa + q_2) \tau_t$. To that end, it is clear that new optimal Λ^* that to be attained once (r_s^*, τ_t^*) are updated will be much lower.

Given the E_t equation form, to arrive at the new optimal Λ^* , shift in r_s^* in the reverse direction of the change in κ is more effective compared to a change in τ_t^* . Hence, r_s^* moves significantly in the reverse direction of the κ change. Albeit not at the order of r_s^* , τ_t^* also shifts significantly with changes in κ . However, in contrast to r_s^* , τ_t^* moves in the same direction as κ , as indicated in rows 1-3 of Table 3.5.

Sensitivity to PU Birth Rate β

PU birth rate β is the second parameter that optimal Λ^* is most sensitive to. All else being equal (incl. base value (r_s^*, τ_t^*)), a rise in β increases both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$, as given in Table 3.4. This is driven by the fact that a surge in β drives down both

P_{idle} probability for SSA as well as P_{np} probability for SFT once SSA is achieved.

Adjustments on (r_s, τ_t) have little to no impact on P_{idle} and P_{fa} probabilities that drive the likelihood of SSA. However, both decision variables are directly relevant for $P_{np} = e^{-\rho_p S(1-e^{-\tau_t \beta})}$, which is the relevant probability to complete SFT after SSA is achieved. To optimize the reward function, system chooses to increase $E\{W\}$ via increasing r_s and reduce chances for PU arrival during frame transmission by decreasing τ_t , as indicated in rows 4-6 of Table 3.5. Consequently, both r_s^* and τ_t^* shift respectably as a response to changes in β . r_s^* moves in the same direction as β , while τ_t^* shifts to the reverse. It is worth noting that an increase in r_s^* is possible here, despite an increase in β , driven by the low base value for PU node density ρ_p .

Sensitivity to PU Death Rate α

PU death rate α is the third parameter that optimal reward is most sensitive to. All else being equal (incl. base value (r_s^*, τ_t^*)) when α rises, both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$ decrease, as revealed in Table 3.4. This is driven by the fact that a surge in α boosts P_{idle} hence likelihood of SSA.

To arrive at new optimal Λ^* , since average idle time of each channel is distributed exponentially with mean $1/\beta$, there is no reason for τ_t to increase as α increases. On the other hand, an increase in α surges P_{idle} to help SSA, yet has no impact on P_{np} that is relevant for SFT after SSA is achieved. Therefore, optimization seeks to maximize $P_{np} = e^{-\rho_p S(1-e^{-\tau_t \beta})}$, which is the probability that no PU arrives to the channel during frame transmission. It is much easier to raise this probability by decreasing τ_t rather than r_s , as S is a slowly increasing function with respect to r_s . To that end, all else being equal, τ_t^* and r_s^* are negatively correlated with α , with τ_t^* having a much higher sensitivity, as in rows 7-9 of Table 3.5.

Sensitivity to PU Node Density ρ_p

Optimal $(r_s^*, \tau_t^*, \Lambda^*)$ are slightly sensitive to changes in PU node density ρ_p . All else being equal (incl. base value (r_s^*, τ_t^*)) when ρ_p is increased, both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$

rise, as tabulated in Table 3.4. This in turn decreases the attainable Λ value.

Parameter ρ_p essentially has no direct link to likelihood of SSA, yet is impacting $P_{np} = e^{-\rho_p S(1-e^{-\tau_t \beta})}$ and hence SFT probability through the S term, which is defined as $S(w, r_p) = 2\pi r_p^2 - 2r_p \cos^{-1}\left(\frac{w}{2r_p}\right) + \frac{w}{2} \sqrt{4r_p^2 - w^2}$. Here, $E\{W\} = w$ for a given r_s . As shown in rows 10-12 of Table 3.5, when r_p is changed, to arrive at new optimal Λ^* , τ_t^* is mildly shifted in the opposite direction while r_s^* is slightly moved in the same direction. In that sense, as ρ_p rises, system chooses progressing more meters toward the sink at each hop at the expense of reduced transmission time. It should be noted that an increase in r_s^* is possible here, despite an increase in ρ_p , driven by the low base value for ρ_p .

Sensitivity to PU Guardring Radius r_p

Optimal $(r_s^*, \tau_t^*, \Lambda^*)$ are slightly sensitive to changes in PU guardring radius r_p . All else being equal (incl. base value (r_s^*, τ_t^*)) when r_p is increased, both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$ rise, as in Table 3.4. It is clear that reward value decreases with increasing r_p .

r_p has a similar story as ρ_p , having no direct link to likelihood of SSA but impacting the P_{np} probability for SFT completion. Hence, as a response to changes in r_p , τ_t^* is significantly shifted in the opposite direction while r_s^* is mildly moved in the same direction, as disclosed in rows 13-15 of Table 3.5. In fact, τ_t^* is most sensitive to r_p out of all the 9 parameters in scrutiny.

Sensitivity to Reference SU SNR γ_0

Optimal Λ^* is slightly and r_s^* is significantly sensitive to changes in γ_0 , while τ_t^* is not impacted at all. When γ_0 is increased, all else being equal (incl. base value (r_s^*, τ_t^*)), both the achievable rate R and expected energy consumption per SFT $E\{E_{sft}\}$ increase, as seen in Table 3.4.

Obviously, γ_0 has no direct link to SSA or SFT probabilities, yet mainly impacts the consumed energy during attempted transmissions, $E_t = (q_1 r_s^\kappa + q_2) \tau_t$. Here, γ_0 is

a linear multiplier term within q_1 . As γ_0 is changed, to arrive at the new optimal Λ^* , system prefers to adjust r_s^* in the opposite direction of the change of γ_0 , rather than τ_t^* . This is driven by the fact that r_s has an exponent $\kappa = 2.5$ in the E_t equation, whereas τ_t is a linear multiplier. Hence, r_s^* is significantly sensitive to and negatively correlated with γ_0 , while τ_t^* is uncorrelated, as shown in rows 16-18 of Table 3.5.

Sensitivity to Spectrum Sensing Power Consumption P_s

Optimal Λ^* is very sensitive to power consumption during spectrum sensing P_s . When P_s is increased, all else being equal (incl. base value (r_s^*, τ_t^*)), only $E\{E_{sft}\}$ increases, while no impact is observed on other terms of the Λ^* expression from Eq. (3.24), as seen in Table 3.4.

All three possible scenario sets A , B and C contain the term $2E_s = 2P_s\tau_s$ in their energy consumption in each slot. To that end, $2E_s$ is sort of a fixed cost in $E\{E_{sft}\}$ which is consumed until SFT. To that end, when P_s is increased, the system seeks to arrive at the new optimal Λ^* via increasing r_s^* . This increase in r_s^* is driven by the fact that system would like to achieve more of either $\tau_t R$ or $E\{W\}$ as fixed cost in the energy component $E\{E_{sft}\}$ becomes even more important. Raising $\tau_t R$ via increasing τ_t^* is not very feasible, as this increases likelihood of failed transmissions without SFT, thus increasing $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$. In contrast, r_s^* is linked to SFT probability through the S term in $P_{np} = e^{-\rho_p S(1-e^{-\tau_t^\beta})}$, which is a slowly increasing function of r_s . Hence, r_s^* shifts significantly in the same direction as γ_0 , whereas τ_t^* remains intact, as shown in rows 19-21 of Table 3.5.

Sensitivity to SU-PU Collision Probability P_{col}

Optimal Λ^* is significantly sensitive to the maximum SU-PU collision probability that a PU can tolerate P_{col} . When P_{col} is increased, the constraint on maximum SU-PU collision probability is relaxed. Consequently, all else being equal (incl. base value (r_s^*, τ_t^*)), both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$ decrease as seen in Table 3.4, increasing Λ as a consequence.

We have seen in Fig. 3.5(a) that as P_{col} decreases, required τ_s value significantly increases. Consequently, this increases not only the fixed cost component of energy consumption of amount $2E_s$ but also the time consumed for spectrum sensing τ_s , in each slot. To that end, as P_{col} decreases, to arrive at the new optimal Λ^* , the system shifts both r_s^* and τ_t^* in the reverse direction. r_s^* is slightly and τ_t^* is mildly sensitive to changes in P_{col} , as shown in rows 22-24 of Table 3.5.

Sensitivity to Ratio of PU Signal Variance to Noise Variance σ_p^2/σ_n^2

Optimal Λ^* is significantly sensitive to PU signal variance as a ratio to noise variance σ_p^2/σ_n^2 . When σ_p^2/σ_n^2 is increased, all else being equal (incl. base value (r_s^*, τ_t^*)), both $E\{\tau_{sft}\}$ and $E\{E_{sft}\}$ decrease as shown in Table 3.4, raising Λ as a result.

Assuming σ_n^2 fixed, increasing σ_p^2 for a given τ_s decreases the false alarm probability P_{fa} in Eq. (3.1) and the mis-detection probability P_{md} in Eq. (3.4). Hence, to achieve a fixed target $P_{fa} = P_{md}$, τ_s can be reduced when σ_p^2/σ_n^2 is increased. This curtails not only the $2E_s$ amount of energy consumption but also the time consumed for spectrum sensing τ_s , in each slot. Hence, r_s^* and τ_t^* are negatively correlated with σ_p^2/σ_n^2 , with τ_t^* being more sensitive, as shown in rows 25-27 of Table 3.5.

3.6.5 Sensitivity of r_s^* , τ_t^* , Λ^* to κ , β and α

As revealed in Table 3.5, the Λ function is most sensitive to changes in κ , β and α , respectively. To that end, we investigate the combined affect of changes in κ , β and α to $(r_s^*, \tau_t^*, \Lambda^*)$ in Fig. 3.9. In this representation, each row stands for a fixed α value ($\alpha = 10$, $\alpha = 20$ and $\alpha = 30$ for first, second and third rows, respectively) and each column stands for a fixed β value ($\beta = 10$, $\beta = 20$ and $\beta = 30$ for first, second and third columns, respectively). Additionally, each sub-figure displays 15 data points, each corresponding to a case on $\kappa = 2$ to $\kappa = 3.4$ with 0.1 steps. We also fix the PU guardring radius as $r_p = 100\text{m}$, with the condition that $r_s \leq r_p$, which is the usual case for CRSNs.

Accordingly, each sub-figure starts the data points with $\kappa = 2.0$ scenario, which

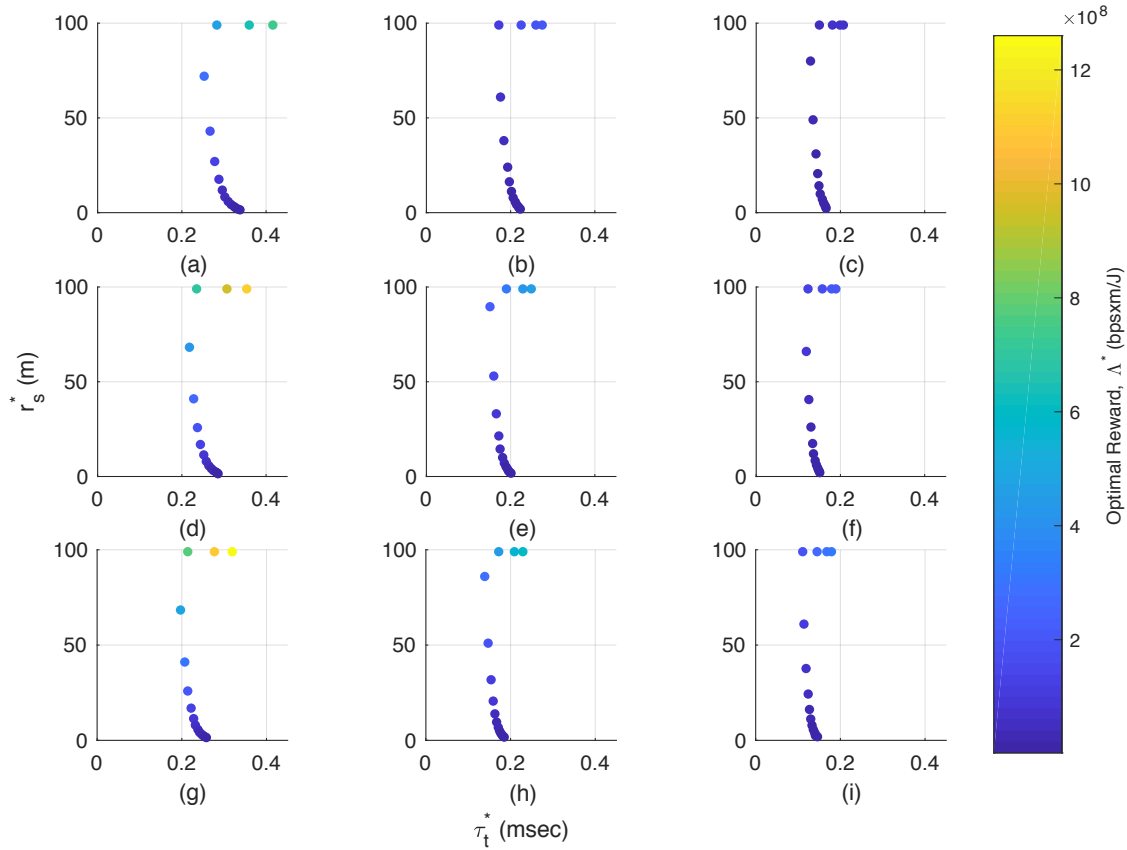


Figure 3.9: Optimal $(r_s^*, \tau_t^*, \Lambda^*)$ with respect to (a) $\alpha = 10, \beta = 10$; (b) $\alpha = 10, \beta = 20$; (c) $\alpha = 10, \beta = 30$; (d) $\alpha = 20, \beta = 10$; (e) $\alpha = 20, \beta = 20$; (f) $\alpha = 20, \beta = 30$; (g) $\alpha = 30, \beta = 10$; (h) $\alpha = 30, \beta = 20$; (i) $\alpha = 30, \beta = 30$, when κ changes from 2.0 to 3.4 with 0.1 steps, and PU guarding radius set as $r_p = 100$ m.

yields $r_s^* = 100$ m in all 9 cases. Again in all cases, r_s^* stays at 100 m for 3-4 data points corresponding to $\kappa = 2.0, 2.1, 2.2, 2.3$ scenarios, while τ_t^* is decreased. After $\kappa = 2.4$ or $\kappa = 2.5$ depending on the case, r_s^* is consistently reduced and τ_t^* is consistently boosted as we move towards the $\kappa = 3.4$ scenario.

As previously revealed in Section 3.6.4, r_s^* moves in a very sensitive manner in the reverse direction of the change in κ , whereas τ_t^* moves significantly in the same direction as κ . Hence, in each sub-figure, we see an increase in r_s^* until 100 m, as κ is decreased from 3.4 to 2.3-2.4. After that point, further decrease in κ cannot trigger any more r_s^* increase as it is capped by $r_p = 100$ m. Therefore, we observe an increase in τ_t^* as we further reduce κ from 2.3-2.4 levels to 2.0, despite they are being

positively correlated under normal circumstances. Key highlights on sensitivities of $(r_s^*, \tau_t^*, \Lambda^*)$ on κ , β and α are as follows:

- All three are most sensitive to κ . r_s^* and Λ^* move in reverse direction of κ , whereas τ_t^* follows κ with the exception of when r_s^* is capped to 100 m.
- Increasing β decreases τ_t^* significantly irrespective of κ value (except when r_s^* is capped to 100 m). Assuming $\alpha = 10$, increasing β from 10 to 20 and from 10 to 30 reduces τ_t^* by $\sim 33\%$ and $\sim 50\%$, respectively. Boosting α smoothens this decrease. As an example, when $\alpha = 30$, τ_t^* decreases by $\sim 29\%$ and $\sim 45\%$ for the same increases in β .
- Increasing α decreases τ_t^* significantly irrespective of κ value. Assuming $\beta = 10$, increasing α from 10 to 20 and from 10 to 30 reduces τ_t^* by $\sim 15\%$ and $\sim 24\%$, respectively. Boosting β smoothens this decrease. As an example, when $\beta = 30$, τ_t^* decreases by $\sim 9\%$ and $\sim 13\%$ for the same increases in α .

Chapter 4

COMMON CONTROL INTERFACE IN COGNITIVE RADIO NETWORKS: WHEN TO USE IT?

4.1 Introduction

Dynamic Spectrum Access (DSA) technique stand as a promising solution to inefficient spectrum utilization in wireless communications of the current fixed spectrum assignment policies. Cognitive Radio (CR) is the key technology for realization of DSA in wireless communications [2, 61]. With CR capabilities, wireless nodes can tune their communication parameters, communicate over vacant spectrum bands via carrying out cognitive cycle operations, thereby increasing overall spectrum-efficiency.

Improved spectrum utilization and communication capacity through CR comes at the expense of intermittent communication in wireless networks. Secondary Users (SUs), who are equipped with CR capabilities, get interrupted in their communication when a Primary User (PU) arrives on the spectrum they accessed previously. In such a case, SU has to switch its spectrum band (i.e. perform spectrum handoff), start sensing this new portion of the spectrum and try to access it consequently. Even with abundant spectrum opportunities, communications in Cognitive Radio Ad Hoc Networks (CRAHN) and Cognitive Radio Sensor Networks (CRSN) are still interrupted by periodic spectrum sensing functionality to detect available spectrum portions.

When spectrum availability is low, accessing the spectrum and then achieving successful communication might take a significantly longer time. In such scenarios, numerous channel switching and spectrum sensing periods take place consecutively, until finally a portion of the spectrum is accessed and the frame in consideration

is successfully transmitted from the transmitter SU to the receiver SU. Such cases represent significant communication delays, as well as spectrum and energy inefficiencies. In that sense, analyses on the intermittent communications in CRAHN and CRSN, with its impact on delay and efficiency, are essential to devise communication architectures.

Coordination of spectrum information among communicating entities, i.e. distributing spectrum sensing results and executing spectrum decision over a common medium, is essential in CRAHN and CRSN [62–64]. This coordination can be performed over the existing CR interface [63, 65–67], or over a dedicated Common Control Interface (CCI), which can be separately integrated into the architecture of CR nodes [64, 68, 69]. CCI can add value in the efficiency of CRAHN and CRSN, depending on critical network parameters and extent of PU activity. Without CCI, nodes cannot gather spectrum availability information during transmissions. Hence, without CCI, they would need to wait for their current transmission period to end for further spectrum coordination. However, with CCI, nodes can gather information about the availability of the spectrum they are using, simultaneously with transmissions, allowing them to cease transmission instantly upon any PU detection information. CCI inherently consumes additional energy, however savings in energy consumption through instantaneous spectrum coordination can compensate it in certain conditions. Goodput is certainly boosted with CCI, given the decrease in delay to achieve successful frame transmission. However, the extent of the goodput increase depends highly on network and communication parameters. On the other hand, CCI would also have nontrivial implications on distributed event sensing performance in CRSN. Hence, potential improvements provided by CCI needs to be thoroughly assessed, with respect to critical parameters such as PU activity rates, spectrum sensing and frame transmission durations and their respective power consumption levels. Such an assessment is beneficial to reveal conditions under which CCI would be worthy of its cost in terms of gains achieved in key performance indicators.

In this chapter, we present an analytical framework to assess the impact of CCI

on communication and distributed event sensing performance in CRAHN and CRSN [52], [53]. The developed framework reveals achievable delay, spectrum and energy-efficiency in CRAHN with and without CCI. On CRSN side, gains in distributed event sensing performance via CCI incorporation is also analyzed. To this end, Successful Frame Transmission (SFT) delay is formulated for a CR-equipped node pair and its closed-form expressions, with and without CCI, for any given channel sequence under heterogeneous PU activity, bit error rate, spectrum sensing and frame transmission durations are derived. The framework can also be used with any wireless channel model. Additionally, building on the SFT delay expressions, the energy-efficiency and spectrum-efficiency with and without CCI are also analytically characterized. SFT delay analyses are also embedded into distributed event sensing performance analysis in CRSN. Numerical results disclose achievable gains by inclusion of CCI, and reveal spectrum and network conditions that would render a CCI worthy of its costs.

The remainder of this chapter is organized as follows: In Section 4.2, the related work on spectrum handoff performance in CRAHN is presented and contributions of this chapter are listed. Section 4.3 lays out the system model adopted in our work. Section 4.4.1 derives the Spectrum Access (SA) delay in a dynamic radio environment. Sections 4.4.2 and 4.4.3 contain the analysis on the time elapsed until a frame is successfully transmitted, called Successful Frame Transmission (SFT) delay, without and with CCI, respectively. Sections 4.4.4, 4.4.5 and 4.4.6 analytically characterize the impact of CCI on energy-efficiency, spectrum-efficiency and distributed event sensing performance. The effects of CCI on network performance are assessed in Section 4.5 by extensive numerical evaluations.

4.2 Related Work and Contributions

In this section, we lay out the existing studies on the spectrum handoff performance in CRAHN and CRSN, and reveal contributions of this chapter with that respect. In [70], authors investigate impact of delay driven by consecutive spectrum handoffs

on the total data communication duration. A Preemptive Resume Priority (PRP) M/G/1 queuing system is proposed to model the delay caused by multiple handoffs. Total data communication duration, which is extended due to interruptions by PU arrivals, is modeled to help devise admission control strategies for secondary networks according to specific delay requirements. Instead of employing CCI for spectrum coordination, a frequency hopping-based spectrum handoff scheme is proposed in [71]. Accordingly, delay caused by spectrum handoff process in CRAHN under homogeneous PU activity at each channel is studied. A three dimensional discrete-time Markov chain is proposed to capture behavior of spectrum handoffs and assess performance of SUs according to communication requirements. In [72], a PRP M/G/1 queuing network model is formulated to study spectrum utilization under connection-oriented spectrum handoffs. Total data communication delay is derived and analyzed. In [73], authors investigate prioritized SU traffic. In this work, a general setting is considered, in which the PU transmissions can happen at any instant. Centralized and distributed schemes are considered to manage the prioritized SU traffic through different handoff mechanisms. The proposed DSA schemes are analyzed using a continuous time Markov chain. The mean handoff delay, for SUs with different priority segments is derived. In [74], the effects of interruptions due to multiple spectrum handoffs on connection-based channel usage is investigated based on a PRP M/G/1 queuing model. In [75], spectrum handoff process is analyzed. The average number of spectrum handoffs is derived as a key performance indicator.

In [70]- [72] and [74]- [75], the impact of PU activity on the SU communication delay is assessed, without CCI for spectrum coordination. In other studies, a CCI or a similar mechanism is assumed to be available for spectrum coordination [73], [76] or for disseminating spectrum sensing results [77]. This chapter differs from the related work by assessing the benefits of CCI on the spectrum handoff performance in terms of communication delay, and its consequent impact on network performance metrics for CRAHN and distributed event sensing performance of CRSN. Hence, this chapter fills the gap on the analyses regarding the cost-benefit trade-offs of CCI incorporation

in CRAHN and CRSN.

The contributions of our chapter, on top of the related work, are three fold:

- To the best of our knowledge, for the first time in the literature, the impact of employing a CCI for spectrum coordination on successful frame transmission delay is analytically characterized.
- Secondly, in contrast to related work, the impact of CCI is not limited only to the delay perspective. Potential gains in energy and spectrum-efficiency, as well as distributed event sensing performance through CCI inclusion are also studied analytically and numerically.
- Finally, the impact of CCI is thoroughly assessed by numerical results, revealing the critical network and spectrum conditions under which gains brought by CCI become material.

Thanks to these contributions, designers of CR equipped networks can have valuable design guidelines on when to include of CCI, under varying the parameters on the network and the spectrum.

4.3 System Model

In CRAHN and CRSN, unlicensed users (i.e. SUs) coexist with licensed users (i.e. PUs). Each SU has a CR interface, which carries out cognitive cycle operations such as spectrum sensing, decision and handoff. SUs perform periodic spectrum sensing to detect PU activity. During this period, the SU cannot communicate with its intended receiver. After spectrum sensing, in case the sensed spectrum band is found to be idle, the SU accesses this band and starts its transmission. However, if PU communication is detected during spectrum sensing, the SU performs spectrum handoff. During spectrum handoff, SU moves to the next channel in its pre-determined circular channel list, later restarting spectrum sensing at this new channel. If PU communication is not detected at a channel throughout a whole sensing period, the SU

resumes its communication in this channel, otherwise spectrum handoff is performed consecutively until a vacant channel is found.

The circular channel list that is used during spectrum handoff can be provided to nodes by a centralized spectrum sharing mechanism, or it may be constructed by means of cooperative spectrum sensing or each node may decide individually which channel to sense next [78]. Optimization of this channel list is beyond the scope of this chapter. The effect of channel list on CCI's impact on key performance indicators is normalized by using the same channel list for both with and without CCI cases.

We assume that PU activity on all channels follows an independent and identically distributed (i.i.d.) ON/OFF random process [56, 79–81]. The channel is occupied by a PU during an ON period. In contrast, the OFF periods represent PU absence at that respective channel. As per this model, PU inter-arrival and inter-departure durations, with an arrival rate of λ_c and departure rate of β_c on a specific channel c , are exponentially distributed with a mean of $\frac{1}{\lambda_c}$ and $\frac{1}{\beta_c}$, respectively.

4.4 Analysis Framework

In this section, we formulate

- Spectrum Access (SA) delay,
- Successful Frame Transmission (SFT) delay without CCI,
- Successful Frame Transmission (SFT) delay with CCI,
- Energy-efficiency during SFT with and without CCI,
- Spectrum-efficiency during SFT with and without CCI,
- Distributed event sensing performance in CRSN with and without CCI.

Without CCI, PU arrivals can be detected only after completion of current frame transmission; whereas with CCI, SUs can be instantly informed about PU arrivals by

neighbor nodes. We first formulate the Spectrum Access (SA) delay, i.e. the duration until an SU node gets a spectrum opportunity. Then, SFT delays with and without CCI are analytically characterized. Finally, energy-efficiency, spectrum-efficiency and distributed event sensing performance with and without CCI are analytically derived.

4.4.1 Spectrum Access (SA) Delay

In this subsection, we analytically model the delay experienced by an SU until it achieves Spectrum Access (SA) in a predetermined circular list of channels. This delay comprises the cumulative time elapsed until a full spectrum sensing period is completed at one of the channels without PU interruption. Required spectrum sensing duration for each channel would typically vary based on the channel signal-to-noise ratio (SNR), PU arrival λ_c and departure β_c rates [56]. To represent these effects, we assume that the required spectrum sensing intervals ($\tau_s^{(c)}$) for different channels are outcomes of a uniformly distributed random variable. We drop channel index c and consider spectrum sensing intervals for each channel c in the circular list at different access instants as an independent and identically distributed (i.i.d.) random variable τ_s , distributed as

$$f_{\tau_s}(\tau) = \begin{cases} \frac{1}{\tau_s^{\max} - \tau_s^{\min}}, & \tau_s^{\min} \leq \tau \leq \tau_s^{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4.1)$$

where $\tau_s^{\max}$, $\tau_s^{\min}$ are assumed maximum and minimum spectrum sensing durations.

To derive the probability distribution function for the SA delay τ_{sa} , we define $\tau_{is}^{(c)}$ as the random variable representing the duration of an interrupted spectrum sensing period in channel c . If a PU arrives before the completion of the sensing period τ_s , SU stops spectrum sensing, switches its channel and restarts spectrum sensing in the new channel. This is repeated until a full spectrum sensing period is completed at a channel without any PU arrival. At this point, spectrum is seized at that channel, and the frame transmission begins. In accordance with the PU model described in

Section 4.3, the conditional cumulative distribution function of interrupted sensing duration $\tau_{\text{is}}^{(c)}$ is given by

$$\begin{aligned} F_{\tau_{\text{is}}^{(c)}}(\tau|\tau_s) &= \frac{P(t \leq \tau \cap t \leq \tau_s)}{P(t \leq \tau_s)} \\ &= \frac{1 - e^{-\lambda_c \tau}}{1 - e^{-\lambda_c \tau_s}}, \quad 0 < \tau \leq \tau_s \end{aligned} \quad (4.2)$$

where λ_c is the PU arrival rate in channel c . Consequently, the conditional probability distribution function of $\tau_{\text{is}}^{(c)}$ is

$$f_{\tau_{\text{is}}^{(c)}}(\tau|\tau_s) = \frac{\lambda_c e^{-\lambda_c \tau}}{1 - e^{-\lambda_c \tau_s}}, \quad 0 < \tau \leq \tau_s. \quad (4.3)$$

After de-conditioning, using the probability distribution function of τ_s given in Eq. (4.1), the probability distribution function of $\tau_{\text{is}}^{(c)}$ is obtained for $0 < \tau \leq \tau_s^{\min}$ as

$$\begin{aligned} f_{\tau_{\text{is}}^{(c)}}(\tau) &= \int_{\tau_s^{\min}}^{\tau_s^{\max}} \frac{\lambda_c e^{-\lambda_c \tau}}{1 - e^{-\lambda_c \tau_s}} \frac{1}{\tau_s^{\max} - \tau_s^{\min}} d\tau_s \\ &= \frac{e^{-\lambda_c \tau}}{\tau_s^{\max} - \tau_s^{\min}} \ln \left(\frac{e^{\lambda_c \tau_s^{\max}} - 1}{e^{\lambda_c \tau_s^{\min}} - 1} \right) \end{aligned} \quad (4.4)$$

and for $\tau_s^{\min} \leq \tau \leq \tau_s^{\max}$ as

$$\begin{aligned} f_{\tau_{\text{is}}^{(c)}}(\tau) &= \int_{\tau}^{\tau_s^{\max}} \frac{(\lambda_c e^{-\lambda_c \tau})}{(1 - e^{-\lambda_c \tau_s})(\tau_s^{\max} - \tau_s^{\min})} d\tau_s \\ &= \frac{e^{-\lambda_c \tau}}{\tau_s^{\max} - \tau_s^{\min}} \ln \left(\frac{e^{\lambda_c \tau_s^{\max}} - 1}{e^{\lambda_c \tau} - 1} \right) \end{aligned} \quad (4.5)$$

Final form of $f_{\tau_{\text{is}}^{(c)}}$ is given below:

$$f_{\tau_{\text{is}}^{(c)}}(\tau) = \begin{cases} \frac{e^{-\lambda_c \tau}}{\tau_s^{\max} - \tau_s^{\min}} \ln \left(\frac{e^{\lambda_c \tau_s^{\max}} - 1}{e^{\lambda_c \tau_s^{\min}} - 1} \right), & 0 \leq \tau \leq \tau_s^{\min} \\ \frac{e^{-\lambda_c \tau}}{\tau_s^{\max} - \tau_s^{\min}} \ln \left(\frac{e^{\lambda_c \tau_s^{\max}} - 1}{e^{\lambda_c \tau} - 1} \right), & \tau_s^{\min} < \tau \leq \tau_s^{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4.6)$$

Table 4.1 illustrates the SA delay τ_{sa} for different spectrum handoff cases, where

Table 4.1: Illustration of Spectrum Access (SA) delay for a channel list composed of 3 channels.

# of visited channels	Cumulative SA delay	Probability
1	τ_s	$P_+^{(1)}$
2	$\tau_{\text{is}}^{(1)} + \tau_s$	$P_-^{(1)} \cdot P_+^{(2)}$
3	$\tau_{\text{is}}^{(1)} + \tau_{\text{is}}^{(2)} + \tau_s$	$P_-^{(1)} \cdot P_-^{(2)} \cdot P_+^{(3)}$
4	$\tau_{\text{is}}^{(1)} + \tau_{\text{is}}^{(2)} + \tau_{\text{is}}^{(3)} + \tau_s$	$P_-^{(1)} \cdot P_-^{(2)} \cdot P_-^{(3)} \cdot P_+^{(1)}$
5	$\tau_{\text{is}}^{(1)} + \tau_{\text{is}}^{(2)} + \tau_{\text{is}}^{(3)} + \tau_{\text{is}}^{(1)} + \tau_s$	$P_-^{(1)} \cdot P_-^{(2)} \cdot P_-^{(3)} \cdot P_-^{(1)} \cdot P_+^{(2)}$
...	...	...

$P_+^{(c)}$ and $P_-^{(c)}$ represent the probabilities of finding channel c vacant and occupied, respectively. Delay due to channel switching, τ_{sw} , is assumed to be accounted for in the random spectrum sensing duration τ_s . Accordingly, $P_+^{(c)}$ during a spectrum sensing interval τ_s is

$$\begin{aligned}
 P_+^{(c)} &= \int_{\tau_s^{\min}}^{\tau_s^{\max}} e^{-\lambda_c \tau_s} \frac{1}{\tau_s^{\max} - \tau_s^{\min}} d\tau_s \\
 &= \frac{e^{-\lambda_c \tau_s^{\min}} - e^{-\lambda_c \tau_s^{\max}}}{\lambda_c (\tau_s^{\max} - \tau_s^{\min})}
 \end{aligned} \tag{4.7}$$

Clearly, $P_-^{(c)} = 1 - P_+^{(c)}$. τ_s and $\tau_{\text{is}}^{(c)}$ are mutually independent random variables.

As shown in Table 4.1, τ_{sa} is the sum of independent τ_{is} random variables and its probability distribution function can be calculated by the convolution of the probability distribution functions of summed random variables as

$$\begin{aligned}
 f_{\tau_{\text{sa}}}(\tau) &= \left[1 + \left(\prod_{c=1}^C P_-^{(c)} \right) f_{\tau_{\text{sa}}^{(\text{tr}(C))}}(\tau) \right. \\
 &\quad \left. + \left(\prod_{c=1}^C P_-^{(c)} \right)^2 \left(f_{\tau_{\text{sa}}^{(\text{tr}(C))}} * f_{\tau_{\text{sa}}^{(\text{tr}(C))}} \right)(\tau) + \dots \right] \\
 &\quad * \left[\sum_{c=1}^C \left(\prod_{l=1}^{c-1} P_-^{(l)} \right) P_+^{(c)} \left(f_{\tau_{\text{sa}}^{(\text{tr}(c-1))}} * f_{\tau_s} \right)(\tau) \right]
 \end{aligned} \tag{4.8}$$

where $(f_{\tau_{\text{sa}}^{\text{tr}(c-1)}} * f_{\tau_s})(\tau)$ is the conditional probability distribution function of SA delay given that channel c is vacant. Here $f_{\tau_{\text{sa}}^{\text{tr}(c-1)}}(\tau)$ terms corresponds to the probability distribution function of the total duration spent for seeking a vacant channel until spectrum sensing starts at channel c and is equal to $f_{\tau_{\text{is}}^{(1)}} * f_{\tau_{\text{is}}^{(2)}} * \dots * f_{\tau_{\text{is}}^{(c-1)}}$. For tractability, convolutions can be turned into multiplication by taking the Laplace transform of the expression. In the resulting Laplace domain expression, since $\left| \prod_{c=1}^C (P_-^{(c)} f_{\tau_{\text{is}}^{(c)}}(s)) \right| < 1$, $\mathcal{F}_{\tau_{\text{sa}}}(s)$ can be found via geometric series interpretation on infinite series as

$$\mathcal{F}_{\tau_{\text{sa}}}(s) = \frac{\sum_{c=1}^C \left(\prod_{l=1}^{c-1} P_-^{(l)} \right) P_+^{(c)} \left(\prod_{k=1}^{c-1} \mathcal{F}_{\tau_{\text{is}}^{(k)}}(\tau) \right) \mathcal{F}_{\tau_s}(s)}{1 - \prod_{c=1}^C \left(P_-^{(c)} \mathcal{F}_{\tau_{\text{is}}^{(c)}}(s) \right)} \quad (4.9)$$

where $\mathcal{F}_{\tau_{\text{sa}}}(s)$, $\mathcal{F}_{\tau_s}(s)$ and $\mathcal{F}_{\tau_{\text{is}}^{(c)}}(s)$ are the Laplace transforms of $f_{\tau_{\text{sa}}}(\tau)$, $f_{\tau_s}(\tau)$ and $f_{\tau_{\text{is}}^{(c)}}(\tau)$, respectively. It is assumed that $P_-^{(0)} = 1$ and $f_{\tau_{\text{is}}^{(0)}}(s) = 1$.

4.4.2 Successful Frame Transmission (SFT) Delay Without CCI

To extend our SA delay analysis to SFT delay for the case where a CCI is not available, we define a new random variable $\tau_{\text{sft}}^{\text{sa}}$, corresponding to having $k-1$ frame transmission trials, each interrupted by PU arrivals during transmission periods, before having a successful frame transmission. The mean of the SFT delay without CCI, i.e. $E\{\tau_{\text{sft}}^{\text{sa}}\}$, is given by

$$E\{\tau_{\text{sft}}^{\text{sa}}\} = \sum_{k=1}^{\infty} k(E\{\tau_{\text{sa}}\} + \tau_f) e^{-\lambda_k \tau_f} \prod_{c=1}^{k-1} (1 - e^{-\lambda_c \tau_f}) \quad (4.10)$$

where k is the number of frame transmissions until successfully delivering a single frame, and τ_f is the deterministic duration for transmission of a single frame.

4.4.3 Successful Frame Transmission (SFT) Delay With CCI

In this subsection, we analytically model the SFT delay with CCI, i.e. τ_{sft} , which comprises the cumulative time elapsed, both for spectrum sensing and frame transmission, until a frame is successfully delivered on the link, under spectrum coordination with a CCI.

As a definition, $\tau_s^{(c)} + \tau_f$ is called as the cognitive cycle duration $\tau_{\text{cyc}}^{(c)}$. Each cognitive cycle $\tau_{\text{cyc}}^{(c)}$ in channel c starts with a spectrum sensing period $\tau_s^{(c)}$ to seek a spectrum opportunity. If a vacancy is found, the frame transmission interval τ_f starts, otherwise spectrum handoff to the next channel takes place. If no PU arrives during the frame transmission period, the frame is said to be successfully transmitted, otherwise it is retransmitted with the next spectrum vacancy. Thanks to the presence of CCI, frame transmission can also be ceased upon PU arrival, in contrast to the without CCI case.

Since $\tau_s^{(c)}$ is an i.i.d. random variable and τ_f is deterministic, $\tau_{\text{cyc}}^{(c)}$ is also an i.i.d. random variable for each channel at each access instant, and hence, channel index c is dropped from it. Using Eq. (4.1), probability distribution function of $\tau_{\text{cyc}}^{(c)}$ is found as

$$f_{\tau_{\text{cyc}}}(\tau) = \begin{cases} \frac{1}{\tau_{\text{cyc}}^{\max} - \tau_{\text{cyc}}^{\min}}, & \tau_{\text{cyc}}^{\min} \leq \tau \leq \tau_{\text{cyc}}^{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4.11)$$

where $\tau_{\text{cyc}}^{\max}$ and $\tau_{\text{cyc}}^{\min}$ are equal to $\tau_s^{\max} + \tau_f$, and $\tau_s^{\min} + \tau_f$.

We define the random variable $\tau_{\text{icyc}}^{(c)}$ which corresponds to the interrupted cognitive cycle duration in channel c . The probabilities of finding channel c in an ON, i.e., busy, and OFF, i.e., hole, state are $P_b^{(c)} = \frac{\lambda_c}{\lambda_c + \beta_c}$, and $P_h^{(c)} = \frac{\beta_c}{\lambda_c + \beta_c}$, respectively. Similar to derivation of $\tau_{\text{is}}^{(c)}$ in Section 4.4.2, τ_{icyc} is found to be distributed by

$$f_{\tau_{\text{icyc}}^{(c)}}(\tau) = \begin{cases} \frac{e^{-\lambda_c \tau}}{\tau_{\text{cyc}}^{\max} - \tau_{\text{cyc}}^{\min}} \ln \left(\frac{e^{\lambda_c \tau_{\text{cyc}}^{\max}} - 1}{e^{\lambda_c \tau_{\text{cyc}}^{\min}} - 1} \right), & 0 < \tau \leq \tau_{\text{cyc}}^{\min} \\ \frac{e^{-\lambda_c \tau}}{\tau_{\text{cyc}}^{\max} - \tau_{\text{cyc}}^{\min}} \ln \left(\frac{e^{\lambda_c \tau_{\text{cyc}}^{\max}} - 1}{e^{\lambda_c \tau} - 1} \right), & \tau_{\text{cyc}}^{\min} \leq \tau \leq \tau_{\text{cyc}}^{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4.12)$$

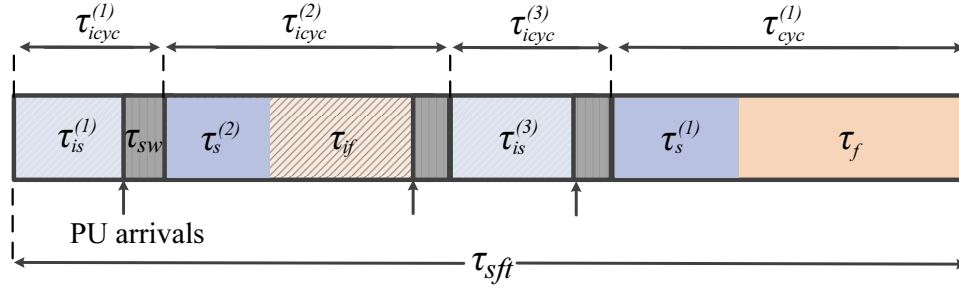

 Figure 4.1: Illustration of Successful Frame Transmission (SFT) delay τ_{sft} with CCI.

Fig. 4.1 illustrates SFT delay with CCI, denoted as τ_{sft} , whose probability distribution function $f_{\tau_{sft}}(\tau)$ can be derived as

$$\begin{aligned}
 f_{\tau_{sft}}(\tau) = & \left[1 + \left(\prod_{c=1}^C P_b^{(c)} \right) f_{\tau_{sft}^{tr(C)}}(\tau) \right. \\
 & + \left(\prod_{c=1}^C P_b^{(c)} \right)^2 \left(f_{\tau_{sft}^{tr(C)}} * f_{\tau_{sft}^{tr(C)}} \right)(\tau) + \dots \Big] \\
 & * \left[\sum_{c=1}^C \left(\prod_{l=1}^{c-1} P_b^{(l)} \right) P_h^{(c)} \left(f_{\tau_{sft}^{tr(c-1)}} * f_{\tau_{cyc}} \right)(\tau) \right]
 \end{aligned} \tag{4.13}$$

where $\left(f_{\tau_{sft}^{tr(c-1)}} * f_{\tau_{cyc}} \right)(\tau)$ is the conditional probability distribution function of SFT delay given that accessed channel c has been used successfully for the transmission of the frame. $f_{\tau_{sft}^{tr(c-1)}}(\tau)$ is the probability distribution function of the total interrupted cognitive cycle durations until cognitive cycle starts at channel c . Hence, $f_{\tau_{sft}^{tr(c-1)}}(\tau) = f_{\tau_{icyc}^{(1)}} * f_{\tau_{icyc}^{(2)}} * \dots * f_{\tau_{icyc}^{(c-1)}}$. Following the derivation of $\mathcal{F}_{\tau_{sa}}(s)$ in Section 4.4.2, since $\left| \prod_{c=1}^C (P_b^{(c)} \mathcal{F}_{\tau_{icyc}^{(c)}}(s)) \right| < 1$, $\mathcal{F}_{\tau_{sft}}(s)$ can be found via Laplace transform of Eq. (4.13) as

$$\mathcal{F}_{\tau_{sft}}(s) = \frac{\sum_{c=1}^C \left(\prod_{l=1}^{c-1} P_b^{(l)} \right) P_h^{(c)} \left(\prod_{k=1}^{c-1} \mathcal{F}_{\tau_{icyc}^{(k)}}(s) \right) \mathcal{F}_{\tau_{cyc}}(s)}{1 - \prod_{c=1}^C \left(P_b^{(c)} \mathcal{F}_{\tau_{icyc}^{(c)}}(s) \right)}, \tag{4.14}$$

where $\mathcal{F}_{\tau_{sft}}(s)$, $\mathcal{F}_{\tau_{cyc}}(s)$, and $\mathcal{F}_{\tau_{icyc}^{(c)}}(s)$ are the Laplace transforms of $f_{\tau_{sft}}(\tau)$, $f_{\tau_{cyc}}(\tau)$ and $f_{\tau_{icyc}^{(c)}}(\tau)$, respectively. It is assumed that $P_b^{(0)} = 1$ and $f_{\tau_{icyc}^{(0)}}(s) = 1$.

Eq. (4.10) and (4.13) are used to assess potential improvements in delay via incorporation of CCI in Section 4.5, and are also the starting point for the formulation of energy and spectrum-efficiency and distributed event sensing performance analyses in Sections 4.4.4, 4.4.5 and 4.4.6, respectively.

4.4.4 Energy-Efficiency During SFT With and Without CCI

While CCI saves energy by instant cessation of frame transmission in case of PU arrival, it consumes extra energy as a dedicated radio. To investigate whether incorporation of CCI brings a net positive effect in energy-efficiency, we study energy consumption to achieve an SFT with and without CCI. Power consumption levels for spectrum sensing, main radio frame transmission and CCI radio operation are denoted as Υ_s , Υ_f and Υ_{cci} respectively. The mean energy consumption to achieve an SFT without CCI is then

$$\Omega_{\text{sft}}^{\text{sa}} = \sum_{k=1}^{\infty} k(E\{\tau_{\text{sa}}\} \cdot \Upsilon_s + \tau_f \cdot \Upsilon_f) \cdot e^{-\lambda_c \tau_f} (1 - e^{-\lambda_c \tau_f})^{k-1} \quad (4.15)$$

and the energy-efficiency of achieving an SFT without CCI, i.e. the ratio of useful to total energy consumption is defined as $\Gamma_{\text{sft}}^{\text{sa}} = (\Upsilon_f \tau_f) / \Omega_{\text{sft}}^{\text{sa}}$.

On the other hand, when a CCI is present, assuming CCI is kept off during spectrum sensing and is always on during transmissions, one can approximate the total energy consumption to achieve an SFT as

$$\Omega_{\text{sft}} = E\{\tau_{\text{ss}}\} \cdot \Upsilon_s + (E\{\tau_{\text{sft}}\} - E\{\tau_{\text{ss}}\}) \cdot (\Upsilon_f + \Upsilon_{cci}) \quad (4.16)$$

where $E\{\tau_{\text{ss}}\} = E\{\tau_{\text{sa}}\} \cdot \frac{E\{\tau_{\text{sft}}^{\text{sa}}\}}{E\{\tau_{\text{sa}}\} + \tau_f}$ is the total time spent on spectrum sensing until successfully transmitting a frame. Energy-efficiency when a CCI is available is then $\Gamma_{\text{sft}} = (\Upsilon_f \tau_f) / \Omega_{\text{sft}}$. Using Γ_{sft} and $\Gamma_{\text{sft}}^{\text{sa}}$, we reveal potential improvements in energy-efficiency thanks to a dedicated CCI in Section 4.5.3.

4.4.5 Spectrum-Efficiency During SFT With and Without CCI

In this subsection, we characterize the spectrum-efficiency with and without CCI. Assuming an SU always has data to send, transmission starting at $t = 0$ successfully delivers $D(t)$ frames, achieving a goodput of $G(t) = D(t) \cdot R \cdot \tau_f / t$ in a duration of t for a communication rate of R . Steady-state goodput G of such spectrum utilization is found as

$$G = \lim_{t \rightarrow \infty} \frac{E\{D(t)\} \cdot R \cdot \tau_f}{t} \quad (4.17)$$

where $E\{D(t)\}$ is the mean of the number of frames transmitted successfully in $[0, t]$ for $t \rightarrow \infty$.

Achieved goodput during an SFT, i.e., $D(t) = 1$, in a duration of $\tau_{\text{sft}}^{\text{sa}}(i)$, i.e., in a realization of i.i.d random variable for SFT delay without CCI, is $G_{\text{sft}}^{\text{sa}} = R \cdot (\tau_f / \tau_{\text{sft}}^{\text{sa}}(i))$. Since each $\tau_{\text{sft}}^{\text{sa}}(i)$ ends with an SFT, the evolution of SFT delay $\tau_{\text{sft}}^{\text{sa}}(i)$ can be taken as a Markov regenerative process [82], with a reward of one frame. From the renewal theorem, asymptotic goodput without CCI is $G_{\text{sft}}^{\text{sa}} = R \cdot (\tau_f / E\{\tau_{\text{sft}}^{\text{sa}}\})$ where $E\{\tau_{\text{sft}}^{\text{sa}}\}$ is the mean of SFT delay without CCI derived in Section 4.4.2. Similarly, asymptotic goodput with CCI is $G_{\text{sft}} = R \cdot (\tau_f / E\{\tau_{\text{sft}}\})$ where $E\{\tau_{\text{sft}}\}$ is the mean of SFT delay with CCI derived in Section 4.4.3. We benefit from $G_{\text{sft}}^{\text{sa}}$ and G_{sft} to study the impact of CCI on spectrum-efficiency in Section 4.5.4.

4.4.6 Event Sensing Performance in CRSN With and Without CCI

CRSN is the offspring of the merger of Wireless Sensor Networks (WSN) and CR, provide distributed event sensing over a field of interest, and convey such information with DSA [34], [42], [43], [49], [83], [84], [85]. CRSN nodes report their event observations to the sink through multiple hops, with DSA. Using samples collected from individual nodes, the sink generates an estimate of the sensed signal, i.e., the event, at periodic intervals. In this subsection, we assess the distributed event sensing performance of a CRSN with and without CCI.

The average event communication duration in an estimation interval τ_e is $E\{\tau_e^{\text{non}}\} = \lfloor \tau_e / E\{\tau_{\text{sft}}^{\text{sa}}\} \rfloor \tau_f$ and $E\{\tau_e^{\text{cci}}\} = \lfloor \tau_e / E\{\tau_{\text{sft}}\} \rfloor \tau_f$ for without and with CCI cases, respectively. This duration reveals the average time spent for communication of event features with the sink. We use $E\{\tau_e^{\text{cci}}\}$ and $E\{\tau_e^{\text{non}}\}$ to calculate received samples from sensor nodes at sink at the end of each estimation interval.

We formulate estimation distortion with respect to event communication duration in an estimation interval under DSA in CRSN. For this analysis, the Best Linear Unbiased Estimator (BLUE) [86] is used to assess estimation distortion without any specific assumption on the distribution of the sensed signal and sensing noise. Sensor k obtains distorted version $s_k(t)$ of the event signal $\theta(t)$ at time t due to signal attenuation factor $h_k(t)$ and sensing noise $n_k^s(t)$, i.e., $s_k(t) = h_k(t)\theta(t) + n_k^s(t)$. We also assume that $h_k(t)$, $\theta(t)$ and $n_k^s(t)$ are i.i.d. over time. $\hat{\theta}(t)$ is estimated from the received samples from sensor nodes. Another assumption is that the mean of $\theta(t)$ and $n_k^s(t)$ are zero, and sensing noise $n_k^s(t)$ has a power of σ_k^2 . We can drop the time indices in accordance with independence over time assumption. Using the event communication duration, number of received samples X from sensor k is found as $X_{\text{non}} = (E\{\tau_e^{\text{non}}\} / \tau_f) \cdot S_f$ and $X_{\text{cci}} = (E\{\tau_e^{\text{cci}}\} / \tau_f) \cdot S_f$ for without and with CCI cases, where S_f is the number of samples per frame. Received sample vector $\mathbf{r}_k$ from sensor k is as $\mathbf{r}_k = \mathbf{h}_k \theta + \mathbf{n}_k^s$, where $\mathbf{r}_k = [r_1, \dots, r_{X_k}]^T$, $\mathbf{h}_k = [h_1, \dots, h_{X_k}]^T$, and $\mathbf{n}_k^s = [n_1^s, \dots, n_{X_k}^s]^T$. Covariance matrix $\mathbf{R}_k$ for received samples from CRSN node k is a X_k dimensional rectangular diagonal matrix whose diagonal entries are σ_k^2 for $1 \dots X_k$, and other entries are zero. Hence, the Mean Square Error (MSE) for sensor k is [86]

$$\begin{aligned} \text{Var} [\hat{\theta}_k] &= [\mathbf{h}_k^T \mathbf{R}_k^{-1} \mathbf{h}_k]^{-1} \\ &= \left(\sum_{i=1}^{X_k} \frac{1}{\sigma_k^2} \right)^{-1} \end{aligned} \tag{4.18}$$

and overall estimation distortion at sink for the BLUE is

$$\text{Var} [\hat{\theta}] = \left(\sum_{k=1}^K \frac{X_k}{\sigma_k^2} \right)^{-1} \quad (4.19)$$

We use Eq. (4.19) for distortion assessment of distributed sensing over CRSN with and without CCI in Section 4.5.5.

4.5 Numerical Results

In this section, potential improvements through CCI utilization are examined numerically from SFT delay, energy-efficiency, spectrum-efficiency and distributed event estimation performance perspectives. We first provide the assumed network setup and investigate accuracy of the presented delay analysis in Sections 4.4.2 and 4.4.3. Then, effects of CCI utilization on various performance metrics in CRAHN and CRSN are examined.

4.5.1 Network Setup

For the numerical analyses, frame duration τ_f and mean spectrum sensing duration $E\{\tau_s\}$ are assumed to be 100 ms and 5 ms [90], PU packet arrival rate λ is varied from 2 to 20 packets/sec, SNR for AWGN channels is assumed to be 20 dB, unless otherwise stated.

The Nearest Neighbor Model

Nodes are assumed to communicate to their nearest neighbor. We assume that both PU and SU nodes are distributed randomly according to a stationary homogeneous Poisson Point Process (PPP), with intensities λ_{PU} and λ_{SU} , respectively. With PPP, probability that an SU node has less than k neighbors in a neighborhood of r is $P(k) = \sum_{i=0}^{k-1} e^{-\lambda_{\text{SU}} \pi r^2} \frac{(\lambda_{\text{SU}} \pi r^2)^i}{i!}$. The probability distribution function for the distance to the nearest SU neighbor is then $f_{r_1} = 2\pi \lambda_{\text{SU}} r_1 \cdot e^{-\lambda_{\text{SU}} \pi r_1^2}$ [87]. Given the distance r_1 between two nearest neighbors, and assuming each PU node has an average packet

arrival rate of α , we approximate the total PU traffic in channel c by arrivals in a radius $R_c = 2r_1$ neighborhood of the SU link as $E\{\lambda_{R_c}\} = \alpha(\lambda_{PU}\pi R_c^2)$.

Wireless Channel Model

Effects of CCI are studied for both AWGN and flat fading channels. Assumed modulation scheme for SU communications is Quadrature Amplitude Modulation (4-QAM). Bit Error Rate (BER) for AWGN channel with 4-QAM modulation is $P_b^{\text{awgn}} = \frac{3}{4}Q(\sqrt{4\frac{E_b}{N_0}/5})$, where E_b is energy per bit and noise has flat power spectral density of $N_0/2$ W/Hz [88]. The BER for 4-QAM over a fading channel is given by $P_b^{\text{fade}} = \frac{2}{\pi} \int_0^{\pi/2} M_\gamma\left(-\frac{9}{2\sin^2(\phi)}\right) d\phi - \frac{1}{\pi} \int_0^{\pi/4} M_\gamma\left(-\frac{9}{2\sin^2(\phi)}\right) d\phi$, where $M_\gamma(s) = (1 - s\bar{\gamma})^{-1}$ and $\bar{\gamma}$ is the SNR [89].

4.5.2 Evaluation of SFT Delay

Accuracy of the analytical expressions for $E\{\tau_{\text{sft}}^{\text{sa}}\}$ and $E\{\tau_{\text{sft}}\}$, derived in Sections 4.4.2 and 4.4.3 are first evaluated via time-step simulations. Then, we examine in detail the SFT delays with and without CCI under the implications of the network model.

Validation of SFT Delay Analyses

The analytical SFT delay expressions are constructed in Mathematica, and compared with results of a realistic time-step simulation framework in MATLAB for validation. The PU packet departure rate is set to 30, in accordance with expectation of a stable channel. The analytical and simulation results for $E\{\tau_{\text{sft}}^{\text{sa}}\}$ and $E\{\tau_{\text{sft}}\}$ are compared in Fig. 4.2. The difference between the analytical and simulation results are below 5% for all given λ values, increasing slightly as λ 's become comparable to the β value.

Evaluation of SFT Delay With and Without CCI

Let us first define the SFT delays under the network setup, which are denoted by τ_{cci} and $\tau_{\text{n-cci}}$ for the cases with and without CCI, respectively. It is assumed that whole

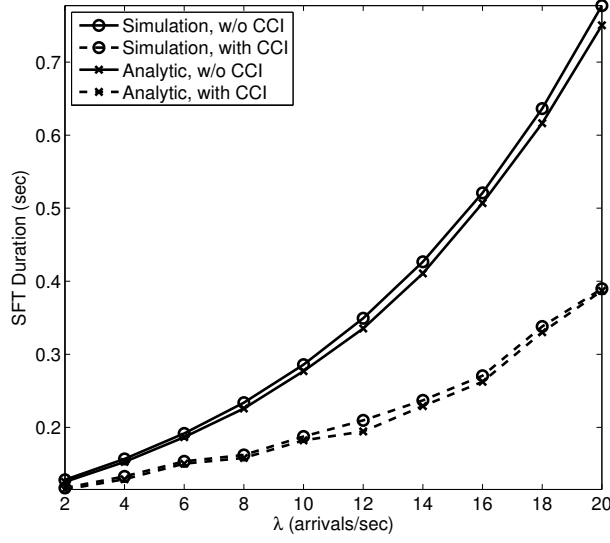


Figure 4.2: Comparison of analytical and simulation based results on SFT delay.

τ_{sft} and $\tau_{\text{sft}}^{\text{sa}}$ have to be repeated in case of a wireless error in these two scenarios. $E\{\tau_{\text{cci}}\}$ is given by

$$E\{\tau_{\text{cci}}\} = \sum_{k=1}^{\infty} k \cdot E\{\tau_{\text{sft}}\} \cdot P_{\text{cf}}(1 - P_{\text{cf}})^{k-1} \quad (4.20)$$

where $P_{\text{cf}} = (1 - \text{BER})^{R \cdot \tau_f}$ is the correct transmission probability and BER and R are the bit error rate and transmission rate, respectively. $E\{\tau_{\text{n-cci}}\}$ is defined similarly, with $E\{\tau_{\text{sft}}^{\text{sa}}\}$ replaced by $E\{\tau_{\text{sft}}\}$.

In Fig. 4.3, the SFT delays with and without CCI, for both AWGN and flat fading channels are revealed. As the $\frac{\tau_f}{E\{\tau_s\}}$ ratio increases, the CCI case becomes highly advantageous. This advantage becomes significant when $\tau_f > 5 \cdot E\{\tau_s\}$, following an exponential trend.

With fading, the improvement in delay by CCI is higher than the improvement in delay by CCI under AWGN only. Hence, CCI incorporation is even more useful, from a delay perspective, under fading. However, the improvements in delay vanish under low PU arrival rates or unconventionally low $\frac{\tau_f}{E\{\tau_s\}}$ ratios.

Fig. 4.4 illustrates the impact of SU node density on SFT delay. Here, the SU

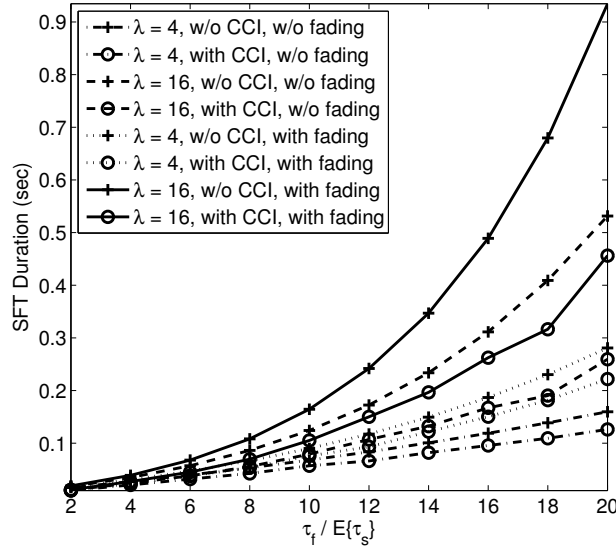


Figure 4.3: SFT delay with and without CCI, with respect to $\frac{\tau_f}{E\{\tau_s\}}$ ratio, for different PU arrival rates.

nearest neighbor link distance probability distribution function given in Section 4.5.1 is used. As this distance is changed, statistics of E_b/N_0 and BER are determined assuming constant SU transmit power. SFT delay is investigated for two α values from Section 4.5.1, namely, 4 and 1, for the SU node density $\lambda_{SU} = 0.01$. These packet arrival rates result in total PU traffic values of 16 and 4 at that SU density. The SNR and total PU packet arrival rates for other λ_{SU} values are calculated according to their R_c values. Accordingly, in Fig. 4.4, increasing the λ_{SU} decreases the SFT delay, since SU nodes become closer and links bear less PU traffic. CCI provides same SFT delay as the non-CCI case, at a lower SU density. This result is important for network design choices, as moderate node density and CCI incorporation could perform equally well with a design without CCI yet having a higher the node density.

Fig. 4.5 is a further investigation of the economic aspects of CCI incorporation. Inherently, the CCI has a hardware cost. Besides, an additional cost for the network-wide CCI channel allocation and maintenance is also very likely. These factors make it hard to estimate the cost of incorporating CCI. To this end, we investigate the total equivalent node cost per meter-square to achieve a given SFT delay, for varying

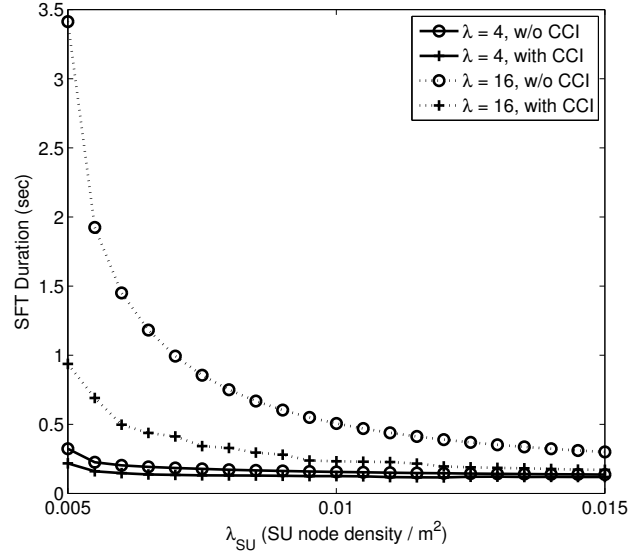


Figure 4.4: SFT delay with respect to λ_{SU} , with and without CCI. $E_b/N_0 = 10$ dB at $\lambda_{\text{SU}} = 0.005$, λ_{PU} is assumed to be 0.0125.

CCI costs in terms of the regular full node cost. The non-CCI case can become advantageous for certain CCI cost values, as the SFT delay constraint is relaxed. Logically, on the low SU density side, to meet a not very strict SFT requirement, increasing node density should be the choice, since it has higher incremental utility in

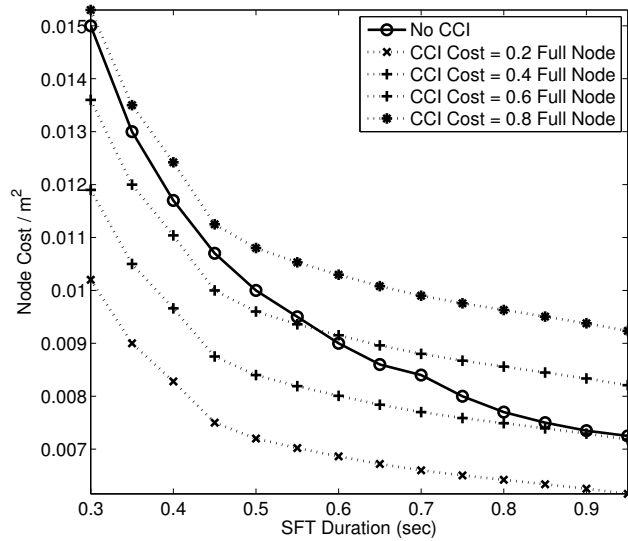


Figure 4.5: Equivalent node cost, to achieve a given SFT delay requirement, for different CCI cost to full node cost ratios.

delay gains by enhancing the received power (and consequently BER) of transmission. For stringent SFT requirements, incremental utility of increasing node density on communication performance is monotonically decreasing. In that case, one needs CCI to obtain additional material improvements in delay. The conclusion is that, to achieve strict SFT delay requirements, CCI is the key to a reduced cost, whereas with loose SFT constraints, solely increasing node density could be less costly.

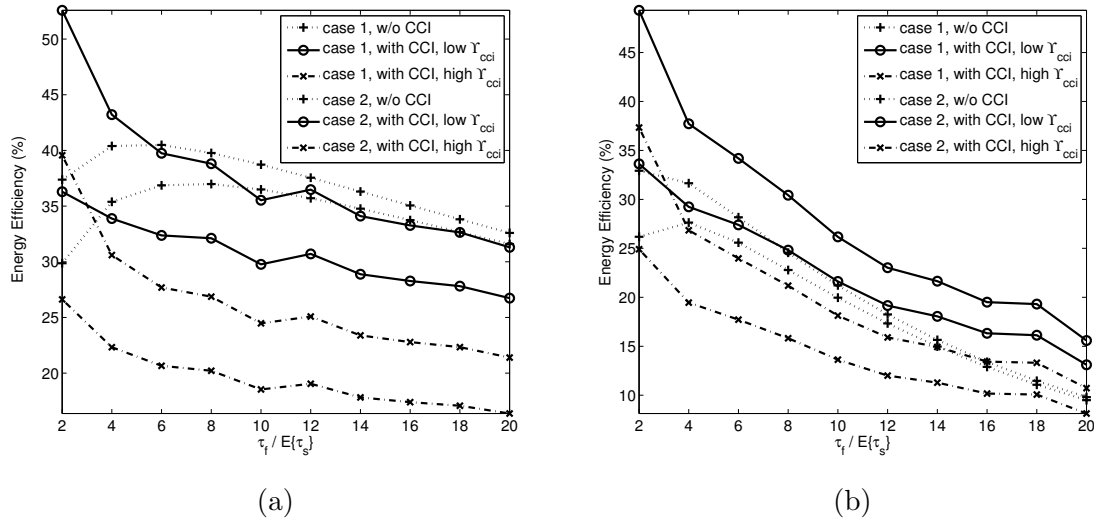


Figure 4.6: Energy-efficiency with and without CCI, for two transmission and sensing power level pairs under low (a) and high (b) PU traffic.

4.5.3 Evaluation of Energy-Efficiency With and Without CCI

Energy-efficiency depends on spectrum sensing period τ_s , frame transmission period τ_f , and power consumption levels in these periods. We use two different spectrum sensing Υ_s and frame transmission Υ_f power level pairs, i.e., $\{\Upsilon_{s_1} = 180 \text{ mW}, \Upsilon_{f_1} = 320 \text{ mW}\}$, and $\{\Upsilon_{s_2} = 240 \text{ mW}, \Upsilon_{f_2} = 200 \text{ mW}\}$, two CCI power consumption levels ($\Upsilon_{cci_1} = 100 \text{ mW}$ and $\Upsilon_{cci_2} = 300 \text{ mW}$), and assume fixed $E\{\tau_s\}$ of 5 ms. The effects of CCI on energy-efficiency are nontrivial, driven by the fact that CCI consumes energy on its own, yet also helps to decrease SFT delay; consequently saving energy to some extent. Fig. 4.6(a) depicts the energy-efficiency for low traffic scenario ($\lambda = 2$). When CCI consumes more power than regular radio transmission, energy-efficiency is almost

always degraded by CCI; meaning that under low traffic, improvements brought by CCI cannot compensate its own energy consumption. However, when CCI consumes less power than regular radio, the node either outperforms or performs equally well in terms of energy-efficiency as its non-CCI counterparts. Fig. 4.6(b) depicts the high traffic ($\lambda = 8$) scenario, in which low CCI power consumption cases usually outperform the regular non-CCI setups and high power CCI cases either underperform or perform equally well as regular non-CCI setups.

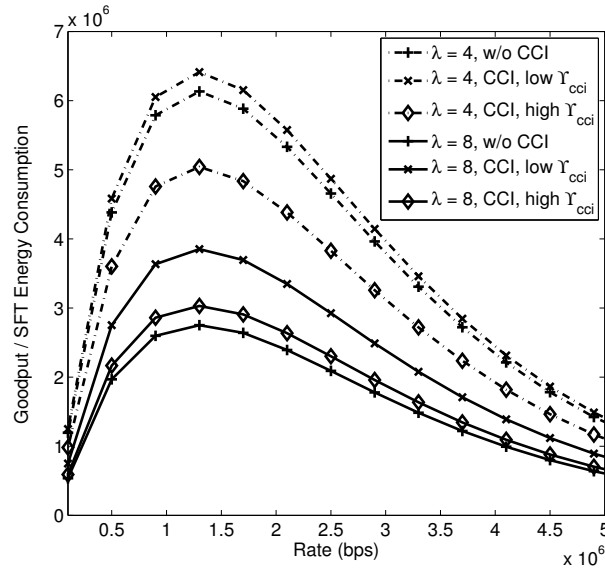


Figure 4.7: Goodput/(SFT energy consumption) ratio (in bits/(second·Joule)), under different PU traffic and CCI power levels.

4.5.4 Evaluation of Spectrum-Efficiency With and Without CCI

The goodput performance of CR equipped networks with and without CCI is assessed in Fig. 4.7. We consider the goodput per energy consumed for an SFT as the spectrum-efficiency metric. This ratio provides more meaningful results than bare goodput, assessing whether CCI brings in extra goodput in an energy-efficient manner. Considered scenario comprises two PU traffic cases and two CCI power consumption levels. In the lower traffic case, depicted by dashed lines, CCI inclusion with power consumption equal to that of regular transmission results in lower goodput per

energy, whereas CCI with low power consumption and non-CCI cases perform almost identically. Under high traffic, however, the CCI that requires lower power consumption than regular transmission brings in energy-efficient goodput, whereas high power demanding CCI and non-CCI cases perform similarly. It is revealed that if the CCI energy consumption per node, including CCI radio consumption and common control channel maintenance cost across the network or local area, cannot be kept at a lower level than regular transmission consumption, then goodput improvements through CCI comes at the cost of increased energy consumption.

4.5.5 Evaluation of Event Sensing Performance With and Without CCI

Distributed event sensing distortion results are presented in Fig. 4.8. Sensing noise σ_k^2 is taken to be 0.01 for all sensors. Estimation interval τ_e is assumed as 1 sec. Number of samples per frame S_p is set to 10. Achieved Mean Square Error (MSE) with CCI is expectedly lower than the one achieved without CCI. Nevertheless, the improvements in MSE through CCI heavily depend on $\frac{\tau_f}{E\{\tau_s\}}$ ratio and PU traffic

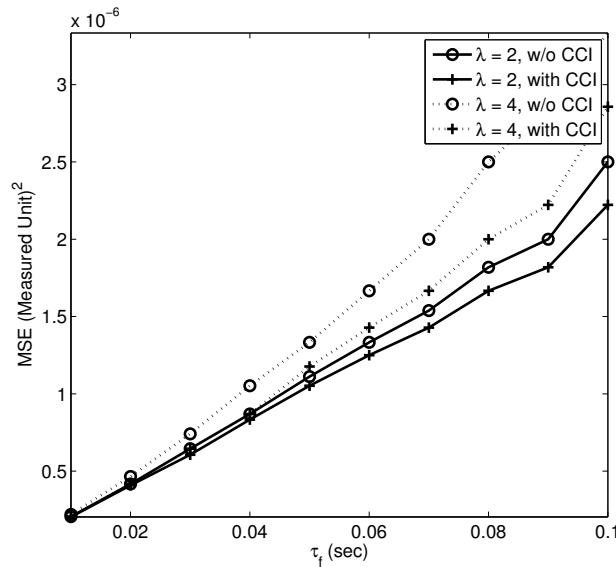


Figure 4.8: Distributed sensing distortion performance with and without CCI, under different PU arrival rates.

conditions. Since sensor nodes have lower ranges, we consider a lower PU traffic than their CRN counterparts. Gains in MSE with CCI are not that bright, when τ_f is not an order greater than τ_s . Improvements in MSE with CCI increase with higher PU traffic, however not as drastically as their CRAHN equivalents.



Chapter 5

SELF-OPTIMIZATION FOR CAPACITY IN LOW TERAHERTZ BAND INDOOR WIRELESS NETWORKS

5.1 Introduction

Throughput and network capacity requirements of the Beyond Fourth Generation (B4G) systems are very challenging to accomplish within the highly congested sub-6 gigahertz (GHz) bands. To reach data rates on the order of 10 gigabits per second (Gbps), as the current spectral efficiency methods are already at its limits, vacant portions of the spectrum need to be leveraged. Although higher frequencies have been out of general usage due to their greater path losses, ultra-densification of devices is likely to make them the new spectrum norm in the near future.

Low terahertz (THz) band, centered around 0.3 THz, could be especially suitable for indoor Fifth Generation (5G) networks [91]. To study achievable rates and overall network capacity for indoor 5G applications, we consider a network equipped with Full Dimensional MIMO (FD-MIMO) capabilities, operating in the low terahertz (THz) band. Such a network is very suitable for the requirements of upcoming indoor 5G applications, for the reasons listed below:

- Given the abundance of bandwidth at low THz band and availability of idle spectrum, much higher data rates can be achieved.
- FD-MIMO capabilities in the form of an antenna array provides flexibility of beam-forming in almost any direction.
- Given the high frequency of operation, even an array of 64 antennas (in a 8×8 formation) would only require an area of $0.4 \text{ cm} \times 0.4 \text{ cm}$, making it very feasible

for an indoor environment in contrast to their lower frequency counterparts [92].

- The path loss is very high in the low THz band, hence needed antenna radiation is highly directional [93]. Although this is a challenge for coverage, directionality limits the intra-cell interference among User Equipments (UEs), which in turn gives UEs the flexibility to use the full channel bandwidth of that cell.

To that end, although such a network is promising for future 5G applications, there is a need for a self-optimizing approach that intelligently allocates a power budget among Base Station (BS) transmitters in the FD-MIMO arrays, such that overall network capacity is maximized while interference among communicating pairs is managed. In this chapter, we investigate self-optimization for capacity in FD-MIMO equipped indoor wireless networks that operate in the low THz band. Main contributions are the following:

- A self-optimization approach for capacity in realities of the low THz band has been developed for the first time.
- A self-optimization protocol for capacity have been developed, its complexity vs. utility in terms of additional capacity in low THz environment has been studied.
- Our research group's previous work [94] on FD-MIMO based capacity analyses for indoor 5G networks have been extended to randomly deployed networks with self-organization for capacity.

The remainder of this chapter is organized as follows. System model is depicted in Section 5.2. In Section 5.3, mathematical expressions for capacity in FD-MIMO equipped low THz band indoor wireless networks are studied. Then, water filling-based capacity optimization is discussed and a self-optimization protocol thereof is proposed. Numerical results are provided in Section 5.4 to assess the benefits of this protocol.

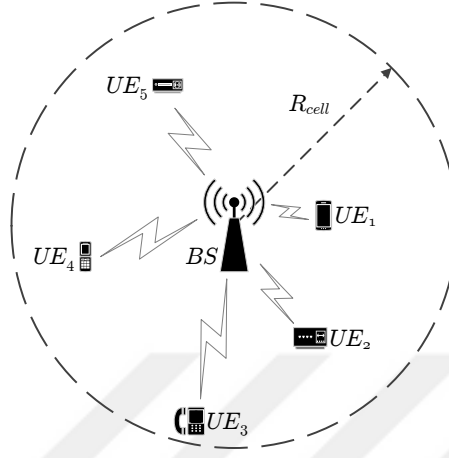


Figure 5.1: Assumed network model, with FD-MIMO capable BS at the center of the cell, UEs are deployed randomly

5.2 System Model

In this section, we first describe the network model that we employed while developing the self-optimization protocol for network capacity. The antenna radiation pattern and the interference model details are provided afterwards.

5.2.1 Network Model

For development of the self-optimization protocol for network capacity, we consider an indoor IoT network with a Base Station (BS) and a number of User Equipments (UEs). Assumed network is depicted in Fig. 5.1. Following characteristics are considered in our detailed network model:

- Both the BS and UEs are placed on an Euclidean plane $\mathbb{R}^2$ that is assumed to be indoor.
- BS is placed at the center of the Euclidean plane, whereas UEs, $UE_1, \dots, UE_N$, are randomly deployed.
- BS and UEs use the low THz band, with center frequency of 300 GHz and bandwidth B_c of 4.48 GHz [95, 96].

- BS is assumed to be equipped with FD-MIMO type antenna arrays, providing flexibility to the BS to communicate concurrently with N UEs.
- A planar antenna array can typically cover 120° . Hence, three arrays are placed in a triangular geometry to provide full 360° coverage of the cell [97].
- The BS is assumed to run a geographical location service periodically, to identify the coordinates of each UE_k [55].
- Each transmitter TX_k ($k = 1, 2, \dots, N$) in the array of the BS serves the corresponding UE_k with a direct beam.
- Each transmitter TX_k has a directional antenna radiation pattern, which is detailed in Section 5.2.2.

5.2.2 Antenna Radiation Pattern

As the low THz band is employed for communications, the need for highly directional beams arises. With such radiations, given limited interference caused by side lobes of the beam, all TX_k can theoretically use the full channel bandwidth for concurrent transmissions to all UEs within the cell. Link availability between TX_k - UE_k then depends on whether there is an obstacle on the Line of Sight (LoS). In case it is available, the link's achievable rate depends on its Signal-to-Interference-plus-Noise ratio ($SINR_k$). Interference is induced by the side lobes in beams of all other TX_i , $i \neq k$. It is possible to reduce this interference via techniques on diffraction limited optics [98]. Yet such techniques come at the expense of significant system complexity, hence are not considered in this chapter.

A *sinc* waveform is a good approximation to a highly directional radiation pattern, as in Fig. 5.2(a). However, as the ideal *sinc* returns to null, it does not represent a realistic transmitter. Hence, we model the antenna radiation pattern with a modified *sinc* function characterized by three parameters; Half power beam width β_{3dB} , main

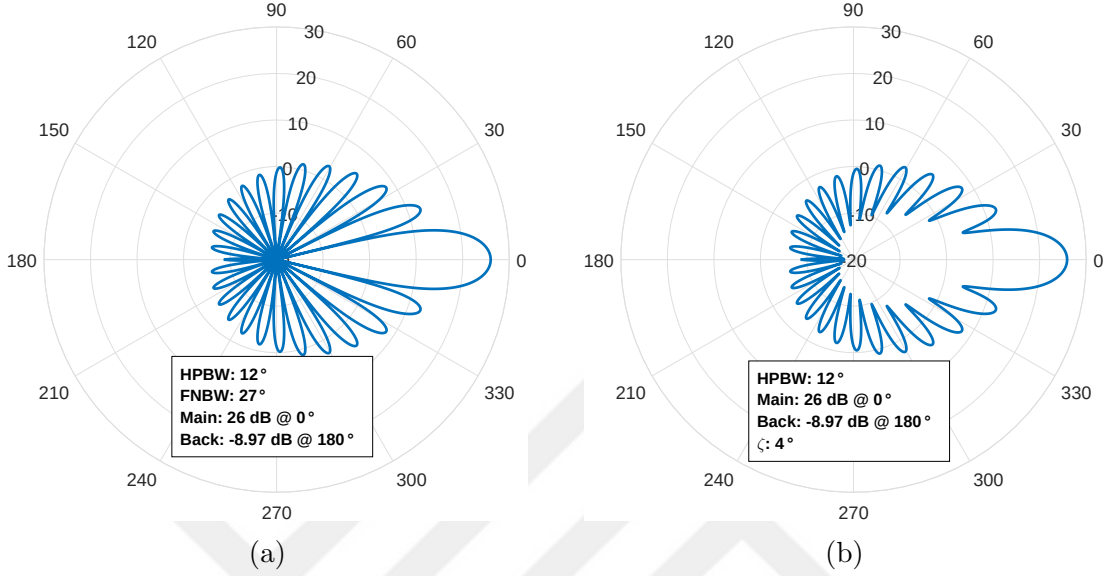


Figure 5.2: (a) Radiation pattern of a single antenna with ideal *sinc* function, directed at $\theta = 0$; (b) Realistic radiation pattern of a single antenna with modified *sinc* function, directed at $\theta = 0$

lobe gain constant G_t and damping factor ζ . Consequently, the radiation equation is

$$G(\theta) = \frac{1}{2\zeta + 1} \sum_{m=-\zeta}^{\zeta} G_t \operatorname{sinc}^2\left(\frac{0.8854}{\beta_{3dB}} \theta + m\right). \quad (5.1)$$

This pattern is illustrated in Fig. 5.2(b) and used in our further analyses for network capacity in low THz band.

5.2.3 Channel and Interference Model

It is assumed that UE_k is aligned with the main lobe of the radiation pattern of TX_k through beam alignment. Since transmissions are concurrent for all UEs on the same bandwidth B_c , UE_k faces interference from side lobes in beams of TX_i , $i \neq k$. As the low THz radiation is significantly directional as in Fig. 5.2(b), interference is insignificant when UEs are placed far apart from each other. However, when UEs are randomly placed, there is a need for a self-organizing mechanism, which identifies locations of UE_k , $k = 1, 2, \dots, N$ and allocates power P_k such that intra-interference

would be manageable and network capacity can be maximized. The interference imposed upon UE_k can be expressed as

$$I_{UE_k} = \sum_{i=1, i \neq k}^N P_i G_r G(\theta_{ik}) \left(\frac{c}{4\pi f r_k} \right)^n, \quad (5.2)$$

where P_i is the power allocated to TX_i for its directed beam towards UE_i , G_r is the constant gain of receiver equipment on all UEs, $G(\theta_{ik})$ is the gain of the beam between TX_i - UE_i in the direction of UE_k as per Eq. (5.1) (θ_{ik} being the angle between the directed beams of TX_i - UE_i and TX_k - UE_k couples), c is the speed of light, f is the frequency of transmission and r_k is the distance between the transmitter array and UE_k . It should be noted that the UEs and antenna arrays on the BS are assumed to be points on the $\mathbb{R}^2$ plane, and G_r is considered constant, irrespective of the angle θ_{ik} .

Obstacles have a significant impact on link availability, as they can absorb transmitted signals in the range of 10-30 dB in low THz band [93]. In our simulations in Section 5.4, we also consider randomly deployed obstacles. In case of an obstacle between the TX_k - UE_k LoS, this link is assumed to be unavailable and the self-optimization protocol allocates power to other transmitters in the array accordingly.

5.3 Self-Optimization for Capacity in Low THz Band Indoor Wireless Networks

5.3.1 Analysis Framework

To enable data rates that are expected to be demanded by future 5G applications, transmissions in our system model take place on the same full coherence bandwidth B_c of 4.48 GHz. Under these conditions, network capacity scales with $SINR$ values of all concurrent transmissions within the cell. The SINR at the receiver of UE_k can

be calculated as

$$SINR_k = \frac{P_k G_r G(0) \left(\frac{c}{4\pi f r_k} \right)^n}{N_0 + \sum_{i=1, i \neq k}^N P_i G_r G(\theta_{ik}) \left(\frac{c}{4\pi f r_k} \right)^n}. \quad (5.3)$$

where $G(0)$ represents the maximum point of the directed beam as per the antenna radiation pattern between TX_k - UE_k . The total ergodic capacity of the low THz wireless network is represented as $C_{total} = B_c \sum_{k=1}^N \log_2 (1 + SINR_k)$ [99]. Therefore, the optimization problem for total network capacity by allocating a total power budget of P_{total} to TX_k - UE_k pairs, $k = 1, 2, \dots, N$, can be formulated as

$$\begin{aligned} \max_{P_k} \quad & \sum_{k=1}^N \log_2 \left(\frac{N_0 + \sum_{i=1}^N P_i G_r G(\theta_{ik}) \left(\frac{c}{4\pi f r_k} \right)^n}{N_0 + \sum_{i=1, i \neq k}^N P_i G_r G(\theta_{ik}) \left(\frac{c}{4\pi f r_k} \right)^n} \right) \\ \text{sub. to} \quad & \sum_{k=1}^N P_k \leq P_{total}, \\ & P_k \geq 0, \end{aligned} \quad (5.4)$$

where constant B_c has been omitted and $\theta_{kk} = 0$ by definition. It is clear that the optimization function from Eq. (5.4) is non-convex and very hard to solve. Therefore, optimal P_k values cannot be determined through straightforward water-filling [100]. To turn the optimization problem into convex, interference term in the denominator of Eq. (5.4) is often treated as constant [101, 102]. Following this assumption, the Lagrangian is given as in Eq. (5.5) below.

$$\mathcal{L}(P_k, \lambda) = \sum_{k=1}^N \log_2 \left(\frac{N_0 + \sum_{i=1}^N P_i h(i, k)}{N_0 + \sum_{i=1, i \neq k}^N P_i h(i, k)} \right) - \lambda (P_{total} - \sum_{k=1}^N P_k) \quad (5.5)$$

On the other hand, partial derivative of Lagrangian is provided in Eq. (5.6) below,

$$\frac{\partial \mathcal{L}(P_k, \lambda)}{\partial P_k} = \frac{1}{\ln(2)} \left(\frac{h(k, k)}{N_0 + \sum_{i=1}^N P_i h(i, k)} - \sum_{j=1, j \neq k}^N \frac{\gamma_j P_j h(j, k)}{\sum_{i=1}^N P_i h(i, j) + N_0} \right) - \lambda \quad (5.6)$$

where $\gamma_k = h(k, k) / (\sum_{i=1, i \neq k}^N h(i, k) P_i + N_0)$. Consequently,

$$P_k = \max\{\xi - 1/\gamma_k, 0\}, \quad (5.7)$$

where

$$\xi = \frac{1}{\lambda \ln(2) + \left(\sum_{j=1, j \neq k}^N \frac{\gamma_j P_j h(j, k)}{\sum_{i=1}^N P_i h(i, j) + N_0} \right)} \quad (5.8)$$

5.3.2 Proposed Approximate Water Filling (AWF) Protocol

It is possible to find the optimal P_k^* values, as per Eq. (5.6-5.7), through iterations as in [101]. The Approximate Water Filling (AWF) procedure to find optimal P_k^* is detailed in Algorithm 1.

5.4 Numerical Results

We study the performance and benefits of Approximate Water Filling (AWF) algorithm when applied to the system model described in Section 5.2. To that end, simulations have been performed on MATLAB, using the parameters in Table 5.1. Gain of antenna used is 26 dBi with a HPBW of 12°, which is easily achievable in THz Band [93, 103]. The path loss of 2 is statistically determined in previous experiments [95]. 0 dBm of total BS power is achievable as shown in [104].

Accordingly, the benefit of the AWF algorithm on total network capacity with respect to number of UEs N can be seen in Fig. 5.3(a). With $P_{total} = 1$ mW and cell radius $R_{cell} = 4$ m, the AWF algorithm brings $\sim 5\%$ to overall network

Algorithm 1 Approximate Water Filling (AWF) for downlink capacity maximization in a low THz cell

Inputs: P_{total}, r_k, θ_k for $k = 1, 2, \dots, N$; Δ as step size

Outputs: P_k for $k = 1, 2, \dots, N$

Once the BS finds about locations of all UE_k , it runs this algorithm to maximize network capacity through power allocation

compute all $\theta_{i,k}$, $k = 1, 2, \dots, N$

initialize power allocation as $P_{k,1} = P_{total}/N$

compute all $\gamma_{k,1} = h(k, k)/(\sum_{i=1, i \neq k}^N h(i, k)P_{i,1} + N_0)$ for this first iteration, $k = 1, 2, \dots, N$

compute first $\xi_1 = \frac{1}{N} \left(P_{total} + \sum_{k=1}^N \frac{1}{\gamma_{k,1}} \right)$

assign $l = 2$

while l **do**

 Iterate $P_{k,l}$ until convergence in this loop

compute $P_{k,l} = \max\{\xi_{l-1} - \frac{1}{\gamma_{k,l-1}}\}$

compute $\xi_l = \xi_{l-1} + \frac{\Delta}{N} \left(P_{total} - \sum_{i=1}^N P_{k,l} \right)$

if $\sum_{k=1}^N |P_{k,l} - P_{k,l-1}| < 10^{-10}$ **then**

finalize $P_k = P_{k,l}$ for $k = 1, 2, \dots, N$

break

else

$l = l + 1$

end if

end while

return P_k for $k = 1, 2, \dots, N$

capacity, compared to equal allocation of power to all TX_k - UE_k pairs. The utility of AWF increases to $\sim 20\%$ contribution to total network capacity in the range of $15 \leq N \leq 30$. This phenomenon can be explained by the following facts:

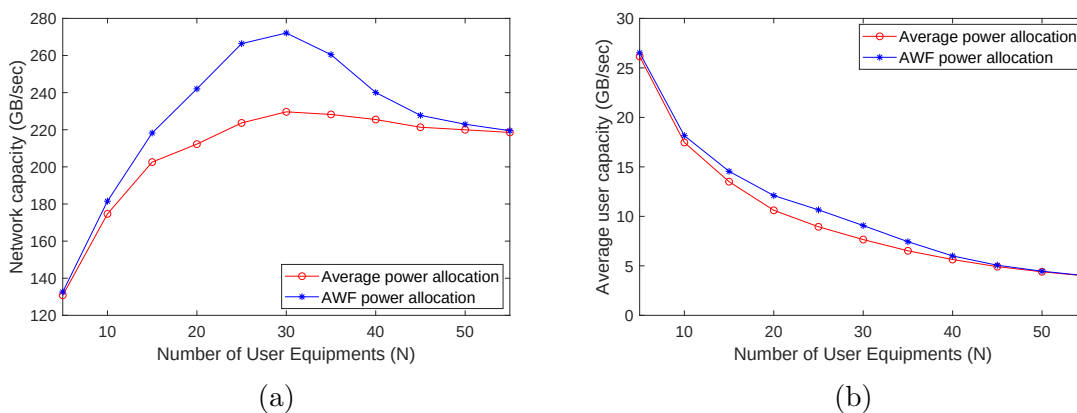
- Utility of AWF is not very high when $N < 10$, as random deployment seldom puts two nodes in close proximity. Given the high directivity in beams, a deployment with low N usually does not generate significant interference among UEs, thereby limiting achievable utility by AWF.

Table 5.1: System parameters used in simulations for capacity optimization in low THz band.

Parameter	Symbol	Value
Bandwidth	B_c	4.48 GHz
Frequency	f	300 GHz
Speed of light	c	299792458 m/s
White noise	N_0	-70 dBm
Gain	G_r, G_t	26 dBi
Pattern smoothing	ζ	4°
HPBW	β_{3dB}	12°
Path loss exponent	n	2
Transmitter power	P_t	0 dBm
Network radius	R_{cell}	4 m

- When $15 \leq N \leq 30$, the interference among UEs starts to kick-in, where the AWF has potential to add much higher value to total network capacity.

Interestingly, network capacity that is obtained by P_k , $k = 1, 2, \dots, N$ delivered by AWF starts to decrease after $N > 30$. In fact, number of iterations as well as the run-time until convergence of AWF increases exponentially with N , as can be seen in Fig. 5.4 for $N \leq 30$. However, given the requirement of tolerable delay in self-organization


 Figure 5.3: (a) Total network capacity with respect to N for equal power allocation vs. AWF; (b) Average user capacity with respect to N for equal power allocation vs. AWF

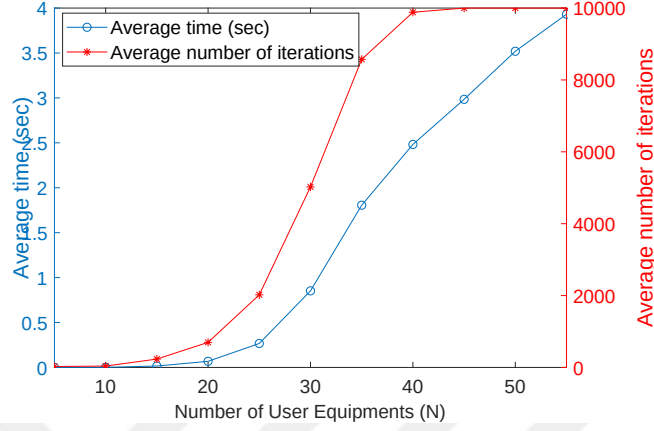


Figure 5.4: Average number of iterations and algorithm run-time with respect to N

algorithms in 5G, a maximum of 1000 iterations and 4 seconds of run-time has been allowed in our simulations. In a significant portion of random deployments where $N > 30$, AWF has not fully converge within given limits. This limit is also the reason why run-time and iteration number curves in Fig. 5.4 turn from convex to concave. In such cases, AWF returns a sub-optimal solution within given time. Under a fixed time limit, performance of these sub-optimal solutions deteriorate fast with increasing N above 30, as seen in both Fig. 5.3(a) on network capacity and in Fig. 5.3(b) on average user capacity.

Similar water-filling algorithms might deliver 2-4 fold benefit on network capacity at $\text{SNR} > 20$ dB, in traditional sub-6 GHz spectrum. However, majority of the network capacity in low THz setting is driven by bandwidth, considering that SINR is usually $\text{SINR}_k < 0$ dB. To that end, gains from AWF are limited by nature, as shown in our case. However, to contribute in progress towards data rate requirements of 5G networks, AWF could prove very useful, with a manageable complexity, in indoor low THz networks with a maximum number of connected users $N = 30$.

Chapter 6

HARVESTING-THROUGHPUT TRADE-OFF FOR WIRELESS-POWERED SMART GRID IOT APPLICATIONS: AN EXPERIMENTAL STUDY

6.1 Introduction

Internet of Things (IoT) provides a link between cyber and physical world by connecting devices over the Internet [105]. One of the key enabling technologies for IoT is Wireless Sensor Networks (WSNs), where wireless sensor nodes monitor the physical world, digitize their observations and send their readings to an Access Point (AP) that is connected to the Internet. Such a set-up enables sensor observations to be accessed remotely, whenever needed. However, operating wireless sensor nodes in an IoT setting with energy-limited batteries could be challenging as they are often deployed in large numbers and require frequent battery replenishments. In this context, Energy Harvesting (EH) come into prominence to extend the lifetime of wireless devices by exploiting a stray source or converting useful energy from one form to another.

When an industrial facility in Smart Grid (SG) context is envisioned, continuity and reliability in operation are of utmost importance, as interruption in power provisioning cannot be tolerated in most cases. This necessitates real time monitoring of power systems over the Internet to better anticipate potential threats before they yield undesired consequences. Considering the harsh conditions in transformer rooms, which are the main pillars of SG applications, battery replenishments for wireless sensor nodes cannot be carried out without a system shut down, which in turn would interrupt power provisioning. Hence, EH approaches can be employed in SG context to altogether eliminate lifetime constraints of the sensors, evolving them to battery-

less and self-operable entities, while ensuring reliable and interruption-free operation of SG pillars.

In simple terms, EH alleviates the limited lifetime problem of energy-constrained devices such as wireless sensor nodes [106]. The availability of conventional harvestable sources, namely light propagation, temperature gradient, motion variations and electromagnetic waves, are significantly affected by environmental variables or other external factors. Given this randomness, recent research interests on EH is focusing on investigating application specific solutions that are reliable, durable and profitable. In SG applications, alternatives of EH are highly limited for transformer room scenarios given the lack of ambient resources (e.g., light, air flow and/or temperature gradient) thereof. In addition, as there is neither noise nor motion variations in transformer rooms, electromagnetic (EM) fields in abundance are the only options for EH. When EM field-based techniques are envisioned for these settings, electric-field (E-field) oriented methods stand out compared to their magnetic-field (M-field) based counterparts, as they do not necessitate any galvanic contact with the transformer to harvest energy. To that end, E-field EH (EFEH) method is less complex, safer and less costly solution for battery-less operations of wireless devices in monitoring SG assets.

Preliminary works on EFEH were first performed on middle and high voltage transmission lines, exploiting the E-field in abundance [107]. The empirical findings revealed the efficiency of EFEH in providing advanced condition monitoring and remote control for SG architectures. EFEH has also been applied to low voltage systems as mounting commercial AC power cords with conductive metallic sheaths [108]. Results disclosed that it is possible to constitute an applicable procedure for applications in which the E-field intensity is considerably low. As an alternative to the previous studies, [109] has brought a new perspective with the proposed parallel plate model. This work introduced placement of a copper plate between troffers and ground plain to exploit the E-field flow around overhead fluorescent fixtures. The plate acts as a generic current source, and supports autonomous operation of wireless devices to be

attached. The idea presented in [109] has been taken as the basis of the experimental set-up of this chapter as well.

While studies that characterize the harvestable energy profile through EFEH are at early stages, many other works focused on optimal utilization of the harvested energy by the sensing and communicating entities. [110] is an excellent introduction to that matter, laying the various cases for optimal use of the scavenged energy in terms of different utility functions such as maximum throughput or minimum outage probability. [111–118] are all great examples on optimal utilization of the extracted energy in wireless and cognitive radio ad hoc networks. However, none of these studies are based on neither an EFEH scheme nor an experiment-backed real energy profile. Additionally, these works consider that entirety of the exploited energy is consumed for communications; neglecting the energy consumed for nodal operations, such as waking up, sensing and processing. These operations actually account for a significant portion of the harvested energy according to our experimental results.

In this chapter, we investigate an EFEH method for a point-to-point communication architecture used in an IoT-assisted transformer room monitoring application. Through an experimental set-up, for the first time in the literature, we reveal the profile of harvestable energy in a transformer room with our proposed EFEH method. Building on this energy profile, in the second part, we theoretically study the harvesting-throughput trade-off for an EFEH wireless sensor node that is employed in a SG setting. The sensor node is assumed to periodically send its readings, in a single hop, to a distant AP that is connected to the Internet. To that end, the sensor node is following a “harvest-then-transmit” cycle; where it first harvests energy of amount E_H during a period of T_H ; then uses this energy for wake-up, sensing, processing and transmission operations. The entire cycle is completed in a total duration of T_{Total} and is periodically repeated. The aim is to ensure battery-less operation of the sensor node and to maximize the throughput function θ (i.e. the expected number of bits that are reliably conveyed to the AP in a single “harvest-then-transmit” cycle) while complying with a requirement on minimum frequency of reporting related

to conditions of the transformer room. We reveal the changes in θ with respect to harvesting period T_H , transmission period T_T and other critical system parameters.

The remainder of this chapter is organized as follows. We first introduce the system model in Section 6.2. Basic principles of EFEH are reviewed alongside a detailed description of the experimental set-up that we utilized for EFEH in Section 6.2.1. Theoretical model on nodal operations and communication set-up employed in our work are described in Section 6.2.2. In Section 6.3, we define a throughput objective function θ to shed light on harvesting-throughput trade-off specific to IoT-assisted SG applications. In Section 6.4, we find the optimal harvesting period T_H and transmission period T_T which maximize throughput θ , and investigate how they vary with respect to changes in critical network parameters.

6.2 System Model

In this section, we first describe the experimental set-up that we utilized for EFEH in a transformer room, as a case study for IoT-assisted SG applications. System model details of the point-to-point communicating wireless sensor node that is powered with harvested energy are provided afterwards.

6.2.1 Experimental Set-up for the Proposed EFEH Method

According to electrostatics, electric charges distributed in a closed surface result in an E-field. In AC, this time varying field yields a displacement current, I_d , which can be defined with Maxwell's equation $\mathbf{I}_d = \epsilon \iint_s \frac{d\mathbf{E}}{dt} d\mathbf{S}$, where ϵ is absolute permittivity, and $\mathbf{E}$ is the electric field intensity. The electric charges are dispatched and then collected in a storage element, C_s , via I_d . The energy stored in C_s can be expressed as $E = \frac{1}{2}CV_s^2$, where C is capacitance of C_s and V_s is the voltage accumulated. As energy is gathered by exploiting the ambient field, this method can be referred to as Electric-field Energy Harvesting (EFEH) [107–109].

Fig. 6.1(a) illustrates the diagram of the set-up that we built in a transformer room to harvest energy through EFEH method. A copper plate, i.e., the E-field

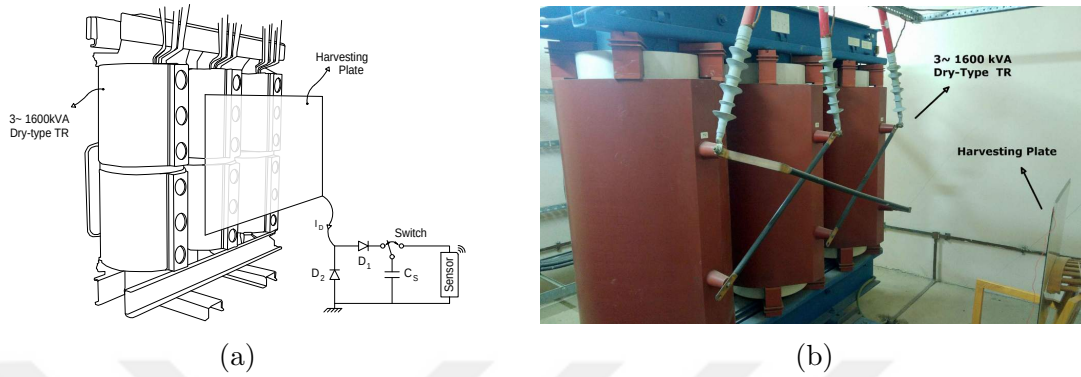


Figure 6.1: (a) Representative diagram of the proposed EFEH method; (b) Experimental set-up of the proposed EFEH method.

energy harvester, is placed between the concrete wall and the primary side of a 3-phase 1600 kVA dry-type transformer. The main objective of this plate is to collect the transformer (i.e. E-field) induced AC charges in C_s by draining I_d , until the gathered voltage V_s becomes adequate for sensor to perform wake up, sensing, processing, and data transmission processes. The diodes, D_1 and D_2 , serve both to convert the harnessed energy from AC to DC, and to prevent it from backfeeding. The given ‘Switch’ model refers to an autonomous connection circuit to be employed for interchange between the operation stages, i.e., for the regulation of energy usage. This circuit allows charge conveyance when the accumulated energy E_H is high enough for transmission, and switches off the sensory circuit when the voltage V_s on C_s drops below a certain threshold. This operation not only prevents C_s from discharge down to 0 V, but also enables better throughput by shortening the time spent on harvesting stage [108].

The actual implementation of this set-up is depicted in Fig. 6.1(b). As shown, a copper plate of $50 \times 50 \text{ cm}^2$ size is used in the set-up as an E-field energy harvester. The location of the plate, or the safest possible distance for operation, is determined ideally to take advantage of capacitive coupling. As a result of high tension presented on the operation area, sharp corners of the plate are smoothened to avoid possible drawbacks related to corona-like electrical discharges. Voltage and energy accumulated on C_s with respect to time are revealed in detailed later in Section 6.4.

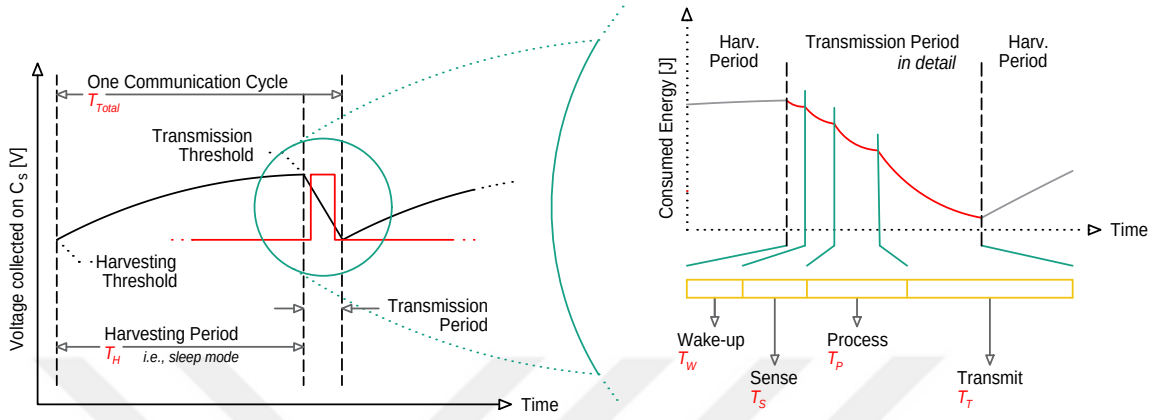


Figure 6.2: Operation cycle of the proposed EFEH method.

6.2.2 Nodal Operation and Communication Model

The wireless sensor node is assumed to be fed with the energy harvested through the proposed method of EFEH to carry out its nodal operations and communicate with an AP in a single hop, regarding the conditions of the transformer room and the transformer itself. This would enable notification of an upper level authority for decision-making procedures. By this means, preclusive actions can be performed against any unexpected events. With these actions, interruption-free operation of the SG medium can be guaranteed. The following set of facts and assumptions are taken into account regarding the operations of this node:

- The node is fed with the energy scavenged by the EFEH method, profile of which is determined through the experimental set-up described in Section 6.2.1. It is assumed that there is no other external source of energy.
- The node's operations follow a cyclical pattern of duration T_{Total} . Each cycle starts with a harvesting period, followed by a transmission period, as depicted in Fig. 6.2.
- The node operates in a half-duplex fashion given its hardware limitations, meaning it can only be in harvesting or transmission mode at any time [116].

- In each cycle, EFEH is carried out for a duration of T_H , accumulating a total energy of E_H .
- Transmission period consists of wake-up, sensing, processing and transmission operations, as in Fig. 6.2. These operations occupy durations of T_W , T_S , T_P and T_T in each cycle, respectively.
- The node is assumed to have data to send throughout the period T_T in each cycle.
- Total duration of one cycle is $T_{Total} = T_H + T_W + T_S + T_P + T_T$.
- The node is assumed to be asleep during the harvesting period, consuming a constant power P_Z [119], resulting in an energy consumption of $E_Z = P_Z T_H$.
- The node is assumed to consume a constant $E_N = E_W + E_S + E_P$ amount of energy for its nodal operations in each cycle.
- Due to energy causality constraint, consumed energy by the node within a cycle for wake-up, sensing, processing and transmission cannot exceed the energy available at that cycle after harvesting period is completed [112].
- The node follows an offline transmission policy, i.e., it has full knowledge of the amount and arrival time of harvested energy [111]. In fact, considering very low variability of the E-field in a transformer room, we assume the harvested energy profile does not change in time.
- The communication channel is assumed to be imperfect with Rayleigh fading.

6.3 Harvesting-Throughput Trade-off Problem Formulation

Building on the system model presented in Section 6.2, in this section we design one communication cycle of our EFEH-powered node. In each cycle, there are three

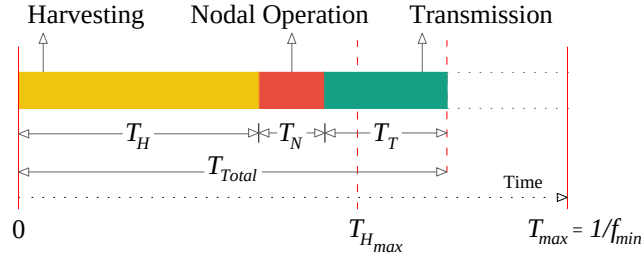


Figure 6.3: Harvest-then-transmit cycle of the EFEH-powered node.

orthogonal periods reserved for energy harvesting, nodal operations and transmission, as in Fig. 6.3. These periods serve the following:

- During the period T_H , the node exploits the E-field around the transformer, and subsequently charges C_s , reserving energy of amount E_H . Part of this reserve is already used during T_H , as the node consumes energy of amount $E_Z = P_Z T_H$ during sleep.
- $T_N = T_W + T_S + T_P$ is allocated for wake up, sensing and processing operations. Both the duration T_N and energy consumption E_N of this period are constants [119].
- In the last period of the cycle, the node has an energy budget of $E_H - E_Z - E_N$, which it needs to allocate in terms of transmission power P_T and duration T_T , to send the maximum expected number of bits to the AP.

As part of the offline transmission policy, gathered energy E_H in harvesting period T_H is completely utilized within the same cycle, for energy consumptions in sleep mode during harvesting, nodal operations and transmission. Hence, we investigate one communication cycle of the node in operation, and look for optimal T_H and T_T values that would maximize the expected number of bits sent to the AP per second.

The harvesting period T_H determines the energy captured in C_s . As T_H is extended, scavenged energy E_H increases non-linearly, with a decreasing rate as time progresses. There is an upper bound on the voltage V_s that can be accumulated on

C_s , which in turn limits the maximum amount of E_H . To that end, there is no logical reason to extend harvesting period beyond $T_{H_{max}}$ when V_s saturates, as seen in Fig. 6.3. Initial energy reserve of E_H is reduced to $E_H - E_Z - E_N$ by the time the transmission period is started. During this period, the node needs to find an optimal T_T (hence P_T) to allocate this energy budget, maximizing the expected throughput.

In our optimization problem, the objective function is defined as the throughput of the communication channel between the wireless sensor node and the AP, denoted as θ :

$$\theta = \frac{T_T \text{Blog}_2(1 + \gamma)(1 - \text{BER})^{(R_b T_T)}}{T_{Total}}, \quad (6.1)$$

where $T_{Total} = T_H + T_N + T_T$, BER is the bit error rate, B is the communication bandwidth, γ is the received signal-to-noise ratio (SNR) and $R_b = \text{Blog}_2(1 + \gamma)$ is the assumed channel bit rate. The optimization problem is formulated as:

$$\begin{aligned} \max_{T_H, T_T} \quad & \frac{T_T \text{Blog}_2(1 + \gamma)(1 - \text{BER})^{(R_b T_T)}}{T_{Total}} \\ \text{s.t.} \quad & T_{Total} \leq \frac{1}{f_{min}}, \end{aligned} \quad (6.2)$$

$$E_H > P_Z T_H + E_N$$

For this optimization problem, we account for the following two constraints:

1. T_{Total} must be less than or equal to inverse of minimum reporting frequency f_{min} , which depends on requirements of the IoT-assisted SG monitoring application.
2. T_H should be such that harvested energy E_H would not be less than sum of the energy consumptions during sleep mode ($E_Z = P_Z T_H$) and nodal operations (E_N).

For the reliability of the packet, we first determine bit error rate (BER), which depends on SNR per bit, γ_b . The packet length is $R_b T_T$. Hence, the packet error

probability is defined as $(1 - BER(\gamma_b))^{R_b T_T}$, where $R_b = B \log_2(1 + \gamma)$ is the bit rate of the channel and is assumed equal to the channel capacity. For Rayleigh fading channel, average BER for binary phase-shift keying (BPSK) is expressed as [88]

$$BER(\gamma_b) = \frac{1}{2} \left(1 - \sqrt{\frac{\gamma_b}{1 + \gamma_b}} \right). \quad (6.3)$$

SNR is defined as the ratio of the received signal power to the noise power, i.e., $\gamma = \frac{P_R}{N_0 B}$. Accordingly, $\gamma_b = \frac{E_b}{N_0}$ refers to energy per bit over noise power density, i.e.,

$$\frac{E_b}{N_0} = \frac{\gamma}{R_b} = \frac{P_R}{N_0 B R_b}, \quad (6.4)$$

where P_R is the power received by the access point [43]. P_R is defined as

$$P_R = \frac{P_T G_t G_r \lambda^2}{(4\pi)^2 L d^\nu}, \quad (6.5)$$

where λ is the wavelength, L is the receiver implementation losses, d is the distance between transmitter and receiver, ν is the path loss exponent, and G_t and G_r are the antenna gains of the transmitter and the receiver, respectively. The transmit power P_T is provided from the residue of the harvested energy in one communication cycle. Hence, it is determined as

$$P_T = \frac{E_H - E_Z - E_N}{T_T}. \quad (6.6)$$

Throughput expression, θ , corresponds to the expected number of bits that are correctly transmitted to the AP over one communication period. We find T_H and T_T values that would maximize θ for different parameters such as hop distance (d), path loss exponent (ν) and minimum reporting frequency (f_{min}) in Section 6.4. We denote the optimal harvesting and transmission periods as T_H^* and T_T^* , respectively.

Table 6.1: System parameters used for throughput optimization in proposed system [119].

Parameter	Value
P_Z	6.2×10^{-5} W
E_W	1.6×10^{-5} J
E_S	6.4×10^{-4} J
E_P	1.6×10^{-4} J
T_W	10^{-2} sec
T_S	6.7×10^{-2} sec
T_P	2.2×10^{-3} sec
B	10 kHz
N_0	4.17×10^{-21} W/Hz
λ	0.125 m
G_t, G_r, L	1

6.4 Numerical Results

In this section, we study the optimal harvesting and transmission periods, T_H^* and T_T^* , that would maximize the objective function θ , which corresponds to the throughput in between the sensor node-AP system in one communication cycle. The constants that are used for the numerical study of the θ function are given in Table 6.1.

6.4.1 Theoretical and Practical Battery-less Operating Regions

As the first step, we reveal the harvested energy E_H with respect to harvesting period T_H , according to experimental results from the EFEH set-up illustrated earlier in Fig. 6.1(a). Based on measurements from the actual transformer room visualized in Fig. 6.1(b), for a capacitance of $2.2 \mu\text{F}$, the accumulated voltage V_s on C_s increases up to 195 V and saturates when the harvesting time reaches to 900 sec as seen in Fig. 6.4(a). In a similar fashion, the accumulated energy E_H saturates at 40 mJ in the same time period. The shapes of V_s - T_H and E_H - T_H curves can be altered by employing different capacitors in the set-up, and accumulated energy at saturation would be higher for a bigger capacitance, assuming the voltage induced by the EFEH source (which acts

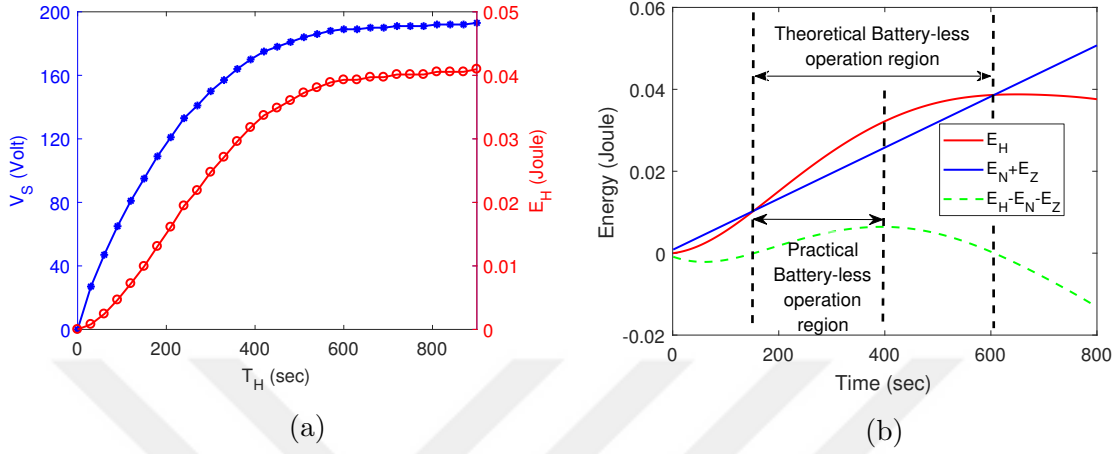


Figure 6.4: (a) Accumulated voltage V_s and energy E_H on capacitor C with respect to T_H ; (b) Accumulated energy E_H and consumed energy for sleep mode and nodal operations $E_N + E_Z$ with respect to T_H .

as a non-ideal current source) does not exceed the safe voltage limit of the capacitor.

The E_H - T_H curve is specifically important not only because it dictates the region(s) where the sensor node-AP system can operate in a battery-less manner, but also influences where this system would have the optimal throughput as defined by the θ function. For battery-less operation, E_H should be higher than the energy consumed for sleep mode during harvesting, E_Z , and the energy consumed for subsequent nodal operations of wake up, sensing and processing, i.e., $E_N = E_W + E_S + E_P$. As illustrated in Fig. 6.4(b), this is achieved when T_H lies in the 154-605 sec interval. This is the theoretical battery-less operation region of our system. However, as depicted by the dashed line in Fig. 6.4(b), there is no logical reason to extend T_H in a cycle beyond 390 sec, as the reserve energy $E_H - E_N - E_Z$ starts to decrease at that point. Hence, 154-390 sec interval is the practical battery-less operation region of our system. In summary, any transformer room monitoring application that requires a reporting frequency of more than once in 154 sec cannot be realized with the EFEH set-up in consideration. Additionally, if the application's reporting frequency requirement is less than once in 390 sec, the system is still better off from a throughput perspective to do reporting sometime in the interval of 154-390 sec in each cycle.

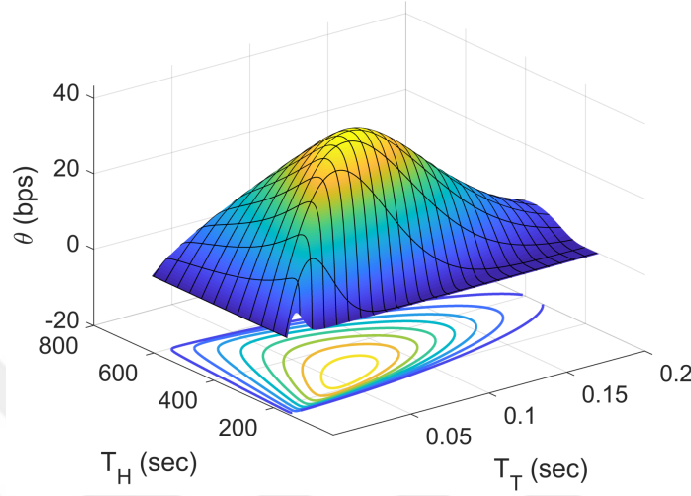


Figure 6.5: Value of objective function θ with respect to T_H and T_T .

6.4.2 Behavior of Throughput Function θ With Respect to T_H , T_T

As the second step, building on the identified E_H - T_H curve and the battery-less operation regions, we examine the variation of throughput function θ with respect to T_H and T_T . With the distance between the wireless sensor node and AP set to $d = 5$ m, a path loss exponent of $\nu = 2.5$ and a minimum reporting frequency of $f_{min} = \frac{1}{800}$; the θ function with respect to T_H and T_T is depicted in Fig. 6.5. Sections of θ for different T_H and T_T values are concave functions, and θ has one global maximum. For this case, optimal values that maximize θ are $T_H^* = 266.5$ sec and $T_T^* = 60.5$ msec. Consequently, $\theta = 43.24$ bps and in a duration of $T_{Total} = T_H + T_N + T_T = 266.64$ sec, sensor node can share 11.53 kbits with the AP in this optimal setting.

6.4.3 Sensitivity of Optimal T_H^* , T_T^* to System Parameters

As the third step, we study how T_H^* and T_T^* values that maximize throughput θ vary with respect to distance d between the sensor node-AP system and the path loss exponent ν of the channel. Fig. 6.6 reveals these optimal values for (a) $\nu = 2$, (b) $\nu = 2.5$ and (c) $\nu = 3$; where in each case d is swept from 3 m to 250 m. The first and foremost result is that T_H^* is very robust to changes in d and ν . In fact, T_H^* varies very slightly from 266 sec to 267 sec, as d is increased from 3 m to 250 m and ν is

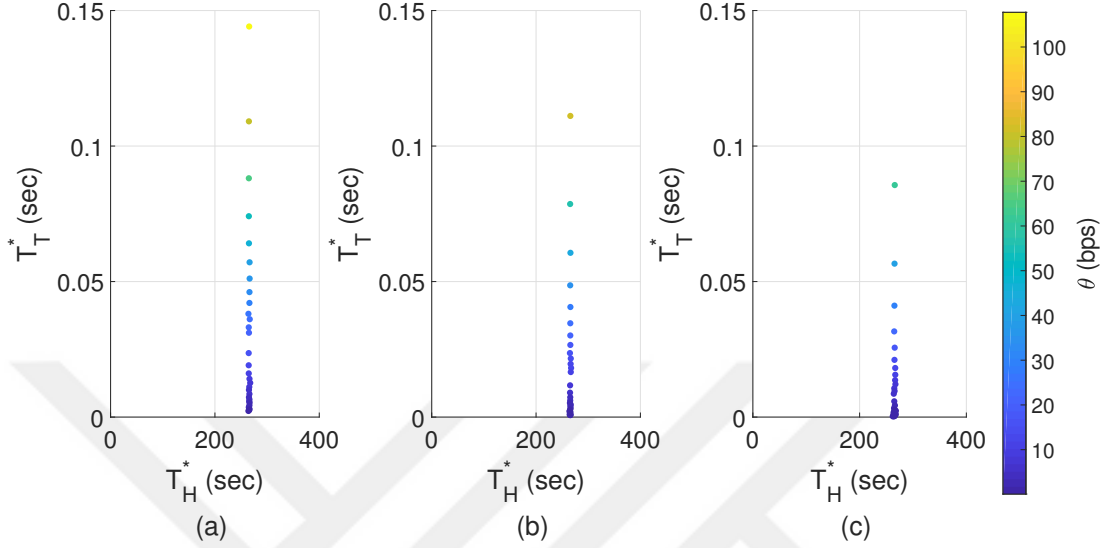


Figure 6.6: T_H^* and T_T^* values that maximize θ , for path loss exponents of (a) $\nu = 2$, (b) $\nu = 2.5$ and (c) $\nu = 3$, respectively.

increased from 2 to 3. This stems from the fact that denominator of θ in Eq. (6.2) is dominated by T_H , as $T_H \gg T_T$, given the energy harvesting speed of the proposed EFEH method. Consequently, θ is maximized in close proximity of $T_H \approx 266$ sec, where the E_H - T_H curve turns from concave upward to concave downward. At this exact point, reserve energy $E_H - E_N - E_Z$ available for transmission in ratio to T_H is maximized. In short, energy profile E_H - T_H is the dominant driver of T_H^* , and it can be altered by varying the capacitance in the circuitry of Fig. 6.1(a). A different capacitance would change the T_H^* , as it represents the inflection point of the E_H - T_H curve. On the other hand, T_T^* is very sensitive to changes in d and ν , as seen in Fig. 6.6. Taking $\nu = 2.5$ case as an example, it varies from $T_T^* = 111$ msec for $d = 3$ m to $T_T^* = 16.5$ msec for $d = 250$ m. As conditions get harsher with a longer distance and a higher path loss exponent, T_T^* is contracted to obtain relatively high SNR values in order to keep the BER manageable.

As the last step, effect of minimum reporting frequency f_{min} on T_H^* and T_T^* is investigated. This f_{min} depends on the variable that is being monitored by the sensor node and translates into an upper limit on T_{Total} as $T_{Total} = T_H + T_N + T_T \leq \frac{1}{f_{min}}$. As seen already in Fig. 6.4(b), any application requirement that implies $\frac{1}{f_{min}} = T_{max} <$

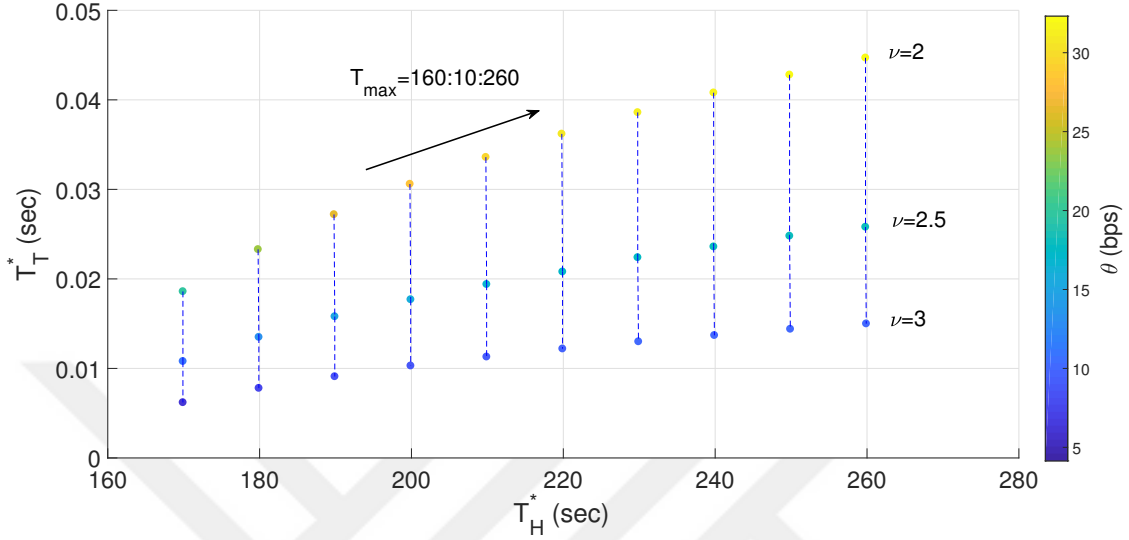


Figure 6.7: T_H^* and T_T^* values that maximize θ , for path loss exponents of $\nu = 2 : 0.5 : 3$, where T_{max} is swept from 160 to 300 sec. Points with same T_{max} are connected with dashed lines.

154 sec cannot be fulfilled with the proposed system model. For $d = 10$ m, Fig. 6.7 illustrates how T_H^* and T_T^* varies with T_{max} and ν . As an example, without a restricting requirement on reporting frequency, $T_H^* = 266.3$ sec and $T_T^* = 46$ msec for $d = 10$ and $\nu = 2$. Under same channel conditions, with $f_{min} = \frac{1}{200}$ requirement set, these values shift to $T_H^* = 199.89$ sec and $T_T^* = 30.6$ msec. In a sense, a restricting requirement on f_{min} forces the system to end up in a sub-optimal setting in terms of θ . That being said, loss in θ compared to its global maximum due to this sub-optimality is less than 20% for all $f_{min} < \frac{1}{190}$. Additionally, as f_{min} requirement is gradually relaxed, the trajectory of the T_T^* is concave downward, with diminishing returns on θ . As an example, θ increases almost by a factor $2\times$ when f_{min} is relaxed from $\frac{1}{160}$ to $\frac{1}{200}$, whereas it increases by a mere 10% as it is further extended to $\frac{1}{300}$.

Part II

Energy and Spectrum-Efficient Communication Techniques in Nanonetworks for IoBNT

Chapter 7

MODULATION IN MOLECULAR COMMUNICATIONS: A LOOK ON METHODOLOGIES

7.1 Introduction

Field of nanotechnology envisages development of very small scale entities, called nanomachines, that are able to carry out tasks such as communication, sensing and computation. Capabilities of a nanomachine on its own is very limited. To that end, making nanomachines collaborate efficiently is a key design objective to achieve complex tasks. Nanonetworking aims to accomplish this objective by defining a set of rules for communication and collaboration between a collection of nanomachines, so as to achieve bigger tasks. The capabilities that numerous nanomachines can achieve via nanonetworking have been exemplified in [120]. For instance, chemical nanosensors will be able to perform complex drug delivery in human body via collaboration. In other scenarios, in cases of deployment over large areas, nanomachines will be able to interact with each other via either multihop or broadcast communication mechanisms.

Networking and collaboration of nanomachines to achieve complex tasks inherently requires solid communication schemes. Communication between nanomachines is a difficult task requiring multifaceted research efforts. The challenges ahead of nanonetworking can be briefly listed as follows:

1. Downsizing of classical communication circuitry to nanoscale is not possible, hence design of nanomachines that are able to carry out communication functionalities is required.
2. Mobility of both the communicating nanomachines and the messengers they use for communication are highly dependent on the medium they reside in

and the application they are designed for, bringing a number of channel and communication models to study.

3. Some nanocommunications and nanonetworking applications are designed to operate in living organisms, in which case both the nanomachines and their communication mechanism has to be biocompatible.

Various communication types, namely; acoustic, electromagnetic, nanomechanical and molecular communications have been proposed for nanonetworking. Acoustic and electromagnetic communications types are challenging to be realized until entities that can carry out acoustic or electromagnetic operations at nanoscale can be developed. Nanomechanical communications, on the other hand, require direct contact between communicating parties and is not suitable for distant applications [121]. In this context, Molecular Communications (MC) becomes a viable method for nanocommunications, backed by the fact that similar mechanisms are already present in many living organisms and environments.

In MC, two or more nanomachines are envisaged to communicate by the help of messenger molecule exchanges [122]. Transmitter nanomachine can encode informa-

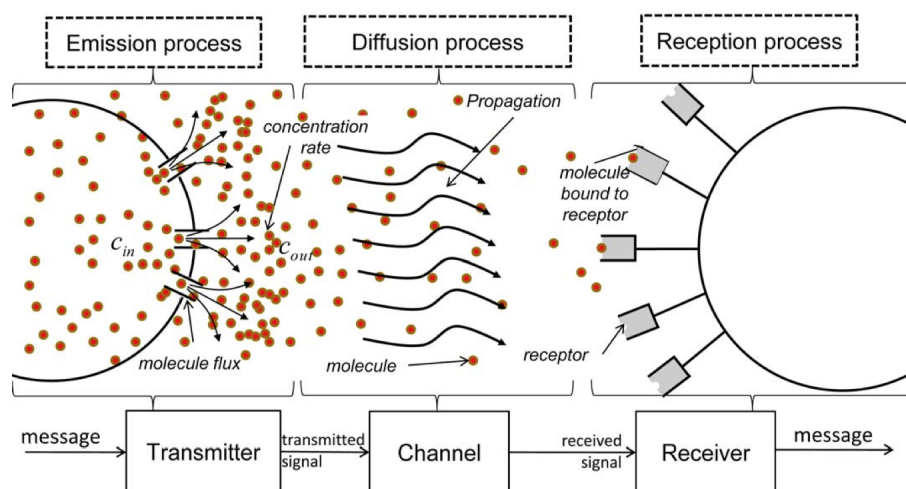


Figure 7.1: Diffusion based molecular communication [140]

Table 7.1: Comparison of traditional vs. molecular communications

	Traditional communications	Molecular communications
Communicating parties	Electronic devices	Nanomachines
Signal types	Electrical or optical	Chemical
Propagation speed	Speed of light	Very slow
Propagation range	On the order of $m - km$	On the order of $nm - \mu m$
Communication medium	Air and/or wire	Aqueous

tion, i.e. carry out modulation in molecular communication, via various techniques. The messenger molecules, via which the nanomachines communicate, then are released to the medium according to the preferred modulation technique. The messenger molecules propagate through the medium in which the nanomachines reside, from the transmitter nanomachine towards the receiver nanomachine. These molecule(s) eventually arrive at the receiver nanomachine, and the message is decoded at this nanomachine accordingly. A basic molecular communication system is depicted in Figure 7.1. The comparison of MC and traditional communications for communication entities and properties is given in Table 7.1 [120], [122], [123], [124], [137].

Molecular communication faces different challenges across networking layers. Routing needs reconsideration given the aqueous medium and the slow propagation characteristics of MC. How to encode messages into molecules and address different nanomachines present significant challenges. Synchronization, multiple access into the channel, channel capacity and interference are just a few other areas that need consideration in molecular communication, since results from traditional communication architectures and solutions are not directly applicable [137]. In this chapter we will focus on a very important physical layer issue; namely, modulation; and review the proposed modulation techniques for MC according to their categorization, and discuss their strengths and open issues [156].

7.2 Motivation for Modulation Research in Molecular Communications

MC can be broadly categorized under three main groups according to how molecules propagate in the medium.

- The first category is walkway-based or active transport MC, where molecules follow predefined paths to the receiver.
- The second category is the flow based MC, where molecules are directed by flows towards the receiver nanomachine.
- Lastly, and the most common molecular communication technique is diffusion-based molecular communication, where the molecules carrying information are left to freely diffuse in the transmission medium to convey the messages from the transmitter to the receiver.

The majority of this chapter is focused on diffusion based MC, which rely on the random movement of molecules in the direction of the concentration gradient. Because of this stochastic nature of diffusion, MC require a different perspective than the traditional wireless communications using electromagnetic waves as information carrier. The fundamental challenges in MC compared to traditional electromagnetic communications are very low transmission speeds (few millimeters per second), long and varying delays (up to hours), discretization of the signal to a number of molecules, relatively short ranges (few millimeters) and high amount of distortion. Besides the challenges arising from the diffusion process, the restricted information processing and memory capabilities of nanomachines also limit the complexity of design solutions for MC.

Modulation of the information in MC is central to many of these challenges. Given the low transmission speeds and long delays, having an efficient modulation scheme for MC is highly necessary. To this end, many modulation schemes have been proposed. The research efforts in this field can be broadly categorized under four main groups:

1. The first group of modulation schemes is called **concentration based modulation**. In such schemes, the information message is encoded into the concentration of a certain type of molecule within the aqueous medium. When the sender nanomachine intends to send a digital 1, it releases a certain number of that

molecule such that the concentration of this molecule is higher than a predefined threshold. When these molecules propagate to the receiver, hopefully their concentration will still be above this threshold, and the receiver nanomachine will correctly decode the bit as 1.

2. The second group comprises **molecule type based modulation** schemes. In a molecule type based modulation, the information is encoded on the bits of the molecule to be released into the medium. Examples of such messenger molecules are isomers, where the H and F atoms can be considered to represent bits 0 and 1, respectively.
3. The third group corresponds to the **molecule release time** or **molecule release order based modulation** schemes. In such schemes, the information is encoded into the release time or order of the messenger molecules.
4. The last group are **hybrid modulation** schemes, where two or more of the modulation schemes pertaining to the above groups are employed at the same time to increase diversity.

Considering the challenges of MC, it is crucial to understand for which types of applications each of these modulation groups would be most suitable, in terms of achieved communication rate. Besides the rate, need for synchronization could be another important factor for choosing one modulation type over another, since synchronizing the communicating entities may not always be possible. Furthermore, for applications requiring different communication ranges, different modulation schemes are proposed to increase the capacity. For molecular communication, the range of a nanomachine can be short, medium or long which are on the scale of $nm - \mu m$, $\mu m - mm$ and $mm - m$, respectively. A short range molecular communication takes place inside of a cell or in between two neighbor cells. Quorum sensing in a bacteria population can be considered as medium range molecular communication where the bacteria release a specific substance whose concentration gives the information on

the size of the population. For long range molecular communication, the hormonal communication between organs meters apart is a good example.

Applications of MC inside human body would also require the messenger molecules to be biocompatible, since otherwise they would cause biological complexities for the functions of the human body. Given these perspectives, it is necessary to understand what the current landscape in modulation schemes of MC is, for which applications what kind of modulation type would be most suitable and what are the open issues and remaining challenges towards realization of more efficient MC.

7.3 Modulation in Molecular Communications

In this section, different types of modulation techniques for diffusion-based molecular communication are presented. Some of the techniques have been implemented such as the Time Elapse Communications (TEC) modulation in bacteria [141] and On-Off Keying (OOK) modulation in [125]. However, since molecular communication field is still in its early stages, there is not any system which is able to implement and compare all the modulation techniques presented here. This categorized list point outs how the modulation techniques for electromagnetic communication can be adapted to molecular communication and how the peculiarities of molecular communication offer new perspectives for modulation.

7.3.1 Concentration Based Modulation Schemes

In this part, we will explain the different modulation techniques which basically change the shape of a molecular concentration wave. Similar to electromagnetic waves, both analog and digital modulation schemes can be constructed to encode information on molecular concentration waves.

Amplitude and Frequency Modulation

The idea of using analog modulation in molecular communication is proposed in [126] based on the inherent Amplitude Modulation (AM) and Frequency Modulation (FM)

of intercellular calcium waves propagating through the gap junctions between cells [127]. Calcium ions are the messenger molecules of intercell communication systems existing in living organisms. By the diffusion of Intercellular Calcium Waves (ICW), groups of cells coordinate with each other and regulate the cellular activities such as chemical secretion, fertilization, neural signaling, contraction and death [127].

One of the regulations of ICWs takes place in immune system cells, which generate high or low concentration Ca^{2+} waves causing different reactions in cells. A naive B lymphocyte (a type of immune system cell) having a contact with the antigen (a substance which provokes an immune system response) for the first time creates a high concentration Ca^{2+} wave inducing reproduction of the cell. However, a B lymphocyte which has encountered the same antigen before creates a low concentration Ca^{2+} wave blocking the reproduction. This process illustrates an AM signaling for diffusion based molecular communication where the same immune system's cells (B lymphocytes) using the same messenger molecule (Ca^{2+} waves) can control two different processes by adjusting the concentration level of Ca^{2+} waves [127].

Another regulation mechanism of ICWs uses the frequency of Ca^{2+} waves instead of the amplitude. The frequency of the regular oscillations in Ca^{2+} concentration can be varied by the concentration of an input signal in order to control the rate of different cellular processes such as fluid secretion by salivary glands or glycogen metabolism in liver cells. These processes corresponds to FM signaling for diffusion based MC.

A nanoscale communication system using ICWs can be formed via a group of neighboring cells capable of propagating ICWs. A source cell encodes the information on the Ca^{2+} wave using either AM or FM modulation and transmits the wave to the nearest intermediary cell in the channel from the source cell to the destination cell. Each intermediary cell in the channel propagates the wave to its neighbor until it reaches the destination cell where the modulated signal is decoded [128].

As calcium signaling is the main intercellular communication mechanism in biological cells [138], it is definitely a biocompatible modulation technique. Furthermore the

transmitter, receiver and intermediary cells are inherently synchronized for the propagation of continuous calcium ion waves in living organisms. Thus, while designing a molecular communication system interacting with living organisms via calcium ion waves, this system must be synchronized to cellular processes in that living organism.

On-Off Keying (OOK)

On-Off Keying (OOK) is the basic digital modulation scheme used in diffusion based MC. In this scheme, time is divided into slots for transmission. At the beginning of each time slot or during the whole time slot, a predefined number of molecules is released from the transmitter to the communication medium to represent bit 1, and no molecule is released to represent bit 0. The released molecules diffuse in medium stochastically and a portion of the released molecules arrive to the vicinity of the receiver. The receiver is capable of sensing the number of molecules in its vicinity and decides whether a 0 or 1 has been transmitted according to a concentration threshold.

OOK is mostly considered to characterize the diffusion based molecular communication channel and to calculate its capacity. In [139], OOK is used to provide an information theoretical approach for MC, taking into account the binding of the molecules to the receiver and considering an exponential decay for molecule concentrations. Molecules are assumed to be released from the transmitter at a fixed concentration level during the whole time slot corresponding to a square pulse. Then, the mutual information between the point-like transmitter and the point-like receiver is analyzed. Being one of the first capacity investigations for molecular communication, [139] provides valuable insight. Nevertheless, [139] models the random propagation of molecules as an exponential decay with time, which is a coarse assumption that was later replaced by Brownian Motion assumption in most of the studies.

[129] elaborates the capacity analysis by considering both the Brownian Motion propagation and the stability of the molecule in the environment. First, a naive modulation scheme is considered where only one molecule is released to represent

bit 1 at the beginning of the time slot and no molecule is transmitted to represent bit 0 corresponding to impulse signal. The receiver decides that a 1 was sent if it receives a molecule in the current time slot. Otherwise, it decides that a 0 was sent. Since the probability that one molecule reaches the receiver in the same time slot while following a Brownian Motion is really low, this scheme yields low capacity. In order to achieve higher capacities, an extended modulation scheme is proposed. A high number of the copies of the same molecule are transmitted to represent bit 1 to increase the probability of reception at the receiver and different number of molecules are considered as the threshold for decoding. By the help of this extended scheme, when the slot time, the number of released molecules for bit 1 and the reception threshold are properly adjusted, capacities in the order of 0.03 bits per channel use can be achieved.

For OOK, several different pulse shapes are investigated in [130] other than the square pulse considered in [139] and the impulse in [129]. A square, a cosine and a Gaussian pulse having the same time slot duration and total energy are compared. Although it is observed that the Gaussian pulse has the lowest distortion and the maximum received peak level, the pulse shape does not have a significant effect on the total received energy. In addition, it is shown in [130] that to achieve higher bandwidth for longer ranges, the pulse duration should be the minimum possible duration. Considering all of these features, an optimal pulse shape may be chosen according to the decoding scheme.

For all of the modulation schemes proposed for OOK, synchronizing the time slots of transmitters and receivers is a major issue. Taking into account the low complexity of nanomachines, keeping the time for each communicating party is a challenge. Furthermore, because of the long and random propagation delays there will always be an inconsistency between the clocks of transmitter and receiver even if they do some efforts to synchronize via the diffusion channel.

Concentration Shift Keying (CSK)

Concentration Shift Keying (CSK) is analogous to Amplitude Shift Keying (ASK) in traditional electromagnetic communications. Here, the concentration of the molecules are the amplitudes of the signals for short and medium range molecular communication. CSK, first proposed in [131], is based on the same assumptions as OOK. Again, the information sequence is spread over the time slots using one symbol per time slot. However this time, there are more than 2 binary symbols and one symbol represents more than one bit. The concentration level of the transmitted pulse is varied according to the groups of sequential bits.

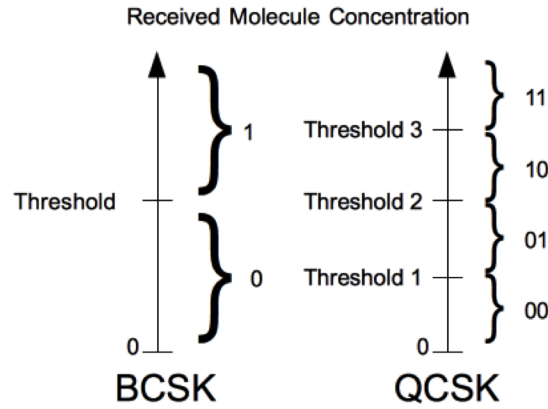


Figure 7.2: Detection by thresholds for Binary CSK and Quadruple CSK

To represent n bits, 2^n different molecular concentration levels must be generated at the transmitter and $2^n - 1$ threshold levels must be considered at the receiver as shown in Figure 7.2. Because of the diffusion dynamics, molecules may not arrive to the receiver in the time slot that they are released, causing incorrect decoding of the current symbol. If they reach the receiver in the following time slots, they cause Inter Symbol Interference (ISI) for the next symbols. Hence, although increasing concentration levels improves bit per symbol rate, it also rises symbol errors because of ISI.

Transmitted molecules exhibit Brownian Motion with reception probability depending on the distance between the transmitter and receiver nanomachines and the

duration of the time slots. Receiver nanomachines sense the number of molecules arriving to it, i.e., the concentration level, which is modeled with a Binomial distribution. The noise in this scheme arises from the residual molecules from previous time slots and the molecules present in the environment sent by other information sources and it is modeled as Additive White Gaussian Noise (AWGN). To investigate the effect of CSK modulation on capacity, authors in [131] compare two cases where $n = 1$ (Binary CSK) and $n = 2$ (Quadruple CSK). For higher SNR values both cases achieve the theoretical limits. However when the noise levels are significant, Binary CSK outperforms Quadruple CSK since for QCSK case the concentration intervals between the thresholds are smaller so it is more vulnerable to noise.

Pass Band Modulation (M-AM and FSK)

For the analog modulation schemes, we covered AM and FM and for the digital modulation schemes we covered base-band modulation techniques such as OOK and CSK. In this part, we consider pass band modulation where the amplitude and the frequency of the molecular concentration oscillations are discretely modulated to obtain Multilevel Amplitude Modulation (M-AM) and Frequency Shift Keying (FSK) for medium-to-long range molecular communication.

In [132], binary amplitude modulation and binary frequency shift keying in pass band is considered where the time is slotted. The propagation of messenger molecules are modeled as Brownian Motion, the messenger molecules are assumed to be pheromones propagating long distances in the air. The ligand-binding process of the molecules to the receiver is taken into account. The signal distortion is analyzed by simulating the transmitted and the received signals. Furthermore, different detection schemes for M-AM case are proposed. Sampling based detection and energy detection schemes are investigated where in the former the receiver nanomachine senses the concentration of the messenger molecules only for a short period of time and in the latter the receiver counts all of the molecules during the current time slot corresponding to the energy of the signal. In [132], the capacity of the proposed modulation schemes is not

considered but this work stands as an exemplary modulation scheme for long range molecular communication.

The synchronization is an essential part of this modulation scheme. The transmitter and the receiver must both know perfectly the operation frequency and the start and end times of the time slots which is a difficult task for nanomachines with limited capabilities.

The pass band modulation schemes can be designed to be biocompatible if existing pheromone molecules are chosen as messenger molecules as in [132]. In this case, for the biocompatibility analysis, the pheromone concentration levels must be adjusted so that it does not affect the biological processes in existing organisms depending on that specific pheromones or the molecular communication system must be isolated from those organisms.

7.3.2 Molecule Type Based Modulation Schemes

In concentration based modulation, the modulation of the signal is obtained by varying the amplitude or frequency of the messenger molecules' concentration while using the same type of molecule for all different symbols. To improve the capacity of molecular channels, molecular diversity can be used as an additional degree of freedom by constructing a molecule type based modulation scheme.

Encoding the information on the type of the messenger molecules is introduced in [133], where the modulation scheme exploits the vast number of different pheromone molecules present in the nature. In one hand, the molecule diversity can be used to create separate non-interfering channels by using different type of molecules for different channels defined as Molecule Division Multiple Access (MDMA) [133] similar to FDMA or TDMA in traditional electromagnetic communications. On the other hand, different types of molecules can be used in the same channel to boost the capacity of the channel defined as Molecule Shift Keying (MoSK) [131] similar to using orthogonal symbols in traditional electromagnetic communications.

To transfer n bits of information, MoSK utilizes 2^n different types of molecules

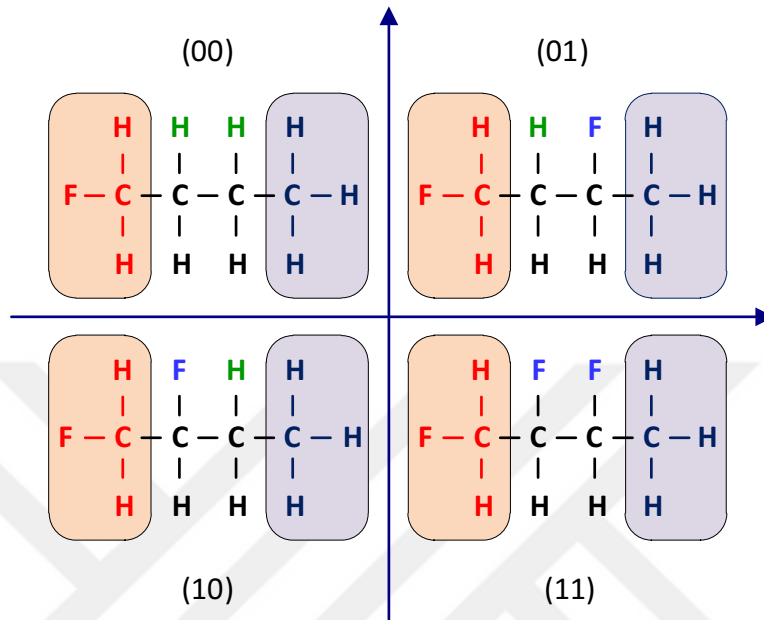


Figure 7.3: Constellation diagram for QMoSK [134]

to represent each combination of 2^n different n bit sequences. The time is divided in equal sized time slots and a large number of one type of these 2^n different molecules are released from the transmitter at the beginning of each time slot. If the concentration of one type of molecule exceeds the threshold in the current time slot, n bits represented by that molecule type is decoded. If the concentration of more than one molecule is larger than the threshold or none of the molecule type's concentration exceeds the threshold, an error occurs. Also, the residual molecules from the previous time slots cause ISI [131].

In [131], hydrofluorocarbons are introduced as candidate messenger molecules for MoSK for which n can reach up to 10^9 [134]. However, n should be small enough that the nanomachines can process the molecules and the molecules can propagate in the medium. For example, for intra-body applications, hydrocarbons molecules with n larger than 10^6 cannot propagate in the blood [136].

For hydrocarbon molecules, by attaching H or F atoms to a long carbon chain representing bits 0 and 1, information can be encoded onto molecules with similar physical and chemical properties. In this way, by altering n atoms of a molecule, 2^n different symbols can be produced. An illustrative constellation diagram for Quadru-

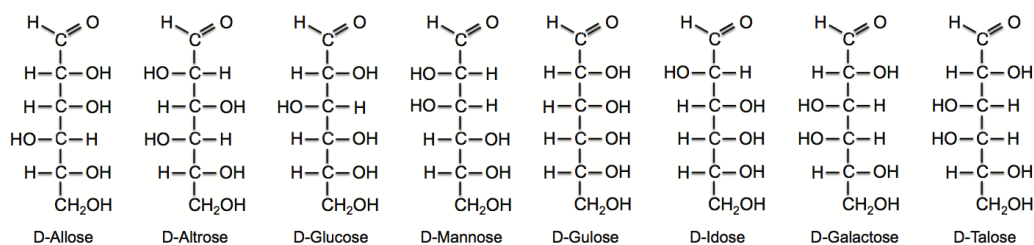


Figure 7.4: Isomers of aldohexoses

ple MoSK is shown in Figure 7.3. In [131], it is shown that the capacity of diffusion based molecular channels using Binary MoSK ($n = 1$) may be improved by using a higher order modulation such as Quadruple MoSK ($n = 2$) due to the orthogonality of symbols. Thus, as the modulation depth increases the capacity improves. Nevertheless, while the modulation depth increases, the messenger molecules become larger and start to diffuse more slowly reducing the reception probability. In [134], the optimal modulation depth and the conditions yielding it are investigated. By tuning the time slot duration and the concentration level according to the properties of the messenger molecules and the communication medium, the capacity is optimized.

Another technique for molecule type based modulation is using isomer molecules as symbols. In [135], isomers are proposed as messenger molecules since isomers contain same number and type of molecules which facilitates the synthesis of molecules for the transmitter. Specifically, [135] analyses the use isomers of aldohexoses illustrated in Figure 7.4, which are monosaccharides with the chemical formula $C_6H_{12}O_6$, for 32-MoSK to increase achievable rate. Using aldohexose isomers is also a biocompatible technique since monosaccharides are naturally found in human body. On the contrary, for hydrofluorocarbons proposed in [131], a toxicity analysis is required to choose the biocompatible sets of messenger molecules since some of the hydrocarbons may be toxic to human body or flammable.

When compared to CSK, MoSK technique is more noise tolerant than CSK since when the modulation depth in CSK increases, the detection thresholds for the concentration of the messenger molecule get closer and errors occur more easily. However, in MoSK, detection depends on a single concentration threshold of the molecule type

in use for that symbol.

The challenge in molecule type based modulation is the difficulty of producing, storing and receiving multiple different molecules in transmitter and receiver nanomachines which has limited source and functionality. A developed nanomachine capable of handling more than one type of molecules is required. Besides, every molecule representing a symbol has different physical and chemical properties, thus the propagation of the molecule and the interaction of the molecule with the environment varies favoring some symbols over the others. In this context, some molecules may even be toxic for in-body applications, therefore additional research is needed to ensure biocompatibility in such schemes. Moreover, similar to the concentration based modulation, synchronization of time slots poses an important challenge.

7.3.3 Release Time Based Modulation Schemes

In this category of modulation schemes, the information is encoded into the release time of the molecules that are released from the sender nanomachine towards the receiver nanomachine. A very important work on time based modulation in MC is [141]. Authors have made a significant contribution to the molecular communication research not only in the sense of the modulation considered, but also because they used genetically engineered bacteria in an experimental molecular communication setting. This second point is important in the sense to verify that such bacteria can be used as nanomachines, and since they can be made biocompatible via genetic engineering, biocompatibility at least for the nanomachines in such a setting has been proven to be possible.

Apart from biocompatibility, another result from [141] is that authors employed time based modulation. Before arriving to the time based modulation, however, authors first showed On Off Keying in bacteria. The setup considered in this paper is as follows: The sender has access to a bacterial population, which, when stimulated, releases molecular signals. These molecular signals propagate towards the bacterial population that resides at the receiver site. When these signals are interpreted by

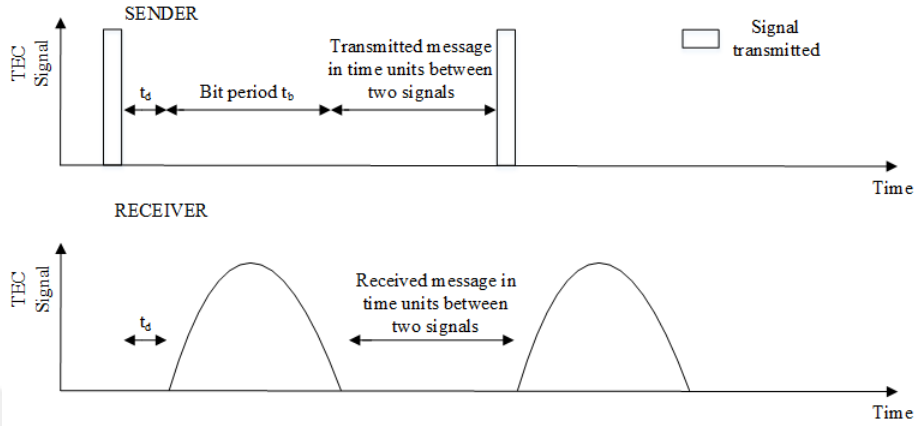


Figure 7.5: Illustration of TEC signaling at the sender and receiver sides [141].

the bacteria population at the receiver side, these bacteria respond with fluorescence, which is then interpreted by the receiver circuitry. As a consequence, via bacteria populations, one is able to convey information from sender to receiver.

Authors in [141] have shown that the bit duration has to be significantly large (≈ 450 minutes long) in OOK. This inherently lowers the data rate performance of OOK in bacteria based MC. To this end, *Time Elapse Communications* (TEC) have been proposed. In TEC, the information is encoded in the time interval between two consecutive signals. In this scheme, it is assumed that both the sender and the receiver have the same clock rate f_c , although synchronization of clocks between these two entities is not required. The number of generated molecular signals is always two in this scheme: The first molecular signal is the start signal, and the second molecular signal is the stop signal. The information is encoded in the time between these molecules. Assuming that the information to be sent is represented by v , the start and stop signals are separated by a time period of $\frac{v}{f_c}$, i.e. the bits in v is encoded by the number of $\frac{v}{f_c}$ clock cycles. This modulation scheme is illustrated in Figure 7.5.

Another point is that the stop signal can be arranged to be the start signal of the next transmission, further increasing rate. To have a better data rate in TEC with respect to OOK, the clock rate f_c has to be higher than the inverse of the achievable minimum bit duration in OOK, which is denoted as t_b . Although the above illustration is depicted for an error-free communication channel, the molecular

start and stop signals can be degraded due to noise, causing errors in the bits encoded in time. Despite this, authors in [141] proved that an improvement of approximately an order in data rate is achievable by employing TEC rather than OOK in bacteria based MC.

[142] is another work on time based modulation of information in MC. However, the authors study neither starts with time based modulations, nor stops at that point. The study begins with the consideration of how molecule type based modulation works, then consider how would a release time based modulation scheme could be implemented, and then the authors combine the molecule type based and the release time based modulation schemes to construct a mixed type modulation. This last combined scheme will be presented in Section 7.3.4. For the release time based modulation, authors consider the following model: The two molecule transmission are separated in time by a Bernoulli random variable S , i.e. $P(S = s_0) = \pi_0$ and $P(S = s_1) = 1 - \pi_0$. We assume that M_0 is encoded in case of an s_0 separation in time, and M_1 is encoded in case of an s_1 separation in time. Then, consecutive molecules arrive at the receiver nanomachine, and the information is decided based on the time separation between consecutive molecules, which would be modified until these molecules arrive at the receiver according to the mobility model assumed. The biggest advantage of this model is that it is asynchronous, hence does not require any synchronization nor does it require exact same clocks at the sender and the receiver. Only the receiver is expected to count the time difference between two consecutive molecule arrivals. The model indeed surpasses the need for synchronization or same clocks, however could be prone to error propagation: When a molecule arrives later than expected, not only it would be decoded in error, but also it would affect the timing considerations for the upcoming molecules and may cause them to be decoded in error as well.

Another modulation scheme that relies not on the release time of molecules but on the release order of molecules is presented in [143]. In the described modulation scheme, the release order of different types of molecules determines what message is

being transmitted. For instance, let us assume that molecules of type A and B are available. Then, release order of $A - B$ can encode bit 0, while $B - A$ order would encode bit 1. An illustration of this scheme is given in Figure 7.6. Only the order of these molecules are changed, while the inter transmission time for two different symbols T is kept the same. In this modulation scheme, there are two sources of errors:

- For instance, molecules may be sent in order of $A - B$, but they may arrive to the receiver in the order $B - A$.
- Molecules of the previous bit can interfere with the current bit, causing ISI.

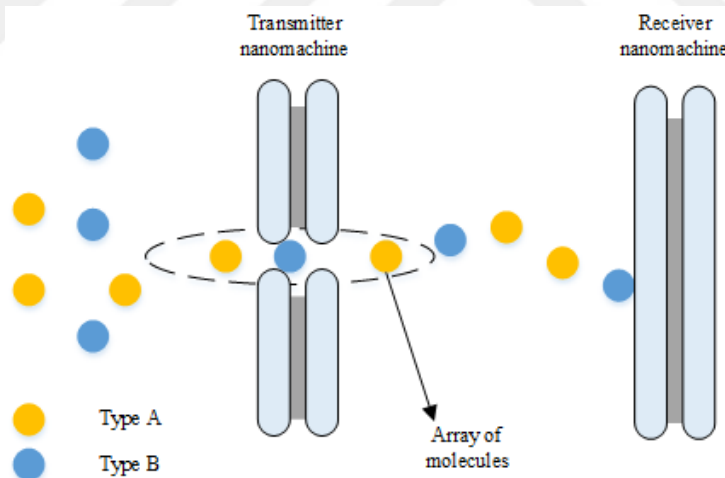


Figure 7.6: Molecular array modulation, when only two molecule types are considered.

The second source of error can be made insignificant by adjusting the inter transmission time for two symbols T . For the first source of error, one has to increase the inter transmission time for two molecules in a symbol, denoted as t_e . However, in [143] it is shown that as the receiver to transmitter distance increases, the probability of correct decoding at the receiver (i.e. $A - B$ received when $A - B$ sent, $B - A$ received when $B - A$ sent) converges to 0.5, i.e. the correct transmission becomes totally random since the medium characteristics become the dominant factor on the order of the molecules in a symbol. We can deduce that, although the modulation

scheme brings great advantages for molecular communication in terms of need for synchronization and achieved capacity, it is not suitable for long range molecular communication networks.

7.3.4 Mixed Type Modulation Schemes

Recently, new modulation schemes that employ more than one of these preceding techniques (i.e. concentration based, molecule type based or molecule release time based) have been proposed in the literature.

Authors in [144] have proposed Molecular Concentration Shift Keying (MOCSK). In Concentration Shift Keying (CSK), to transmit b bits at a time, there has to be 2^b different concentration rates. However, since the molecule is the same for all transmissions in CSK, previous transmission can affect the current transmission and cause the decoding to be false. MOCSK brings MOSK idea to the CSK picture. Assuming there exists two molecule types A_1 and A_2 , the MOCSK transmitter carries out CSK with A_1 in odd slots, and A_2 in even slots. This alternating molecule type in consecutive slots reduces error probability in decoding since now the receiver also differentiates molecules based on the slot it is currently in. MOCSK, as CSK and many other MOSK schemes, requires synchronization between the sender and the receiver nanomachines.

On the other hand, the study in [142] combines time based modulation and molecule type based modulation, i.e. MOSK. Different molecule types are used to encode different bits of information. Moreover, the inter transmission durations are assumed a simple Bernoulli model. In that sense, for instance, when the time separation between arrivals of two molecules is lower than a threshold, the receiver can decide on bit 0, and when it is higher than this threshold, it can decide on bit 1. Besides this time modulation, the receiver can also make use of the data encoded on the molecules used. Analysis in [142] suggest that encoding information in inter transmission durations improves the capacity of the regular MOSK channels. Nevertheless, error propagations can occur in the information encoded, especially those

encoded on the inter transmission durations, since one outlier in terms of delay would affect the upcoming transmissions too.

Diffusion in different fluids such as water, air or blood has different characteristics. Despite the modulation schemes presented in this section are valid for all types of fluid, the parameters of the modulation schemes such as time slot duration or amplitude of the concentrations should be adjusted in order to accommodate the peculiarities of the propagation medium.

7.3.5 Modulation and Nanomachines in Molecular Communication

All the modulation techniques presented in this section requires complex machinery at nanoscale. There are three approaches for building nanomachines in the literature, namely, the top-down approach, the bottom-up approach and the bio-hybrid approach [120].

The top-down approach suggests miniaturization of current microelectro-mechanic systems down to nanoscale. However, at nanoscale level, the direct downscaling is hindered by the atomic interactions. Even though nanomicroelectro-mechanical parts can be fabricated, producing batteries which can support these nanoscale parts is still a challenge.

The bottom-up approach constitutes of building up nanomachines by chemical interactions of individual molecules. With today's technology it is possible to fabricate molecular components such as molecular gears and molecular switches. Nevertheless, the assembly of these components is not straightforward.

The third approach is the bio-hybrid approach suggesting using biological entities which inherently behaves as nanomachines. By tearing apart the useful parts of the cells or reprogramming them as in synthetic biology, a complete nanomachine capable of manipulating molecular communication signals can be obtained. As of now, this approach is the most promising approach since cells already constitute complete nanomachines communicating with each other via molecules.

The modulation techniques discussed in this chapter can be implemented by fab-

ricating artificial nanomachines with the top-down or bottom-up approach which will result in producing electrical circuits at atomic level. Furthermore, by using the bio-hybrid approach, the modification of natural cells may lead us to achieve biocompatible nanomachines.

7.4 Segmentation of Related Work on Key Criteria

As the molecular communication signals are transferred with molecule concentration levels, the most instinctive modulation technique is concentration based modulation where the amplitude of concentration signals are varied according to the message to be conveyed. All of the concentration based modulation schemes are prone to errors because of the accumulation of molecules from previous transmissions and digital concentration based modulation techniques suffer severe ISI as the modulation depth increases. The biocompatibility of these techniques depends on the molecule type preferred as messenger molecule. If the messenger molecules are chosen from the molecules which do not interact with biological processes in their communication environment, they may be classified as biocompatible. However, due to the slotted time structure of these techniques, synchronization among communicating parties stands as a major issue.

Another way of modulation for molecular communication channels relies on the usage of different molecules for different symbols. In this case, ISI effects are lowered in expense of the complexity of synthesizing, storing multiple molecules at the transmitter side and the need for multiple receptors and the complexity of decoding at the receiver side. All the proposed molecule type based modulation schemes in the literature takes into account the biocompatibility by intelligently choosing messenger molecule sets. As in the concentration based modulation, molecule type based modulation requires also tight synchronization of transmitters and receivers due to the slotted time assumption.

To address the synchronization problem of the first two techniques, time/order based modulation schemes are introduced. Despite severe ISI, by cleverly encoding

the information not on the concentration level of the molecular signal but onto the timing of the signal or the order of signals, we can bypass the need of synchronization. Nevertheless, the disadvantage of these techniques compared to concentration based techniques is the propagation of error over symbols.

The last category of modulation techniques constitutes the techniques combining some of the techniques from the first three category in order to address all of the challenges of molecular communication. These mixed type techniques aim at increasing the rate and capacity, reducing error rates and avoid synchronization issues. Mixed type techniques outperforms the previous techniques in all terms.

All of these four category of modulation are designed for short-to-medium range molecular communication. Designing modulation schemes for long range molecular communication remains as a research challenge.

Even though theoretically numerous modulation schemes can be suggested, the practically of each of them in the nanoscale communication environment is still an important question. The determination of which techniques lies in the limits of operation of actual nanomachines and which techniques operate more robustly in the communication environment that might even be a living organisms is the next step in modulation research for molecular communication.

Table 2 summarizes the research efforts on modulation in MC, which have also been summarized in Section 7.3.

Table 7.2: Research efforts in modulation in molecular communications.

Category	Sub-category	Examples	ISI level	Bio compatibility	No need for sync.
Concentration based	On Off Keying	[139]	Mid	N/A	-
	FM and AM signaling	[126]	N/A	✓	-
	CSK	[131]	High	N/A	-
	Pass Band Modulation	[132]	High	N/A	-
Type based	Pheromones	[133]	Low	✓	-
	Hydrofluorocarbons	[131]	Low	✓	-
	Aldohexoses	[135]	Low	✓	-
Time / order based	Time Elapse	[141]	Mid	✓	✓
	Inter transmission coded	[142]	High	-	✓
	Molecular array	[143]	High	-	✓
Mixed type	MOCSK	[144]	Low	-	-
	Time + MOSK	[142]	Mid	-	✓

Chapter 8

POLYETHYLENE: A PROMISING MESSENGER FOR HIGH PERFORMANCE MOLECULAR COMMUNICATIONS

8.1 Introduction

Molecular Communications (MC) is a recent bio-inspired paradigm [124]. Existing biological phenomena, such as intra-cellular communications through transportation of vesicles and hormones, have lately inspired the communications via molecules in the context of nanomachines. Nanomachines are nanoscale devices, and can be naturally existent (i.e. biological cells) or artificially synthesized (i.e. bioengineered cells) [129]. MC allows nanomachines to encode information on to transmitted molecules, to concede these molecules to propagate in an aqueous medium and to decode information from the received molecules. MC is very convenient for nanomachines not only because of its undemanding nature in terms of energy and complexity, but also for its power in networking simple nanomachines to carry out more complex tasks. Potential application areas of MC in networked nanomachines include but are not limited to medicine, monitoring, quality control and event detection.

In this chapter, a messenger-based MC model is introduced [134], [145]. As per this model, information is encoded on the atoms of messenger polyethylene molecules, which are in the form of $CH_3(CHX)_nCH_2F$. Here X is either an H or F atom, representing a 0 and a 1 bit, respectively. This MC model is analogous to the chemical communications carried out by insects, through pheromone release and detection processes [154]. Akin to communications by pheromone release and reception, messenger polyethylene molecules are first encoded by the sender nanomachine, then released

into the the aqueous medium to propagate towards the receiver nanomachine. We assume the propagation of these messenger molecules follows a Brownian motion pattern, and model this sort of communication as an erasure channel. We later derive the capacity of this channel and reveal the impact of critical communication parameters on this capacity.

The remainder of this chapter is organized as follows. In Section 8.2, contributions of this chapter are revealed. In Section 8.3, the system model for messenger-based MC that employs multiple bit carrying polyethylene molecules is introduced. The channel model and derivations on its capacity are given in Section 8.4. The impact of critical communication parameters on the channel capacity of this messenger-based MC model are provided in Section 8.5.

8.2 Related Work and Contributions

Related work on MC span a variety of topics from channel modeling to protocols. The capacity of MC under additive inverse Gaussian noise is examined in [155]. Our previous work [156], presented in *Chapter 7: Modulation in Molecular Communications: A Look on Methodologies*, provides a very thorough overview of all existing schemes in MC.

The novelty of this chapter, compared to related work can be summarized as follows:

1. In contrast to existing schemes, multiple bits can be encoded on a single polyethylene molecule and conveyed between two nanomachines in a single propagation session.
2. Bit encoding on messenger molecules can also include identifications for transmitter and receiver nanomachines and related sequence numbers. Such a model would allow us to move one step forward from physical layer to the data link layer in MC.

3. As the messenger molecules can include data packet sequence numbers, there is no need for synchronization among nanomachines.
4. Numerical results disclose the conditions for which a multiple bit carrying messenger molecule is a feasible and convenient method for MC, for a use case scenario.

8.3 System Model

Related work on diffusion-based MC models encode information on the concentration of the molecules that are released to the aqueous medium [129, 133]. Instead, in this chapter, we examine the viability of encoding information into multiple atoms of a messenger molecule to carry out MC.

For this model, we consider the partially fluorinated polyethylene molecule, in the form of $CH_3(CHX)_nCH_2F$, as the messenger carrying n bits of information in a single propagation session. The atom defined as X is either a Hydrogen (H) or a Fluorine (F) atom, to represent the bits 0 and 1, respectively. Select forms of fluorinated polyethylene molecules have been used as carriers of oxygen in artificial blood formation applications [147]. Therefore, we consider in our model molecules called fluorinated hydrocarbon perfusants, assuming that they propagate in human blood as a use case.

Number of bits n that can be encoded on the polyethylene molecule can reach up to 10^9 [136], making transfer of incredibly large amounts of information possible in a single propagation session. Nevertheless, as the messenger molecule gets larger, its diffusion slows down, thereby increasing the propagation delay. Therefore, it is of interest to reveal an optimum value for n that would maximize capacity of this channel.

Other details of the assumed system model are listed below:

- Transmitters are nanomachines that are capable of producing polyethylene molecules. Information that is desired to be conveyed to the receiver nanomachine is en-

coded on the molecule during its production.

- These messenger molecules are emitted into the aqueous medium by the transmitter nanomachine.
- Entire system is assumed to be immersed in an aqueous medium, where the messenger molecules propagate with a Brownian motion.
- Receiver nanomachines have specific receptors capable of capturing polyethylene molecules arriving to its whereabouts and decoding the relevant information on these molecules.

We now detail the five main processes to achieve MC in our system model; namely; information encoding, transmission, propagation, reception and information decoding.

8.3.1 Encoding Information on Polyethylene Molecules

The transmitter nanomachines generate polyethylene messenger molecules, which carry n bits of information. A messenger molecule would ideally possess the following properties:

- It must have a distinctive head, to be easily recognized by nano receptors. This way, the entire message does not need to be read to identify the intended recipient.
- It must be bioinactive, in order not to be easily damaged by various processes throughout diffusion, before it reaches the receiver nanomachine.
- After being decoded by the receiver nanomachine, it must be easily detachable from its receptors and degradable without difficulty [136]. These features would prevent molecule accumulation on nano receptors, ensuring sustainability of the system.

The partially fluorinated polyethylene molecules [136] have been suggested as such candidate messenger molecules in nanocomputer memory systems [146]. Such a molecule is in the form $CH_3(CHX)_nCH_2F$, where atom X corresponds to an H or F atom, representing the 0 and 1 bits, respectively. CH_3 is the distinctive head and CH_2F is the distinctive tail of the molecule, which provide a hint to the receiver nanomachine on where to start decoding. To that end, headers such as identifiers, parity and checksum bits, flags and alike can be added right after the distinctive head. Such encoding enables to embrace a data link layer concept in MC, paving the way for analogies between MC and classical wireless communications for concepts in higher layers such as routing and rate control.

For the sake of reliability and ease of decoding, we assume the encoded atoms are present only in one side of the chain, whereas the other side is assumed to be full of H atoms. This way, the receiver nanomachine would identify which side to decode. It is also assumed that, on the encoded side, the average distribution between H - F atoms are equal, allowing us to assume equal diffusion coefficients for each encoded molecule.

Regarding the compatibility of such molecules for a use case in blood, long hydrocarbon molecules become toxic to human body. On a positive note, a certain amount of F atoms in the messenger molecule prevents this toxicity. Consequently, message molecule density $r_{message}$ is $\sim 1000 \text{ kg/m}^3$ and message molecule volume $V_{message}$ is $\sim (3.82 \times 10^{-29})n \text{ m}^3$ [136]. Hence the information density of the fluorinated polyethylene molecule is $d_{message} \sim 26 \text{ bits/nm}^3$, significantly higher than DNA, whose information density is $\sim 1 \text{ bit/nm}^3$ [136]. This fact led us to prefer fluorinated polyethylene over DNA for this MC use case.

While number of bits that can be conveyed in a single propagation session increases with increasing n , this makes the messenger molecule larger, which in turn slows down its diffusion in the aqueous medium. Hydrofluorocarbon molecules with $n > 20$, as in the case of considered messenger molecule in this chapter, can be folded into a compact spherical shape during encoding, in order to ensure a high density [146]. Einstein-

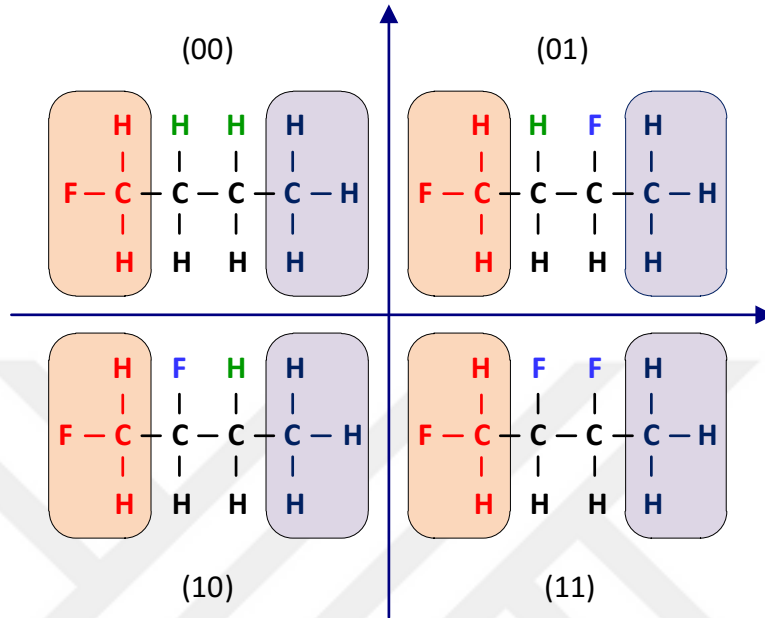


Figure 8.1: Illustration of Quadruple MoSK modulation, for molecule size $n = 2$.

Stokes equation governs the diffusion coefficient for spherical particles through an aqueous medium as [149]

$$D = \frac{k_B T}{6\pi\eta r}, \quad (8.1)$$

where k_B is the Boltzmann's constant, T is the absolute temperature, η is viscosity of the medium and r is the radius of the nanomachine.

Molecule Shift Keying (MoSK), an MC modulation scheme [131], can be exploited for the encoding of information on messenger molecules. This scheme encompasses 2^n distinct molecules to represent n bits of information. To send a desired sequence, corresponding molecule of this sequence is released to the aqueous medium. An example diagram for $n = 2$ bits Quadruple MoSK is illustrated in Fig. 8.1, with respective fluorinated polyethylene molecules [131]. In [131], authors considered a messenger based MC channel for only $n = 1$ and $n = 2$ cases. In this chapter, we generalize the case for different n .

8.3.2 Transmission of Polyethylene Messenger Molecules

Both transmitter and receiver nanomachines are assumed to be spherical. Accordingly, messenger molecules are uniformly discharged from the spherical boundary of the transmitter nanomachine to the aqueous medium, as shown in Fig. 8.2. If the nanomachine is a living cell, this emission boundary is the actual cell membrane.

As per the Brownian Motion propagation model, the probability that a single messenger molecule reaches the receiver in a given time is low. To that end, the transmitter nanomachine emits multiple copies of the encoded molecule, namely N_M such copies. This redundancy raises the reception probability for the envisaged MC channel. As a result, the transmission rate for this MC channel can be calculated as

$$R^{(TX)} = \frac{n}{T_M}, \quad (8.2)$$

where T_M is the period required to prepare N_M copies of the molecule. This period is required as the transmitter nanomachine can not prepare the molecules in advance and can not store molecules for future use.

It is assumed that molecule emission is instantaneous, thereby sharply increasing the concentration of the messenger molecules in the aqueous medium in a very short period. Such phenomena is observed in the nature, as in the form of instantaneous emission of pheromones in animals under alarming situations [150].

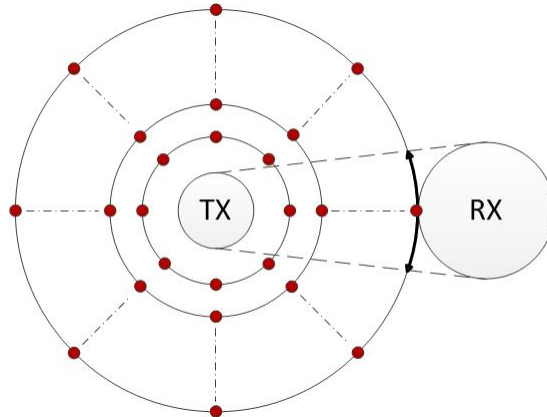


Figure 8.2: Illustration of transmitter and receiver nanomachines, messenger molecule emission and reception.

8.3.3 Propagation of Polyethylene Messenger Molecules

It is assumed that released messenger molecules propagate in the aqueous medium through diffusion, following a Brownian Motion. Inherently, total concentration of messenger molecules are insignificant within entirety of all molecules in the aqueous medium. Therefore, the molecule-to-molecule interactions among messengers can be neglected and the propagation of each messenger molecule can be assumed independent. Hence, the messenger molecule concentration in time and space, based on instantaneous emission assumption in Section 8.3.2, can be calculated through by Fick's law of diffusion [133] as below

$$U(r, t) = \frac{N_M}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(\frac{-r^2}{4Dt}\right), \quad (8.3)$$

where D is the coefficient of diffusion and r represents the distance with respect to the emission boundary.

Denoting the distance between the emission boundary of the transmitter nanomachine and the nanoreceptors of the receiver nanomachine as d , then the propagation delay t of a single molecule to cover this distance is a random variable, which is distributed with an inverse Gaussian probability density function [129] as seen below,

$$f(t) = \frac{d}{\sqrt{4\pi Dt^3}} \exp\left(\frac{-d^2}{4Dt}\right), \quad t \geq 0. \quad (8.4)$$

8.3.4 Reception of Polyethylene Messenger Molecules

As a messenger molecule completes its propagation and touches the boundary of the receiver nanomachine, the nano-receptors bind to it instantaneously. The captured molecules are not re-emitted to the medium after decoding is completed.

Since the messenger molecules randomly propagate in the aqueous medium in all directions, not all the emitted molecules reach the receiver nanomachine. Besides, the messenger molecules can be degraded by the aqueous medium. Only a part of them arrive to the boundary of the receiver nanomachine. To that end, the transmitter

releases N_M copies of the messenger molecule, to increase reliability by redundancy. The booster effect of redundant molecules on channel capacity in MC, where the information is encoded in the concentration of the molecules is illustrated in [129]. As per the adopted system model, reception of only one of the N_M copies of the messenger molecule would suffice to correctly decode the intended message. To that end, a mechanism is needed to identify the redundant copies that arrive to the receiver, after the reception of the initial molecule of that kind. Additionally, messenger molecules from the previous messages may arrive to the receiver while current message is a completely different one, thereby causing inter-symbol interference of some sort. To overcome these issues, additional identifiers on the messenger molecule, right after its distinctive head CH_3 , can be added. This way, the receiver would be able to identify the first messenger molecule of its kind, skipping to decode other copies belonging to the same message.

8.3.5 Decoding of Polyethylene Messenger Molecules

The decoding process takes place after the reception of the incoming messenger molecules. Thanks to the distinctive head and tail, as well as the fact that only one side of the messenger molecule being encoded, the receiver nanomachine easily identifies from which end of the molecule to start decoding and which side should be actually decoded. Decoding takes place as per demodulation of the messenger molecule according to MoSK rules described earlier in Section 8.3.2. An exemplary decoder could be the antennas in the olfactory systems of moths, which are capable to differentiate among thousands of varying pheromone molecules [152]. Much simpler alternative decoder is one that decodes the messenger molecule one bit at a time. For such a case, a probe of form C_5H_5B or $C_3H_3N_2B$ can detect different interaction energy levels between its boron B atom and the coded H and F atoms [153], thereby revealing whether a 0 or 1 has been encoded.

8.4 Molecular Communications as an Erasure Channel

As per the system model presented in Section 8.3, in this section we analytically characterize the polyethylene messenger-based MC channel.

Fluorinated polyethylene molecules, which are employed as messengers in the system model here, are often biocompatible. However, a detailed examination to reveal potential toxicity of varying polyethylene molecules is necessary [136]. Select types such as Fluosol-DA are approved by Food and Drug Administration, to be employed in artificial blood formation applications [147]. Therefore, it can be considered that the system model described here functions in human body. Fluorinated polyethylene molecules are considered stable [147], however human immune system would take actions to clear them from the blood in time. Related work shows that monocytes or macrophages clear these messenger molecules from the blood [148]. For a messenger molecule to be successfully received, it has to hit the boundary of the receiver nanomachine before being cleared from the blood by such mechanisms. If not, the message carried by the corresponding molecule is erased on the channel.

The phagocytosis of messenger molecules by the immune system during the propagation process constitutes an error source in this channel model. However, the transmitter and receiver nanomachines are assumed to encode and decode the information on the polyethylene molecules without any errors. Therefore, if one the N_M identical copies manages to reach the receiver before being cleared out, the message would be completely decoded without error. This assumption is logical since it is unlikely for the environment to change the composition of a fluorinated polyethylene molecule, as these molecules are chemically inert [147]. On the other hand, if all N_M copies fail to reach the receiver due to randomness in the movement or due to being cleared out by the immune system, this piece of information is completely erased. Consequently, we model the communication in this case as a block erasure channel, as depicted in Fig. 8.3. Accordingly, all of the n bits in are either conveyed together with probability $(1 - \alpha)$ or erased together with probability α .

As shown in our previous work [156], presented in *Chapter 7: Modulation in*

Molecular Communications: A Look on Methodologies, synchronization is an important issue in MC. That being said, system model presented here does not require synchronization, since fluorinated polyethylene molecules bear identifiers right after their distinctive head. Hence, the receiver nanomachine can identify which messages it successfully received which of them have been erased, by tracking the sequence numbers.

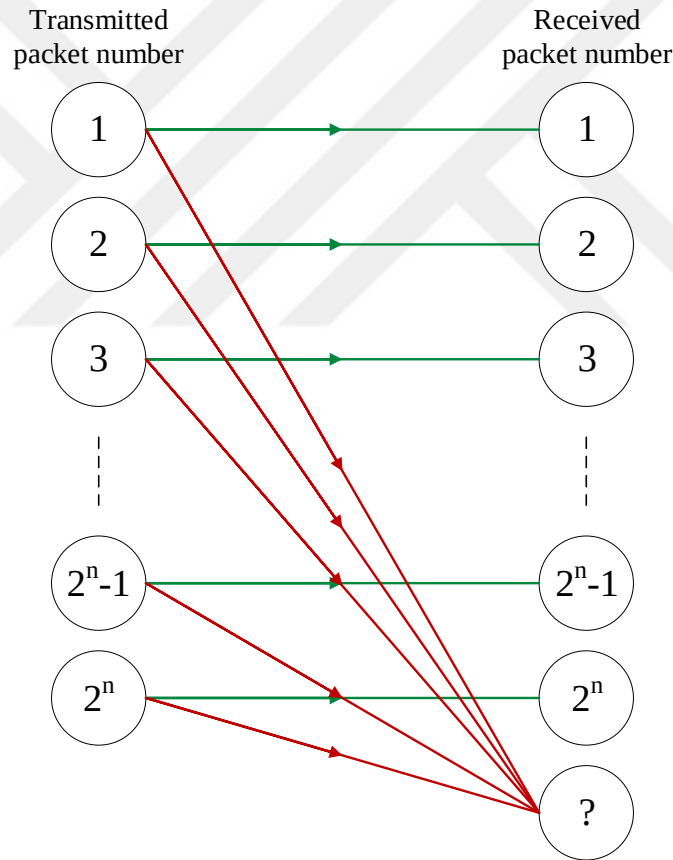


Figure 8.3: Block erasure channel representation.

The fluorinated polyethylene molecules can be assumed to be cleared from the blood at an exponential rate λ [147]. As a result, an emitted messenger molecule would have a lifetime $\tau = 1/\lambda$, after which it can be considered to be cleared from the blood. The block erasure probability, α , can be found by finding the probability that none of the N_M copies of the same messenger molecule arrive to the receiver

before being captured by the immune system agents. To do so, first we need to find the successful delivery of one of the N_M released molecules before lifetime of the molecule, i.e., in a τ amount of time. Let us denote the probability that a molecule reaches the receiver in τ seconds as q . Then,

$$q = \int_0^\tau f(t)dt \quad (8.5)$$

where $f(t)$ is the probability density function of the delay of the molecule to achieve a progress of d meters for the first time after its emission, given in (8.4). Consequently, α can be formulated as

$$\alpha = (1 - q)^{N_M}. \quad (8.6)$$

The capacity of the block erasure channel is then

$$\begin{aligned} C_e &= (1 - \alpha)n \\ &= \left(1 - \left(1 - \int_0^\tau f(t)dt\right)^{N_M}\right) n \text{ bits/channel use,} \end{aligned} \quad (8.7)$$

where $f(t)$ is given in (8.4). As a result, the capacity of this channel depends on

- $\tau = \frac{1}{\lambda}$, average lifetime of the fluorinated polyethylene molecule in the medium (considered as blood here),
- N_M , number of identical copies emitted for each message,
- D , coefficient of diffusion,
- d , distance between the transmitter nanomachine and receiver nanomachine
- n , number of bits encoded on a fluorinated polyethylene molecule.

Clearly, N_M and τ are positively correlated while d is negatively correlated with the erasure channel capacity. However, the impact of n on channel capacity is not

trivial. As n increases, it is possible to convey more bits in a single propagation session. However, as the molecule gets larger, the coefficient of diffusion, D , and hence the ultimate diffusion speed of the molecules decrease significantly. With slower diffusion, probability of a molecule to reach the receiver before τ seconds is lowered. To that end, it is necessary to investigate n value that would optimize the erasure channel capacity.

8.5 Numerical Results

Throughout the numerical analyses, we reveal the impact of various parameters on the achievable capacity of the erasure channel described in Section 8.4.

8.5.1 Impact of Viscosity of the Medium on Performance

With one polyethylene molecule transmission, n bits of information is conveyed to the receiver, since one side of the molecule is reserved for only H atoms to facilitate reception at the receiver. Fig. 8.4 depicts the achievable erasure channel capacity with respect to molecule size n . For this analysis, molecule lifetime τ is fixed to 100 seconds, which is a reasonable period for such molecules that have been employed in artificial blood formation applications [148]. Additionally, distance between transmitter nanomachine and receiver nanomachine is fixed as $50 \mu m$ [136]. Viscosity of human blood is found to be in the range of $\sim 20 \times 10^{-3}$ to $40 \times 10^{-3} kg/m - sec$ [136], hence we depict erasure channel capacities for varying viscosity levels in Fig. 8.4. Accordingly, optimal n value is significantly sensitive to the viscosity of the medium. Optimal n values shift in the opposite direction of the change in medium viscosity.

8.5.2 Impact of Molecule Lifetime τ on Performance

As fluorinated polyethylene molecules follow a Brownian Motion in their propagation, some (or in select cases, all) of the N_M identical copies that are emitted for an intended symbol do not reach the receiver, before they are cleared from the medium. Here we consider potentially longer lifetime values for the messenger molecule and investigate

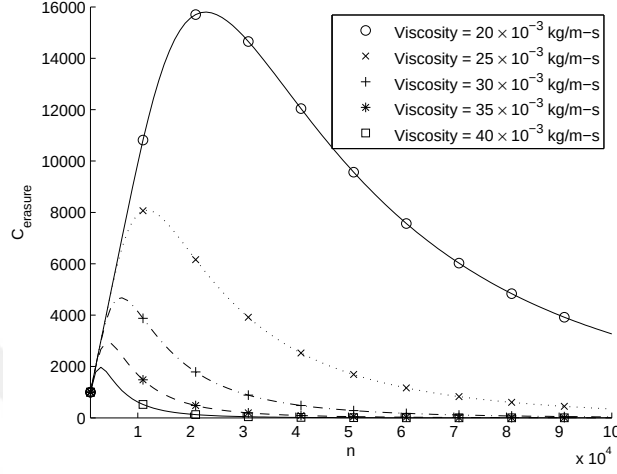


Figure 8.4: Molecular erasure channel capacity with respect to messenger molecule size n , for varying viscosity levels, η .

the impact of this change on the capacity of the block erasure channel. Fig. 8.5 depicts the capacity of the block erasure channel with respect to messenger molecule size n , for varying lifetimes, τ . For this analysis, we fix the viscosity of the medium to $30 \times 10^{-3} \text{ kg/m-sec}$, as the mean viscosity value of blood. The number identical copies that are transmitted for each intended symbol, N_M , is fixed to 1000. Accordingly, as also evident Fig. 8.5, optimal n is very sensitive to molecule lifetime in the aqueous medium, τ . In fact, optimal n is slightly more sensitive to τ than it is to viscosity of the medium. Hence, for network design purposes of this MC system, molecule lifetime would play a very critical role to optimize capacity.

8.5.3 Impact of Number of Redundant Molecules N_M on Performance

As per the system model, some (or in select cases, all) of the N_M identical copies that are emitted for an intended symbol do not reach the receiver, before they are cleared from the medium. However, increasing N_M helps to have at least one molecule with high probability to arrive at the receiver nanomachine. Fig. 8.6 reveals the impact of N_M on both the achievable block erasure channel capacity, as well the optimal messenger molecule size n that would maximize this capacity. Accordingly, both achievable capacity and the optimal molecule size n are very sensitive when N_M is

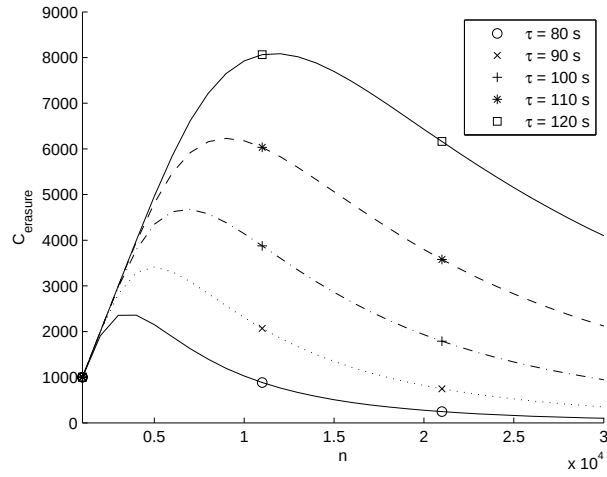


Figure 8.5: Molecular erasure channel capacity with respect to messenger molecule size n , for varying molecule lifetimes, τ .

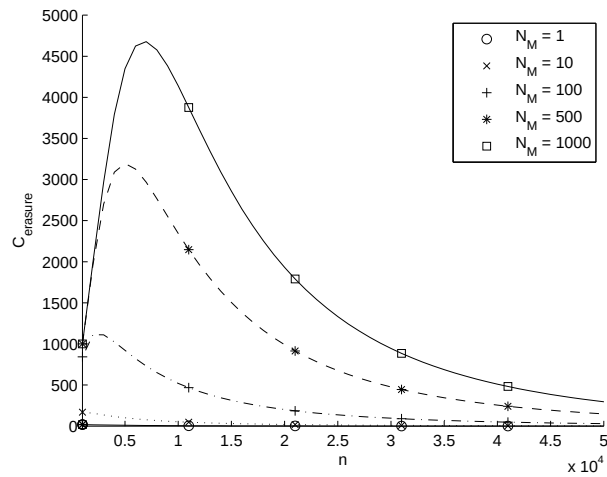


Figure 8.6: Molecular erasure channel capacity with respect to messenger molecule size n , for varying number of identical copies per emission, N_M .

low (in the range of 1-100 copies), even more so compared to both viscosity of the medium η , and molecule lifetime τ . However, as N_M is increased, the sensitivity of both achievable capacity and the optimal molecule size n to N_M reduces significantly.

8.5.4 Closed Form Expression for Optimal Molecule Size

Achievable capacity and molecule size n that maximizes this capacity are negatively correlated with medium viscosity η , and are positively correlated with molecule life-

time τ and number of copies sent for each symbol N_M . Building on the block erasure channel capacity expression in Eq. (8.7), it is not possible to derive an analytical expression for optimal n that maximizes achievable capacity. Therefore, we used MATLAB Optimization and Surface Fitting Toolboxes to obtain a closed-form expression for optimal n through numerical means.

To derive this closed form expression, we first calculate the optimal n with respect to molecule lifetime τ and number of copies for each symbol N_M , and plot the results in 3D. Fitting a surface to these optimal n value, Fig. 8.7 is obtained. Finally, we derive the closed form polynomial expression of third degree for optimal n . Accordingly, this expression is

$$\begin{aligned}
 n_{opt} = & 26,55 - 1.163N_M + 8.532\tau + 0.0023N_M^2 + 0.012N_M\tau \\
 & - 0.3019\tau^2 - 8.758 \times 10^{-7}N_M^3 + 4.362 \times 10^{-5}N_M^2\tau \\
 & - 0.00028N_M\tau^2 + 0.0024\tau^3.
 \end{aligned} \tag{8.8}$$

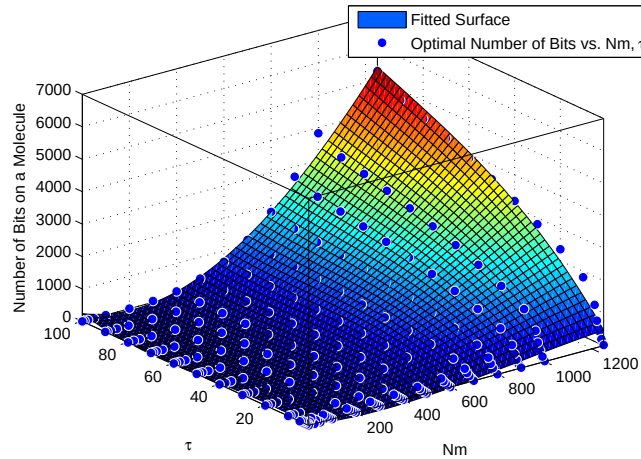


Figure 8.7: The fitted surface for the optimal number of bits, n

It is also important to understand the practicality and feasibility of these optimal n values in potential applications in human body. For instance, messenger molecules

of the type considered in this study, with $n < 5$ typically boil below 37° Celsius. Additionally, molecules with $n > 10^6$ harm the phagocytic functions of agents of the human immune system. Such conditions set natural upper and lower bounds on n . Considering the findings from Sections 8.5.1, 8.5.2 and 8.5.3, for the practical η , τ and N_M values the optimal number of bits n ends up in the interval $0.25 \times 10^4 < n < 1 \times 10^4$. As a matter of fact, this interval is perfectly in conformity with the natural upper and lower bounds on n put forth by the requirements of the human body.

Chapter 9

CONCLUSIONS

In this thesis, we consider next-generation networks that are likely to be the backbones of Internet of Things (IoT) and Internet of Bio-Nano Things (IoBNT). With proliferation in number of devices and dense deployments, such networks are facing challenges in spectrum scarcity, self-sustainability and coordination. To that end, we devise energy and spectrum-efficient techniques for these next-generation networks. Novel techniques brought up in this thesis include but are not limited to energy-efficient transmission range and duration optimization, incorporation of a dedicated radio for efficiency spectrum coordination, self-optimization for capacity while utilizing vacant low THz spectrum, novel energy harvesting methods that enable self-sustainability and high performance modulation schemes in molecular communications with no need for synchronization.

The detailed conclusions reached in each chapter is given in the following:

Part I: Next-Generation Ad Hoc Networks for Internet of Things

Chapter 2: Sensing Coverage and Connectivity in Cognitive Radio Sensor Networks

- Sensing coverage and connectivity are two crucial metrics that assess the QoS in a sensor network. The literature on sensing coverage and connectivity in traditional WSNs is very broad, however these works' applicability to CRSNs had not been assessed so far. Additionally, connectivity in CRAHNs is a nascent field, yet CRSNs need a joint view of sensing coverage and connectivity, as two metrics are impacting together the event detection performance. In this chapter, we first extensively review the literature on sensing coverage and connectivity from WSNs and CRAHNs. Then, building on these reviews, we assess which

models would prove useful to satisfy the unique challenges of CRSN, and provide insights on how to overcome difficulties in different CRSN scenarios. To guide future research, a spatio-temporal distortion based detailed model has been proposed to jointly assess sensing coverage and connectivity in CRSN. The model proposition has been analytically characterized in detail to provide design guidelines in CRSN.

Chapter 3: Energy-Efficient Transmission Range and Duration for CRSN

- In this chapter, for the first time in the literature, we formulate the problem of jointly finding energy-efficient transmission range r_s and transmission duration τ_t for CRSN deployment that would minimize the energy consumed per goodput per meter towards the sink in a greedy forwarding scenario. For that purpose, we first characterize the transmissions in CRSN under DSA; deriving the probability of successful frame transmission at each attempt between one-hop neighbor CRSN nodes, extending our previous work in [51]. Leveraging expected hop progress towards the sink [38] and expected hop distance concepts, we reveal the expected time and energy consumption until successful frame transmission under different spectrum utilization scenario probabilities, building on our work in [52]. Finally, we formulate our optimization problem as a reward function Λ that represents goodput times expected hop progress towards the sink per Joule at each hop. Our findings reveal that
 - Path loss exponent κ is by far the parameter that r_s^* and Λ^* are most sensitive to,
 - Apart from κ , Λ^* is also very sensitive to PU birth rate β , PU death rate α , SU power consumption during spectrum sensing P_s , maximum SU-PU collision probability that a PU can tolerate P_{col} and PU signal variance σ_p^2 ,
 - Apart from κ , r_s^* is also very sensitive to reference SNR needed at SU receiver for demodulation γ_0 and SU power consumption during spectrum sensing P_s ,

- τ_t^* is mostly impacted by κ , PU guardring radius r_p , PU death rate α and PU node density ρ_p .

These relations provide valuable insights for CRSN design prior to deployment.

Chapter 4: Common Control Interface in Cognitive Radio Networks: When to Use It?

- In this chapter, we analyze the impact of incorporating a dedicated Common Control Interface (CCI) in CRAHN and CRSN, in terms of delay, energy and spectrum-efficiency and distributed event sensing performance. Following guidelines were derived for CCI utilization in CRAHN and CRSN:
 - SFT delay can be significantly improved thanks to CCI, especially when $\frac{\tau_f}{E\{\tau_s\}} \geq 10$.
 - With a nearest-neighbor model, in sparse networks, communication links are longer. Assuming constant transmit power, this results in less SNR and more BER at the receiver. Hence, for such networks, increasing node density will often be more cost effective than including dedicated CCI. However, under moderate densities, incorporating CCI brings significant improvements in SFT delay, compared to further increasing node density.
 - The improvement in goodput brought by CCI is not energy efficient, unless PU traffic is significant and CCI consumes less power than the main radio.
 - Incorporating CCI in CRSN solely to increase distributed event sensing performance is not effective, given quite low improvement in Mean Square Error in detection. However, CCI can still offer a valid value proposition in terms of delay, energy-efficiency and goodput/energy improvements for CRSN as well.

Chapter 5: Self-Optimization for Capacity in Low Terahertz Band Indoor Wireless Networks

- In this chapter, we considered an indoor wireless network with Full Dimensional MIMO (FD-MIMO) capabilities operating in the low THz band, with a realistic and highly directional antenna radiation pattern. Building on the previous SINR based water-filling approaches, we proposed a self-optimization protocol where the base station equipped with FD-MIMO capabilities can decide on power allocations P_k depending on the locations of all User Equipments (UEs) to optimize downlink capacity. Protocol is shown to bring 20%+ when the number of UEs are in the range of $10 \leq N \leq 30$, whereas its complexity increases exponentially with N and may limit its utility under run time constraints. Future work can cover extension of AWF to balance network capacity vs. coverage of individual UEs, as well as consideration of priority traffic vs. best effort traffic.

Chapter 6: Harvesting-Throughput Trade-off for Wireless-powered Smart Grid IoT Applications: An Experimental Study

- In this chapter, we propose an Electric Field Energy Harvesting (EFEH) method to enable battery-less execution of sensor-based IoT services for Smart Grid (SG) context. For the first time in the literature, harvestable energy through EFEH is investigated with a transformer room experimental set-up. Our experiments reveal that 40 mJ of energy can be harvested in a period of 900 sec with the proposed EFEH method. Building on this energy profile, we define a throughput objective function θ for a “harvest-then-transmit” type system model, to shed light on the harvesting-throughput trade-off specific to IoT-assisted SG applications. Accordingly, battery-less execution of this system is plausible when harvesting period T_H lies in the interval of 154-390 sec. T_H^* is very robust to changes in conditions such as hop distance d and path loss exponent ν . That being said, T_H^* is mainly driven by the energy profile E_H-T_H , which can be altered by varying the capacitance in the circuitry of the proposed EFEH method. Having a requirement on minimum reporting frequency for the sensor network f_{min} forces the system to end up in a sub-optimal setting in terms of θ . However,

loss in θ compared to its global maximum is not significant, unless it pushes T_{max} very close to lower end of the battery-less operation region.

Part II: Nanonetworks for Internet of Bio-Nano Things

Chapter 7: Modulation in Molecular Communications: A Look on Methodologies

- Molecular Communication (MC) paves the way for the next-generation communication networks at nanoscale. Even though it suffers low data rates and long delays, it carries the potential of directly interacting with nature whose main communication tool consists of molecules. Very diverse biological and biomedical applications can be envisioned from water pollution control to cancer diagnosis/treatment using MC which are delay-tolerant. In this chapter, several modulation techniques for MC addressing different challenges of nanoscale environment are presented. All of these techniques aspire to increase rate and capacity in MC channels, which are usually very low because of the slow and stochastic nature of diffusion. However, while designing modulation schemes for better performances, the peculiarities of MC such as the limited functionality and memory of nanomachines, bio-compatibility, difficulty of synchronization among nanomachines, and mobility should be taken care of. Considering all of these challenges, the most feasible modulation schemes are categorized in four chunks, namely, concentration based modulation, molecule type based modulation, time/order based modulation and mixed modulation. This extensive review led us to come up with a messenger-based MC scheme that employs polyethylenes for modulation, which is presented in detail in Chapter 8.

Chapter 8: Polyethylene: A Promising Messenger for High Performance Molecular Communications

- In this chapter, a novel Molecular Communication (MC) scheme is developed, where multiple-bit carrying fluorinated polyethylene molecules are employed as the messengers among the transmitter and receiver nanomachines. These messenger molecules are assumed to follow random Brownian Motion in their

propagation, and the resultant channel is modeled as a block erasure channel. Main advantages of this scheme are very high capacity per channel use and elimination of the need for synchronization through incorporation of identifier bits on the messenger molecules. Capacity of the envisaged channel is also analytically characterized, and a closed-form expression of the messenger molecule size that maximizes the achievable capacity is derived. Accordingly, when parameters such as redundant molecule emissions per symbol and molecule lifetime can be tuned appropriately, high capacities per channel use can indeed be achieved. Given its high capacity and the biocompatibility of employed messenger molecules, envisaged MC scheme in this chapter is a viable candidate to form nanonetworks that can carry out complex tasks.

BIBLIOGRAPHY

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey," in *Elsevier Computer Networks Journal*, vol. 38, no. 4, pp. 393-422, 2002.
- [2] I. F. Akyildiz, W. Y. Lee, M. C. Vuran and S. Mohanty, "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," in *Elsevier Computer Networks Journal*, vol. 50, no. 13, pp. 2127-2159, 2006.
- [3] H. H. Zhang and J. C. Hou, "Maintaining Sensing Coverage and Connectivity in Large Sensor Networks," in *Journal of Ad Hoc and Sensor Wireless Networks*, vol. 1, no. 1-2, pp. 89-124, 2005.
- [4] R. Cristescu and M. Vetterli, "On the Optimal Density for Real-Time Data Gathering of Spatio-temporal Processes in Sensor Networks," in *International Symposium on Information Processing in Sensor Networks (IPSN)*, pp. 159-164, 2005.
- [5] D. Tian and N. D. Georganas, "A Coverage-Preserving Node Scheduling Scheme for Large Wireless Sensor Networks," in *Proc. of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, pp. 2-41, 2002.
- [6] S. Slijepcevic and M. Potkonjak, "Power Efficient Organization of Wireless Sensor Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 472-476, 2001.
- [7] S. Meguerdichian and M. Potkonjak, "Low Power 0/1 Coverage and Scheduling Techniques in Sensor Networks," in *University of California, Los Angeles, Technical Report 30001*, 2003.
- [8] L. Wang and S. S. Kulkarni, "Sacrificing a Little Coverage Can Substantially Increase Network Lifetime," in *IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, pp. 326-335, 2006.
- [9] Y. Liu and W. Liang, "Approximate Coverage in Wireless Sensor Networks," in *Proc. of Conference on Local Computer Networks (LCN)*, pp. 68-75, 2005.

- [10] P. J. Wan and C. W. Yi, "Coverage by Randomly Deployed Wireless Sensor Networks," in *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2658-2669, 2006.
- [11] B. Liu and D. Towsley, "A Study of the Coverage of Large-scale Sensor Networks," in *Proc. of IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, pp. 475-483, 2004.
- [12] S. Meguerdichian, F. Koushanfar, M. Potkonjak and M. B. Srivastava, "Coverage Problems in Wireless Ad-Hoc Sensor Networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1380-1387, 2001.
- [13] A. Anandkumar, L. Tong and A. Swami, "Optimal Node Density for Detection in Energy-Constrained Random Networks," in *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5232-5245, 2008.
- [14] T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan and K. K. Saluja, "Sensor Deployment Strategy for Target Detection," in *Proc. of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, pp. 42-48, 2002.
- [15] J. Li, K. Li and W. Zhu, "Improving Sensing Coverage of Wireless Sensor Networks by Employing Mobile Robots," in *Proc. of IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 899-903, 2007.
- [16] L. H. Yen, C. W. Yu and Y. M. Cheng, "Expected k-coverage in Wireless Sensor Networks," in *Ad Hoc Networks*, vol. 4, no. 5, pp. 636-650, 2006.
- [17] S. Kumar, T. H. Lai and J. Balogh, "On k-coverage in a Mostly Sleeping Sensor Network," in *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 144-158, 2004.
- [18] Q. Liu, Z. Zhou, C. Yang and Y. Ye, "The Coverage Analysis of Cognitive Radio Network," in *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1-4, 2008.
- [19] C. Bettstetter, "On the connectivity of ad hoc networks," in *The Computer Journal*, vol. 46, no. 4, pp. 432-447, 2004.

- [20] M. Desai and D. Manjunath, "On the Connectivity in Finite Ad Hoc Networks," in *IEEE Communications Letters*, vol. 6, no. 10, pp. 437-439, 2002.
- [21] C. Bettstetter, "On the Connectivity of Wireless Multihop Networks with Homogeneous and Inhomogeneous Range Assignment," in *Proc. of IEEE Vehicular Technology Conference*, vol. 3, pp. 1706-1710, 2002.
- [22] X. Zhou, S. Durrani and H. M. Jones, "Connectivity of Ad Hoc Networks: Is Fading Good or Bad?," in *Proc. of International Conference on Signal Processing and Communication Systems*, pp. 1-5, 2008.
- [23] P. Gupta and P. R. Kumar, "Critical Power for Asymptotic Connectivity," in *Proc. of IEEE Conference on Decision and Control*, vol. 1, pp. 1106-1110, 1998.
- [24] Z. Kong and E. M. Yeh, "Connectivity and Latency in Large-Scale Wireless Networks with Unreliable Links," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 11-15, 2008.
- [25] C. Yin, L. Gao, T. Liu and S. Cui, "Transmission Capacities for Overlaid Wireless Ad Hoc Networks with Outage Constraints," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-5, 2009.
- [26] C. Li and H. Dai, "Transport Capacity and Connectivity of Cognitive Radio Networks with Outage Constraint," in *Proc. of IEEE International Symposium on Information Theory*, pp. 1743-1747, 2010.
- [27] W. Ren, Q. Zhao and A. Swami, "Connectivity of Cognitive Radio Networks: Proximity vs. Opportunity," in *Proc. of ACM Workshop on Cognitive Radio Networks*, pp. 37-42, 2009.
- [28] M. C. Vuran, O. B. Akan and I. F. Akyildiz, "Spatio-temporal Correlation: Theory and Applications for Wireless Sensor Networks," in *Elsevier Computer Networks Journal*, vol. 45, no. 3, pp. 245-259, 2004.
- [29] S. De, "On Hop Count and Euclidean Distance in Greedy Forwarding in Wireless Ad Hoc Networks," in *IEEE Communications Letters*, vol. 9, no. 11, pp. 1000-1002, 2005.

- [30] M. Tanemura, "Statistical Distributions of Poisson Voronoi Cells in Two and Three Dimensions," in *Forma*, 18, pp. 221-247, 2003.
- [31] M. Ozger and O. B. Akan, "Event-Driven Spectrum-Aware Clustering in Cognitive Radio Sensor Networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1483-1491, 2013.
- [32] Haenggi, M., "Delay-Based Connectivity of Wireless Networks," in *Proc. of International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, vol. 5, no. 9, 2010.
- [33] W. Ren, Q. Zhao and A. Swami, "Temporal Traffic Dynamics Improve the Connectivity of Ad Hoc Cognitive Radio Networks," in *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 124-136, 2014.
- [34] O. B. Akan, O. B. Karli and O. Ergul, "Cognitive Radio Sensor Networks," in *IEEE Network*, vol. 23, no.4, pp. 34-40, 2009.
- [35] S.R. Gandham, M. Dawande, R. Prakash and S. Venkatesan, "Energy Efficient Schemes for Wireless Sensor Networks with Multiple Mobile Base Stations," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, pp. 377-381, 2003.
- [36] C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multihop Network," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc)*, pp. 80-91, 2002.
- [37] S. De, "On Hop Count and Euclidean Distance in Greedy Forwarding in Wireless Ad Hoc Networks," in *IEEE Communications Letters*, vol. 9, no. 11, pp. 1000-1002, 2005.
- [38] J. Deng, Y. S. Han, P. N. Chen and P. K. Varshney, "Optimal Transmission Range for Wireless Ad Hoc Networks Based on Energy Efficiency," in *IEEE Transactions on Communications*, vol. 55, no. 7, pp. 1439-1439, 2007.
- [39] P. Chen, B. O'Dea and E. Callaway, "Energy Efficient System Design with Optimum Transmission Range for Wireless Ad Hoc Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 945-952, 2002.

- [40] H. Zhang, Z. Zhang, X. Chen and R. Yin, "Energy Efficient Joint Source and Channel Sensing in Cognitive Radio Sensor Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-6, 2011.
- [41] F. Hu and S. Wang, "Energy Detection for Spectrum Sensing in Cognitive Radio Sensor Network over Fading Channels," in *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1-4, 2009.
- [42] J. A. Han, W. S. Jeon and D. G. Jeong, "Energy-Efficient Channel Management Scheme for Cognitive Radio Sensor Networks," in *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1905-1910, 2011.
- [43] M. C. Oto and O. B. Akan, "Energy-efficient Packet Size Optimization for Cognitive Radio Networks," in *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1544-1553, 2012.
- [44] S. Gao, L. Qian, D. R. Vaman and Q. Qu, "Energy Efficient Adaptive Modulation in Wireless Cognitive Radio Sensor Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 3980-3986, 2007.
- [45] M. H. Rehmani, S. Lohier and A. Rachedi, "Channel Bonding in Cognitive Radio Wireless Sensor Networks," in *Proc. of International Conference on Selected Topics in Mobile and Wireless Networking*, pp. 72-76, 2012.
- [46] Y. Xu, C. Wu, C. He and Lingge Jiang, "A Cluster-based Energy Efficient MAC Protocol for Multi-hop Cognitive Radio Sensor Networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, pp. 537-542, 2012.
- [47] M. Zareei, A. K. M. M. Islam, N. Mansoor, S. Baharun, E. M. Mohamed and S. Sampei, "Cross-layer Mobility-aware MAC Protocol for Cognitive Radio Sensor Network," in *Proc. of International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 1-6, 2015.
- [48] C. Wu, S. Ohzahata and T. Kato, "Dynamic Channel Assignment and Routing for Cognitive Sensor Networks," in *Proc. of International Symposium on Wireless Communication Systems (ISWCS)*, pp. 86-90, 2012.

- [49] A. O. Bicen, V. C. Gungor and O. B. Akan, "Delay-sensitive and Multimedia Communication in Cognitive Radio Sensor Networks," in *Elsevier Ad Hoc Networks*, vol. 10, no. 5, pp. 816-830, 2012.
- [50] A. O. Bicen and O. B. Akan, "Reliability and Congestion Control in Cognitive Radio Sensor Networks," in *Elsevier Ad Hoc Networks*, vol. 9, no. 7, pp. 1154-1164, 2011.
- [51] M. Ozger and O. B. Akan, "On the Utilization of Spectrum Opportunity in Cognitive Radio Networks," in *IEEE Communications Letters*, vol. 20, no. 1, pp. 157-160, 2016.
- [52] A. O. Bicen, E. B. Pehlivanoglu, S. Galmes and O. B. Akan, "Dedicated Radio Utilization for Spectrum Handoff and Efficiency in Cognitive Radio Networks," in *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5251-5259, 2015.
- [53] A. O. Bicen, "Spectrum-Aware and Reliable Distributed Sensing in Cognitive Radio Networks," M.Sc. Thesis, Koc University, Istanbul, Turkey, 2012.
- [54] B. Karp and H.T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," in *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 243-254, 2000.
- [55] J. Li, J. Jannotti, D. S. J. De Couto, D. R. Karger and R. Morris, "A Scalable Location Service for Geographical Ad Hoc Routing," in *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pp. 120-130, 2000.
- [56] W. Y. Lee and I. F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," in *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845-3857, 2008.
- [57] R. Zhang, J. M. Gorce, R. Dong and K. Jaffres-Runser, "Energy efficiency of Opportunistic Routing with Unreliable Links," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6, 2009.

- [58] IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: “Low-Rate Wireless Personal Area Networks (LR-WPANs),” in *IEEE Std 802.15.4-2011 (Revision of IEEE Std 802.15.4-2006)*, pp.1-314, 2011.
- [59] R. Zhang and J. M. Gorce, “Optimal Transmission Range for Minimum Energy Consumption in Wireless Sensor Networks,” in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 757-762, 2008.
- [60] J. M. Gorce, R. Zhang and H. Parvery, “Impact of Radio Unreliability on the Connectivity of Wireless Sensor Networks,” in *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007.
- [61] S. Haykin, “Cognitive Radio: Brain-empowered Wireless Communications,” in *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201-220, 2005.
- [62] B. F. Lo, “A Survey of Common Control Channel Design in Cognitive Radio Networks,” in *Elsevier Physical Communication*, vol. 1, no. 4, pp. 26-39, 2011.
- [63] K. R. Chowdhury and I. F. Akyildiz, “OFDM-based Common Control Channel Design for Cognitive Radio Ad Hoc Networks,” in *IEEE Transactions on Mobile Computing*, vol. 10, no. 2, pp. 228-238, 2011.
- [64] B.F. Lo, I.F. Akyildiz and A.M. Al-Dhelaan, “Efficient Recovery Control Channel Design in Cognitive Radio Ad Hoc Networks,” in *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4513-4526, 2010.
- [65] N. Baldo, A. Asterjadhi and M. Zorzi, “Dynamic Spectrum Access Using a Network Coded Cognitive Control Channel,” in *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2575-2587, 2010.
- [66] L. Le and E. Hossain, “OSA-MAC: A MAC Protocol for Opportunistic spectrum access in cognitive radio networks,” in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1426-1430, 2008.
- [67] Q. Zhao, L. Tong, A. Swami and Y. Chen, “Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework,” in *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589-600, 2007.

- [68] Y. R. Kondareddy and P. Agrawal, "Synchronized MAC Protocol For Multi-hop Cognitive Radio Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 3198-3202, 2008.
- [69] H. Su and X. Zhang, "Cross-layer Based Opportunistic MAC Protocols for QoS Provisionings Over Cognitive Radio Wireless Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 118-129, 2008.
- [70] C. Wang, L. Wang and F. Adachi, "Modeling and Analysis for Reactive-decision Spectrum Handoff in Cognitive Radio Networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2010.
- [71] Y. Song and J. Xie, "Performance Analysis of Spectrum Handoff for Cognitive Radio Ad Hoc Networks without Common Control Channel under Homogeneous Primary Traffic," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 3011-3019, 2011.
- [72] L. Wang, C. Wang and C. Chang, "Modeling and Analysis for Spectrum Handoffs in Cognitive Radio Networks," in *IEEE Transactions on Mobile Computing*, vol. 11, no. 9, pp. 1499-1513, 2012.
- [73] V. K. Tumuluru, P. Wang, D. Niyato and W. Song, "Performance Analysis of Cognitive Radio Spectrum Access with Prioritized Traffic," in *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1895-1906, 2012.
- [74] L. Wang, C. Wang and K. Feng, "A Queueing-theoretical Framework for QoS-enhanced Spectrum Management in Cognitive Radio Network," in *IEEE Wireless Communications*, vol. 18, no. 6, pp. 18-26, 2011.
- [75] Y. Zhang, "Spectrum Handoff in Cognitive Radio Networks: Opportunistic and Negotiated Situations," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-6, 2009.
- [76] D. Willkomm, J. Gross and A. Wolisz, "Reliable Link Maintenance in Cognitive Radio Systems," in *Proc. of First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp.371-378, 2005.

- [77] S. U. Yoon and E. Ekici, "Voluntary Spectrum Handoff: A Novel Approach to Spectrum Management in CRNs," in *Proc. of IEEE International Conference on Communications (ICC)*, pp.1-5, 2010.
- [78] I.F. Akyildiz, B. F. Lo and R. Balakrishnan, "Cooperative Spectrum Sensing in Cognitive Radio Networks: A Survey," in *Elsevier Physical Communication*, vol. 4, no. 1, pp. 40-62, 2011.
- [79] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexer for voice and data," in *IEEE Journal of Selected Areas in Communications*, vol. 4, no. 6, pp. 833-846, 1986.
- [80] Y. Chen, Q. Zhao and A. Swami, "Joint Design and Separation Principle for Opportunistic Spectrum Access," in *Proc. of IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 696-700, 2006.
- [81] C. Chou, S. Shankar, H. Kim and K. G. Shin, "What and How much to Gain by Spectrum Agility?", in *IEEE Journal of Selected Areas in Communications*, vol. 25, no. 3, pp. 576-588, 2007.
- [82] D. Logothetis, K. S. Trivedi and A. Puliafito, "Markov Regenerative Models," in *Proc. of IEEE International Computer Performance and Dependability Symposium*, pp. 134-143, 1995.
- [83] A. O. Bicen, V. C. Gungor and O. B. Akan, "Spectrum-aware and Cognitive Sensor Networks for Smart Grid Applications," in *IEEE Communications Magazine*, vol. 50, no. 5, pp. 158-165, 2012.
- [84] Z. Liang, S. Feng, D. Zhao and X. Shen, "Delay Performance Analysis for Supporting Real-Time Traffic in a Cognitive Radio Sensor Network," in *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 325-335, 2011.
- [85] A. O. Bicen and O. B. Akan, "Reliability and Congestion Control in Cognitive Radio Sensor Networks," in *Elsevier Ad Hoc Networks*, vol. 9, no. 7, pp. 1154-1164, 2011.
- [86] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

- [87] M. Haenggi, "On Distances in Uniformly Random Networks," in *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3584-3586, 2005.
- [88] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [89] M. K. Simon and M. S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. Hoboken, NJ: 1st ed., Wiley, 2000.
- [90] Y. C. Liang, Y. Zeng, E. C. Y. Peh and A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 5330-5335, 2007.
- [91] N. Khalid, N. A. Abbasi and O. B. Akan, "300 GHz Broadband Transceiver Design for Low-THz Band Wireless Communications in Indoor Internet of Things," in *Proc. of IEEE iThings*, 2017.
- [92] Y. Kim et al., "Full dimension MIMO (FD-MIMO): The Next Evolution of MIMO in LTE Systems," in *IEEE Wireless Communications*, vol. 21, no. 2, pp. 26-33, 2014.
- [93] N. Khalid and O. B. Akan, "Wideband THz Communication Channel Measurements for 5G Indoor Wireless Networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-6, 2016.
- [94] N. Khalid, N. A. Abbasi and O. B. Akan, "Capacity and Coverage Analysis for FD-MIMO based THz Band 5G Indoor Internet of Things," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2017.
- [95] S. Kim and A. G. Zajic, "Statistical Characterization of 300-GHz Propagation on a Desktop," in *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3330-3338, 2015.
- [96] S. Priebe, M. Kannicht, M. Jacob and T. Krner, "Ultra Broadband Indoor Channel Measurements and Calibrated Ray Tracing Propagation Modeling at THz Frequencies," in *Journal of Communications and Networks*, vol. 15, no. 6, pp. 547-558, 2013.

- [97] Y. H. Nam et al., "Full-dimension MIMO (FD-MIMO) for Next Generation Cellular Technology," in *IEEE Communications Magazine*, vol. 51, no. 6, pp. 172-179, 2013.
- [98] N. Khalid and O. B. Akan, "Experimental Throughput Analysis of low-THz MIMO Communication Channel in 5G Wireless Networks," in *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 616-619, 2016.
- [99] C. Shannon, "A Mathematical Theory of Communication," in *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, 1948.
- [100] Q. Qi, A. Minturn and Y. Yang, "An Efficient Water-Filling Algorithm for Power Allocation in OFDM-based Cognitive Radio Systems," in *Proc. of International Conference on Systems and Informatics (ICSAI)*, pp. 2069-2073, 2012.
- [101] X. Su, B. Liu, X. Zhu, J. Zeng and C. Xiao, "Low-Complexity Iterative Approximated Water-Filling Based Power Allocation in an Ultra-Dense Network," in *Entropy*, vol. 18, no. 5, 2016.
- [102] J. Kim, H. W. Lee and S. Chong, "Virtual Cell Beamforming in Cooperative Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1126-1138, 2014.
- [103] W. Roh et al., "Millimeter-wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results," in *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106-113, 2014.
- [104] I. Kallfass et al., "Towards MMIC-based 300GHz Indoor Wireless Communication Systems," in *IEICE Transactions on Electronics*, vol. E98.C, no. 12, pp. 1081-1090, 2015.
- [105] L. Atzori, A. Iera and G. Morabito, "The Internet of Things: A Survey," in *Elsevier Computer Networks Journal*, vol. 54, no. 15, pp. 2787-2805, 2010.
- [106] S. Sudevalayam and P. Kulkarni, "Energy Harvesting Sensor Nodes: Survey and Implications," in *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443-461, 2011.

- [107] R. Moghe, Y. Yang, F. Lambert and D. Divan, "A Scoping Study of Electric and Magnetic Field Energy Harvesting for Wireless Sensor Networks in Power System Applications," in *Proc. of IEEE Energy Conversion Congress and Exposition*, pp. 3550-3557, 2009.
- [108] O. Cetinkaya and O. B. Akan, "Electric-field Energy Harvesting Wireless Networks," in *IEEE Wireless Communications*, vol. 24, no. 2, pp. 34-41, 2017.
- [109] O. Cetinkaya and O. B. Akan, "Electric-field Energy Harvesting from Lighting Elements for Battery-less Internet of Things, in *IEEE Access*, vol. 5, pp. 7423-7434, 2017.
- [110] H. Li, J. Xu, R. Zhang and S. Cui, "A General Utility Optimization Framework for Energy-Harvesting-based Wireless Communications," in *IEEE Communications Magazine*, vol. 53, no. 4, pp. 79-85, 2015.
- [111] Y. He, X. Cheng, W. Peng and G. L. Stuber, "A Survey of Energy Harvesting Communications: Models and Offline Optimal Policies," in *IEEE Communications Magazine*, vol. 53, no. 6, pp. 79-85, 2015.
- [112] O. Ozel, K. Tutuncuoglu, S. Ulukus and A. Yener, "Fundamental Limits of Energy Harvesting Communications," in *IEEE Communications Magazine*, vol. 53, no. 4, pp. 126-132, 2015.
- [113] A. Sultan, "Sensing and Transmit Energy Optimization for an Energy Harvesting Cognitive Radio," in *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 500-503, 2012.
- [114] S. Park, S. Lee, B. Kim, D. Hong and J. Lee, "Energy-efficient Opportunistic Spectrum Access in Cognitive Radio Networks with Energy harvesting, in *Proc. of International Conference on Cognitive Radio and Advanced Spectrum Management*, pp. 15, 2011.
- [115] S. Yin, E. Zhang, L. Yin and S. Li, "Optimal Saving-Sensing-Transmitting Structure in Self-powered Cognitive Radio Systems with Wireless Energy Harvesting," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 2807-2811, 2013.

- [116] M. Zheng, C. Xu, W. Liang and H. Yu, "Harvesting-Throughput Tradeoff for RF-powered Underlay Cognitive Radio Networks," in *Electronics Letters*, vol. 52, no. 10, pp. 881-883, 2016.
- [117] S. Yin, E. Zhang, L. Yin and S. Li, "Saving-Sensing-Throughput Tradeoff in Cognitive Radio Systems with Wireless Energy Harvesting," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, pp. 1032-1037, 2013.
- [118] M. Zheng, W. Liang and H. Yu, "Harvesting-Throughput Tradeoff for CDMA-Based Underlay Cognitive Radio Networks With Wireless Energy Harvesting," in *IEEE Systems Journal*, vol. PP, no. 99, pp. 1-4, 2016.
- [119] S. Chamanian *et al.*, "Powering-up Wireless Sensor Nodes Utilizing Rechargeable Batteries and an Electromagnetic Vibration Energy Harvesting System," in *Energies*, vol. 7, no. 10, pp. 6323-6339, 2014.
- [120] I.F. Akyildiz, F. Brunetti and C. Blazquez, "Nanonetworks: A New Communication Paradigm," in *Elsevier Computer Networks Journal*, vol. 52, no. 12, pp. 2260-2279, 2008.
- [121] A. Guney, B. Atakan and O. B. Akan, "Mobile Ad Hoc Nanonetworks with Collision-Based Molecular Communication," in *IEEE Transactions on Mobile Computing*, vol.11, no.3, pp. 353-366, 2012.
- [122] T. Suda, M. Moore, T. Nakano, R. Egashira and A. Enomoto, "Exploratory Research on Molecular Communication Between Nanomachines," in *Proc. of Genetic and Evolutionary Computation Conference (GECCO)*, 2005.
- [123] S. Hiyama and Y. Moritani, "Molecular Communication: Harnessing Biochemical Materials to Engineer Biomimetic Communication Systems," in *Nano Communication Networks*, vol. 1, no. 1, pp. 20-30, 2010.
- [124] S. Hiyama, Y. Moritani, T. Suda, R. Egashira, A. Enomoto, M. Moore and T. Nakano, "Molecular Communication," in *NSTI Nanotech*, vol. 3, pp. 391-394, 2005.
- [125] N. Farsad, W. Guo and A. W. Eckford, "Tabletop Molecular Communication: Text Messages Through Chemical Signals," in *PLoS One*, vol. 8, no. 12, p. e82935, 2013.

- [126] T. Nakano, T. Suda, M. Moore, R. Egashira, A. Enomoto and K. Arima, "Molecular Communication for Nanomachines Using Intercellular Calcium Signaling," in *IEEE Conference on Nanotechnology*, pp. 478-481, 2005.
- [127] M.J. Berridge, "The AM and FM of calcium signalling," in *Nature*, vol. 386, pp. 759-760, 1997.
- [128] M. S. Kuran, T. Tugcu and B. E. Ozerman, "Calcium Signaling: Overview and Research Directions of a Molecular Communication Paradigm," in *IEEE Wireless Communications*, vol. 19, no. 5, pp. 20-27, 2012.
- [129] T. Nakano, Y. Okaie and J. Q. Liu, "Channel Model and Capacity Analysis of Molecular Communication with Brownian Motion," in *IEEE Communications Letters*, vol. 16, no. 6, pp 797-800, 2012.
- [130] N. Garralda, I. Llatser, A. Cabellos-Aparicio and M. Pierobon, "Simulation-based Evaluation of the Diffusion-based Physical Channel in Molecular Nanonetworks," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 443-448, 2011.
- [131] M. S. Kuran, H. B. Yilmaz, T. Tugcu and I. F. Akyildiz, "Modulation Techniques for Communication via Diffusion in Nanonetworks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-5, 2011.
- [132] M. U. Mahfuz, D. Makrakis and H. Mouftah, "Spatiotemporal Distribution and Modulation Schemes for Concentration-Encoded Medium-to-Long Range Molecular Communication," in *25th Biennial Symposium on Communications*, pp. 100-105, 2010.
- [133] L. Parcerisa, I.F. Akyildiz, "Molecular Communication Options for Long Range Nanonetworks," in *Elsevier Computer Networks Journal*, vol. 53, no. 16, pp. 2753-2766, 2009.
- [134] B. D. Unluturk, E. B. Pehlivanoglu and O. B. Akan, "Molecular Channel Model with Multiple Bit Carrying Molecules," in *International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 79-83, 2013.

- [135] N. R. Kim and C. B. Chae, "Novel Modulation Techniques Using Isomers as Messenger Molecules for Molecular Communication via Diffusion," in *IEEE International Conference on Communications (ICC)*, pp. 6146-6150, 2012.
- [136] R.A. Freitas, *Nanomedicine, Volume I: Basic Capabilities*, Landes Bioscience, Georgetown, TX, 1999.
- [137] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos and J. Shuai, "Molecular Communication and Networking: Opportunities and Challenges," in *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 135-148, 2012.
- [138] E. Scemes and G. Christian, "Astrocyte Calcium Waves: What They Are and What They Do," in *Glia*, vol. 54, no. 7, pp. 71625, 2006.
- [139] B. Atakan and O. B. Akan, "An Information Theoretical Approach for Molecular Communication," in *Proc. of Bio-Inspired Models of Network, Information and Computing Systems (BIONETICS)*, pp. 33-40, 2007.
- [140] I. F. Akyildiz, J. M. Jornet and M. Pierobon, "Propagation Models for Nanocommunication Networks," in *Proc. of the Fourth European Conference on Antennas and Propagation*, pp. 1-5, 2010.
- [141] B. Krishnaswamy et al., "Time-Elapse Communication: Bacterial Communication on a Microfluidic Chip," in *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 5139-5151, 2013.
- [142] Y. P. Hsieh, P. J. Shih, Y. C. Lee, P. C. Yeh and K. C. Chen, "An Asynchronous Communication Scheme for Molecular Communication," in *IEEE International Conference on Communications (ICC)*, pp. 6177-6182, 2012.
- [143] B. Atakan, S. Galmes and O. B. Akan, "Nanoscale Communication With Molecular Arrays in Nanonetworks," in *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 149-160, 2012.
- [144] H. Arjmandi, A. Gohari, M. N. Kenari and F. Bateni, "Diffusion-Based Nanonetworking: A New Modulation Technique and Performance Analysis," in *IEEE Communications Letters*, vol. 17, no. 4, pp. 645-648, 2013.

- [145] B.D. Unluturk, "An Information Theoretical Study on Nanoscale Communication Channels with Molecule Diversity," M.Sc. Thesis, Koc University, Istanbul, Turkey, 2013.
- [146] K. E Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, John Wiley & Sons, NY, 1992.
- [147] T. F. Zuck and J. G. Riess, "Current Status of Injectable Oxygen Carriers," in *Critical Reviews in Clinical Laboratory Sciences*, no. 31, pp. 295-324, 1994.
- [148] R. M. Winslow, "Current Status of Oxygen Carriers (Blood Substitutes)," in *Vox Sanguinis*, no. 91, pp. 102-110, 2006.
- [149] S. Lindsay, *Introduction to Nanoscience*, Oxford University Press, Oxford, UK, 2009.
- [150] T. D. Wyatt, *Pheromones and Animal Behaviour: Communication by Smell and Taste*, Cambridge University Press, New York, 2003.
- [151] W.H. Bossert and E. Wilson, "Analysis of Olfactory Communication Among Animals," in *Journal of Theoretical Biology*, no. 5, pp. 443-469, 1963.
- [152] W. Leal, "Pheromone Reception," in *The Chemistry of Pheromones and Other Semiochemicals II*, pp. 1-36, 2005.
- [153] C. W. Bauschlicher Jr., A. Ricca and R. Merkle, "Chemical Storage of Data," in *Nanotechnology*, vol. 8, no. 1, pp. 1-5, 1997.
- [154] J. Billen, "Signal Variety and Communication in Social Insects," in *Proc. of the Netherlands Entomological Society Meeting*, no. 17, pp. 9-25, 2006.
- [155] M.N. Khormuji, "On the Capacity of Molecular Communication over the AIGN Channel," in *Proc. of 45th Annual Conference on Information Sciences and Systems (CISS)*, pp.1-4, 2011.
- [156] E.B. Pehlivanoglu, B.D. Unluturk and O.B. Akan, "Modulation in Molecular Communications: A Look on Methodologies," in *Modeling, Methodologies and Tools for Molecular and Nano-scale Communications*, Modeling and Optimization in Science and Technologies, vol 9. Springer, 2017.